International Journal of Emerging Engineering and Technology (IJEET)
ISSN (e): 2958-3764
Volume: 2, Number: 1, Pages: 110- 119, Year: 2023

SELF ASSESSMENT INTERVIEW BOT

Shereen Gul[1], Sumaira Gul[2], Muhammad Sheraz Gul[3], Mirza Adnan Baig[4] and Shahbaz Younis[4]

[1]Govt. Graduate College for Women Rahim Yar Khan, 64200, Pakistan
[2]Govt Girls Higher Secondary School 1/P Khanpur Distt. Rahim Yar Khan, 64200, Pakistan
[3]The University of Lahore, Pakistan
[4]The Islamia University Bahawalpur Rahim Yar Khan Campus, 64200, Pakistan

Corresponding author: First Author (e-mail: shereengul.it@gmail.com).

Abstract- We live in the age of automation, and businesses have made great strides in recent years toward fully mechanizing their industrial processes. In the proposed research, we use AI to help us carry out responsibilities and jobs normally performed by humans, and we advise automating the required procedures for both the industry and the people involved. These days, candidates often try to have their interview prep done in advance. Those unfamiliar with the process turn to books and multiple-choice questions (MCQs) online to help them get ready. Nevertheless, they still need to figure out what questions are acceptable and how to pass the interview. In this modern age, our proposed study adds to the human self-evaluation gap between interview-related emotions and question responses. The self-interview bot analyses the candidate's previous interview footage and provides a synopsis of his interview preparation in this work. Emotion classification in the proposed study was done using Fear-2013, a publically available dataset. In preparation for the next interview, people can take a pre-test; the results summary may include issue monitors for applicants, depending on the test. We can analyze the video effectively using CNN for image processing and NLP for semantic analysis of the candidates' interview questions and answers. Using natural language processing for semantic similarity and convolutional neural networks (CNNs) for image processing, we were able to build a model that accurately answers queries and uses them. Finally, we can assess the candidate's readiness for the next interview by combining the two sets of results. Future research findings and the AI bot's current assessment of the applicant. The person-in-the-loop evaluation method involves a human assessor who converses with the bot and gives it comments about how it did. The evaluator can accomplish this by posing a series of questions to the bot or starting a dialogue with it, after which the bot will be evaluated based on its responses and overall performance.

Keywords: Machine learning, NLP, CNN, automation, AI, self-assessment.

I. INTRODUCTION

One of the techniques that organizations can use to perform competency assessments is a behavioural event interview (BEI). BEI is a systematic interviewing method carried out in five stages: introducing and explaining the interview process, extracting information about job responsibilities, extracting information about behavioural events, exploring the characteristics required for the job, and finally, drawing conclusions and summaries. In contrast to ordinary interviews, considered less reliable in predicting candidates who will perform well, BEIs can reveal detailed behavioural descriptions of how someone does their job and thus overcome the ineffectiveness of typical interviews. However, although BEI is a suitable method for measuring competencies, it also has some weaknesses. For example, the interview process takes longer, requires many certified interviewers—trained experts are still limited in number—requires high organizing costs, and introduces potential interviewer bias in the assessment. Meanwhile, the need to conduct competency assessments increases in line with the growing awareness of conducting competency assessments in companies and organizations. In conducting BEIs, interviewers look for clues about the interviewee's past experiences. In English, these can be identified from past-tense responses. However, Indonesian grammar is simple in the sense that it has no past-tense pattern. Instead, it uses keywords that indicate the time or frequency of one's actions. Therefore, developing an interview bot that "speaks" Indonesian presents a different challenge and requires a different approach. The fourth Industrial Revolution has impacted automation technology development in various fields, and technological advances are being integrated with every aspect of our lives [1] Artificial intelligence (AI) facilitates decision-making, creates integrative systems, and simplifies complex mechanisms through automation[2] The interview bot application for assessing competency levels is a further development of chatbot technology. A chatbot is a computer system that operates as an interface between human users and software applications, using natural written and oral language to communicate. Some examples of chatbots that have been developed are Siri, IBM Watson, and Google Assistant. A chatbot has several advantages, including ease of access, efficiency, availability, scalability, cost, and insight. Chatbot technology has been applied in various fields, such as handling e-commerce queries [3].

II. MOTIVATION:

This study focuses on exploring fundamental concepts related to the proposed study, which revolves around self-assessment in the context of interview preparation. The central challenge identified is individuals' difficulty objectively evaluating their body language, emotional expressions, confidence levels, and knowledge gaps when gearing up for job interviews. To address these challenges, the study suggests leveraging interview assessment tools, particularly those powered by automation and chatbots. By automating the interview process, organizations can evaluate candidates based on their skills, cultural fit, emotional intelligence, and other relevant criteria, ultimately aiding in data-driven employment decisions. The chapter outlines the potential of chatbots in automating initial candidate interactions, collecting applicant data, screening candidates, providing information, and maintaining continuous communication. The proposed study aims to contribute to human self-assessment by incorporating emotion analysis and question-reply interactions before interviews, utilizing candidates' prior interview footage and a self-assessment interview bot to provide summarized preparation feedback. The overarching goal is to offer candidates a pre-assessment for upcoming interviews and highlight specific areas for improvement, ultimately enhancing the effectiveness of the interview process.

In summary, the study introduces the challenges in self-assessment during interview preparation and proposes using automated interview assessment tools, specifically chatbots, to address these issues. The study emphasizes the potential benefits of leveraging technology to streamline the hiring process and improve candidates' interview readiness through innovative self-assessment methods.

III. RESEARCH PROBLEM

In the journey towards landing that dream job, just like you, we crave a self-assessment tool that acts as our trusted ally before stepping into the actual interview battleground. Imagine having the power to evaluate and prepare yourself thoroughly, understanding your emotional strengths and weaknesses, boosting confidence, and fine-tuning your domain knowledge. The reality is that many candidates need help with emotional resilience and need more confidence in their expertise. They yearn for a solution that goes beyond the surface, a way to uncover and address their deficiencies in real time. Enter the proposed study to revolutionize the pre-interview game by providing individuals with a comprehensive self-assessment mechanism, ensuring they step into interviews well-prepared and armed with the confidence to shine. Let's explore how this study seeks to be the game-changer in the realm of interview self-analysis.

IV. STATEMENTS OF OBJECTIVES

The objectives of the proposed study are mentioned below:

- To provide the solution for the candidate self-assessment bot for interview.
- To provide the solution to test his/her emotional deficiencies.
- To provide the solution to test his/her domain knowledge related to his/her interview.
- To provide the mechanism for the candidates and the organizations to access the self/candidate assessment.
- To provide the solution an organization analyzes the domain knowledge of a candidate
- To provide the solution to an organization to analyze candidate emotions.

V. RESEARCH QUESTION

1. How are emotions collected from the model?
2. How do you get emotions from video content with audio and text?
3. How do we evaluate candidates from the system-generated results?

VI. LITERATURE

Innovative approach to hiring using artificial intelligence and machine learning, aiming to enhance efficiency and reduce bias in the process. The proposed system leverages AI-powered photorealistic bots to conduct one-on-one interviews, evaluating candidates' skills, experience, confidence, and body language. By automating the initial screening and assessment, this approach can save time and lead to better-quality hires. The paper discusses this method's potential benefits and ethical considerations, offering a promising solution to transform the hiring landscape in the era of technological advancement.[4] ChatGPT, an AI chatbot released by OpenAI, and its role in advancing user-friendly, versatile AI chatbots capable of engaging in personalized conversations. The transition from earlier, more limited AI chatbots to ChatGPT's capabilities is discussed, highlighting the role of human feedback in its development.

The author engages in a point-counterpoint conversation with ChatGPT to discuss the potential applications of AI in translational medicine, providing valuable insights into the healthcare sector. The paper also presents an AI-generated artwork inspired by Andy Warhol, created using DALL-E 2, showcasing the creative possibilities of AI technology. Overall, this article offers a comprehensive look at the impact and potential of AI in both medical and artistic domains.[5] human-bot collaboration within software development, specifically in the context of GitHub pull requests. It identifies that developers generally prefer personable bots with limited autonomy. Notably, more experienced developers tend to lean towards more autonomous bots. The study recommends offering increased configuration options for bots to cater to individual developer preferences and adapt to diverse project cultures. This paper sheds light on crucial aspects of bot behaviour that can impact developer satisfaction, contributing to a more productive and harmonious collaboration in software development [6].

AI in the early detection of depression, a critical area in behavioural health. It underscores the pressing need to address barriers like cost, location, stigma, and a shortage of mental health professionals that leave a substantial population undiagnosed. The proposed solution advocates for a mass screening integrative approach driven by AI to identify individuals with depression at an early stage. Such an approach is essential in preventing potential mental health crises and ensuring timely intervention. The paper emphasizes the significance of leveraging AI to bridge this gap in behavioural health interventions, ultimately improving the well-being of individuals in need.[7] Small-scale automation services, or SE Bots, have become integral in daily software development, aiming to enhance productivity and alleviate the burdens faced by elite developers. This paper builds upon prior research in BotSE and interviews with elite developers to explore the design and implementation of SE bots tailored to their workflows and expectations. It outlines six key design guidelines addressing concerns such as noise, security, and simplicity, crucial for accommodating elite developers' unique needs. Additionally, the paper delves into the future prospects of SE bots, emphasizing their role in alleviating the mounting workload of elite developers in response to increasing demands in the software development landscape.[8] students' adoption intention (ADI) and actual usage (ATU) of AI-based teacher bots (T-bots) in the educational context. It utilizes the Technology Adoption Model (TAM) and incorporates context-specific variables to explore the factors influencing students' acceptance and effective utilization of these AI-driven teaching assistants. The study contributes valuable insights into the dynamics of AI technology adoption in the realm of education, shedding light on the critical determinants that shape students' decisions to engage with and benefit from AI-enhanced learning.

By applying TAM and considering context-specific variables, this research provides a comprehensive framework for understanding the adoption of AI-based teacher bots in the context of modern education. [9] Interview Bot is a novel system designed to conduct hybrid-domain interviews with foreign students applying to U.S. colleges and evaluate their academic and cultural readiness. The model is built upon a neural-based end-to-end dialogue system trained on a substantial dataset of interviews. To address the limitations of transformer-based models, the paper proposes innovative techniques—context attention and topic storing—to ensure coherent and relevant interactions. The final model undergoes rigorous testing, showcasing high fluency and context awareness levels. It offers a promising and efficient solution for enhancing the interview process in U.S. college admissions, catering to the needs of both students and institutions and opening avenues for further developments in the dynamic conversation integration [10]. Ladderbot is a text-based conversational agent designed to conduct online laddering interviews to uncover user goals and underlying values. Traditional laddering interviews are valuable but resource-intensive, prompting the development of Ladderbot to make them more scalable. An experimental study involving 256 participants compares Ladderbot to established survey-based laddering methods in exploring user values related to smartphone use. Results indicate that CA-based laddering interviews conducted by Ladderbot yield twice as many and longer answers on average, with higher learnability than survey-based approaches. However, survey-based laddering reliably uncovers value-related ladders, while CA-based laddering delves into negative gains and attribute-consequence-value structures. This research introduces a new CA-based approach and suggests the potential for combining survey and CA-based laddering methods for a more comprehensive understanding of user values [11]. The issue of cheating in university assessments is aggravated by the widespread availability of generative AI tools like ChatGPT. The research conducts a "quality assurance" exercise by comparing student scripts with ChatGPT-generated scripts for end-of-module assessments across a university's computer science curriculum. Surprisingly, in the majority of cases, ChatGPT-generated scripts perform on par with or even outperform undergraduate scripts, regardless of question formats, topics, or study levels. This study underscores the disruptive impact of generative AI on quality assurance processes and highlights the necessity for more stringent measures to combat academic cheating in the age of AI-driven solutions [12]. landscape of chronic heart failure (CHF) management, highlighting the distressing and disabling symptoms that persist despite optimal treatments. It underscores the pivotal role of patients in adhering to complex self-care regimens, encompassing maintenance, monitoring, and management activities. The study emphasizes that disciplined adherence to these regimens is essential for symptom control and improving patients' quality of life. However, it also addresses the under-recognized issue of non-adherence, exploring the intricate interplay between symptom burden and self-care. This paper acknowledges the need to better understand the factors contributing to non-adherence and provides valuable insights into improving CHF management, aiming for enhanced patient outcomes and quality of life [13].

VII. METHODOLOGY

The Interview Bot incorporates video and audio processing for detecting human emotions and expressions during interviews. The primary objective is to design and build a highly accurate model, utilizing it to predict emotions and expressions from interview videos. Figure 1.0 illustrates the step-by-step process of utilizing the dataset to train and save the AI model for the Interview Bot application, outlining a clear roadmap to achieve the proposed study's goals.

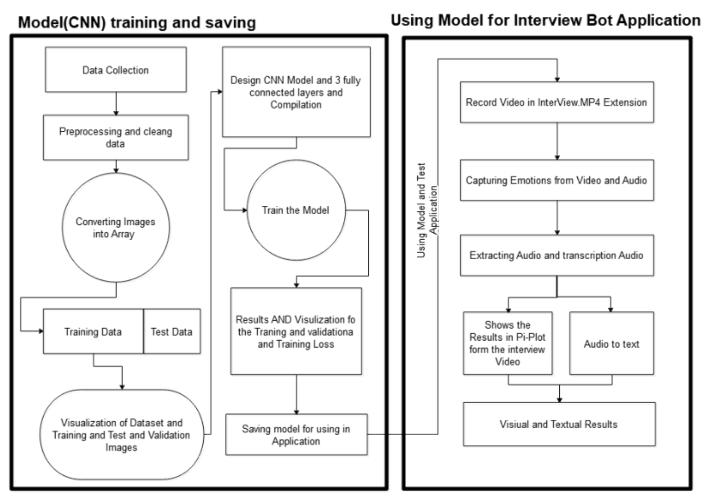

Figure 1.0 *Methodology Diagram*

Diagram Description:

The initial phase involves dataset collection and preprocessing, where data augmentation is performed to enhance image datasets. The images are then converted into array data stored in a CSV file, as depicted in Figure 4.13. The subsequent step employs Convolutional Neural Networks (CNN) for feature extraction within the Deep Learning model, enabling emotion classification based on each image. Following model training, the classification results are obtained. The subsequent stage incorporates video interview processing, saving, and answer extraction. Emotions from the video are extracted, utilizing a Deep Learning model for prediction, providing a percentage breakdown of each emotion throughout the interview. Additionally, audio answers are extracted, converted to text using Google API, and assessed for semantic similarity against pre-written HR answers, yielding a percentage score. The final step involves visualizing results, as shown in Figure 5.6, to evaluate both HR and Bot assessments and gauge the effectiveness of the Interview Bot. This comprehensive process ensures a thorough evaluation of candidate performance and the Bot's effectiveness in comparison to HR standards.

Processes Model Description:

The proposed study unfolds in two pivotal phases: firstly, it constructs the model utilizing Convolutional Neural Networks (CNN) for effective feature extraction and accurate class prediction based on the trained data. Upon successfully saving the model, the second phase focuses on practical application. The saved model is deployed for real-time testing on interview videos. This structured approach ensures a meticulous progression, emphasizing initial model creation followed by its targeted utilization in the subsequent phase for real-world scenarios, particularly in the analysis of interview videos.

Dataset Description:

The dataset utilized in this study is sourced from Kaggle, specifically the publicly available "Human Emotion Collection Dataset." Named fer2013.csv, the dataset comprises images capturing seven distinct emotions. Displayed in array form, the data is stored in a comma-separated file.

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training

Figure 2.0 *Dataset View*

Figure 2.0 offers a snapshot of the dataset, revealing three attributes or columns. The first column denotes the emotion or facial expression, the second column represents the pixel values corresponding to that emotion, and the third column categorizes the usage of the image into three distinct labels. This organized dataset forms the foundation for the subsequent model-building and training phases in the proposed study.

Emotions in Dataset:

The dataset encompasses seven distinct expressions, each revealing a unique facet of human emotion. These emotions, namely Natural, Angry, Happy, Surprised, Sad, Fear, and Disgusting, are inherently self-explanatory. For instance, a candidate exhibiting a "Natural" emotion during an interview signifies a genuine and authentic demeanor. Similarly, emotions like "Angry," "Happy," and "Surprised" convey their intuitive meanings, providing a comprehensive spectrum of human responses. Sourced from the widely recognized platform Kaggle, this freely available dataset serves as a rich repository for understanding and analyzing emotional states. Emotions, as encapsulated in this dataset, serve as valuable indicators of our sentiments during conversations and interviews. By leveraging this resource, the proposed study aims to empower individuals in self-assessment, enabling a deeper understanding of their emotional responses and fostering personal growth in communication and interview scenarios.

Visualization of Dataset:

The dataset exhibits a diverse range of emotions, and their distribution is visually depicted in Figure 3.0. Among the seven emotions, 'Happy' emerges as the most prevalent, boasting the highest number of instances. Following closely, 'Natural' claims the second-largest share in the dataset, while 'Surprise' secures the third position. The visualization further reveals that the 'Happy' emotion attains the maximum frequency with a count of 9000, signifying its prominence. On the other end of the spectrum, 'Disgusting' holds the minimum count, standing at 500, thus representing the least prevalent emotion in the dataset. This insightful breakdown of emotional values within the training data sets the stage for the proposed study's subsequent modelling and analysis phases.

Figure 3.0 *Emotion Contributions.*

Training and Test Dataset:

Training and the test data is used to train the model and test the model for this in the figure 4.0 show that the training and test images in the dataset.

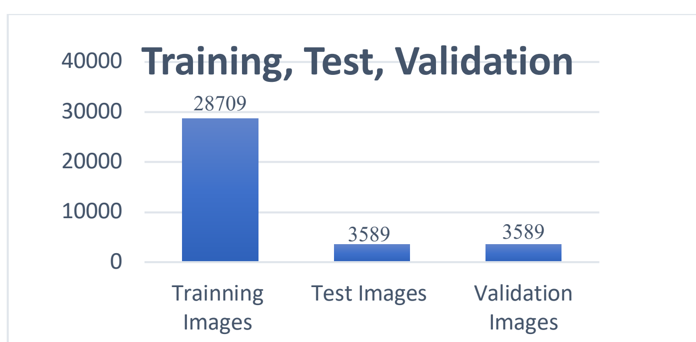

Figure 4.0 *Training and Test data values*

Building CCN model for training:
The model construction process involves the utilization of Convolutional Neural Networks (CNN) alongside MobileNetV3 for effective feature extraction. Detailed insights into this process are elaborated in the upcoming section.

Description of CNN:
A Convolutional Neural Network (CNN) stands out as a specialized neural network particularly adept at image classification tasks. The term "convolutional" is derived from its utilization of the convolution operation, a mathematical process that amalgamates small data chunks, such as image pixels, to generate a more intricate output. The crux of convolution lies in the employment of filters—compact weight matrices that traverse the input data, conducting a dot product at each position. Subsequently, the dot product undergoes activation through a designated function, determining the significance of the output as a noteworthy feature. In essence, a single convolutional layer in a CNN can be succinctly described through a mathematical equation, encapsulating the essence of this powerful image-processing network.

Mathematical representations of CNN:

Convolution operation:
Given an input volume with dimensions W1 x H1 x D1, a convolutional layer employs F filters of size FH x FW with a stride S, applying the following operation to each subregion of the input. This results in an output volume characterized by dimensions W2 x H2 x D2, as detailed by Equation 1:

$$W2 = \frac{W1 - FW}{S + 1}$$
$$H2 = \frac{H1 - FH}{S + 1}$$
$$D2 = F$$

Equation 1 encapsulates the mathematical representation of the convolutional layer's operation, illustrating how the shape of the output volume is influenced by the input and Output filters. The actual computation is denoted by:

$$[i,j,f] = \sum Input[s:s+FH, t:t+FW, :] * Weight[f,:,:,:] + Bias[f]$$

Here, (i, j) denotes the index of the output volume, (s, t) represents the index of the input volume, and f signifies the index of the filter. This equation provides a comprehensive understanding of the convolutional layer's intricate mathematical underpinnings.

Pooling operation:
The pooling layer operates by applying a specific operation to every W2 x H2 subregion of the input, utilizing a window size of W2 x H2 and a specified stride, resulting in an output volume characterized by dimensions W3 x H3 x D3. As expressed in Equation 2, when no padding is employed, the output dimensions remain unchanged, aligning with the objective of preserving the size in this particular scenario.

$$W3 = (W2 - W2) / S + 1$$

$$H3 = (H2 - H2) / S + 1$$
$$D3 = D2$$

Equation 1: *Equation of the layer's size.*

Output [i, j, k] = max (Input [s: s+W2, t:t+H2, k]) where (i, j) is the index of the output volume, (s, t) is the index of the input volume, and k is the index of the channel.

Fully-connected layer:
In the context of a fully-connected (FC) layer, comprising M neurons and receiving an input of size N, the operational procedure is delineated by Equation 3. This equation illustrates that within a fully connected layer, the input undergoes multiplication by a weight matrix, followed by the addition of a bias vector.

$$Output[i] = sum(Input[j]) * weight[i, j] + Bias[i]$$

where i is the index of the output neuron and j is the index of the input neuron. Equation.4 shows layers are followed by one or more fully connected layers. As the name suggested, all neurons in the fully connected layer connect to all the neurons in the previous layer

$$O[i, j] = f(\sum\{k = 0\}\hat{}\{n - 1\} \sum\{1 = 0\}^{\{m-1\}I}[i + k + J + l] + b[i, j])$$

In this context, O represents the output derived from the convolutional layer, with I being the input, W serving as the filter matrix (also known as the kernel or weights), b representing the bias term, and f denoting the activation function. The output, O, is generated by applying the filter W to the input I, incorporating the bias term b, and subsequently subjecting the result to the activation function f. Indices i and j are utilized for the output matrix, while k and l are employed for indexing the filter matrix. The size of the output, O, is contingent upon the dimensions of the input I, the filter size W, and the stride (indicating the number of pixels the filter moves at each step). If the stride is greater than 1, the output will be smaller than the input, whereas a stride less than 1 will result in a larger output compared to the input.

Image Processing by Using CNN:
Convolutional Neural Networks (CNNs) excel in image classification tasks due to their ability to automatically acquire hierarchical representations of images. This is facilitated by convolutional layers, where a convolution operation is applied to the input image, discerning patterns and features across different scales.

Certainly, here's a revised version:

1. Preprocess Input Images:

- Resize images to a standardized dimension.
- Convert images to a designated color space.

- Normalize pixel values.

2. Define CNN Architecture:

- Determine the number of convolutional and pooling layers.
- Specify filter and pooling window sizes.
- Select activation functions.
- Define the neuron count in fully connected (FC) layers.

3. Train the CNN:

- Feed preprocessed images and corresponding labels to the CNN.
- Utilize an optimization algorithm like stochastic gradient descent (SGD) to iteratively adjust weights, minimizing classification errors.

4. Test the CNN:

- Evaluate CNN performance using a separate test set.

5. Utilize CNN for Classification:

- For classifying new images, leverage the CNN to predict class probabilities.
- Choose the class with the highest probability as the final prediction.

Figure 5.0 illustrates a comprehensive overview of CNN utilization in image classification.

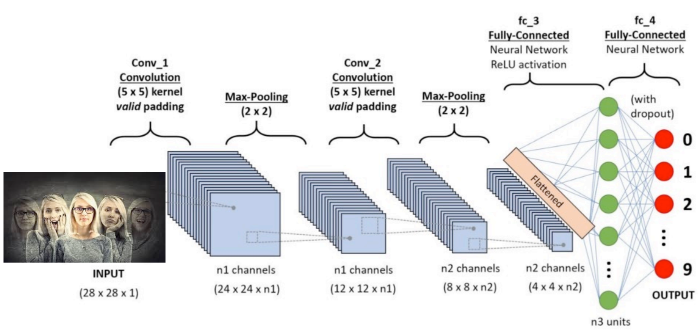

Figure 5.0 *Architectural diagram of CNN*

Steps for Image Classification in CNN (Convolutional Neural Network):
1. Preprocess Input Images:

- Resize images to a fixed size.
- Convert images to a specified color space.
- Normalize pixel values.

2. Define CNN Architecture:

- Choose the number of convolutional and pooling layers.
- Specify filter and pooling window sizes.
- Select activation functions and determine the number of neurons in fully connected (FC) layers.

3. Load Input Data and Labels:

- Use data loading libraries like TORCHVISION to retrieve data from a dataset or file.

4. Convert Data and Labels to Tensors:

- Utilize libraries such as numpy for converting data and labels to tensors.
- Use torch library to ensure tensors are in the appropriate data type.

5. Define Loss Function and Optimizer:

- Choose a loss function, e.g., cross-entropy loss.
- Select an optimizer, e.g., stochastic gradient descent (SGD), to optimize CNN weights.

6. Train the CNN:

- Feed input tensors to the CNN with corresponding labels.
- Use the optimizer to update weights and minimize classification error.

7. Test the CNN:

- Evaluate CNN performance on a separate test set.

8. Utilize CNN for Classification:

- Predict class probabilities for new images.
- Select the class with the highest probability as the final prediction.

Building CNN Model:
Layer-1:

In the initial layer construction, a 2D convolutional operation is applied to the input image using 64 filters. The activation function employed is rectified linear unit (ReLU), followed by max pooling with a pool size of 2 and a stride of 2.

Layer-2:

In the subsequent layer, a grayscale image is processed using 128 filters through a 2D convolution. The activation function remains rectified linear unit (ReLU).

Layer-3:

The third layer is configured with 256 filters, applying a rectified linear unit activation function to grayscale images.

Output Layer:

The output layer represents the classification into 7 emotions, utilizing the softmax activation function.

Compilation of Model:
During the model compilation, a learning rate of 0.0001 is specified, and the chosen metric for evaluation is the accuracy of the model.

History of the Model:
The model is compiled with a batch size of 64 and trained for 40 epochs. Consideration is given to the possibility of increasing epochs to 60 or 80, but this might extend the processing time, and there's a point where the validation accuracy could plateau or even decrease. The dataset is divided into train and test sets, denoted by Train_X and Train_Y.

$$\text{Trian\_X} = \text{train\_images}$$
$$\text{Train\_Y} = \text{train\_labels}$$

Equation 5 illustrates the correspondence between the names of the train and test datasets. The training dataset comprises 28,709 images and labels, while the test dataset consists of 3,589 images and labels. Additionally, there are validation images (3,589) and labels (3,589). A summarized overview of the compiled CNN model is presented in Figure 6.0.

Summary of the complied model:

```
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 46, 46, 64)        640

 batch_normalization (BatchN  (None, 46, 46, 64)       256
 ormalization)

 conv2d_1 (Conv2D)           (None, 44, 44, 64)        36928

 batch_normalization_1 (Batc  (None, 44, 44, 64)       256
 hNormalization)

 max_pooling2d (MaxPooling2D  (None, 22, 22, 64)       0
 )

 dropout (Dropout)           (None, 22, 22, 64)        0

 conv2d_2 (Conv2D)           (None, 20, 20, 128)       73856

 batch_normalization_2 (Batc  (None, 20, 20, 128)      512
 hNormalization)

 conv2d_3 (Conv2D)           (None, 18, 18, 128)       147584
...
Total params: 1,938,887
Trainable params: 1,937,095
Non-trainable params: 1,792
```

Figure 6.0 *Model summary of the CNN*

Emotional Model results:
Model training starts at 64% accuracy but gradually declines to 40 epochs. Corrections in dataset, methods, or parameters are necessary. Figure 7.0 depicts a model accuracy of 64%, with a corresponding drop in validation accuracy from 63%, declining after 25 epochs. Due to time constraints, this accuracy is accepted for now, and focus shifts to the next steps in model initialization. Figure 7.0 illustrates the training and validation loss, indicating a suboptimal validation loss of 2.0. Efforts will be made to address and reduce this after model deployment.

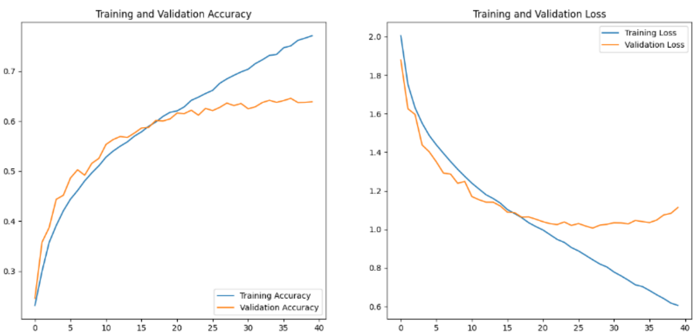

Figure 7.0 *Model Training and Test Curves.*

Classification report of the model:
In the Jupyter Notebook, a classification report is generated, utilizing precision, recall, and F1-Score to evaluate the model. Figure 8.0 displays this report. Analyzing the classification outcomes, "Happy" and "Surprised" classes exhibit accuracy surpassing 70%, while the remaining classes require enhancement for improved accuracy. The algorithm achieves an overall performance of 63%.

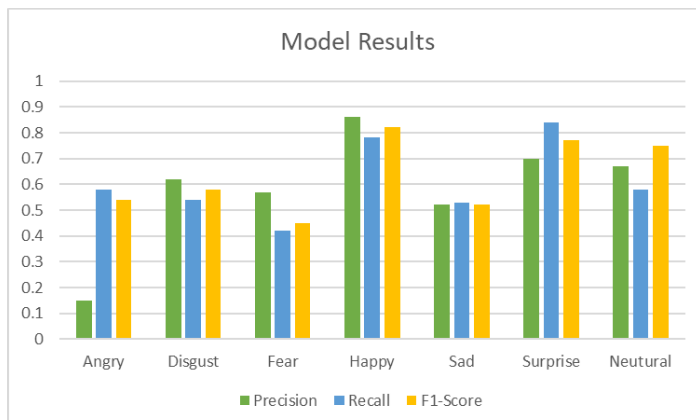

Figure 8.0 *Classification report of the model*

Deployment of the model as App:
The model is deployed with the existing accuracy. A demo interview video is utilized, incorporating Google Speech recognition for voice transcription. The video is captured, stored in an array, and the overall results after video processing are presented in Figure 9.0.

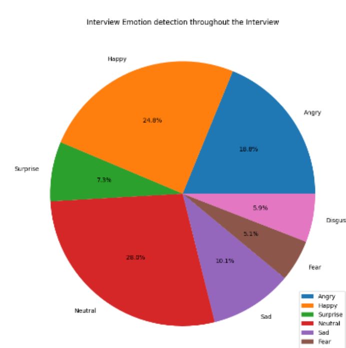

Figure 9.0 *Demo interview video emotion results*

The analysis of the interview with the candidate reveals that the predominant emotion exhibited is natural, followed by happiness as the second most prominent, and surprisingly, anger as the third emotion. This anomaly raises concerns as anger is not an ideal emotion for a candidate during an interview. It's important to note that these results are based on a demo video, and with a model accuracy of 63%, there is room for improvement. Recognizing the need for enhancement, we plan to focus on refining the model in future iterations to ensure more accurate emotion detection in interview scenarios.

Textual semantic similarity:
Analyzing questions and answers involves utilizing a Python package to assess the semantic similarity between the interviewer's and candidate's responses. Figure 10.0 provides a comprehensive overview of the textual semantic similarity between the content of the two text files, as described earlier.

```
File Answers.txt :
2581 lines,
375 words,
194 distinct words
File Answers copy.txt :
2334 lines,
360 words,
179 distinct words
The distance between the documents is:  0.088023 (radians)
```

Figure 10.0 *Semantic similarity of the two text files.*

Steps for the analysis of Candidate Answers during an interview:

a) Provide responses to the predefined interview questions for Interview Bot.

b) Extract the audio from the available interview video.

c) Assess the semantic similarity between both text files.

d) Utilize the similarity score to grade the answers. Combining both reports, the candidate's emotional activity and answer scores will aid in evaluating skills and confidence.

Figure 11.0 illustrates a comparative view between the text file containing the actual answers and the text file containing the candidate's responses.

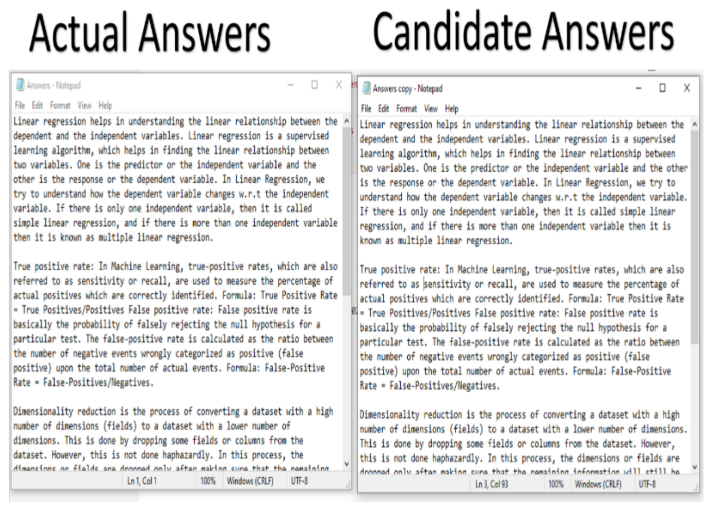

Figure 11.0 *View of the Actual answers text file and candidate answers text file*

RESULT & DISCUSSION

Model Training Accuracy and Validation Results:
The model's training accuracy experiences a decline from 64% and reaches 40 epochs, indicating a need for adjustments in the dataset, methods, or parameters. The model currently attains an accuracy of 64%, and the validation accuracy drops from 63%, persisting below the 64% mark after 25 epochs. Given time constraints, we acknowledge the current accuracy and proceed to the next steps for model initialization. Figure 12.0 displays the training and validation loss, revealing a validation loss of 2.0. Post-deployment, efforts will be made to mitigate and reduce this loss.

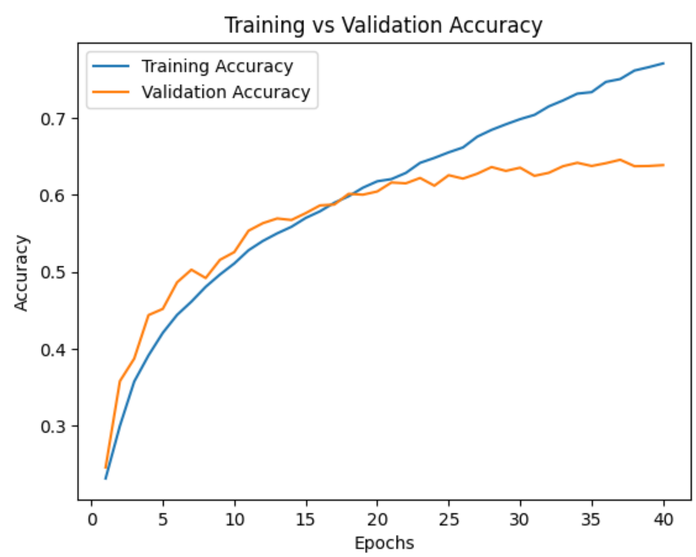

Figure 12.0 *Results of Training and validation of the model*

Figure 13.0 depicts a graph illustrating the training process of the model. Initially, the training accuracy is consistently at 100%, showcasing a strong start to the training. However, in the test accuracy, the model's performance is optimal until around epoch 20, after which there is a gradual decline, impacting the overall test accuracy. This decline in accuracy indicates a potential effect on the model's overall performance.

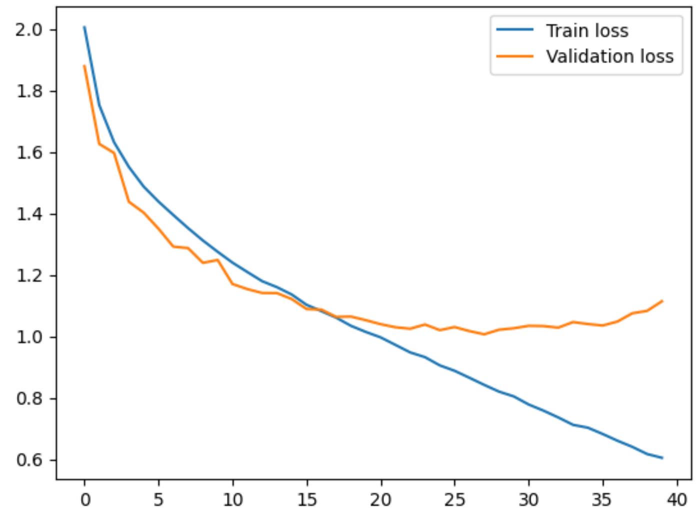

Figure 13.0 *Results Training and Validation Loss*

The training loss remains consistently low, contributing to the high training accuracy. Conversely, in the test phase, the loss starts increasing from epoch 15 and continues to rise until the completion of the epochs. This rise in test loss further influences the overall effectiveness of the model.

Our proposed method offers emotion activity tracking and incorporates a feature for analyzing the relevance of candidate answers to the actual answers. This study aligns with the current age and contributes to human self-assessment in terms of emotions and question responses before an interview. Leveraging prior interview footage of candidates, the system processes videos and provides a summarized overview of their interview preparation. Figure 5.6 illustrates multiple candidates across various domains undergoing interviews conducted by human resource personnel, with percentages assigned to each candidate, similar to the Interview Bot.

Table 1.0 Multiple candidates Test results.

C.No	Angry	disgust	Fear	happy	sad	surprise	natural	SSAS	SSAS %	HR ASMNT
Candidate-01	1%	5%	10 %	25 %	12 %	10 %	37 %	0.18	18 %	30%
Candidate-02	2%	2%	15 %	24 %	5 %	19 %	33 %	0.88	88 %	90%
Candidate-03	5%	6%	12 %	26 %	2 %	18 %	31 %	0.59	59 %	65%
Candidate-04	3%	3%	13 %	23 %	3 %	15 %	40 %	0.89	89 %	80%
Candidate-05	4%	4%	10 %	25 %	6 %	6%	45 %	0.87	87 %	80%
Candidate-06	4%	8%	14 %	27 %	4 %	13 %	30 %	0.88	88 %	90%
Candidate-07	2%	4%	12 %	24 %	3 %	7%	48 %	0.59	59 %	50%

The graph **Figure 14.0** shows the comparison of the actual Human resource person assessment and the Bot assessment.

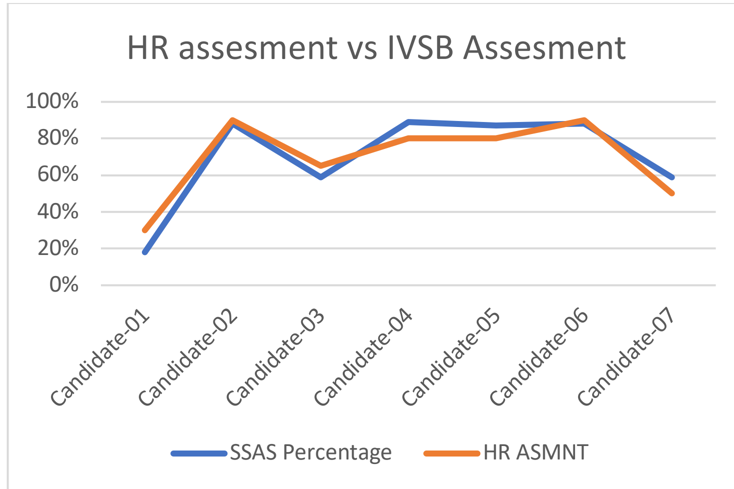

Figure 14.0 *HR and Bot assessment*

CONCLUSIONS

In the collection of grayscale images for the dataset, it is observed that the dataset contains sufficient data points. However, during model training and validation, it is noted that while 3 out of 7 classes are correctly classified, 4 classes exhibit low precision-recall and F1-Score, requiring improvement. The validation loss, reaching 2.0, is primarily attributed to these 4 classes with low classification accuracy.

The study, focusing on emotion detection from interview videos, initially constructs a model from the dataset. It is acknowledged that certain classes face challenges in classification due to low training accuracy. Future efforts will concentrate on enhancing model accuracy. Subsequently, the model is applied to a demo interview video, yielding overall emotion results that form a summary for candidate assessment. This summary informs decisions for the next interview based on emotional outcomes.

The semantic similarity analysis of questions and answers aids in evaluating candidates for subsequent interviews. The combination of emotional results and semantic similarity scores enhances confidence in assessing candidates' skills and confidence levels.

For future enhancements, several methods can be explored, such as improving dataset values and implementing cross-validation on the CNN and PCA. The proposed solution proves beneficial for HR companies and aids candidates in self-assessment before interviews, fostering confidence through repeated practice.

Future Work:

Future endeavors involve embedding the artificial intelligence model into a mobile app for production deployment, focusing on accuracy and productivity improvement. Automation for HR companies with cloud-based requirement systems is envisioned, allowing for simultaneous interviews with multiple candidates.

REFERENCES

[1] M. Berawi, "The Fourth Industrial Revolution: Managing Technology Development for Competitiveness," *Int. J. Technol.*, vol. 9, p. 1, Jan. 2018, doi: 10.14716/ijtech.v9i1.1504.

[2] M. A. Berawi, "Managing Artificial Intelligence Technology for Added Value," IJTech - International Journal of Technology. Accessed: Nov. 03, 2023. [Online]. Available: https://ijtech.eng.ui.ac.id/article/view/3889

[3] C. Pricilla, D. P. Lestari, and D. Dharma, "Designing Interaction for Chatbot-Based Conversational Commerce with User-Centered Design," in *2018 5th International Conference on Advanced Informatics: Concept Theory and Applications (ICAICTA)*, Aug. 2018, pp. 244–249. doi: 10.1109/ICAICTA.2018.8541320.

[4] V. Narkhede, O. Surushe, S. Kulkarni, H. Solanki, T. Ekbote, and D. Joshi, "AVA: A Photorealistic AI Bot for Human-like Interaction and Extended Reality," in *2023 International Conference on Communication System, Computing and IT Applications (CSCITA)*, IEEE, 2023, pp. 5–9. Accessed: Nov. 01, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10104672/

[5] D. L. Mann, "Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine," *JACC Basic Transl. Sci.*, vol. 8, no. 2, pp. 221–223, Feb. 2023, doi: 10.1016/j.jacbts.2023.01.001.

[6] A. Ghorbani *et al.*, "Autonomy is an acquired taste: Exploring developer preferences for GitHub bots," in *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, IEEE, 2023, pp. 1405–1417. Accessed: Nov. 01, 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10172895/

[7] P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao, and B. Gu, "Early detection of depression using a conversational AI bot: A non-clinical trial," *Plos One*, vol. 18, no. 2, p. e0279743, 2023.

[8] Z. Wang, Y. Wang, and D. Redmiles, "Optimizing Workflow for Elite Developers: Perspectives on Leveraging SE Bots." arXiv, Apr. 28, 2023. Accessed: Nov. 01, 2023. [Online]. Available: http://arxiv.org/abs/2304.14828

[9] R. Pillai, B. Sivathanu, B. Metri, and N. Kaushik, "Students' adoption of AI-based teacher-bots (T-bots) for learning in higher education," *Inf. Technol. People*, 2023, Accessed: Nov. 01, 2023. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/ITP-02-2021-0152/full/html

[10] Z. Wang, N. Keyes, T. Crawford, and J. D. Choi, "InterviewBot: Real-Time End-to-End Dialogue System for Interviewing Students for College Admission," *Information*, vol. 14, no. 8, p. 460, 2023.

[11] T. Rietz and A. Maedche, "Ladderbot—A conversational agent for human-like online laddering interviews," *Int. J. Hum.-Comput. Stud.*, vol. 171, p. 102969, 2023.

[12] M. Richards, K. Waugh, M. Slaymaker, M. Petre, and D. Gooch, "Bob or Bot: Exploring ChatGPT's answers to University Computer Science Assessment," *ACM*, p. Pre-Print, 2023.

[13] "Qualitative interviews results from heart failure survey respondents on the interaction between symptoms and burden of self-care work," *J. Clin. Nurs.*, vol. 32, no. 15–16, pp. 4649–4662, Aug. 2023, doi: 10.1111/jocn.16484.