Monitoring Student Attendance Through Vision Transformer-based Iris Recognition

Slimane Ennajar, Walid Bouarifi
Mathematical Team and Information Processing-National School of Applied Sciences, SAFI Cadi AYYAD University
Marrakech, Morocco

***Abstract*—In the context of the ongoing digital transformation, the effective monitoring of student attendance holds paramount significance for educational establishments. This study presents an innovative approach using Vision Transformer technology for iris recognition to automate student attendance tracking. We fine-tuned Vision Transformer models, specifically ViT-B16, ViT-B32, ViT-L16, and ViT-L32, using the CASIA-Iris-Syn dataset and focused on overcoming challenges related to intra-class variation through data augmentation techniques, including rotation, shearing, and brightness adjustments. The results reveal that ViT-L16 is the most proficient, achieving an impressive accuracy of 95.69%. Comparative analysis with prior methodologies, specifically those employing Vision Transformer with Convolutional Neural Network, underscores the superiority of our proposed ViT-L16 model. This superiority is evident across various metrics, including accuracy, precision, recall, and F1 score. The experimental setup involves the use of Jupyter Notebook, Python technologies, TensorFlow, and Keras, emphasizing evaluations based on loss, accuracy, and Confusion Matrix. ViT-L16 consistently outshines other models, showcasing its resilience in iris recognition for student attendance. This research marks a significant step towards modernizing attendance systems, offering an accurate and automated solution suitable for the evolving needs of educational settings. Future work could explore integrating additional biometric modalities and refining Vision Transformer architecture for enhanced performance and broader application in educational environments.**

Keywords*—Iris Recognition; Vision transformer; student attendance; vision transformer models; educational technology

I. INTRODUCTION

A Student Attendance System is a digital solution designed to track and manage the attendance of students in educational institutions. It offers an efficient and accurate way to record and monitor student attendance, replacing traditional manual methods. Additionally, applying Student Attendance System significantly enhance the organization, precision, and transparency within educational institutions. It can also serve as a tool for analyzing attendance patterns, identifying areas for improvement, and fostering communication between educators and parents.

Accurate and timely attendance records are fundamental for identifying student absenteeism, serving as a crucial component in promoting student retention and academic achievement [1]. By meticulously tracking attendance, educators can readily implement necessary interventions for at-risk students, facilitating their academic success.

Many educational institutions still employ manual student attendance tracking, such as roll call or sign-in sheets. These methods are inefficient, as they are time-consuming and susceptible to human error. Additionally, they lack real-time capabilities, hindering the timely identification and resolution of attendance issues. Furthermore, manual methods offer limited security and privacy compared to biometric solutions. Vision Transformer (ViT) technology seeks to address these shortcomings by automating attendance, potentially reducing educator workload [2]. ViTs leverage iris recognition, a biometric approach offering greater accuracy, security, and real-time monitoring than manual methods. This transformation can streamline administrative processes and align with the ongoing integration of technology within the educational sector.

Traditional attendance tracking solutions often rely on manual methods (roll calls, sign-in sheets), card-based systems (RFID [3][4]), or biometric technologies [5][6]. While these methods offer varying degrees of utility, they frequently encounter limitations in efficiency, security, and accuracy. The ViT-based approach demonstrates a technological evolution in attendance management, leveraging sophisticated image processing and machine learning for iris recognition. This translates to superior security and precision, minimizing the potential for fraudulent activity. Furthermore, process automation makes the ViT approach a uniquely efficient and dependable solution within educational and organizational settings.Haut du formulaire.

Iris recognition technology is being integrated into student attendance systems to automate and enhance monitoring and documentation processes. This biometric approach involves capturing and analyzing the unique patterns within the iris (the colored ring surrounding the pupil) [7]. Extracted features, including crypts, furrows, and freckles, serve as the basis for generating unique biometric templates for each individual.

Integrating AI, especially Vision Transformer technology, demonstrates a commitment to technological advancement within the educational institution. It positions the institution at the forefront of leveraging innovative solutions for routine tasks.

Addressing this challenge, we propose an innovative method for handling student attendance in educational institutions through the application of Computer Vision. Our strategy involves the detection and recognition of students' irises in classrooms utilizing a VIT. The primary focus of this paper is the creation of a transformer model designed specifically for the identification and recognition of iris images.

The specific tasks were carried out according to the following steps:

- Fine-tuning various Vision Transformer models to evaluate their performance in iris image classification.
- Utilizing a dataset from CASIA-Iris-Syn to assess the effectiveness of the proposed method.
- Evaluating the performance and accuracy of different Vision Transformer models, including ViT-B16, ViT-B32, ViT-L16, and ViT-L32, for the identification and classification of iris images.
- The results achieved demonstrated high performance, with an accuracy rate of 95.69% for iris image classification.

The subsequent sections of this paper are organized as follows: Section 2 offers a review of the existing studies correlated to iris image recognition in attendance systems and investigates ViT applications in image processing. Section 3 outlines the materials and methods utilized in the experimental approach. Moving forward to Section 4, the paper examines the results obtained and conducts a performance evaluation. Section 5 provides a comparative analysis of the proposed models. Lastly, Section 6 encapsulates the conclusions derived from this study.

II. Related Work

Various studies have explored different methods for monitoring attendance, Okokpujie et al. [8] implemented a Student Attendance System that utilizes Iris Biometric Recognition. The experimental findings indicate that the system operates through a web-based platform. Student identification is achieved by comparing the acquired iris image with the database entries. The system assigns an integer value of (1) for a successful match and (0) for no match, with these outcomes are then stored in a MySQL-created database.

Shaban et al. [9] proposed a multimodal system utilizing ear and iris biometrics at the feature fusion stage to recognize students in electronic examinations (E-exams) amid the COVID-19 pandemic. The approach attained a precision rate of 92.6%.

Hassan et al. [10] devised a technique for iris segmentation comprising two stages. Initially, it identifies the outer iris boundary, followed by the detection of the inner iris boundary in the second stage. The method underwent testing on CASIA iris image datasets V1 and V4, yielding accuracy results of 100% and 99.16% respectively.

Trabelsi & Shuaib, [11] proposed a biometric attendance system using fingerprint and iris recognition to improve accuracy and security in educational settings. This system addresses limitations of manual methods by offering reliable and efficient student identification, enhancing overall attendance recording processes. Similarly, Adamu, [12] introduced an advanced system integrating fingerprint and iris biometrics for attendance management in higher education. This system replaces traditional methods with a secure, efficient, and accurate approach. Utilizing fingerprint and iris scanners at lecture entrances, it verifies student identities against stored biometric data, enabling real-time tracking and reporting.

Kadry & Smaili, [13] implemented a wireless attendance management system incorporating Daugman's algorithm (Daugman, 2003) for iris recognition. This biometrics-based system, integrated with wireless technology, addresses issues related to inaccurate attendance records and surpasses the challenges associated with establishing a dedicated network for this purpose.

Khatun et al. [14] introduced the Iris Recognition Attendance Management System, which employs a camera to capture real-time images of the human iris, and storing this data in a database. The system utilizes the Gray-coding algorithm in MATLAB data analysis software to compute the iris radius. Employing MATLAB, it compares the radius of each individual with the previously stored value and automatically sends the attendance report to a predefined email address, eliminating the need for human intervention.

Sujatha et al. [15] proposed a solution for a biometric-based attendance system utilizing iris recognition, interfaced with NI MYRIO. The proposal emphasizes the robustness of iris recognition, highlighting its reliability, accuracy, and efficiency attributed to the unique and immutable characteristics of the iris. Furthermore, NI LABVIEW, a graphical user interface-based software, facilitates real-time monitoring and attendance management. The integration of features such as SMS notifications for absentees and the generation of Excel sheets enhances the overall functionality of the system.

Joshy & Jalaja. [16] introduced a biometric authentication system based on the Internet of Things (IoT), and emphasizes the use of iris recognition for its unparalleled accuracy and security. The proposed system incorporates a hybrid encryption algorithm (Blowfish and RSA) for securing data transmitted over the Internet and implements a two-step authentication process. Developed as an embedded system for secure employee authentication.

Lad & More, [17] developed a student attendance system leveraging iris detection technology, which is acknowledged as the most reliable and accurate form of biometric identification. This initiative aims to address the shortcomings of commercial systems by offering an open-source alternative. The system employs the Hough transform for automatic iris segmentation, normalizes the iris region, and uses 1D Log-Gabor filters for feature extraction. These steps are designed to enhance the efficiency and accuracy of attendance tracking in educational contexts.

In [18], the authors presented a multimodal biometric system utilizing Convolutional Neural Networks (CNN) and transfer learning for iris recognition. It aims to overcome limitations in unimodal biometric methods by focusing on deep learning models for analyzing both left and right irises. Employing back-propagation with Adam's optimization, the system demonstrates high accuracy on public datasets, IITD and CASIA-Iris-V3 Interval, achieving up to 99% accuracy. This study underscores the effectiveness of combining CNN characteristics and transfer learning in real-time iris recognition, enhancing security and identification processes in various conditions.

In recent studies, the Vision Transformer has been employed for image classification and identification, representing a neural network architecture tailored specifically for image processing in computer vision applications [19].The ViT is a neural network crafted for image processing in computer vision. It employs a self-attention mechanism commonly found in natural language processing, setting it apart from traditional image processing architectures like CNNs and RNNs. Introduced to address limitations in handling image data, ViT offers robust image feature representation and requires fewer computational resources for training compared to CNNs [20].

Elpina & Kusuma,[21] introduced a Swin Transformer model for feature extraction in food image classification, incorporating an SVM classifier. The methodology underwent training and evaluation utilizing the Food-101 Dataset, resulting in an impressive accuracy (ACC) of 97.61%.

Mehta et al.[22] introduced a method for ear recognition that exercises the ViT network architecture, attaining a recognition accuracy surpassing 99.36%.

Latif et al.[23] introduced a hybrid model combining ViT and Convolutional Neural Network (CNN) for the identification and verification of iris images. The hybrid model demonstrated an accuracy of up to 93.66% in recognizing iris patterns.

Ha et al.[24] employed the Vision Transformer architecture to extract data features and categorize X-ray images as either pneumonia-positive or negative. Experimental findings reveal that the Vision Transformer algorithm consistently yields favorable classification outcomes, achieving an accuracy of approximately 94%.

III. Materials and Methods

A. Proposed Methodologies

This paper investigates the enhancement of student attendance management through a novel ViT-based iris recognition approach. This approach leverages automation to improve the accuracy and efficiency of attendance tracking.

Fig. 1 illustrates our proposed Vision Transformers approach for iris identification and recognition.

B. Dataset

Our study utilized a dataset sourced from CASIA-Iris-Syn, featuring 8533 artificially generated iris images distributed across 50 classes, as illustrated in Fig. 2. The textures of these iris images were automatically synthesized from a subset of CASIA-IrisV1.Subsequently, the iris ring regions were incorporated into authentic iris images, augmenting the realism of the artificial iris images. Intra-class variations, including deformation, blurring, and rotation, were introduced into the synthesized iris dataset. The training dataset comprises 5814 iris images in JPG format. The validation and test sets were assessed using new databases, consisting of 1027 images and 1692 images, respectively. A graphical illustration of the configuration of this dataset as depicted in Fig. 3.

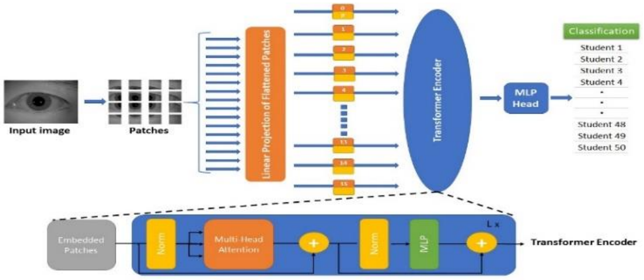


Fig. 1. Visualization of our proposed Vision Transformer (ViT) model for identifying and recognizing iris images. Initially, the input image undergoes segmentation into fixed-size patches, which are subsequently flattened. Following this, position embeddings are introduced, and the resulting sequence of vectors is then passed through a standard Transformer encoder. The inspiration for this illustration is derived from [2].

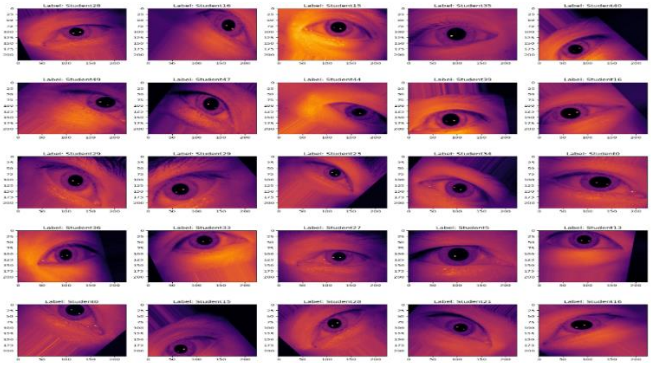


Fig. 2. Example of iris images in the Iris Dataset (50 classes).

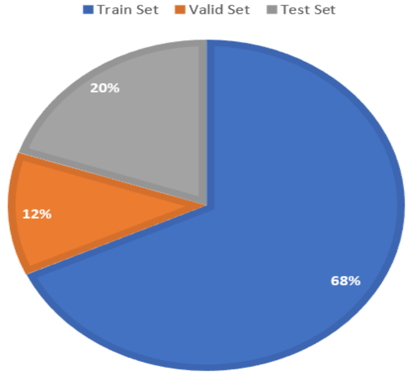


Fig. 3. Distribution of the Iris dataset (50 classes).

C. Data Augmentation Technique

Data augmentation serves as a pivotal technique in machine learning, aimed at artificially expanding the scale of a training dataset through the application of diverse transformations to the existing data. This strategy proves instrumental in enhancing the generalization and resilience of machine learning models. Its significance becomes more pronounced when dealing with a restricted training dataset size. By creating novel variations to pre-existing data, data augmentation serves a dual purpose of mitigating overfitting risks and enabling the model to capture more resilient features. Widely utilized across domains like image classification, object detection, and segmentation, data augmentation emerges as a fundamental tool for bolstering the performance and adaptability of machine learning models.

In this research, our emphasis was directed towards various data augmentation methods. The precise parameters selected for each operation are detailed in Table I.

The Table I outlines the specific parameters employed for various data augmentation operations. Rotation is set at 30 degrees, shearing at 0.2 radians, and zooming within a range of 0.2. Horizontal and vertical flips are enabled, and brightness is varied within the range of 0.4 to 1.5. These meticulously chosen parameters contribute to the augmentation of the training dataset, and ultimately bolstering the model's robustness and performance.

TABLE I. DATA AUGMENTATION PARAMETERS

Operations	Values
Rotation	30 degrees
Shearing	0.2 radians
Zooming (range)	0.2
Horizontal flip	True
Vertical flip	True
Brightness	[0.4, 1.5]

D. Vision Transformer (ViT)

The Vision Transformer presents a revolutionary deep learning architecture designed to tackle computer vision tasks, challenging the conventional prominence of convolutional neural networks. Originating from the paper titled "An Image is Worth 16x16 Words: Transformers for Image Recognition" by Alexey Dosovitskiy et al.[2], ViT extends the transformer architecture, initially crafted for natural language processing, into the realm of images. This adaptation involves the incorporation of self-attention mechanisms, enabling the model to adeptly capture long-range dependencies within the input data. ViT's introduction marks a paradigm shift, opening up new possibilities for image recognition and paving the way for diverse applications beyond the confines of traditional CNN-based approaches.

In contrast to processing the entire image in a holistic manner, the Vision Transformer adopts a strategy of partitioning the input image into fixed-size, non-overlapping patches. Subsequently, each of these patches undergoes a linear embedding, transforming it into a flat vector and composing the input sequence for the transformer. To preserve spatial information, positional embeddings are introduced to the patch embeddings. This addition enables the model to discern the spatial relationships existing between distinct patches, ensuring a nuanced understanding of the overall image structure. The incorporation of such mechanisms enhances ViT's capacity to effectively process and interpret intricate spatial features within images.

ViT models are typically pre-trained on large datasets, such as ImageNet, using a contrastive learning framework. This pre-training helps the model learn rich visual representations. The pre-trained ViT model is fine-tuned for specific tasks by adding a linear classification head on top. The model can be fine-tuned for various computer vision tasks such as image classification, object detection, and segmentation.

ViT has shown good scalability, performing well on both small and large datasets. This scalability is advantageous for adapting the model to different tasks.

In this research, the Vision Transformer architecture is crafted with adjustable dimensions to suit specific requirements. Additionally, each parameter in the vision transformer holds a crucial role, and their descriptions are outlined as follows:

- image_size=224: This parameter defines the preferred dimensions (width and height) of the input images for the model. In this instance, the images are expected to have dimensions of 224x224 pixels.
- patch_size=16: The images undergo segmentation into smaller patches, and this parameter determines the size (width and height) of each patch. In this case, each patch measures 16x16 pixels.
- num_classes=50: This parameter signifies the number of classes involved in the classification task. In this particular example, the model is configured to categorize inputs into 50 classes.
- dropout=0.2: This parameter governs the dropout rate, a regularization technique employed to mitigate overfitting. It involves randomly setting a fraction of input units to 0 during training.

E. Evaluation Metrics

The assessment of prediction algorithms in this study relies on various performance metrics. The paper examines the subsequent evaluation metrics to gauge the efficacy of the proposed model:

1) Accuracy score: The accuracy score is a performance metric used to measure the overall correctness of a predictive model. It is calculated by dividing the number of correct predictions by the total number of predictions and is often expressed as a percentage[25]. The formula for accuracy is shown in equation (1).

$$\mathrm{Accuracy} = (\mathrm{TP}+\mathrm{TN})/(\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}) \quad (1)$$

2) Precision: Precision is a performance metric used in classification tasks to assess the accuracy of the positive predictions made by a model. It is defined as the ratio of true positive predictions to the total number of positive predictions (both true positives and false positives) [25].The formula for precision is shown in equation (2).

$$\mathrm{Precision} = (\mathrm{TP})/(\mathrm{TP}+\mathrm{FP}) \quad (2)$$

3) Recall: Recall, also known as sensitivity or true positive rate, is a performance metric used in classification tasks to evaluate a model's ability to correctly identify all relevant instances of a particular class. It is the ratio of true positive predictions to the total number of actual positive instances (including both true positives and false negatives) [25].The formula for recall is shown in equation (3):

$$\text{Recall} = (\text{TP})/(\text{TP} + \text{FN}) \qquad (3)$$

4) F1-score: The F1 score is a metric commonly used in classification tasks that combines both precision and recall into a single measure. It is particularly useful when there is an uneven class distribution (imbalanced datasets) and provides a balance between the precision and recall metrics [25].The formula for the F1 score is shown in equation (4):

$$\text{F1} = 2 * (\text{precision} + \text{recall})/(\text{precision} + \text{recall}) \quad (4)$$

5) Matthews correlation coefficient: The Matthews correlation coefficient (MCC) is a metric used to evaluate the performance of binary classification models, particularly when dealing with imbalanced datasets. It takes into account true positives, true negatives, false positives, and false negatives. The formula for Matthews correlation coefficient is shown in equation (5):

$$\text{MCC} = \frac{\text{TP}\times\text{TN}-\text{FP}\times\text{FN}}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \times 100 \qquad (5)$$

IV. Results and Discussions

The primary objective of this research is to develop a transformer model designed for the identification and recognition of iris images. The model underwent training using both regular and augmented images, with where data augmentation employed to enhance the training dataset. Training involved 5814 images, validation with 1027 images, and testing utilized 1692 images. The network layers were subsequently frozen and fine-tuned with dense layers containing 1024, 512, 256, and 50 neurons, respectively.

Table II summarizes the performance metrics of different ViT models, each trained for 50 epochs, in the context of iris image identification and recognition. Notably, the ViT-L16 model emerges as the top performer with an accuracy score of 95.69%, demonstrating its effectiveness in accurately classifying iris images. ViT-L32, ViT-B32, and ViT-B16 also exhibit commendable performance, achieving accuracy scores of 94.03%, 93.26%, and 92.02% respectively. In terms of precision, recall, F1 score, and Matthews Correlation Coefficient (MCC), ViT-L16 consistently outperforms the other models, emphasizing its robustness in various evaluation criteria. These results indicate that the ViT-L16 model, with its customizable dimensions and advanced architecture, proves to be particularly effective in iris recognition tasks, demonstrating its potential for applications such as student attendance using Vision Transformer technology.

A. *Experimental Setup*

The experimental setup utilized Jupyter Notebook along with Python technologies such as NumPy, Pandas, and OpenCV for image processing tasks. For implementing classifiers, Scikit-Learn, Anaconda, and Python 3.9 were employed. The Vision Transformer model underwent training and testing processes using TensorFlow and Keras, leveraging Google Colab PRO T4-GPU with reported memory at 51GB and storage space at 166.77GB for refined computational capabilities.

TABLE II. Accuracy Score, Precision Score, Recall Score, F1 Score and MCC of our Vision Transformers Models

Model	Number of epochs	Accuracy Score (%)	Precision Score (%)	Recall Score (%)	F1Score (%)	MCC (%)
ViT-B16	50	92.02	93.83	92.02	91.85	91.91
ViT-B32	50	93.26	93.87	93.26	9327	93.14
ViT-L16	50	95.69	96.08	95.69	95.64	95.61
ViT-L32	50	94.03	94.65	94.03	93.88	93.93

B. *Loss and Accuracy*

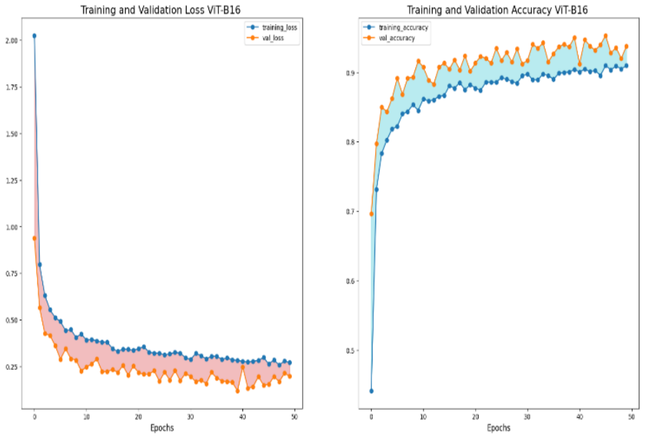
1) Loss: serves as an indicator of the model's performance on training data, gauging the discrepancy between predicted values and actual ground truth. The training objective involves minimizing the loss, with a lower value indicating closer alignment between model predictions and actual values.

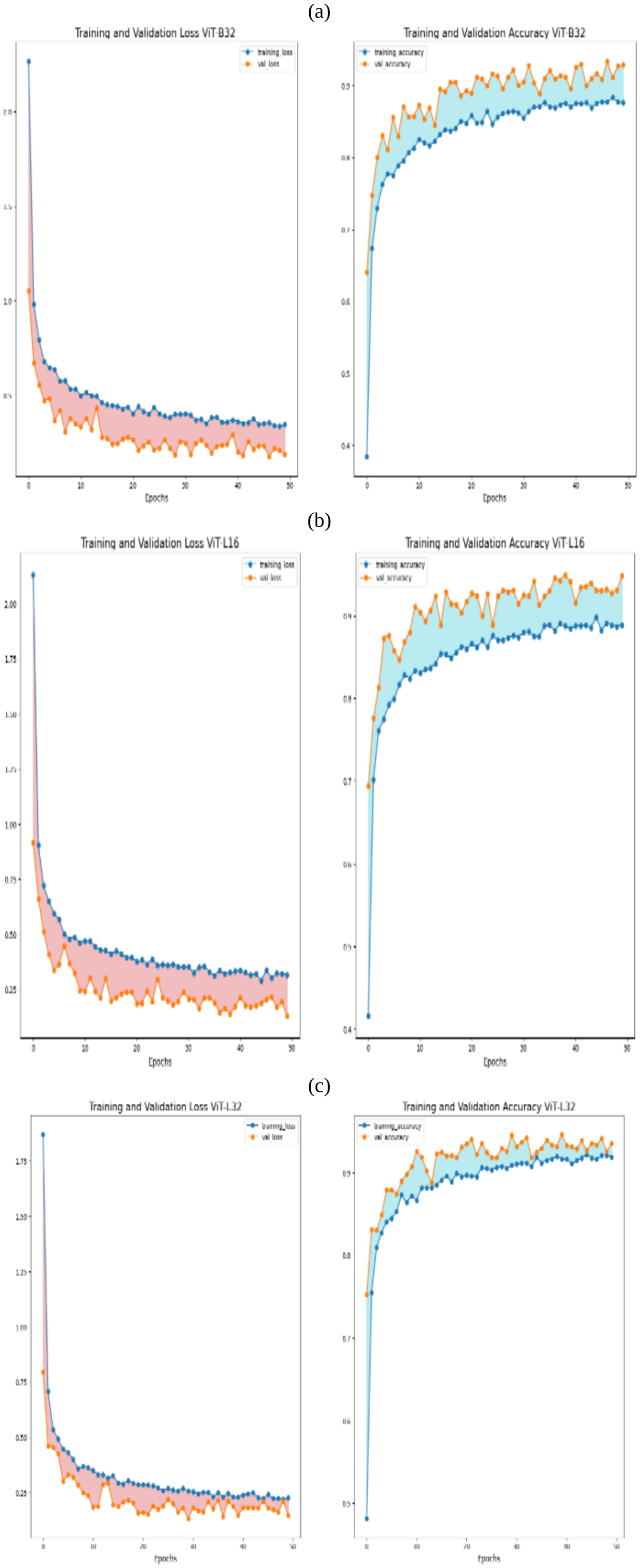
2) Accuracy: serves as a metric for the overall correctness of the model, determining the ratio of correctly predicted instances to the total instances. The goal in both training and testing phases is to maximize accuracy, as a higher value signifies a greater proportion of correct predictions.

In Fig. 4, the evaluation of loss and accuracy is depicted for the ViT-B16, ViT-B32, ViT-L16, and ViT-L32 models. The results clearly indicate that the ViT-L16 model exhibits superior performance, confirming its heightened effectiveness when compared to the other models.

C. *Confusion Matrix*

An additional evaluation metric, the Confusion Matrix, was utilized to assess the overall effectiveness of a classification model. The Confusion Matrix serves as a tabular summary, offering a detailed breakdown of the model's predictions in comparison to the actual class labels. The evaluation outcomes for the mentioned algorithms, using these criteria, are illustrated in Fig. 5.





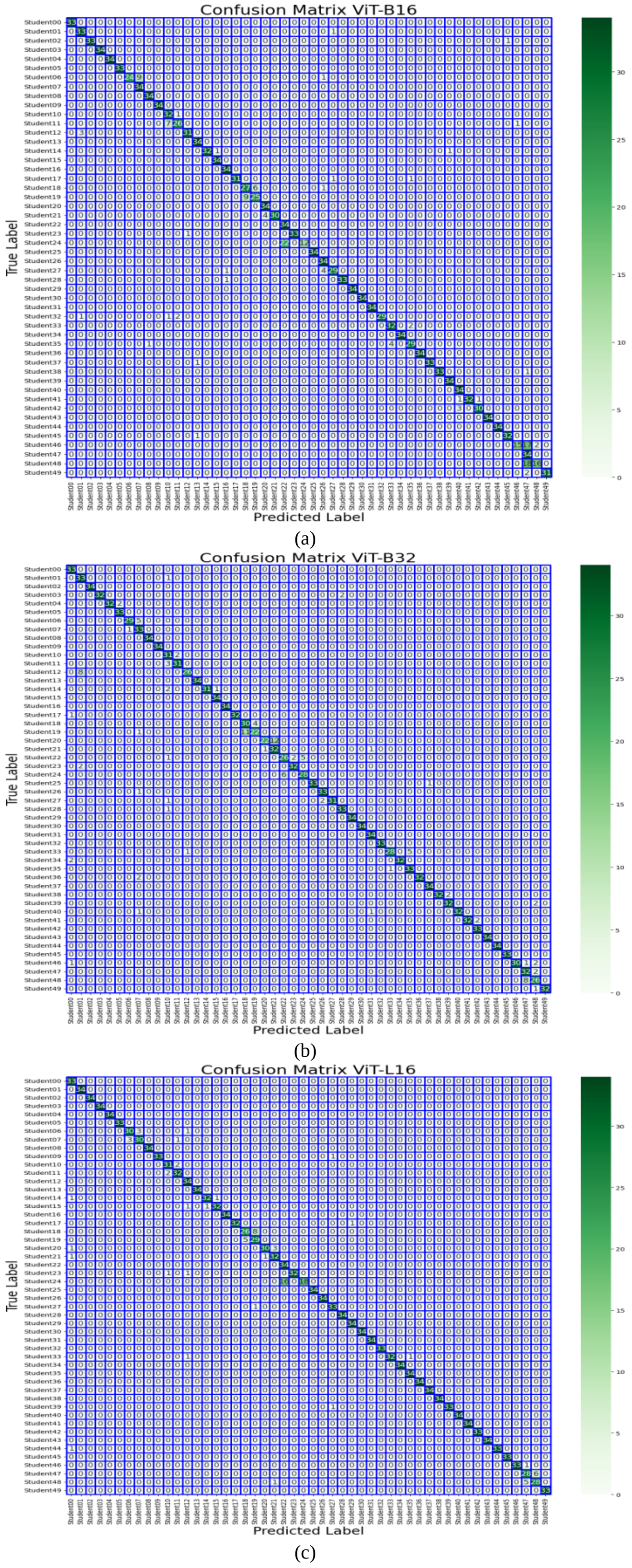
(a)

(b)

(c)

(d)

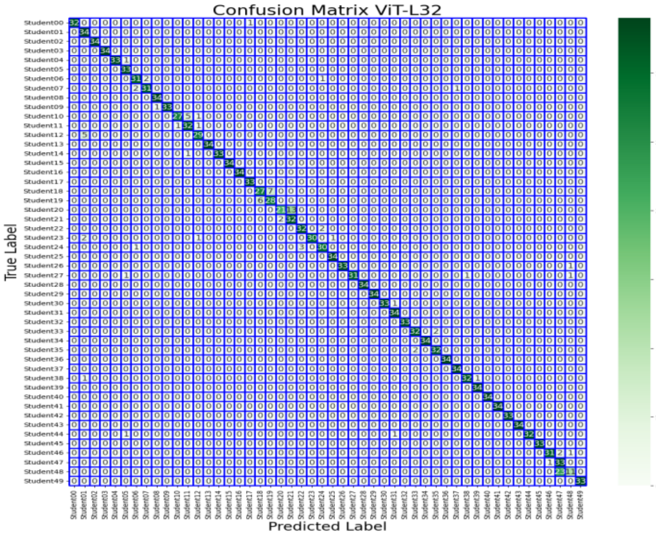
Fig. 4. Loss and Accuracy of ViT-B16 (a), ViT-B32 (b), ViT-L16 (c), and ViT-L32 (d).



(a)

(b)

(c)



(d)

Fig. 5. Confusion Matrix of ViT-B16 (a), ViT-B32 (b), ViT-L16 (c), and ViT-L32 (d).

Fig. 5(a) presented the recognition outcomes achieved using the ViT-B16 model. The average count of accurate recognitions for each category was 31.16. The expected correct recognition count ranged between 33 and 34. As a result, the average accuracy attained by the ViT-B16 model was 92.02%.

In Fig. 5(b), the ViT-B32 model's recognition results were presented. The mean correct recognition count for each category was 31.56, with the expected correct recognition count ranging 33 and 34. As a result, the average accuracy of the ViT-B32 model was 93.26%.

Moving to Fig. 5(c), it demonstrated the recognition outcomes of the ViT-L16 model. The mean correct recognition number for each category was 32.44, with the expected correct recognition number ranging between 33 and 34. The ViT-L16 model achieved an average accuracy of 95.69%.

In Fig. 5(d), the recognition results of the ViT-L32 model were depicted. The mean correct recognition number for each category was 31.82, and the expected correct recognition number ranged between 30 and 34. The average accuracy of the ViT-L32 model was 94.03%.

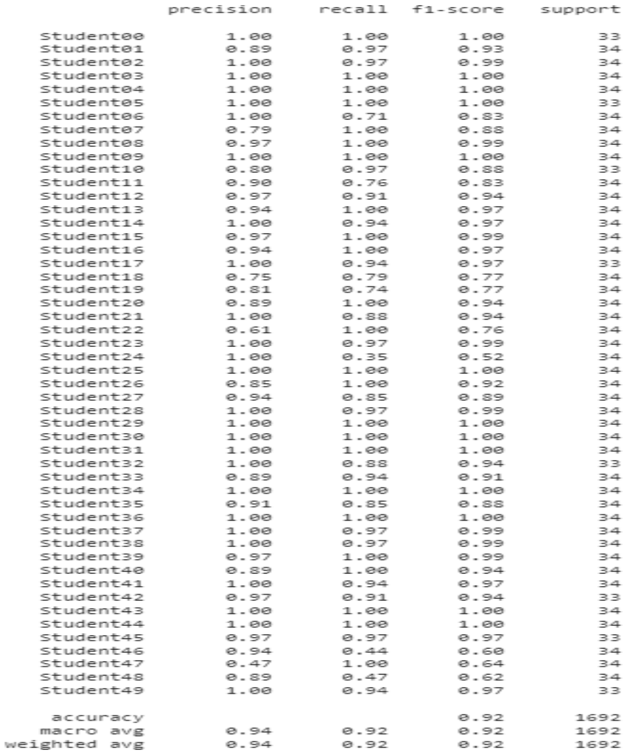
D. Classification Report

Fig. 6(a) displays the classification report for the ViT-B16 model, indicating precision values for iris classes ranging from 0.47 to 1. Additionally, the recall performance values for iris classes fall within the range of 0.35 to 1, with corresponding support values between 33 and 34. F1 scores for the iris classes vary from 0.52 to 1. The ViT-B16 model achieves an accuracy of 0.92 (92%) based on the F1 score, considering 1692 support values. The macro and weighted averages for precision and recall are 0.94, 0.92, 0.94, and 0.92, and the f1 scores are 0.92 and 0.92, each with support values of 1692.

Fig. 6(b) exhibits the classification report of the ViT-B32 model, revealing precision values within the range of 0.73 to 1 for iris classes. The recall performance spans from 0.65 to 1 across iris classes, accompanied by corresponding support values ranging from 33 to 34. F1 scores for the iris classes vary from 0.73 to 1. The ViT-B32 model attains an accuracy of 0.93 (93%) based on the F1 score, taking into account 1692 support values. The macro and weighted averages for precision and recall stand at 0.94, 0.93, 0.94, and 0.93, with f1 scores of 0.93 and 0.93, respectively, supported by 1692 instances.

Fig. 6(c) illustrates the ViT-L16 model's classification report, delineating precision values for iris classes ranging from 0.84 to 1. Likewise, the recall performance for iris classes spans from 0.53 to 1, accompanied by support values ranging between 33 and 34. F1 scores for the iris classes vary between 0.69 and 1. The accuracy of the ViT-B16 model stands at 0.96 (96%) based on the F1 score, considering 1692 support values. The macro and weighted averages for precision and recall are 0.96, 0.96, 0.96, and 0.96, respectively, with f1 scores of 0.96 and 0.96, supported by 1692 instances.

In Fig. 6(d), the classification report of the ViT-L32 model illustrates precision values for iris classes spanning from 0.57 to 1. Moreover, recall performance values for iris classes range from 0.32 to 1, with corresponding support values falling between 33 and 34. F1 scores for the iris classes are distributed within the range of 0.46 to 1. The ViT-B16 model attains an accuracy of 0.94 (94%) based on the F1 score, taking into account 1692 support values. The precision and recall values for both macro and weighted averages are 0.95, 0.94, 0.95, and 0.94, with F1 scores of 0.94 and 0.94, respectively, supported by 1692 instances.

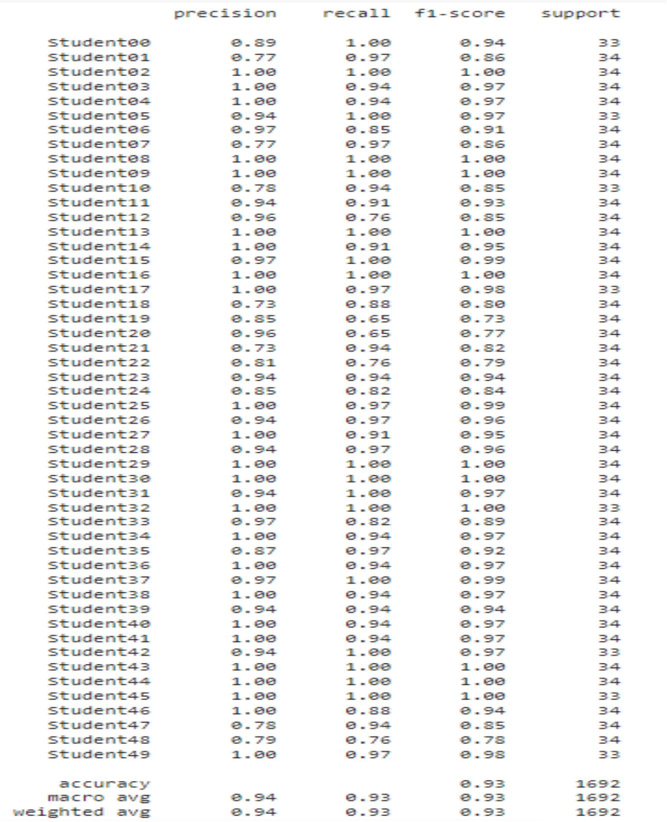


```
              precision    recall  f1-score   support

   Student00       1.00      1.00      1.00        33
   Student01       0.89      0.97      0.93        34
   Student02       1.00      0.97      0.99        34
   Student03       1.00      1.00      1.00        34
   Student04       1.00      1.00      1.00        34
   Student05       1.00      1.00      1.00        33
   Student06       1.00      0.71      0.83        34
   Student07       0.79      1.00      0.88        34
   Student08       0.97      1.00      0.99        34
   Student09       1.00      1.00      1.00        34
   Student10       0.80      0.97      0.88        33
   Student11       0.90      0.76      0.83        34
   Student12       0.97      0.91      0.94        34
   Student13       0.94      1.00      0.97        34
   Student14       1.00      0.94      0.97        34
   Student15       0.97      1.00      0.99        34
   Student16       0.94      1.00      0.97        34
   Student17       1.00      0.94      0.97        33
   Student18       0.75      0.79      0.77        34
   Student19       0.81      0.74      0.77        34
   Student20       0.89      1.00      0.94        34
   Student21       1.00      0.88      0.94        34
   Student22       0.61      1.00      0.76        34
   Student23       1.00      0.97      0.99        34
   Student24       1.00      0.35      0.52        34
   Student25       1.00      1.00      1.00        34
   Student26       0.85      1.00      0.92        34
   Student27       0.94      0.85      0.89        34
   Student28       1.00      0.97      0.99        34
   Student29       1.00      1.00      1.00        34
   Student30       1.00      1.00      1.00        34
   Student31       1.00      1.00      1.00        34
   Student32       1.00      0.88      0.94        33
   Student33       0.89      0.94      0.91        34
   Student34       1.00      1.00      1.00        34
   Student35       0.91      0.85      0.88        34
   Student36       1.00      1.00      1.00        34
   Student37       1.00      0.97      0.99        34
   Student38       1.00      0.97      0.99        34
   Student39       0.97      1.00      0.99        34
   Student40       0.89      1.00      0.94        34
   Student41       1.00      0.94      0.97        34
   Student42       0.97      0.91      0.94        33
   Student43       1.00      1.00      1.00        34
   Student44       1.00      1.00      1.00        34
   Student45       0.97      0.97      0.97        33
   Student46       0.94      0.44      0.60        34
   Student47       0.47      1.00      0.64        34
   Student48       0.89      0.47      0.62        34
   Student49       1.00      0.94      0.97        33

    accuracy                           0.92      1692
   macro avg       0.94      0.92      0.92      1692
weighted avg       0.94      0.92      0.92      1692
```

(a)

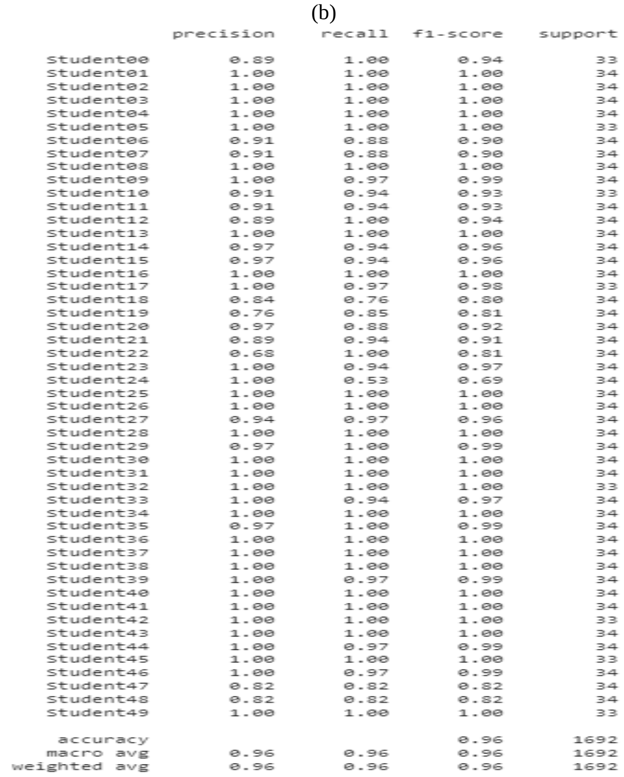


```
              precision    recall  f1-score   support

   Student00       0.89      1.00      0.94        33
   Student01       0.77      0.97      0.86        34
   Student02       1.00      1.00      1.00        34
   Student03       1.00      0.94      0.97        34
   Student04       1.00      0.94      0.97        34
   Student05       0.94      1.00      0.97        33
   Student06       0.97      0.85      0.91        34
   Student07       0.77      0.97      0.86        34
   Student08       1.00      1.00      1.00        34
   Student09       1.00      1.00      1.00        34
   Student10       0.78      0.94      0.85        33
   Student11       0.94      0.91      0.93        34
   Student12       0.96      0.76      0.85        34
   Student13       1.00      1.00      1.00        34
   Student14       1.00      0.91      0.95        34
   Student15       0.97      1.00      0.99        34
   Student16       1.00      1.00      1.00        34
   Student17       1.00      0.97      0.98        33
   Student18       0.73      0.88      0.80        34
   Student19       0.85      0.65      0.73        34
   Student20       0.96      0.65      0.77        34
   Student21       0.73      0.94      0.82        34
   Student22       0.81      0.76      0.79        34
   Student23       0.94      0.94      0.94        34
   Student24       0.85      0.82      0.84        34
   Student25       1.00      0.97      0.99        34
   Student26       0.94      0.97      0.96        34
   Student27       1.00      0.91      0.95        34
   Student28       0.94      0.97      0.96        34
   Student29       1.00      1.00      1.00        34
   Student30       1.00      1.00      1.00        34
   Student31       0.94      1.00      0.97        34
   Student32       1.00      1.00      1.00        33
   Student33       0.97      0.82      0.89        34
   Student34       1.00      0.94      0.97        34
   Student35       0.87      0.97      0.92        34
   Student36       1.00      0.94      0.97        34
   Student37       0.97      1.00      0.99        34
   Student38       1.00      0.94      0.97        34
   Student39       0.94      0.94      0.94        34
   Student40       1.00      0.94      0.97        34
   Student41       1.00      0.94      0.97        34
   Student42       0.94      1.00      0.97        33
   Student43       1.00      1.00      1.00        34
   Student44       1.00      1.00      1.00        34
   Student45       1.00      1.00      1.00        33
   Student46       1.00      0.88      0.94        34
   Student47       0.78      0.94      0.85        34
   Student48       0.79      0.76      0.78        34
   Student49       1.00      0.97      0.98        33

    accuracy                           0.93      1692
   macro avg       0.94      0.93      0.93      1692
weighted avg       0.94      0.93      0.93      1692
```

(b)

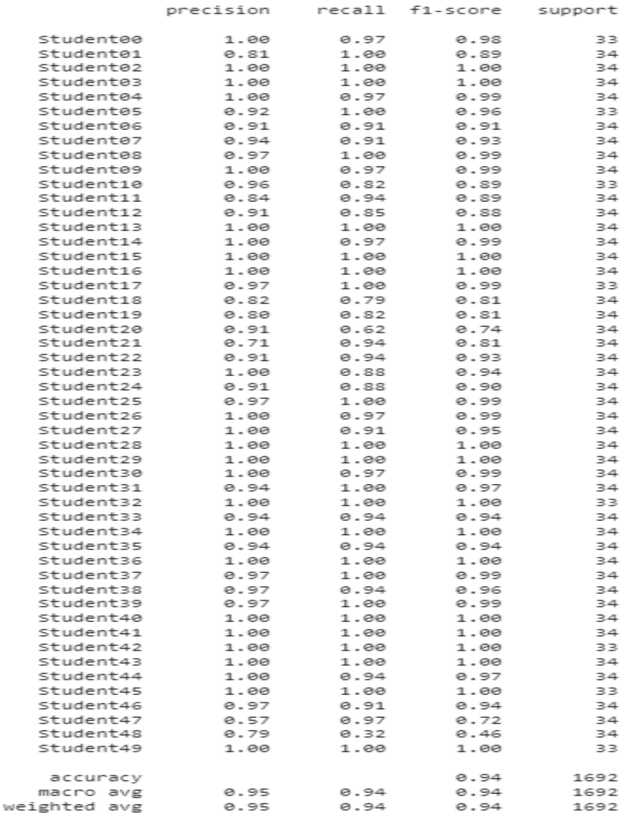


```
              precision    recall  f1-score   support

   Student00       0.89      1.00      0.94        33
   Student01       1.00      1.00      1.00        34
   Student02       1.00      1.00      1.00        34
   Student03       1.00      1.00      1.00        34
   Student04       1.00      1.00      1.00        34
   Student05       1.00      1.00      1.00        33
   Student06       0.91      0.88      0.90        34
   Student07       0.91      0.88      0.90        34
   Student08       1.00      1.00      1.00        34
   Student09       1.00      0.97      0.99        34
   Student10       0.91      0.94      0.93        33
   Student11       0.91      0.94      0.93        34
   Student12       0.89      1.00      0.94        34
   Student13       1.00      1.00      1.00        34
   Student14       0.97      0.94      0.96        34
   Student15       0.97      0.94      0.96        34
   Student16       1.00      1.00      1.00        34
   Student17       1.00      0.97      0.98        33
   Student18       0.84      0.76      0.80        34
   Student19       0.76      0.85      0.81        34
   Student20       0.97      0.88      0.92        34
   Student21       0.89      0.94      0.91        34
   Student22       0.68      1.00      0.81        34
   Student23       1.00      0.94      0.97        34
   Student24       1.00      0.53      0.69        34
   Student25       1.00      1.00      1.00        34
   Student26       1.00      1.00      1.00        34
   Student27       0.94      0.97      0.96        34
   Student28       1.00      1.00      1.00        34
   Student29       0.97      1.00      0.99        34
   Student30       1.00      1.00      1.00        34
   Student31       1.00      1.00      1.00        34
   Student32       1.00      1.00      1.00        33
   Student33       1.00      0.94      0.97        34
   Student34       1.00      1.00      1.00        34
   Student35       0.97      1.00      0.99        34
   Student36       1.00      1.00      1.00        34
   Student37       1.00      1.00      1.00        34
   Student38       1.00      1.00      1.00        34
   Student39       1.00      0.97      0.99        34
   Student40       1.00      1.00      1.00        34
   Student41       1.00      1.00      1.00        34
   Student42       1.00      1.00      1.00        33
   Student43       1.00      1.00      1.00        34
   Student44       1.00      0.97      0.99        34
   Student45       1.00      1.00      1.00        33
   Student46       1.00      0.97      0.99        34
   Student47       0.82      0.82      0.82        34
   Student48       0.82      0.82      0.82        34
   Student49       1.00      1.00      1.00        33

    accuracy                           0.96      1692
   macro avg       0.96      0.96      0.96      1692
weighted avg       0.96      0.96      0.96      1692
```

(c)



```
              precision    recall  f1-score   support

   Student00       1.00      0.97      0.98        33
   Student01       0.81      1.00      0.89        34
   Student02       1.00      1.00      1.00        34
   Student03       1.00      1.00      1.00        34
   Student04       1.00      0.97      0.99        34
   Student05       0.92      1.00      0.96        33
   Student06       0.91      0.91      0.91        34
   Student07       0.94      0.91      0.93        34
   Student08       0.97      1.00      0.99        34
   Student09       1.00      0.97      0.99        34
   Student10       0.96      0.82      0.89        33
   Student11       0.84      0.94      0.89        34
   Student12       0.91      0.85      0.88        34
   Student13       1.00      1.00      1.00        34
   Student14       1.00      0.97      0.99        34
   Student15       1.00      1.00      1.00        34
   Student16       1.00      1.00      1.00        34
   Student17       0.97      1.00      0.99        33
   Student18       0.82      0.79      0.81        34
   Student19       0.80      0.82      0.81        34
   Student20       0.91      0.62      0.74        34
   Student21       0.71      0.94      0.81        34
   Student22       0.91      0.94      0.93        34
   Student23       1.00      0.88      0.94        34
   Student24       0.91      0.88      0.90        34
   Student25       0.97      1.00      0.99        34
   Student26       1.00      0.97      0.99        34
   Student27       1.00      0.91      0.95        34
   Student28       1.00      1.00      1.00        34
   Student29       1.00      1.00      1.00        34
   Student30       1.00      0.97      0.99        34
   Student31       0.94      1.00      0.97        34
   Student32       1.00      1.00      1.00        33
   Student33       0.94      0.94      0.94        34
   Student34       1.00      1.00      1.00        34
   Student35       0.94      0.94      0.94        34
   Student36       1.00      1.00      1.00        34
   Student37       0.97      1.00      0.99        34
   Student38       0.97      0.94      0.96        34
   Student39       0.97      1.00      0.99        34
   Student40       1.00      1.00      1.00        34
   Student41       1.00      1.00      1.00        34
   Student42       1.00      1.00      1.00        33
   Student43       1.00      1.00      1.00        34
   Student44       1.00      0.94      0.97        34
   Student45       1.00      1.00      1.00        33
   Student46       0.97      0.91      0.94        34
   Student47       0.57      0.97      0.72        34
   Student48       0.79      0.32      0.46        34
   Student49       1.00      1.00      1.00        33

    accuracy                           0.94      1692
   macro avg       0.95      0.94      0.94      1692
weighted avg       0.95      0.94      0.94      1692
```

(d)

Fig. 6. Classification report of ViT-B16 (a), ViT-B32 (b), ViT-L16 (c), and ViT-L32 (d).

The research team faced challenges during the fine-tuning process due to intra-class variations within the synthesized iris dataset. To address these limitations, data augmentation techniques such as deformation, blurring, and rotation were employed. These augmentations significantly enhanced the robustness of the Vision Transformer models, particularly the ViT-L16, by introducing artificial variations within the training data. This resulted in improved model performance on real-world iris patterns with diverse characteristics, leading to higher accuracy and reliability in iris recognition tasks. These findings demonstrate the effectiveness of data augmentation in mitigating the effects of intra-class variations and highlight the adaptability of Vision Transformer architectures for tasks like student attendance monitoring using iris recognition.

This research explored the implementation of various Vision Transformer models (ViT-B16, ViT-B32, ViT-L16, and ViT-L32) for iris-based student attendance tracking. The ViT-L16 model demonstrated the highest performance in terms of accuracy, precision, and recall. Additionally, the study confirmed the adaptability of Vision Transformer architectures for iris recognition, underscoring the importance of data augmentation in improving model robustness.

V. Comparative Analysis

Table III presents a comparative examination of outcomes derived from our approach, employing Vision Transformer models ViT-L16 and ViT-L32, juxtaposed with findings from a preceding investigation utilizing ViT with CNN. The assessment hinges on pivotal metrics, including accuracy, precision, recall, and F1 score, observed across a span of 50 epochs.

In the study delineated by [23], the ViT+CNN model attained an accuracy of 93.66% after 50 epochs, although explicit figures for precision, recall, and F1 score remain undisclosed. Our methodology, leveraging ViT-L16, outperformed these results, manifesting an elevated accuracy of 95.69%. Furthermore, precision, recall, and F1 score for ViT-L16 registered at 96.08%, 95.69%, and 95.64%, sequentially. This signifies an amelioration in our model's capacity to accurately discern and categorize instances.

The ViT-L16 model's exceptional performance likely stems from its architecture, which excels at processing global image features – a vital aspect of accurate iris recognition. Unlike hybrid ViT+CNN models, which may introduce redundancies or inefficiencies through convolutional layers, the ViT-L16 relies exclusively on self-attention mechanisms. This enables a more direct and focused learning process that emphasizes the most pertinent features without the limitations inherent in convolutional operations.

Comparative results across the ViT-L16, ViT-L32, and ViT+CNN models demonstrate a clear pattern: the pure transformer-based models (ViT-L16, ViT-L32) consistently outperform the hybrid ViT+CNN model in accuracy, precision, recall, and F1 score. This finding suggests that self-attention mechanisms within transformers may be intrinsically better suited for iris recognition in attendance systems compared to a hybrid approach. Furthermore, these results highlight the potential of pure transformer models for driving improvements in biometric recognition systems.

TABLE III. Comparison of Results with Previous Works

Authors	[23]	**Our method**	
Model	ViT+CNN	**ViT-L16**	ViT-L32
Epochs	50	50	50
Accuracy (%)	93.66	**95.69**	94.03
Precision (%)	--	96.08	94.65
Recall (%)	--	95.69	95.64
F1 score (%)	--	94.03	93.88

VI. Conclusions

In this research, we formulated diverse models utilizing Vision Transformers, including ViT-B16, ViT-B32, ViT-L16, and ViT-L32, to manage student attendance in educational institutions. The most effective model turned out to be ViT-L16. ViT-L16 underwent fine-tuning to perform identification and recognition of students' iris images, leveraging a dataset comprising 8533 iris images.

Furthermore, this research has formulated four models using the Vision Transformer methodology. An evaluation of all the models revealed that, although ViT-B16, ViT-B32, and ViT-L32 performed satisfactorily, the ViT-L16 transformer outshone all the models in terms of accuracy. A comparison of F1 score, precision, and recall provides supporting evidence that the ViT-L16 transformer surpasses all other models. Additionally, when compared with all the models, the ViT-L16 transformer required less training time with an equal number of epochs.

The Vision Transformer model demonstrated the effectiveness of ViT-L16 in identifying and recognizing students' iris patterns. The proposed method represents a notable contribution to advancing the development of a student attendance system capable of recording and monitoring attendance through iris images.

Our experiments manifested the inherent adaptability of Vision Transformer architectures to iris recognition tasks. ViT models displayed robust feature extraction capabilities, allowing for accurate and reliable identification of unique iris patterns. The incorporation of data augmentation techniques, including deformation, blurring, and rotation, played a crucial role in enhancing the robustness of the models. This approach effectively mitigated the effects of intra-class variations within the synthesized iris dataset.

Future research avenues could explore the integration of additional biometric modalities to enhance the overall security and accuracy of attendance systems. Additionally, the refinement of the Vision Transformer architecture, specifically tailored to the unique requirements of educational settings, holds potential for advancing the continuous improvement of biometric-based attendance solutions.

References

[1] F. Bakhri, H. Mohd Ekhsan, and J. N. Hamid, “Students’ Attendance Monitoring System with SMS Notification,” J. Comput. Res. Innov., vol. 5, no. 1, pp. 19–24, 2020, doi: 10.24191/jcrinn.v5i1.159.

[2] A. Dosovitskiy et al., “an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale,” ICLR 2021 - 9th Int. Conf. Learn. Represent., 2021.

[3] T. S. Lim, S. C. Sim, and M. M. Mansor, “RFID based attendance system,” in 2009 IEEE Symposium on Industrial Electronics & Applications, 2009, vol. 2, pp. 778–782. doi: 10.1109/ISIEA.2009.5356360.

[4] K. A. Alnajjar and O. Hegy, “Attendance System Based on Biometrics and RFID,” in 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 596–599. doi: 10.1109/ICIIP47207.2019.8985745.

[5] K. Jayakumar, V. Surendar, A. Sheela, P. Javagar, K. A. Riyas, and K. Dhanush, “Internet of Things based Biometric Smart Attendance System,” in 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022, pp. 1492–1497. doi: 10.1109/ICSCDS53736.2022.9761045.

[6] S. Ennajar and W. Bouarifi, “Deep Transfer Learning Approach for Student Attendance System During the COVID-19 Pandemic,” J. Comput. Sci., vol. 20, no. 3, pp. 229–238, 2024, doi: 10.3844/jcssp.2024.229.238.

[7] P. S. Bhagat and P. S. Y. Chincholikar, “Biometric Attendance System using Iris Recognition,” pp. 263–266, 2016, [Online]. Available: http://www.ijirmf.com/wp-content/uploads/2016/11/201611047.pdf

[8] K. O. Okokpujie, E. Noma-Osaghae, O. J. Okesola, S. N. John, and O. Robert, “Design and Implementation of a Student Attendance System Using Iris Biometric Recognition,” in Proceedings - 2017 International Conference on Computational Science and Computational Intelligence, CSCI 2017, Dec. 2018, pp. 563–567. doi: 10.1109/CSCI.2017.96.

[9] S. A. Shaban, H. M. M. Ahmed, and D. L. Elsheweikh, “A Novel Fusion System Based on Iris and Ear Biometrics for E-exams,” Intell. Autom. Soft Comput., vol. 35, no. 3, pp. 3295–3315, 2023, doi: 10.32604/iasc.2023.030237.

[10] I. A. Hassan, S. A. Ali, and H. K. Obayes, “Enhance iris segmentation method for person recognition based on image processing techniques,” Telkomnika (Telecommunication Comput. Electron. Control., vol. 21, no. 2, pp. 364–373, 2023, doi: 10.12928/TELKOMNIKA.v21i2.23567.

[11] Z. Trabelsi and K. Shuaib, “Implementation of an effective and secure biometrics-based student attendance system,” Int. J. Comput. Appl., vol. 33, no. 2, pp. 144–153, 2011, doi: 10.2316/Journal.202.2011.2.202-2928.

[12] A. Adamu, “Attendance Management System Using Fingerprint and Iris Biometric,” Rabit J. Teknol. dan Sist. Inf. Univrab, vol. 3, no. 4, pp. 427–433, 2019.

[13] S. Kadry and M. Smaili, “Wireless attendance management system based on iris recognition,” Sci. Res. Essays, vol. 5, no. 12, pp. 1428–1435, 2010.

[14] A. Khatun, A. K. M. F. Haque, S. Ahmed, and M. M. Rahman, “Design and implementation of iris recognition based attendance management system,” 2nd Int. Conf. Electr. Eng. Inf. Commun. Technol. iCEEiCT 2015, no. May, pp. 21–23, 2015, doi: 10.1109/ICEEICT.2015.7307458.

[15] M. Sujatha et al., “Attendance management system using iris recognition,” Int. J. Pharm. Res., vol. 11, no. 1, pp. 451–459, 2019, doi: 10.31838/ijpr/2019.11.01.060.

[16] A. Joshy and M. J. Jalaja, “Design and implementation of an IoT based secure biometric authentication system,” in 2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), Aug. 2017, pp. 1–13. doi: 10.1109/SPICES.2017.8091360.

[17] P. Lad and S. More, “Student Attendance System Using Iris Detection,” no. 2, pp. 3293–3298, 2017.

[18] H. M. Therar, L. D. E. A. Mohammed, and A. P. D. A. J. Ali, “Multibiometric System for Iris Recognition Based Convolutional Neural Network and Transfer Learning,” IOP Conf. Ser. Mater. Sci. Eng., vol. 1105, no. 1, p. 012032, 2021, doi: 10.1088/1757-899x/1105/1/012032.

[19] K. Han et al., “A Survey on Vision Transformer,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 87–110, 2023, doi: 10.1109/TPAMI.2022.3152247.

[20] K. Al-hammuri, F. Gebali, A. Kanan, and I. T. Chelvan, “Vision transformer architecture and applications in digital health: a tutorial and survey,” Vis. Comput. Ind. Biomed. Art, vol. 6, no. 1, 2023, doi: 10.1186/s42492-023-00140-9.

[21] Elpina and G. P. Kusuma, “Revolutionizing Computer Vision: Enhanced Food Image Classification With Swin Transformer and Svm Classifier,” J. Theor. Appl. Inf. Technol., vol. 101, no. 23, pp. 7549–7561, 2023.

[22] R. Mehta, S. Shukla, J. Pradhan, K. K. Singh, and A. Kumar, “A vision transformer-based automated human identification using ear biometrics,” J. Inf. Secur. Appl., vol. 78, p. 103599, 2023, doi: https://doi.org/10.1016/j.jisa.2023.103599.

[23] S. A. Latif, K. A. Sidek, and A. H. A. Hashim, “An Efficient Iris Recognition Technique using CNN and Vision Transformer,” J. Adv. Res. Appl. Sci. Eng. Technol., vol. 34, no. 2, pp. 235–245, 2024, doi: 10.37934/araset.34.2.235245.

[24] P. N. Ha, A. Doucet, and G. S. Tran, “Vision Transformer for Pneumonia Classification in X-Ray Images,” in Proceedings of the 2023 8th International Conference on Intelligent Information Technology, 2023, pp. 185–192. doi: 10.1145/3591569.3591602.

[25] A. Kulkarni, D. Chong, and F. A. Batarseh, “Foundations of data imbalance and solutions for a data democracy,” Data Democr. Nexus Artif. Intell. Softw. Dev. Knowl. Eng., pp. 83–106, 2020, doi: 10.1016/B978-0-12-818366-3.00005-8.