ISSN No:-2456-2165

Facial Authentication using Deep-Learning: An Advanced Biosecure Login Model Employing an Integrated Deep-Learning Approach to Enhance the Robustness and Security of the Login Authentication Process

Deeksha Patel[1]; Jyostna Parasabaktula[2]**;** Shiv Mangal Yadav[3]; Ritendu Bhattacharyya[4]; Bharani Kumar Depuru[5*]
[1]Research Associate, Innodatatics, Hyderabad, India.
[2]Research Associate, Innodatatics, Hyderabad, India.
[3]Research Associate, Innodatatics, Hyderabad, India
[4]Team Leader, Research and Development, Innodatatics, Hyderabad, India.
[5]Director, Innodatatics, Hyderabad, India

Corresponding Author:- Bharani Kumar Depuru[5*] ORC ID: 0009-0003-4338-8914

Abstract:- Face recognition is a concept of the safest way of logging on; it entails that our facial images are acquired, detected, and subsequently authenticated by the particular interface. In this present digital generation, safe authentication of the interfaces is the primary cautionary aspect that should be maintained, and this model suggests a secure and strong authentication system. This paper recommends a face recognition login interface that involves deep learning models to provide a strong and secure authentication mechanism. It involves the extraction of facial images, proposes a solution to enhance accuracy and trustworthiness, and presents a weighty improvement over the traditional username-password login method. This gives us a user-friendly login experience along with the highest level of security. These facial authentication models are being used in numerous fields, ranging from security, healthcare, marketing, retail, public events, payments, door unlocking and video monitoring systems, user authentication on devices, etc. It is also useful for multi-class classification problems. This paper includes face recognition techniques from convolutional neural networks (CNN) and transformer models like ViT (vision transformer), VGG16, RestNet50, Inception V3, and EfficientNetB0. It proposes that the best model will be deployed using Streamlit.

Keywords:- *Secured Authentication System, Facial Recognition, Deep Learning, Vision Transformer, VGG16, Image Classification, Streamlit.*

I. INTRODUCTION

Initially, we built this project for face recognition[4] for logging in to the education portal. But we can use this project for many purposes such as Door access authentication, attendance mark system, Image classification, and Searching people in the crowd. Face recognition provides security and does not require one to remember or write a username and password on paper. It helps to protect from fraud and duplicate access. We already discussed that, for this project, we used 5 pre-trained models which are ViT (vision transformer), VGG16, RestNet50, Inception V3, and EfficientNetB0. Now we will discuss the project process in short. After that, we discuss each step/process in detail.

The project methodology followed here is the open-sourceCRISP-ML(Q)methodologyfrom360DigiTMG. CRISP-ML (Q) [Fig.1][1] stands for Cross Industry Standard Practice for Machine Learning with Quality Assurance. CRISP-ML (Q) can broadly be defined as a methodology designed to deal with a Machine Learning solution's project lifecycle.

ISSN No:-2456-2165

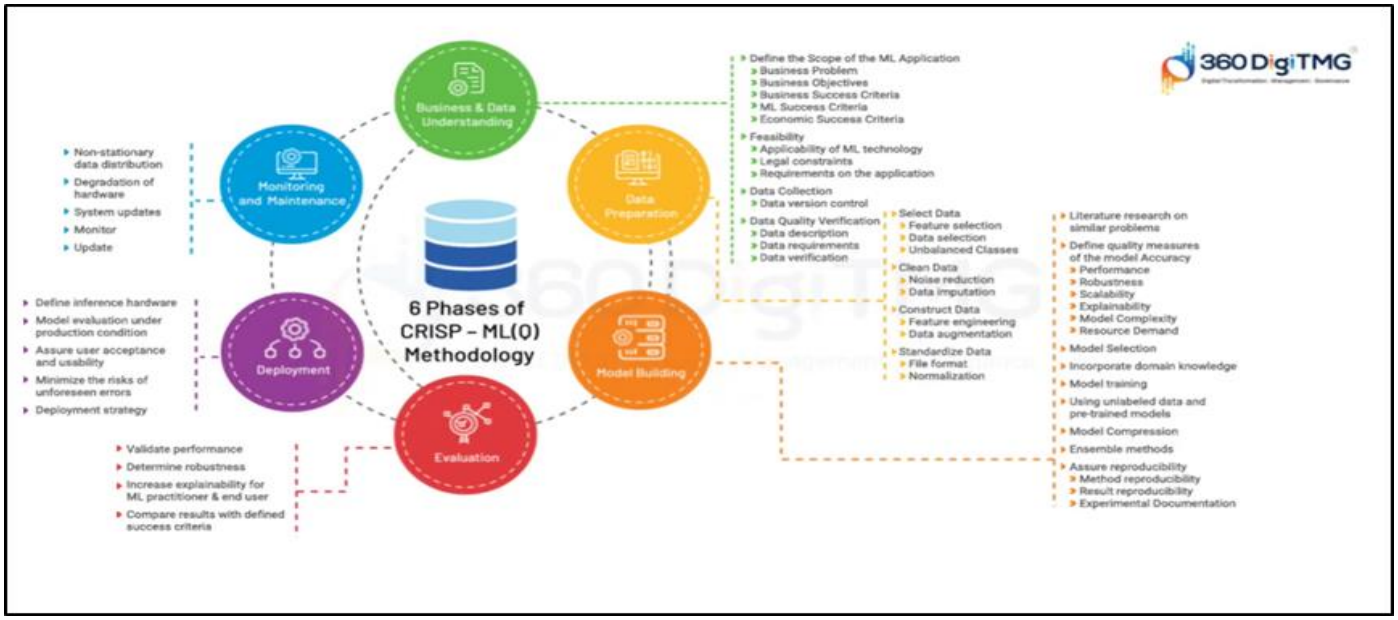

Fig 1 CRISP-ML (Q) Methodological Framework, Outlining its Key Components and Steps Visually
(Source:-Mind Map - 360DigiTMG)

First, we have collected a video dataset for 38 users, including self-video. When we collect videos from users, we decide that all videos should be in the landscape. After the collection of the video dataset, we decided to collect frames from the videos, at the time of collection, we extracted only distinct frames. After extraction, we check whether all image class datasets are balanced or not. So first, we balance the entire class image count.

After that, we divide all datasets into 3 parts: train, validate, and test, in a ratio of 80:10:10. After splitting the data set, we use five pre-trained models and record their accuracy. When required, we tune the hyperparameters as per the requirement and compare the accuracy of all five models. We choose the best model and deploy the best model with the help of Streamlit.

II. METHODS AND TECHNOLOGY

A. System Requirements (Computer Hardware and Software) used:

Table 1 System Requirements (Computer Hardware and Software)

Operating System	**Windows 11**
RAM	8/16 GB
Processor	i5
IDE	Spyder, Jupyter Notebook, Google Colab

B. Model Architecture:

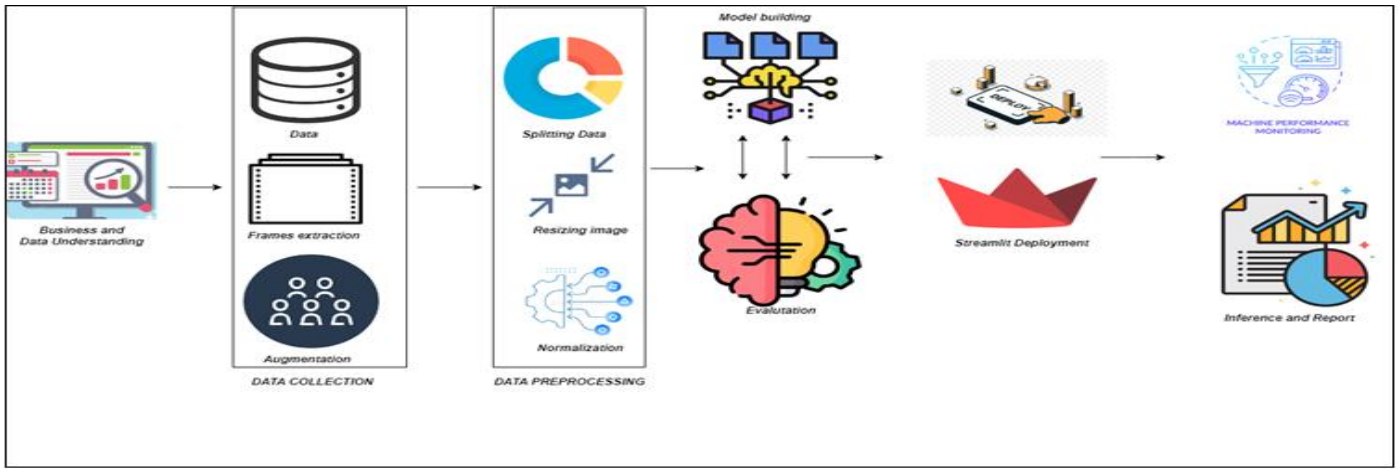

Fig 2 Architecture Diagram: Explanation of the Workflow of the Face Recognition Project
(Source:- https://360digitmg.com/ml-workflow)

The project architecture[Fig.2] explains how the entire project has been conducted. According to the business problem, relevant data has been collected as videos. With the necessary preprocessing steps data was given to the model. Once the model was finalized, with the help of Streamlit and the great architecture of AWS, the application was deployed for the end users.

AWS was chosen because of the facility of scalability and server loads. For the inference, one UI will open, where using the webcam one image will be captured which will be a numpy array. That image will be passed to the model for inference. It will place the name and the confidence in the UI.

C. Data Collection:

The data is primary data. Captured video shots of the people in a specific way, so that all of the area of the face will be covered. A total of 38 videos we have collected. Using Python those videos were split into frames in jpg format. Those images have been stored in separate folders.

D. Dataset Dimension:

Table 2 Dataset Dimension

Video File format	**.mp4, .mkv, .mov**
Image format	jpg
Number of Classes	38
Frames extracted	1591
Size of the data set	1.18 GB
Frames after Augmentation	38000

E. Data Preprocessing:

The image library is harnessed to establish a series of augmentation techniques[12] intended for application to the images[3]. The resulting images are methodically stored within a predefined output folder, exhibiting a structure that mirrors the organization of the input folders. As part of this process, the images undergo resizing to conform to passport-size dimensions. While not obligatory, this resizing step confers several advantages in terms of both processing efficiency and potential enhancements to model performance.

It is crucial to keep the dataset balanced, that is all the subdirectories will have the same number of images. Here we have 1000 images in each subdirectory[Fig.3]. It is ensured to prevent biases and also to ensure fair and accurate recognition across diverse individuals. An imbalanced dataset can lead to biased models favoring the majority class, resulting in unfair and discriminatory outcomes. A balanced dataset will promote equal representation during training, enabling the model to learn diverse features associated with different individuals and improving its generalization capability. Also, it increases the robustness to variations in lighting, poses, and expressions, contributing to reliable performance in real-world scenarios.

As per the preprocessing steps before passing to the model all of the images have been normalized[2]. This method leads the images scaled from 0 to 1. This will help to build a great model.

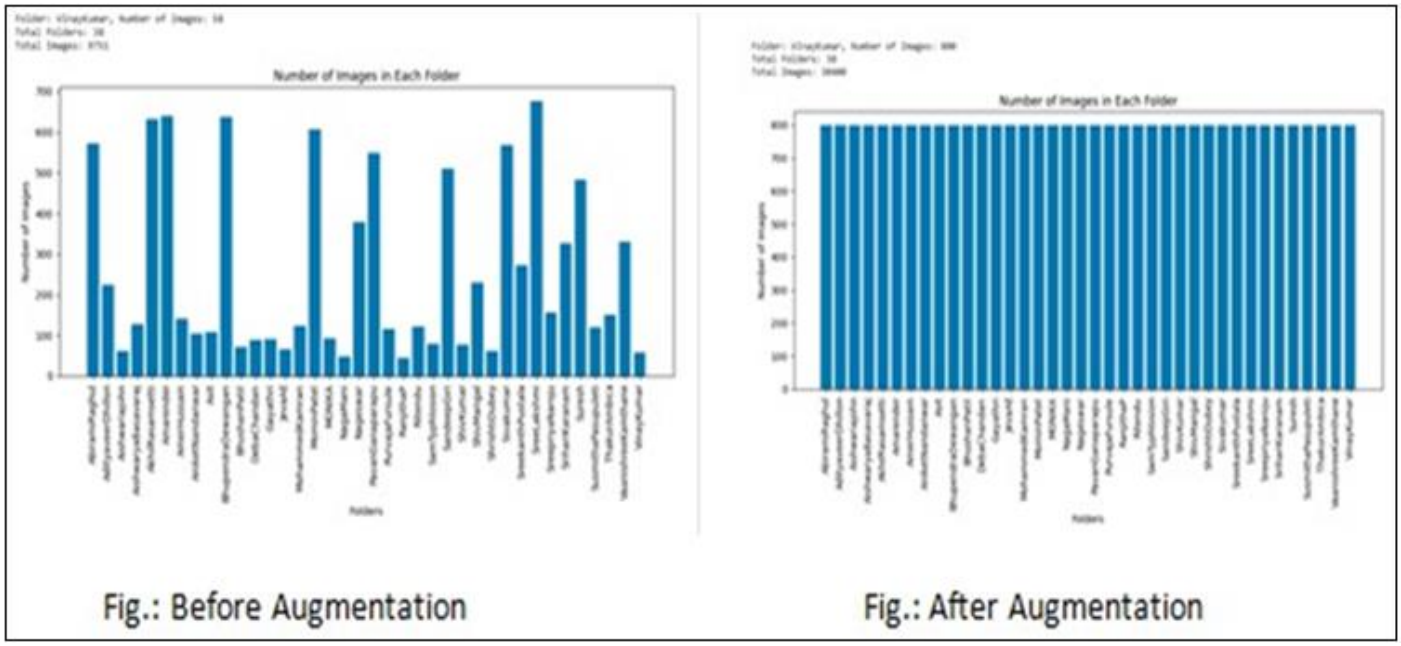

Fig 3 Histogram of Files before and after Augmentation it Shows the Balancing of the Dataset.

After processing these steps, the data was split into 3 parts, i.e., training, testing, and validation. During the splitting procedure, industry best practice was followed, i.e., it was split in the ratio of 80%, 10%, and 10%. So here 80% was decided as a training dataset, 10% was decided for the testing dataset, and another 10% was decided for the validation dataset. Following this split methodology to avoid the data leak. Here the test will be completely unknown to the model. For that reason, it can be used for accuracy inference.

ISSN No:-2456-2165

Table 3 Testing and Validation

Splitting of Dataset (80:10:10)	
Train	30,400
Test	3800
Validation	3800

F. Model Building:

In the model-building phase, 5 models were tried out i.e., VGG16, VISION TRANSFORMER, RESNET50, INCEPTION V3, and EFFICIENT NET B0. After researching and checking sevaral articles, this models has been listed out.

➢ *VGG16:*

It's a well-known pre-trained AI model[6][7], designed for image data. Due to the high and decent accuracy, this research concludes this model to take as an experiment. It has 16 layers divided into 13 convolutions and 3 fully connected layers. It was trained on a huge dataset known as imagenet. Face recognition can be obtained using vgg16 by changing the output layer with the required classes and some hyperparameter changes if needed. Training with this configuration can provide a strong face authentication model.

➢ *Res Net50:*

Res Net 50[7][10][11] is a well-known pre-trained model, which has a 50-layer deep CNN[5] structure, the integration of multiple blocks each incorporating a series of layers significantly amplifies the models' proficiency in discerning nuanced facial patterns with exceptional accuracy. This robust architectural design fairly establishes the model as an exemplary solution for achieving optional accuracy in face recognition applications.

➢ *Inception V3:*

A constituent of Google's Inception series, Inception v3 [8]signifies a remarkable advancement in convolutional neural network (CNN) architectures, purposefully crafted for image classification. Through thorough training on expansive datasets like ImageNet, the model attains proficiency in the analysis and categorization of images. Its adaptability is evident with the provision of fine-tuning options, enabling customization for specific applications. Pre-trained on extensive datasets, such as ImageNet, and Inception v3 demonstrates a remarkable ability to generalize across diverse visual patterns. This adaptability, in tandem with fine-tuning flexibility, firmly establishes Inception v3 as a pivotal element in the realm of computer vision.

➢ *Efficient Net B0:*

Efficient Net B0[13] is a compact member of the EfficientNet family of neural network architectures and finds novel applications in the domain of face recognition. Tailored for optimal performance in tasks requiring precise facial identification, EfficientNet B0 capitalizes on its efficient design and computational prowess. Leveraging pre-trained features on facial datasets, the model adapts seamlessly to the intricacies of face recognition tasks. With a focus on computational efficiency, it stands as an efficient solution for scenarios where accurate face recognition is paramount, making it particularly suitable for deployment on resource-constrained devices like mobile and edge platforms.

➢ *Vision Transformer:*

In the realm of AI, It's an extraordinary model[9]. Classifying images is the base tech for this algorithm, although it's a transformer-based model. According to the architecture diagram [Fig.5]it uses basic CNN to break the entire image into some small patches. Those patches will then be represented as vectors and fed sequentially to the transformer model. With the help of the self-attention technology approach, it creates an incredible sense of the image.

This information has been received for further evaluation. Experiments with learning rate changes were also conducted [Fig.4]. We decided on the best model on account of the top accuracy.

S/No	Model	Optimizer	Learning rate	Loss	Metrices	Train_Accuracy	Validation_Accuracy	Ephocs	Test_Accuracy
1	vgg16	adam	0.001	categorical_cross entropy	accuracy	92.13	93.44	10	92.11
2	ResNet50	adam	0.0001	categorical_cross entropy	accuracy	11.2	15.25	10	Not Try
3	Inception	adam	0.0001	categorical_cross entropy	accuracy	87.95	89.67	10	88.15
4	EfficientNetB0	adam	0.0001	categorical_cross entropy	accuracy	0.94	1.64	10	Not Try
5	VIT	adam	0.001	categorical_cross entropy	accuracy	100	99.33	10	100

Fig 4 All 5 Model Results and Choose the Best One for Deployment

- *Vision Transformer has been Chosen According to the Selection Logic Mentioned.*

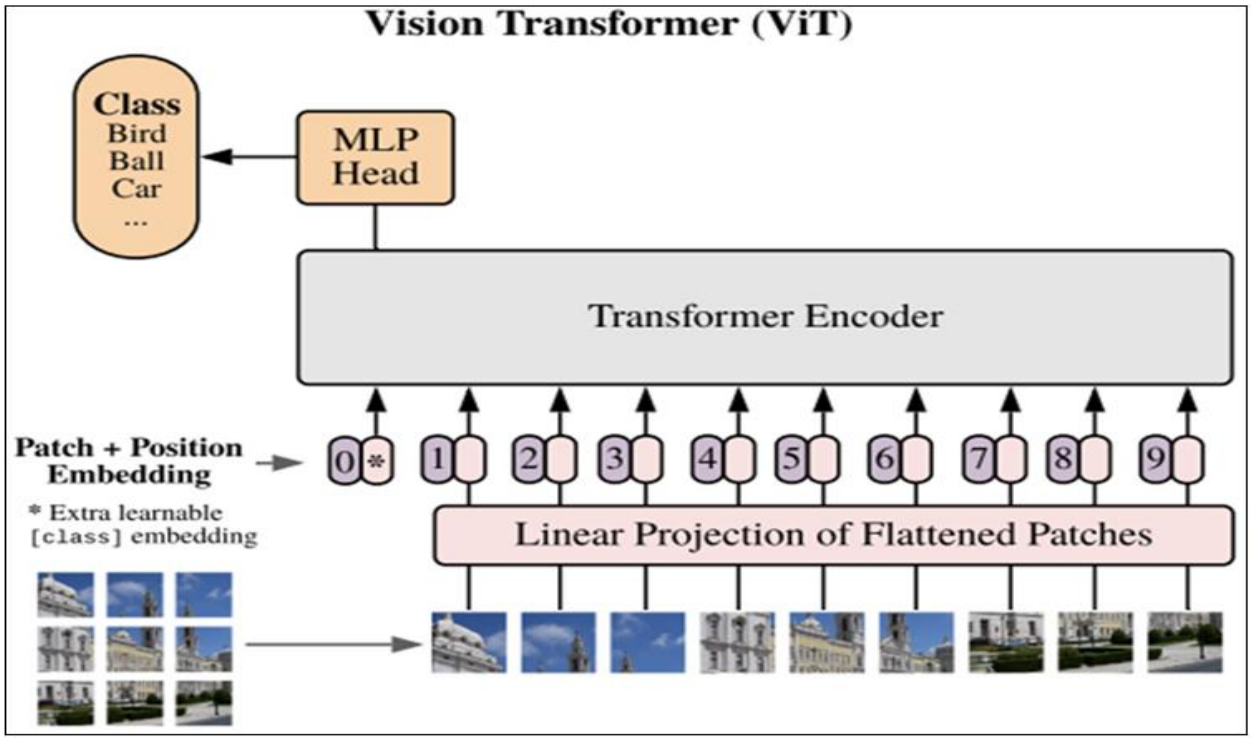

Fig 5 Vision Transformer Architecture
(Source:-https://paperswithcode.com/method/vision-transformer)

In this study, we employed a pre-trained Vision Transformer (ViT) architecture and customized its output layer to align with the specific classes in our dataset. Our methodology involved meticulous research to optimize hyperparameters, leading us to adopt a batch size of 10 and a learning rate of 0.0001. The model underwent a 10-epoch training regimen, culminating in an achieved accuracy score of 0.00002. This research underscores the adaptability of the ViT architecture for tailored datasets, showcasing the effectiveness of our approach in addressing the unique challenges posed by our dataset. The results affirm the viability of pre-trained ViT models with customized output layers as a robust solution in computer vision applications.

III. RESULTS AND DISCUSSION

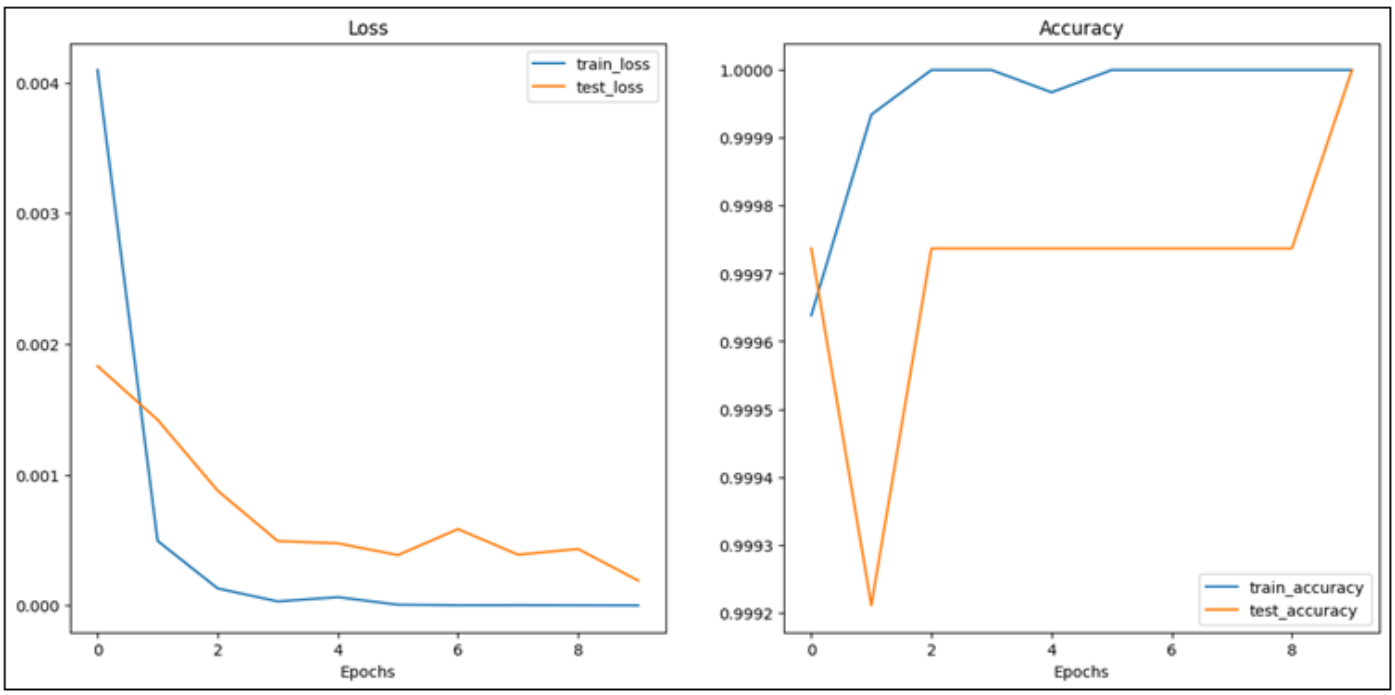

Fig 6 Output Plotted in the Graph

Upon successfully obtaining the optimized model, we preserved its parameters and validated its predictive accuracy through image assessments within predefined classes. Subsequently, the focus shifted towards implementing a user-friendly interface, accomplished through the integration of Streamlit [Fig.7] [14]. The devised interface features a secure login mechanism, wherein users input their credentials. Upon authentication, the webcam initiates facial recognition for user verification, thus granting access for further interactions. This streamlined and secure approach to model deployment enhances the practicality and accessibility of our system, ensuring a seamless user experience in real-world applications.

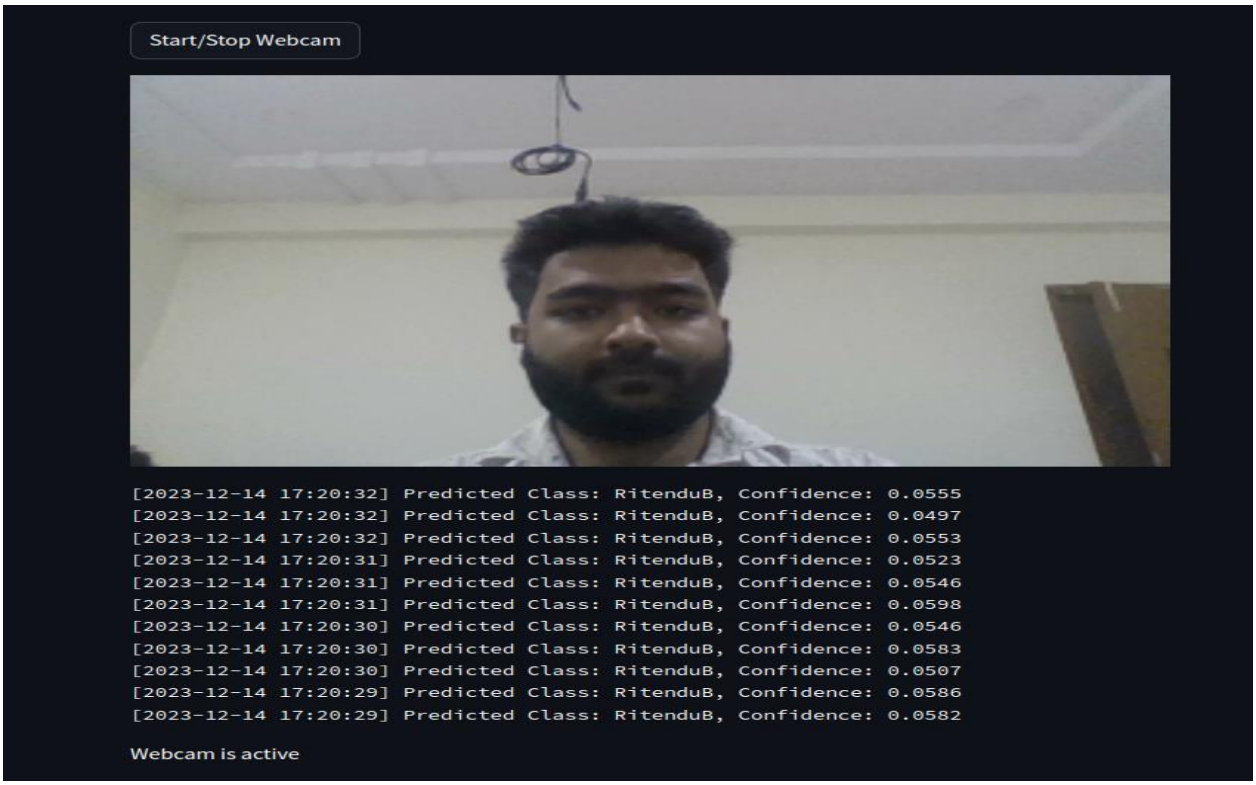

Fig 7 The above Image Shows the Sample Implementation Output Predicting the Person and also Shows the Confidence Score with the Exact Date and Time

IV. CONCLUSION

In the contemporary landscape, projects centered around image processing play a pivotal role in bridging gaps across diverse sectors. Particularly noteworthy is the integration of authenticated login mechanisms within web browsers, as well as the establishment of robust systems for presence collection and person identification. In the current technological milieu, artificial intelligence (AI) stands out as a preeminent field, leveraging advanced deep learning techniques to address complex challenges. With a particular emphasis on resolving intricate vision-related issues, AI emerges as a formidable force, offering efficient and accessible solutions to a myriad of problems across various domains. This research underscores the transformative impact of AI, particularly within the realm of deep learning, in ushering in innovative and practical approaches to image-related projects and problem-solving applications.

ACKNOWLEDGMENTS

We affirm our use of the CRISP-ML(Q) and ML Workflow which are openly available on the official 360digitmg website with the explicit consent from 360digitmg.

REFERENCES

[1]. Stefan Studer, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters and Klaus-Robert Muller, Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology, 2021, Volume 3, Issue 2. https://doi.org/10.3390/make3020020

[2]. Graham D. Finlayson, Bernt Schiele & James L. Crowley, Comprehensive color image normalization, 2006, Part of the Lecture Notes in Computer Science book series (LNCS, volume 1406), https://link.springer.com/chapter/10.1007/BFb0055685

[3]. Mingle Xu, Sook Yoon, Alvaro Fuentes, Dong Sun Park, A Comprehensive Survey of Image Augmentation Techniques for Deep Learning, 2023, Volume 137, https://doi.org/10.1016/j.patcog.2023.109347

[4]. Mei Wang, Weihong Deng, Deep face recognition: A survey, 2021, Neurocomputing, Volume 429, Pages 215-244. https://doi.org/10.1016/j.neucom.2020.10.081

[5]. Zhiming Xie, Junjie Li, and Hui Shi, A Face Recognition Method Based on CNN Journal of Physics: Conference Series, 2019, *J. Phys.: Conf. Ser*. 1395 012006, DOI: 10.1088/1742-6596/1395/1/012006. A Face Recognition Method Based on CNN - IOPscience

[6]. R.Meena Prakash, N. Thenmoezhi, M. Gayathri, Face Recognition with Convolutional Neural Network and Transfer Learning, 2020, Published in 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Publisher: IEEE, DOI: 10.1109/ICSSIT46314.2019.8987899

[7]. Bendjillali Ridha Ilyas, Beladgham Mohammed, Merit Khaled; Kamline Miloud, Enhanced Face Recognition System Based on Deep CNN, 2020, Published in 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), Publisher: IEEE, DOI: 10.1109/ISPA48434.2019.8966797

[8]. Rajasekaran Thangaraj, P Pandiyan; T Pavithra, V.K Manavalasundaram, R Sivaramakrishnan, Vishnu Kumar Kaliappan, Deep Learning based Real-Time Face Detection and Gender Classification using OpenCV and Inception v3, 2022, Published in: 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Publisher: IEEE, DOI: 10.1109/ICAECA52838.2021.9675635

[9]. Chi-Phuc Tran, Anh-Khoa Nguyen Vu, Vinh-Tiep Nguyen, Baby Learning with Vision Transformer for Face Recognition, 2022, Published in 2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), Publisher: IEEE, DOI: 10.1109/MAPR56351.2022.9924795

[10]. Yohanssen Pratama, Lit Malem Ginting, Emma Hannisa Laurencia Nainggolan, Ade Erispra Rismanda, Face recognition for presence system by using residual networks-50 architecture, 2021, International Journal of Electrical and Computer Engineering (IJECE) Vol. 11, No. 6, December 2021, pp. 5488~5496 ISSN: 2088-8708, DOI: http://doi.org/10.11591/ijece.v11i6.pp5488-5496

[11]. Md. Mehedi Hasan, Md. Ali Hossain, Azmain Yakin Srizon; Abu Sayeed, Mohiuddin Ahmed, Md Rakibul Haquek, Improving Performance of a Pre-trained ResNet-50 Based VGGFace Recognition System by Utilizing Retraining as a Heuristic Step, 2022, Published in: 2021 24th International Conference on Computer and Information Technology (ICCIT), Publisher: IEEE, DOI: 10.1109/ICCIT54785.2021.9689918

[12]. Divyarajsinh N. Parmar, Brijesh B. Mehta, Face Recognition Methods & Applications, 2014, Journal reference: International Journal of Computer Technology & Applications, Vol 4(1), pp. 84-86, Jan-Feb 2013, https://doi.org/10.48550/arXiv.1403.0485

[13]. Antonio Bruno, Davide Moroni, Massimo Martinelli, Efficient Adaptive Ensembling for Image Classification, 2023, Journal: Expert Systems, reference: (2023), https://doi.org/10.48550/arXiv.2206.07394

[14]. Himangi Dani, Pooja Bhople, Hariom Waghmare, Kartik Munginwar, Prof. Ankush Patil, Review on Frameworks Used for Deployment of Machine Learning Model, 2022, Volume:10, Issue: II, https://doi.org/10.22214/ijraset.2022.40222