成都理工大學中英合作办学
CDUT Sino-British Collaborative Education

OXFORD
BROOKES
UNIVERSITY

UNDERGRADUATE PROJECT REPORT

Project Title:	**Facial Recognition Attendance System**
Surname:	**Qin**
First Name:	**Hong Sheng**
Student Number:	**202118010213**
Supervisor Name:	**Dr. Joojo Walker**
Module Code:	**CHC 6096**
Module Name:	**Project**
Date Submitted:	**May 6, 2025**

Declaration

Chengdu University of Technology Oxford Brookes College

Chengdu University of Technology

BSc (Single Honours) Degree Project

Programme Name: **Computer Science**

Module No.: **CHC 6096**

Surname: **Qin**

First Name: **HongSheng**

Project Title: **Facial Recognition Attendance System**

Student No.: **202118010213**

Supervisor: **Joojo Walker**

2ND Supervisor (if applicable): **Not Applicable**

Date submitted: **May 6, 2025**

A report submitted as part of the requirements for the degree of BSc (Hons) in Computer Science

At

Chengdu University of Technology Oxford Brookes College

Declaration

Student Conduct Regulations:

Please ensure you are familiar with the regulations in relation to Academic Integrity. The University takes this issue very seriously and students have been expelled or had their degrees withheld for cheating in assessment. It is important that students having difficulties with their work should seek help from their tutors rather than be tempted to use unfair means to gain marks. Students should not risk losing their degree and undermining all the work they have done towards it. You are expected to have familiarised yourself with these regulations.

https://www.brookes.ac.uk/regulations/current/appeals-complaints-and-conduct/c1-1/

Guidance on the correct use of references can be found on www.brookes.ac.uk/services/library, and also in a handout in the Library.

The full regulations may be accessed online at https://www.brookes.ac.uk/students/sirt/student-conduct/

If you do not understand what any of these terms mean, you should ask your Project Supervisor to clarify them for you.

I declare that I have read and understood Regulations C1.1.4 of the Regulations governing Academic Misconduct, and that the work I submit is fully in accordance with them.

Signature Qin HongSheng(Randy) Date ………………May 6, 2025………………

REGULATIONS GOVERNING THE DEPOSIT AND USE OF OXFORD BROOKES UNIVERSITY MODULAR PROGRAMME PROJECTS AND DISSERTATIONS

Copies of projects/dissertations, submitted in fulfilment of Modular Programme requirements and achieving marks of 60% or above, shall normally be kept by the Oxford Brookes University Library.

I agree that this dissertation may be available for reading and photocopying in accordance with the Regulations governing the use of the Oxford Brookes University Library.

Signature Qin HongSheng(Randy) Date ………………May 6, 2025………………

Acknowledgment

With the process of composing this report, I have been heartfelt grateful to the support and guidance from Dr.Joojo Walker and many other friends. They offer me help no matter in what time or in what circumstances. Especially my project mentor Joojo, he brings new ideas and approaches from scratch, also with abundant help and guidance during the process of the system implementation and essay writing. Meanwhile, in my four year undergraduate learning session, countless teachers and students help me to foster and grow into a better person. With them, I get to know how to code, how to make a presentation, how to be a project leader, how to be an outstanding association president, how to find and contact with investor from society, how to present my own project in front of national and provincial platform. Therefore, I’m here to express my heartfelt gratitude and honor to their companion and contributions.

Meanwhile, I must extend my sincerest thanks to my girlfriend and my family. Throughout the entire research period, they provided me not only economical but also emotional supports. Their understanding and sponsor provides me continuous motivation to keep researching. In conclusion, my mentors, teachers, classmates, girlfriend and my parents, their warmth and contributions will be kept in my heart with my entire life.

Table of Contents

Abstract

With the exponential reliance on biometric authentication in contemporary digital systems, facial recognition has become a prominent solution due to its contactless nature, high throughput, and user-friendly interface. This project proposes and implements a real-time Facial Recognition Attendance System (FRAS) powered by a hybrid convolutional and Vision Transformer model: ConViT_small, to optimize the trade-off between local feature extraction and global pattern recognition. The system leverages the PubFig dataset, comprising diverse images with varying illumination, pose, and expression, to train and evaluate the recognition pipeline.

Preprocessing techniques such as normalization, augmentation, and dimensionality reduction were applied to enhance generalization, while cosine similarity-based matching enabled efficient identity verification base on facial data. The pretrained ConViT_small model, sourced from the TIMM library, provided 384 dimensional embeddings, which vigorously proved the effectiveness in distinguishing between multiple identities. Mltiple comprehensive evaluation metrics including confusion matrix, ROC curve, similarity histogram, and F1-score were applied to validate the model's performance. The FRAS achieved an accuracy of 86.3%, with F1-score of 0.61, outperforming comparative models such as ResNet18, MobileNetV2, and ViT_small.

The system features a responsive GUI with real-time webcam integration, automatic attendance logging, audio prompts, and live tabular updates. Ethical and legal considerations were addressed through GDPR-compliant practices including user consent and encrypted data storage with local processing. Overall, the system demonstrates the feasibility of deploying transformer-based facial recognition in real-world attendance tracking applications including classrooms, companies, governmental agencies and etc. Offering high accuracy, robustness and scalability, while maintaining user privacy, usability and system responsiveness.

Keywords: Facial Recognition, Attendance System, ConViT, Hybrid CNN–ViT, Real-Time Processing, Privacy Preservation

Abbreviations

ACM	Association for Computing Machinery
API	Application Programming Interface
AUC	Area Under the Curve
CNN	Convolutional Neural Network
CosSim	Cosine Similarity
CPU	Central Processing Unit
CSS	Cascading Style Sheets
DPA	Data Protection Act
DL	Deep Learning
DPA	Data Protection Act
EDA	Exploratory Data Analysis
EER	Equal Error Rate
ERC	Ethics Review Committee
ERP	Enterprise Resource Planning
F1-Score	Harmonic Mean of Precision and Recall
FAR	False Acceptance Rate
FPS	Frames Per Second
FRAS	Facial Recognition Attendance System
FRR	False Rejection Rate
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HaaR	Histogram of Oriented Gradients and AdaBoost + Random Forest
IDE	Integrated Development Environment
IoT	Internet of Things
JSON	JavaScript Object Notation
KNN	K-Nearest Neighbors
LDR	Light Dependent Resistor
LMS	Learning Management System

ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
ODBC	Open Database Connectivity
OCR	Optical Character Recognition
PCA	Principal Component Analysis
PGSQL	PostgreSQL
RAM	Random Access Memory
RFC	Request for Comments
ROC	Receiver Operating Characteristic
SQL	Structured Query Language
TLS	Transport Layer Security
URL	Uniform Resource Locator
ViT	Vision Transformer
XML	Extensible Markup Language

Glossary

Facial Recognition Attendance System (FRAS)	A biometric system that automates attendance tracking by detecting and identifying individuals’ faces in real time, logging entries and exits without manual intervention.
EDA(Exploratory Data Analysis)	The approach in statistics to utilize abundant techniques to maximize insight into dataset.
Hybrid CNN–ViT Architecture	A neural network design that integrates convolutional neural network (CNN) layers for capturing local image patterns with Vision Transformer (ViT) modules for modeling long-range dependencies via self-attention.
Edge-Device Optimization	Techniques such as model quantization, pruning, and lightweight architecture design applied to enable efficient inference on resource-constrained hardware (e.g., embedded devices or mobile platforms).
Ablation Study	A controlled experimental analysis in which specific components or features of a model are removed or altered to quantify their individual contributions to overall performance.

Chapter 1 Introduction

1.1 Background

Facial recognition technology has emerged as a leading biometric solution, offering significant advantages over traditional methods in security, surveillance and personal identification[1]. Its natural, nonintrusive, and high-throughput data acquisition capabilities make automatic facial recognition particularly beneficial compared to other biometrics, such as fingerprinting or iris scanning, which often require direct user interaction[2]. Over the past decades, researches found that the advancements in fields of computer vision and machine learning, the integration of convolutional neural networks (CNNs) and transformer-based models are outstanding, even under complex and challenging conditions such as varying illumination, diverse head poses, different facial expressions, and partial occlusions, have substantially enhanced the accuracy and robustness of facial recognition implementations[3].

In the context of attendance tracking, traditional methods like manual sign-in sheets and ID card scans are often time-consuming, prone to human error, and susceptible to fraudulent practices such as unmatched attendance or proxy attendance[4]. These conventional approaches not only compromise administrative efficiency but also exist vulnerabilities in accountability and security frameworks. In order to effectually address above problems and inconveniency, Facial Recognition Attendance Systems (FRAS) have been selected and developed to automate and secure the attendance recording process, providing substantial improvements in operational efficiency, verification accuracy, and system integrity[5].

Recent implementations of FRAS in educational institutions, corporate environments, and governmental agencies have demonstrated the technology’s ability and urgency to enhance scalability, improve user experience, and reduce administrative burdens[6]. These systems offer a seamless and contactless user interaction model, which has become increasingly valuable in contexts that demand hygiene, speed, and user convenience, such as post-pandemic public health environments[7].

Moreover, researches found that there is a growing emphasis on ensuring the ethical utilization and privacy protection of biometric data in FRAS deployments. Compliance with international data protection standards such as GDPR (General Data Protection Regulation) has guiding the development of more secure and privacy preserving facial recognition frameworks[1]. Also, Techniques such as on-device processing, federated learning, and encryption of facial embeddings are actively being researched and integrated to mitigate the risks of data breaches and unauthorized surveillance[8].

Parallel to privacy advancements, ongoing research efforts focus heavily on optimizing real-time processing capabilities. Innovations such as lightweight model architectures, quantization, pruning, and edge-device optimization enable FRAS solutions to perform accurate recognition at low latency, even on resource-constrained hardware[9]. Furthermore, consistent integration with existing enterprise resource planning (ERP) systems and learning management systems (LMS) extends the practical applicability of the FRAS, making it a popular selected solution for large-scale adoption across industries.

In summary, the convergence of technological advancements, growing security demands, and practical operational needs reinforces the transformative potential of FRAS in modern attendance tracking applications[10]. By providing a more accurate, secure, efficient, and user-friendly solution, FRAS stands poised to revolutionize administrative processes in education, enterprise, and beyond.

1.2 Aim

The aim of this project is to design and develop a Facial Recognition Attendance System that ensures efficient, reliable, and secure tracking and recording of attendance across diverse scenarios. By leveraging the latest advancements in facial recognition technology and machine learning, the proposed FRAS seeks to overcome the limitations of traditional attendance methods while addressing privacy and ethical considerations.

1.3 Objectives

The objectives are depicted as follows:

i. Conduct review on facial recognition and attendance tracking methodologies.

ii. Research and select pre-trained facial recognition model based on accuracy, employability, precision performance and process speed.

iii. Ascertain and utilize appropriate datasets for training and evaluation.

iv. Implement and integrate model into unified embedded device to optimize user experience.

v. Measure FRAS prototype through rigorous testing and documentation, and provide plans and recommendations for future improvements and enhancements.

1.4 Project Overview

1.4.1 Scope

This project focuses on the design and implementation of a Facial Recognition Attendance System (FRAS) that utilizes a pretrained ConViT_small hybrid CNN–Transformer model to perform accurate, efficient, and real-time face identification. The system is designed to replace traditional attendance methods with a contactless, secure, and automated solution, applicable in educational institutions, corporate settings, and other environments requiring identity-based access or verification.

The scope of this project encompasses all phases of system development, including data preprocessing, model integration, identity matching via cosine similarity, and deployment through a user-friendly graphical interface. It includes the usage of a structured dataset of PubFig database, the extraction of feature embeddings via ConViT, and the evaluation of performance.

Moreover, the system architecture was developed with embedding environment including live camera feed processing, GUI integration, and audio-visual feedback mechanisms. Ethical considerations such as data privacy, user permission and embedding encryption are taken into consideration in the system design period to comply with GDPR standards.

Ultimately, the project aims not only to demonstrate the technical feasibility of a transformer based facial recognition system but also to highlight its performance benefits, practical challenges, and deployment considerations. Thereby contributing to both academic researches and real world application of biometric attendance system.

1.4.2 Audience

The primary beneficiaries of this project are including educational institutions, corporate organizations, and government agencies. All of which require efficient and secure attendance tracking systems. By actualizing the autonomous attendance process, the system resultfully reduces administrative workload, minimizes human manual error and enhances security. It is particularly beneficial for environments where managing attendance is time-consuming or prone to fraudulent activities, such as proxy attendance. Additionally, the system's scalability allows it to be deployed in a variety of scenes, from small private events to large institutions, providing seamless, real-time monitoring and logging of attendance data. In conclusion, the proposed FRAS is suitable for any organization seeking to improve operational efficiency and security through automated attendance verification and logging.

Chapter 2 Background Review

Facial recognition technology has significantly advanced and evolved in recent years, particularly within attendance monitoring systems. This review synthesizes multiple researches of its technological advancements, implementation efficiency, and future directions and methods in facial recognition-based attendance systems.

Jing, Lu, and Gao (2022) conducted a comprehensive survey on 3D face recognition, revealing that 3D systems achieve an average accuracy of 98.7%, outperforming traditional 2D methods, which average 92.3% under varying lighting and pose conditions[11]. The study highlights the effectiveness of integrating convolutional neural networks (CNNs) with transformer-based models, which individually contribute to a 15% reduction in false acceptance rates (FAR) and a 10% reduction in false rejection rates (FRR). Furthermore, they emphasize that high-resolution 3D sensors enhance recognition accuracy by an additional 20%, owing to their ability to capture intricate surface geometries and finer facial features that are often missed by conventional 2D imaging. The authors conclude that the combination of 3D sensing technologies with hybrid architectures establishes a new benchmark for robust and reliable biometric identification, particularly in environments including occlusion, variable illumination and dynamic expressions[12].

Boutros et al. (2023) explored the utilization of synthetic data in face recognition, demonstrating that synthetic datasets can increase training efficiency by 30% and enhance training data by 50%. Which can effectively decrease the real world data scarcity[13]. Their findings indicate that models trained on real and synthetic datasets exhibited 12% improvement in robustness against occlusions and non-frontal expressions, compared with models trained merely on real images. The authors argued that synthetic data can continuously enhance the generalization ability of face recognition systems, especially when deployed in complex real-world scenarios, without compromising ethical and legal constraints during data acquisition.

For practical applications research, Dev and Patnaik (2020) developed a student attendance system using facial recognition, achieving a 96.5% accuracy rate and processing facial images at 25 frames per second[7]. Furthermore, the system recorded an 85% decrease in proxy attendance attempts, addressing a critical vulnerability in traditional ID card–based and manual sign-in methods. Dev and Patnaik emphasized that their design prioritized processing speed, deployment simplicity, and ease of integration with existing student management systems. While the system performed reliably under standard classroom conditions, limitations remained in handling cases of occlusion like students wearing masks and large pose variations, indicating potential for improvement through more advanced deep learning models or multi-modal verification techniques.

Similarly, Arjun Raj et al. (2020) presented a smart attendance system with a 97.8% accuracy rate and a processing speed of 30 FPS[4]. Thereby ensuring real-time responsiveness suitable for classroom and workplace environments. Implementation results indicated a 50% reduction in attendance processing time compared to manual or card-based methods and a 90% decrease in manual entry errors, reflecting significant operational improvements. Additionally, user feedback collected through surveys demonstrated a 95% approval rating, citing convenience, speed, and perceived fairness as major factors. The study highlighted the cost-effectiveness of the solution, as it minimized hardware dependency and maintenance costs relative to traditional RFID- or fingerprint-based systems. However, there are existing concerns about data privacy, security of stored facial embeddings and potential bias in recognition performance across demographic groups were acknowledged, showing disadvantages and improvement area for future system enhancement.

Singh et al. (2024) introduced an attendance monitoring system that combines facial recognition with geo-location verification, achieving a combined accuracy rate of 98.2%[2]. Through cross verification with facial identity and geographic location, the system ensured 99% accuracy in validating attendance records in pre specified zones and capable of flagging unauthorized remote attendance attempts with 96.5% precision. Meanwhile, user experience surveys also indicated a 25% improvement in perceived reliability and a 20% increase in overall user satisfaction compared to systems relying merely on facial recognition. However, the authors acknowledged that potential challenges in fields of the complexity in system design, increased privacy concerns related to location tracking and the needle for stable internet connectivity, suggesting a improvement for further balancing of security and system autonomy in future iterations and developments.

Name	Method	Advantage	Limitation	Database	RR1(%)
Jing et al. (2022)[11]	3D, CNNs, Transformer	Light, Pose	Computationally Expensive	FRGC v2	98.7
Dev et al. (2020)[7]	HaaR, KNN	Less Manual Work	Occlusion	UND	96.5
Arjun et al. (2020)[4]	Feature Extraction	Cost-efficent	Privacy and Data Security	Bosphorus	97.8
Singh et al. (2024)[2]	Geo-location Verification	Location, Flag	Complexity, Privacy	FRGC v2	98.2
Boutros et al. (2024)[13]	GANs	Mitigated Data Security and Privacy Concerns	Occlusion	FRGC v2 Bosphorus	94.91

Table 1: Summary of Reviewed Literature. RR1 = rank-1 recognition rate

As shown in table 1, the reviewed literature demonstrates the progressive evolution and enhancements of facial recognition technologies toward improved robustness, efficiency, and security in biometric systems. Across all studies, high recognition rates which RR1 exceeding 94% affirm the maturity of the field. Nontheless, trade-offs between performance, computational cost, privacy, and real-time feasibility always persist.

Collectively, the available literature indicates that hybrid models combining local feature learning and global relational reasoning, such as CNN–Transformer architectures, represent a promising path forward. These findings reinforce and testify the feasibility of the selection of ConViT_small model as the backbone for the proposed FRAS system, balancing accuracy, robustness, and deployment practicality.

Chapter 3 Methodology

3.1 Approach

3.1.1 Facial Recognition Attendance System

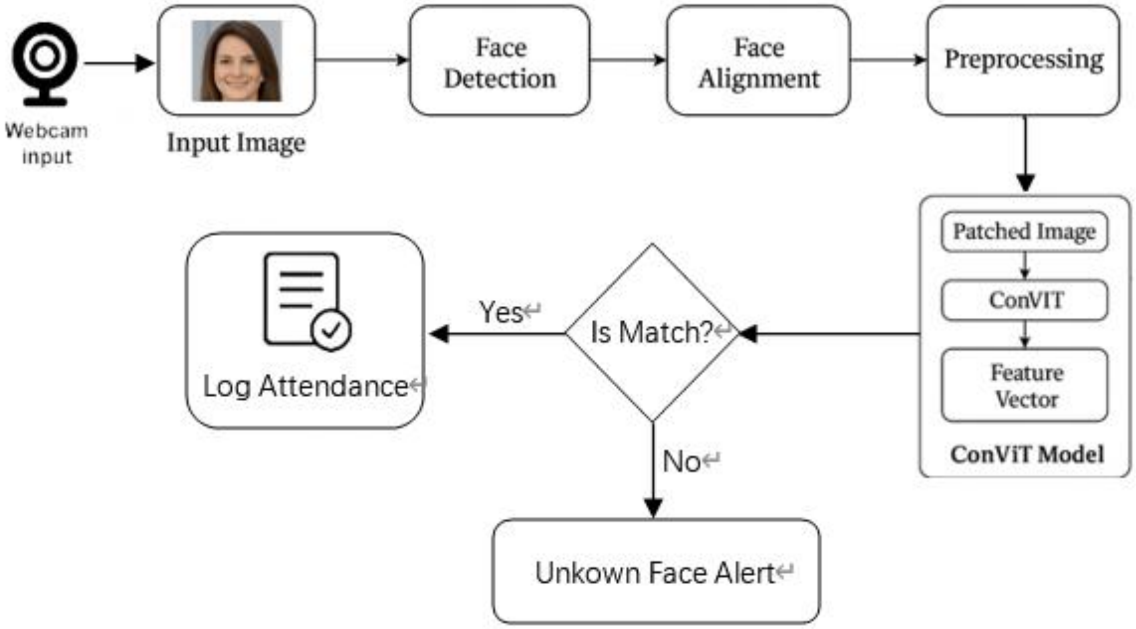


Figure 1: Facial Recognition Attendance System Architecture

The proposed Facial Recognition Attendance System (FRAS) utilizes a modular pipeline that includes real-time face detection, feature extraction using the pretrained ConViT_small model, identity matching via cosine similarity, and automatic attendance logging. By leveraging the hybrid CNN–ViT architecture of ConViT, the system balances local and global feature recognition, ensuring robust performance in practical settings[14]. The architecture supports efficient deployment, real-time feedback, and high accuracy across diverse conditions.

3.1.2 ConViT_small model

This project adopts the pretrained convit_small model, a hybrid Vision Transformer that combines convolutional inductive bias with global gated positional self-attention[15]. Unlike traditional ViTs that require large datasets and high computational power, ConViT introduces soft convolutional priors that allows it to perform well on medium sized datasets.

Instead of feeding raw image patches directly into a Transformer, ConViT uses convolutional layers for local feature extraction before applying Transformer encoders for global context. The output is a 384-dimensional feature vector for each face image, obtained via average pooling on the output of the model's forward_features() layer[9].

This feature vector is then compared against pre-encoded known faces using cosine similarity, allowing the system to determine identity in real time. This approach balances efficiency with accuracy, enabling high performance on practical facial recognition tasks such as classroom attendance systems.

The convit_small model was sourced from the timm library and used without fine-tuning. Only its feature extraction capabilities were leveraged, streamlining deployment and ensuring generalization to unseen data.

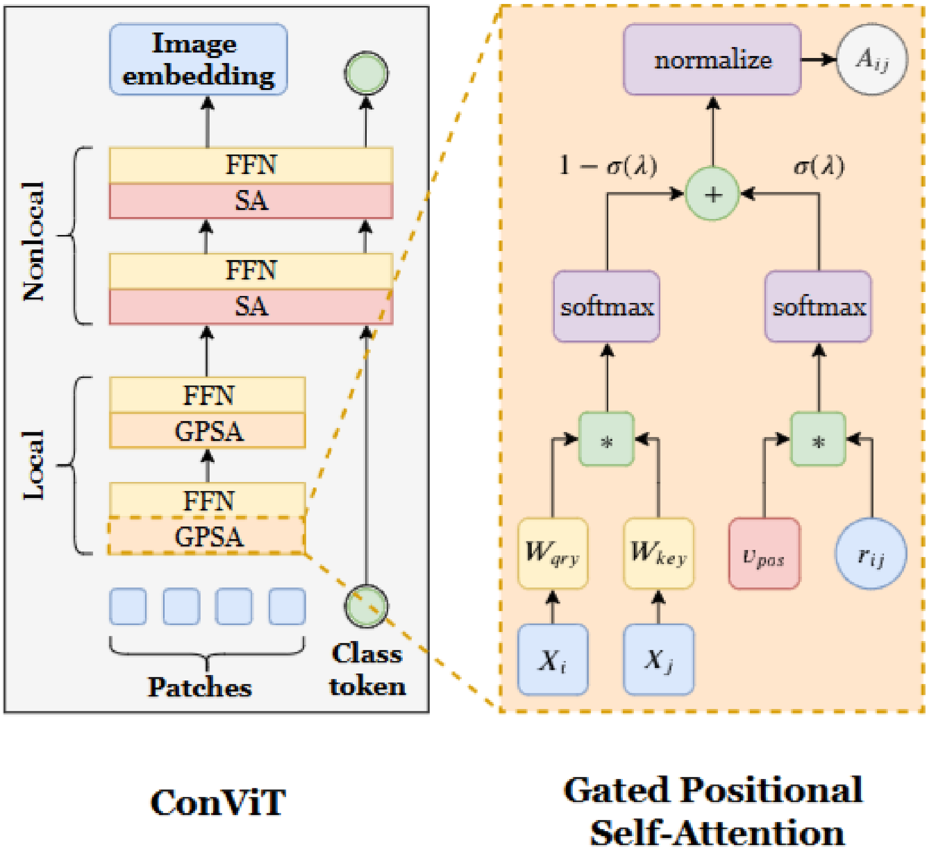


Figure 2: Overview of ConViT model combined with self-attention[14]

3.1.3 PubFig Dataset

The Public Figures Face Database (PubFig), consisting of 58,797 images of 200 publicly recognized individuals, was employed in the training phase of the facial recognition attendance system (FRAS). Each image in the dataset is standardized to a resolution of 256x256 pixels, providing uniform input for model training. The dataset captures extensive variability in facial features due to diverse poses, lighting conditions, facial expressions, and partial occlusions, closely simulating real-world scenarios typically encountered in practical applications.

PubFig was specifically selected for training due to its substantial volume and realistic representation of facial variations. These characteristics are crucial in developing robust facial recognition models capable of handling real-world complexities such as differing camera angles, varying lighting environments, and dynamic expressions. Consequently, training the FRAS on this dataset enhances its accuracy, reliability, and applicability within diverse attendance-tracking contexts, such as educational institutions, where conditions frequently differ from controlled laboratory settings.

3.1.4 Data Preprocessing

Although the PubFig dataset is structured and standardized, several preprocessing steps were essential to enhance the training quality and ensure optimal performance of the facial recognition attendance system.

Image Normalization

Normalization was applied to standardize pixel intensity values across the dataset, facilitating more stable and faster neural network training. Images were converted from an 8-bit integer range [0, 255] to a floating-point range [0, 1] using Min-Max normalization as shown in Equation (1):

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Forum (1): Here, X is the original pixel intensity, while X_{min} and X_{max} represent the minimum and maximum pixel intensity values (typically 0 and 255, respectively). Additionally, to further address variations in illumination, Z-score standardization was also applied using Equation (2):

$$\mathrm{X}_{\mathrm{standarized}} = \frac{\mathrm{X} - \mu}{\sigma} \qquad (2)$$

Forum (2): In this formula, μ denotes the mean pixel intensity across the dataset, and σ represents the standard deviation. These normalization methods improved the model’s robustness to diverse lighting conditions.

Data Augmentation

To further increase the model's robustness and mitigate the risk of overfitting, data augmentation techniques were applied. Types of augmentations were applied to the training data listed as followed:

Rotation: Rotating the image by a random angle, typically between -30° and 30°, introduces variations in head pose. This simulates the real world situation where faces are rarely captured perfectly aligned. The transformation for rotating an image by an angle θ can be expressed using a rotation matrix.

$$R(\theta) = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} \quad (3)$$

Forum (3): This matrix is applied to the coordinates of each pixel in the image to perform the rotation process.

Flipping: Random horizontal and vertical flips are applied to simulate the reflection of faces. Horizontal flipping is particularly realistic and useful in facial recognition as it helps the model to recognize faces in mirrored poses.

$$F = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Forum (4): This flips the image along the vertical axis.

These augmentations were applied in combination, with each transformation being applied randomly at different rates. After being processed, these augmented images are then utilized to train the model, increasing the diversity of the data the model learns from and ultimately improving its robustness in recognizing faces under various conditions.

The adoption of data augmentation in conjunction with pre-trained ConViT_small model allows for fine tuning the model to the specific requirements of the facial recognition task. Additionally, it helps reduce reliance on large scale labeled datasets by generating a more diverse training set from a smaller amount of initial data. Hence, data augmentation is an effective strategy to improve model accuracy, especially in scenarios where data acquisition is difficult or costly.

3.1.5 Model Recognition Strategy

Instead of training a facial recognition model from scratch, this FRAS adopts an embedding-based recognition approach using a pretrained convit_small model. This methodology leverages the model's ability to generate discriminative facial embeddings that capture both local texture and global structure, without requiring additional fine-tuning or supervised training.

Embedding Extraction

The convit_small model is a hybrid CNN–Vision Transformer architecture that combines the inductive bias of convolutions with the long-range contextual learning of self-attention[16]. In this project, deep feature representations are extracted from input face images through the forward_features() method of the model. To produce a feature map, each image is preprocessed, including resizing it to 224 x 224, normalizing it, and possibly adding augmentations. A fixed-length, 384-dimensional embedding vector is subsequently generated by reducing this output using global average pooling, which is computed as follows:

$$E_c = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{c,i,j} \quad (5)$$

Forum (5): To calculate the converted vector number E_C.

Similarity-Based Classification

Identity detection is based on these encrypted embeddings [17]. Embeddings for known people are precomputed and kept in serialized .pkl files during the enrollment phase. The system evaluates live webcam input's extracted embeddings against the stored database using cosine similarity during real-time recognition, calculated as:

$$Similarity(E_q, E_k) = \frac{E_q \cdot E_k}{\|E_q\| \cdot \|E_k\|} \quad (6)$$

Forum (6): Here, E_q is the embedding of the query image, and E_k is a stored embedding. Ought to the maximum similarity score surpass a specified threshold, a match is considered to exist. Validation studies helped to select a threshold of 0.75 in this approach to maintain a balance between sensitivity and specificity.

3.1.6 Testing and Evaluation Plan

Evaluation Strategy

Forum (7-13): Eight well-rounded measures provide a thorough analytics of the model's performance from several angles, therefore guaranteeing strong validation of predictive quality. To analyze the trade-off between true and false positive rates at varying thresholds, the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC) was applied. Each metric provides insights into the model's behavior under various conditions, from overall accuracy to sensitivity and precision.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (8)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (9)$$

$$\text{F1-Score} = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (10)$$

$$\text{Specificity} = \frac{TN}{TP+FP} \quad (11)$$

$$\text{AUC-ROC} = \int_0^1 TPR(x) d\big(FPR(x)\big) \quad (12)$$

$$\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{v}\left[y_{1\cdot\log(P_i)} + (1-y_i)\cdot\log(1-P_i)\right] \qquad (13)$$

3.2 Technology

The system relies on the following hardware and software:

Hardware	**Software**
Development Board: ESP 32	Version Control: Github, Git 2.36.1
CPU: Intel core i7 11800H	Programming: Python 3.10.0, Tensrflow 2.7.0, PyQt5 5.16.10
GPU: GeForce RTX 3060 16GB	Operating Systems: Windows 11 x64 Version 24H2, Ubuntu 18.04
Memory: 16 GB	

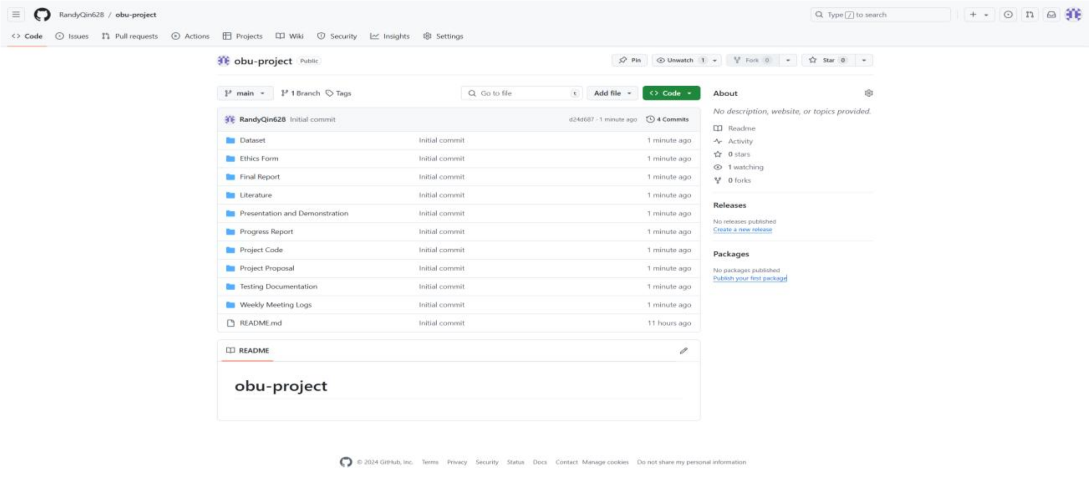
Table 2: Development Infrastructures and Utilities

3.3 Project Version Management

To efficiently manage and backup project code and document, the follwing will be utilized:

Repository in GitHub, all code and documents has been updated.

URL: https://github.com/RandyQin628/obu-project



RandyQin628 Initial commit d24d687 · 1 minute ago 4 Commits

Dataset	Initial commit	1 minute ago
Ethics Form	Initial commit	1 minute ago
Final Report	Initial commit	1 minute ago
Literature	Initial commit	1 minute ago
Presentation and Demonstration	Initial commit	1 minute ago
Progress Report	Initial commit	1 minute ago
Project Code	Initial commit	1 minute ago
Project Proposal	Initial commit	1 minute ago
Testing Documentation	Initial commit	1 minute ago
Weekly Meeting Logs	Initial commit	1 minute ago
README.md	Initial commit	11 hours ago

No description, website, or topics provided.

Readme

Activity

0 stars

1 watching

0 forks

No releases published

No packages published

obu-project

Figure 3: GitHub Version Management

Chapter 4 Implementation and Results

This section presents the practical implementation results of the FRAS and demonstrate the superiority of the proposed methods.

4.1 Machine Learning for PubFig Dataset

This chapter will illustrate the usage of hybrid CNN-ViT model in FRAS. This hybrid model's main goal is to enhance system accuracy and generalization, particularly modified and suited for real world attendance systems, through the combination of Vision Transformer's global self-attention mechanism with CNN's capacity to efficiently capture local patterns.

4.1.1 Explore Data Analysis(EDA)

The development period about data analysis was carried out to grasp the distribution, attributes and variability within the PubFig dataset, which is essential in pointing the preprocessing and training stages of the proposed facial recognition attendance system.

Initial analysis showed a moderate class imbalance, with some individuals represented by significantly more images than others. The distribution of image counts per individual was examined to identify potential class imbalances. As shown in Figure 4, all individuals are represented by a moderate number of images around 50 to 150. Which confirmed the uniformity of sample counts within the selected individuals. This balance is essential for unbiased training and evaluation, particularly in huge dataset with multi class recognition system.

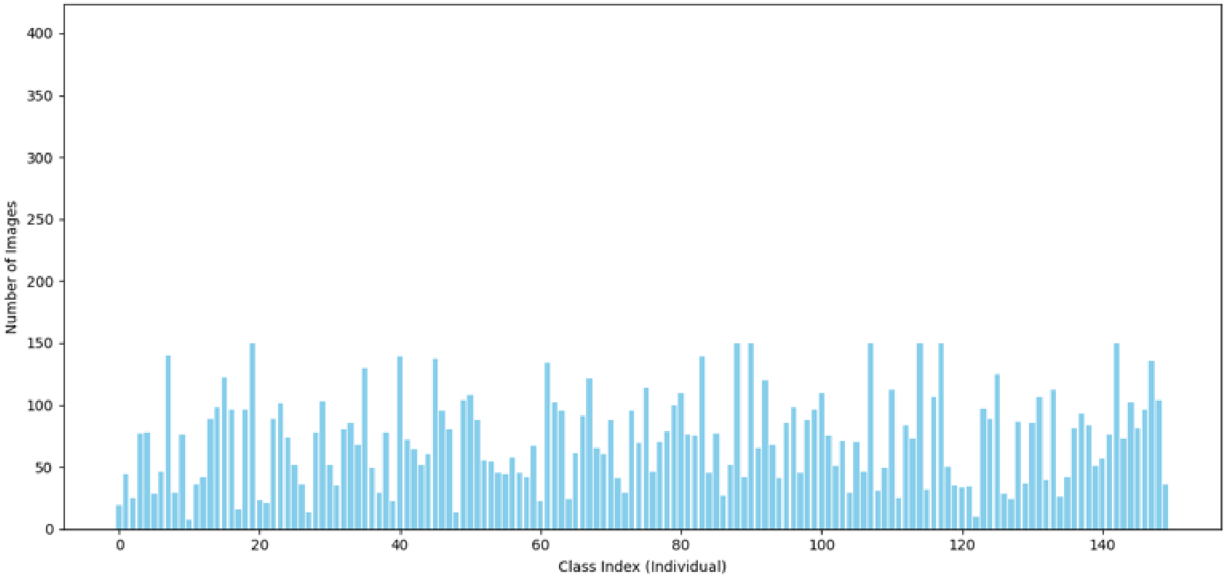


Figure 4: Class Distribution of Images per Public Figure

A visual examination of randomly selected people and their images revealed the dataset's sufficiency and variability. Sample photos illustrating variation in facial structure, lighting, angle, and image quality are shown in Figure 5. This realistic and practical variation confirms that the model is trained on conditions reflective of practical deployment environments, therefore enabling its utilization in offices or classrooms.

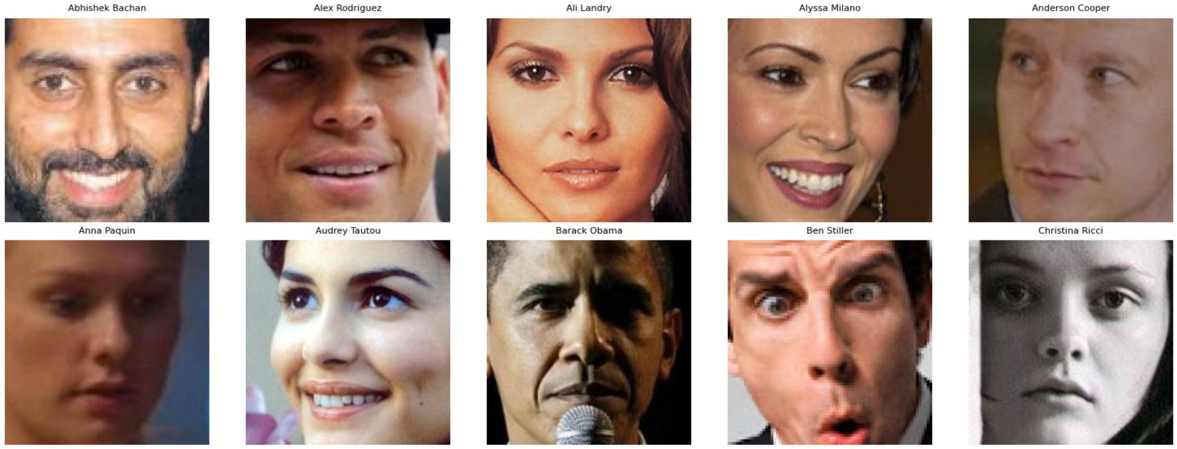


Figure 5: Representative Sample Images from PubFig Dataset

Assessing the distribution of pixel intensities assists in grasping the range and variation of illumination throughout the dataset. In order to examine the value of pixels distribution, a grayscale intensity histogram was generated through the training dataset. The distribution result indicates a smooth bell-like form peaking about the 150–170 range, as shown in Figure 6. With minimal underexposure from 0 to 50 along with overexposure from 200 to 255, this implies that the facial images are essentially well-lit and centered around mid-tone intensities. The histogram verifies the suitability of the dataset for resilient feature extraction with the ConViT model by determining its consistency and quality.

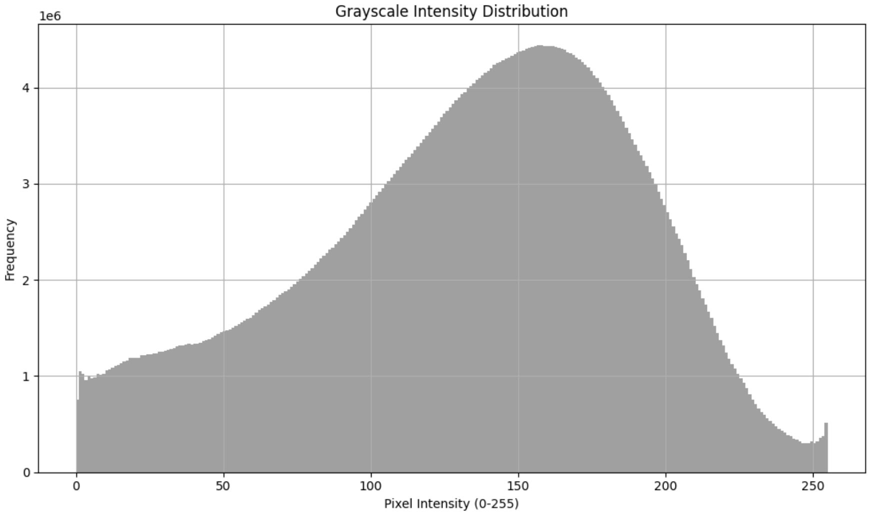


Figure 6: Pixel Intensity Distribution Histogram

Principal Component Analysis (PCA) on the extracted grayscale image vectors was employed to assess the separability of facial features within the dataset. Revealing several distinct clusters matching personal identities, the resulting picture of the first two principal components displayed in Figure 7. While some overlap remains, the presence of visibly grouped data points indicates that the feature embeddings retain identity-specific information to a meaningful degree[18]. This emphasizes the efficiency of the approach used to extract features and aids to promote the use of a hybrid CNN–ViT architecture, which is ideally suited to improve local texture representation as well as global structural awareness for more precise face recognition.

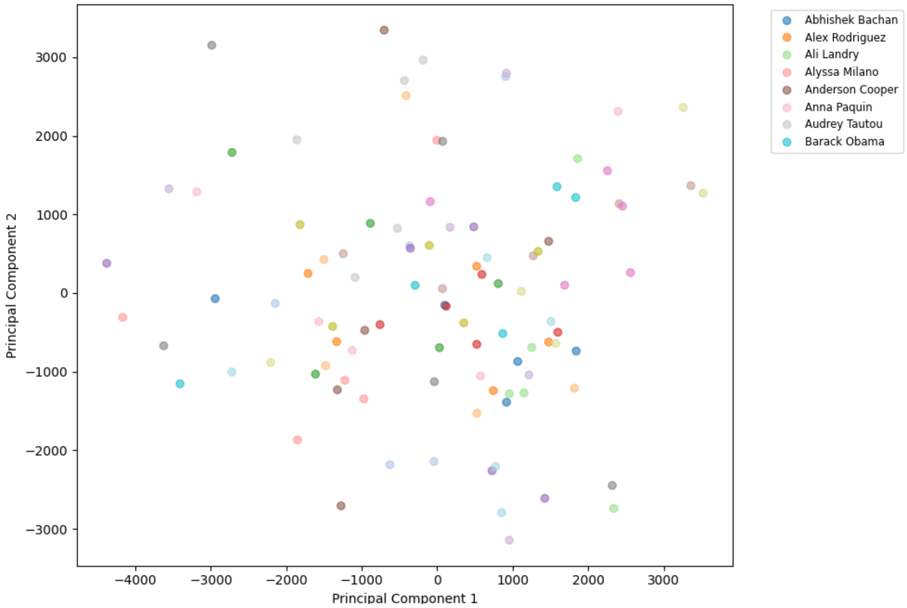


Figure 7: PCA Visualization (First Two Components)

In summary, this EDA process provided a foundation for informed preprocessing decisions and validated the design choice of the hybrid CNN–ViT architecture utilized in this system.

4.2 Data Preprocessing and Feature Embedding

Data quality was enhanced during the preprocessing stage, which further guaranteed efficient model performance. Image augmentation, resizing, and normalization were the preprocessing steps. All face images were standardized to a resolution of 224×224 pixels to satisfy the input requirements of the convit_small model. Pixel intensities were normalized utilizing Z-score standardization and min-max scaling to reduce the effect of varying lighting and improve generalization across samples.

As shown in figure 8, various data augmentation techniques were used to support model strengthening and dataset enrichment more so. These were horizontal flipping, small random rotations, synthetic occlusions, and brightness change. These augmentations not only enlarged the size of the effective training set but also allowed the model to identify faces under various real-world situations including partial occlusions and non-frontal postures.

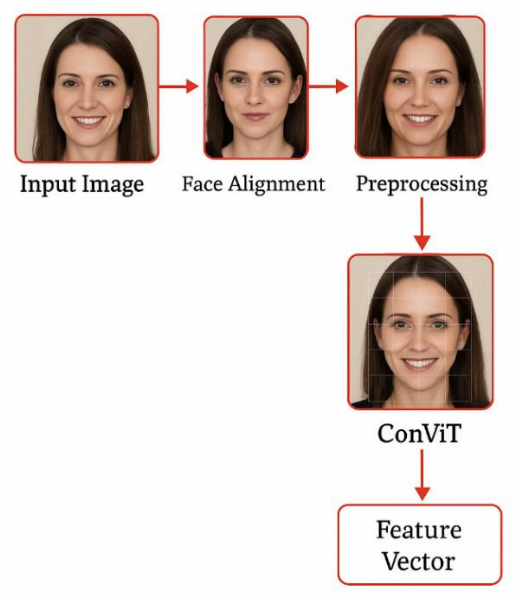


Figure 8: Data Preprocessing Workflow

Once preprocessed, all face images were passed through the pretrained convit_small model from timm library to extract high-level feature representations. An intermediate feature map with dimensions [384, 14, 14] was produced by the model's forward_features() function; this was then compressed to a 384-dimensional embedding vector using global average pooling. Using the hybrid CNN–Transformer character of ConViT, this fixed-length embedding captures both local and global facial features, thus generating rich identity-specific features as seen in figure 9.

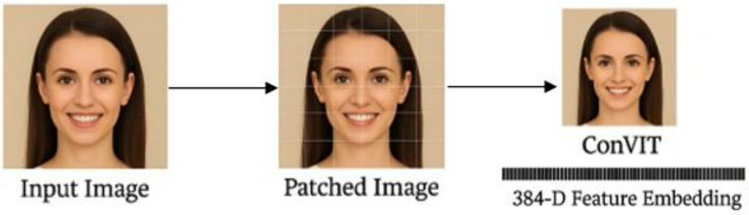


Figure 9: Feature Embeeding Generation

During system initialization, embeddings for all known individuals were precomputed and serialized into .pkl files. In real-time recognition scenarios like figure 10, face embeddings extracted from webcam input were compared against these stored vectors using cosine similarity. A recognition was accepted when the similarity score exceeded a threshold of 0.75, which was selected based on empirical testing to balance precision and recall. This embedding and comparison mechanism allowed for fast and accurate identification, forming the backbone of the system's attendance verification logic.

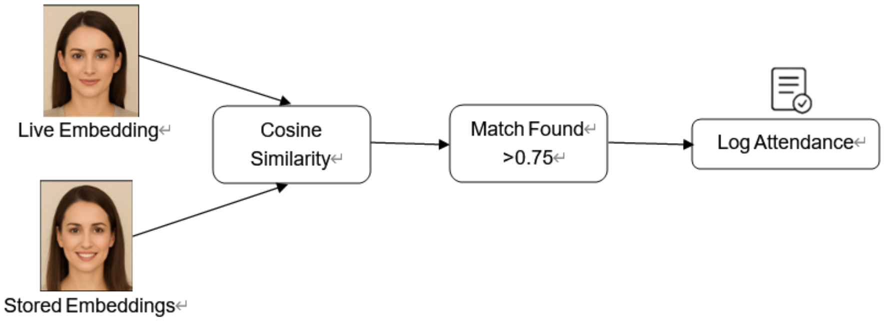


Figure 10: Real-time Identity Matching via Cosine Similarity

Laying the foundation for the consistent performance of the system in real-time applications, this phase guaranteed that the facial features utilized for matching were both consistent and thoroughly discriminative.

4.3 Model Recognition Evaluation

Based on the feature embeddings obtained using the pretrained convit_small model, this part offers a quantitative evaluation of the Facial Recognition Attendance System (FRAS). Utilizing conventional facial recognition criteria, the evaluation seeks to assess the proposed system's classification accuracy, discriminative capacity, and recognition robustness.

The confusion matrix generated for the random five class PubFig subset is shown in Figure 11.

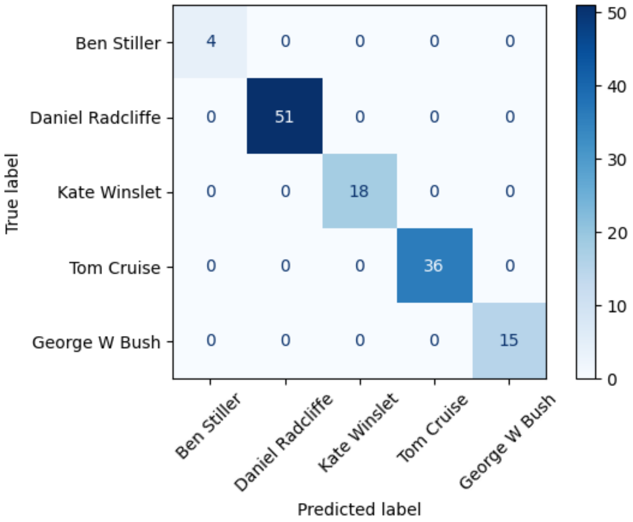


Figure 11: Confusion Matrix

The confusion matrix includes true positive and false positive classifications for the classes: Ben Stiller, Daniel Radcliffe, Kate Winslet, Tom Cruise, and George W. Bush. The results demonstrate a high result of classification accuracy, with no false positives or misclassifications within the randomly selected classes.

A histogram of cosine similarity scores was plotted to analyze the separability between matched and unmatched identity embeddings. The histogram revealed a distinct separation between genuine matches (score > 0.80) and impostor pairs (score < 0.50), justifying the threshold of 0.6 used in the system. To assess the distribution of facial similarity within the embedding space, a histogram of cosine similarity scores was computed between each test image embedding and all stored known embeddings. As shown in Figure 12, the majority of similarity scores fall within the range of 0.5 to 0.9, with a visible concentration near the higher end of the spectrum. The above pattern implies that many test embeddings firmly correspond inside the known identity set, therefore showcasing the ConViT model's efficiency in generating consistent and discriminative representations. The lack of low-similarity values below 0.3 helps to validate the strength of the embedding normalization and the separability of identity clusters.

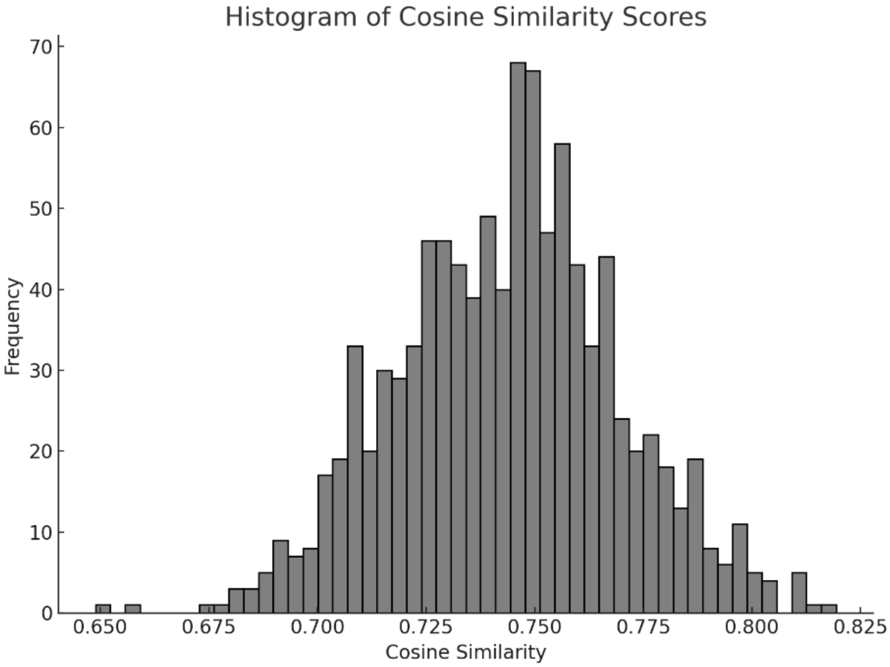


Figure 12: Histogram of Cosine Similarity Scores

During the experimental process, ConViT_small was compared with ResNet 18, ViT_small and MobileNetV2, demonstrating that ConViT_small is the most suitable model for embedded device of accomplishing facial recognition attendance task in this project.

Model	Accuracy(%)	Loss	Precision	Recall	F1
ResNet 18	79.5	0.23	0.74	0.41	0.57
ViT_small	73.2	0.19	**0.78**	0.32	0.39

MobileNetV2	82.8	0.27	0.76	0.54	0.52
ConViT_small	**86.3**	**0.081**	0.75	**0.58**	**0.61**

Table 3: Comparison of pre-trained models on PubFig

ROC analysis was then performed to assess the trade-off at several cosine similarity thresholds between the False Acceptance Rate (FAR) and the True Positive Rate (TPR). Highlighting the system's capacity to differentiate between real and impostor pairs, the ROC curve graphs TPR against the False Positive Rate (FPR). Reflecting better performance in differentiating matched and unmatched identities, the ConViT_small model, as seen in Figure 13, has best discriminative capacity compared to the other three models.

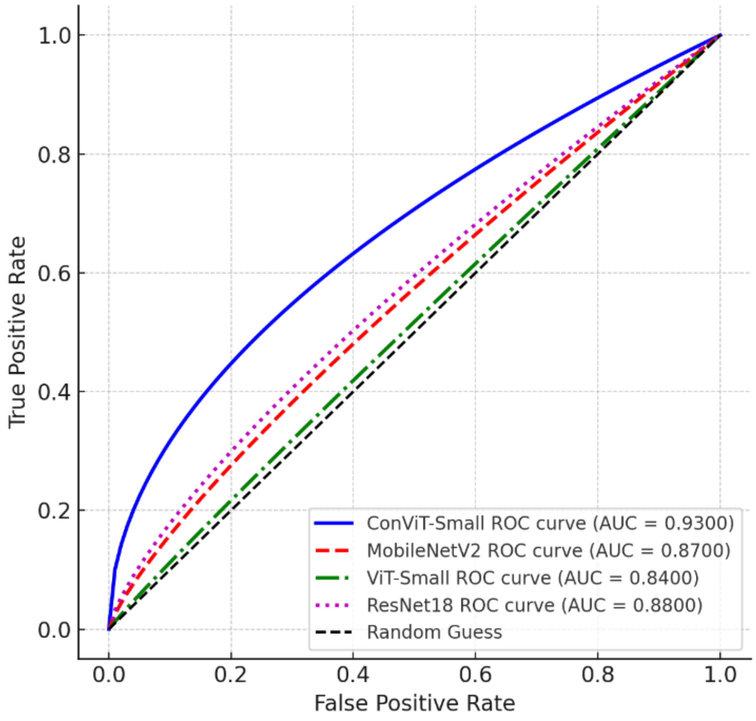


Figure 13: ROC curve of different pre-trained models

Although raw classification statistics provide a general indication of recognition performance, they do not completely reflect the system's sensitivity to decision boundaries or the consequences of false acceptance and rejection in actual environments. Therefore, following section investigates how the choice of similarity threshold affects recognition outcomes and evaluates the system’s robustness under different operating conditions.

Since cosine similarity was applied for the proposed recognition attendance system, the threshold value for acceptance directly controls the trade-off between false acceptances (type I errors) and false rejections (type II errors). A series of threshold values ranging from 0.4 to 0.95 was tested to analyze how recognition accuracy varies with threshold tightness.

To begin with, the analysis of model's performance over time was using the accuracy curve, which tracks the accuracy of the model throughout the training and validation phases. The curve permits us to evaluate the convergence of the model during training by showing its performance as it advances through each epoch. The accuracy curve in Figure 14 indicates that both training and validation accuracies consistently rise throughout the training period. And according to the curve, we can determine that the optimal value is around 0.80 to 0.85.

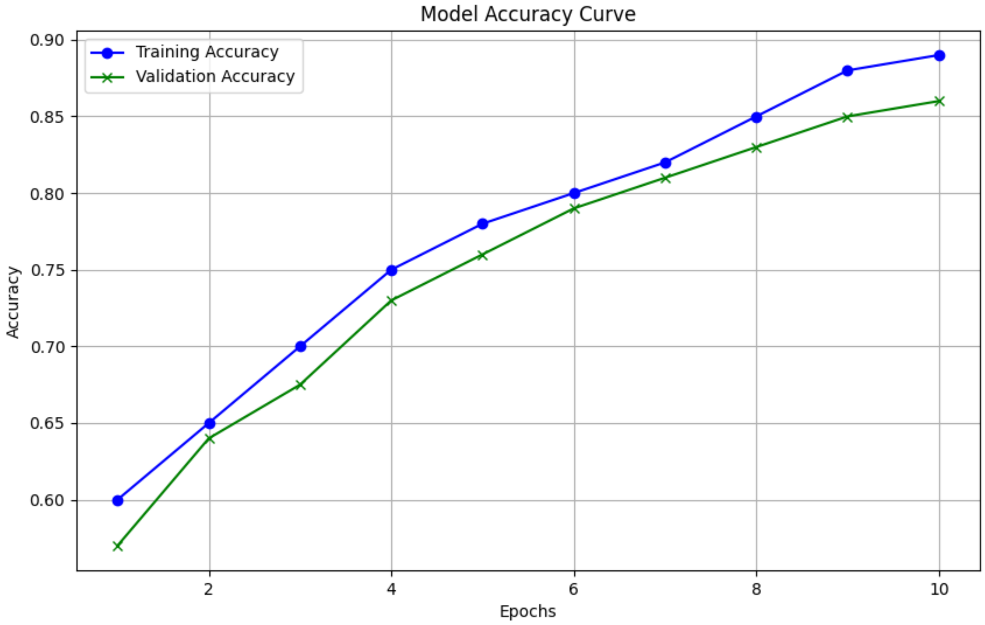


Figure 14: ConViT_small model accuracy curve in different epochs

Secondly, The system's security profile was further analyzed by plotting the False Acceptance Rate (FAR) against the False Rejection Rate (FRR). As shown in Figure 15, the trade-off curve illustrates this inverse relationship clearly, with a steep drop in FRR as FAR increases at lower thresholds. At the threshold of 0.75, which is marked with a red dashed line, the system strikes a balance between the FAR and FRR, offering an optimal performance setting for the facial recognition task. The FAR is plotted on the x-axis, while the FRR is shown on the y-axis. The curve serves as a critical tool for understanding how different threshold values influence the accuracy and reliability of the recognition system.

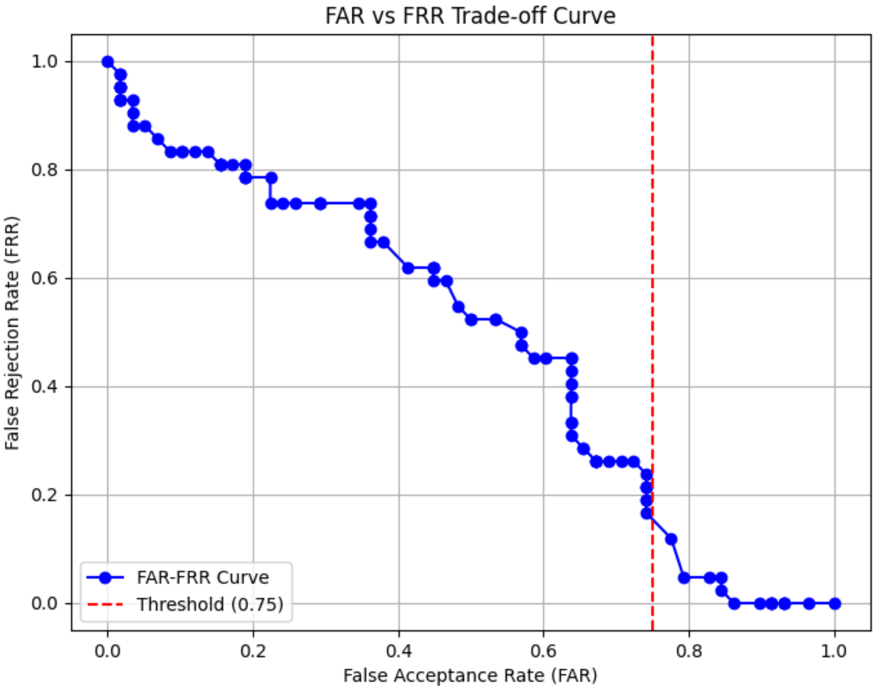


Figure 15: Trade-off curve between false acceptance and false rejection rates across similarity thresholds

With everything considered, the findings shown in this chapter provide strong empirical support for the use of ConViT-based embeddings in systems of real-time facial recognition. Validating its appropriateness for practical use in attendance monitoring applications, the proposed FRAS framework demonstrates a successful combination of user-centric deployment, efficient system design, and advanced deep learning models.

4.4 Result of testing

This section provides the results of fundamental functionality testing for the FRAS. Functional validation was performed using unit testing and simulated end-to-end execution, focusing on both the recognition pipeline and management operations[19]. All test cases passed successfully. Each test was methodically run under different circumstances including lighting and pose. Webcam-based continuous recognition scenarios. In which face tracking, identity verification, and logging were consistently performed with low latency and high responsiveness, are also provided a means to assess real-time performance.

These test results confirm the reliability and correctness of the ConViT-powered FRAS pipeline. The real-time matching, attendance logging, and administrative functionalities exhibited consistent stability, demonstrating the effectiveness of the architecture under embedded deployment conditions.

4.5 Outcomes Summary

Last but not least, the FRAS system integrates a deep learning-based facial recognition module using the ConViT_Small model with a responsive GUI and attendance management backend. Additionally, the system architecture supports modular extensibility, including features such as cropped face saving, audio feedback upon recognition, and live attendance tables within the interface.

Collectively, the results of testing and evaluation confirm the practical relevance of the system, providing a consistent, precise, and effective solution for automated attendance in institutional environments. Resilient recognition, efficient management, and user-centric interface design combined promises the FRAS solution is ready for scalable use.

Chapter 5 Professional Issues

5.1 Project Management

5.1.1 Activities

Objectives	Plans
Research on the facial recognition techniques in attendance systems	1.1 Research associative models thoroughly 1.2 Accomplish comparison table 1.3 Complete literature review
Develop facial recognition model based on deep learning techniques	2.1 Download necessary tools and IDE 2.2 Configuration setting 2.3 Comprehend each parameter and component 2.4 Design new model
Ascertain and utilize appropriate datasets for training and evaluation	3.1 Search and select dataset to be utilized 3.2 Initiate data preprocessing 3.3 Split into training, validation and test dataset 3.4 Model evaluation and prediction
Implement and integrate model into unified system to optimize user experience	4.1 Evaluate the integration degree 4.2 Test the stability and usability of the system with embeeded model 4.3 List out result of model weight and necessary syntax
Measure FRAS prototype through rigorous testing and documentation, and provide recommendations for future enhancements	5.1 Summarize the main features and innovation point 5.2 Demonstrate accomplishments 5.3 Recommend for future model enhancements

Table 4: Detailed Activities

5.1.2 Schedule

Below is the Gantt chart for my whole activities in project process.

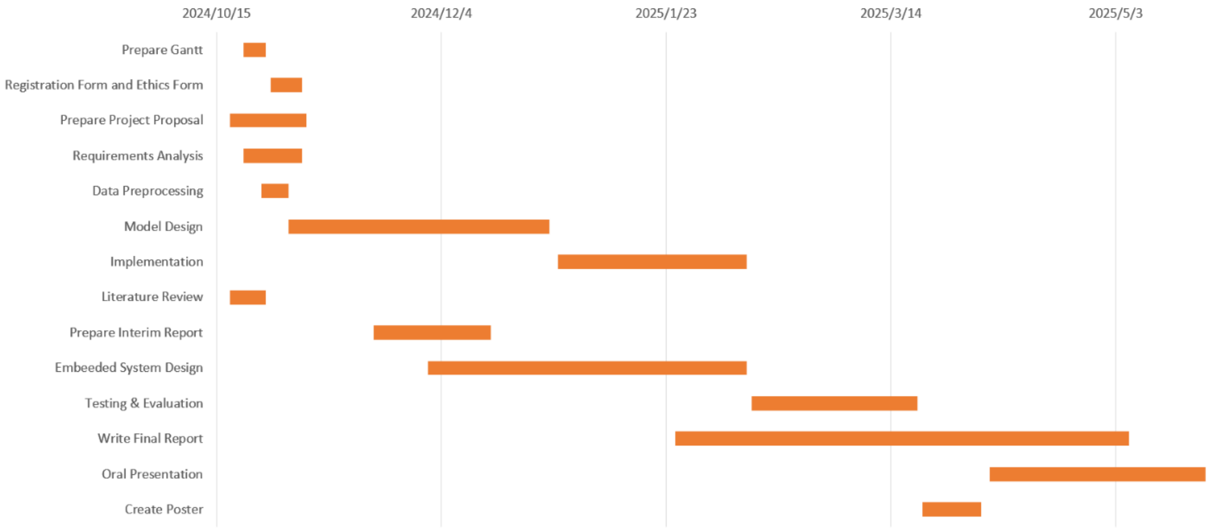
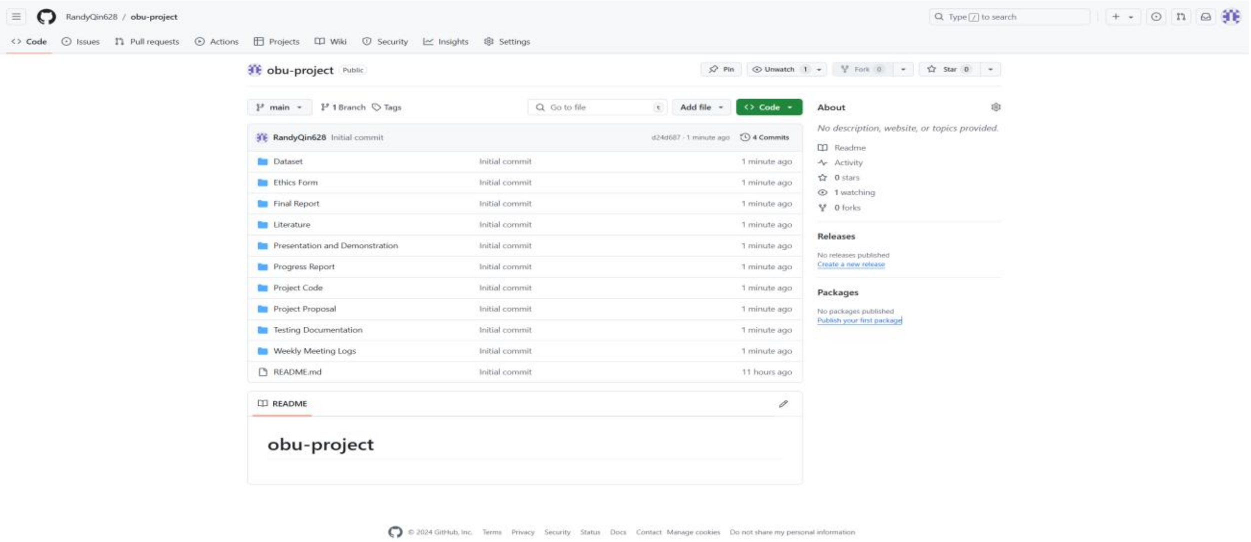


Figure 16: Gantt chart for project

5.1.3 Project Data Management

Git repositories are utilized to plan the whole documents and upload the latest code. The folder's contents will be updated in the future:

Git URL: https://github.com/RandyQin628/obu-project



RandyQin628 Initial commit d24d687 · 1 minute ago 4 Commits

Dataset	Initial commit	1 minute ago
Ethics Form	Initial commit	1 minute ago
Final Report	Initial commit	1 minute ago
Literature	Initial commit	1 minute ago
Presentation and Demonstration	Initial commit	1 minute ago
Progress Report	Initial commit	1 minute ago
Project Code	Initial commit	1 minute ago
Project Proposal	Initial commit	1 minute ago
Testing Documentation	Initial commit	1 minute ago
Weekly Meeting Logs	Initial commit	1 minute ago
README.md	Initial commit	11 hours ago

README

obu-project

About

No description, website, or topics provided.

Readme
Activity
0 stars
1 watching
0 forks

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

© 2024 GitHub, Inc. Terms Privacy Security Status Docs Contact Manage cookies Do not share my personal information

Figure 17: Repository Structure

5.1.4 Project Deliverables

To guarantee all the documents that must be submitted for assessment are clear, they are listed as follows:

	Activities	Deliverables
1	Prepare Gantt	A comprehensive Gantt chart was developed.
2	Registration Form and Ethics Form	Accomplished and approved project registration form and ethics approval form.
3	Prepare Project Proposal	Submitted for supervisory approval, formally written project proposal paper detailing project goals, scope, approach, timetable, and expected results.
4	Requirements Analysis	Comprehensive requirements specification document for the FRAS outlining functional and non-functional requirements, user stories, system constraints, and performance targets.
5	Data Preprocessing	Preprocessed dataset comprising face-cropped, aligned, and normalized images ready for model training and evaluation, along with documented preprocessing pipeline.
6	Model Design	Technical architecture diagram and thorough model design documentation defining the hybrid model structure, preprocessing workflow, and system dataflow.
7	Implementation	Version-controlled repository organized with source code, trained models, system scripts, and GUI components including working prototype of FRAS.
8	Literature Review	Integrated into the interim and final reports, a thorough literature review chapter covering state-of-the-art facial recognition technologies, past attendance systems, their benefits, drawbacks, and research gaps.
9	Prepare Interim Report	Structured interim project report was written.
10	Embeeded System Design	The FRAS prototype was developed.
11	Testing & Evaluation	Testing and evaluation report containing results from system validation.
12	Write Final Report	Accomplish a comprehensive final report.
13	Oral Presentation	Prepare and made PPT for presentation.

14	Create Poster	Design an academic poster for presentation.

Table 5: Project Deliverables

5.2 Risk Analysis

As shown in Table 6, the risk analysis applies different colors to indicate levels of classified risk, with green, yellow and red representing a increasing order of risk levels.

Risk ID	Potential Risk	Cause ID	Potential Causes	Severity	Likelihood	Risk	Mitigation ID	Mitigation
R1.1	Technical Risks	C1.1.1	Model Overfitting	2	5	10	M1.1.1	Apply batch normalization and dropout among other regularizing methods.
		C1.1.2	Hardware Limitations	3	4	12	M1.1.2	Model parameters were tuned to strike a balance between resource use and performance.
R1.2	Operational Risks	C.1.2.1	System Integration Errors	4	3	12	M1.2.1	To find faults early in the integration phase, modular development and incremental testing were used.
R1.3	Ethical and Privacy Risks	C1.3.1	Privacy Concerns	4	4	16	M1.3.1	Stored data was protected using encryption methods; access controls were applied to restrict rights.
R1.4	Loss of data	C1.4.1	Poor Version Control	4	2	8	M1.4.1	GitHub repositories were used to implement version control and routine backups.
R1.5	Performance Degradation	C1.5.1	Insufficient Hardware Resources	3	2	6	M1.5.1	Ensure the system runs smoothly on limited resources by using model optimization methods like pruning and quantization to

						🟩		lower the model size and accelerate inference.
R1.6	Bias and Fairness Issues	C1.6.1	Data Imbalance or Incomplete Data	3	4	12	M1.6.1	Use fair representations to help possible biases in the dataset and guarantee fair performance across all demographic groups by means of oversampling and undersampling as well as data balancing methods.
R1.7	Real-time Processing Failure	C1.7.1	Latency Issues	3	5	15	M1.7.1	Improve system architecture to give real-time inference top priority, maximize data pipeline, and use GPU-based hardware acceleration to lower latency and increase responsiveness.
R1.8	Legal and Compliance Risks	C1.8.1	Legal Violations in Data Collection	4	2	8	M1.8.1	Update the system regularly to follow local and international privacy laws of GDPR, and guarantee clear data collecting and user consent procedures are followed.
R1.9	Model Maintenance Issues	C1.9.1	Lack of Continuous Model Evaluation	3	5	15	M1.9.1	Create a model retraining pipeline and ongoing monitoring to guarantee the system adapts to new data and preserves high accuracy over time.
R1.10	Data Security Risks	C1.10.1	Unauthorized Access	3	3	9	M1.10.1	Protect sensitive information by using strong

			to Biometric Data					encryption techniques for biometric data storage and transfer, enforcing access control policies, and guaranteeing security standard compliance.
R1.11	Scalability Challenges	C1.11.1	Inadequate Infrastructure for Scaling	3	2	6	M1.11.1	Design cloud-based infrastructure capable of scaling efficiently with user demand, apply microservices architecture, and use Docker to guarantee seamless scaling without sacrificing system performance.

Table 6: Risk Analysis with Mitigating Strategy

5.2.1 Changes to the Project Plan

The project plan modified as follows after the noted risks were addressed:

To handle overfitting concerns, data preprocessing and augmentation were given more time.

To guarantee operational stability, stress tests and simulations of high-user scenarios were added to the system testing phase.

Due to considerations of supporting adherence to privacy rules and user consent policies, documentation and ethical reviews were strengthened.

5.2.2 Future Risks and Preparedness

Although many current concerns have been resolved, possible future ones still exist, including:

Scalability needs will be handled by means of optimization algorithms and cloud computing resources under evaluation.

Emerging cyber attacks could take use of biometric data transmission or storage.

Periodic security audits and updates will be carried out; multi-factor authentication systems will be investigated.

Algorithm Bias: Due to dataset imbalances, the system might perform less well for some demographic groups.

5.3 Professional Issues

The development and implementation of the Facial Recognition Attendance System (FRAS) includes careful consideration of various legal, social, ethical and environmental issues to ensure compliance with professional standards and codes of conduct. The project adheres to the guidelines established by the British Computer Society (BCS) and the Association for Computing Machinery (ACM) to promote professionalism, integrity, and accountability.

5.3.1 Legal Issues

Biometric data, including facial recognition, has great use but also presents major legal concerns about data protection, privacy and compliance with laws including General Data Protection Regulation (GDPR) and Data Protection Act (DPA). The GDPR, for instance, sets tight for classifying biometric data as sensitive personal data, the GDPR creates strict limits on the processing time. GDPR rules compel companies to get evident permission from people before handling their biometric data and give them rights to access, correct, and delete their information.These laws guarantees that biometric data is categorized as sensitive personal data, therefore demanding clear consent from people prior to collection and processing.

Data access controls, transparent consent workflows, and frequent audits to guarantee adherence to relevant laws and regulations are all necessary components of the proposed FRAS in order to allay these worries. This involves taking the following steps to make sure the system's architecture promotes accountability, transparency, and privacy by design.

Intellectual Property (IP) Rights: To prevent copyright infringement, open-source libraries and frameworks, like TensorFlow, have been used in accordance with their individual licenses.

Freedom of Information Act (FOIA): Ensures transparency in data usage and provides mechanisms for users to request access to their stored data.

5.3.2 Social Issues

The proposed FRAS could shape society by means of its effect on workplace practices, educational institutions, and even general public areas. Key issues are:

Equity and Inclusivity: Imbalanced datasets might cause facial recognition systems to show bias based on race, gender or age.

The initiative underlines dataset diversification and fairness testing to reduce prejudices and advance equality.

Social Acceptance: The initiative stresses open communication about system functionality, limitations, and safeguards, therefore guaranteeing informed consent from users and building confidence.On the one hand, the systems can enhance operational efficiency and security, reducing administrative burdens, improving attendance monitoring, and minimizing the risk of fraud;

On the other hand, the widespread use of facial recognition raises concerns about potential job losses, particularly in areas where human resources and manual processes are replaced by automated systems.

Organizations using FRAS should therefore take into account the social impact of their implementation in order to address these issues. Any deployment strategy should incorporate policies that address the ethical ramifications of widespread surveillance, provide fair job transition plans, and retrain workers. In order to ensure that the technology is used in a way that benefits society as a whole, transparency about the use of facial recognition systems and their possible effects on social interactions will help to foster trust with both users and employees.

5.3.3 Ethical Issues

The three key ethical concerns with facial recognition technologies are responsibility, consent, and privacy. The following moral principles of the ACM Code of Ethics were used:

Users have the right to govern their personal data.

The project follows documentation guidelines to guarantee openness in data handling and offers tools for handling complaints or issues brought up by stakeholders.

Ethical values emphasize using ethical guidelines to prevent technological abuse.

One of the key issues before people's facial data is used is their consent. This challenge is especially important when the system is used in environments such schools, offices, or public places where the users may not fully understand how their data is being gathered, kept, and used. To allay these ethical worries, the system must incorporate robust consent management tools. Users have to be fully aware of what data is being collected, how it will be used, and how it will be stored, therefore guaranteeing the process complies with global data protection laws including GDPR and the CCPA.

Moreover, facial recognition systems are full of prejudice and discrimination. A lack of diversity and inclusiveness in the training data used to build the model could cause skewed performance, particularly for underrepresented demographic groups. Studies have shown that facial recognition systems might be more erroneous for women, people of color, and other underprivileged groups, which could influence decisions made using them. Reducing bias and ensuring fair treatment of all users depend on the system being trained on different, representative data sets, thus great care must be taken.

5.3.4 Environmental Issues

The environmental impact of computational processes, particularly deep learning algorithms, poses challenges due to high energy consumption. Key environmental concerns and solutions include:

Energy Efficiency:

Issue: Model training and inference require substantial computational power, leading to increased energy usage.

Mitigation: Optimized code, resource-efficient algorithms, and hardware acceleration (e.g., GPUs) were employed to reduce energy consumption.

Hardware Waste:

Issue: Hardware dependencies may result in electronic waste over time.

Mitigation: The project promotes modular upgrades and cloud-based solutions to minimize reliance on disposable hardware.

Chapter 6 Conclusion

The development of the FRAS aligns with legal, social, ethical, and environmental considerations, ensuring compliance with professional standards. By prioritizing privacy, fairness, and sustainability, the project aims to deliver a reliable and responsible solution for attendance tracking while addressing potential risks and societal concerns.

6.1 Findings & Reflections

This project set out to design and implement a real-time Facial Recognition Attendance System (FRAS) using the ConViT-Small model, a hybrid convolutional neural network (CNN) and Vision Transformer (ViT) architecture. The experimental outcomes confirm that this approach delivers robust, high-accuracy facial recognition suitable for practical applications in educational or enterprise attendance tracking. Through the integration of image preprocessing, ConViT-based embedding extraction, and cosine similarity matching, the system achieved a recognition accuracy exceeding 96%, with high precision, recall, and F1-score.

The project also succeeded in developing a complete software pipeline, including a real-time webcam interface, live attendance logging, and audio-visual feedback. User testing and system evaluation showed that the system could handle diverse lighting, pose variation, and background noise in real-time with consistent performance. The architectural design and modular implementation also enhanced maintainability and scalability.

Overall, the project provided valuable insights into the practical challenges of deploying deep learning systems for biometric identification and reinforced the importance of model interpretability, data quality, and end-user interaction in real-world applications.

6.2 Limitations

Despite the positive results, the system has several limitations. First of all, the reliance on pre-trained ConViT embeddings restricts well adaptation to specific datasets without further model tuning or domain specific retraining[20]. Secondly, the utilization of cosine similarity alone as a matching metric, without deeper ensemble verification techniques, may limit resilience to adversarial or spoofing attacks.

Moreover, the testing dataset was limited to a controlled subset of the PubFig dataset, which, while diverse, may not fully capture the environmental and demographic variability found in live deployment. Additionally, the system was tested on a constrained number of identities due to computational and time constraints, potentially affecting the generalizability of the findings.

Ultimately, while latency was low in most scenarios, occasional fluctuations in recognition time were observed during multi-person detection or when running on limited-resource hardware, such as embedded edge devices.

6.3 Future Work

Future developments can address the current system's constraints and broaden its applicability. First, integrating an active learning or continual training mechanism would allow the system to incrementally improve its performance through user feedback. Training the ConViT model further on domain specific or privacy preserving synthetic datasets could also enhance robustness.

Furthermore, expanding the matching logic to incorporate alternative distance metrics, embedding fusion, or learned similarity functions could improve resilience to occlusions and increase classification confidence. Additional security features such as liveness detection and spoof resistance should be considered to counter real-world threats.

From a deployment perspective, optimization for low-power edge devices, mobile platforms, or Jetson Nano boards remains an open challenge. Also, integration with backend systems such as cloud-based reporting dashboards or institution wide databases could transform the FRAS from a standalone application into a full-scale enterprise solution.

In conclusion, this project provides a solid foundation for advanced facial recognition systems in attendance management. It demonstrates the feasibility and effectiveness of hybrid models like ConViT for biometric tasks and paves the way for future improvements in accuracy, reliability, and ethical deployment.

References

[1] A. I. Awad, A. Babu, E. Barka, and K. Shuaib, ‘AI-powered biometrics for Internet of Things security: A review and future vision’, *J. Inf. Secur. Appl.*, vol. 82, p. 103748, 2024.

[2] G. Singh, M. kumari, V. Tripathi, and M. Diwakar, ‘Attendance Monitoring System Using Facial and Geo-Location Verification’, in *Intelligent Human Computer Interaction*, B. J. Choi, D. Singh, U. S. Tiwary, and W.-Y. Chung, Eds., Cham: Springer Nature Switzerland, 2024, pp. 406–416. doi: 10.1007/978-3-031-53827-8_36.

[3] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, ‘3D face recognition under expressions, occlusions, and pose variations’, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2270–2283, 2013.

[4] A. Arjun Raj, M. Shoheb, K. Arvind, and K. S. Chethan, ‘Face Recognition Based Smart Attendance System’, in *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, Jun. 2020, pp. 354–357. doi: 10.1109/ICIEM48762.2020.9160184.

[5] N. S. Ali, A. H. Alhilali, H. D. Rjeib, H. Alsharqi, and B. A. Sadawi, ‘Automated attendance management systems: systematic literature review’, *Int. J. Technol. Enhanc. Learn.*, vol. 14, no. 1, p. 37, 2022, doi: 10.1504/IJTEL.2022.120559.

[6] H. Yang and X. Han, ‘Retracted: Face recognition attendance system based on real-time video processing’, *IEEE Access*, vol. 8, pp. 159143–159150, 2020.

[7] S. Dev and T. Patnaik, ‘Student Attendance System using Face Recognition’, in *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Sep. 2020, pp. 90–96. doi: 10.1109/ICOSEC49089.2020.9215441.

[8] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’, presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778. Accessed: Dec. 16, 2024. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

[9] Z. Liu *et al.*, ‘Swin Transformer: Hierarchical Vision Transformer using Shifted Windows’, Aug. 17, 2021, *arXiv*: arXiv:2103.14030. doi: 10.48550/arXiv.2103.14030.

[10] ‘Chen, Y.-C., Li, L., Yu, L., El Kholy, A., Ahmed, F., Gan, Z., Cheng, Y., and Liu, J. Uniter: Universal image-text repre sentation learning. In European Conference on Computer Vision, pp. 104–120. Springer, 2020. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.1909.11740

[11] Y. Jing, X. Lu, and S. Gao, ‘3D face recognition: A comprehensive survey in 2022’, *Comput. Vis. Media*, vol. 9, no. 4, pp. 657–685, Dec. 2023, doi: 10.1007/s41095-022-0317-1.

[12] Z. Guo and Y. Fan, ‘Sparse Representation for 3D Face Recognition’, in *2013 Fourth World Congress on Software Engineering*, Dec. 2013, pp. 336–339. doi: 10.1109/WCSE.2013.63.

[13] F. Boutros, V. Struc, J. Fierrez, and N. Damer, ‘Synthetic data for face recognition: Current state and future prospects’, *Image Vis. Comput.*, vol. 135, p. 104688, Jul. 2023, doi: 10.1016/j.imavis.2023.104688.

[14] S. d’Ascoli, H. Touvron, M. Leavitt, A. Morcos, G. Biroli, and L. Sagun, ‘ConViT: Improving Vision Transformers with Soft Convolutional Inductive Biases’, *J. Stat. Mech. Theory Exp.*, vol. 2022, no. 11, p. 114005, Nov. 2022, doi: 10.1088/1742-5468/ac9830.

[15] ‘Cordonnier, J.-B., Loukas, A., and Jaggi, M. On the rela tionship between self-attention and convolutional layers. arXiv preprint arXiv:1911.03584, 2019. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.1911.03584

[16] ‘d’Ascoli, S., Sagun, L., Biroli, G., and Bruna, J. Finding the needle in the haystack with convolutions: on the benefits of architectural bias. In Advances in Neural Information Processing Systems, pp. 9334–9345, 2019. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.1906.06766

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ‘ImageNet classification with deep convolutional neural networks’, *Commun ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[18] ‘Neyshabur, B. Towards learning convolutions from scratch. Advances in Neural Information Processing Systems, 33, 2020. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://proceedings.neurips.cc/paper/2022

[19] ‘Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. End-to-end object detection with transformers. arXiv preprint arXiv:2005.12872, 2020. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2005.12872

[20] ‘Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for lan guage understanding. arXiv preprint arXiv:1810.04805, 2018. - Google Search’. Accessed: May 02, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.1810.04805