Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

A Review of Personality Trait Recognition with Deep Learning Techniques

Nurrul Akma Mahamad Amin[1,*], Nilam Nur Amir Sjarif[1], Siti Sophiayati Yuhaniz[1], Inass Shahadha Hussein[2]

[1] Department of Advanced Informatic, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia
[2] Technical Institute of Baquba, Middle Technical University, Baghdad Governorate, Iraq

ABSTRACT

Deep learning (DL) has proven to be highly successful in various classification tasks using extracted data from images, text, or sound. DL also enables computers to understand and interpret massive amounts of data, leading to applications such as self-driving cars, medical imaging analysis, and video surveillance. The maturity of deep learning techniques has also helped solve tasks in the field of affective computing, allowing computers to understand, interpret, and respond to human emotions. This has paved the way for the development of automated personality recognition, in which computers can recognize human personality traits using video analysis. A variety of deep learning techniques have been developed to learn and extract meaningful patterns and representations from audio-visual data for automatic personality trait recognition (PTR). This study aimed to explore deep learning techniques that have been used and modified in previous studies for PTR. Initially, this paper presents detailed explanations of the general process of personality trait recognition and the data modalities used. Next, this paper discusses the latest deep learning techniques that have been modified to solve personality recognition tasks. Based on the review of previous studies, the development of a PTR model using deep learning techniques combined with audio-visual modalities yielded impressive prediction results.

Keywords:
Automatic personality recognition; Personality computing; Big five; Deep learning; Job interview video

1. Introduction

In psychology study, the term “personality” refers to the distinct and permanent patterns of thoughts, feelings, and behaviours that characterize an individual [1]. A personality is made up of several elements known as traits, where traits refer to various characteristics of an individual that contribute to constructing a personality. According to Rauthmann *et al.,* [2], personality is associated with both the situations people encounter and how they interpret them. Personality traits also reflect the characteristics of people that determine a person's personality and self-quality where some examples include being outgoing, confident in oneself, talkative, friendly, shy, or lacking in self-confidence. These traits have been used to describe and to get a better understanding of people’s personalities. The way a person acts or responds to some stimuli reflects his or her personality,

* *Corresponding author.*
E-mail address: nurrulakma@graduate.utm.my

https://doi.org/10.37934/araset.57.1.168181

making one individual different from another. In the study conducted by Mehta *et al.,* [3], several personality models have been used for personality detection using deep learning. The personality models include Myers-Briggs Type Indicator (MBTI) [4], The Sixteen Personality Factor Questionnaire (16PF) [5], Eysenck Personality Questionnaire-Revised (EPQ-R) [6], Three Traits Personality Model (PEN) [7] and Big 5 Model [8]. The interpretation of human personality represented by each of these models differs from each other. The Big 5 Model is the most widely used predictors of personality traits measured in psychology as well as affective computing domain [3,9-11]. This model describes human personality with five traits that can be remembered with the acronym OCEAN: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. The Big 5 Model measures personality based on the following explainable scales [12]:

i. <u>Openness:</u> person who is inventive, imaginative, insightful, original, with wide interests, artistic, curious, etc.
ii. <u>Conscientiousness:</u> person who is efficient, responsible, reliable, organized, planful, etc.
iii. <u>Extraversion:</u> person who is outgoing, talkative, active, assertive, energetic, etc.
iv. <u>Agreeableness:</u> person who is trustworthy, straightforward, appreciative, generous, sympathetic, forgiving, etc.
v. <u>Neuroticism:</u> person who is sensitive, worrying, self-pitying, nervous, tense, anxious, unstable, etc.

The Covid-19 pandemic, which struck at the end of 2019, has caused destruction of people worldwide. In addition to threats to public health, economic and social activities are severely affected. Human Resources Management (HRM) is also not exempted from the effects of this pandemic, including the impact on traditional talent acquisition methods and the rise of demand for new forms of employment. There are many challenges faced by HRM due to this pandemic, including staffing [13], communication and collaboration [14], and increasing the complexity of the onboarding process. With social and physical distancing order taking place during the pandemic, normal recruitment, and selection processes, such as face-to-face interviews, could not be held as usual. Thus, organizations were eagerly exploring and adopting AI-based tools such as asynchronous video interviews for employment screening [15,16]. With the availability of advanced technology, software, and computational hardware, an organization can innovate personality trait recognition in the job-screening process. Moreover, tough competition among job seekers forces employers to use various methods to screen and select their employees. The evolution of computer vision technology and maturity of deep learning techniques have led to the development of automatic personality trait recognition. Thus, automatic personality trait recognition can help employers mimic the interviewer's role during screening and automatically recognize personality traits among the interviewees.

Typically, the scheme for development of personality trait recognition system goes through several stages, including data preparation and pre-processing, features extraction and selection, classification modelling and final prediction [17-20]. The main concern in developing an automatic personality recognition model is the process of extracting and selecting relevant features from different modalities to obtain a better final classification score. The challenges arise during development when models try to learn and extract meaningful patterns and representations from audio-visual features. The examples of audio features include pitch, intensity, shimmer, loudness, and simmer, whereas eyes, mouths, and noses are examples of visual features. Audio-visual features are relevant for personality judgements as they provide information about human expression and behaviour. For example, people with higher score of conscientiousness have greater fluctuations in their pupil size while people who blink faster are more neurotic. As for acoustic features, people with low-pitched deep voices are dominant, confident, assertive. On the other hand, voice intensity is positively related to extraversion personality. The potential of each audio-visual feature to automatically describe human personality has prompted research in the development of personality trait recognition. The literature shows that the development of personality trait recognition can be performed in many ways using deep learning techniques. Utilizing various of deep learning techniques, like convolutional neural networks (CNN), residual networks (ResNet), long short-term memory (LSTM), and extreme learning machines (ELM), in a hybrid approach is one of the most well-liked and efficient approaches [10,11,21-24].

This study aimed to explore the trend of learning models in the development of personality trait recognition. In depth, this study is intended to identify more recent approaches and advancements in deep learning techniques that have been applied for model development. Additionally, we sought to discover the modality behaviour and personality model used in employment screening. This study was conducted according to the guidelines of Kitchenham and Charters guidelines of systematic literature review (SLR) method [25]. This study selected previous studies published between 2016 and 2023 through academic databases, including IEEEXplore, Web of Science, Scopus, SpringerLink, ScienceDirect, and Google Scholar, using the following keywords: "automatic personality recognition", "personality computing", "affective computing", "multimodal recognition", "personality detection", "job interview video," and “employment screening”. We identified several key papers that were strongly related to personality recognition during employment screening. This review paper is organized into six sections: Introduction, Previous Studies in Personality Trait Recognition (PTR), General Process of PTR, Data Modality Behaviour, Deep Learning Techniques and Conclusion. The first section explains the background and context of this study. The second section describes the existing research on PTR. The details of personality trait recognition processes are explained in the third section. In the next two sections, the data modality behaviour and deep learning techniques are discussed, respectively. Finally, the findings and future work of this study are discussed in the conclusion section.

2. The Previous Study in Personality Trait Recognition (PTR)

Prior to 2015, research efforts in affective computing focused more on sentiment polarity and emotion recognition than on personality trait recognition. This is because of the difficulty in obtaining adequate datasets coupled with the complexity of personality recognition processes [10]. However, in 2016, the organization of ChaLearn Looking at People Challenge 2016 (ECCV2016) became an ideal beginning for researchers to investigate the problem of personality analysis. This challenge aimed to recognize OCEAN traits in the videos of people speaking in front of the camera. The challenging aspect of personality trait recognition system development is associated with numerous circumstances that may have an impact on the final outcomes, including individual and cultural differences, random noise present in the video input, distinctive facial features, and various articulation styles [20,26]. In addition, recruiters have discovered that manually processing the vast number of video interviews they receive at once is burdensome. Moreover, conventional screening methods that rely on human intervention are the most costly and time-consuming parts of the recruitment process [27,28]. Hemamou *et al.,* [29] developed their own dataset of French asynchronous video interviews, which consists of only 7938 candidates applying for 475 sales positions. On the other hand, Suen *et al.,* [30], developed asynchronous video interview (AVI) software, involved 120 participants in human resources field with video duration of each. Owing to certain restrictions, a self-collected data set was not available for further analysis and research. Based on a literature study, there are not many datasets available for personality traits prediction, so most previous research used the ChaLearn dataset with different modalities and learning algorithm approaches [18,20,21,24,31]. The ChaLearn dataset consists of 10,000 short videos with a duration of 15 s and is labelled with the Big-5 of OCEAN traits.

Additionally, inaccurate predictions of personality traits classification may occur due to various factors, including inadequate data for model development and poor labelling process of the raw data [32]. One of the previous studies which used datasets from limited domain (sales job interview videos) only achieved an 0.6450 of accuracy score for personality recognition based on audio-visual and text [27]. The other previous studies also mentioned that behavioural cues, including verbal and non-verbal cues, play an important role in creating favourable personality impressions [33,34]. Recently, automated behaviour cue analysis using video interviews in recruitment has gained momentum, aligned with the expansion of studies in affective computing. According to Holman and Hughes, the Big-5 OCEAN personality constructs are valid predictors of certain job performance criteria (leadership, interpersonal behaviours, and counterproductive behaviour) [35]. The automation process of personality trait recognition is believed to reduce time and cost while conducting mass recruitment. Compared to automated machine-based evaluators, human recruiters are well trained in evaluating an interviewee’s personality through well-established metrics such as the Big-5 trait taxonomy, also known as OCEAN. Thus, comprehensive training on the development of personality recognition models is important for achieving higher accuracy. In the automation of personality traits, it is crucial to extract and select the most relevant features from the different modalities (visual and audio) to avoid inaccurate predictions. Feature selection and dimensionality reduction are often used to reduce the number of features in a dataset and have been proven to improve classification accuracy [36,37]. With the growth and expansion of technology in computer vision and artificial intelligence (AI), AI-based tools with CNN architectures are becoming the newest approaches to automate personality testing and mitigate issues in traditional practices. However, current research is still focused on how to enable learning algorithms to fully understand personality feature characteristics and improve recognition accuracy. The main challenge in developing an automatic personality recognition model is extracting and selecting relevant features from audio and visual modalities that contribute to a better classification score of the OCEAN traits [10,31,38].

3. General Process of Personality Trait Recognition (PTR)

Personality Trait Recognition is defined as the approach to automatically detecting individual personality traits based on behavioural signals using a computational process. Personality recognition also concerned with the automatic inference of a person's self-reported traits based on behavioural signals or cues, such as facial expressions, body movements, gestures, speech, vocal representation, mobile phone data, digital footprints, wearables, and online games [39]. The recognition of personality traits can also detect personality disorders such as bipolar disorder because certain personality traits are associated with certain disorder symptoms [40]. It aims to adopt computer vision and artificial intelligence technology to model personality trait recognition problems in cognitive science [41]. Recently, personality computing has become an active and significant field of study that focuses on the development of computational algorithms to address automated human personality recognition. Personality trait recognition is concerned with the automated classification of a person's personality traits based on their own generated content such as images, videos, text, and audio [3]. Generally, the primary goal of personality trait recognition is to develop an algorithm that can extract features from user-generated content and make automatic judgments on human personality. Automatic human personality judgements have been used to enhance recognition systems in many applications and practical solutions in social computing. For example, personality recognition is helpful for predicting sentiment polarity and subjectivity on Twitter [42]. Subsequently, a virtual robot called Zara was developed by Fung *et al.,* [43], who used emotion and sentiment recognition to interact with the users. Zara is also able to give a personality analysis based on users' pronouncement of speech.

Personality trait recognition has been exploited in user profiling, product personalization, and business marketing. For example, Micu *et al.,* [44] proposed an AI platform that automatically gathers customer data during store visits to create novel solutions that are customized to specific client demands and desires. Generally, in the development of the personality trait recognition model, there are several main processes that are carried out consecutively, namely data pre-processing, features extraction and selection, classification modelling and final prediction. Figure 1 illustrates a general scheme used for the systematic plan of the personality trait recognition system used in previous studies. The maturity of this scheme is proven based on the implementation carried out by prior researchers [10,18-20,24].

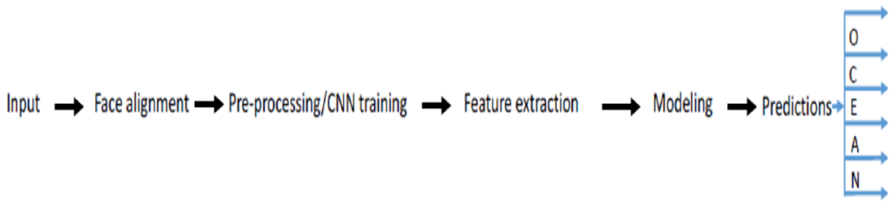

Fig. 1. Illustration of general scheme for personality trait recognition

The initial step of data preprocessing is the process of extracting information related to visual and audio data from a raw data source (video). Visual and audio data were extracted separately for filtering, segmentation, and normalization [45]. Noise reduction, normalization of data, and compression of the retained information are required during preprocessing. The purpose of the preprocessing step is to produce clean facial images and audio signals that can be used in the next step of feature extraction. Features extraction and selection steps are to remove redundancy from data. Specifically, feature extraction is a method that can be applied to extract features from the modality input as representations. Feature selection is related to choosing the most relevant features to improve classification accuracy and reduce computational resources [46]. Following the extraction and selection of relevant features, classification steps utilize the data to determine feature classes based on the characteristics of the features. For example, in the visual modality, the process was to obtain the original face image by detecting facial landmarks and cropping the facial image to train the classifier. An illustration of facial landmark detection is shown in Figure 2. As discussed in the next section, several deep learning modelling techniques can be used. The final step of the personality trait recognition model is to classify subjects into personality traits classes or category which are openness, conscientiousness, extraversion, agreeableness, and neuroticism.

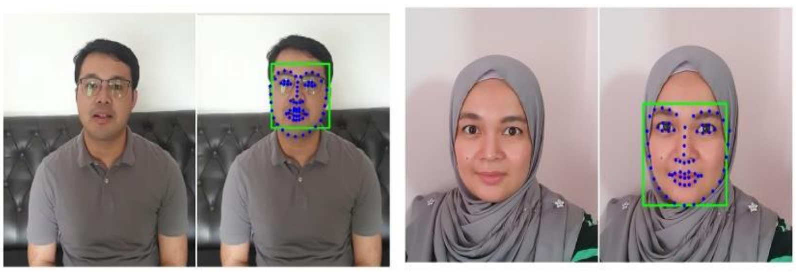

Fig. 2. Illustration for facial landmarks detection

4. Data Modality Behaviour

In nutshell, the development of automatic personality trait recognition incorporates deep learning techniques and the theory of personality models from psychology field. The behaviour of the data used as input in model building is a critical fundamental issue in deep learning that requires attention. This is because deep learning requires specific inputs that instruct them on how to recognize objects to build an accurate prediction model. Based on the type of input data used for model development in earlier studies, it is either a single modality or multiple modalities [47]. Specifically, a single modality involves a single visual, audio, or text feature and multimodal personality trait recognition, combining more than one data modality, such as audio-visual or audio-visual-text features. The primary goal of an automatic personality trait recognition model is to understand and predict individual personalities. One of the most recent challenges in this field is to automatically identify the five personality traits (OCEAN) from recordings of speakers speaking in front of cameras [10,19,24,30,48]. Table 1 (visual), Table 2 (visual and audio) and Table 3 (visual, audio and text) show a summary of data modality behaviour used in existing studies related to personality trait recognition.

According to Table 1, visual or image-based inputs are frequently used in the single-modality prediction of automated personality traits detection. A video comprises a series of consecutive moving images called frames. In general, one second of a video may contain approximately 24 frames per second. Each frame consists of still images and temporal information as inputs for the development of the model. Because deep learning demands a massive amount of input data, each frame may yield significant information. To process the input data and create a prediction model that is more accurate, resources and high computing power are required. Thus, feature extraction is a useful technique for reducing the number of resources required without losing crucial data.

Features extracted from images can be categorized as local or global. Global features are commonly utilized in object detection and image retrieval, where the goal of detection is to determine whether an object is present in a video image. On the other hand, local features are more beneficial in image recognition, where recognition aims to recognize a person or object [49,50]. Most previous studies that have focused on a single modality have used visual-based features for personality trait recognition. Popular approaches, including the use of local facial cues by detecting the face region, face alignment, and eye localization in the images [51,52], and 86 facial landmarks were tracked using OpenCV and Dlib [53]. As for the audio-based modality, feature extraction is often implemented, such as audio or prosodic feature extraction (speaking manners, tones, intensity, and pitch) [19,20,54], and the Python library librosa to extract spectral features Mel Frequency Cepstrum Coefficients (MFCCs), pitch, and tonality [55]. In addition to using visual and audio modalities separately, previous researchers have combined both modalities. As an example, the study extracts facial features using OpenFace to detect 17 face activation units in 15 frames per second plus head position and prosody features using OpenSMILE to detect rhythms, stress, and intonation [32]. The baseline model for automatic personality trait recognition combines audio, scene, and facial features as data input for model development [18]. The scenes and background in the images were treated as global features, whereas facial cues were treated as local features. Generally, global features are utilized for lower-level applications, such as object detection and image classification, while local features are used for higher-level applications, such as object or image recognition. Thus, previous studies related to personality trait recognition have tended to extract local facial cues rather than global features. Other existing studies also proved that a multi-modality that integrates the audio-visual modality results in a higher accuracy prediction compared to a single modality [18,54].

5. Deep Learning Techniques

Deep learning is a subset of machine learning, and machine learning is a subset of artificial intelligence, which is an umbrella term for any computer system that can perform tasks that normally require human intelligence. Machine learning allows systems to learn on their own and improve based on their experiences, without having to be programmed. It is an excellent approach when faced with a task or problem that involves a large amount of data and many variables. In general, machine learning begins by analysing the input data and then looking for patterns within it. It will attempt to predict eventual decisions based on these patterns. In the end, all the decisions made throughout the process are kept, to ensure that learning occurs. This learning was then used to predict a new set of input data without having to go through the entire process. Machine learning approaches are effective in solving many problems where a set of feature selections can easily be extracted by a human expert [56]. However, when it comes to difficult tasks, such as image classification, object detection, or speech recognition, it is extremely difficult to extract features from the input data using traditional machine learning approaches. Thus, deep learning attempts to address this weakness by making a prediction based on the additional learning of hierarchical features from the raw input data [57]. Deep learning is a technique that consists of hierarchical learning algorithms with many layers and is inspired by the structure of the brain. Deep learning has recently become an emerging research topic, especially pertaining to the computer vision domain in the development of numerous applications related to image classification, object detection, automatic identification and recognition, language processing, segmentation, and speech recognition [58-63]. Deep learning mimics the functioning of the human brain in processing data and constructing knowledge patterns for use in prediction or decision making.

Personality trait recognition using deep learning is one of the latest interests in the fields of computer vision, data science, and artificial intelligence (AI). Deep learning provides a system with the ability to learn from the data, identify knowledge patterns, and make decisions with minimal human intervention. Currently, deep learning and machine learning techniques have been used to innovate new digital solutions in many fields, such as medical services (e.g., cancer detection), financial services (e.g., fraud detection, customer spending patterns, and market analysis), image processing, sentiment analysis, email classification, and spam filtering. Besides personality trait recognition, deep learning is also being used in different domains such as in development of wind resource assessment (WRA) model to evaluate the wind power potential [64]. Deep learning is capable of learning from unstructured or unlabelled data, and it can automatically identify the features that are crucial for the classification process. Artificial neural networks (ANNs) are one of the underlying technologies in deep learning, inspired by the structure and function of the human brain [65]. Artificial neurons in deep learning imitate the way biological neurons work in the human brain. ANNs are extensively used in medical science to provide intelligent approaches for disease detection and diagnosis [66]. The architecture of a deep neural network (DNN) is designed to extract complex structures and build representations from its inputs. Additionally, unsupervised DNNs are used to capture unknown patterns in input data and categorize them based on the patterns learned. Supervised DNNs are used for pattern classification. Existing research has shown that DNN is suitable for recognition and classification tasks where DNN can automatically extract features from data sources during the training process. Convolution Neural Network (CNN) is proposed to train the system to identify personality traits from a face in images [67]. The literature has also shown that automatic personality trait recognition can be achieved through computer vision technology, which includes hybrid deep learning approaches. Table 1 until Table 3 show the summary of previous studies based on deep learning techniques in personality trait recognition using different modality.

The summary table consists of information regarding modalities, dataset, algorithm, personality model and accuracy achievement.

Table 1
Summary of Previous PTR Studies with Visual Modality

Ref.	Dataset	Algorithm	Personality Model	Accuracy
[30]	AVI Project	Convolutional Neural Network (CNN)	Big five	0.9536
[31]	VHQ Project	Dynamic Facial Neural Network (DFNN)	Big five	0.8467
[51]	ChaLearn	Support Vector Regression (SVR)	Big five	0.8748
[52]	ChaLearn	Temporal face texture	Big five	0.9116
[68]	SALSA Dataset	Convolutional Neural Network (CNN)	Big five	0.8582
[69]	IMM Face Database	Deep Belief Networks (DBN)	Extroversion, Openness, Agreeable, Rigor	0.8529

Table 2
Summary of Previous PTR Studies with combination of Visual and Audio Modality

Ref.	Dataset	Algorithm	Personality Model	Accuracy
[10]	ChaLearn	Residual Network (ResNet) + Long Short-Term Memory (LSTM)	Big five	0.9207
[11]	ChaLearn	CNN + Bi-LSTM + Transformer network	Big five	0.9167
[18]	ChaLearn	Deep Convolutional Neural Network (DCNN) +Extreme Learning Machines (ELM)	Big five	0.9173
[19]	ChaLearn	Long Short-Term Memory (LSTM)	Big five	0.9120
[20]	ChaLearn	Descriptor Aggregation Network (DANs)	Big five	0.9130
[22]	ChaLearn	Residual Network (ResNet)	Big five	0.9118
[70]	Music Video dataset	Convolutional Neural Network (CNN)	6 emotion signals	0.8856

Table 3
Summary of Previous PTR Studies Based with combination of Visual, Audio and Text Modality

Ref.	Dataset	Algorithm	Personality Model	Accuracy
[17]	Speed Interviews project	Multilayer Perceptron (MLP) Neural Network	Big five	0.8918
[21]	ChaLearn	(ResNet + VGGish + ELM) + LSTM	Big five	0.9180
[24]	ChaLearn	MTCNN + VGGish CNN + CNN	Big five	0.9143
[27]	Video Interviews project	Recurrent Neural Network (RNN)	Hire & not hire	0.6450
[54]	ChaLearn	Convolutional neural networks (CNN) + Classification-regression network (CR-Net)	Big five	0.9188
[71]	ChaLearn	Residual Network (ResNet)	Big five	0.9118
[72]	ChaLearn	Gradient boosting regression	Big five	0.9013
[73]	CMUMOSEI Dataset	Bidirectional Long-Short Term Memory (BLSTM) + Deep Neural Networks (DNNs)	6 emotion signals	0.9060

Zhang *et al.,* [20] modified the traditional CNN architecture by discarding the fully connected layer, which is called a Descriptor Aggregation Network (DAN). The last convolutional layer then aggregated both the average- and max-pooling layers before concatenating them into the final image representation for regression. The DAN model was successful in achieving first place in ChaLearn Looking at People Challenge 2016 (ECCV2016) with an accuracy of 0.913. Some of the benefits of DAN include a reduced model size, reduced final feature dimensions, and faster model training.

Furthermore, according to Ahmad *et al.,* [74], the use of dimensionality reduction techniques in model development has the potential to increase the accuracy of personality traits classification.

Subramaniam *et al.,* [19] proposed two types of deep learning model which are Long Short-Term Memory (LSTM) based model and 3D CNN based model. Both models learned temporal patterns in video sequences and then concatenated the audio and visual features for the final personality predictions. In the 3D CNN model, only visual features (3D face aligned images) were used to define the temporal relationship, whereas in LSTM both audio and visual features were considered. As a result, LSTM performed better with an accuracy of 0.912 and secured second place in ECCV2016. The base line model of personality trait recognition was developed using Deep Convolutional Neural Network (DCNNs) and Extreme Learning Machine (ELM). DCNNs were used for feature extraction. Inspired from ECCV2016, Kaya *et al.,* [18] used scenes feature together with audio and facial images for personality predictions. All the three features are handled in separate channels using Extreme Learning Machine (ELM) classifiers to evaluate each channel. ELM was used due to the learning speed and accuracy of the algorithm. This model achieved accuracy 0.913 and ranks in first place at ChaLearn Job Candidate Screening Coopetition (CVPR2017).

Zhao *et al.,* [11] introduced a new method for multimodal personality trait recognition by using hybrid deep learning includes CNN, bi-directional long short-term memory network (Bi-LSTM), and the Transformer network. All these models captured high-level audio-visual spatio-temporal feature representations for personality trait recognition. Pre-trained CNN model was used to learn high-level global scene features, local face features and audio features. Next, all the extracted features were fed into Bi-LSTM and Transformer network. Finally, a linear regression layer was adopted to predict the personality traits. As results, this model achieved accuracy 0.9167 when implementing on ChaLearn dataset. On the other hand, Sun *et al.,* [10] is not only concerned on face images and audio features but also make used of eye gaze distribution features to determine personality traits [10]. Their study proposed a multimodal attention network with Category-based mean square error (CBMSE) which simultaneously implemented residual network (ResNet) with attention mechanisms and a loss function with a higher penalty for the fuzzy boundary in personality assessment. By leveraging ResNet with attention and LSTM, this model achieved accuracy 0.9207 which outperformed the base line model and other models as well. Performance detail in terms of accuracy achievements in existing personality trait recognition model is shown in Table 1 until Table 3. Figure 3 shows the taxonomy of deep learning techniques which have been applied to personality trait recognition study.

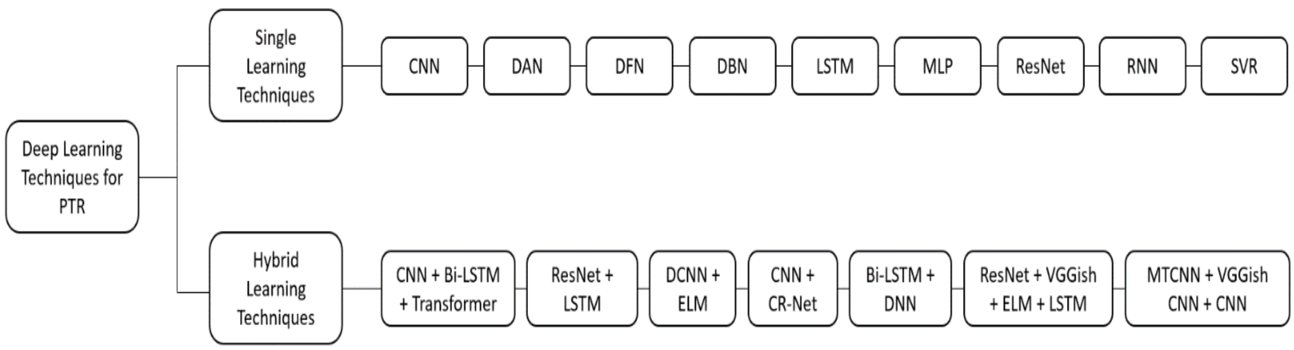

Fig. 3. Taxonomy of existing deep learning techniques for personality trait recognition model

6. Conclusion

Due to the advantages in strong feature learning and extraction ability with high recognition accuracy, deep learning is widely used not only in image recognition but also in video-based analysis. Personality trait recognition from videos is considered a challenging task due to the complexity of detecting human personality in short inputs. Although, the current study on personality trait recognition has achieved a promising accuracy performance by adopting deep learning techniques in model development. As discussed in this paper, there are many types of deep learning techniques which can be applied in solving video-based personality trait recognition problems. However, the issue of accuracy performance degradation lies on the poor process of extracting and selecting relevant features from input sources. The challenge is to enable the model to fully understand personality feature characteristics which mostly carried by audio-visual modalities. Several of CNN-based model have been developed in previous studies with various of enhancement and combination. CNN-based model also often used for single or multi-class classification tasks because it's able to provide highly accurate results. But then, there is still ongoing research on how implementing CNNs for multi-label classification problem such as weather conditions in single image [75] and multiple heart diseases in a single ECG signal [76] and multi-label emotion and sentiment in an image [77]. Current research also concerns neural network layer configuration for multi-label classification, effect of dependencies between label and model accuracy achievement [78,79]. Thus, there is huge potential to enhance and modify CNN-based model to generate more efficient solutions for personality trait recognition problems.

With the help of technological advances in computer vision and artificial intelligence, personality assessment is possible to be done automatically using hybrid of deep neural network. Audio-visual as input signals to the network are the key features of non-verbal communication for human beings that allow them to express emotion, feelings and intentions as well as subtly describe person's personality traits. One of significant findings from literature study is that audio-visual features are not only used for personality trait recognition but widely used in previous study of emotion recognition. Both personality and emotion recognition are central to affective computing where it concerns recognizing human emotions based on facial expressions, voice tone or body language. Based on the accuracy obtained in previous studies, the fusion of audio-visual features modalities produced an outstanding prediction result. Deep learning techniques like CNN-based models have shown promising accuracy results in emotion recognition based on audio-visual data modalities such as facial features and acoustic signals. However, there are still gaps in accuracy performance achieved between emotion and personality recognition based on audio-visual features. Thus, as mentioned that the effort on this study is to explore the trend of learning models in the development of personality trait recognition and to identify audio-visual features that pertain to personality dimensions. The findings from literature study, there is significant correlation between audio-visual features with personality dimensions, extracting more relevant features will help to improve personality trait recognition performance. In future, it is interesting to explore more combination of audio-visual features other than only focus on audio, global scene, and local face marks. Other audio-visual features like eye gaze distribution, upper body movement, head nod, forehead line, pitch, shimmer, intensity, etc. are worth exploring in future research. It is also an important direction to investigate a hybrid of deep learning techniques that can learn and extract as much information from video sequences. In future, it is interesting to explore more on modification and enhancement of deep learning techniques for audio-visual features extraction to improve personality recognition performance.

Acknowledgement
This research was not funded by a grant.

References

[1] Khan, Alam Sher, Hussain Ahmad, Muhammad Zubair Asghar, Furqan Khan Saddozai, Areeba Arif, and Hassan Ali Khalid. "Personality classification from online text using machine learning approach." *International journal of advanced computer science and applications* 11, no. 3 (2020): 460-476. https://doi.org/10.14569/IJACSA.2020.0110358
[2] Rauthmann, John F., Ryne A. Sherman, Christopher S. Nave, and David C. Funder. "Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life." *Journal of Research in Personality* 55 (2015): 98-111. https://doi.org/10.1016/j.jrp.2015.02.003
[3] Mehta, Yash, Navonil Majumder, Alexander Gelbukh, and Erik Cambria. "Recent trends in deep learning based personality detection." *Artificial Intelligence Review* 53, no. 4 (2020): 2313-2339. https://doi.org/10.1007/s10462-019-09770-z
[4] Quenk, Naomi L. *Essentials of Myers-Briggs type indicator assessment*. John Wiley & Sons, 2009.
[5] Boyle, Gregory J., Gerald Matthews, and Donald H. Saklofske, eds. *The SAGE Handbook of Personality Theory and Assessment: Personality Measurement and Testing (Volume 2)*. Vol. 2. Sage, 2008. https://doi.org/10.4135/9781849200479
[6] Eysenck, H. J., and S. B. G. Eysenck. "Eysenck personality questionnaire-revised." (1984). https://doi.org/10.1037/t12641-000
[7] Eysenck, Hans Jurgen, ed. *A model for personality*. Springer Science & Business Media, 2012.
[8] Digman, John M. "Personality structure: Emergence of the five-factor model." *Annual review of psychology* 41, no. 1 (1990): 417-440. https://doi.org/10.1146/annurev.ps.41.020190.002221
[9] Kachur, Alexander, Evgeny Osin, Denis Davydov, Konstantin Shutilov, and Alexey Novokshonov. "Assessing the Big Five personality traits using real-life static facial images." *Scientific Reports* 10, no. 1 (2020): 8487. https://doi.org/10.1038/s41598-020-65358-6
[10] Sun, Xiao, Jie Huang, Shixin Zheng, Xuanheng Rao, and Meng Wang. "Personality assessment based on multimodal attention network learning with category-based mean square error." *IEEE Transactions on Image Processing* 31 (2022): 2162-2174. https://doi.org/10.1109/TIP.2022.3152049
[11] Zhao, Xiaoming, Yuehui Liao, Zhiwei Tang, Yicheng Xu, Xin Tao, Dandan Wang, Guoyu Wang, and Hongsheng Lu. "Integrating audio and visual modalities for multimodal personality trait recognition via hybrid deep learning." *Frontiers in Neuroscience* 16 (2023): 1107284. https://doi.org/10.3389/fnins.2022.1107284
[12] McCrea, Robert R. "An introduction to the five-factor model and its applications." *J Pers* 60 (1992): 175-215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x
[13] Hamouche, Salima. "Human resource management and the COVID-19 crisis: Implications, challenges, opportunities, and future organizational directions." *Journal of Management & Organization* 29, no. 5 (2023): 799-814. https://doi.org/10.1017/jmo.2021.15
[14] Rodeghero, Paige, Thomas Zimmermann, Brian Houck, and Denae Ford. "Please turn your cameras on: Remote onboarding of software developers during a pandemic." In *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pp. 41-50. IEEE, 2021. https://doi.org/10.1109/ICSE-SEIP52600.2021.00013
[15] Lukacik, Eden-Raye, Joshua S. Bourdage, and Nicolas Roulin. "Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews." *Human Resource Management Review* 32, no. 1 (2022): 100789. https://doi.org/10.1016/j.hrmr.2020.100789
[16] Gonçalves, Sónia P., Joana Vieira dos Santos, Isabel S. Silva, Ana Veloso, Catarina Brandão, and Rita Moura. "COVID-19 and people management: The view of human resource managers." *Administrative Sciences* 11, no. 3 (2021): 69. https://doi.org/10.3390/admsci11030069
[17] Gorbova, Jelena, Iiris Lusi, Andre Litvin, and Gholamreza Anbarjafari. "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. 2017. https://doi.org/10.1109/CVPRW.2017.214
[18] Kaya, Heysem, Furkan Gurpinar, and Albert Ali Salah. "Multi-modal score fusion and decision trees for explainable automatic job candidate screening from video cvs." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 1-9. 2017. https://doi.org/10.1109/CVPRW.2017.210
[19] Subramaniam, Arulkumar, Vismay Patel, Ashish Mishra, Prashanth Balasubramanian, and Anurag Mittal. "Bi-modal first impressions recognition using temporally ordered deep audio and stochastic visual features." In *Computer*

Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14, pp. 337-348. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-49409-8_27

[20] Zhang, Chen-Lin, Hao Zhang, Xiu-Shen Wei, and Jianxin Wu. "Deep bimodal regression for apparent personality analysis." In *European conference on computer vision*, pp. 311-324. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-49409-8_25

[21] Aslan, Süleyman, Uğur Güdükbay, and Hamdi Dibeklioğlu. "Multimodal assessment of apparent personality using feature attention and error consistency constraint." *Image and Vision Computing* 110 (2021): 104163. https://doi.org/10.1016/j.imavis.2021.104163

[22] Güçlütürk, Yağmur, Umut Güçlü, Marcel AJ van Gerven, and Rob van Lier. "Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition." In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pp. 349-358. Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-49409-8_28

[23] Gürpinar, Furkan, Heysem Kaya, and Albert Ali Salah. "Multimodal fusion of audio, scene, and face features for first impression estimation." In *2016 23rd International conference on pattern recognition (ICPR)*, pp. 43-48. IEEE, 2016. https://doi.org/10.1109/ICPR.2016.7899605

[24] Suman, Chanchal, Sriparna Saha, Aditya Gupta, Saurabh Kumar Pandey, and Pushpak Bhattacharyya. "A multi-modal personality prediction system." *Knowledge-Based Systems* 236 (2022): 107715. https://doi.org/10.1016/j.knosys.2021.107715

[25] Brereton, Pearl, Barbara A. Kitchenham, David Budgen, Mark Turner, and Mohamed Khalil. "Lessons from applying the systematic literature review process within the software engineering domain." *Journal of systems and software* 80, no. 4 (2007): 571-583. https://doi.org/10.1016/j.jss.2006.07.009

[26] Veinović, Mladen. "Apparent Personality Analysis Based on Aggregation Model." In *Sinteza 2021-International Scientific Conference on Information Technology and Data Related Research*, pp. 220-225. Singidunum University, 2021.

[27] Hemamou, Léo, Ghazi Felhi, Vincent Vandenbussche, Jean-Claude Martin, and Chloé Clavel. "Hirenet: A hierarchical attention model for the automatic analysis of asynchronous video job interviews." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, pp. 573-581. 2019. https://doi.org/10.1609/aaai.v33i01.3301573

[28] Campion, Michael C., Emily D. Campion, and Michael A. Campion. "Using practice employment tests to improve recruitment and personnel selection outcomes for organizations and job seekers." *Journal of applied psychology* 104, no. 9 (2019): 1089. https://doi.org/10.1037/apl0000401

[29] Köchling, Alina, Shirin Riazy, Marius Claus Wehner, and Katharina Simbeck. "Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context." *Business & Information Systems Engineering* 63 (2021): 39-54. https://doi.org/10.1007/s12599-020-00673-w

[30] Suen, Hung-Yue, Kuo-En Hung, and Chien-Liang Lin. "TensorFlow-based automatic personality recognition used in asynchronous video interviews." *IEEE Access* 7 (2019): 61018-61023. https://doi.org/10.1109/ACCESS.2019.2902863

[31] Song, Siyang, Shashank Jaiswal, Enrique Sanchez, Georgios Tzimiropoulos, Linlin Shen, and Michel Valstar. "Self-supervised learning of person-specific facial dynamics for automatic personality recognition." *IEEE Transactions on Affective Computing* 14, no. 1 (2021): 178-195. https://doi.org/10.1109/TAFFC.2021.3064601

[32] Singhania, Abhishek, Abhishek Unnam, and Varun Aggarwal. "Grading video interviews with fairness considerations." *arXiv preprint arXiv:2007.05461* (2020).

[33] Agrawal, Anumeha, Rosa Anil George, and Selvan Sunitha Ravi. "Leveraging multimodal behavioral analytics for automated job interview performance assessment and feedback." *arXiv preprint arXiv:2006.07909* (2020). https://doi.org/10.18653/v1/2020.challengehml-1.6

[34] Jackob, Nikolaus, Thomas Roessing, and Thomas Petersen. "The effects of verbal and nonverbal elements in persuasive communication: Findings from two multi-method experiments." (2011): 245-271. https://doi.org/10.1515/comm.2011.012

[35] Holman, David J., and David J. Hughes. "Transactions between Big-5 personality traits and job characteristics across 20 years." *Journal of Occupational and Organizational Psychology* 94, no. 3 (2021): 762-788. https://doi.org/10.1111/joop.12332

[36] Ayesha, Shaeela, Muhammad Kashif Hanif, and Ramzan Talib. "Overview and comparative study of dimensionality reduction techniques for high dimensional data." *Information Fusion* 59 (2020): 44-58. https://doi.org/10.1016/j.inffus.2020.01.005

[37] Junior, José Jair A. Mendes, Melissa LB Freitas, Hugo V. Siqueira, André E. Lazzaretti, Sergio F. Pichorim, and Sergio L. Stevan Jr. "Feature selection and dimensionality reduction: An extensive comparison in hand gesture

classification by sEMG in eight channels armband approach." *Biomedical Signal Processing and Control* 59 (2020): 101920. https://doi.org/10.1016/j.bspc.2020.101920
[38] Stachl, Clemens, Florian Pargent, Sven Hilbert, Gabriella M. Harari, Ramona Schoedel, Sumer Vaid, Samuel D. Gosling, and Markus Bühner. "Personality research and assessment in the era of machine learning." *European Journal of Personality* 34, no. 5 (2020): 613-631. https://doi.org/10.1002/per.2257
[39] Phan, Le Vy, and John F. Rauthmann. "Personality computing: New frontiers in personality assessment." *Social and personality psychology compass* 15, no. 7 (2021): e12624. https://doi.org/10.1111/spc3.12624
[40] Quilty, Lena Catherine, Martin Sellbom, Jennifer Lee Tackett, and Robert Michael Bagby. "Personality trait predictors of bipolar disorder symptoms." *Psychiatry Research* 169, no. 2 (2009): 159-163. https://doi.org/10.1016/j.psychres.2008.07.004
[41] Vinciarelli, Alessandro, and Gelareh Mohammadi. "A survey of personality computing." *IEEE Transactions on Affective Computing* 5, no. 3 (2014): 273-291. https://doi.org/10.1109/TAFFC.2014.2330816
[42] Celli, Fabio, and Cristina Zaga. "Be conscientious, express your sentiment!." *training* 5747, no. 495 (2013): 5252.
[43] Fung, Pascale, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Dario Bertero, Yan Wan, Ricky Ho Yin Chan, and Chien-Sheng Wu. "Zara: A virtual interactive dialogue system incorporating emotion, sentiment and personality recognition." In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pp. 278-281. 2016.
[44] Micu, Adrian, Alexandru Capatina, Dragos Sebastian Cristea, Dan Munteanu, Angela-Eliza Micu, and Daniela Ancuta Sarpe. "Assessing an on-site customer profiling and hyper-personalization system prototype based on a deep learning approach." *Technological Forecasting and Social Change* 174 (2022): 121289. https://doi.org/10.1016/j.techfore.2021.121289
[45] Berkovsky, Shlomo, Ronnie Taib, Irena Koprinska, Eileen Wang, Yucheng Zeng, Jingjie Li, and Sabina Kleitman. "Detecting personality traits using eye-tracking data." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1-12. 2019. https://doi.org/10.1145/3290605.3300451
[46] Saleem, Faizan, Muhammad Attique Khan, Majed Alhaisoni, Usman Tariq, Ammar Armghan, Fayadh Alenezi, Jung-In Choi, and Seifedine Kadry. "Human gait recognition: A single stream optimal deep learning features fusion." *Sensors* 21, no. 22 (2021): 7584. https://doi.org/10.3390/s21227584
[47] Zhao, Xiaoming, Zhiwei Tang, and Shiqing Zhang. "Deep personality trait recognition: a survey." *Frontiers in Psychology* 13 (2022): 839619. https://doi.org/10.3389/fpsyg.2022.839619
[48] Yang, Karen, S. Mall, and N. Glaser. "Prediction of personality first impressions with deep bimodal LSTM." *Technical report, arXiv* (2017): 1-10.
[49] Krithiga, R., and P. Geetha. "Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review." *Archives of Computational Methods in Engineering* 28, no. 4 (2021): 2607-2619. https://doi.org/10.1007/s11831-020-09470-w
[50] Kuban, Kadhim H., and Wasan M. Jwaid. "A Novel Modification Of Surf Algorithm For Fingerprint Matching." *Journal of Theoretical & Applied Information Technology* 96, no. 6 (2018).
[51] Benlamoudi, Azeddine. "Frame-Difference & Multi-blocks for Job Candidate Screening Competition" *ChaLearn LAP2017* (2017). https://ChaLearnlap.cvc.uab.cat/challenge/23/track/22/result
[52] Eddine Bekhouche, Salah, Fadi Dornaika, Abdelkrim Ouafi, and Abdelmalik Taleb-Ahmed. "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. 2017. https://doi.org/10.1109/CVPRW.2017.211
[53] Suen, Hung-Yue, Kuo-En Hung, and Chien-Liang Lin. "Intelligent video interview agent used to predict communication skill and perceived personality traits." *Human-centric Computing and Information Sciences* 10, no. 1 (2020): 3. https://doi.org/10.1186/s13673-020-0208-3
[54] Li, Yunan, Jun Wan, Qiguang Miao, Sergio Escalera, Huijuan Fang, Huizhou Chen, Xiangda Qi, and Guodong Guo. "Cr-net: A deep classification-regression network for multimodal apparent personality analysis." *International Journal of Computer Vision* 128 (2020): 2763-2780. https://doi.org/10.1007/s11263-020-01309-y
[55] Gorbova, Jelena, Egils Avots, Iiris Lüsi, Mark Fishel, Sergio Escalera, and Gholamreza Anbarjafari. "Integrating vision and language for first-impression personality analysis." *IEEE MultiMedia* 25, no. 2 (2018): 24-33. https://doi.org/10.1109/MMUL.2018.023121162
[56] Zöller, Marc-André, and Marco F. Huber. "Benchmark and survey of automated machine learning frameworks." *Journal of artificial intelligence research* 70 (2021): 409-472. https://doi.org/10.1613/jair.1.11854
[57] Chilimbi, Trishul, Yutaka Suzue, Johnson Apacible, and Karthik Kalyanaraman. "Project adam: Building an efficient and scalable deep learning training system." In *11th USENIX symposium on operating systems design and implementation (OSDI 14)*, pp. 571-582. 2014.
[58] Haris, Malik, and Adam Glowacz. "Road object detection: A comparative study of deep learning-based algorithms." *Electronics* 10, no. 16 (2021): 1932. https://doi.org/10.3390/electronics10161932

[59] Lee, Wookey, Jessica Jiwon Seong, Busra Ozlu, Bong Sup Shim, Azizbek Marakhimov, and Suan Lee. "Biosignal sensors and deep learning-based speech recognition: A review." *Sensors* 21, no. 4 (2021): 1399. https://doi.org/10.3390/s21041399
[60] Ohri, Kriti, and Mukesh Kumar. "Review on self-supervised image recognition using deep neural networks." *Knowledge-Based Systems* 224 (2021): 107090. https://doi.org/10.1016/j.knosys.2021.107090
[61] Qi, Shaohua, Xin Ning, Guowei Yang, Liping Zhang, Peng Long, Weiwei Cai, and Weijun Li. "Review of multi-view 3D object recognition methods based on deep learning." *Displays* 69 (2021): 102053. https://doi.org/10.1016/j.displa.2021.102053
[62] Sethi, Shilpa, Mamta Kathuria, and Trilok Kaushik. "Face mask detection using deep learning: An approach to reduce risk of Coronavirus spread." *Journal of biomedical informatics* 120 (2021): 103848. https://doi.org/10.1016/j.jbi.2021.103848
[63] Khaw Li Wen, Shahrum Shah Abdullah. "MRI Brain Image Classification Using Convolutional Neural Networks And Transfer Learning." *Journal of Advanced Research in Computing and Applications 31*, no. 1 (2023): 20-26.
[64] Lawan, Salisu Muhammad, Wan Azlan Wan Zainal Abidin, and Maitama Yusif Hotoro. "Development of an improved hybrid back propagation ANN for low wind speed prediction and wind energy evaluation." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 75, no. 1 (2020): 112-126. https://doi.org/10.37934/arfmts.75.1.112126
[65] Cichy, Radoslaw M., and Daniel Kaiser. "Deep neural networks as scientific models." *Trends in cognitive sciences* 23, no. 4 (2019): 305-317. https://doi.org/10.1016/j.tics.2019.01.009
[66] Parveen, R., M. Nabi, F. A. Memon, S. Zaman, and M. Ali. "A review and survey of artificial neural network in medical science." *Journal of Advanced Research in Computing and Applications* 3, no. 1 (2016): 7-16.
[67] Ilmini, Kalani, and T. Fernando. "Persons' personality traits recognition using machine learning algorithms and image processing techniques." *Advances in Computer Science: an International Journal* 5, no. 1 (2016): 40-44.
[68] Dotti, Dario, Mirela Popa, and Stylianos Asteriadis. "Being the center of attention: A person-context cnn framework for personality recognition." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, no. 3 (2020): 1-20. https://doi.org/10.1145/3338245
[69] Fu, JinFeng, and Hongli Zhang. "Personality trait detection based on ASM localization and deep learning." *Scientific Programming* 2021, no. 1 (2021): 5675917. https://doi.org/10.1155/2021/5675917
[70] Pandeya, Yagya Raj, and Joonwhoan Lee. "Deep learning-based late fusion of multimodal information for emotion classification of music video." *Multimedia Tools and Applications* 80, no. 2 (2021): 2887-2905. https://doi.org/10.1007/s11042-020-08836-3
[71] Güçlütürk, Yağmur, Umut Güçlü, Xavier Baro, Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Marcel AJ Van Gerven, and Rob Van Lier. "Multimodal first impression analysis with deep residual networks." *IEEE Transactions on Affective Computing* 9, no. 3 (2017): 316-329. https://doi.org/10.1109/TAFFC.2017.2751469
[72] Tanveer, M. I., Naim, I., Hossain, S., Ali, R., & Hoque, M. E. "Interpretable Prediction using Hand-picked Features." *ChaLearn LAP2017.* (2017). https://ChaLearnlap.cvc.uab.cat/challenge/23/track/22/result
[73] Williams, Jennifer, Steven Kleinegesse, Ramona Comanescu, and Oana Radu. "Recognizing emotions in video using multimodal dnn feature fusion." In *Proceedings of Grand Challenge and Workshop on Human Multimodal Language (Challenge-HML)*, pp. 11-19. 2018. https://doi.org/10.18653/v1/W18-3302
[74] Ahmad, Hussain, Muhammad Zubair Asghar, Alam Sher Khan, and Anam Habib. "A systematic literature review of personality trait classification from textual content." *Open Computer Science* 10, no. 1 (2020): 175-193. https://doi.org/10.1515/comp-2020-0188
[75] Zhao, Guozhen, Yan Ge, Biying Shen, Xingjie Wei, and Hao Wang. "Emotion analysis for personality inference from EEG signals." *IEEE transactions on affective computing* 9, no. 3 (2017): 362-371. https://doi.org/10.1109/TAFFC.2017.2786207
[76] Ganeshkumar, M., Vinayakumar Ravi, V. Sowmya, E. A. Gopalakrishnan, and K. P. Soman. "Explainable deep learning-based approach for multilabel classification of electrocardiogram." *IEEE Transactions on Engineering Management* 70, no. 8 (2021): 2787-2799. https://doi.org/10.1109/TEM.2021.3104751
[77] Asakawa, Tetsuya, and Masaki Aono. "Multi-label prediction for visual sentiment analysis using eight different emotions based on psychology." In *Proceedings of the 4th international conference on control and computer vision*, pp. 142-146. 2021. https://doi.org/10.1145/3484274.3484296
[78] Lydia, A. Agnes, and F. Sagayaraj Francis. "Multi-label classification using deep convolutional neural network." In *2020 international conference on innovative trends in information technology (ICITIIT)*, pp. 1-6. IEEE, 2020. https://doi.org/10.1109/ICITIIT49094.2020.9071539
[79] Wang, Jiang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. "Cnn-rnn: A unified framework for multi-label image classification." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2285-2294. 2016. https://doi.org/10.1109/CVPR.2016.251