Advancing Virtual Interviews: AI-Driven Facial Emotion Recognition for Better Recruitment

Rohini Mehta[1]; Pulicharla Sai Pravalika[2]; Bellamkonda Venkata Naga Durga Sai[3]; Bharath Kumar P[4]
Ritendu Bhattacharyya[5]; Bharani Kumar Depuru[6*] (ORC ID: 0009-0003-4338-8914)
[1]Research Associate, Aispry, Hyderabad, India.
[2]Research Associate, Aispry, Hyderabad, India.
[3]Research Associate, Aispry, Hyderabad, India
[4]Research Associate, Aispry, Hyderabad, India
[5]Team Leader, Research and Development, Aispry, Hyderabad, India.
[6]Director, Aispry, Hyderabad, India

Corresponding Author:- Bharani Kumar Depuru[6*]

Abstract:- Behavior analysis involves the detailed process of identifying, modeling, and comprehending the various nuances and patterns of emotional expressions exhibited by individuals. It poses a significant challenge to accurately detect and predict facial emotions, especially in contexts like remote interviews, which have become increasingly prevalent. Notably, many participants struggle to convey their thoughts to interviewers with a happy expression and good posture, which may unfairly diminish their chances of employment, despite their qualifications. To address this challenge, artificial intelligence techniques such as image classification offer promising solutions. By leveraging AI models, behavior analysis can be applied to perceive and interpret facial reactions, thereby paving the way to anticipate future behaviors based on learned patterns to the participants. Despite existing works on facial emotion recognition (FER) using image classification, there is limited research focused on platforms like remote interviews and online courses. In this paper, our primary focus lies on emotions such as happiness, sadness, anger, surprise, eye contact, neutrality, smile, confusion, and stooped posture. We have curated our dataset, comprising a diverse range of sample interviews captured through participants' video recordings and other images documenting facial expressions and speech during interviews. Additionally, we have integrated existing datasets such as FER 2013 and the Celebrity Emotions dataset. Through our investigation, we explore a variety of AI and deep learning methodologies, including VGG19, ResNet50V2, ResNet152V2, Inception-ResNetV2, Xception, EfficientNet B0, and YOLO V8 to analyze facial patterns and predict emotions. Our results demonstrate an accuracy of 73% using the YOLO v8 model. However, we discovered that the categories of happy and smile, as well as surprised and confused, are not disjoint, leading to potential inaccuracies in classification. Furthermore, we considered stooped posture as a non-essential class since the interviews are conducted via webcam, which does not allow for the observation of posture. By removing these overlapping categories, we achieved a remarkable accuracy increase to around 76.88% using the YOLO v8 model.

Keywords:- *Behavior Analysis, Computer Vision, YOLO, Keras, Tensorflow, Image Classification.*

I. INTRODUCTION

Analyzing people's behavior [1], especially through facial reactions, is essential for insight into human interaction and communicative processes playing a vital role in situations like remote interviews, where in-person interaction is absent. Correctly analyzing affective responses delivered through facial actions in such cases is a considerable challenge. Adding to the complexity is the expectation in scenarios such as job interviews for individuals to project happiness and self-assurance, qualities that may not be readily apparent via facial cues alone on a digital platform. Misreading these specific facial expressions might result in misunderstandings and inequitable outcomes.

Artificial intelligence, especially methodologies such as image identification [2] [3], can assist in addressing this issue. AI models can learn to comprehend facial cues that can help in the prediction of individuals' future behavior based on their emotional states [4]. While some research has already made progress in recognizing emotions through images, there has been limited focus on contexts like remote interviews and online courses, where emotions such as happy, sad, smile, surprise, anger, eye contact, and neutrality are really important.

In this paper, we present our efforts to bridge this gap by focusing on the nuanced emotions crucial in remote interviews and online learning environments. We employed the project methodology following the open-source CRISP-ML(Q) methodology from 360DigiTMG [Fig.1]. CRISP-ML (Q) [5] encompasses Cross Industry Standard Practice for Machine Learning with Quality Assurance and is specifically crafted to guide the project lifecycle of Machine Learning solutions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

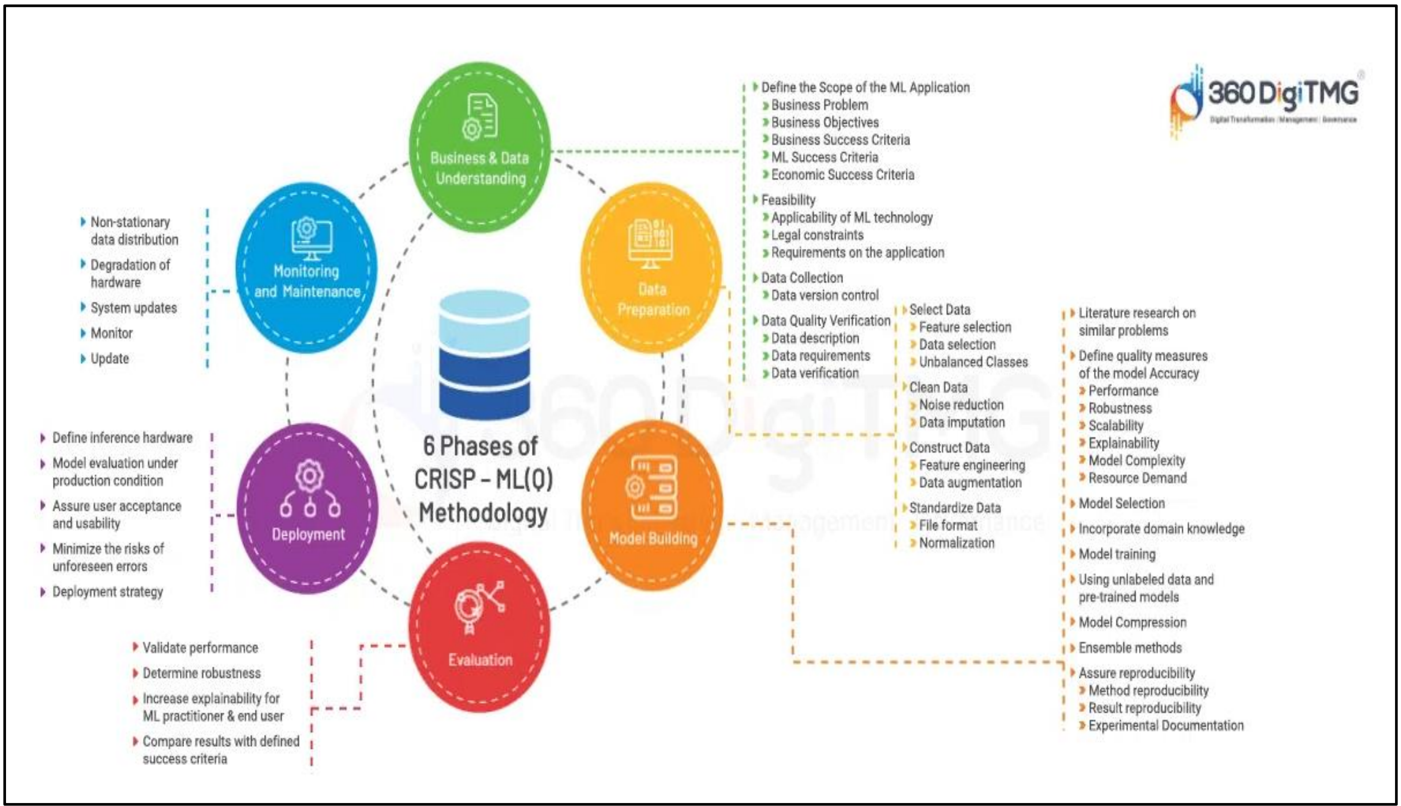


Fig 1 Show Cases the CRISP-ML (Q) Methodological Framework, Offering a Visual Breakdown of its Pivotal Elements and Sequential Procedures.
Source: - Mind Map 360DigiTMG

We have curated an exhaustive dataset combining sample interview recordings and images capturing facial gestures and posture. Additionally, we have integrated existing datasets such as FER 2013, IIITM faces emotion dataset, ISAFE dataset, and the celebrity emotions dataset to enrich our analysis. Post-data collection, we meticulously validated the balance of image class datasets to ensure parity in their respective counts.

Thereafter, we partitioned the complete dataset into 3 parts: train, validate, and test, in a ratio of 60:20:20 [6] [7]. We employed 7 pre-trained models: VGG19 [8][9][11][13], ResNet50V2 [10][11], ResNet152V2 [12], Inception ResNetV2 [13], Xception [12][14], EfficientNet B0 [15], and YOLO [16] classification and recorded their accuracies. We have adjusted hyperparameters as per the requirements and compared the accuracies of the models. After intensive testing, our findings demonstrate the effectiveness of our approach, achieving an accuracy of 73% using the YOLO v8 model. We found that the categories “happy and smile” and “surprised and confused” overlap, causing classification errors. Additionally, since posture can’t be assessed via webcam, we removed “stooped posture” as a category. Eliminating these overlapping categories improved our YOLO v8 model’s accuracy to about 77%. Furthermore, we deployed the model using Streamlit.

This research improves our grasp and delivers valuable advice for dealing with emotional factors in digital interactions, like virtual interviews for recruitment.

II. METHODS AND TECHNOLOGY

A. System Requirements:

Table 1 Data Processing was Performed using a Combination of Local and Cloud Resources, Including the Specified OS, Hardware, and Software Configurations.

Resource	Specification
OS	Windows 10 & above
RAM / GPU	16 GB / 16 GB
Server/Cloud instance	EC2 Instance, Instance Type - g4dn.xlarge
Storage	100 GB
IDE	Jupyter, Spyder & Google Colab

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

In various scenarios, we processed our data using a robust set of resources. Our operating system of choice was Windows 10 and above, ensuring compatibility and reliability. We utilized a system equipped with 16 GB of RAM and a 16 GB GPU to handle intensive computational tasks efficiently. For cloud-based processing, we relied on an EC2 Instance, specifically the g4dn.xlarge type, which provided the necessary scalability and performance. Data storage needs were met with a capacity of 100 GB, ample for our extensive datasets. For development and data analysis, we employed a combination of IDEs, including Jupyter, Spyder, and Google Colab, which offered flexibility and powerful tools for coding and visualization. This setup enabled us to manage and process data effectively across different scenarios.

B. Model Architecture:

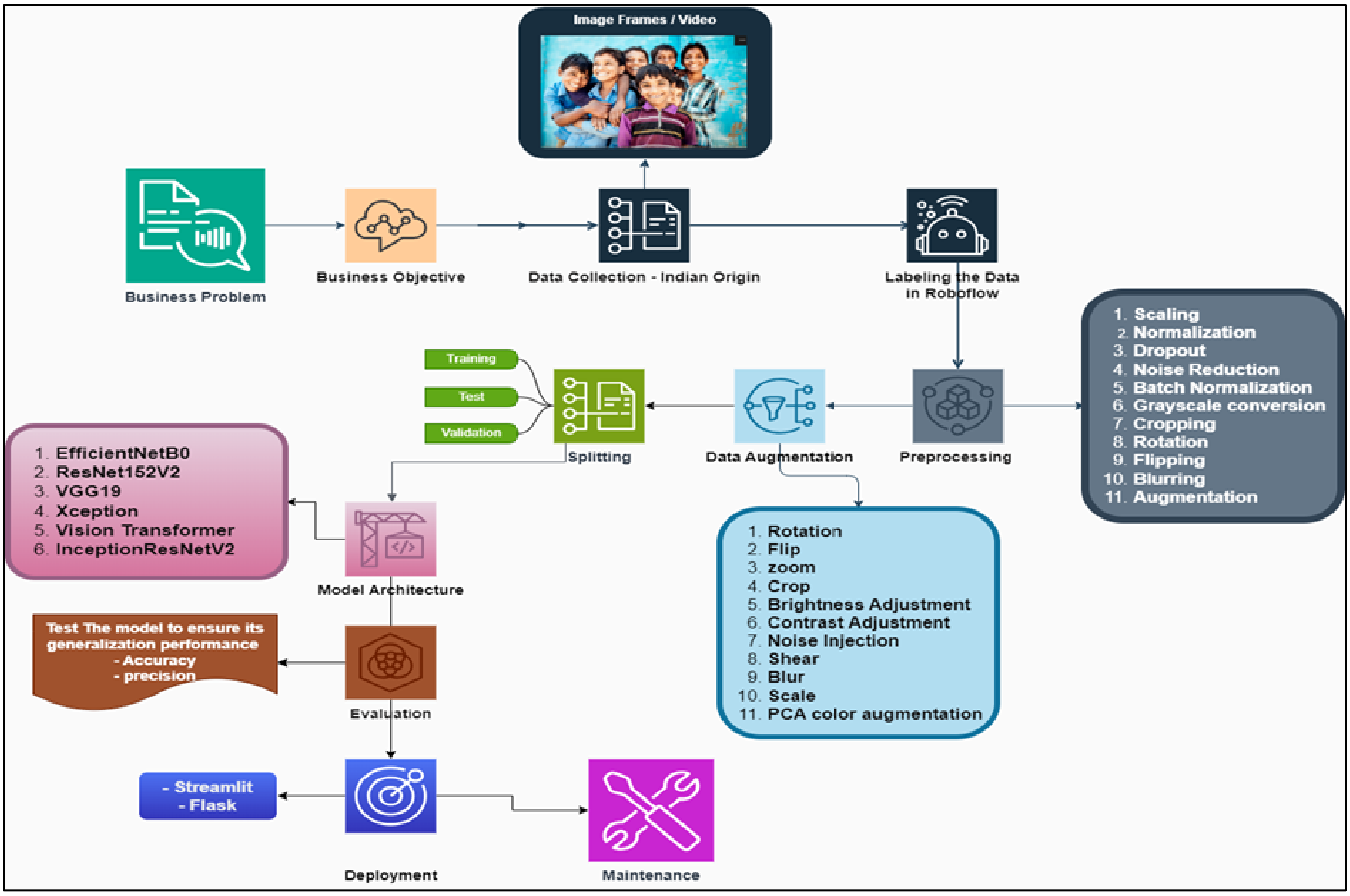


Fig 2 Deep Learning Architecture for Categorizing Images by Emotional Expressions.

CRISP-ML(Q) [Fig.1] serves as the mainstay of the study, hence, getting started with comprehending the problem and its objective. Let's delve into a detailed interpretation of the project architecture depicted in [Fig.2].

During interviews, it's vital to look beyond candidates' technical prowess. Despite impressive credentials, some may lack the spark of charisma and confidence, which could dampen their potential. Lackluster interactions may dim their appeal to employers, thereby diminishing their prospects in the job market. This emphasizes the need for comprehensive evaluations to gauge candidates' overall capabilities.

To tackle the mentioned challenge, the business aims to develop a system capable of analyzing candidates' non-verbal signals, including facial expressions, posture, and eye contact, throughout the interview process. Through the utilization of machine learning methods, this system seeks to offer insights into candidates' overall demeanor and communication abilities, assisting interviewers in making well-informed hiring decisions. Data collection involves extracting images from video recordings of interview sessions, which are subsequently annotated and labelled to identify pertinent non-verbal cues. Preprocessing techniques [18] are then employed to standardize and improve the quality of the image data, while data augmentation strategies are implemented to diversify the dataset, accommodating various interview scenarios and environments.

Following the dataset's division into training, testing, and validation sets, the focus shifts to crafting and evaluating a resilient machine-learning model. This involves thorough research into computer vision algorithms capable of adeptly recognizing and interpreting non-verbal cues from image data. The model's selection hinges on this exploration.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

Finally, the evaluation of the models involves rigorous testing to ensure they accurately interpret non-verbal cues and provide reliable insights for interviewers.

Let's delve into the detailed description of each step.

- *Problem Overview and Objective:*

During interviews, numerous participants convey their thoughts to the interviewer with a poised smile, impeccable posture, and attentive eye contact. However, despite their worthiness, the presence of a subdued, melancholic, or apprehensive demeanor often diminishes their likelihood of securing the job.

If someone acts bored, unhappy, or worried during interviews, it might make interviewers think they're not confident or excited, even if they're good at the job. Interviewers might think these behaviors mean the person isn't very motivated, isn't good with people, or just isn't the right fit for the job.

So, the study aims to enhance the probability of securing employment while at the same time lessening the mental stress associated with not being successful. To accomplish this, facial expressions were detected by extracting frames from video footage recorded during interviews.

- *Gathering and Comprehending Data:*

Facial cues are paramount in the study. Data were sourced from platforms like YouTube, Instagram, Facebook, etc. Image frames extracted from these sources were then categorized according to different facial cues using Roboflow [17], facilitating their organization into distinct classes for analysis.

Table 2 The Challenges Faced while Collecting/Understanding the Data.

Challenge	**Solution**
Data collection specific to Indian origin	Research, requested to authors for data.
Correct labeling of data	Re-validating
Quality of the images	low-quality images were discarded
Quantity for some classes was less as compared to other	collected more data specific to the classes

We met several encounters [Table.2] while building their dataset. First, gathering data specific to Indian origin was tough, so they reached out to authors and did extensive research to compile the necessary information. Next, we had to ensure the data was correctly labeled, which involved a meticulous re-validation process. We also encountered many low-quality images, which they chose to discard to maintain high standards. This led to some classes having fewer images than others, so focused on collecting more data for those specific classes. Through determination and innovative solutions, a robust and balanced dataset was successfully created.

- *Data Preprocessing.*

We have collected 76,703 images in total. To further step of data pre-processing [18], we were very much clear about the data imbalance. To ensure that the data are balanced, we applied different augmentation techniques [19] to increase no. of images in each class with an equal split of images in all the classes. Our target classes are happy, sad, angry, surprise, neutral, eye contact, smile, confused, and stooped posture. Each class has a different set of data as in visual [Fig.3]. With this no of images in each class, we will be facing model bias and inaccurate recognition of class during the prediction process. Also, during model building these imbalanced data would favor the highest class and resulting inaccurate prediction. So always the balanced dataset will represent equal learning patterns across all targeted classes and features which in turn gives us the best class prediction for our model building.

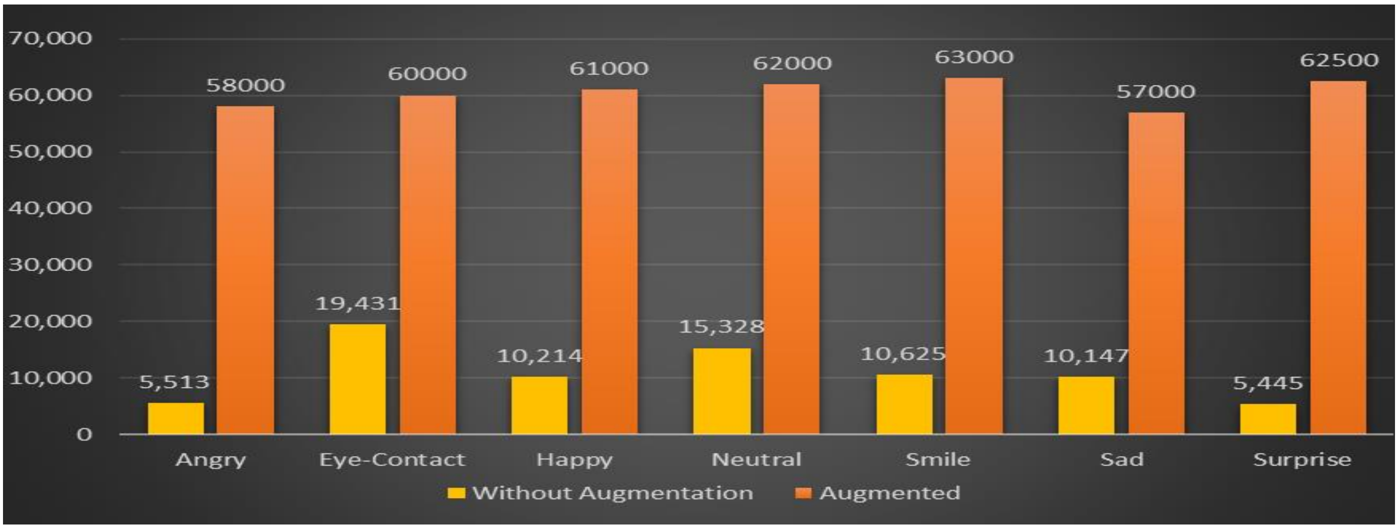


Fig 3 Visual Shows No. of Images before and after Augmentation, Imbalanced & Balanced Data

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

The visual [Fig.3] portrays the no. of images associated with each class.

After we finalized the target classes for our study, we came across a challenging phase during the analysis of each and every class Emotions like happiness & smiles show us a similar pattern of expressions. confused & surprised emotions coincide with each other. So, to tackle this kind of scenario, we narrowed it down to a happy & surprise instead of happy, smile, confused, and surprise.

Once all the primary data was ready, different augmentation techniques were used:

Rescaling: 51 x 51 mm: No need to code as already taken care in the roboflow.

- Shift: width_shift_range and height_shift_range dictate the scope of potential horizontal and vertical image movement, akin to a dance across the canvas of pixels.
- Shear Range: shear_range introduces a playful twist to the image as if it were a sheet of paper being gently skewed along to its axis.
- Brightness Range: brightness_range plays with the luminosity of the images, bringing forth a spectrum of moods within the specified brightness spectrum.
- Preprocessing Function (Adding Noise): preprocessing_function injects a touch of chaos into each image, akin to a sprinkle of stardust, by applying a bespoke function. Here, it conjures random noise into the images using NumPy.
- Channel Shift Range: channel_shift_range whimsically shuffles the color palette of the images, as if each picture is exploring a new chromatic dimension.

Once the augmentation technique step was completed, we normalized all images and scaled-down all pixel values to a standard range between 0 to 1 to build the best model for class prediction.

Table 3 Dataset Dimension

Attribute	Details
Image format	.jpeg; .jpg; .png
No. of Classes	6
Class name	Angry, Sad, Neutral, Surprise, Eye-Contact, Happy
Total images before pre-processing	76,703
No. of images after Augmentation	3,58, 928
No. of Images in each class	Angry: 59,846; Eye-Contact: 60,582; Happy: 59,932; Neutral: 61,102; Sad: 58,022; Surprise: 59,444

The details of dataset [Table.3] comprises images in various formats (.jpeg, .jpg, .png) categorized into six classes. Post-augmentation, the dataset expanded significantly, ensuring balanced representation across all classes.

Moving forward, the dataset was then subdivided into 3 partitions i.e., train, valid, & test. While doing so, the sub-divided ratio was 60% in train 20% in test, and 20% in valid [6][7]. Also here in this phase, we kept 20 of the test datasets as unknown to come up with the model to avoid data breaches to derive insights.

➢ *Model Building:*

After analyzing the statistics and studying pertinent papers, we tested seven distinct models: VGG19, ResNet50V2, ResNet152V2, Inception-ResNetV2, Xception, EfficientNet B0, and YOLO V8.

- *VGG19*

VGG19 [8][9][11][13] is a sophisticated convnet equipped with pre-trained layers that possess a comprehensive grasp of image attributes such as shape, color, and structure. Having undergone training on millions of diverse images across intricate classification tasks, VGG19 boasts significant depth and expertise.

- *ResNet50V2:*

ResNet50V2 [10][11], stands as a groundbreaking evolution in deep learning architecture, elevating the achievements of the original ResNet model. Through its enhanced depth, resilient connections, and bottleneck blocks, it amplifies performance and efficacy across diverse computer vision assignments, rendering it an indispensable asset embraced by both pioneers and professionals in the field.

- *ResNet152V2*

ResNet152V2's [12] extensive layering empowers the extraction of intricate image features, rendering it adept at tackling sophisticated visual recognition challenges. By incorporating skip connections, it mitigates the vanishing gradient issue, easing the training burden on deep neural networks. Due to its resilience and adaptability, this model finds widespread usage across research and industry realms in diverse computer vision applications.

- *Inception-ResNetV2*

InceptionResNetV2 [13] represents a major leap in deep learning architecture by blending the strengths of Inception modules and ResNet networks. This groundbreaking hybrid architecture demonstrates exceptional proficiency in managing diverse and demanding tasks while guaranteeing streamlined training, rendering it a Transfigurative asset within the realm of visual perception.

- *Xception*

Xception [12][14], a breakthrough from Google, enhanced the convnet structure to improve productivity and act within visual computing assignments. By employing depth-wise separable convolutions, it extends the capabilities beyond those of the Inception architecture. This distinctive methodology enriches the model's ability to comprehend intricate nuances and details while still preserving computational efficiency, thereby establishing it as a potent resource for endeavors like visual recognition and entity localization.

- *EfficientNetB0*

Google's EfficientNetB0 [15] is notable for striking a fine balance for achieving both accuracy and computational efficiency in tasks like image classification within computer vision. As one of the model alternatives, it employs intelligent scaling techniques to optimize performance rendering it a preferred option for deployment across various platforms and scenarios.

- *Yolo V8*

YOLO [16] models have been a game-changer in Computer Vision, offering unmatched performance. With YOLOv8, this trend continues, providing swift and accurate object recognition in images, perfect for applications such as self-driving cars and surveillance. Its advancements mark a notable progression in the field.

➢ *Model Evaluation:*

During our pursuit, we faced numerous challenges, especially concerning data collection, as highlighted in [Table.2]. Overcoming these obstacles proved vital in crafting a resilient and effective model for our project.

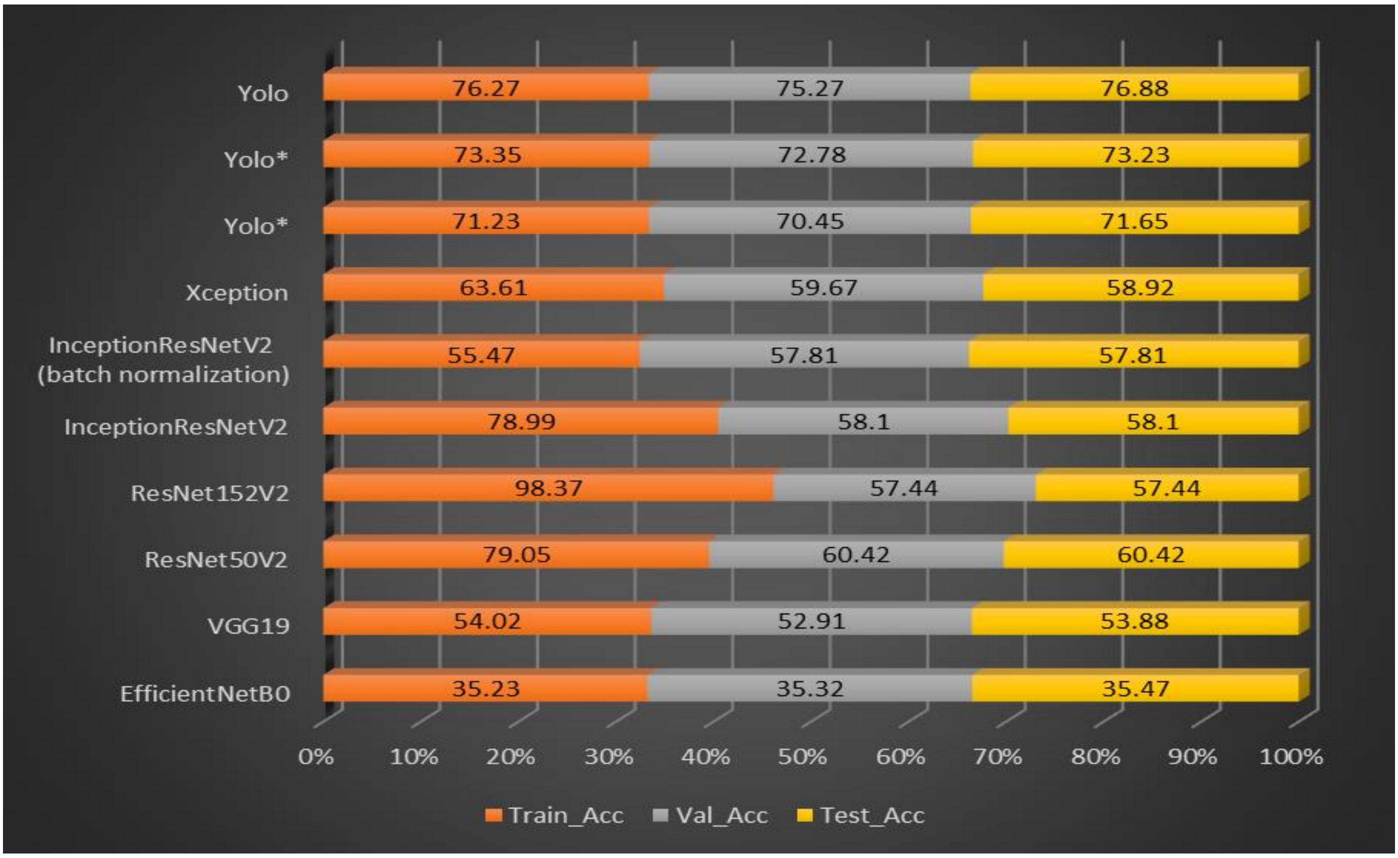


Fig 4 Chart Illustrating the Accuracies of Diverse Models.

Throughout our investigation, we embarked on a journey of continuous model refinement to elevate the precision of emotion recognition within remote interviews and online learning environments. Initially, employing the ResNet50V2 and ResNet152V2 models with a data split of 98:1:1, we achieved commendable training accuracies of 57% and test accuracies of 53% [Fig.4]. However, the presence of overfitting, where the models excelled on the training data but struggled with unseen data, prompted a strategic shift. Transitioning to a 60:20:20 data split yielded enhancements, with test accuracies rising to 58% and 66% for the respective models[Fig.4]. Despite these improvements, overfitting persisted, leading us to explore alternative avenues. Our foray into the YOLOv8 model resulted in a modest accuracy of 69%[Fig.4]. Undeterred, we pursued further optimization, leveraging 70% of the data with the same data split, culminating in a notable accuracy surge to 73%*[Fig.4]. Additionally, by recognizing the overlap between emotions like happy and smile, and between surprised and confused, and excluding stooped posture due to the nature of webcam interviews, we achieved an impressive accuracy of 76.88% using the YOLOv8 model[Fig.4]. To ensure seamless user interaction and accessibility, we integrated our model deployment into the Streamlit framework, enhancing user experience and facilitating effortless engagement for candidates.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

III. RESULTS AND DISCUSSION

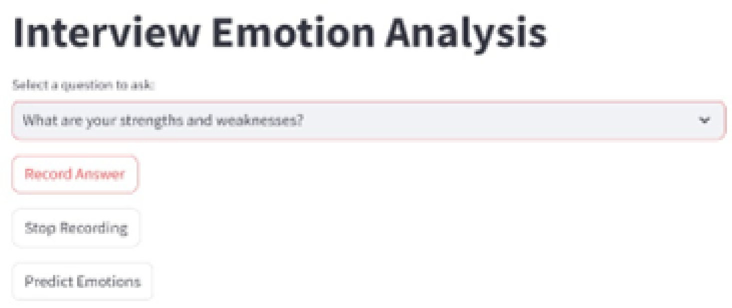
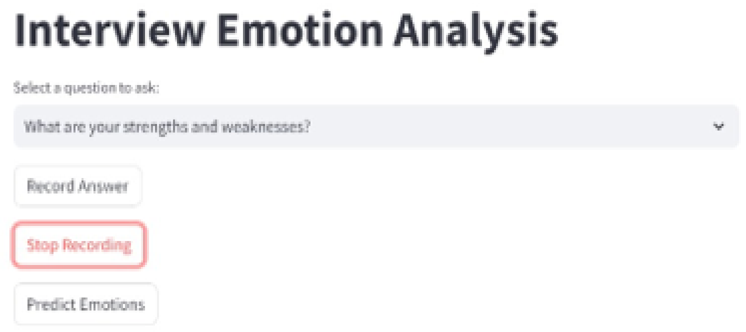
Selecting an intriguing question and beginning the recording session.	
Capturing the candidate's answer and stopping the recording.	
Analyzing the answer with the "Predict emotions" button.	
A colorful pie chart showcasing the emotional insights from the candidate's response.	

Fig 5 An Engaging UI Interface Designed to Detect Emotions from Recordings based on the Provided Question.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

After a thorough exploration of available models [Fig.4], we devised a user interface showcasing a variety of interview questions alongside their corresponding video recordings [Fig.5]. The snapshots of UI present charts as an analysis of emotional cues detected during the interviews.

Recording and analyzing a candidate's answer is both intuitive and fascinating. Here's how it unfolds: First, choose the question that sparks the candidate's interest. With a simple click on "Start recording," as the candidate speaks, sharing their thoughts and insights. Once they're done, a quick tap on "Stop recording" captures their response.

Now, click on "Predict emotions." Instantly, the system dives into the recorded answer, decoding the emotional undertones. The result? A vibrant pie chart pops up, revealing a spectrum of emotions embedded in the candidate's words. This visual representation not only makes it easy to understand the emotional dynamics but also adds a layer of depth to the evaluation process.

We meticulously scrutinized the recorded videos in comparison with the analysis outputs, assessing the actual emotions against the predicted ones for each model. Following this exhaustive evaluation, YOLOv8 emerged as the most optimal choice.

IV. CONCLUSION

In conclusion, our study underscores the importance of iterative model refinement in enhancing emotion recognition accuracy for remote interviews. Through multiple iterations, we observed significant improvements in accuracy, despite initial challenges with overfitting. Key findings include the effectiveness of adjusting data splits and exploring alternative model architectures in mitigating overfitting and improving accuracy. Addressing limitations such as dataset constraints and exploring additional model architectures remain crucial for advancing emotion recognition technology. Among the models evaluated, the YOLO v8 model achieved an appreciable accuracy of 76.88%. Moreover, the deployment of user-friendly interfaces, exemplified by our integration of the Streamlit framework, holds promise for enhancing accessibility and usability in real-world applications. Overall, our research contributes valuable insights into the practical implementation of emotion recognition systems, with implications for improving user experiences and facilitating seamless interactions in remote contexts.

ACKNOWLEDGMENTS

We acknowledge that with the consent from 360DigiTMG, we have used the CRISP-ML(Q) and the ML Workflow which are available as open-source in the official website of 360DigiTMG methodology.

REFERENCES

[1]. Jia Lu, Minh Nguyen, Wei Qi Yan; **Deep Learning Methods for Human Behavior Recognition;** 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ); https://ieeexplore.ieee.org/abstract/document/9290640

[2]. N. Mehendale; **Facial Emotion Recognition Using Convolutional Neural Networks (FERC);** SN Appl. Sci., vol. 2, no.3, 2020; https://link.springer.com/article/10.1007/s42452-020-2234-1

[3]. Shrey Srivastava, Amit Vishvas Divekar, Chandu Anilkumar, Ishika Naik, Ved Kulkarni and V. Pattabiraman; **Comparative Analysis Of Deep Learning Image Detection Algorithms;** Journal of Big Data volume 8, Article number: 66 (2021); https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00434-w

[4]. Liam Schoneveld a, Alice Othmani b, Hazem Abdelkawy b; **Leveraging Recent Advances In Deep Learning For Audio-Visual Emotion Recognition**; Pattern Recognition Letters, Volume 146, June 2021, Pages 1-7; https://www.sciencedirect.com/science/article/abs/pii/S0167865521000878

[5]. Stefan Studer, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters and Klaus-Robert Müller; **Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology;** Mach. Learn. Knowl. Extr. 2021, 3, 392–413; https://doi.org/10.3390/make3020020

[6]. Ismail Olaniyi Muraina; **Ideal Dataset Splitting Ratios In Machine Learning Algorithms: General Concerns For Data Scientists And Data Analysts**; 7th International Mardin Artuklu Scientific Researches Conference; https://scholar.google.com/citations?view_op=view_citation&hl=en&user=rXa9qAgAAAAJ&citation_for_view=rXa9qAgAAAAJ:hFOr9nPyWt4C

[7]. Ahatsham Hayat, Fernando Morgado-Dias; **Deep Learning-Based Automatic Safety Helmet Detection System for Construction Safety**; Appl. Sci. 2022, 12(16), 8268; https://doi.org/10.3390/app12168268

[8]. Gaurav Meena, Krishna Kumar Mohbey, Ajay Indian, Sunil Kumar; **Sentiment Analysis from Images using VGG19 based Transfer Learning Approach**; Procedia Computer Science, Volume 204, 2022; https://www.sciencedirect.com/science/article/pii/S1877050922007888

[9]. N. Abbassi, R. Helaly, M. A. Hajjaji, A. Mtibaa**; A Deep Learning Facial Emotion Classification system: a VGGNet-19 based approach;** 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Monastir, Tunisia, 2020; https://ieeexplore.ieee.org/document/9329355

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL721

[10]. Evander Banjarnahor, Yohanes Adi Saputra Abraham, Yoseph Siahaan; **Fundus Image Classification for Diabetic Retinopathy Using ResNet50V2 and InceptionV3**2024 IEEE International Conference on Artificial Intelligence and Mechatronics Systems (AIMS); https://ieeexplore.ieee.org/abstract/document/10512970

[11]. Usman Haruna, Rozniza Ali, Mustafa Man; **A new modification CNN using VGG19 and ResNet50V2 for classification of COVID-19 from X-ray radiograph images**; Indonesian Journal of Electrical Engineering and Computer Science Vol. 31, No. 1, July 2023; https://ijeecs.iaescore.com/index.php/IJEECS/article/view/30991

[12]. H. Kishan Kondaveeti and M. Vishal Goud; **Emotion Detection using Deep Facial Features**; *IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, Buldhana, India, 2020; https://ieeexplore.ieee.org/abstract/document/9398439

[13]. Almeida, José, and Fátima Rodrigues; **Facial Expression Recognition System for Stress Detection with Deep Learning**; *ICEIS (1)*. 2021; https://www.scitepress.org/Papers/2021/104742/104742.pdf

[14]. Priyadarshini D. Kalwad, Suvarna G. Kanakaraddi, Ashok K. Chikaraddi, T. Preeti & Karuna C. Gull (2022); **XCEPTION: Facial Expression Detection Using Deep Learning Techniques**; Advances in Intelligent Systems and Computing, vol 1415. Springer, Singapore; https://doi.org/10.1007/978-981-16-7330-6_26

[15]. R. Angeline, A. Alice Nithya**; Deep Human Facial Emotion Recognition: A Transfer Learning Approach Using Efficientnetb0 Model;** Journal of Theoretical and Applied Information Technology 102.8 (2024); https://www.jatit.org/volumes/Vol102No8/37Vol102No8.pdf

[16]. Mahmoud Jameel Atta Daasan, Mohamad Hafis Izran Bin Ishak; **Enhancing Face Recognition Accuracy through Integration of YOLO v8 and Deep Learning: A Custom Recognition Model Approach**; Methods and Applications for Modeling and Simulation of Complex Systems, AsiaSim 2023, CCIS, vol. 1911, Springer, Singapore; https://doi.org/10.1007/978-981-99-7240-1_19

[17]. Kakani, Palak, and Shreya Vyas; **Automated Catalog Generation Using Deep Learning**; International Research Journal of Modernization in Engineering Technology and Science Volume:05/Issue:08/August-2023; https://www.irjmets.com/uploadedfiles/paper/issue_8_august_2023/44010/final/fin_irjmets1692089966.pdf

[18]. K. K. Pal and K. S. Sudeep; **Preprocessing For Image Classification By Convolutional Neural Networks**; *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, Bangalore, India, 2016; https://ieeexplore.ieee.org/abstract/document/7808140

[19]. Kiran Maharana, Surajit Mondal, Bhushankumar Nemade; **A Review: Data Pre-Processing And Data Augmentation Techniques**; Global Transitions Proceedings, Volume 3, Issue 1, 2022,ISSN 2666-285X; https://www.sciencedirect.com/science/article/pii/S2666285X22000565