Survey On Interview Performance Prediction and Analysis using Audio Features and NLP

P.Srihari[1], Suravarapu.Shivani[2], Nimmalapudi.Trisha Lakshmi[3], Paidi.Sirisha[4],Tumula.Pooja[5]

[1]Assistant Professor, [2]19341A12B1, [3]19341A1280 [4]19341A1281, [5]19341A12B9

Department of IT, GMRIT, AP

ABSTRACT

The traditional face-to-face interview is the most common recruitment interviews. However,Besides being time-consuming,it can be difficult for some interviewers to read candidates because there is no way to know what their voice sounds like under stress or how to properly grade an applicant based on how they responded.In contrast with Virtual interviews that don't take place in person.The interviewer within short period of time judges the entire personality of the candidate only basis of the answers given by the interviewee.Our framework takes audio as an input and predicts the interview score followed by feedback about candidates personal and emotional traits based on prosodic features and lexical model. By using prosodic features our framework considers emotions like anger, happiness, anxiety and sorrow.While our lexical model works on the major personality characteristics.There will be a marked psychological traits such neuroticism and extraversion. RAVDESS Emotional Speech and Song data set is used for audio analysis.To train our lexical model we used a Stream-of-consciousness data set. Therefore, in our project we use time distributed CNN along with LSTM for audio analysis and NLP with RNN is used to build lexical model.For final score prediction we choose Decision Tree Regression Model.Finally, our framework is less susceptible to human biases and thus improving the interviewers ability to make a hiring decision.

DOI Number: 10.14704/nq.2022.20.9.NQ44767 **Neuro Quantology 2022; 20(9):6567-6573**

I. INTRODUCTION

Online video interviews have become very popular in recent years, making the hiring process quicker and more effective. Many Fortune 500 firms are said to use HireVue, one of the market's established players that has been performing well in the online video interview hosting space. Online video-based interviews have numerous advantages over traditional ones, including the ability for employers to interview multiple candidates at once because they are simple and quick to do. Respondents must coordinate a variety of behaviors in the little time available, including linguistic content, prosody, and nonverbal or facial clues, in order to successfully communicate their qualities. As we know, the interview score is the result of adding personality or quantitative ratings to technical expertise in the relevant sector. During a live interview,we can quickly decipher the implications of both verbal and nonverbal conduct, but we lack a framework for understanding how many characteristics affect the final score in various ways,making the interview process rather challenging.

Enhancing human judgment with an automated evaluation of interview performance is a novel replacement for the conventional human interview scoring approach. Speaking style, prosody, facial traits, lexical features, and language all reveal important details about your character and emotional state.We therefore made an effort to comprehend the significance of each of these distinct characteristics and developed a model that can score your interviews in accordance with a set of accepted guidelines, freeing them from the prejudice inherent in human selection processes.Our research revealed that facial expressions contribute very little to the final interview score and should thus be disregarded, so our approach quantifies HR interviews based on linguistic and

prosodic aspects.

The prediction framework automatically extracts a variety of other factors and assesses both the general interview performance that will decide the chance of getting hired as well as 14 additional social traits crucial to the hiring process.The concept of interview score prediction would be expanded such that it functions more like a virtual coach or a tool that would assist the applicants in Utilize the tool's feedback and ideas to make them better It would be a form of interactive tool that would combine the actual elements that our computational framework would require and try to provide them ideas, summary feedback.

II. LITERATURE SURVEY

In this study, the author developed an AVI-AI platform that combines an asynchronous video interview (AVI) platform with an artificial intelligence (AI) decision agent built on a TensorFlow convolutional neural network (CNN). This platform may be used to precisely anticipate a job candidate's communication skills and personality traits as well as to partially replace the work of human raters during the initial stages of employment screening. Using our own dataset, AVI is utilised to extract the interviewees' facial expressions from each frame of the video data. By monitoring 86 landmark locations on the face every frame, OpenCV and Dlib were used to identify facial features.They employed the Pearson correlation coefficient (R), explained variation (R2), and MSE to evaluate the good reliability of the AVI-AI.

As determined by human raters, the AVI-AI was able to learn and predict neuroticism, interpersonal communication skills, openness, and agreeableness but not conscientiousness and extraversion. The findings suggest that a person's facial emotion characteristics may influence the communication skills that HR professionals evaluated during structured behavioral interviews.Additionally,our prediction program learned where to identify if a person will be interactive, pleasant, or neurotic throughout the interview. Despite this, our algorithm was unable to recognise applicants that are most certain to be conscientious and extraverted.[1]

The author proposed a system that automatically gives candidates summative comments after an interview.The suggested technology proactively predicts an actionable response, providing the applicant with a concrete takeaway at the conclusion of the interview.They reached out to placement trainers and compiled a list of the most frequent comments made during training in an effort to predict them in advance and collected information from more than 145 participants in a setting resembling an interview.The ground truth is provided by a number of cutting-edge machine learning algorithms with human annotation. The prediction models were created with a focus on visual communication cues in order to make it easier to solve the challenges in spoken languages interpretation and task design.Candidates were thought to have more assurance when they responded clearly and loudly with fewer pauses.

Video includes didn't show a significant relationship since most candidates kept neutral facial movements throughout the interview.Results show that auditory and prosodic factors have a stronger correlation with applicant evaluation than audio and lexical aspects.When compared to the baseline accuracy of 77%, our highest results had an accuracy of 95%.Finally, make an effort was made to forecast the candidate's capacity for self-promotion.This was more difficult because the information gathered was from a simulated interview with generic behavioral questions.[2]

The communication skills of applicants in various kinds of behavioral interviews are methodically studied and automatically measured in this research.A dataset of 100 participant structured interviews are gathered.Two human expert annotators independently evaluate these interviews using evaluation criteria tailored to each of the situationsUsing common methods for machine learning like SVM,Random Forests, a prediction model was created that features extracted mixed factors like lexical, visual, and aural information. The skilled and proficient presenters are separated from the subpar and mediocre presenters using a binary classification task.

The combined findings of comparisons among the three interview mode types are presented.

The following consequences are based on how each setting's communication skill is perceived by experts. When individuals' performances in the written and video interviews were compared, 36% did best in the written format, 35% performed best in the video version, and 29% scored equally in both methods.In all three situations, the best model performs with a binary classification task accuracy of 75%.Experiments on a huge and diverse population can be carried out as a future direction for this subject. The same characteristics of cognitive assessments can be studied using enhanced feature groups in all contexts. One can attempt a more thorough user study.[3]

One such piece of software or equipment is the Good Interviewing System, which simplifies the traditional interviewing process utilizing contemporary NLP techniques and deep learning software.For interviewers and HR management candidates working for certain firms who conduct technology-related interviews, the system can be extremely helpful.The method utilizes writing and voice patterns from people. The computerised interviewing method uses content inputs that the system translates from human language into inputs that the computer can understand. After that, a system verifies the validity of the answers that applicants provided during oral interviews, technical interviews, and written tests. Eventually, the system uses deep learning principles to automatically forecast rankings for each solution. Recruiters can reduce the effort they must do in order to select the candidates who are most capable of working for their organization.

This method uses complex evaluation factors rather than human-oriented evaluations to evaluate and compare candidates more properly at each level of the interview. A trained model that uses the linear regression approach is used to detect personality traits. The version's acquired accuracy is 85%. The component for processing written tests uses the Doc2Vec model to determine whether any sentences provided to it have any similarities. This is a research article about the specially designed Smart Interviewing System web application for interviewers, HR management staff, and candidates who may access through interviewers, HR management staff. Using this technique, businesses may quickly identify the best and most valued candidates during the hiring process.[4]

This study attempts to develop a mechanism for forecasting college students' employability using machine learning. To comprehend how college students get employment, three learning algorithms Decision Trees , Random Forest and Support Vector Machine have been applied.Using the total performance matrix, the three algorithms were assessed for accuracy, precision and recall, f1-score, and support.measures.Learning about algorithms was greatly influenced by the preparation phase.

Researchers came to the conclusion that Support Vector Machine generates a predictive model that achieved the best precision and recall metrics of.991 or 91.15% and 91%, respectively, with an accuracy of 91.22%. The study's authors came to the conclusion that the learning curve and validation curve were no longer over- or under-fit. The skill sets that will be given more priority ratings in forecasting college students' employability will be the subject of future research by the researchers. To verify its correctness, the researcher will utilize the results of the trials in a real-world setting. The approach might be used to create a system that forecasts engineering students' employability and determines what skill sets need to be enhanced..[5]

By evaluating the language that was taken from the candidate's tweets, this method aims to simplify the psychological assessment process and provides generic employment roles that the candidates would be a good fit for. Big Five and the Myers-Briggs Type Indicator (MBTI) are two personality analysis models that are commonly utilised.Instead of utilizing the conventional questionnaire designed to measure the attributes, which might be rigged by the candidate, the personality traits of the applicant are ascertained by using their tweets. It is also impractical to fill out the surveys to track personality shifts over time.The application uses the most current tweets to assess the most recent personality change.

The available data is divided into two equal parts, with 70% of the data used for training and 30% for testing, to form a training dataset and a

testing dataset.The TF-IDF weighting method was used to create the word vectors, which represent the word frequency in each text and across the entire corpus. The findings demonstrate each of the four binary MBTI models' accuracy, precision, and recall.In terms of all performance parameters, the support vector model offers the best overall result for all attributes.A metric of the personality analysis of the applied candidates will be displayed to the recruiting company's employer.[6]

The strategy outlined in this paper uses automatic personality screening based on visual, auditory, and lexical data from short video clips to cut down on the time needed for applicant selection at the pre-employment stage. The system's objective is to forecast a candidate's Big Five Personality Trait scores and to anticipate the final decision on whether to invite the person in the video clip to the job interview.A set of important features is taken for each channel and utilized to train individually using Deep Learning.In order to create a time series, the speech signal is first removed from the video file. Because the clips we're working with in this study are so brief (15 seconds or less), all of the predicted values are mostly dependent on first impressions. If so, how someone expresses their emotions through speech depends on their personality attributes and interview values.Therefore, personality analysis and

the assessment of the Interview value may both make use of auditory characteristics, which are widely used for disciplined, etc.), so it has the highest regression weight and correlation value, is seen in the relationship between automatic emotion identification. Researchers used speaking rates, zero crossing rates, MFCCs, and spectral energy distribution characteristics in this investigation.These are frequently used in studies on personality analysis as well as the field of emotion recognition.The final stage combines the results from the visual, audio, and lexical stages to determine the ultimate score. On the basis of the first impression database, the experiment was run, and considerable performance was obtained.[7]

In an online video interview, they are aiming to automatically evaluate a candidate's proficiency using text features from transcripts of automatic voice recognition.Asynchronous video interviews (AVIs) and person-job (or talent-job) fit are two studies that are becoming more and more popular due to the rapid development of online recruitment platforms. An interview in AVI consists of a series of questions and responses that include important social cues. AVI models will extract characteristics from video, text, and spoken answers in order to more completely evaluate candidates. Here is the proposed Hierarchical Reasoning Graph Neural Network.The HRGNN is made up of four interconnected parts: a competency classifier, a sentence-level GAT encoder, a sentence-level GCN encoder, and a gated recurrent unit encoder. They proposed a hierarchical logic neural network (HRGNN) in this paper for automatically scoring answer transcriptions (ASAT) in video interviews. The purpose of the ASAT assignment is to evaluate a candidate's proficiency using a variety of text-based question-answer pairings.The Relational Graph Convolutional Network (RGCN) module was created expressly to explain the connections between inquiry and response phrases.. It generates internal temporal dependencies and interaction dependencies of questions and answers.We then use a GRU-based classifier to assess the candidate's competence. On the Chinese real-world dataset, empirical findings and ablation investigations employing 10 random seeds demonstrate the model's state-of-the-art performance (CHNAT).[8]

The purpose of this research was to look into how personality factors affect how well employees perform in their jobs. It is undeniably true that personality qualities have an effect on an employee's ability to accomplish their job duties, and many firms make use of this impact. The research information takes both independent and dependent variables into account. Personality attribute was chosen as the independent variable in this study while job performance was chosen as the dependent variable. The results show that a positive correlation between personality traits and job performance, with conscientiousness being an attribute closely linked to job performance for employees carrying the most weight.The first personality trait, conscientiousness, has attributes that are directly related to an employee's job performance (responsible,

organized,conscientiousness and the structure of both jobs performance.Extraversion, the second personality characteristic, is a good and reliable indicator of work satisfaction.The third personality trait, openness to experience, works creatively and innovatively, so he does better than others at his job, which differentiates his work from others The fourth personality trait, agreeableness, is a reliable and significant predictor. Employees that are understanding and supportive of their supervisors, clients, and coworkers work more effectively. The fifth personality framework was created to investigate employee personality traits and job performance based on the table below.[9]

Analyzing an interviewee's overall performance is a challenging and complex undertaking. Our suggested approach builds multiclass category models to automate this operation. The suggested system receives video that was recorded throughout an interview and uses it to extract frames and audio. The first version, a face emotion identification model, is given frames. It uses the HaarCascade classifier, Gabor filters, and Convolution Neural Network to classify facial emotions as seven different types, including happiness, surprise, anger, disgust, neutrality, fear and sorrow.The second model receives audio and employs logistic regression and Mel frequency cepstral coefficient characteristics to classify speech into four training categories: fluency, stuttering, clutter, and pauses. The interviewee may receive an overall performance score based on predictions of such trends. The Gabor Filter-based approach that we utilised provided superior accuracy with a narrower range of hidden layers and a great deal less training time when compared to simply CNN-based and Deep Neural Network-based facial emotion identification.We have offered a method for awarding a score to an interviewee.[10]

III. COMPARISION TABLE :

S.no	Author	Methodology/ Techniques	Advantages
1.	Kuo- En Hung et.al(2020)	Tensorflow, Convlutional neural network(CNN)	In order to automatically anticipate interviewee communication abilities and personality qualities, this study utilized a DL model with high concurrent validity and accuracy, which might be an important step in bridging the gap between human imagination and computer reality.
2.	Shruthi Kukal et.al(2017)	Logistic Regression, Linear discriminant analysis(LDA)and Support Vector Machine(SVM)	Support Vector Machine predicts a candidate's self- confidence with a 95% accuracy rate.
3.	Pooja Rao S.B et.al(2017)	SVM, Logistic Regression, Random Forests,nlp	Combination of lexical and sentimental features gives the best by using the Boruta feature selection method.
4.	Malaka Silva et.al(2021)	Support Vector Machine (SVM) and Multinomial NaïveBayes, word2vector, TF-IDF	In comparison to human- oriented evaluations, this technology checks and evaluates candidates more precisely at every stage of the interview. Process mentioned above causes the system to discover more human errors.
5.	Cherry Casuat et.al(2 019)	RandomFore st (RF),Decision Trees (DT), and Support Vector	The Support Vector Machine (SVM) generates a prediction model with the maximum precision and recall metrics of.991 or 91.15%

		Machine (SVM)	respectively, and an accuracy of 91.22%.
6.	Roshal Moraes et.al(2020)	SVM, Decision Tree, KNN andNaïve Bayes	Suggested online personality assessment system is straightforward, affordable, embeddable system easily customizable for use in a variety of application fields..
7.	Jelena Gorbova et.al(2017)	Perceptron neural network, Deep Learning	Combining the predictions made using text, audio,and video features produces a more significant and consistent result.
8.	Kai Chen et.al(2022)	Hierarchical Reasoning Graph Neural Network(HR GNN)	The Relational Graph CNN module builds internal temporal dependency and question-answer interaction dependency represents higher capacity to learn the logic of human scoring.
9.	Vianny Jeniston et.al(2019)	Multiple regression	The results of the relationship between personality and performance demonstrate that personality has a direct impact on task, contextual, and overall job performance.
10.	Yashwanth Adepu et.al(2020)	Yashwanth Adepu et.al(2020)	Adam optimizer gives better accuracy than the rmsprop.

IV. CONCLUSION

In this study, we suggested a methodology for predicting interview performance based on audio and linguistic data..Interview analysis is used by human resources departments and businesses to determine a person's suitability for the job. This project can even be used by students to train and prepare for interviews based on the feedback obtained by our model. For audio analysis we use cnn-lstm technique and Nlp with Rnn is used for lexical analysis provide the best accuracy.This project also covers the use of quite different areas of Deep learning, so we're going to learn something completely new. We are hoping that this project will fulfill company standards and be used in the real world.

REFERENCES

[1].Suen,H.Y.,Hung, K. E., & Lin, C L.(2020).Intelligent video interview agent used to predict communication skill and perceived personality traits.Human-centric Computing and Information Sciences.

[2].Nambiar, S.K.,Das, R., Rasipuram, S., & Jayagopi,D.B. (2017, November).Automatic generation of actionable feedback towards improving social competency in job interviews.In Proceedings of the 1st ACM SIGCHI International Workshop on Multimodal Interaction for Education.

[3].RaoS.B,P.,Rasipuram,S.,Das,R.,Jayagopi,D.B.(2017, November).Automatic assessment of communication skill in non-conventional interview settings:a comparative study.In Proceedings of the 19th ACM international conference on multimodal interaction.

[4].Senarathne,P.,Silva,M.,Methmini,A.,Kavinda,D.,&Thelijjagoda,S.(2021,April)Automate Traditional Interviewing Process Using Natural Language Processing and Machine Learning.In 2021 6th International Conference for Convergence in Technology IEEE.

[5].Casuat,C.D.,&Festijo,E.D.(2019,December)."Predicting students' employability using machine learning approach.In 2019 IEEE 6thInternational Conference on Engineering Technologies and Applied Sciences (ICETAS). IEEE.

[6].Moraes,R.,Pinto,L.L.,Pilankar,M.,&Rane,P.(2020,April).Personality Assessment Using Social Media for Hiring Candidates.In 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA) (pp. 192-197). IEEE.

[7]. Gorbova, J., Lusi, I., Litvin, A., & Anbarjafari, G. (2017). Automated screening of job candidates based on multimodal video processing. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops.

[8].Chen,K.,Niu,M., & Chen, Q. (2022). A hierarchical reasoning graph neural network for the automatic scoring of answer transcriptions in video job interviews.International Journal of Machine Learning and Cybernetics, 1-11.

[9].DELIMA,V.J. (2019). Impact of Personality Traits on Employees' Job Performance in Batticaloa Teaching Hospital.Iconic Research and Engineering Journals, 2(12).

[10].Adepu, Y., Boga, V. R., & Sairam, U. (2020, November). Interviewee performance analyzer using facial emotion recognition and speech fluency recognition. In 2020 IEEE International Conference for Innovation in Technology (INOCON) (pp. 1-5). IEEE.