Multimodal Sentiment Analysis for Personality Prediction

Hans Petter Fauchald Taralrud*, Abdulfatah Abdi Salah*, Ali Shariq Imran* *Member, IEEE*, and Zenun Kastrati†
*Department of Computer Science, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway
†Department of Computer Science and Media Technology, Linnaeus University, 351 95 Växjö, Sweden

***Abstract*—Personality is a set of traits and unique characteristics that give both consistency and individuality to a person's behavior. As personality is accepted as an indicator of job performance, recruiters aim to retrieve these behavior traits in the screening process. One issue is that using personality questionnaires is less favored by applicants and negatively affects the pace of the recruitment process. Therefore, it is desired to infer one's personality traits during an interview. Nowadays, companies tend to use Asynchronous Video Interviews (AVIs), an interview setting where applicants record their answers to pre-defined questions and an interviewer is not present. These recordings are valuable to recruiters since emotions expressed in the AVIs reveal personality insights. With the increasing interest in understanding human behavior, Multimodal Sentiment Analysis (MSA) has emerged as a popular research field. MSA aims to recognize the expressed emotions and sentiment in multimodal data. In this paper, we investigate how MSA can assist personality prediction in AVIs. We propose a novel emotion-to-personality method where dimensions of the Big Five model of personality are predicted based on the emotion distribution in an interview. We collect an AVI dataset in order to test our method. The use of self-assessment i.e. personality questionnaire shows a unique and novel way of validating the personality prediction model on interviews from the AVI dataset.**

***Index Terms*—multimodal sentiment analysis, emotion recognition, personality prediction, multimodal feature extraction, multimodal fusion, deep learning**

I. Introduction

Personality is a set of traits and unique characteristics that give both consistency and individuality to a person's behavior [1]. These sets of traits are associated with important life outcomes and choices. Specifically, it is stated that personality is closely related to job performance, job satisfaction, and tenure intention [2]. When there is consistency between one's personality and career, the work is more enjoyable for the employee and thus valuable to the employer. However, finding the most suitable candidates that fit well with the job description is a challenging task. The Human Resources (HR) team rely on personality questionnaires based on widely accepted personality taxonomies such as the Big Five personality traits (Big Five) model to retrieve the candidates' personality. Using these types of personality tests in the recruitment process is costly and time-consuming to complete. Additionally, it can negatively affect the applicants' perception of the company because personality questionnaires are less favored by candidates than interviews [3]. A good recruitment team should interpret an applicant's behavioral traits during the interview.

As with any other aspect of business, recruitment, and selection depend on speed and accuracy [4]. Recruiters are dependent on finding ways to effectively screen applicants while selecting the most appropriate candidates. With the advancements in Machine Learning (ML), Deep Learning (DL), and Artificial Intelligence (AI), these components have an impact on the HR industry. Utilizing systems that leverage AI capabilities can speed up the recruitment process. Some of these systems include applicant tracking, questionnaires, and Asynchronous Video Interviews (AVIs). Many companies have applied the latter as a part of their recruitment process. AVIs, also known as pre-recorded interviews, on-demand interviews, web interviews, etc., is a setting where applicants record their responses to interview questions [5]. During the interview, candidates express their emotions through verbal and nonverbal communication such as gestures and vocal intonation which can reveal personality insights. Mainly, what distinguishes an AVI from regular interviews is that there is no synchronous back-and-forth communication stream and that the interviewer is not present. The recorded videos are rated at a later time by human assessors, AI systems, or both. Using AVIs as a part of the screening process provides several benefits including the ability to assess more applicants and reduce subjective bias (e.g. attractiveness, visual stigmas) [5].

With the tremendous proliferation of social media and a shift from unimodal to multimodal data such as video clips, Multimodal Sentiment Analysis (MSA) has become a popular field of research [6]. MSA aims to investigate the expressed emotions and sentiment in multimodal data where different modalities such as visual, acoustic, and text are fused together. The process of combining, filtering, and extracting required features from the data collected from different modalities has resulted in several fusing methods to emerge [7].

In this paper, we investigate how MSA can be used to predict behavioral traits in AVIs. State-of-the-art MSA models are employed to retrieve emotions expressed in videos. These emotions can be linked to different behavioral traits. The issue is that there do not exist (to the best of our knowledge) any multimodal datasets with annotated emotions and personality traits. Therefore, we propose a multimodal approach to personality inference by predicting personality traits based on emotions. Lastly, we conduct experiments inviting participants

Corresponding author: H.P.F. Taralrud (email: hptaralr@stud.ntnu.no).

to perform an AVI and a personality test. The results are then compared to conclude how the personality predictor performs in contrast to the observers' own assessment.

Contribution. The major contributions of this article are the following:

- Proposed an approach to predicting personality traits in AVIs based on emotions.
- Validation of detected traits via Big Five questionnaire.
- Establishing a multimodal feature extraction pipeline for video interviews.
- Providing interesting insights into assessing job applicants in today's work environment.

II. Related Work

Multimodal Sentiment Analysis. MSA has been a topic of interest for several years. Researchers have performed sentiment analysis and emotion classification on multimodal data for various application domains such as spoken reviews and vlogs [8], human–machine and human–human interactions [9], and unstructured conversations [10]. Most approaches found in the literature have performed sentiment analysis and emotion recognition using deep learning where data is processed through a neural network. The abundance of literature on the subject cited by Gandhi et al. [11] and others [12] – [14] to present surveys on MSA approaches to analyze multimodal data for polarity assessment, i.e. positive, neutral, negative, and predicting emotions (i.e. happy, sad, frustrated, anger, etc.). Most fusion approaches found in previous work concern early and late fusion methods. For example, the authors in [15] proposed a Tensor Fusion Network that learned inter- and intra-modality end to end. As predicting emotions in multiparty conversations is a popular area of research, Wang et al. [16] established an attention-based fusion framework (Af-CAN) that leverages the conversational information from both target and the other speaker for utterance-level emotion detection. Majumder et al. [17] proposed a hierarchical fusion strategy for single utterance emotion prediction. The authors in [18] investigated how multimodal fusion schemes can extract meaningful information while preserving their mutual independence. Here, a Bi-Bimodal Fusion Network (BBFN) is used to balance the contribution of different modality pairs.

Linking emotions to personality traits. Some efforts have been made into investigating how emotions and personality traits are linked. Zhao et al. [19] analyzed the brainwaves of participants when they were exposed to various emotional video clips. The authors in [20] used three questionnaires to explore to what extent primary emotions act as a foundation for personality and higher-order emotions. Moreover, Li et al. [21] propose a multitask-learning approach that predicts emotions and personality traits simultaneously on unimodal data. Information is shared between the two tasks using Softmax Gate (SoG) and a CNN is used for prediction.

III. Research Method

The framework proposed in this study consists of three parts, as depicted in Figure 1. The first part entails employing a state-of-the-art MSA model Contextualized GNN-based Multimodal Emotion Recognition (COGMEN) [22] trained on the IEMOCAP benchmark for emotion prediction. In addition, a multimodal feature extraction pipeline is developed to extract audio, video, and text features from a video. Based on the output from emotion recognition, we establish a personality prediction method based on previous research in the second part. Lastly, in the third part, we collect a private AVI dataset to test the proposed personality prediction method. Also, participants answer the Big Five personality test in order to validate our prediction model.

A. Data Preprocessing

The data preprocessing step involves data alignment, which divides the data into segments based on each utterance. The segmentation is determined by factors such as speech pauses and question transitions. In this article, each AVI sample is segmented by utterances, with each video segment ranging from 7 to 20 seconds in length.

B. Feature Extraction

After completing the data preprocessing step, the data preparation process is performed for each of the three modalities. In this study, a feature extraction pipeline has been developed to efficiently prepare the data and extract the relevant information from the different modalities. The pipeline consists of three primary sections, namely the acoustic, visual, and textual feature extractions as shown in the right side of Fig. 1.

1) Acoustic feature: OpenSmile [30] is used to extract the low-level descriptors (LLDs) of each segment. To ensure consistency in the data preparation, the approach described in [31] is utilized, which applies Z-score standardization to the LLDs for voice normalization. The normalized features are fed to a convolutional neural network (CNN) model with a fully connected (FC) layer containing 100 neurons. The 100-dimensional feature vector extracted from the FC layer is used as the audio feature.

The CNN model is trained on the CMU-MOSEI [32] dataset and is solely used for audio feature extraction. Once the audio features are extracted, they are combined with visual and textual features and fed into the MSA model.

2) Text feature: Each segment is first transcribed using the Google Cloud Speech API. This creates a text transcription of each video segment. After the segments are transcribed, they are directly fed into the pre-trained sentence-BERT (sBERT) model [39] to extract the text feature. By feeding each transcribed text segment into the sBERT model, the high-dimensional text feature vector that encodes the semantic meaning of the text is extracted.

3) Visual feature: Extracting the facial feature involves multiple steps. First, the video segment is converted into a series of frames. Next, the Multi-task Cascaded Convolutional Network (MTCNN) model [27] is used to identify the face

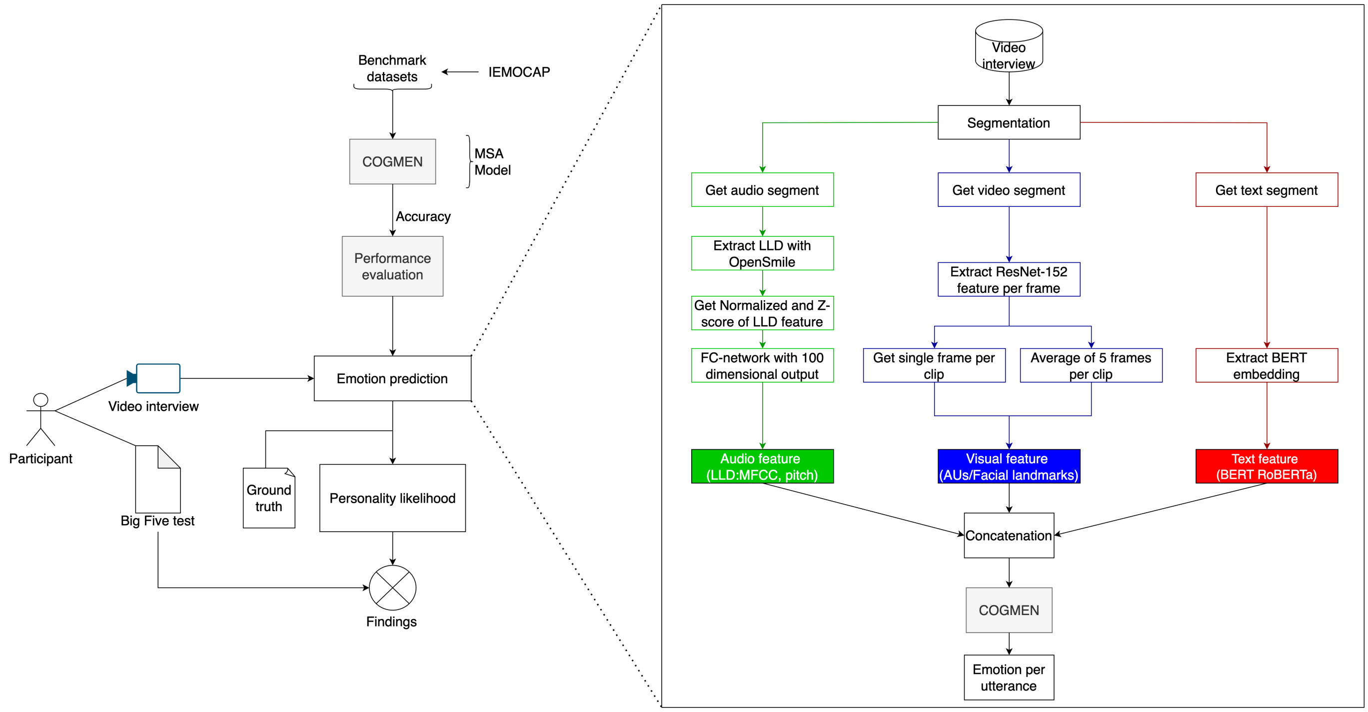

Fig. 1: High-level architecture of the proposed framework.

in each frame. The MTCNN model creates a bounding box around the face in each frame, which is used to crop each frame image before feeding it to the ResNet model.

The ResNet-152 model [25] is then used to extract a 512-dimensional feature vector for each frame. Since the analysis is conducted on a per-segment basis, only one frame is extracted from each video segment. The ResNet-152 model is pre-trained on the imagenet-1k dataset [26], which consists of over 1,2 million training images, 50,000 validation images, and 100,000 test images. It is important to note that the imagenet-1k dataset was intended for image classification, rather than emotion or sentiment analysis. As the ResNet-152 model is exclusively utilized for facial feature extraction, it is unnecessary to train it for sentiment analysis purposes.

C. Data Acquisition

To the best of our knowledge, few multimodal datasets concerning AVIs exist. Therefore, we create a novel dataset, named the *AVI dataset*. Data was collected by inviting participants to record themselves answering interview questions. Details about the established dataset are presented in Table I. *Modalities* denotes the different modalities of A(coustic), V(isual), and T(ext). *#v* indicates the total number of video segments, *Lang* denotes language, and #s designates the total number of speakers. *TL* is the total length of the dataset. In the following subsections, we describe important decisions when creating the AVI dataset.

1) Interview setting: The AVIs design, which conveys the interview setting, is an important part of the AVI. It can be designed in various fashions and is dependent on the desired outcome of the interview [23]. AVI design features include question formatting, question timers, response preparation time, re-recording possibility, interrupted interview completion, length of allowed response, etc. In this research, our ambition is to enable participants to express as many emotions as possible. Therefore, we have set the following AVI design features:

TABLE I: AVI dataset statistics.

Name	Year	Modalities	#v	#s	Lang	TL (hh:mm:ss)
AVI dataset	2023	A + V + T	180	6	English	00:33:58

- Response preparation time: Participants can view the questions right before the recording happens. This allows interviewees to show their initial communication skills as well as express emotions through nonverbal communication channels. However, since textual modality is the most important, it is essential that participants are able to communicate through questions and answers.
- Re-recording responses: Participants are allowed to re-record each response up to three times. This is to mitigate the difficulties of having no preparation time. Additionally, allowing applicants to re-record their responses increases the fairness perception of the AVI [23].
- Interrupted interview completion: Participants should complete the interview in one session because we do not want them to practice answers in their breaks.
- Length of allowed responses: The length of each response is set to one minute.

2) Interview questions: We chose the following interview questions:

1) What motivates you? What are you passionate about?
2) Not everyone agrees all the time. Have you had a peer, teammate or friend disagree with you? What did you do?
3) Give an example of a time you have gone over and above to achieve something. Why was it important for you to achieve this?
4) Sometimes things don't always go to plan. Describe a time when you failed to meet a deadline or personal commitment. What did you do? How did that make you feel?

These questions are open-ended, enabling participants to express their emotions based on past behavior [2].

3) Participants: Participants that were invited to perform the interview are university students at the end of their study program. For that reason, it was natural to ask these people who are in a job-seeking position to participate in the interview. A common denominator among the interviewees is that they are in an informatics study program. When they apply for a job position, all desire to fit the same applicant profile. Further details about ethical and legal considerations are described in Section VI.

D. Personality Prediction Approach

The use of MSA for personality prediction is something new. Previous work is limited to predicting personality based on language. Exploring the possibilities of predicting behavioral traits in AVIs can be of great value in job recruitment. We aim to predict personality based on the Big Five personality traits, a widely used model in contemporary psychology [28]. The model consists of five basic dimensions which are *openness* (to experience), *conscientiousness*, *extraversion*, *agreeableness*, and *neuroticism*. These dimensions outline the structure of human personality [29]. People who score high on openness to experience are characterized by greater openness to feelings and emotions, imagination, appreciation of art and beauty, adventurousness, and liberalism [28], [40]. People who score high on conscientiousness are characterized by competence, orderliness, achievement-striving, dutifulness, self-discipline, and deliberation compared to low conscientiousness individuals. Individuals who are extrovert report greater activity, need for stimulation, seeking social interactions, gregariousness, warmth, and tendency to experience positive emotions than introvert individuals. People who score high on agreeableness report greater tender-mindedness, straightforwardness, modesty, trust, compliance, and altruism than individuals with low agreeableness. Individuals who score high on neuroticism are characterized by more vulnerability to stress, the tendency to experience negative emotions, self-consciousness, and impulsivity compared to people who are more emotionally stable.

The personality prediction approach is based on the emotion distribution in an interview (i.e. the number of various emotions expressed in the video). By looking at which emotions are most dominant, we predict the likelihood of the three strongest traits of each participant. The basis of our prediction is the work conducted in [33]–[38]. We can say that each trait has the following relations to emotions (illustrated in Fig.2):

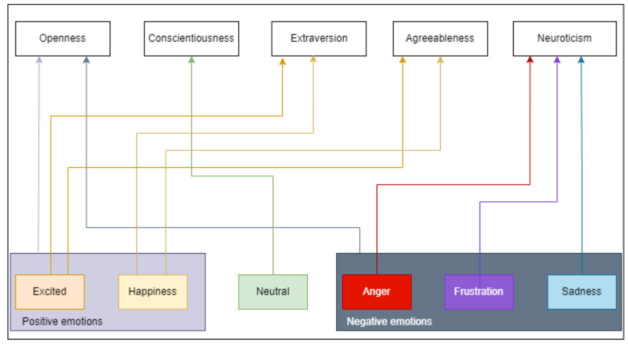

Fig. 2: Relationship between coarse-grained/fine-grained emotions and personality traits.

- Openness: Associated with a mix of positive and negative emotions. In addition, neutral emotions are associated with a higher likelihood of openness trait.
- Conscientiousness: Associated with neutral emotions and a lower negative affect.
- Extraversion: Associated with positive emotions.
- Agreeableness: Associated with positive emotions and lower negative affect. This trait is also overlapping with conscientiousness, meaning that a higher likelihood of conscientiousness increases the probability of being agreeable.
- Neuroticism: Associated with negative emotions.

E. Validation Approach

Given the lack of ground truth i.e. labeled personality traits for the collected dataset, participants pursue a personality test that works as a validation mechanism for the predicted behavior traits. The participants answered the Big Five personality test which is based on the 120-item IPIP NEO-PI-R [24]. The personality inventory consists of 120 statements that are rated on a 5-point scale from strongly agree to strongly disagree. This validation approach is used to examine whether a predicted personality trait reflects the trait via self-assessment the same or not.

IV. Results and Discussion

The COGMEN model, which achieves state-of-the-art performance on IEMOCAP for emotion prediction, was used to detect the emotions expressed in the AVIs dataset. The features extracted in Section III-B are fed to the COGMEN for prediction. Figure 3 depicts the distribution of expressed emotions in an interview for the interviewees with the likelihood of the three strongest behavioral traits. Emotions distribution for all participants (P1-P6) is presented in Fig. 4.

The emotions are normalized in the range of 0 - 1 by finding the sum of an emotion per video over the total number of emotions for a given video. Predicted personality traits are

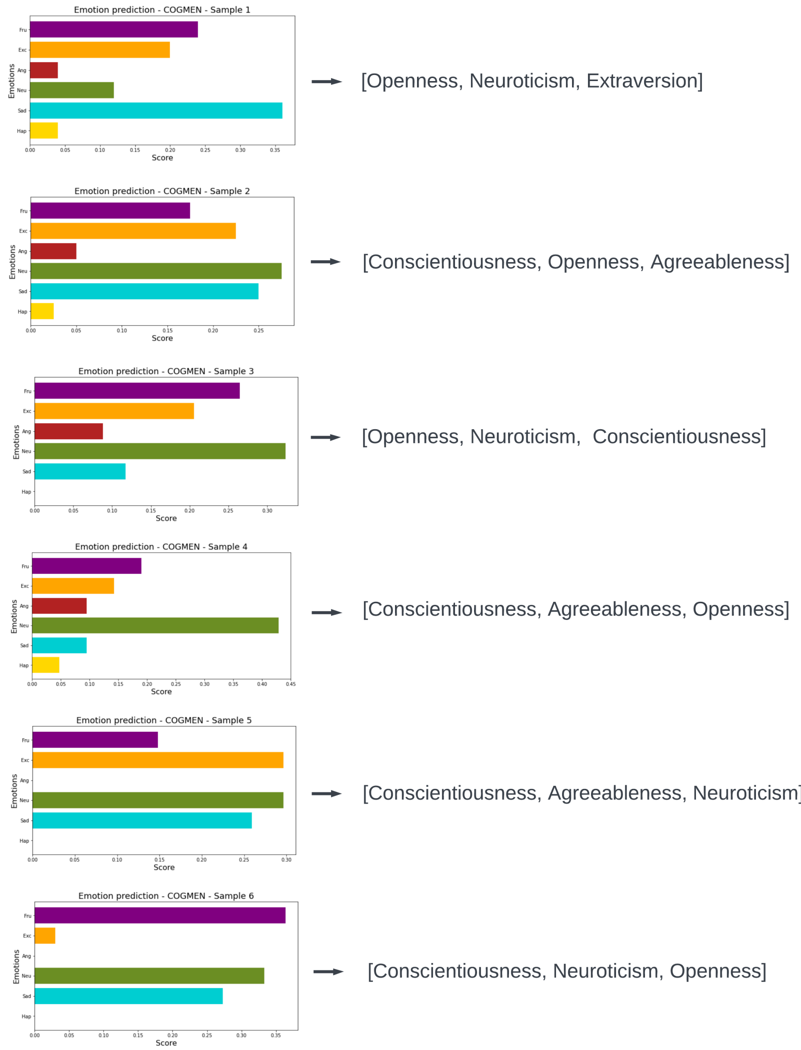

Fig. 3: Emotion distribution with predicted personality traits.

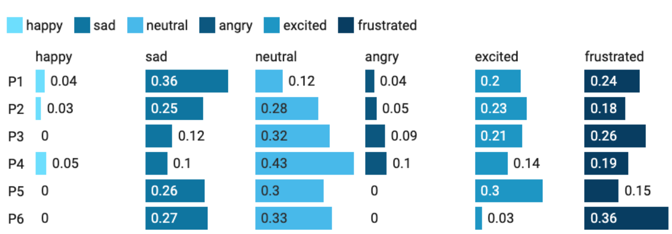

Fig. 4: Emotion distribution of all samples.

presented with the strongest trait first. As shown from the figure, few participants expressed happy emotions during the interview, whereas more people showed excited emotions. On the other hand, interviewees expressed a higher degree of negative emotions (i.e. frustration, anger, sadness) as well as neutral emotion. The reason for this emotion distribution could be that participants were unfamiliar with this interview setting. It may have resulted in interview nerves because interview questions were presented right before the recording of the responses.

Further, out of the predicted personality traits, conscientiousness, openness, and neuroticism are the most reoccurring traits across samples. It is interesting to note that a higher degree of neutral and excited emotion provides a larger likelihood of a participant being conscientious. When a mix of positive, negative, and neutral emotions is present, there is a higher probability of openness. Neuroticism is only associated with negative emotions. An important notion is that when anger is expressed, the likelihood increases of a person with neurotic traits. Agreeableness and extraversion are the least predicted traits and the reason for this is that these traits are only associated with positive emotions. However, due to the fact that agreeableness has overlap with conscientiousness, it can be easier to predict that someone is agreeable than someone being extrovert/introvert.

Results from the Big Five personality test for all participants

(P1-P6) are shown in Fig. 5. This works as a validation mechanism for our detected personality traits. By comparing the predicted personality traits with the Big Five test scores, we are able to predict the three highest traits with an accuracy of 67 %. For example, in sample 4, all predicted traits match the test results. Especially in this sample, the neutral emotion is dominant. However, sample 3 predicts one true label which is conscientiousness. The other two true labels are extroversion and agreeableness. These are the most challenging traits to predict since they are solely associated with positive emotions.

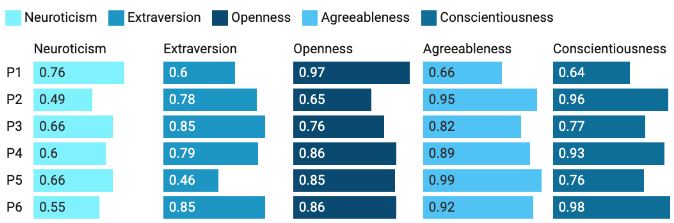

Fig. 5: Personality distribution of all samples.

Following the detected personality traits by the proposed approach and the analysis of results presented in previous subsections, it is safe to assume that MSA can assist in the prediction of behavioral traits of candidates in AVIs. It is difficult to say to what extent, as the number of interviews was less than for any statistically significant observations. It was observed that a neutral emotion is a good predictor of conscientiousness and anger for neuroticism. A mix of positive, neutral, and negative emotions can predict openness, whereas agreeableness can be forecast by an excited emotion. In addition, agreeableness tends to be prevalent in samples that have a strong indication of conscientiousness. Extroversion is the most difficult trait to predict due to its correlation with happy and excited emotions.

A. Limitations

Even though the work carried out in this study draws a reasonable portrait of personality prediction based on MSA, it is impossible to cover all aspects within this domain. The following are the limitations of our work:

1) The collected dataset from six participants is not too large. Predictions from a larger sample size may yield different results.
2) The AVI dataset lacks ground truth i.e. labeled emotions and personality traits. Therefore, we are dependent on the accuracy of COGMEN for emotion recognition and the validation method for personality detection.
3) This work utilizes the Big Five personality test as a validation procedure to check if our prediction reflects self-assessment. As the use of personality tests is accepted, they are prone to response bias. Some candidates may respond inaccurately or untruthfully to the questionnaires to make them look more favorable than others. This aspect can threaten the validity of the study.

V. Conclusion and Future Work

This paper aimed to investigate how MSA can assist in the prediction of personality traits in asynchronous video interviews (AVI). Deep learning COGMEN pre-trained on the IEMOCAP benchmark was utilized for the emotion classification of each utterance in a video. A multimodal feature extraction pipeline is developed for retrieving acoustic, textual, and visual features. A private AVI dataset was collected in this study. Emotions expressed by six interviewees are analyzed. This article also presented a unique way of predicting personality traits based on the emotion distribution in an interview. We further validated the proposed approach by comparing the predicted traits with self-assessment utilizing the Big Five personality test. Our findings showed that MSA can be utilized to predict behavioral traits. The results also indicated that the different dimensions of the Big Five can be correlated with coarse-grained emotions (sentiment level) as well as fine-grained emotions. Openness is correlated with a mix of positive, negative, and neutral emotions. Conscientiousness is prevalent when there is a higher degree of neutral emotions present in an interview. Extroversion is solely correlated with happiness and excited emotions. Agreeableness is related to positive emotions. In addition, it is an overlapping trait with conscientiousness, increasing the probability of agreeableness in line with conscientiousness. Neuroticism is associated with negative emotions including sadness, anger, and frustration. Our novel personality prediction approach achieves 67 % accuracy in detecting the three highest personality traits in candidates. In the future, it would be interesting to expand the AVI dataset with a larger sample size as well as annotated emotions and personality labels. When a large-scale AVI dataset is created, it paves the way to create Deep learning models for multi-class classification since emotion recognition and personality detection are complementary tasks.

VI. Ethical and Legal Considerations

The AVI dataset collected for this project contains video information from the participants and thus requires strict security measures to ensure confidentiality. These measures include secure storage of the data and the deletion of the data once the project is completed. To ensure anonymity, the data has been de-identified and assigned unique identifiers instead of participant names and includes information about the participants' current occupation or field of study. In addition, we obtained signed agreement forms from all participants before using their data. The processing of the data was deemed lawful by the Norwegian Centre for Research Data (NSD) on the basis of the General Data Protection Regulation (GDPR). The collected data contains only low-risk information regarding the data subjects.

References

[1] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria and S. Eetemadi, "Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features," 2020 IEEE International Conference on Data Mining (ICDM), Sorrento, Italy, 2020, pp. 1184-1189, doi: 10.1109/ICDM50108.2020.00146.

[2] M. Jayaratne and B. Jayatilleke, ”Predicting Personality Using Answers to Open-Ended Interview Questions,” in IEEE Access, vol. 8, pp. 115345-115355, 2020, doi: 10.1109/ACCESS.2020.3004002.

[3] N. Anderson, J.F. Salgado and U.R. Hülsheger, ”Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity”, Int. J. Selection Assessment, vol. 18, no. 3, pp. 291-304, 2010, doi:10.1111/j.1468-2389.2010.00512.

[4] C. Sołek-Borowska and M. Wilczewska, ”New Technologies in the Recruitment Process,” Economics and Culture, vol. 15, pp. 25-33, 2018, doi:10.2478/jec-2018-0017.

[5] E. Lukacik, J.S. Bourdage, N. Roulin, ”Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews, ,” Human Resource Management Review, vol. 32, no. 1, pp. 100789, 2022, ISSN 1053-4822, doi:10.1016/j.hrmr.2020.100789.

[6] W. Han, H. Chen, A. Gelbukh, A. Zadeh, L.P. Morency, and S. Poria, ”Bi-Bimodal Modality Fusion for Correlation-Controlled Multimodal Sentiment Analysis,” In Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI ’21), Association for Computing Machinery, pp. 6–15, 2021, doi:10.1145/3462244.3479919.

[7] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, A. Hussain, ”Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” Information Fusion, vol. 91, pp. 424-444, ISSN 1566-2535, 2023, doi:10.1016/j.inffus.2022.09.025.

[8] M. Wöllmer et al., ”YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context,” in IEEE Intelligent Systems, vol. 28, no. 3, pp. 46-53, May-June 2013, doi: 10.1109/MIS.2013.34.

[9] C. Langlet and C. Clavel, ”Adapting sentiment analysis to face-to-face human-agent interactions: From the detection to the evaluation issues,” 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi’an, China, 2015, pp. 14-20, doi: 10.1109/ACII.2015.7344545.

[10] T. Wang, Y. Hou, D. Zhou and Q. Zhang, ”A Contextual Attention Network for Multimodal Emotion Recognition in Conversation,” 2021 International Joint Conference on Neural Networks (IJCNN), Shenzhen, China, 2021, pp. 1-7, doi: 10.1109/IJCNN52387.2021.9533718.

[11] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, A. Hussain, ”Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” Information Fusion, vol. 91, pp 424-444, ISSN 1566-2535, 2023, doi: 10.1016/j.inffus.2022.09.025.

[12] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.F. Chang, M. Pantic, ”A survey of multimodal sentiment analysis,” Image and Vision Computing, vol. 65, pp. 3-14, ISSN 0262-8856, 2017, doi: 10.1016/j.imavis.2017.08.003.

[13] A. Gandhi, K. Adhvaryu and V. Khanduja, ”Multimodal Sentiment Analysis: Review, Application Domains and Future Directions,” 2021 IEEE Pune Section International Conference (PuneCon), Pune, India, 2021, pp. 1-5, doi: 10.1109/PuneCon52575.2021.9686504.

[14] N. Ahmed, Z.A. Aghbari, S. Girija, ”A systematic survey on multimodal emotion recognition using learning algorithms,” Intelligent Systems with Applications, vol. 17, ISSN 2667-3053, 2023, doi: 10.1016/j.iswa.2022.200171.

[15] A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.P. Morency, ”Tensor Fusion Network for Multimodal Sentiment Analysis,” In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1103–1114, Copenhagen, Denmark, Association for Computational Linguistics, 2017, doi: 10.18653/v1/D17-1115.

[16] T. Wang, Y. Hou, D. Zhou and Q. Zhang, ”A contextual attention network for multimodal emotion recognition in conversation,” in 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1–7. doi: 10.1109/IJCNN52387.2021.9533718.

[17] N. Majumder, D. Hazarika, A. Gelbukh, E. Cambria, and S. Poria, ”Multimodal sentiment analysis using hierarchical fusion with context modeling,” Knowledge-Based Systems, vol. 161, pp. 124-133, ISSN 0950-7051, 2018, doi: 10.1016/j.knosys.2018.07.041.

[18] W. Han, H. Chen, A. Gelbukh, A. Zadeh, L.P Morency, and S. Poria, ”Bi-Bimodal Modality Fusion for Correlation-Controlled Multimodal Sentiment Analysis,” In Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI ’21), Association for Computing Machinery, New York, NY, USA, pp. 6–15, 2021, doi: 10.1145/3462244.3479919.

[19] G. Zhao, Y. Ge, B. Shen, X. Wei and H. Wang, ”Emotion analysis for personality inference from eeg signals,” IEEE Transactions on Affective Computing, vol. 9, no. 3, pp. 362–371, 2018. doi: 10.1109/TAFFC.2017.2786207.

[20] M. Hiebler-Ragger, J. Fuchshuber, H. Dröscher, C. Vajda, A. Fink and H. F. Unterrainer, ”Personality influences the relationship between primary emotions and religious/spiritual well-being,” Frontiers in Psychology, vol. 9, 2018, ISSN: 1664-1078, doi: 10.3389/fpsyg.2018.00370.

[21] Y. Li, A. Kazemeini, Y. Mehta and E. Cambria, ”Multitask learning for emotion and personality traits detection,” Neurocomputing, vol. 493, pp. 340–350, 2022, ISSN: 0925-2312. doi: https://doi.org/10.1016/j.neucom.

[22] A. Joshi, A. Bhat, A. Jain, A. Singh, and A. Modi, ”COGMEN: COntextualized GNN based Multimodal Emotion recognitioN,” In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4148–4164, Seattle, United States, Association for Computational Linguistics, 2022, doi: 10.18653/v1/2022.naacl-main.306.

[23] E.-R. Lukacik, J. S. Bourdage and N. Roulin, ”Into the void: A conceptual model and research agenda for the design and use of asynchronous video interviews,” Human Resource Management Review, vol. 32, no. 1, p. 100 789, 2022, ISSN: 1053-4822, DOI: 10.1016/j.hrmr.2020.100789.

[24] J. A. Johnson, ”Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120,” Journal of Research in Personality, vol. 51, pp. 78–89, 2014, ISSN: 0092-6566, DOI: 10.1016/j.jrp.2014.05.003.

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

[26] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... and Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. International journal of computer vision, 115, 211-252.

[27] N. Zhang, J. Luo and W. Gao, ”Research on Face Detection Technology Based on MTCNN,” 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi’an, China, 2020, pp. 154-158, doi: 10.1109/ICCNEA50255.2020.00040.

[28] R. R. McCrae and P. T. Costa, ”Validation of the five-factor model of personality across instruments and observers.,” Journal of personality and social psychology, vol. 52, no. 1, p. 81, 1987.

[29] M. Gungea, V. Jaunky and V. Ramesh, ‘Personality traits and juvenile delinquency: A critical analysis,’ International Journal of Conceptions on Management and Social Sciences, vol. 5, pp. 42–46, Mar. 2017.

[30] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In Proceedings of the 18th ACM international conference on Multimedia (MM ’10). Association for Computing Machinery, New York, NY, USA, 1459–1462. https://doi.org/10.1145/1873951.1874246

[31] S. Poria, E. Cambria, D. Hazarika, N. Mazumder, A. Zadeh and L. -P. Morency, ”Multi-level Multiple Attentions for Contextual Multimodal Sentiment Analysis,” 2017 IEEE International Conference on Data Mining (ICDM), New Orleans, LA, USA, 2017, pp. 1033-1038, doi: 10.1109/ICDM.2017.134.

[32] A. B. Zadeh, P. P. Liang, S. Poria, E. Cambria and L.-P. Morency, ”Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018, pp. 2236–2246.

[33] C. G. DeYoung, L. C. Quilty and J. B. Peterson, ”Between facets and domains: 10 aspects of the big five.,” Journal of personality and social psychology, vol. 93, no. 5, p. 880, 2007.

[34] W. Revelle and K. R. Scherer, ”Personality and emotion,” Oxford companion to emotion and the affective sciences, vol. 1, pp. 304–306, 2009.

[35] E. Komulainen, K. Meskanen, J. Lipsanen, J. M. Lahti, P. Jylhä, T. Melartin, M. Wichers, E. Isometsä and J. Ekelund, ”The effect of personality on daily life emotional processes,” PLoS One, vol. 9, no. 10, 2014.

[36] M. Hiebler-Ragger, J. Fuchshuber, H. Dröscher, C. Vajda, A. Fink and H. F.Unterrainer, ”Personality influences the relationship between primary emotions and religious/spiritual well-being,” Frontiers in Psychology, vol. 9, 2018, ISSN: 1664-1078. DOI:10.3389/fpsyg.2018.00370.

[37] K. A. Barford and L. D. Smillie, ”Openness and other big five traits in relation to dispositional mixed emotions,” Personality and individual differences, vol. 102, pp. 118–122, 2016.

[38] G. Mohammadi and P. Vuilleumier, ”A Multi-Componential Approach to Emotion Recognition and the Effect of Personality,” in IEEE Transac-

tions on Affective Computing, vol. 13, no. 3, pp. 1127-1139, 1 July-Sept. 2022, DOI: 10.1109/TAFFC.2020.3028109.
[39] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," arXiv preprint arXiv:1908.10084, 2019.
[40] U. Baranczuk, "The five factor model of personality and emotion regulation: A meta-analysis," Personality and Individual Differences, vol. 139, pp. 217–227, 2019, ISSN: 0191-8869. DOI: 10.1016/j.paid.2018.11.025.