Received: 30 September 2024 | Revised: 12 May 2025 | Accepted: 26 June 2025 | Published online: 8 September 2025

RESEARCH ARTICLE

Artificial Intelligence and Applications
2025, Vol. 00(00) 1–13
DOI: 10.47852/bonview52024469
BON VIEW PUBLISHING

# Facial and Gesture Recognition-based Deep Learning Model for Academic Time Attendance at PKRU Demonstration School

**Tanagrit Chansaeng**[1] and **Pita Jarupunphol**[1,*]

[1] *Department of Digital Technology, Phuket Rajabhat University, Thailand*

**Abstract:** This paper describes the design and implementation of an automated academic attendance system using facial and gesture recognition deep learning technologies at the PKRU Demonstration School. In solving issues caused by manual attendance systems, the solution adapts Multi-task Cascaded Convolutional Neural Network (MTCNN) for face detection, implements FaceNet for feature extraction, and performs identity verification with NuSVC. To reduce false positive errors, an additional verification layer-based gesture recognition was incorporated. Over 150 days of operation, the system recorded and processed 8700 instances of attendance with biometric authentication across 58 faculty members and conducted these on standard computing hardware. The system achieved a controlled environment recognition accuracy rate of 100%. Although these results demonstrate the efficiency in the constrained setting accuracy testing environments, expansive or more diverse population studies would be required to benchmark performance in heterogeneous settings. User interaction through gestures directed deliberate confirmation, which improved engagement with routine actions, helping fortify reliability within system functions. In process efficiency metrics, compared to manual methods, the system eliminated 93% of time per event processing while achieving an average per event completion time of 7 s. Following Thailand's Personal Data Protection Act (PDPA), the system was designed to ensure ethical compliance, incorporating explicit consent protocols and secure data handling through Secure Sockets Layer and Transport Layer Security (SSL/TLS) encryption. The proposed solution offers a scalable and cost-effective approach for educational institutions aiming to adopt contactless attendance systems. Future work will focus on expanding the dataset, optimizing performance in crowded or dynamic environments, and integrating additional biometric modalities to enhance accuracy and fairness.

**Keywords:** deep learning, facial recognition, feature extraction, MTCNN, educational technology, demonstration school

## 1. Introduction

Phuket Rajabhat University (PKRU) Demonstration School, established in 1998, operates under the university's educational governance framework. The school offers early childhood education grounded in Waldorf's pedagogical principles, emphasizing age-appropriate developmental stages and experiential learning. This philosophy supports the holistic growth of students through active engagement and discovery-based activities. The curriculum prioritizes academic excellence in science, mathematics, languages, and technology at the secondary level, equipping students with the foundational competencies necessary for higher education and future professional pathways. However, despite its pedagogical innovations, the school continues to rely on traditional administrative practices. For instance, student registration is conducted manually, requiring parents to complete physical forms and submit them on-site. Similarly, attendance tracking for teachers and staff involves handwritten logs, which are transcribed into digital spreadsheets by designated personnel. These legacy processes, established alongside the school's founding, remain in place due to their familiarity and ease of implementation among long-serving staff members.

The significant initiatives undertaken by the PKRU Demonstration School involve adopting facial recognition technology for managing time attendance. This transition eliminates the need for physical contact associated with conventional methods such as signature logs or fingerprint scanners. As Martinez [1] notes, facial recognition technology is grounded in studying how biological systems perceive faces, a process now being emulated by computers. With advances in artificial intelligence and machine learning, modern facial recognition systems can reliably identify individuals from digital images or video frames, making them well-suited for applications in attendance tracking. The present study focuses on developing a cost-effective facial recognition solution, using an existing, underused computer system for academic time attendance. The primary goal is to design a practical, efficient, and budget-friendly system that can be deployed in environments with limited resources. By integrating advanced feature extraction and classification techniques, the system addresses common challenges in traditional attendance methods, such as manual error and inefficiency.

The system uses Multi-task Cascaded Convolutional Neural Networks (MTCNNs) to detect facial regions and employs the FaceNet model to extract robust facial features. This combination enables fast, accurate identification while maintaining low hardware requirements. Beyond facial recognition, the system incorporates gesture recognition as a secondary verification mechanism to improve accuracy and foster active user engagement. This gesture-based interaction is particularly beneficial in school environments, where attendance processes must be intuitive, fast, and minimally disruptive. While facial and gesture recognition technologies are not new, their application in educational institutions presents specific challenges, including resource limitations, data privacy compliance, and the need for user-friendly interfaces. This research directly addresses these challenges by designing a system tailored to the operational context of the PKRU Demonstration School. The outcome is a scalable and replicable model suitable for institutions facing similar financial and administrative constraints. The key research objectives of this study are as follows:

***Corresponding author:** Pita Jarupunphol, Department of Digital Technology, Phuket Rajabhat University, Thailand. Email: p.jarupunphol@pkru.ac.th

1) to develop a facial recognition system that processes attendance data efficiently and identifies individuals with high accuracy; and
2) to construct a system that utilizes existing resources, such as an underutilized computer, thereby minimizing additional financial burden.

## 2. Literature Review

This literature review aims to provide the background of issues associated with the manual attendance system at PKRU Demonstration School and the foundation for understanding the various facets of facial recognition technology and machine learning operations, which are crucial to developing the proposed system.

### 2.1. Attendance monitoring at PKRU Demonstration School

At the PKRU Demonstration School, the manual attendance system has been in place for quite some time. However, this method continues to present several ongoing challenges, particularly in terms of accuracy, efficiency, and the ability to scale with the school's needs. One major issue is the possibility of human error when teachers and staff record attendance times. Mistakes can happen unintentionally, or there may be cases where someone marks an on-time arrival even if a student or employee arrived late, affecting the reliability of the records. These problems are often made worse when handwritten attendance logs are later typed into a computer system. During this process, administrative staff might accidentally enter incorrect information, leading to further inaccuracies. On top of that, the whole system is quite time-consuming. A lot of effort goes into collecting, checking, and entering the data, time that could be better spent on other tasks within the school.

One major drawback of the manual attendance system is the lack of instant access to attendance data. This delay can make it harder to resolve issues quickly or produce accurate records when needed, affecting tasks like payroll processing and compliance with labor regulations. As the school grows or updates its attendance policies, managing the system manually becomes even more challenging. In this case, sticking with outdated methods can make the school appear behind the times. Transitioning to an automated attendance system would be a practical upgrade. It would boost accuracy, lighten administrative workloads, and offer real-time access to essential data, helping the school stay organized and better aligned with its operational goals.

### 2.2. Facial recognition technology applications

Facial recognition technology is increasingly finding its place in various fields such as security, healthcare, and education, valued for its ability to simplify identity verification and automate monitoring processes. In educational settings, its use has grown particularly around automating attendance tracking—a task that, when done manually, can be time-consuming and prone to error. By adopting such systems, schools and universities hope to ease administrative workloads and enhance the accuracy of attendance records. Researchers have explored various methods to implement facial recognition in educational environments. One notable example is the work by Alhanaee et al. [2], who developed a system based on deep learning, utilizing pre-trained convolutional neural networks like AlexNet and GoogleNet. Their approach achieved impressive accuracy but required considerable computing resources, which may not be feasible for all institutions.

In response to such limitations, Rao et al. [3] proposed AttenFace, which uses periodic image snapshots rather than constant video streams for facial recognition. This method offers real-time tracking benefits while using less processing power, but it necessitates frequent image captures, increasing energy use. Other researchers, such as Ashritha et al. [4], have attempted to strike a balance by using convolutional neural networks (CNNs) suited to classroom environments, though these models still demand relatively robust hardware. Other teams have looked into more cost-effective alternatives. Kar et al. [5] employed Principal Component Analysis (PCA) for facial recognition, which works well under stable conditions but tends to struggle when lighting or environmental variables shift. Jadhav et al. [6] combined the Viola-Jones method for detection with PCA and Support Vector Machines (SVMs) for classification. Their model was more budget-friendly but showed limitations in performance, especially with larger or more variable datasets.

Facial recognition research has evolved beyond simple attendance tracking. Yao [7] explored systems that involve stages like face detection, alignment, and feature extraction. His findings highlight that CNNs consistently outperform older methods such as Haar cascades and HOG by automatically learning complex visual patterns from image data. Yao also raised key concerns around fairness, data privacy, and vulnerabilities like adversarial attacks. He suggests that future work should explore multi-modal learning and advocate for stronger policies to ensure ethical and secure use of these technologies.

In a different field, Chen, Wang, and Zhang [8] investigated how generative adversarial networks (GANs) are being used in finance. By combining GANs with models like long short-term memory (LSTM) and multilayer perceptron (MLP), they showed improvements in predicting market trends, detecting irregularities, and handling time-series data. These advances suggest that GANs have the potential to transform financial analysis by making it more precise and flexible. In an era where education is becoming increasingly digital, sticking with outdated methods can undermine an institution's image and progress. This makes it even more critical to adopt automated attendance solutions that are accurate, efficient, and in line with broader goals for technological and institutional growth.

### 2.3. TIPB and DLB approaches for face recognition

Facial recognition technology has evolved significantly over the past few decades, transitioning from basic pattern recognition techniques to sophisticated algorithms capable of identifying and verifying facial features in diverse conditions. Several research works highlight two prevalent combinations, image processing-based (TIPB) [9–11] and deep learning-based (DLB) approaches [2, 12, 13]. TIPB primarily focused on simple geometric feature-based models, employing the Viola-Jones algorithm for face detection by scanning the input image with varying window sizes [12]. Haar features, representing edges and lines, are extracted from each window and fed into a cascade classifier to ascertain the presence of a face. Depending on the method, the detected face region is then processed for recognition using either the Local Binary Pattern Histogram (LBPH) algorithm or PCA and compared with known faces in a database for face recognition [9, 11].

However, TIPB has since been augmented and replaced by advanced machine learning and deep learning techniques. These modern techniques offer higher accuracy and robustness, especially in unconstrained environments. The DLB approach begins with MTCNN for face detection [14–17]. The input image is resized and scanned with windows at different scales. The P-Net, a fully convolutional network (FCN) [13], identifies potential face regions, while the R-Net, a CNN [12, 18, 19], filters out false positives. Another CNN, the O-Net, finalizes face detection and locates facial landmarks. FaceNet, a Deep Neural Network (DNN) [20], extracts facial features and inputs into a classifier trained on known faces for recognition [21, 22].

Although TIPB and DLB approaches offer the advantage of readily available libraries and pre-trained models online, comparative studies indicate that the DLB approach excels in unconstrained environments and surpasses other performance methods [2, 15, 20]. Those utilizing CNNs outperform TIPB approaches in terms of accuracy and adaptability. For example, Agarwal et al. [12] developed an automatic attendance system using facial recognition, highlighting the efficiency of deep learning models in real-world scenarios. Besides, Alhanaee et al. [2] investigated the use of deep transfer learning for facial recognition in attendance systems, demonstrating significant accuracy and processing speed improvements. Bah and Ming [13] improved face recognition algorithms and their application in attendance management, focusing on enhancing robustness against variations in facial expressions and lighting conditions. Razzaq et al. [22] also present a facial recognition-based attendance system designed for educational institutions to improve accuracy and efficiency in tracking student attendance using advanced facial recognition technology. The study demonstrates the implementation and effectiveness of the system, highlighting its potential to streamline attendance management and reduce manual errors. In particular, the combination of MTCNN for face detection and FaceNet for feature extraction has demonstrated superior performance in various studies, providing high accuracy in unconstrained environments [17]. Therefore, we have chosen the DLB approach to implement our proposed system. Table 1 compares TIPB and DLB approaches regarding method, robustness, speed, accuracy, implementation complexity, and applications. This comparison provides a clear understanding of the trade-offs between TIPB methods and DLB approaches, highlighting the advantages of adopting MTCNN in environments where accuracy and adaptability are prioritized over speed and simplicity.

## 2.4. MLOps in facial recognition

Machine Learning Operations (MLOps) is a set of practices aimed at managing the entire lifecycle of machine learning models [23]. This is especially crucial when models need frequent retraining and redeployment, like in facial recognition systems used for attendance. These systems must adapt over time, since people's appearances can change due to aging, hairstyles, or weight fluctuations. Without regular updates, accuracy can drop, leading to errors in identifying individuals. Updating the model with new facial data helps maintain its performance and ensures attendance records stay reliable. MLOps also involves organizing facial image datasets efficiently so that updates can happen quickly. On the deployment side, it ensures new model versions are rolled out smoothly without interrupting system operations. More recently, MLOps practices have started addressing privacy concerns, especially in biometric data systems. Watson and Larson [24] noted that MLOps transforms machine learning from a one-time task into a repeatable, structured process. Their proposed framework brings together principles from DevOps and DataOps, covering everything from data handling to deployment and ongoing monitoring. Using a recommendation system as an example, they show how MLOps can boost reliability, improve teamwork, and help align technical work with business goals.

## 2.5. Privacy, explainability, and cloud intelligence in biometrics

Growing concerns about privacy in facial recognition have pushed developers to create models that are accurate, transparent, and respectful of user data. Addressing this, Shankar et al. [25] introduced a novel method that combines explainable artificial intelligence (XAI) with synthetic data generation. Their solution aims to protect individual identities while still preserving the usefulness of biometric data, a promising move toward privacy-friendly AI. Similarly, Khadidos et al. [26] highlighted the value of synthetic data in sensitive fields like healthcare. Their study showed that adversarial networks can create synthetic datasets that are just as useful as real ones, paving the way for privacy-conscious recognition systems.

Although the innovations are especially relevant for schools, they must comply with strict data protection laws like Thailand's Personal Data Protection Act (PDPA) and the EU's General Data Protection Regulation (GDPR). On top of algorithm development, researchers are also exploring the infrastructure needed to handle large-scale attendance data. For example, Selvarajan et al. [27] proposed a hybrid model using k-Nearest Neighbors (k-NN), Decision Trees, and Deep Q-learning to manage big data in industrial cloud environments. This approach could be adapted for education. In a broader context, Idoko et al. [28] examined the social impacts of AI and big data, particularly in the workforce. Their findings spotlight two major concerns: job losses from automation and ethical questions around collecting biometric data. As facial recognition and fingerprint scanning become more common, there is a growing need for strong regulations and secure systems to ensure that tech advancement does not come at the cost of privacy or job stability.

**Table 1**
**Comparison of TIPB (Viola-Jones algorithm) and DLB (MTCNN) for face detection**

| Feature | TIPB (Viola-Jones algorithm) | DLB (MTCNN) |
|---|---|---|
| Method | Uses Haar-like features and a cascade classifier for efficient face detection. | A deep learning-based approach that improves detection accuracy by jointly optimizing face detection and landmark localization. |
| Robustness | Known for its robustness in real-time applications. | Highly accurate due to deep learning techniques. |
| Speed | Very fast and suitable for real-time applications. | Slower compared to Viola-Jones but provides better accuracy and reliability. |
| Accuracy | Moderate accuracy; can be affected by lighting and pose variations. | High accuracy due to the use of convolutional neural networks. |
| Implementation Complexity | Relatively simple to implement with well-defined steps and readily available libraries. | More complex due to the involvement of multiple neural networks and advanced optimization techniques. |
| Applications | Suitable for applications requiring fast and robust face detection with moderate accuracy, such as real-time face tracking. | Suitable for applications where high accuracy is crucial, such as facial recognition and biometric authentication systems. |

## 3. Research Methodology

This section outlines the methodological steps to develop, implement, and evaluate the proposed facial recognition system for time attendance at PKRU Demonstration School. The methodology comprises system analysis, design, implementation, and evaluation phases.

### 3.1. Requirements gathering and analysis

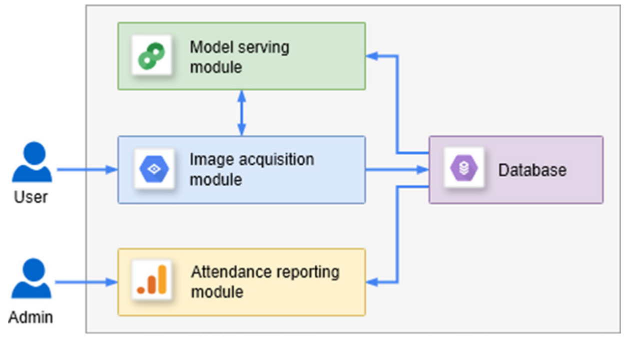
The development of our facial recognition system commenced with a thorough analysis of requirements gathered from key stakeholders. Two main user groups for the system were identified: 1) the teachers and staff at the demonstration school who would utilize the system for workplace sign-in, and 2) administrators responsible for generating attendance reports. Given the increasing application of facial recognition technology, we have also addressed ethical and privacy concerns, which have gained significant attention. Data protection, user consent, and the potential for technology misuse are critically examined in recent literature, underscoring the importance of adhering to ethical guidelines and regulatory frameworks.

During the requirement analysis, we recognized that the proposed system would handle sensitive data, specifically the facial images of users. In compliance with Thailand's PDPA, transparency regarding the storage and use of personal data must be assured. This act mandates that the intended use of personal data be communicated to users, prohibiting the unauthorized use of their data [29]. Accordingly, all users must provide explicit permission before incorporating their data into the system. Next, the hardware specifications required to run the facial recognition system were determined. We opted for a computer equipped with an Intel Core 2 Duo processor and a VGA camera, setting this configuration as the minimum system requirement. This choice was influenced by the need to balance performance with cost-effectiveness.

### 3.2. System design

The system architecture has been thoughtfully designed to meet both the functional and practical needs of the proposed solution. The architecture is built around the MTCNN framework [15] and includes three main components: Image Acquisition (IA), Model Serving (MS), and Attendance Reporting (AR). Each module contributes to the system's smooth operation and user experience. The IA module is responsible for capturing facial images and prepping them for analysis. These processed images are then sent to the MS module, which handles face detection and identity recognition using MTCNN and related models. Once a face is recognized, the AR module generates and organizes attendance records in a clear, easy-to-use format for administrative purposes. Figure 1 shows a top-level view of how these modules work together. The following sections break down each module's design and role in more detail.

**Figure 1**
**An architecture of the proposed system**




#### *3.2.1. IA module*

The IA module acts as the main interface between users and the facial recognition system, capturing live video during the recognition process. It bridges the hardware and software, ensuring image capture is both fast and reliable. Built with OpenCV, a well-known open-source computer vision library, the module connects directly to the camera, capturing frames in resolutions optimized for facial recognition. By using OpenCV's advanced features, it maintains high image quality and supports real-time performance. A key feature is its frame buffering system, which ensures continuous video capture, even in busy environments with many users. It prevents frame loss due to brief processing or network delays. The module's multi-threaded design further boosts performance by separating image capture from image processing. This separation minimizes delays and ensures smooth, uninterrupted operation, which is critical in places like schools where quick interactions are common.

Once video frames are captured, they are sent to the Management and Storage (MS) module for facial identification. The IA and MS modules work together in real time, enabling quick feedback after a recognition event. When a user is successfully identified, the IA module logs their ID along with a timestamp in the attendance system. The IA module is a dependable entry point for the facial recognition system. It plays a central role in handling user-facing image capture and data transfer, directly influencing the system's speed and accuracy. It is also designed to work with various hardware setups, making it a versatile option for different institutional environments.

#### *3.2.2. MS module*

The MS module is an essential part of the facial recognition system, responsible for analyzing images and confirming user identities. When it receives input from the IA module, it extracts unique facial features and compares them to a database of stored profiles to verify the match. Once a user's identity is confirmed, the results are sent back to the IA module for the next steps. The MS module can handle high-performance tasks and multiple users at once. This modular design reduces system slowdowns and makes resource management more efficient. Tasks are managed in a first-in-first-out (FIFO) queue, so each recognition request is handled in order. Moreover, the system caches facial data of frequently seen users, allowing for faster recognition during busy times when speed matters most. It also includes a self-updating feature that refreshes the recognition model with new user data. This helps the system adjust to natural changes in appearance, like aging or new hairstyles, so it stays accurate over time. Regular updates ensure that the system remains dependable and delivers a smooth, user-friendly experience.

#### *3.2.3. AR module*

The AR module is a key part of the backend in a facial recognition-based attendance system. Its primary role is handling administrative and analytical tasks related to tracking attendance. Working alongside the IA module that captures images and timestamps when a face is recognized, the AR module manages how attendance data is recorded, retrieved, and reported. Administrators can generate customized attendance reports by day, week, or month. These reports help spot attendance trends, ensure that rules are followed, and highlight unusual patterns. The system uses an SQLite database, which offers a lightweight yet efficient way to store attendance data, ideal for schools or institutions with limited resources. The AR module is built with a RESTful API, making it simple to connect with other systems like HR or payroll platforms. This flexibility makes the system practical for various educational settings. The module also includes data visualization tools, including dashboards and performance metrics. These features help administrators quickly understand attendance behavior, identify high and low attendance periods, monitor absenteeism, and track engagement over time.

## 3.3. Model development

The implementation phase of our model development was systematically divided into two main processes: model training and testing and model validation. Below is a detailed mathematical representation of each stage, specifically tailored to the machine-learning context of facial recognition.

*3.3.1. Model training and testing*

The dataset $D$ prepared for this study consisted of a total of 290 images, where each user $u_i(where\ i = 1, 2, \ldots, 58)$ had five images $I_{i,j}(where\ j = 1, 2, \ldots, 5)$ The images were also stored in GitHub repository [30]. Out of these five images per user, four images $I_{i,1}, I_{i,2}, I_{i,3}, I_{i,4}$ were allocated to the training set $D_{train}$, and one image $I_{i,5}$ was allocated to the testing set $D_{test}$. The total number of images in the training set $|D_{train}| = 232$ and in the testing set $|D_{test}| = 58$. The dataset $D$ is structured in Equations (1) and (2) to represent the training process where $N_u$ denotes the number of users, each having $N_i$ images. The first $N_t$ images per user are used for training ($T$), and the remaining one image is used for testing V.

$$D = \bigcup_{j=1}^{N_u} I_{uj}, \quad I_{uj} = \{i_{j1},\ i_{j2},\ \ldots,\ i_{jNi}\} \tag{1}$$

$$T = \bigcup_{j=1}^{N_u} \{i_1,\ \ldots,\ i_{jNt}\}, \quad V = \bigcup_{j=1}^{N_u} \{i_j(N_t + 1)\} \tag{2}$$

The MTCNN model $M_{MTCNN}$ was used to detect faces in each image $I \in D$. For each detected image $I$, the model performed face cropping, producing a cropped face image $I'$. This step can be mathematically represented as Equation (3).

$$I' = M_{MTCNN}(I) \tag{3}$$

Subsequently, the FaceNet model $M_{FaceNet}$ was employed to extract facial embeddings $\mathrm{e}$ from each cropped face image $I'$. The embedding $\mathrm{e}$ is a 128-dimensional vector representing the unique facial features of the individual. The extraction process is described in Equation (4).

$$\mathrm{e} = M_{FaceNet}(I') \tag{4}$$

Each image $i$ undergoes a two-step transformation process as Equation (5).First, the MTCNN function $f_{MTCNN}$detects and isolates the facial region, producing a cropped image $i_c$. This cropped image is then passed to the FaceNet model $f_{FaceNet}$, which generates a feature vector $\mathrm{e} \in R^{128}$.

$$i_c = f_{MTCNN}(i), \quad \underset{e \,\in\, \mathbb{R}^{128}}{e = f_{FaceNet}(i_c),} \tag{5}$$

The classification model $M_{classifier}$ is then trained on the training set $D_{train}$ to map embeddings $\mathrm{e}$ to their respective user labels $y$. The goal of the classifier is to minimize the loss function $L(\theta)$, where all $\theta$ are the model parameters optimized using gradient descent methods. The objective function for the model training can be expressed as Equation (6).

$$\theta^* = \arg\ min_\theta \frac{1}{|D_{train}|} \sum_{(\mathrm{e},y)} L(M_{classifier}(\mathrm{e}; \theta), y) \tag{6}$$

The trained model $M_{classifier}(\theta^*)$ was then validated on the test set $D_{test}$, where each test image's facial embedding $\mathrm{e}'$ was obtained using the same process described above. The model's performance $P$ was evaluated based on its accuracy $A$, defined as the ratio of correctly predicted labels $\hat{y}$ to the total number of test samples $|D_{test}|$, where I, the indicator function, returns 1 if the prediction is correct and 0 if otherwise. This process can be explained in Equation (7).

$$A = \frac{1}{|D_{test}|} \sum_{(\mathrm{e}', y') \in D_{test}} \mathbb{I}\left(M_{classifier}\left(\mathrm{e}'; \theta^*\right) = y'\right) \tag{7}$$

The developed system demonstrated its capability to identify an individual in the presence of up to two additional individuals in the camera frame, maintaining high accuracy. However, the performance metric A was observed to decrease when more than three individuals were present. This indicates the model's sensitivity to the number of faces in the input frame, necessitating future improvements to handle such scenarios. In this case, the facial recognition system was constructed using facial embeddings $\mathrm{e}$ and their corresponding labels $y$, which together formed the foundation for model training and evaluation. This end-to-end development process, from data preparation to model validation, was conducted in alignment with established machine learning best practices, ensuring the system's accuracy and reliability. As illustrated in Figure 2, the workflow begins with preprocessing raw images from the training and testing datasets. These images are initially processed using the MTCNN to detect and crop facial regions. The cropped facial images are passed through the FaceNet model to generate fixed-length embedding vectors. These embeddings, paired with their respective identity labels, serve as input for training and evaluating the classification model, which identifies individuals during system operation.

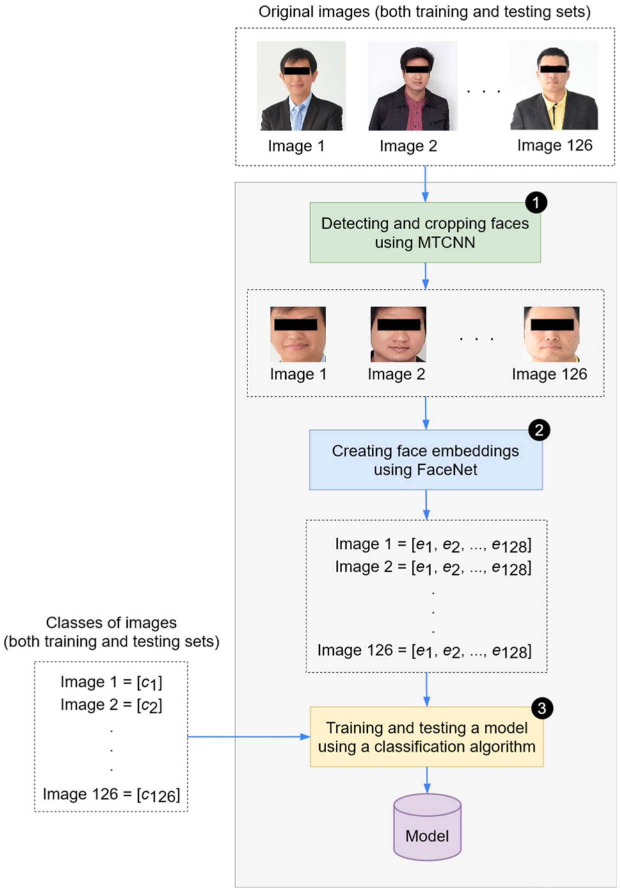
*3.3.2. Model validation*

This study examines four classification algorithms to determine the most appropriate model for facial and gesture recognition in time attendance systems. The algorithms considered Logistic Regression, Support Vector Machine (SVM), Random Forest, and Nu-Support Vector Classification (NuSVC) for exhibiting different strengths and limitations, which make them suitable for different machine learning tasks and types of data distributions [31–36].

Logistic Regression is a commonly used method for binary classification tasks. It estimates the probability of a certain outcome by modeling a linear relationship between the input variables and the log-odds of that outcome [31]. The method is valued for its clarity, speed, and simplicity in implementation. SVMs are robust classifiers, particularly effective in high-dimensional datasets or cases with more features than samples [31, 32]. SVM works by identifying a hyperplane that best separates different classes while maximizing the margin between them. With the right kernel and regularization, SVM can generalize well and is less likely to overfit [32]. Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs for final predictions [33–35]. It is especially effective for handling complex, high-dimensional datasets. This method works effectively with continuous and categorical data and is known for its high accuracy and resistance to overfitting, thanks to its averaging process.

NuSVC is a variant of the standard SVM algorithm that replaces the conventional C parameter with a $\nu$ (*nu*) parameter. This formulation allows the user to control the upper bound on the fraction of margin errors and the lower bound on the fraction of support vectors [36]. The result is a more flexible mechanism for balancing model complexity

**Figure 2**
**Creating a face-recognition model**




with generalization performance. Like traditional SVM, NuSVC performs well in high-dimensional settings and maintains strong resistance to overfitting. These models were evaluated to identify the most efficient and accurate algorithm for the proposed facial and gesture recognition-based attendance system. Table 2 presents a detailed summary of the mathematical formulations and operational characteristics of the algorithms examined in this study, including Logistic Regression, SVM, k-Nearest Neighbors (k-NN), Random Forest, and NuSVC.

We computed the accuracy $A$ for each model using the testing dataset $D_{test}$ to evaluate the performance of each model, where $M(\mathrm{e}_i; \theta^*)$ is the predicted label of the model for a test sample $\mathrm{e}_i$ as described in Equation (8).

$$A = \frac{1}{|D_{test}|} \sum_{i=1}^{|D_{test}|} \mathbb{I}(M(\mathrm{e}_i; \theta^*) = y_i) \tag{8}$$

The performance evaluation results are summarized in Table 3. Among the models assessed, the NuSVC model is denoted as $M_{nusvc}(\theta^*_{nusvc})$, exhibited the highest performance, achieving an accuracy rate of 100% on the testing set. This outstanding result underscores the model's effectiveness in accurately classifying facial identities, making it suitable for deployment in a real-world attendance system. Therefore, given its superior performance, the NuSVC model was selected as the core classification component of the proposed facial recognition framework. This choice reflects a commitment to prioritizing accuracy and reliability, which are critical for maintaining the integrity and usability of automated attendance systems in operational environments.

Upon testing, the NuSVC model exhibited exceptional performance, achieving 100% accuracy in facial and gesture recognition tasks. This outcome highlights the model's effectiveness in supporting the automated time attendance system. The selection of NuSVC as the system's primary classifier was driven by its consistent ability to deliver accurate and reliable results, an essential requirement for real-world deployment in educational environments.

## 3.4. Software development

The IA and MS modules were developed in Python, focusing on smooth interaction between these key parts of the facial recognition system. The process starts with the IA module capturing a real-time image of the user. This image is then passed to the MS module for detailed analysis based on the following steps.

1) Face Detection and Cropping: The MTCNN algorithm scans the image, detects the face, and extracts just the facial area for further analysis.
2) Embedding Generation: The cropped face is processed using the FaceNet model, which converts it into an embedding, a unique vector representing the person's facial features.
3) User Identification: This embedding is fed into a NuSVC classifier (trained during development) that compares it with known embeddings to identify the user.
4) Result Transmission: The system sends the recognized username back to the IA module after the user is identified. This enables actions like marking attendance to be carried out.

Following this, the IA module performs a visually informative action by overlaying a red frame with the predicted name over the user's face in the image. It then diligently records the image, the identified username, and the exact attendance time in a database, completing the attendance recording process. To ensure the system's adaptability and accuracy over time, the MS module incorporates a continuous training feature. This feature is programmed to automatically extract users' images from the database and periodically retrain the model using these new images every month. On the other hand, the AR module was developed using the PHP programming language. In this module, an administrator initiates the process by logging in and selecting a specific date for which they wish to generate an attendance report. The AR module then retrieves a list of users from the database on the selected date.

## 3.5. System testing and validation

To ensure a smooth rollout of the facial recognition system, users were given clear guidance from the start. Instructions were shared verbally and posted in printed form next to the system interface. This two-way approach made the system more user-friendly, catering to different learning styles and helping users feel comfortable with the new technology.

### *3.5.1. Initial deployment and user training*

User training was conducted to ensure seamless system adoption. This includes two primary approaches: verbal explanations $U_v$ and printed instructions $U_p$. The goal was to minimize user errors $E_u$ during system usage, where $f$ represents the function relating user training to system usage errors as shown in Equation (9).

**Table 2**
**Comparison of Logistic Regression, SVM, Random Forest, and NuSVC**

| Algorithm | Description | Equation |
|---|---|---|
| Logistic Regression | A linear model for binary classification that uses a logistic function to model the probability of the default class. The training objective is to find parameters $\theta_{lr}$ that minimize the following cost function, where $\log(M_{lr}(e_i;\theta_{lr}))$ is the natural logarithm of the predicted probability for class 1; and $\log(1-M_{lr}(e_i;\theta_{lr}))$ is the natural logarithm of the predicted probability for class 0. | $\theta^*_{lr} = \arg\min_{\theta_{lr}} \frac{1}{N}\sum_{i=1}^{N}[y_i \log(M_{lr}(e_i;\theta_{lr})) + (1-y_i)\log(1-M_{lr}(e_i;\theta_{lr}))]$ |
| SVM | A classifier that finds the hyperplane that best separates the data into two classes, C is a regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error. | $\theta^*_{svm} = arg\ min_{w,b}\frac{1}{2}\lVert w\rVert^2 + c\sum_{i=1}^{N} max\left(0, 1-y_i\left(w^T e_i + b\right)\right)$ |
| k-NN | A non-parametric algorithm is used for classification and regression. It classifies a data point based on the majority class among its k nearest neighbors, where $N_k(x)$ denotes the set of k nearest neighbors of x. | $\hat{y} = mode(y_i\ for\ i\ \in N_k(x))$ |
| Random Forest | An ensemble method that builds multiple decision trees and merges them to get a more accurate and stable prediction. $\mathbb{I}$ is an indicator function that returns 1 if the predicted label does not match the actual label and 0 otherwise. | $\theta^*_{rf} = arg\ min_{\theta_{rf}}\frac{1}{T}\sum_{t=1}^{T}\sum_{e_i,y_i\in D_t}\mathbb{I}\left(M^t_{rf}(e_i;\theta_{rf}) \neq y_i\right)$ |
| NuSVC | A variant of SVM that uses a parameter ν to control the number of support vectors and margin errors, where $v$ is a parameter that controls the support vectors, $\rho$ represents the margin, and $\xi_i$are slack variables that allow some misclassifications. | $\theta^*_{nusvc} = arg\ min_{w,b,\xi,\rho}\frac{1}{2}\lVert w\rVert^2 - v\rho + \frac{1}{N}\sum_{i=1}^{N}\xi_i$ |

**Table 3**
**Table of abbreviations and variables**

| Model | Acc (%) | Pre (%) | Rec (%) | F1 (%) |
|---|---|---|---|---|
| Logistic Regression | 98.4 | 97.4 | 98.3 | 97.7 |
| SVM | 98.9 | 98.7 | 98.5 | 98.9 |
| Random Forest | 96.8 | 94.8 | *96.6* | 95.4 |
| k-NN | 95.2 | 92.2 | 94.8 | 93.1 |
| NuSVC | 100 | 100 | *100* | 100 |

$$E_u = f(U_v,\ U_p) \tag{9}$$

While most users could check in without issues, the system failed to recognize four individuals correctly. After reviewing the errors, it became clear that the main problem was changes in those users' physical appearances compared to the images used during training. Factors like new hairstyles, weight changes, or other facial differences affected the system's accuracy. This highlighted the importance of developing ways for the model to adjust to natural, real-life changes in appearance over time.

*3.5.2. Error analysis and model retraining*

In the first phase of testing, the system accurately identified most users. However, it failed to recognize four users correctly $U_{mis} = \{u_1,\ u_2,\ u_3,\ u_4\}$ due to changes in their physical appearance $C(u_i)$, such as hairstyle or weight variations. The error rate $E$ increased for these users, leading to an analysis of discrepancies between current and training data. The set of discrepancies $D$ for each user $u_i$ can be defined as Equation (10), where $F_{current}(u_i)$ represents the current facial features of user $u_i$, and $F_{train}(u_i)$ represents the features used during initial model training.

$$D(u_i) = |F_{current}(u_i) = F_{train}(u_i)| \tag{10}$$

To rectify this, three new images $I'u_i = \{I'i_1, I'i_2, I'i_3\}$ were collected for each affected user $u_i$. These new images were used to update the training dataset $D_{train}$, and the model was retrained to minimize the discrepancy set $D$. Equation (11) represents the updated model parameters $\theta^*$ obtained by retraining, where $N'$ is the number of new training samples. $L$ is the loss function, and $y'_{u_i}$ are the updated labels.

$$\theta^* = argmin_\theta \frac{1}{N'} L\left(M_{classifier}\left(M_{Facenet}\left(I'_{u_i}\right);\theta\right), y'_{u_i}\right) \tag{11}$$

*3.5.3. Post-retraining validation*

After retraining, the system's performance was re-evaluated using the updated model $M_{updated}(\theta^*)$. The validation phase involved an additional month of testing. Equation (12) shows that each new test sample $I_{test}$ was processed through the system to verify its accuracy $A_{updated}$, where $\mathbb{I}$, the indicator function, returns 1 if the prediction is correct and 0 if otherwise. The post-retraining phase showed that the system correctly identified all users, achieving a perfect accuracy of 100% $A_{updated}$=1.

$$A_{uodated} = \frac{1}{|D_{test}|}\sum_{i=1}^{|D_{test}|}\mathbb{I}(M_{updated}(M_{Facenet}(I_{test});\theta^*) = y_{test}) \quad (12)$$

*3.5.4. Gesture recognition and system interactivity*

A key feature of the system is selective data recording based on gesture recognition. The system only records attendance if the user performs a specific gesture $G$ (e.g., thumbs-up). The gesture detection model $M_{gesture}$ operates as Equation (13), where $G_{detected}$ is the detected gesture. The system records the attendance time T if and only if the gesture matches the expected input, as in Equation (14).

$$G_{detected} = M_{gesture}(I)\ if\ G_{detected} = thumbs - up \quad (13)$$

$$T = \begin{cases} Record(t) & if\ G_{detected} = thmbs - up \\ No\ Recoed & otherwise \end{cases} \quad (14)$$

This feature ensures precise and efficient attendance recording, enhancing user engagement by responding to specific gestures, such as raising hands or fingers. The gesture recognition accuracy $A_{gesture}$ contributes to the overall system's interactivity as in Equation (15).

$$A_{gesture} = \frac{Number\ of\ correct\ gestures\ detected}{Total\ gestures\ performed} \quad (15)$$

This development makes the system much more interactive, practical, and functionally engaging to users. During an extended testing phase, the system demonstrated flawless user identification and gesture recognition performance, achieving perfect accuracy during the entire testing window. These results underscore its reliability and flexibility within real-world classroom settings. The gesture recognition capability helps minimize errors as intentional confirmations, ensuring that shared or crowded spaces do not lead to a surge in inaccuracies. Unlike passive methods that could anger due to unintended gestures, users preferred more active forms involving simple thumbs-up signals, while contextual to education.

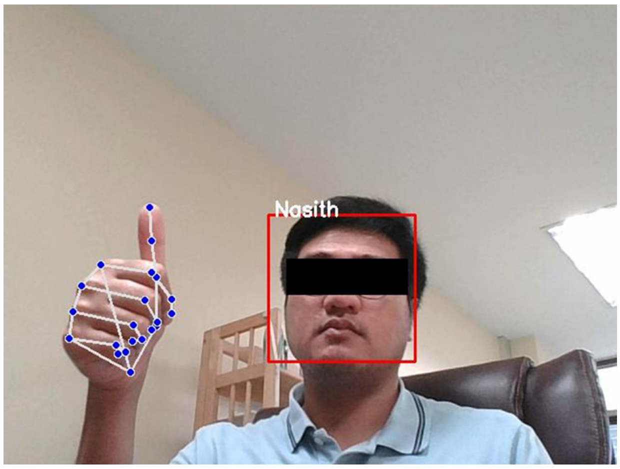
The system demonstrates hand detection capabilities where specific movements such as raising one's hands or extending fingers are tracked through some identified points on the hand, like in Figure 3. With face recognition functions wherein a teacher can be named and recognized biographically, gesture input supplements verification steps, strengthening them. Attaining high accuracy using both facial and gesture recognition makes the system ideal for secure attendance within dynamic learning environments due to its precision and reliability.

## 3.6. System implementation

Over 150 days, the PKRU Demonstration School used a facial and gesture recognition system to monitor academic staff attendance. This extended testing phase offered valuable insights into the system's real-world performance, allowing for fine-tuning and validation. The results showed consistently high reliability, with the system maintaining perfect accuracy during the entire period. The gesture recognition feature was especially effective, responding within roughly 3 s and helping streamline check-ins while minimizing errors.

In the first week, some staff hesitated to switch from familiar manual methods, a common reaction to new technology. However, this resistance faded quickly. By the end of the week, most users had adapted and recognized the system's speed and reliability. As it became part of the school's daily routine, the system noticeably reduced the administrative workload tied to attendance and became a reliable operational tool. This case study highlights both the technical reliability of the system and the importance of supporting user adaptation during implementation. The

**Figure 3**
**Gesture recognition interface of the system**

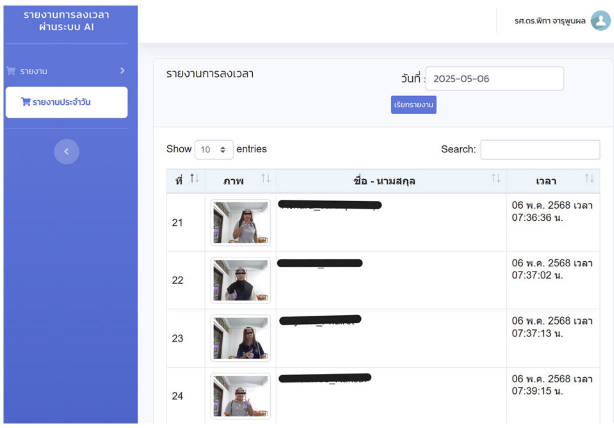



experience suggests that similar systems can be successfully adopted in other educational settings with proper onboarding. Figure 4 shows the system's login screen and a sample admin report, with personal data and faces blurred for privacy.

## 3.7. Ethical considerations

Ethical considerations were central to this study, mainly due to the sensitive nature of collecting biometric data. Facial and gesture recognition required gathering facial images from faculty members at PKRU Demonstration School. To ensure transparency and compliance, all data collection followed the school's Personal Data Protection Policy, which took effect on March 1, 2024. Before any data was gathered, informed consent was obtained from all 58 participating staff members. Each received a formal consent form titled "Document of Intent to Consent to Disclosure of Information and Personal Identity of the Data Provider." The consent process took place from September 6 to 14, 2024. Every form required three signatures: one from the staff member confirming their voluntary participation, one from the principal investigator to ensure ethical conduct, and one from the deputy director as a witness. This thorough process ensured that all participants were well-informed about the system's goals, their data use, and the safeguards to protect their privacy.

**Figure 4**
**System implementation at PKRU Demonstration School**



The system's user interface also displayed the data protection policy to promote ongoing transparency, allowing staff continuous access to relevant information. Participants retained the right to withdraw consent without penalty, and any associated data would be immediately removed from the system upon request. All public-facing materials anonymized identifiable information by blurring names and masking facial features with black bars in illustrative figures. All data were collected under the direct supervision of the principal investigator and stored securely in full compliance with Thailand's PDPA. Access was strictly limited to authorized personnel, and no breaches or unauthorized disclosures occurred throughout the study.

## 4. Results

The facial recognition attendance system was officially submitted to the Office of Academic Resources and Information Technology (ARIT) at PKRU. ARIT, a central administrative unit equivalent to a faculty, oversees the university's smart infrastructure and digital platforms. ARIT has assumed responsibility for its ongoing management and maintenance since the system's initial deployment at the PKRU Demonstration School. This section presents the evaluation results of the facial recognition system's performance in the context of academic staff attendance tracking. The assessment focused on key performance indicators, including recognition accuracy, processing speed, and user adaptability, to determine the system's effectiveness and practical utility within an educational environment.

### 4.1. Functional requirements

The facial recognition system was evaluated to determine how well it met its functional goals, with a key focus on user authentication. Testing involved 58 academic staff members, and the system successfully identified and verified each participant using facial recognition. This demonstrated high accuracy, essential for any reliable automated attendance system. The gesture recognition feature was also assessed. Participants were asked to give a thumbs-up gesture as a second verification step. The system consistently detected this gesture, enabling successful attendance logging. This step added an intentional action, reducing the risk of accidental entries and ensuring user-initiated check-ins.

The combined facial and gesture recognition workflow was observed to evaluate the complete attendance process. After completing both steps, the system automatically logged attendance. This efficient process showed the system's ability to manage multiple tasks smoothly and accurately. Data handling and security were also key areas of focus. Attendance records were transmitted and stored using Secure Sockets Layer and Transport Layer Security (SSL/TLS) protocols, ensuring data protection. Verification checks confirmed that all records were safely stored in a secure database, in line with data privacy regulations. Moreover, the system generated administrative reports that were accurate and thorough, proving useful for school record-keeping and strategic planning.

### 4.2. Non-functional requirements

To ensure a robust and reliable solution, we comprehensively evaluated the facial recognition attendance system's non-functional requirements, assessing its performance, user satisfaction, and operational reliability. We were particularly impressed with its perfect recognition accuracy during initial deployment. This system stands out by integrating both facial and gesture recognition, which allows it to outperform traditional single-mode systems and maintain high recognition rates regardless of lighting changes or user appearance. Such consistent and accurate attendance recording reduces errors and ensures data reliability.

While the NuSVC model achieved 100% accuracy within the test environment, these findings should be interpreted with caution. The evaluation was based on a relatively small and controlled dataset — 290 images from 58 users. As with similar machine learning applications, exceptional performance in a limited setting does not automatically guarantee similar outcomes in larger, more diverse populations without additional training data and model refinement.

We also assessed the system's responsiveness by its processing speed. Attendance logging is consistently completed in under 7 s, meeting our targets for user convenience and efficiency. This performance held steady over a 150-day deployment, handling 8700 attendance events (58 users daily) without any slowdown, demonstrating strong scalability. For data security, all user data was rigorously encrypted (in transit and at rest with SSL/TLS protocols [37]), ensuring regulatory compliance and protecting privacy.

User experience testing showed that the system interface was intuitive and required minimal instruction, leading to a smooth transition for academic staff. Most users became comfortable with the system shortly after it was introduced. The design followed privacy-by-design principles, collecting only necessary data and maintaining compliance with privacy standards, which helped foster user trust. The gesture-based confirmation, specifically the thumbs-up signal, functioned effectively in confirming intentional user engagement, thereby mitigating the risk of false positives, an issue common in facial recognition systems operating independently. Notably, the system could log attendance within 3 s of detecting the gesture, underscoring its real-time responsiveness.

The use of MTCNN for facial detection and FaceNet for feature extraction significantly contributed to the system's resilience in the face of everyday challenges such as poor lighting and changes in facial expressions, consistent with findings reported in previous studies [8, 10]. Additionally, the continuous retraining of the NuSVC model allowed the system to adjust to changes in user appearance—such as hairstyle alterations or signs of aging—helping reduce model drift, a frequent limitation in static machine learning systems.

The system's deployment at PKRU Demonstration School significantly improved administrative workflows. By moving from manual processes to AI-based attendance tracking, the school reduced staff workload, improved data accuracy, and boosted overall efficiency. These results support the idea that AI-driven tools can be effectively used in educational settings and demonstrate their wider potential. The system functioned reliably on low-spec hardware, such as a Core 2 Duo processor, while handling real-world conditions, underscoring its value for institutions with limited resources. Future upgrades will focus on more complex environments, using better hardware, larger datasets, and additional biometric features. These enhancements aim to improve performance in speed, accuracy, and adaptability. This study shows how AI can transform school management and paves the way for future innovations in education technology [21].

### 4.3. Research contributions

This study offers several significant practical contributions. The system we introduced provides an automated, contactless way to monitor attendance, a particularly relevant innovation in post-pandemic education, where minimizing physical interaction remains crucial. By leveraging existing hardware, our solution presents a budget-friendly alternative to traditional attendance procedures, making it especially suitable for institutions with financial limitations. The system has also proven highly accurate and reliable, enhancing administrative workflows by simplifying tasks like attendance recording and payroll, ultimately easing the burden on educational staff.

Moreover, integrating gesture recognition adds an interactive element that could improve user experience and encourage wider adoption [21]. Of course, implementing such systems in schools comes with practical hurdles, including hardware limitations, data privacy concerns, and the need for users to adapt to new tech. Our research tackles these issues head-on by proposing a scalable and adaptable solution, specifically designed to meet the operational demands of academic institutions. Ultimately, these findings contribute meaningfully to educational technology, offering a replicable model for other organizations looking to modernize their attendance systems efficiently and affordably.

*4.3.1. Limited hardware and budget constraints*

Educational institutions often face inadequate budgets, making it hard for them to adopt advanced technology. For example, many current facial recognition systems need specialized hardware or powerful computers, creating significant cost hurdles for schools and universities. In response to these challenges, the solution in this study focuses on being cost-effective and easy to deploy. It uses standard computing devices and consumer-grade cameras without sacrificing performance. Furthermore, the system runs efficiently on standard or even underutilized hardware. This approach ensures reliable automated attendance management and fits within the financial realities of institutions with limited technology budgets, ultimately helping to modernize administration in under-resourced educational environments.

*4.3.2. Accuracy in diverse and dynamic conditions*

Facial recognition technologies often struggle with accuracy when faced with inconsistent lighting, changes in facial expressions, or even subtle shifts in a user's appearance over time. These issues are especially apparent in schools, where lighting and individual behavior can vary significantly daily. Our system uses MTCNN for face detection and the FaceNet model for feature extraction to overcome these challenges. Both have proven remarkably resilient in diverse, real-world situations. In addition, we have built a continuous retraining framework that adapts to gradual changes in a user's appearance, whether a new hairstyle or simply aging. This adaptive approach is crucial for maintaining recognition accuracy and ensuring reliable performance long-term within the ever-changing environment of educational institutions.

*4.3.3. Reduction of false positives in attendance logging*

Facial recognition systems used for attendance can sometimes log false positives, such as mistakenly recording people passing by. To address this, the proposed system adds a second verification step: gesture recognition. Once a face is identified, the user must perform a specific gesture, like a thumbs-up, to confirm their presence. This extra step ensures that only intentional actions result in recorded attendance, helping reduce errors. Combining facial and gesture authentication can improve accuracy and strengthen data integrity, ensuring each record is tied to a deliberate user interaction.

*4.3.4. User engagement and system usability*

Whether manual or partially automated, traditional attendance systems often feel monotonous and unengaging. This can lead to inconsistent use and more errors. To tackle these issues, our proposed system integrates gesture recognition, encouraging more active participation from users. By adding this interactive element, the system does not just speed up the check-in process; it also makes it more user-friendly and less disruptive. This is especially beneficial in educational settings, where efficient and routine operations are crucial. Moreover, using gestures fosters a greater sense of personal involvement and responsibility, leading to more consistent and reliable system use over time.

*4.3.5. Scalability and real-time processing needs*

Educational attendance systems need to scale with growing student populations while maintaining performance. In this study, the proposed system was tested over an extended period, and large volumes of attendance data were consistently handled with accurate, timely responses. These results indicate that the system is highly scalable and can support institutional expansion without frequent hardware upgrades or major system changes. Its flexibility and reliability make it a strong long-term solution, well-suited to meet the changing needs of educational environments while ensuring smooth, efficient operations.

## 4.4. Comparative results

This section compares our approach with recent automated attendance systems to emphasize our implemented model's robustness, efficiency, and reliability. Regarding detection and recognition, Alhanaee et al. [2] use transfer learning with pre-trained CNN models, specifically AlexNet, GoogleNet, and SqueezeNet, for face recognition, achieving high prediction accuracy through deep learning. However, substantial computational resources are required for CNN models, which may be limited in low-resource educational settings. This is consistent with Ashritha et al. [4], who utilized CNNs for face detection and recognition. While CNNs offer robust accuracy, they can be computationally intensive for real-time applications in resource-constrained environments. In Rao et al. [3], the AttenFace system employs FaceNet and YOLO for object detection and face recognition through snapshots taken every 10 min, ensuring students stay for the class duration. This setup is innovative for continuous monitoring but can be computationally demanding. Alhanaee et al.'s and Rao et al.'s systems rely on CNN architectures, which, while accurate, can be resource-intensive and may struggle with scalability due to hardware demands.

In addition, Kar et al. [5] implemented PCA for face recognition, which is lightweight but less effective under varying lighting conditions and complex backgrounds. Similarly, Jadhav et al. [6] combined Viola-Jones for face detection, PCA for feature extraction, and SVM for classification. Although Viola-Jones is efficient for real-time face detection, the PCA's performance may decline with larger datasets due to its sensitivity to illumination changes. Furthermore, PCA-based systems, such as those in Kar et al., are lightweight but unsuited for dynamic, large-scale environments due to limited adaptability. As such, our approach utilized MTCNN for face detection, which effectively handles diverse lighting and orientation conditions, and FaceNet for face recognition, chosen for its adaptability and precision. The gesture recognition layer further enhances accuracy by requiring users to verify presence intentionally, mitigating false positives, a common issue in automated attendance systems without secondary verification.

For accuracy and performance, the AttenFace and CNN-based systems report high accuracy but require periodic image capture or substantial computational power to maintain real-time performance [2, 3]. Moreover, systems employing PCA, like those of Jadhav et al. [6] and Kar et al. [5], provide moderate accuracy but struggle in real-time performance due to PCA's limitations in feature representation under complex conditions. By integrating MTCNN and FaceNet, our model achieves 100% accuracy in controlled tests while maintaining efficient processing. Additionally, gesture-based verification reduces false positives, a feature absent in most compared systems. Among these systems, most do not detail specific security measures. Ashritha et al. and Rao et al. mention data handling but lack comprehensive security protocols. We prioritize data security by implementing SSL/TLS encryption and ensuring PDPA compliance. Additionally, we gather user consent, which is particularly critical for institutional applications where data privacy is a priority. While

**Table 4**
**Comparative summary of face recognition attendance systems**

| Feature | Alhanaee et al. (2021) | Rao et al. (2022) | Ashritha et al. (2022) | Kar et al. (2012) | Jadhav et al. (2017) | Proposed system |
|---|---|---|---|---|---|---|
| Detection algorithm | AlexNet, GoogleNet | YOLO, FaceNet | CNN | PCA | Viola-Jones | MTCNN |
| Recognition algorithm | Transfer learning CNNs | FaceNet | CNN | PCA | PCA | FaceNet |
| Accuracy | High | High | High | Moderate | *Moderate* | 100% |
| Verification | None | Snapshot monitoring | None | None | *None* | Gesture recognition |
| Data security | Not specified | Limited | Not specified | Not specified | *Not specified* | SSL/TLS, PDPA compliance |
| Scalability | Limited by resources | Demands high resources | Limited by resources | Limited by resources | *Limited* | High, continuous retraining |

our system is benchmarked against academic studies, future work will include empirical comparison with state-of-the-art APIs such as Microsoft Azure Face API, OpenCV DNN modules, and Google's FaceNet library to validate performance under real-world constraints. Table 4 presents a comparative overview of the proposed face recognition attendance system alongside five academic benchmark systems, evaluated across key features including detection algorithm, recognition algorithm, accuracy, verification method, data security, and scalability.

## 4.5. Suggestions and future work

One major limitation of this study is the lack of demographic diversity among participants, as all users were academic staff from a single institution in Thailand. This uniform sample may limit the generalizability of the results and could introduce unintended algorithmic bias. Future research should include a broader range of participants across different ages, ethnicities, and facial features to improve the system's fairness and robustness.

The system was tested with 58 users during the trial and processed around 8700 attendance records in a controlled academic setting. While the system performed accurately, the limited scope makes predicting performance in more complex or larger-scale environments difficult. Future deployments will involve thousands of users in bigger institutions, requiring enhanced computational capabilities. Upgrades like GPU acceleration, parallel processing, and better memory management will be key to maintaining low latency and high processing speed.

Existing literature has consistently shown that both the accuracy and reliability of facial recognition systems improve with the inclusion of larger and more heterogeneous datasets [9, 14]. Further refinement of the face detection and recognition algorithms is planned to address the system's observed performance decline in scenarios involving crowded frames, where multiple individuals appear simultaneously. Future work will investigate the application of more advanced neural network architectures, such as CNNs and recurrent neural networks (RNNs), to better manage real-time identification in densely populated settings [16].

While only a small number of misclassification errors were recorded during the evaluation, four in total, these were associated with substantial changes in users' facial appearance. All cases were successfully resolved through retraining of the model. Moving forward, the system will include mechanisms that enable users to appeal misclassifications. Moreover, the integration of XAI features, such as visual embedding projection tools, will aim to improve transparency and foster greater user trust.

We will also focus on hardware enhancements to improve the system. We will look into using higher-resolution cameras and more powerful processors to keep things running smoothly, even in varied lighting and environments. At the same time, we are committed to stringent data governance. This means implementing stronger encryption, anonymization techniques, and secure communication channels to fully comply with Thailand's PDPA [29] and other relevant international data protection laws. In addition, we are considering adding modalities like voice recognition or fingerprint scanning. Incorporating multi-modal biometric systems could boost security, improve authentication reliability, and make things even more convenient for users [23]. These planned enhancements are crucial for the system's ongoing development, ensuring it remains scalable and effective in the ever-changing landscape of educational institutions.

# 5. Conclusion

This research showcases the real-world applicability and conceptual importance of implementing a deep learning-based facial and gesture recognition system for attendance management at the PKRU Demonstration School. By overcoming the key limitations of traditional attendance methods, the system offers improved accuracy, quick processing, and a user-friendly interface, making it a cost-effective and efficient alternative. Using the MTCNN algorithm for facial detection alongside FaceNet for feature extraction highlights the strength and reliability of these deep learning frameworks. Moreover, the NuSVC classifier performed exceptionally well, achieving perfect accuracy in controlled testing.

Acknowledging that such high performance might not be fully replicable in more dynamic, real-world environments is essential. Nevertheless, these findings provide a strong foundation for future studies and practical applications, especially regarding the wider use of AI-based attendance solutions in educational and professional settings. Future improvements should focus on model adaptability, addressing data privacy concerns, and ensuring scalability to various environments. By refining and building upon the existing system, future implementations have the potential to significantly advance educational technology and the practical deployment of artificial intelligence.

# Recommendations

While the results of this study are promising, several limitations should be acknowledged. A key constraint was the relatively small dataset (only 290 facial images from 58 participants), which may not represent the full diversity found in broader populations. To improve the model's reliability and applicability, future research should draw from larger, more diverse datasets that capture differences in age, ethnicity, and facial features [9]. Another issue was a drop in recognition accuracy when more than three individuals were in the camera's view simultaneously. This highlights potential weaknesses in the current face detection algorithms when dealing with crowded or complex scenes. Although the system was designed for low-cost hardware, its performance could likely be enhanced with stronger computational resources. Additionally, while continuous model training could improve long-term accuracy, it introduces challenges around data management and privacy. Expanding the dataset over time must be done carefully, in line with Thailand's PDPA. Ensuring strong encryption and adequate anonymization will be critical to protect users' personal information [29].

## Acknowledgement

The authors would like to express our sincere gratitude to all the executives and teachers of PKRU Demonstration School for their continuous support and coordination from the system's development phase through its successful implementation at the school.

## Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

The data that support the findings of this study are openly available in GitHub at https://github.com/Nasith/academic-attendance-pkru, reference number [30].

## Author Contribution Statement

**Tanagrit Chansaeng:** Software, Validation, Formal analysis, Resources, Visualization. **Pita Jarupunphol:** Conceptualization, Methodology, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision.

## References

[1] Martinez, A. M. (2009). Face recognition, overview. In S. Z. Li, & A. Jain (Eds.), *Encyclopedia of Biometrics* (pp. 355–359). Springer Boston. https://doi.org/10.1007/978-0-387-73003-5_84

[2] Alhanaee, K., Alhammadi, M., Almenhali, N., & Shatnawi, M. (2021). Face recognition smart attendance system using deep transfer learning. *Procedia Computer Science*, *192*, 4093–4102. https://doi.org/10.1016/j.procs.2021.09.184

[3] Rao, A. (2022). AttenFace: A real-time attendance system using face recognition. In *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*, 1–5. https://ieeexplore.ieee.org/document/9998001

[4] Ashritha, K., Yuktha, K., & Bhukya, S. (2022). Automated attendance system using face recognition. *International Journal for Research in Applied Science & Engineering Technology*, *10*(6), 2096–2099. https://doi.org/10.22214/ijraset.2022.44212

[5] Kar, N., Debbarma, M. K., Saha, A., & Pal, D. R. (2012). Study of implementing automated attendance system using face recognition technique. *International Journal of Computer and Communication Engineering*, *1*(2), 100–103. https://doi.org/10.7763/IJCCE.2012.V1.28

[6] Jadhav, A., Jadhav, A., Ladhe, T., & Yeolekar, K. (2017). Automated attendance system using face recognition. *International Research Journal of Engineering and Technology (IRJET)*, *4*(1), 1467–1471.

[7] Yao, Y. (2024). Research on facial recognition system based on deep learning. *Applied and Computational Engineering*, *34*, 232–236. https://doi.org/10.54254/2755-2721/34/20230332

[8] Chen, Z., Wang, Y., & Zhang, S. (2023). Researches advanced in finance based on generative adversarial networks. *Highlights in Science, Engineering and Technology*, *39,* 452–464. https://doi.org/10.54097/hset.v39i.6568

[9] Asha, N., Syed Fiaz, A. S., Jayashree, J., Vijayashree, J. and Indumathi, J. (2022). Principal component analysis on face recognition using artificial firefly swarm optimization algorithm. *Advances in Engineering Software*, *174*. https://doi.org/10.1016/j.advengsoft.2022.103296

[10] Prasanna, G. S., Pavani, K., & Singh, M. K. (2022). Spliced images detection by using Viola-Jones algorithms method. *Materials Today: Proceedings*, *51*, 924–927. https://doi.org/10.1016/j.matpr.2021.06.300

[11] Yin, H., Chen, Y., Xiong, J., Xia, R., Xie, J. and Yang, K. (2021). An improved local binary pattern method for pollen image classification and recognition. *Computers Electrical Engineering*, *90*. https://doi.org/10.1016/j.compeleceng.2021.106983

[12] Agarwal, R., Jain, R., Regunathan, R. and PavanKumar, C. S. (2019). Automatic attendance system using face recognition technique. In A. J. Kulkarni, S. C. Satapathy, T. Kang, & A. H. Kashan (Eds.), *Proceedings of the 2nd International Conference on Data Engineering and Communication Technology* (pp. 525–533). Springer Singapore. https://doi.org/10.1007/978-981-13-1610-4_53

[13] Bah, S. M. and Ming, F. (2020). An improved face recognition algorithm and its application in attendance management system. *Array*, *5*. https://doi.org/10.1016/j.array.2019.100014

[14] Huang, S. and Luo, H. (2020), Attendance system based on dynamic face recognition. In *2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*, 368–371. https://ieeexplore.ieee.org/document/9258833

[15] Kumar, K. K., Kasiviswanadham, Y., Indira, D. V. S. N. V., Priyanka Palesetti, P. and Bhargavi, C.V. (2023). Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN). *Materials Today: Proceedings*, *80*, 2406–2410. https://doi.org/10.1016/j.matpr.2021.06.373

[16] Liu, Y. and Qu, Y. (2024). Construction of a smart face recognition model for university libraries based on FaceNet-mmar algorithm. *PLOS ONE*, *19*(1).

[17] Chen, B., Qi, X., Zhou, Y., Yang, G., Zheng, Y. and Xiao, B. (2020). Image splicing localization using residual image and residual-based fully convolutional network. *Journal of Visual Communication and Image Representation*, *73*. https://doi.org/10.1016/j.jvcir.2020.102967

[18] Revathy, G., Raj, K. B., Kumar, A., Adibatti, S., Dahiya, P., & Latha, T. M. (2022). Investigation of E-voting system using face recognition using convolutional neural network (CNN). *Theoretical Computer Science*, *925*, 61–67. https://doi.org/10.1016/j.tcs.2022.05.005

[19] Varsha, M. and Chitra Nair, S. (2022). Automatic attendance management system using face detection and face recognition. In P. Nayak, S. Pal and S. L. Peng (Eds.), *IoT and Analytics for*

*Sensor Networks* (pp. 97–106), Springer Singapore. https://doi.org/10.1007/978-981-16-2919-89
[20] Daube, C., Xu, T., Zhan, J., Webb, A., Ince, R. A., Garrod, O. G., & Schyns, P. G. (2021). Grounding deep neural network predictions of human categorization behavior in understandable functional features: The case of face identity. *Patterns*, *2*(10). https://doi.org/10.1016/j.patter.2021.100348
[21] Saleem, S., Shiney, J., Shan, B. P., & Mishra, V. K. (2023). Face recognition using facial features. *Materials Today: Proceedings*, *80*, 3857–3862. https://doi.org/10.1016/j.matpr.2021.07.402
[22] Talekar, S. and Parab, S. (2023). Face recognition-based attendance system for education institute. *International Journal of Advanced Research in Science, Communication and Technology*, *3*(1), 551–562. https://doi.org/10.48175/ijarsct-12083
[23] Sorvisto, D. (2023). Introducing MLOps. In D. Sorvisto (Ed.), *MLOps lifecycle toolkit* (pp. 1–34). Apress. https://doi.org/10.1007/978-1-4842-9642-4_1
[24] Watson, H. J. & Larson, D. (2024). MLOps: From a cottage industry to a factory approach. *International Journal of Business Intelligence Research*, *15*(1), 1–22. https://doi.org/10.4018/IJBIR.358916
[25] Shankar, A., Manoharan, H., Khadidos, A. O., Khadidos, A. O., Selvarajan, S., & Goyal, S. (2025). Transparency and privacy measures of biometric patterns for data processing with synthetic data using explainable artificial intelligence. *Image and Vision Computing*, *154*. https://doi.org/10.1016/j.imavis.2025.105429
[26] Khadidos, A. O., Manoharan, H., Khadidos, A. O., Selvarajan, S., & Singh, S. (2025). Synthetic healthcare data utility with biometric pattern recognition using adversarial networks. *Scientific Reports*, *15*. https://doi.org/10.1038/s41598-025-94572-3
[27] Selvarajan, S., Manoharan, H., Alsowail, R. A., Shankar, A., Pandiaraj, S., Maple, C., & Viriyasitavat, W. (2024). Cognitive intelligent decisions for big data and cloud computing in industrial applications using trifold algorithms. *Cognitive Computation*, *16*(6), 2967–2981. https://doi.org/10.1007/s12559-024-10317-w
[28] Idoko, I. P., Igbede, M. A., Manuel, H. N. N., Adeoye, T. O., Akpa, F. A., & Ukaegbu, C. (2024). Big data and AI in employment: The dual challenge of workforce replacement and protecting customer privacy in biometric data usage. *Global Journal of Engineering and Technology Advances*, *19*(2), 89–106. https://doi.org/10.30574/gjeta.2024.19.2.0080
[29] Bumpenboon, T. (2020). Thailand's personal data protection act: An understanding from the perspectives of the European privacy law. *Thammasat Review of Economic and Social Policy*, *6*(1), 50–82.
[30] Laosen, N. (2025). Academic-attendance-PKRU [Data set]. GitHub. https://github.com/Nasith/academic-attendance-pkru
[31] Elmazi, K., & Elmazi, D. (2024). Evaluating support vector machine and logistic regression: A machine learning-based statistical study for breast cancer diagnosis. In *2024 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, 1–6. https://doi.org/10.1109/CoNTESA64738.2024.10891283
[32] Rodríguez-Pérez, R. and Bajorath, J. (2022). Evolution of support vector machine and regression modeling in chemoinformatics and drug discovery. *Journal of Computer-Aided Molecular Design*, *36*(5), 355–362. https://doi.org/10.1007/s10822-022-00442-9
[33] Rothacher, Y. and Strobl, C. (2023). Identifying informative predictor variables with random forests. *Journal of Educational and Behavioral Statistics*, *49*(4), 595–629. https://doi.org/10.3102/10769986231193327
[34] Galal, A., Talal, M., & Moustafa, A. (2022). Applications of machine learning in metabolomics: Disease modeling and classification. *Frontiers in Genetics*, *13*. https://doi.org/10.3389/fgene.2022.1017340
[35] Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *13*, 6308–6325. https://doi.org/10.1109/JSTARS.2020.3026724
[36] Sudianto, S., Masheli, J. A. A., Nugroho, N., Rumpoko, R. W. A., & Akhmad, Z. (2022). Comparison of support vector machines and k-nearest neighbor algorithm analysis of spam comments on YouTube Covid Omicron. *Jurnal Teknik Informatika*, *15*(2), 110–118. https://doi.org/10.15408/jti.v15i2.24996
[37] Baka, P., & Schatten, J. (2020). *SSL/TLS under lock and key: A guide to understanding SSL/TLS cryptography*. Australia: Keyko.

**How to Cite:** Chansaeng, T., & Jarupunphol, P. (2025). Facial and Gesture Recognition-based Deep Learning Model for Academic Time Attendance at PKRU Demonstration School. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonview52024469