iPersonalityReader: An Application to Assess the Personality of a Person via Face Image using Pre-trained Convolutional Neural Networks

WMKS Ilmini[1] and TGI Fernando[2]

[1]Faculty of Graduate Studies, University of Sri Jayewardenepura, Sri Lanka
[2]Department of Computer Science, Faculty of Applied Sciences, University of Sri Jayewardenepura, Sri Lanka

{wmksilmini@gmail.com, tgi@sjp.ac.lk}

Abstract. The iPersonalityReader is a web-based app which is capable to assess the personality from the outer appearance of a person. To assess the personality, we proposed a convolutional neural network-based model with transfer learning. VGG-16 pre-trained deep learning model is used as a feature extractor to extract facial features. The extracted features are fed into six parallel "ReLu" layers to predict "Big-Five" personality values including "interview" factor. The proposed model achieved a mean squared error which is below 0.03 for all the personality traits including the "interview" factor.

Keywords: Big-Five personality traits, convolutional neural networks, transfer learning, VGG-16

1 Introduction

Assessing the personality of a person through outer appearance is a popular research area in affective computing. The apparent personality detection (APD) is popular because it is very beneficial in the current technology-based world. For an instance in the education sector identification of the psychological state of a student can improve the teaching-learning process. In the health sector, it can increase the effectiveness of the treatment process by having knowledge of the psychological condition of the patient. In the security sector also having the knowledge on the persons' behaviour is helpful in crime prevention and combating terrorist activities. Moreover, in the animation movie industry and design of social robotics we can apply concepts of apparent personality to have better outcomes.

Psychologists define the personality based on Big Five personality schema [1], which is the commonly accepted schema to measure the personality. The Big Five personality schema includes five traits namely Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN). Each personality trait has its own meaning with subordinate personality types. The Big Five personality schema and its subordinate personality types are defined by John and Srivastava [1]. A summary of each trait and sub-traits are discussed in Table 1.

Table 1. Description of the Big Five Personality Schema

Big Five Personality Trait Name	Meaning	Dimension	Facet (and correlated trait adjectives)
Openness (O)	Describes about adventurous.	Openness vs. closedness to experience	Fantasy, Aesthetics, Ideas, Actions, Excitable Feelings, Unconventional Values
Conscientiousness (C)	Describes diligent and carefulness.	Conscientiousness vs. lack of direction	Competence, Dutifulness Achievement striving, Self-discipline, Deliberation
Extraversion (E)	Describes socialability.	Extraversion vs. introversion	Activity level- active and energetic, Dominance, Sociability, Expressiveness, Positive emotionality
Agreeableness (A)	Describes trustfulness, supportive and nature.	Agreeableness vs. antagonism	Trust, Altruism, Compliance, Modesty, Tendermindedness
Neuroticism (N)	Describes emotional stability.	Neuroticism vs. emotional stability	Anxiety, Angry, Depression, Self-consciousness, Impulsiveness, Vulnerability

Researchers have been done research works on personality detection using different feature sets. They are Handwriting (text) [2, 3], speech (audio) [4], image (face images) [5–8], video (speech and images) [9, 10], and social media[11–13] data. After deep learning is initiated research works in the area of personality detection is also increased. Anyhow, because of the high training cost of the deep learning techniques researchers tend to use transfer learning. According to Lu et al. [14] transfer learning provides an effective and quick solution to the new problem by using the knowledge gained from the previously solved similar problem; i.e. transfer learning acts as a feature extractor for the new problem where it reduces the training cost of the new problem (Figure 1). In the transfer learning, the weights of the newly added classification layers are optimized, while keeping the weights of the original model as it is. Whereas, fine-tuning (optimizing) the weights of the newly added classification layers as well as some or all the original model weights.

The main objective of this research work is to develop a web-based application which is compatible with any electronic device to measure the personality of a person when the face image is given. The proposed methodology involved face detection, transfer learning and convolutional neural network techniques.

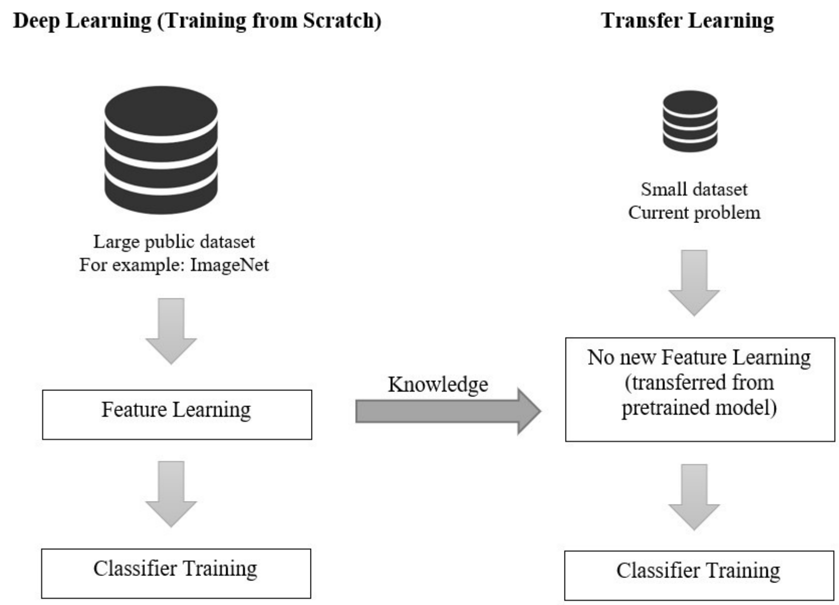

Fig. 1. Process of Transfer Learning

The rest of the paper is organized as follows section 2 discusses the related works, sections 3 describes the methodology followed in the study to develop the CNN model and the web-based app to identify personality traits, section 4 includes discussion and conclusion, and finally section 5 includes future works.

2. Literature Survey

The researchers on the personality detection field used different feature sets with various machine learning techniques. As feature sets handwriting, speech, video, and social media data have been used and before introducing deep learning k-nearest neighbouring (KNN), artificial neural networks (ANN), and support vector machines (SVM) techniques have been mainly used. After the introduction of deep learning, researchers focused on solving personality detection with convolutional neural networks (CNN) and long short-term memory (LSTM). The results of the recent APD researches revealed that the deep learning gave better accuracies than previously used swallow architectures. In deep learning, most of the researches have been used transfer learning to reduce the training cost.

The paper [15] proposed a tri-model stacked CNN architecture to measure the personality from the text, audio and visual data. Three different models have been developed for text, audio and visual streams respectively and connected at the end to predict the personality value. The VGG-16 [16] model is used in this study as a feature extractor to extract features from the video frames. Then three layers added after the VGG-16 model to map personality traits, they are two fully connected layers and one sigmoid layer at the end. The text data is processed and word embeddings are extracted with word2vec [17]. Input to the audio model is raw waveforms. The performance of the developed model is measured using the mean squared error (MSE). According to the findings, Agreeableness is harder to predict, whereas Neuroticism and Extraversion are easier to predict and Conscientiousness and Openness are in between. Although, the researchers concluded that out of the text, audio and visual data, visual data contribute more than the other channels and text is least relevant.

To assess the big five personality values from audio and visual features Subramaniam et al. [18] proposed a bi-model neural network architecture. Two models are developed separately to analyze the audio and visual data, and at the end, results are combined to take the final output. To avoid the bias that might arise from background data the visual data has been pre-processed using “OpenFace” C++ library [19, 20]. The visual analysis has been done using two separate models, a three-dimensional convolutional model and a long short-term memory model, at the end the results are compared from these two models. The outcomes exposed that CNN model took less time than LSTM in training, where the accuracy of LSTM model is better than the convolutional model. They concluded that the LSTM performed well because LSTM is more capable to learn patterns in temporal data.

A new methodology to measure the big five personality values from visual and audio data has been proposed by Gurpinar et al. [21]. In this research, the faces are aligned using supervised decent method [22], and features are analyzed using transfer learning technique with VGG-Face Network [23]. Audio features are analysed using “openSMILE” [24] tool. Then the scene data is analyzed using VGG-VD-19 [25]. At the end outcomes from all the three models are sent through kernel extreme learning machine separately. The results confirm that the audio data does not affect personality recognition. They achieved 91.3% averaged accuracy from the developed model.

Yang and Glaser [10] proposed a bi-model regression LSTM model to predict personality traits. Because the LSTM model is more capable to analyze temporal data and identify relationships. The audio data has been analyzed using the Python library “pyAudioAnalysis” [26]. The visual data is processed using ResNet34 [27]. The features from audio and visual data are combined and sent to LSTM network. They have compared ResNet32 and LSTM models on L1 and L2 loss. They concluded Bi-model LSTM trained on L2 loss attained the best results and ResNet34 overfits the training data.

Most of the APD research have been conducted on the “ChaLearn Looking at People (ChaLearn LAP) First Impression dataset” [28]. These researches used at least two channels such as audio and visual. And at the end, they have concluded that text and audio data does not affect much on personality detection. The current research work is mainly focused on the development of web based personality detector using visual data. Here we gave more focus to visual data rather than audio and text data.

3 Methodology

This section discusses the methodology followed in the current research work to assess the personality traits from the face image. The dataset used in this study to develop a personality assessment model is the First Impressions V2 [28] dataset. The First Impressions V2 dataset contains 10,000 video clips obtained from 3,000 different YouTube videos. This is the largest dataset available with labelled personality values. Each video consists of a person facing to camera and audio is in English. The people are in these videos are differ from nationality, age, gender, and ethnicity. The 10,000 clips are split into 3:1:1 as training, validation and testing data respectively. The videos are labelled with big five personality values which are ranging from 0 to 1. Additionally, there is an extra trait named "interview," which measures the person's capability at an interview.

The steps followed in this section as follows:

a. Prepare a dataset with labelled face images
b. Designing of CNN models with VGG-16 pre-trained deep learning model
c. Design a web-based application using Flask API

3.1 Preparing a dataset with labelled face images

In the preparation of the dataset, video files were split into frames using MoviePy [29] python library and stored them with ground truth values. Then the faces were aligned with the Haar Feature-based Cascade Classifier [30] in order to remove the unwanted background data (Fig. 2).

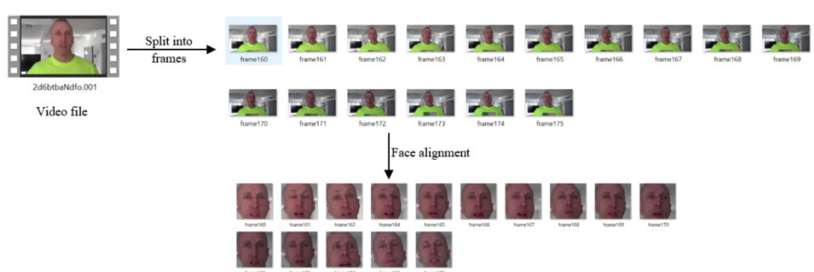

Fig. 2. Process of acquiring images from the videos

The final dataset used in the model development and testing is; training dataset with 85,902 face images, validation dataset with 29,395 face images, and testing dataset with 29,926 face images. The images in the dataset labelled with its ground truth values (Fig. 3).

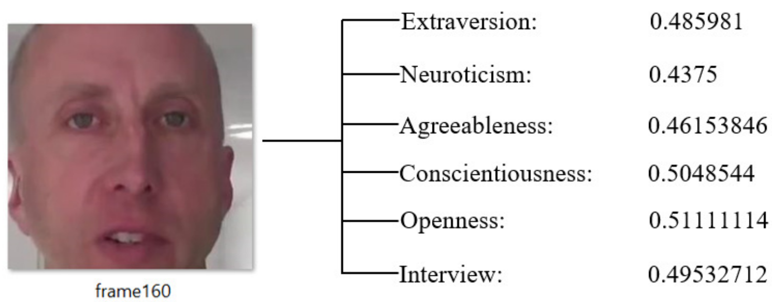

Fig. 3. Sample data with ground truth values

3.2 Designing of CNN models with VGG-16 pre-trained deep learning model

In this research study, we used the transfer learning approach to reduce the training cost using VGG-16 [16] model. VGG-16 is a convolution neural network architecture with very small (3×3) convolution filters. The goal of their work is to improve the accuracy of the network by increasing depth using an architecture with very small convolution filters. In VGG-16 as a preprocessing step, for each image from each pixel subtract the mean RGB value, which is calculated from the training image set. In this architecture, pre-processed images are passed through a stack of convolution layers with small filter size, that is 3×3. The accuracies of the network achieved on ImageNet classification problem are 0.713 and 0.901 as top-1 accuracy and top-5 accuracy respectively. VGG16 network has 138,357,544 number of parameters and depth of the network is 23. Although they have stated that the VGG16 network generalizes well with a wide range of tasks and datasets. VGG16 is outperforming more complex recognition pipelines built less deep image representations.

On top of VGG-16 model, we inserted one ReLu layer followed by six parallel ReLu layers to predict the personality trait values (Fig. 4). We optimized the weights of the newly added final two layers of the network. Six parallel ReLu layers are responsible to produce Big Five personality values and interview trait. The coding is done on Keras [31], an open-source python library with Keras Functional API [32].

The optimizer used in this CNN architecture is Stochastic Gradient Optimization (SGD) with learning rate =0.01, weight decay=1e-6, momentum=0.9, with 'Nesterov' momentum. And the number of epochs is restricted to 2 as increasing the no of epochs has not improved the results. The model is trained and tested on, Intel Core i7-7500CPU with 8GB RAM. The performance of the network is measured using the mean squared error for each trait. Finally, the average mean squared error is calculated using Equation 1.

$$Average\ mean\ squared\ error = \frac{1}{N}\sum_{i=0}^{N} mean\ sqaured\ error_i \qquad \text{(Equation 1)}$$

where i= 1..5 is the number of times the experiment was performed.

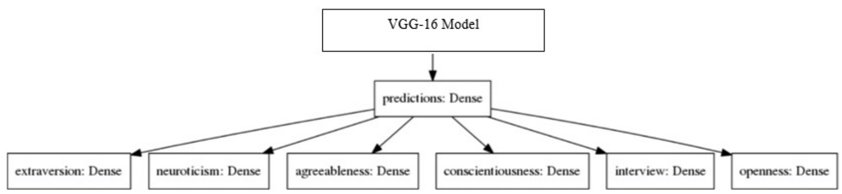

Fig. 4. The architecture of the CNN model

3.3 Design a web-based application using Flask API

The web-based personality detector application: iPeersonalityReader was developed using Flask API [33]. Flask is a web framework which is written in Python. Flask is a micro web framework because Flask keeps the core simple and extensible. Flask does not provide database abstraction layer, form validation and features which can be obtained using existing different libraries. Flask is developed in the way that it is easy to customize to develop machine learning-based web application with Python. The high level architecture of the iPeersonalityReader is shown in Fig. 5. The web app is developed using Flask API and to detect the face OpenCV based module has been used, while personality is defined by deep learning module. The source code of the iPersonalityReader is available at figshare [34].

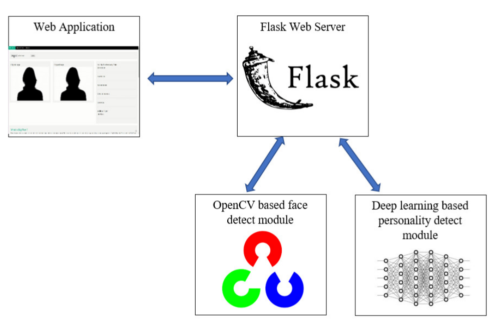

Fig. 5 High level architecture of the system

4 Results

The lowest average MSE in the training, validation and testing is reported by Agreeableness trait followed by Openness trait. Conscientiousness, Extraversion, and Neuroticism achieved approximately equal average mean squared errors (Table 2).

Table 2. Average mean squared error (MSE) and Standard Deviation (Std. Deviation) of the mean squared error

Trait	Measure	Training	Validation	Testing
Openness	Average MSE	.0201200000	.0192200000	.0193549156
	Std. Deviation of the MSE	.00178241409	.00128530152	.00146347698
Conscientiousness	Average MSE	.0222000000	.0224000000	.0213585666
	Std. Deviation of the MSE	.00243207730	.00230976189	.00216412560
Extraversion	Average MSE	.0215200000	.0205400000	.0208343682
	Std. Deviation of the MSE	.00208973683	.00128569048	.00186866291
Agreeableness	Average MSE	**.0177600000**	**.0156400000**	**.0167328826**
	Std. Deviation of the MSE	**.00104307238**	**.00068044103**	**.00069468420**
Neuroticism	Average MSE	.0219000000	.0212200000	.0218756726
	Std. Deviation of the MSE	.00190656760	.00152872496	.00152043401
Interview	Average MSE	.0208800000	.0196400000	.0197245704
	Std. Deviation of MSE	.00216610249	.00174298594	.00180722854

As we noticed VGG-16 is capable to extract features in personality detection image dataset and all average mean squared errors are below 0.03.

The web-based app developed in this study includes an interactive interface where users can input face image and the app automatically detect the face by removing background data from the image. Then the app produces the Big Five personality values for a given face image (Fig. 5). Also, the web-based app includes a description of the Big Five personality schema, to aware the users.

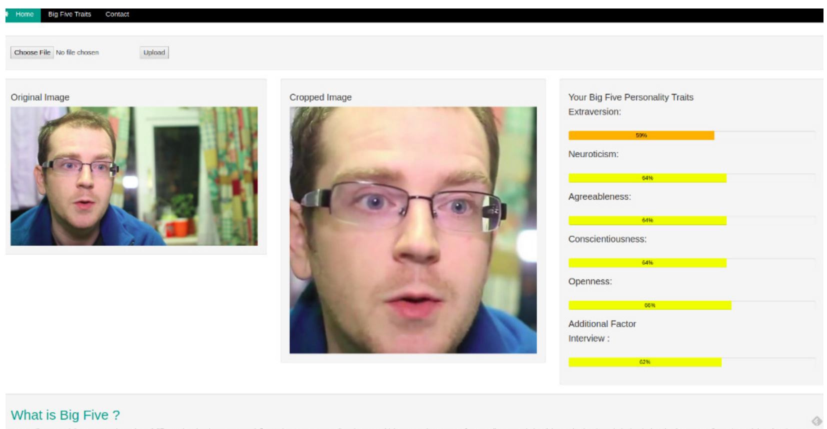

Fig. 6. The caption of the web-based app - personality values

5 Discussion and Conclusion

In this work, we proposed a convolutional neural network-based model with the transfer learning technique to identify apparent personality from the visual features. The results revealed that the transfer learning with VGG-16 model shows good results with selected stochastic gradient decent weight optimizer. Although, in this study, personality is defined using the visual data, it gave acceptable results. This implies visual features are capable to produce the Big-five personality values. Barezi et al. [15] also stated that visual features are more relevant and the contribution of text features are less in the personality detection. Gurpinar et al. [35] also concluded that audio modality does not bring significant improvement in the model they have developed using visual features discussed in the paper [21]. The mean squared errors obtained by each trait is less than 0.03 in the proposed architecture.

The web-based application proposed in this study is capable to predict personality via face image, but the results depend on the quality of the image. Because proposed model's convolutional filters need good quality images to extract features. Although, the First Impressions V2 [28] dataset ground truth values mostly lie in between 3 to 7, no data for extremely low and high personality values, which affects the predictions from the system.

This paper discussed the importance of having personality detection application in the different areas, and how new technologies can be used to develop the application. The outcomes prove that the deep leaning with transfer learning is capable to assess the personality from face images.

6 Future Works

The accuracy of the system can be improved with having more data with personality values lie in the two extreme ends. The web-based app can be improved by having more knowledge of Big-Five personality schema. Also, the web-based app can be improved where different users can gain more knowledge in the respective areas. For an instance for health sector describes more on personality types related to patients' psychology. That is how Big-Five personality schema affects patients' psychology.

References

1. P. John, O., Srivastava, S.: The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives, (1999)
2. Fallah, B., Khotanlou, H.: Detecting features of human personality based on handwriting using learning algorithms. 7
3. Joshi, P., Agarwal, A., Dhavale, A., Suryavanshi, R., Kodolikar, S.: Handwriting Analysis for Detection of Personality Traits using Machine Learning Approach. Int. J. Comput. Appl. 130, 40–45 (2015). https://doi.org/10.5120/ijca2015907189
4. Polzehl, T., Moller, S., Metze, F.: Automatically Assessing Personality from Speech. In: 2010 IEEE Fourth International Conference on Semantic Computing. pp. 134–140. IEEE, Pittsburgh, PA, USA (2010)
5. Liu, L., Preotiuc-Pietro, D., Samani, Z.R., Moghaddam, M.E., Ungar, L.H.: Analyzing Personality through Social Media Profile Picture Choice. In: ICWSM. pp. 211–220 (2016)
6. Martens, L.: Automatic Person and Personality Recognition from Facial Expressions, http://ilk.uvt.nl/downloads/pub/papers/hait/martens2012.pdf, (2012)
7. Kamenskaya, E., Kukharev, G.: Recognition of psychological characteristics from face. Metody Inform. Stosow. 1, 59–73 (2008)
8. Ilmini, K., Fernando, T.: Persons' Personality Traits Recognition using Machine Learning Algorithms and Image Processing Techniques. 5, 5 (2016)
9. Zhang, C.-L., Zhang, H., Wei, X.-S., Wu, J.: Deep Bimodal Regression for Apparent Personality Analysis. In: Computer Vision – ECCV 2016 Workshops. pp. 311–324. Springer, Cham (2016)
10. Yang, K., Glaser, N.: Prediction of Personality First Impressions With Deep Bimodal LSTM. 10 (2017)

11. Farnadi, G., Sitaraman, G., Rohani, M., Kosinski, M., Stillwell, D., Moens, M.-F., Davalos, S., De Cock, M.: How are you doing? Emotions and personality in Facebook. In: Proceedings of the EMPIRE Workshop of the 22nd International Conference on User Modeling, Adaptation and Personalization (UMAP 2014) (2014)
12. Chastagnol, C., Devillers, L.: Personality traits detection using a parallelized modified SFFS algorithm. computing. 15, 16 (2012)
13. Celli, F., Polonio, L.: Relationships between personality and interactions in facebook. Soc. Netw. Recent Trends Emerg. Issues Future Outlook. 41–54 (2013)
14. Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., Zhang, G.: Transfer learning using computational intelligence: A survey. Knowl.-Based Syst. 80, 14–23 (2015). https://doi.org/10.1016/j.knosys.2015.01.010
15. Barezi, E.J., Kampman, O., Bertero, D., Fung, P.: Investigating Audio, Visual, and Text Fusion Methods for End-to-End Automatic Personality Prediction. ArXiv180500705 Cs Stat. (2018)
16. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. ArXiv Prepr. ArXiv14091556. (2014)
17. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed Representations of Words and Phrases and their Compositionality. 9 (2013)
18. Subramaniam, A., Patel, V., Mishra, A., Balasubramanian, P., Mittal, A.: Bi-modal First Impressions Recognition using Temporally Ordered Deep Audio and Stochastic Visual Features. ArXiv161010048 Cs. (2016)
19. Baltrusaitis, T.: OpenFace – a state-of-the art tool intended for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation.: TadasBaltrusaitis/OpenFace. (2019)
20. Baltrusaitis, T., Robinson, P., Morency, L.-P.: OpenFace: An open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1–10. IEEE, Lake Placid, NY, USA (2016)
21. Gurpınar, F., Kaya, H., Salah, A.A.: Combining Deep Facial and Ambient Features for First Impression Estimation. 14 (2016)
22. Xiong, X., De la Torre, F.: Supervised Descent Method and Its Applications to Face Alignment. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp. 532–539. IEEE, Portland, OR, USA (2013)
23. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep Face Recognition. In: Procedings of the British Machine Vision Conference 2015. pp. 41.1-41.12. British Machine Vision Association, Swansea (2015)
24. Eyben, F., Wöllmer, M., Schuller, B.: Opensmile: the munich versatile and fast open-source audio feature extractor. In: Proceedings of the international conference on Multimedia - MM '10. p. 1459. ACM Press, Firenze, Italy (2010)
25. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs. (2014)
26. Giannakopoulos, T.: pyAudioAnalysis: An Open-Source Python Library for Audio Signal Analysis. PLOS ONE. 10, e0144610 (2015). https://doi.org/10.1371/journal.pone.0144610

27. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE, Las Vegas, NV, USA (2016)
28. Ponce-Lopez, V., Chen, B., Oliu, M., Cornearu, C., Guyon, I., Baro, X., Escalante, H.J.: ChaLearn LAP 2016: First Round Challenge on First Impressions - Dataset and Results. 18 (2016)
29. Zulko: moviepy · PyPI. (2019)
30. OpenCV: Face Detection using Haar Cascades, https://docs.opencv.org/3.4.2/d7/d8b/tutorial_py_face_detection.html
31. Home - Keras Documentation, https://keras.io/
32. Chollet, F.: Deep learning with Python. Manning Publications Co, Shelter Island, New York (2018)
33. Ronacher, A.: Welcome to Flask — Flask Documentation (1.1.x), http://flask.palletsprojects.com/en/1.1.x/
34. Ilmini, K.: iPersonalityReader, https://figshare.com/articles/iPersonalityReader/10315742, (2019)
35. Gurpinar, F., Kaya, H., Salah, A.A.: Multimodal fusion of audio, scene, and face features for first impression estimation. Presented at the December (2016)