ISSN No:-2456-2165

An AI-Driven Virtual Preparation Platform for Interviews

H A A Prabhashvara
Department of Information Technology Faculty of Graduate Studies and Research Sri Lanka Institute of Information Technology Colombo, Sri Lanka

Dr. Nuwan Kodagoda
Department of Computer Science & Software Engineering Faculty of Computing Sri Lanka Institute of Information Technology Colombo, Sri Lanka

Abstract:- These days, engaging in an interview might be difficult for many people due to the challenges like lack of social or technical capabilities, insecurity, introversion, etc. Therefore, consistent practice is strongly advised in order to solve the mentioned issues and enhance their performance as a whole. Simulated interviews are one of the greatest methods for getting ready for the interviews, but there aren't many chances to get real-world experience and objective feedback on one's skills and weaknesses before an actual interview. The developed solution enables interviewees to rehearse digitally while having their behavioral responses and answer choices evaluated in real-time, thereby attempting to address the aforementioned issues.

Keywords:- *Digital Human, Facial Feature Extraction, Artificial Intelligence.*

I. INTRODUCTION

Almost all hiring processes include an interview step so that recruiters may speak candidly with candidates, which is frequently helpful in a thorough assessment. Interviewers fre- quently assess applicants in both social and technical aspects. According to surveys, people with stronger social skills are more likely to be chosen for the job they are applying for. However, it is frequently seen that applicants with the best technical talents frequently miss out on opportunities because they lack acceptable interpersonal skills.

Furthermore, When taking part in an interview, people can encounter numerous other difficulties, such as insecurity, intro- version, etc. Continuous practice is therefore strongly advised in order to solve the aforementioned issues and enhance per- formance overall. Even while practicing for mock interviews is one of the best ways to get ready for an interview, there aren't many chances to get real-world experience and analytical feedback on one's abilities and weaknesses before an actual interview. These solutions typically fall below in providing users with a setting that is considerably more participatory and realistic. They frequently deliver queries through text-based interfaces like chatbots, completely ignoring the user's need for social communication training.

A cooperative job interview requires a variety of behavioral characteristics. Prosody, speech, and facial expressions are a few of their most distinctive behavioral characteristics. Face expressions include smiles, head gestures, and other actions. The detection, tracking, and recognition layers are the three main components of head gesture recognition. Head motions can be recognized using interfaces that can record head move- ments. In addition to these unique interfaces, computer vision technologies and deep learning techniques can be used to determine the fundamental structure of the gesture recognition notion [1]. Facial expressions also include the detection of smiles, which can be done using deep learning theories, conventional computer vision methods, and image processing technologies. Several techniques can be used to create the smile facial expression, including OpenCV, TensorFlow and Keras.

A person's speech and prosody are two additional behavioral characteristics that can be used to assess their behavior and level of confidence. These behavioral features can be assessed using a variety of properties, including speaking frequency, loudness, fillers, intonation and pauses. The speaking rate of an individual can be determined using existing approaches. Speech Synthesis Markup Language (SSML) is an XML-basedmarkup language, that can provide text annotations for speech synthesis applications. By enabling users to define volume, speed, pronunciation, pitch, and other variables via markup, SSML offers a standard method to control synthesis process components.

II. BACKGROUND STUDY

Designing an AI-driven virtual platform where users can practice for corporate job interviews requires considerationof numerous different sets of knowledge. The essential areasto explore in recent literature to establish the most optimal strategies are face feature extraction procedures, sentimentanalysis using these extracted features, and the capacity to build an interactiveness-focused virtual 3D character of an interviewer.

Facial Expression Recognition (FER) is a sophisticated but successful method of analyzing sentiment using facial features.It has been proven to be quite useful in a variety of sectors, including emotionally driven robotics, healthcare, and human- computer interaction [2] [3]. Although advancements in FER have expanded its utility, achieving high accuracy remains a difficult task. The six most prevalent human emotions are anger, grief, happiness, disgust, surprise, and fear. Moreover, hatred was recently added to the index of essential emotions [4].

The reliability of FER is totally dependent on the parame- ters utilized, which include lighting variables, age, occlusion, spectacles, and so on. To get significantly higher accuracy, these qualities must be carefully and thoroughly analyzed while creating their FER models. FER systems can be static or dynamic, depending on the picture input type. Dynamic FER combines temporal information from continuous frames, whereas static FER only uses face-point position information from a single image's feature representation. In terms of the suggested system, a dynamic FER system must be constructed while keeping sentiment analysis performance and accuracy, as well as real-time capabilities, in mind. The fundamental procedure of a FER system is depicted in Figure 1.

Fig 1 Process Flow of Static FER System

- *Pre-Processing*

This stage cleans up the dataset by noise removal and compress it. At this level, three major phases can be seen: facial detection, normalization and dimension reduction. The Viola-Jones face identification algorithm has seen considerable use in recent studies. Kar et al. [5] used this strategy to detect faces in photos, then reduced the number of duplicate features across multiple classes using Principal Component Analysis (PCA) and Linear Discrimination Analysis (LDA). This approach has been used for the same purpose by Lopeset al. [6] and Chaudhari et al. [7].

Another approach that can be used to find face features inside an image is PCA, which was used by Islam et al. [8] to exclude identical features. To limit the number of duplicate features, down sampling was used. Later, Luo et al. [9] used PCA to extract fundamental global characteristics from images that are also environment-sensitive for face expressions. To get around this, the Local Binary Pattern (LBP) technique is used to choose specific local features.

LDA and PCA are both methods for minimizing the number of dimensions in a set of data. Similar to how PCA utilizes eigenface, LDA uses fisher face to reduce the dimensionality of the features and identify the face in the image. The fisher face is typically used when the photographs have contrasting lighting. The steps required to produce the fisher face are the same as for PCA, nevertheless it appears to be better [10].

- *Feature Extraction*

Recent literature frequently uses the Gabor Filter as a facial feature extraction function. The characteristic features of a human face were identified by Ramos et al. [11] using the feature extraction function of eigenfaces. Islam et al. [8] used a Gabor Filter bank with five scales and eight orientations afterorienting at various locations. Lopes et al. [6] used a cropped image of an older face and a Garbor Filter function to reducethe overall power required for the feature extraction process. Another function that has been studied recently and has, according to the authors, shown promising results is thelocal binary pattern (LBP). LBP, a relatively new surface descriptor, converts the original pixel value of an image toa decimal number [12]– [13]. It is possible to create labels by thresholding the 3-by-3 neighborhood with a central value andtreating the resulting binary number as a number. However, thefundamental LBP was known to be highly noise-sensitive and to have problems with large structures [14] [15]. Additionally, it is impervious to the rotations and sizes of the features,which increase exponentially with the increase in the number of neighbors.

- *Emotion Classification*

One of the most widely used designs in computer vision and machine learning is the convolutional neural network (CNN). A substantial amount of data must be used for training inorder for it to fully leverage its capacity for solving difficult functions. CNN employs convolution, min-max pooling, and fully connected layers in comparison to a fully connected deepneural network. When all of these layers are combined, the entire structure is created.

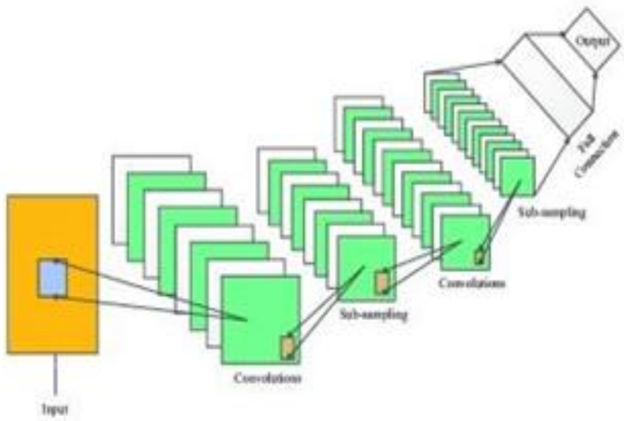

Fig 2 The Common CNN Architecture

CNN has earned recognition in the field of computer vision in recent years, particularly in the area of facial sentiment analysis. The standard of CNN was modified by researchersto perform better. An inception layer was added to a modified CNN by Mollahosseini et al. [16]. Their network architecture has two distinct components. Two conventional CNN archi- tectures, followed by ReLu, both included the Convolutional layer. These modules were followed by the parallel construc- tion of two inception layers with ReLu that were composedof 1-by-1, 3-by-3, and 5-by-5 Convolutional layers.

Kurrup et al. [17] have adopted a different strategy by employing Deep Belief Network (DBN) to fulfill their needs for face feature extraction. They suggested an emotion recog- nition method with fewer features that is semi-supervised. According to the authors, the process has become more accurate and efficient according the aforementioned method. Additionally, they were successful in accurately extracting face characteristics from the CK+, RaFD, and MMI datasets, with accuracy rates of 98.57%, 91.95%, 94.50%, 92.75%, and 98.75%, respectively.

III. METHODOLOGY

The system aims to be a combination of two integrated modules, each with certain roles and functions. These modules have been designed to operate in real time and interact without any issues. With regard to the variety of these modules, the Mi- croservice architecture will serve as the system's fundamental structure. The high-level view of said modules, as well as how they interact with the user and among themselves, is shownin figure 3 below.

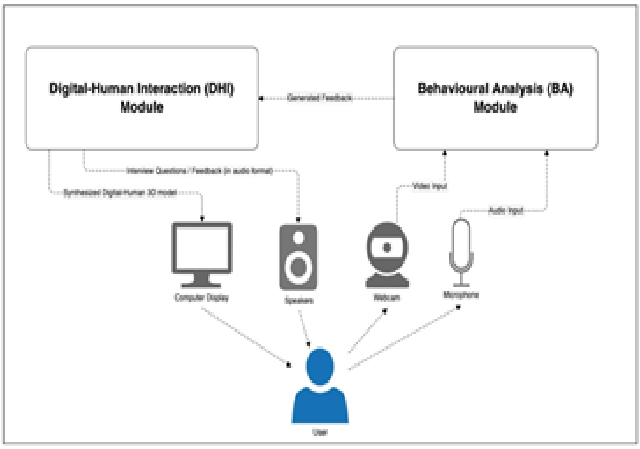

Fig 3 The high-level system architecture

➢ *Behavioral Analysis Module*

During a virtual interview session, the user's vocal and nonverbal behavior is processed and understood using this module. This analysis will be performed in real-time using inputs from web cameras and audio/video input devices like microphones. Therefore, users will be told to keep them on the entire session if it is an online interview, just like in a real job interview. Additionally, in response to the user's real-time behavioral changes, the Digital-Human Interaction Module will generate relevant feedback and transmit it to the user.

Here, the system is intended to monitor the user's typical fa- cial expressions throughout the session, such as how frequently they smile, and instances of subtle behavioral changes, such as when they appear shocked or panicked depending on the circumstance, and to provide real-time advice as necessary. According to the initial literature survey's recommendations, Shore Framework will probably be used for real-time face feature recognition. The classifier will probably be trained using example photographs and the Adaboost method, using the binary classifications of happy and neutral. Boosting will be carried out by utilizing features from various areas of the face. A normalization function that projects the score into a range between 0 and 100 is produced by the classifier. As a result, the degree of smiling on each face photograph will be evaluated on a scale of 0 to 100, with 0 denoting no smile and 100 denoting a full smile. The method will be evaluated using the Extended Cohn-Kanade (CK+) dataset, which consists of 287 images of 97 individuals from all around the United States of America. The performance of the model will be evaluated using the 62 photos from the aforementioned dataset that are categorized as showing "Happiness." To improve performance, this module will be integrated with the OpenCV Python library. We will be able to obtain the anticipated performance with the least amount of latency because this library is capable of processing real-time videos. A real-time database platform like Firebase will retain the score generated by the grin intensity classifier, which will then be utilized to generate the necessary user feedback.

When creating the applicants' meaningful feedback, au- ditory elements will also be taken into account in addition to face ones. Speech and prosody, which are grouped under acoustic characteristics, are the two main categories. Speakingrate and the number of filler words will be taken into account from those categories, respectively. We will almost real-time measure these acoustic properties using open-source software called myprosody. This library is intended primarily for si- multaneous speech and high entropy speaking to persons who speak their native language. The acoustic properties of native speech patterns have been observed and established using machine learning methods. An acoustic algorithm isolates recorded utterances and find formants, syllable boundaries, and fundamental frequency contours at a sample rate of 48 kHz and a bit depth of 32 bits. Its built-in features measure things like how many words are spoken in a minute, how many vowels are found, how many genders are recognized, etc. The score resulting from the average speaking speed and filler word usage over time will be saved in the same database as above and will subsequently be utilized to create improved feedback for the user.

➢ *Digital-Human Interaction Module*

The principal user-to-developed system communication in- terface will be implemented using this module. When required, this module will be in charge of providing the user with the interview questions and pertinent comments. An expertly curated knowledge base of typical interview questions will be used to generate the questions used in the virtual sessions. The authenticity of the user's interaction with the developed system will be enhanced by the use of a digitally constructed 3D model of a human which is referred to as the "Digital- Human" model for the remainder of this document. This will result in a setting that is much more reminiscent of a job interview. Based on the outcomes of the real-time analysis carried out by the BA module, this 3D model will be ableto give the user sufficient verbal and non-verbal reactions. To imitate and uphold the reality of virtual sessions, a background created to resemble an office will also be used.

UneeQ creator has been chosen for implementation after being thoroughly compared to other 3D modelings systems like Blender, Unity3D, Unreal Engine, and SaaS platforms like MetaHuman Creator, Soul Machines, and Hologress. The main benefits to using this for the implementation are its simplicity, the availability of a trial edition for academic use, and the possibility to use a wide range of NLP platforms like Google Dialogflow, AmazonLex, IBM Watson, or even your own conversation platform. Using UneeQ Creator, we can swiftly design, develop, and deploy our own digital beingsthat are powered by AI. They've made it exceedingly easy to change the existing data, chatbot, or other applications into something that seems more like a human interaction. The primary NLP platform will be Google Dialogflow, which will be used in conjunction with UneeQ Creator for NLPprocessing. A conversational user interface can be designed and integrated using Dialogflow into a web application, bot, device, interactive voice response system, etc. Dialogflow is capable of looking at a range of user inputs, including voice and text inputs. Additionally, it offers text or artificial speech as two ways to answer to clients. Throughout a session, Necessary Dialogflow Intends will be utilized to ask questions from a thoroughly vetted knowledge base of typical interview questions. By utilizing phrases like "How am I doing so far?," "Is there anything that can be improved?," or "Can you give me a review on my present performance?," a candidate will be able to query the digital person about his or her performance thus far in the session. Through the UneeQ platform's APIs, the Dialogflow project will get this request. This will start a specific Google Cloud function that is linked to the requisite Dialogflow Intent fulfillment webhook. The required feedback will be generated using Dialogflow's NLP models by searching the real-time database for the requisite scores received for the candidate's behavioral attributes.

➢ *Data Collection Methods*

The capability of the suggested system to advise and train its users in accordance with the expectations and observations of real corporate job recruiters is crucial to the success ofthis research. As a result, speaking with such people and evaluating their main goals and key impressions of a candidate become critical steps in the initial data collection stage. The essential data will be gathered through interviews and questionnaires, which will then be categorized and filtered to include only the most pertinent case studies. They will also be contacted regarding the typical, non-domain-specific questions they typically ask candidates. Then, the most frequently asked and important questions will be determined, and these will be utilized to build the knowledge base that the DHI module will draw from when selecting random questions.

The dataset required for the training and testing of the Machine Learning (ML) model to recognize specific facial ex- pressions can be obtained from a well-known dataset provider like Kaggle. Either the Japanese Female Facial Expression (JFFE) dataset or the Extended Cohn-Kanade (CK+) dataset will probably be used for the process because they both have abig number of photos and a clearly defined division of featureswithin them.

IV. TESTING & EVALUATION

➢ *Test Evaluation Plan*

Following a preliminary survey, 30 individuals in the soft- ware engineering sector were chosen. All 30 applicants will have a planned 15-minute interview with a professional re- cruiter, during which time they will be assessed based on pre-established criteria and given an overall score. Throughout the sessions, speaking rate, filler words, and smile (intensity) will be assessed. After the initial round of interviews, thegroup will be split into two groups of 15, each. The first group will have access to the suggested system so they can train for a week. The other group will be expected to train independently without use of the system for the same amount of time. Following the practice sessions, a second qualified recruiter will assess each applicant while the initial recruiter watches invisibly. Each candidate will receive a score from each recruiter, and the candidate's ultimate score will be determined by taking the mean of those two scores. We will be able to verify that the suggested system works as intended if the applicants who were trained using the proposed system demonstrate a better improvement of their crucial behavioral features for an interview than the other group.

➢ *Hypothesis*

A statistical analysis methodology was used to evaluate the research. I have established a hypothesis as the first step in the statistical analysis before deciding on the research's sample size, design, and sampling technique. A structured population forecast is a statistical hypothesis. Every research hypothesis is transformed into a null hypothesis that cannot be tested and an alternative that can be tested. The alternative hypothesis de-scribes the effect or relationship that the research suggests willexist, whereas the null hypothesis frequently predicts no effect or interaction between variables. In this study, the evaluation of candidate improvements using self-practice sessions and digitally assisted virtual platform practice sessions are coveredby two null hypotheses and one alternative hypothesis.

- *Hypothesis 01:*

✓ **Null hypothesis:** The behavioral attributes necessary for a corporate job interview can be trained and improved using a digitally assisted virtual platform, however, this has no impact on the candidates' improvements.

✓ **Alternative hypothesis:** A digitally assisted virtual platform that trains and improves a person's behavioral attributes necessary for a corporate job interview has an impact on the advancement of candidates.

- *Hypothesis 02:*

✓ **Null hypothesis:** Self-practice has no impact on candidates' improvements.

✓ **Alternative hypothesis:** Self-practice has an impact on ap- plicants' improvements.

ISSN No:-2456-2165

A within-subjects study is the type of design used in this experiment. The goal of the within-subjects experiment is to determine whether a digitally assisted interviewing platform can help a person develop the behavioral skills required for a corporate job interview. For this investigation, repeated measurements from one participant group are gathered.

Participants' initial baseline interview scores were noted. Following that, the candidates took part in mock interviews conducted via the system for a week. Finally, those candidates were instructed to attend a second round of interviewing, and the results of that round were recorded.

The practice interview sessions with assistance from a digital device over the course of one week are the experiment's independent variable, and the interview score both before and after the intervention is its dependent variable.

The self-practice interview sessions over the course of one week are the independent variable for experiment 02, and the interview score both before and after the intervention is the dependent variable.

Five measurement variables were used for this statistical analysis evaluation. Quantitative indicators like age, the num- ber of years of job experience, and interview score data can be employed in the computations, while categorical variables can be used to create groups for comparison testing. The measurement variables that were considered for this research are shown in the following table.

Variable	Type of Data
Gender	Categorical (nominal)
Age	Quantitative (ratio)
Working Experience	Quantitative (ratio)
Baseline Interview Score	Quantitative (interval)

Fig 4 Measuring Variables and their Data Types

The dataset was created from a sample as part of the statistical analysis' second step. A survey that was given out to software engineers in the industry was used to collect data. A minimum of 30 participants were chosen for the analysis because it is standard practice to choose at least 30 units.

The third step in statistical analysis is to summarize the data using descriptive statistics and construct the measures of central tendency. The three primary measures of central tendency that are frequently reported are mode, median, and mean. There are a number of measures of variability that may be assessed using statistical analysis, despite these ones. Some ways to quantify variability include range, interquartile range, standard deviation, and variance.

The descriptive statistics were computed after collecting pretest and posttest data from 30 candidates in the software sector.

The mean score increased following practice sessions using both the self-practice platform and the Digital-Human assisted practice interview session, according to the above table, and the variances of the two scores were comparable. The next step is to perform a statistical analysis to see whether the rise in test scores is statistically significant across the entire population.

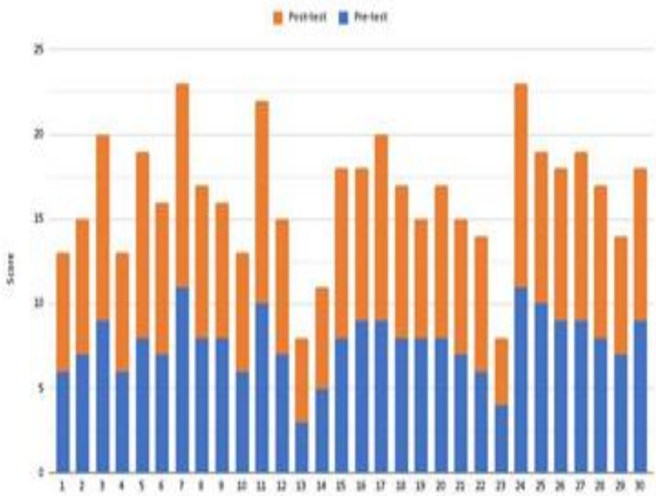

Fig 5 Score distribution of the participants

	Hypothesis 01		Hypothesis 02	
	Pretest Scores	Posttest Scores	Pretest Scores	Posttest Scores
Mean	7.27	8.67	8.13	8.67
Standard Deviation	1.98	2.16	1.68	1.84
Variance	3.92	4.67	2.84	3.38
Range	7	8	4	4
N	15		15	

Fig 6 Pretest and Posttest Calculation Results

Making estimates using inferential statistics and testing the hypothesis are the fourth and final steps of the statistical analysis. In order to determine whether the null hypothesis can be rejected, statistical tests are used to determine whether the null hypothesis is true in the population. When the null hypothesis is considered to be true, statistics show where the sample data would depend on the projected distribution of the sample data. There are two main outcomes from these tests:

- **Test statistic:** Indicates the degree to which the results depart from the null hypothesis.
- **P value:** Indicates the likelihood that the findings will betrue if the population accepts the null hypothesis.

Comparison tests, regression tests, and correlation tests are the three primary types of statistical tests. Comparison tests assess differences in outcomes between groups. While cor- relation tests examine correlations between variables without assuming cause and effect, regression tests evaluate cause and effect linkages between variables.

The dependent paired t-test has been utilized in this study to assess the findings of the hypothesis. This test is categorizedas a comparison test. In comparison tests, group means are typically compared. They could be the averages of numerous groups within a sample, the sample mean, the population mean, or the averages of a single test group gathered overtime. Among comparison tests, the t-test has a small sample size and is intended for exactly 1 to 2 groups (lesser than30 or 30). As a result, the t-test has been chosen to assess the proposed hypothesis. The subtypes of t-tests are basedon the quantity, variety, and type of samples as well as the hypotheses. The dependent paired samples test, one ofthe t-test subtypes that allows within-subjects designs, has been chosen to examine the research's hypothesis. Since the research design is a within-subjects experiment and the pretest and posttest data come from the same group, a dependent (paired) t-test is required. The need for a one-tailed test arises from the expected direction of a change, such as an increasein test results. The equation below can be used to determine the t-test value.

$$t = \frac{\Sigma d}{\sqrt{\frac{n\left(\Sigma d^2\right)-\left(\Sigma d\right)^2}{n-1}}}$$

where d: difference per paired value

n: number of samples

Fig 7 The Equation Used to Determine the T-Test Value

To assess whether the Digital-Human assisted virtual in- terview platform improved a person's behavioral qualities required for a corporate job interview, a dependent-samples, one-tailed t-test is used. The calculation yields 0.000012 as the p-value and -6.55 as the t-value (test statistic).

The research evaluation uses a dependent-sample, one-tailed t-test to ascertain whether self-learning significantly improved behavioral qualities required for a corporate job interview. Thecalculation yields a t-value (test statistic) of -2.48 and a p-valueof 0.03 for this hypothesis.

We can reject the null hypothesis and declare the data to be statistically significant because the p-value is lower when comparing p-values of 0.000012 and 0.03 to the significance level of 0.05.

This indicates that the behavioral attributes that are nec- essary for a corporate job interview can be significantly improved through both self-learning practice sessions and sessions using a Digital-Human assisted virtual interview platform.

Using two samples and test statistics, we may estimate the difference in population means using the confidence interval. The following equation can be used to determine the confi- dence interval.

$$\left(\bar{x}_1 - \bar{x}_2\right) \pm t^*\left(s_p\right)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$s_p = \sqrt{\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}}$$

$$t^* = t_{\left(\frac{\alpha}{2},df\right)}\ ,\ df = n_1+n_2-2$$

Fig 8 The Equation Used to Determine the Confidence Interval

V. CONCLUSION

The results of the literature review revealed a research gap where there are few possibilities to gain actual interviewing experience, decreasing the likelihood of gaining experience and feedback on practice interviews prior to the actual in- teractions. Even though there is some current research and some attempts have been made to solve this particular researchchallenge, we are unable to discover a realistic environment for an interview because their scenarios do not include an interviewer. This study attempted to develop a virtual platform for interview preparation that was AI-driven in order to fill the aforementioned research gap. The two main research areas are machine learning and artificial intelligence. The result of this study is a web application that users can use to sign up for and take part in virtual mock interviews related to the software development industry. The system will continue to track the behavioral traits of the interviewees and will, as needed, provide them with real-time feedback through thedigital interviewer.

The ability of a digitally assisted virtual platform to train and improve a person's behavioral qualities necessary for a corporate job interview was tested using a null hypothesisand an alternative hypothesis. The findings of the statistical analysis that was done demonstrate that the alternative hy- pothesis is the proper hypothesis in opposition to the null hy- pothesis. According to this, candidates will be improved by a digitally assisted virtual platform created to teach and develop a person's behavioral attributes necessary for a corporate job interview.

REFERENCES

[1] N. Simon, "How does ai recognise your hand signs, gestures and movements?" Aug 2019. [Online]. Available: https://mantra.ai/blogs/ ai-in-gesture-recognition/

[2] J. Hathaliya, P. Sharma, S. Tanwar, and R. Gupta, "Blockchain-based remote patient monitoring in healthcare 4.0," in *2019 IEEE 9th Inter- national Conference on Advanced Computing (IACC)*, 2019, pp. 87–91.

[3] J. Vora, P. DevMurari, S. Tanwar, S. Tyagi, N. Kumar, and M. S. Obaidat, "Blind signatures based secured e-healthcare system," in *2018 Inter- national Conference on Computer, Information and Telecommunication Systems (CITS)*, 2018, pp. 1–5.

[4] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363–368, 1992.

[5] N. B. Kar, K. S. Babu, A. K. Sangaiah, and S. Bakshi, "Face expression recognition system based on ripplet transform type II and least square SVM," *Multimedia Tools and Applications*, vol. 78, no. 4, pp. 4789– 4812, 2019.

[6] N. Lopes, A. Silva, S. R. Khanal, A. Reis, J. Barroso, V. Filipe, and J. Sampaio, "Facial emotion recognition in the elderly using a svmclassifier," in *2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW)*, 2018, pp. 1–5.

[7] M. N. Chaudhari, M. Deshmukh, G. Ramrakhiani, and R. Parvatikar, "Face detection using viola jones algorithm and neural networks," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–6.

[8] B. Islam, F. Mahmud, and A. Hossain, "Facial expression region segmentation based approach to emotion recognition using 2d gabor filter and multiclass support vector machine," in *2018 21st International Conference of Computer and Information Technology (ICCIT)*, 2018, pp.1–6.

[9] Y. Luo, C. ming Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of pca and lbp with svm," *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 17, pp. 2767– 2770, 2013. [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S0030402612006560

[10] R. Samanta, B. Sadhukhan, H. Samaddar, S. Sarkar, C. Koner, and M. Ghosh, "Scope of machine learning applications for addressing the challenges in next-generation wireless networks," *CAAI Transactions on Intelligence Technology*, vol. 7, pp. n/a–n/a, 06 2022.

[11] A. L. A. Ramos, B. G. Dadiz, and A. B. G. Santos, "Classifying emotion based on facial expression analysis using gabor filter: A basis for adaptive effective teaching strategy," in *Computational Science and Technology*. Singapore: Springer Singapore, 2020, pp. 469–479.

[12] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[13] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037– 2041, 2006.

[14] I. Revina and W. S. Emmanuel, "A survey on human face expression recognition techniques," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 6, pp. 619–628, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/ S1319157818303379

[15] d. Huang, C. Shan, M. Ardabilian, and L. Chen, "Local binary patterns and its application to facial image analysis: A survey," *IEEE Transac- tions on Systems, Man, and Cybernetics, Part C*, vol. 41, pp. 765–781, 11 2011.

[16] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[17] A. Rajendra Kurup, M. Ajith, and M. Martínez Ramón, "Semi- supervised facial expression recognition using reduced spatial features and deep belief networks," *Neurocomputing*, vol. 367, pp. 188– 197, 2019. [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S0925231219311579

[18] M. Guo, X. Hou, Y. Ma, and X. Wu, "Facial expression recognition using elbp based on covariance matrix transform in klt," *Multimedia Tools and Applications*, vol. 76, 01 2017.

[19] M. Courgeon and C. Clavel, "Marc: A framework that features emo- tion models for facial animation during human–computer interaction," *Journal on Multimodal User Interfaces*, vol. 7, no. 4, p. 311– 319, 2013.

[20] N. H. de Jong and T. Wempe, “Praat script to detect syllable nuclei and measure speech rate automatically,” *Behavior Research Methods*, vol. 41, no. 2, p. 385–390, 2009.

[21] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, “Mach: My automated conversation coach,” *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013.

[22] Y. Jadoul, B. Thompson, and B. de Boer, “Introducing parselmouth: A python interface to praat,” *Journal of Phonetics*, vol. 71, p. 1–15, 2018.

[23] A. Kumari, S. Tanwar, S. Tyagi, and N. Kumar, “Fog computing for healthcare 4.0 environment: Opportunities and challenges,” *Computers & Electrical Engineering*, vol. 72, p. 1–13, 2018.

[24] Y. Luo, C.-m. Wu, and Y. Zhang, “Facial expression recognition based on fusion feature of pca and lbp with svm,” *Optik - International Journal for Light and Electron Optics*, vol. 124, no. 17, p. 2767–2770, 2013.

[25] W. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity (1943),” *Ideas That Created the Future*, p. 79–88, 2021.

[26] V. Takala, T. Ahonen, and M. Pietikäinen, “Block-based methods for image retrieval using local binary patterns,” *Image Analysis*, p. 882–891, 2005.

[27] S. Witt and S. Young, “Phone-level pronunciation scoring and assess- ment for interactive language learning,” *Speech Communication*, vol. 30, no. 2-3, p. 95–108, 2000.

[28] K. Zechner, D. Higgins, X. Xi, and D. M. Williamson, “Automatic scoring of non-native spontaneous speech in tests of spoken english,” *Speech Communication*, vol. 51, no. 10, p. 883–895, 2009.