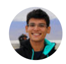
Multimodal Web Application to Infer Emotional Intelligence of Adolescent Counsellor

Conference Paper · September 2019

7 authors, including:


Anupama Ray
Indian Institute of Technology Delhi


Ayush Shah
Indraprastha Institute of Information Technology

Akshay Gugnani
IBM Research

Gargi Dasgupta
IBM

Multimodal Web Application to Infer Emotional Intelligence of Adolescent Counsellor

Prerna Agarwal*, Anupama Ray*, Ayush Shah†, Akshay Gugnani*, Priyanka Halli‡, Shubham Atreja†, Gargi Dasgupta*
IBM Research India*†, EmancipAction India‡
*{preragar, anupamar, aksgug22, gaargidasgupta}@in.ibm.com, †ayush13027@iiitd.ac.in, †shubhamatreja@gmail.com,
‡ priyanka@emancipaction.org

***Abstract*—There are only 0.3 psychiatrists and 0.047 psychologists per 100,000 people in India, compared to a country like the US, which has 29 psychologists per 100,000 people (according to WHO), thereby leading to lack of counselling services and mental health-care. Fortunately, researchers in India have found mental health interventions delivered by lay counsellors rather than specialists to be effective in treating and preventing mental health problems. However, choosing a lay counsellor from a pool of candidates becomes a very important but time-consuming and tedious task because of our deficits in evaluating emotional capabilities, implicit biases and facilitation skills in a resume and standard interview. In this paper, we present a highly scalable web application that can help in hiring emotionally intelligent lay-counselors. The backend framework measures several vital emotional intelligence features that are crucial in a prospective lay counsellor. The framework uses multi-modal data and provides a ranking of potential counsellors. The results and inferencing help establish the importance of each modality and gives insights on features that are key to identify the emotional skills. We compare the predicted rankings to those given by the interviewers (a clinical psychologist and a psychiatrist) and recognize the benefits of automation of the process as well as a need for a deeper analysis of interview questions, discriminative features and importance of multi-modality assessments.**

***Index Terms*—Emotional Intelligence, multi-modal analysis, Machine Learning, Audio Analysis**

I. Introduction

The latest report on global trafficking trends published by the UNODC estimates one-third of trafficked individuals are children (2014), and approximately 2.3 million reside in India[1]. Tragically, even when children are rescued from trafficking and housed in government Child Care Institutions (CCI), the lack of rehabilitative services compounds the high rates of untreated mental health disorders and subsequent re-trafficking. According to a 2011 report by the Indian National Institute of Child Development, a survey of 185 childcare institutions found counseling and therapeutic services lacking in most homes[2]. This may be because the entire mental health workforce in India, comprising clinical psychiatrists, psychologists, psychiatric social workers and psychiatric nurses stands at 7,000, while the actual requirement is 54,750[3].

The lack of mental health specialists in India has made alternative models of care necessary [1]. Psychiatric research has established common factors such as empathy, warmth, and the therapeutic relationship to be highly correlated with successful client outcomes. Thus, it is not surprising that lay counsellors with little experience can deliver mental health interventions successfully if they have high levels of empathy and emotional intelligence (EI). However, finding lay counsellors who have higher emotional intelligence and are more likely to be able to create a therapeutic alliance, is difficult to ascertain from a resume or one standard interview. In addition, human assessments are subjective and interviewers often incorporate their own biases from past education and experience. Artificial Intelligence (AI), however, can be an unbiased tool to assess and objectively identify those with better emotional intelligence suited for the role of an adolescent counsellor. As we create models, however, we do incorporate feedback from multiple human annotators and compute correlation between them to ensure we learn from the knowledge, in the hope of eventually having a reliable and standardized model. This model is crucial for scaling up the lay counsellor delivered mental health intervention to all CCIs in the country, with significant resource savings due to the automated evaluation system which can be accessed from anywhere.

Recent studies have used text and audio to measure personality traits and emotion recognition such as anger, disgust, fear, happiness, sadness, and surprise [2]. Gender bias has also been assessed by building multidimensional scale assessment models [3], [4]. In [2], an algorithm for analyzing anger, panic and normal emotional states is developed using prosodic features. Dellaert et. al [5] explored several statistical pattern recognition techniques to classify utterances according to their emotional content (happiness, sadness, anger and fear). Sato and Morishima [6] used neural networks to classify emotional speech. Building on the previous works, we present an AI based multi-modal interviewing framework to predict the emotional intelligence and other capabilities of candidates. This entire framework is bundled up in a web application, making it highly scalable.

A. Challenges

While building an automated procedure of hiring lay counsellors, the following challenges need to be addressed in order

[1] http://www.unodc.org/res/cld/bibliography/global-report-on-trafficking-in-persons_html/GLOTIP_2014_full_report.pdf

[2] http://nipccd.nic.in/reports/dsmdata.pdf

[3] http://www.businessworld.in/article/Is-India-Suffering-From-A-Deficiency-Of-Mental-Health-Professionals-/11-09-2017-125815/

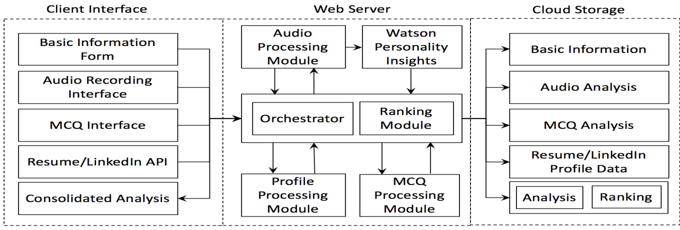


Fig. 1. Block diagram of proposed procedure of evaluating counsellors

to design a reliable, standardized, and unbiased solution.

- Several gold standard questionnaires are available to evaluate EI, but no diagnostic tests exist to assess emotional aptitude directly from a conversation, although a psychiatrist can evaluate humans from a conversation. Such a test or technology would be highly beneficial for the mental health caregivers for assessing conditions and tracking improvements over time.
- It is time consuming and resource heavy to capture multiple personality facets in a face-to-face interview and scaling-up is difficult for wide geographies, emphasizing a need for automation.
- Multi-linguality in audio data mandates translation along with transcription to be able to do text analysis along with audio.
- For sensitive social problems (hiring counselors for rescued children) getting large open source datasets is a huge challenge.

B. Contributions

The main contributions of this paper are as follows:

1) To the best of our knowledge, this work is the first to use multimodal data to predict some emotional intelligence traits such as curiosity, self and social emotional awareness and management, biases and other skills such as communication, facilitation etc.
2) The proposed system is bundled into a web application, where automated interviews can be conducted to assess EI and other traits without a clinician-delivered interview process. This improves scalability, assures unbiased results and helps standardize the procedure of choosing reliable lay counsellors.
3) We employ Resume Parsing and implicit skill identification for better matching and then use the MCQ, text and audio data to predict the EI features and successfully combat the challenges mentioned above.

II. Proposed System Architecture

Previously, the manual process of hiring a lay counsellor began with a resume filtering and the candidates further underwent face-to-face interviews, where interviewers were trained clinicians. The interview consisted of a list of standardized questions and assessment based on them. A web based automated solution for this procedure can help in eliminating the need of physically traveling but ensuring the standards.

Therefore, this paper proposes an AI based interviewing process where we leverage the past work in emotional intelligence assessment to formulate questions and then the evaluation procedure. Our evaluation procedure requires identification of specific features (referred to as *Ground features* in the paper) required in a counsellor, and a distinct set of questions that can capture the *Ground Features* (guided by trained clinicians of the NGO). To accomplish this, the procedure is broken down into two parts: a set of Multiple Choice Questions (MCQ's) and a set of questions for face-to-face interviews.

Figure 1 shows the block diagram of the proposed evaluation procedure. The Client Interface (faced by the candidates) first displays a Basic Information Form, where the candidates fill in details about languages known and educational qualification. It generates a unique-identifier for the candidate through which the responses and analysis is done. This ensures masking of the personal information. The candidate can upload their Resume/LinkedIn profile to be processed by Profile Processing and Job-matching Module. The candidate is then shown a question along with options to record, pause, and upload their audio response which is then sent to the server. The audio received on the server is automatically transcribed to English and sent to Watson Personality Insights. The candidate is then required to answer MCQ questions. The Ranking Module combines the analysis done by different modules in the Web Server and produces a ranking. The consolidated analysis is displayed back on the client interface.

Algorithm 1: Evaluation and Scoring of a Candidate

1 **Input:** Resume R, MCQ responses of 6 sections $M_1,..,M_6$, Audio Recordings of the responses of 10 questions $A_1,...,A_{10}$ of candidate C
2 **Output:** Score S of candidate C
3 $total\_mcq\_score \leftarrow 0$
4 $total\_audio\_score \leftarrow 0$
5 $total\_text\_score \leftarrow 0$
6 $affinity\_score \leftarrow Profile\_Processing(R)$
7 **for** i in range (1,6) **do**
8     $score\_mcq_i \leftarrow Scoring(M_i)$
9     $total\_mcq\_score \leftarrow total\_mcq\_score + score\_mcq_i$
10 **for** i in range (1,10) **do**
11     $A_i^{'} \leftarrow Preprocess(A_i)$
12     $audio\_feature\_vector_i \leftarrow extractFeature(A_i^{'})$
13     $score\_audio_i \leftarrow predictScore_i(audio\_feature\_vector_i)$
14 **for** i in range (1,6) **do**
15     $ground\_feature\_score_i \leftarrow Rule_i(score\_audio_1, ..., score\_audio_{10})$
16     $total\_audio\_score \leftarrow total\_audio\_score + ground\_feature\_score_i$
17 **for** i in range (1,10) **do**
18     $A_i^{'} \leftarrow Preprocess(A_i)$
19     $text_i \leftarrow SpeechToText(A_i^{'})$
20 $text\_scores_1, .., text\_scores_{16} \leftarrow WatsonInsights(text_i, .., text_{10})$
21 $total\_text\_score \leftarrow Sum(text\_scores_1, .., text\_scores_{16})$
22 $S \leftarrow$ $affinity\_score * (total\_mcq\_score + total\_audio\_score + total\_text\_score)$
23 **return** S

Algorithm 1 shows the end-to-end mechanism. The affinity score of a candidate's resume/LinkedIn profile is computed. It then compiles the MCQ responses of 6 sections along with audio responses for 10 questions. The MCQ responses are processed according to the scale outlined in Table I. The audio responses to each question is preprocessed as discussed below and the regression model for each question predicts the score of each question (line 10-13). Audio scores are then used to infer the ground feature scores by using the rules mentioned in Table IV. The scores of *Ground Features* are added up to obtain the total audio score (line 14-16). The audio signals of each question are first preprocessed and then converted into text. The text responses of each question are then assessed with the Watson Personality Insights tool(WPI) [4] to give the scores of the 16 varying features as mentioned in Table II. The scores of these 16 features are added to obtain the total text score (line 17-22). The final score S is obtained by adding up the three scores obtained from MCQ's, ground features and text and multiplying it with the affinity score. The score is computed for each candidate using the Algorithm 1 and sorted in descending order to obtain the final ranking of candidates.

III. Data Collection

MCQ's were selected from different gold standard scales and then deployed into the application to be filled by the candidates. The interview sessions were conducted by clinical experts of the NGO, and the entire audio conversation was recorded with consent. The interviewers rated the candidates for each question as well as for the *Ground Features* (explained in methodology section). The same procedure was also conducted on a group of volunteers (non-job applicants) to study the behavioral differences. The data statistics for both groups are shown in Table **??**. To remove any bias in the order of conductance, the applicants were divided into two groups: One group filled the MCQ's prior to the interview sessions and the other group completed the interview followed by the MCQ's. The dataset is composed of audio recordings in a face-to-face interview setting, where the interviewer asks ten questions for assessing the candidature suitability and the interviewee responds to those questions. There were 16 job applicant interviewees and 14 interviewees from control group, totaling to a set of 300 audio samples (10 questions each) with average 5 minute duration of each sample.

IV. Methodology

In this section, we discuss in detail the methodology used for each module as shown in Figure 1.

A. Profile Processing and Job-matching Module

This module is designed to take inputs in the form of set of user profile documents and job-descriptions [7]. First the resumes are parsed, segmented and tagged based on associative mining rules to classify to Candidate details, Academic details and experience. Using the skill Ontology [7] and deep parsing over each segment, the explicit skills are extracted. The implicit skills are extracted also by reference to similar jobs identified using a pre-trained Doc2Vec model, trained on millions of jobs mined from the web. The Job-Matching module then follows a greedy maximal match algorithm and compute the affinity score. The score is computed as a measure of accounting for frequency, phrase similarity scoring and weather it was an explicit or implicit skill. The scores thus obtained are discussed in the results section.

[4]https://personality-insights-demo.ng.bluemix.net

B. Multi-modal Analysis

The responses of candidates from 3 different modalities i.e., MCQ, Text (Audio transcribed in text) and Audio are analysed through different models.

1) MCQ Analysis: Based on the several gold standard questionnaires, we selected culturally appropriate (in Indian context) MCQ's that had been used to assess emotional, social and relationship strengths. We also incorporated 2 MCQs in the face-to-face interview to assess whether the elaborated answers from audio are semantically related to their corresponding MCQ response. There are 6 counsellor qualities being assessed by MCQs (each quality includes 10 question). Questions for the first four qualities are taken from a standard emotional quotient assessment[5] and their score interpretations are shown in Table I. These sections measures *Emotional Awareness, Emotional Management, Social Emotional Awareness and Relationship Management*. These are important for counsellors to ensure that they will be able to create therapeutic alliance with the traumatized adolescents without burning-out themselves or overpowered by their own emotions. The 5^{th} trait is to assess Gender Responsibilities and section contains 20 questions from the Gender Bias Quiz of the Commonwealth of Learning [8]. The 6^{th} trait measures the past relationship with mother. [9] shows that higher ratings on the *Mother care* subscale of the parental Bonding Instrument [10] correlate to a better working relationship between a counsellor and a patient, thus proving to be an important trait to measure in a potential counsellor. We picked 12 out of the 25 questions from the Mother's form which were appropriate for counselling purposes: six belonging to *care* subscale and six belonging to *control/protect* subscale.

The assessment by MCQs are important to evaluate explicit and implicit biases and EI traits. The candidate's MCQ responses were scored as per the scoring mechanisms provided in each questionnaire as shown in Table I. The response categories included: *Completely Agree, Somewhat Agree, Somewhat Disagree and Completely Disagree*.

2) Audio Analysis: The face-to-face interview questions were created with inputs from the Institute for Healthcare Improvement as well as clinician's. They were curated to assess reliability, biases, empathy, counselling awareness, curiosity, communication skills and a commitment to team work. The 10 interview questions are listed in Table III. The audio recordings for each candidate are clipped to include only the candidates' responses, and de-noised using wavelet transform.

We hypothesize that the quality of communication is measurable via these cues, specifically, that the dimensions of communication and listening skills have measurable acoustic correlates in the timing of speaker responses to questions, and relationship between speech and silence (pauses). To explore this, we extracted the following low-level parameters from each answer: 1) initial response time, 2) number of pauses in the response, 3) mean pause time, 4) standard deviation of

[5]http://www.sdcity.edu/portals/0/cms_editors/mesa/pdfs/emotionalintelligence.pdf

TABLE I
SCALES OF DIFFERENT MCQ SECTIONS

Sections	Total Score Range	Per Question Score	Interpretation
Emotional Awareness, Emotional Management, Social Emotional Awareness, Relationship Management	0-24	1-4	Area for Enrichment: Requires attention and development
	25-34		Effective functioning: consider strengthening
	35-40		Enhanced Skills: use as a leverage to develop weaker areas
Gender Responsibilities	20-35	1-4	Awareness of gender equality is advanced
	36-50		Awareness is advanced but find certain gender concepts difficult to accept
	51-65		Awareness is basic and mixed feelings about responsibility
	66-80		Low awareness and beliefs are shaped by traditional norms
Past Relationship with mother (care)	≥ 13	0-3	caring mother
	< 13		non-caring mother
Past Relationship with mother (protect)	≥ 6	0-3	protective mother
	< 6		non-protective mother

the pause time, 5) skewness of the pause time, 6) total amount of time speaking, 7) total amount of time paused, and 8) the ratio of time spent in pause to time spent speaking. Then, we extracted the following summary features for each speaker to measure communication and listening skills: 1) response time mean, 2) response time standard deviation, 3) response time skewness, 4) ratio of total pause time to total time speaking, 5) mean pause time, 6) pause time standard deviation, and 7) pause time skewness.

To identify pauses in speech, we identify the speaker changes within each of the 10 question-answer exchanges, and applied a voice activity detector with a 20-40 ms duration threshold The spaces between voice activity regions were designated as pauses, with a minimum designated pause duration set to 20 ms. The initial response time to each question, being a direct measure of interaction quality is measured first. For example, longer response times could indicate thoughtful consideration, or possibly confusion; while zero response times could signal eagerness or speaker interruption. Then, pause duration metrics are calculated within each question. Finally, the seven summary features described above were calculated from the marked boundaries of speech and silence for each speaker. This data extracted from the candidates and the control group, and the communication and listening ratings assigned by clinicians were used to fit a two-layer feed-forward neural network with 13 hidden neurons. The resulting model performance was measured in Mean Square Error (MSE), and via regression of predicted vs. actual values.

An audio regression model is also built per question with label as average score given by the clinicians. Low level features like spectrogram and MFCC were extracted to form the mid-level features which later were aggregated to form the high level audio features. These low-level, mid-level and high-level features form the final feature vector for the regression model. We used a linear support vector machine (SVM) with C parameter as the regularization parameter to prevent overfitting. We perform cross-validation using repeated random sampling. After evaluating the regression model, one optimized model is extracted for each learned parameter. We use the predicted scores from the neural network built to identify communication and listening skills along with the audio regression models to give the scores for thr 10 questions, we infer the *Ground Features* such as *Counseling awareness*, *Team work*, *Adolescence concerns*, *Facilitation skills*, *Curiosity* etc as shown in Table IV.

TABLE II
FEATURES EXTRACTED FROM WATSON INSIGHTS TOOL

Variation	Count	Features
Most Variation	16	Emotionality, self-discipline, self-efficacy, Assertiveness, Gregariousness, cooperation, modesty, immoderation, self-consciousness, susceptible to stress, harmony, ideal, love, stability, conservation, openness to change
Least variation	11	Intellect, Dutifulness, Cheerfulness, Altruism, uncompromising, sympathy, trust, prone to worry, liberty, practicality, self-expression

3) Text Analysis: The audio recordings of the candidates were a mix of Hindi, English and other regional Indian languages (Tamil, Konkani and Marathi). Multilingualism makes the audio and text analysis more challenging and requires translation. Therefore, audio recordings of the candidates' responses were manually transcribed verbatim into English to be comprehended by any text processing module. This manual transcription will be replaced with the automatic system to convert multi-lingual speech to text in future to ensure scalability.

To capture key traits and values that clinicians have deemed necessary for a successful counsellor, we used the Watson Personality Insights tool (WPI). WPI provides 47 personality features scores given text as input(as shown in Figure 2) out of which a subset is discriminative and important for a counsellor. The scores range from 0-1 and the presence of each feature is shared in percentages. We extracted the subset of features with maximum variance and tried different combinations within a Random Forest classifier [11] to find important features that distinguish job applicants from non-job applicants. These features are validated with the help of the domain expert. We also identified the set of features with the least variance among job applicants which may represent traits that the majority of such candidates are conditioned by society to express. These features are shown in Table II.

V. RESULTS AND DISCUSSION

In this section, we will show the results from each module as well as the overall results of the system.

A. *Profile Processing Module*

This Module outputs a ranked list of candidates based on their skill suitability with job description. We observe that candidate F is the best match for the job description with an Affinity score of 0.71 based on the resume, as shown in column 3 of table V.

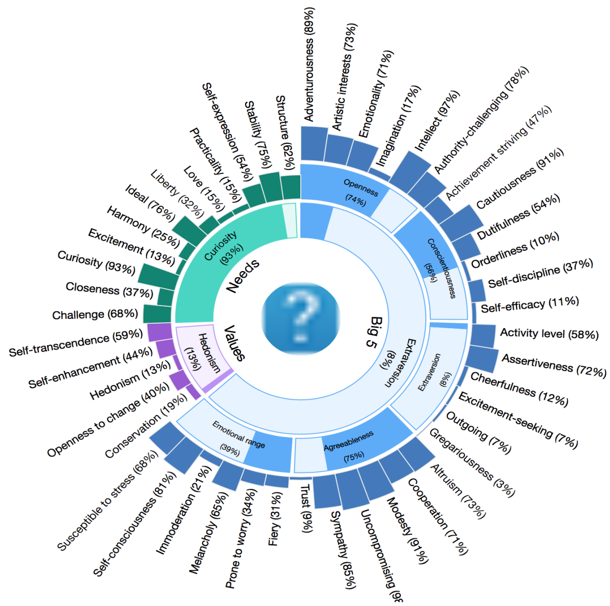


Fig. 2. Watson Personality Insights Tool Feature Visualization

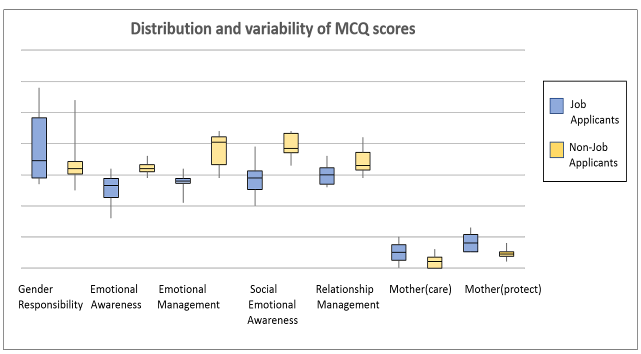


Fig. 3. The Boxplot shows the distribution, spread and variability of MCQ scores per evaluation section in counsellors. The median computed on the values for all candidates is marked with the horizontal line inside the box and the whiskers are the two lines outside the box that extend to the highest and lowest value.

B. MCQ scoring and inferences

The distribution of MCQ scores between the two groups is shown in Figure 3 and is clearly distinguishable. For all sections, the variance in job applicants is high, reinforcing the challenges in selecting them based on only the MCQs. Regarding gender responsibility, the job applicants all agreed with gender equality but more nuanced questions illicited difficulties in accepting higher level gender concepts. Non-job applicants scored significantly better in emotional awareness, emotional management, social emotional awareness and relationship management as compared to the job applicants as they are aware about themselves and were more highly qualified to handle these types of questions prior to the interview.

C. Audio: Inferences

The neural network assessing Listening skills (MSE=0.071), and the Communication skills model (MSE=0.064) performed well. For the audio regression model for the 10 questions, 5-fold cross validation is done on the dataset with 300 audio samples. The dataset was splitted into 80:20 ratio for training and testing. 10% of the train set was chosen randomly for validation. The average MSE/question is shown in Table III. The scores given by the clinical psychologists of the NGO is considered as the ground truth. Each question is mapped to corresponding questions (shown in Column 2 of Table IV). The predicted scores of the corresponding questions were then averaged to infer the score given in the comparative model completed by experts. For example, to assess for *counseling awareness*, interviewers asked candidates about what qualities were necessary in a good counsellor. In our analysis, the predicted score of this question matched accurately with the score given by the expert clinicians.

TABLE III
MSE FROM AUDIO REGRESSION MODEL FOR EACH QUESTION. THE TARGET VARIABLE RANGE IS 0-5.

Questions	Training MSE	Testing MSE
Tell us a bit about yourself	0.10	0.077
Why did you apply for this position	0.44	0.023
Story about how you have helped someone	0.15	0.25
Story about how you have seen someone in a new way	0.82	0.018
Awareness of a bias you have about vulnerable adolescents	0.10	0.20
Teamwork- likes and dislikes	0.10	0.26
Characteristics of a counsellor	0.32	0.06
Anticipated difficulty in dealing with young people	0.10	0.12
T/F: Should women complain if their careers stall during or after a maternity leave	0.15	0.23
T/F: Boys are often better in mathematics and science than girls	0.20	0.32

Each interview also included a *Role Play* where job applicants interacted with an adolescent girl displaying signs of depression. Applicants were evaluated in this interaction by both the clinicians as well as the adolescent client, who rates her level of comfort in the interview and the relatability of the applicant. Analyzing this modality was beyond the scope of the current project as it was much more interactive than interview questions and our text analysis program was not able to compute it. The model appeared to be successful as the (rounded off) predicted values matched the comparative evaluator score for 5 out of top 6 candidates. Questions such as "what do you like or dislike about working in a team?" had a high level of correlation with team work scores and matched for all counsellors perfectly. *Facilitation skills* was more challenging to infer from text analysis. We used the response from the question on the most important characteristics of a counsellor as well as the story told about helping someone to infer this score. The predicted score was highly correlated to interviewer's score however, we hope to validate this further by analyzing communication/listening skills from speech data as well. For the parameter *commitment to work*, although the questions shown in Column 2 of Table IV are used, there are several other parameters such as educational qualifications, expected salary and other job opportunities that influences. Although we had information on their educational background and expected salary, this parameter was also challenging to predict accurately solely with text analysis.

Correlation between Features from different modalities: We computed Pearson's product moment correlation coefficient

TABLE IV
QUESTION MAPPING TO PREDICT GROUND FEATURES

Feature	Mapped Questions	Avg MSE
Counseling Awareness	Characteristics of counsellor	0.0833
Adolescence Concerns	Difficulties with young people, Why applied for the position	0.2083
Team Work	Team like/dislike, story about helping someone	0.375
Facilitation Skills	Characteristics of counsellor, story of helping some-one,communication skills, listening skills	0.333
Curiosity	Why applied for this position, Story of seeing someone in new way, story of helping	0.333
Commitment to work	tell me about yourself, Why applied for this position	0.457

TABLE V
WEIGHTED SCORE RANKING

Candidate	Total Score	Affinity Scores	Weighted Score	Predicted Rank	Actual Rank
A	75.68	0.63	47.68	3	2
B	72.52	0.67	48.58	2	3
C	66.23	0.69	45.70	4	4
D	63.974	0.64	40.94	5	5
E	63.7	0.52	33.12	6	6
F	70.03	0.71	49.72	1	1

[12] for the features of MCQ-text, the 2 MCQ-audio questions and audio-text outputs. For MCQ-text, the correlation was high between the MCQ's sections and the varying features derived from the extracted Watson's Personality traits. The Pearson coefficient shows that a correlation exists between selected personality traits and MCQ's for emotional awareness and relationship management, but did not correlate significantly for Social Emotional Awareness and Emotional Management. For correlation between audio-text features, some features such as teamwork picked up from the audio interview correlated with cooperation and agreeableness; facilitation skills of audio correlated with curiosity, and 6 parameters under extraversion. Both text and audio are used to measure curiosity, so equal weightage is given to both modalities for curiosity. These correlation studies show that the features captured by each modality contribute individually to help assess the suitability of a counsellor. Audio analysis gives insights on counseling awareness, facilitation skills, and adolescent concerns which are difficult to ascertain from MCQ or text. On the other hand, text analysis helps assess personality traits whereas MCQ gives insights on gender bias which can be difficult to analyze from long audio or text inputs.

D. Ranking Module

The Ranking Module combines scores from each modality. The predicted scores are summed to give a total score and the ranked list as shown in Table V. The actual rank in the table refers to the rank given by the interviewer based on the face-to-face interview and the role play. A simple multiplication of total scores and affinity scores is done to obtain the weighted scores. This yields an agreement of 4 out of the top 6 final ranks. However, the significance of the affinity scores and total scores can easily be adjusted by assigning weights to them. The highest ranked candidate according to the interviewers was different from the model primarily due to the commitment to work *Ground Feature*, which is influenced by both salary expectations and other job opportunities. The highest ranked candidate according to the model was rated lower in the *Role Play* due to deficits in listening skills as well as a lower commitment to work, although she got high scores in all the interview questions from the interviewers. We hypothesize that these models will become more appropriate after incorporating Role Play analysis.

VI. CONCLUSION AND FUTURE WORK

This paper presents an end-to-end automated framework to conduct interviews and based on the multimodal data, infer emotional intelligence features and some other skills required in a prospective lay-counsellor. In future, we aim to incorporate video analysis as studies shows that body language, gestures and facial expressions are also important cues to asses EI. This work can also be used for hiring in any other domain. As India struggles with the rising rates of suicide among adolescents, and a 70 percent treatment gap, mental health care is being shifted to lay counselors and community based interventions as there are too few clinicians. Automated systems are necessary to help choose these lay counsellors and improve cost effectiveness, scalabality and also decrease the subjectivity that is inherent in any human evaluator. Models like these provide the first step to build these systems and help ensure that we choose those candidates with the highest likelihood of success to treat this high risk group.

REFERENCES

[1] M. L. Barnett, A. S. Lau, and J. Miranda, "Lay health worker involvement in evidence-based treatment delivery: A conceptual model to address disparities in care," *Annual Review of Clinical Psychology*, vol. 14, no. 1, 2018.

[2] P. B. Dasgupta, "Detection and analysis of human emotions through voice and speech pattern processing," 2017. [Online]. Available: http://arxiv.org/abs/1710.10198

[3] R. B. Parker, P. D. Parker, T. Larkin, and J. Cockburn, "A psychometric evaluation of the gender bias in medical education scale," *BMC Medical Education*. [Online]. Available: https://doi.org/10.1186/s12909-016-0774-2

[4] A. R.-C. Carmen Marfn, "Gender bias in test evaluation."

[5] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," in *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, Oct 1996.

[6] J. Sato and S. Morishima, "Emotion modeling in speech production using emotion space," in *IEEE International Workshop on Robot and Human Communication*, 1996.

[7] A. Gugnani, V. K. R. Kasireddy, and K. Ponnalagu, "Generating unified candidate skill graph for career path recommendation," in *IEEE International Conference on Data Mining Workshops*, 2018.

[8] http://dspace.col.org/bitstream/handle/11599/1740/2015_COL_Gender-Bias-Quiz.pdf?sequence=3&isAllowed=y, 2015.

[9] J. T. M. Anne Grete Hersoug, Per Hoglend and O. E. Havik, "Quality of working alliance in psychotherapy: Therapist variables and patient/therapist similarity as predictors," in *The Journal of Psychotherapy Practice and Research*, vol. 10, 2001.

[10] B. L. Parker G, Tuplin H, "A parental bonding instrument," in *British Journal of Medical Psychology*, 1979, pp. 52:1–10.

[11] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 10, pp. 1619–1630, 2006.

[12] M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," vol. 24, no. 3, pp. 69–71, 2012.