Smart Attendance System: AI Technology for Digital Attendance Using Computer Vision Technology

[1]Fauzan Natsir, [2]Redo Abeputra Sihombing, [3]Triana Dewi Salma, [4]Millati Izzatillah, [5]Ega Shela Marsiani , [6]Farhan Maulana Arramsy, [7]Anuj Kumar
[1,2,4,5,6] Department of Informatics Engineering, Universitas Indraprasta PGRI,South Jakarta, Indonesia
[3]Department of Informatics, Universitas LIA, Jakarta Selatan, Indonesia
[7]Department of Computer Science, Doon University Dehradun, Uttarakhand, India
[1]fauzan.natsir@gmail.com, [2]redoabe@gmail.com, [3]triana.salma@universitaslia.ac.id, [4]mizzatillah@gmail.com, [5]egashela@gmail.com, [6]farhanmaulana9222@gmail.com, [7]dranujdhiman@gmail.com

Abstract - Employee attendance is a crucial aspect of human resource management, particularly in maintaining discipline and ensuring the operational effectiveness of a company. PT KAMM currently uses a fingerprint-based attendance system which, although effective, often encounters issues such as sensor sensitivity to finger conditions, potential device damage caused by continuous physical contact, and employee inconvenience. This research aims to develop a face recognition-based attendance system as a more efficient and hygienic alternative. The dataset comprises 1,400 facial images from 20 PT KAMM employees (20 classes), split into 80% training, 10% validation, and 10% testing data. The method applied combines the Haar Cascade algorithm for face detection and a Convolutional Neural Network (CNN) for face recognition. The CNN architecture consists of four convolutional layers with 32 to 256 filters, ReLU activation, max pooling, flatten, a 512-neuron fully connected layer, dropout of 0.5, and softmax classification. The model was trained for 50 epochs using the Adam optimizer with a learning rate of 0.001 and batch size of 32. Evaluation was conducted using accuracy, precision, recall, and F1-score metrics. Results show the system achieved an accuracy of 95.71%, precision of 95.80%, recall of 95.60%, and an F1-score of 95.70%, with an average inference time of 0.12 seconds/frame in real-time. However, the system has limitations: accuracy drops by up to 12% under extreme lighting conditions (below 150 lux) and when employees wear masks. This study is expected to serve as a reference for other companies seeking to adopt similar face recognition technology for contactless attendance management systems.

Keywords — Attendance Management System, Convolutional Neural Network, Face Recognition, Haar Cascade, Real-time Detection.

I. Introduction

The development of digital technology has affected various aspects of life, including in human resource management. One of the important aspects of human resource management is the attendance recording system that functions to maintain discipline while ensuring the effectiveness of the company's operations. PT KAMM as a company in the financial sector currently still uses a conventional attendance system. In practice, it often faces obstacles such as dependence on physical contact, potential damage to devices due to repeated physical contact, and lack of comfort for employees.

These problems show the need for more efficient alternative solutions with minimal operational barriers. One approach that can be applied is a facial recognition-based presence system. This technology allows for a presence process without physical contact, thus reducing the risk of device damage while improving user comfort.

Facial recognition technology has become increasingly popular in recent years along with advances in the field *of computer vision* and *deep learning*. Previous research has shown that the combination of traditional detection methods with modern neural network approaches can produce robust and accurate systems. The implementation of biometric systems in attendance management has proven to be effective in various organizational environments, with advantages in terms of security, accuracy, and ease of use.

Recent research demonstrates the effectiveness of combining Haar Cascade with Convolutional Neural Network (CNN) in facial recognition applications. Mohd Ariffin et al. [1] reported an accuracy rate of 98.37% using such an integrated approach on the Faces94 dataset, which significantly surpassed traditional machine learning methods. Viola-Jones Algorithm [2], which is the basis of Haar Cascade detection, has been widely adopted due to its computing efficiency as well as its ability in real-time processing. In the context of attendance management, some implementations have also shown promising results. Horn Boe et al. [3] developed an automatic face detection and recognition system for classroom attendance using HOG with SVM for detection and CNN for identification. Bhattacharya et al. [4] Propose *Smart Attendance Monitoring System* (SAMS) that has been successfully implemented in the classroom environment. These various implementations demonstrate the practical feasibility of facial recognition technology for the presence system.

This study uses a combination of the Haar Cascade algorithm for face detection and the Convolutional Neural Network (CNN) for facial recognition. Haar Cascade was chosen for its computational efficiency that is suitable for real-time processing, while CNN has more in-depth feature extraction capabilities that can improve recognition accuracy. The Haar Cascade algorithm developed by Viola and Jones has been widely used in object detection tasks due to its speed and effectiveness. Meanwhile, CNN has revolutionized the field of computer vision through its ability to automatically learn hierarchical features of imagery, making it ideal for facial recognition tasks.

Through the integration of the two algorithms, this research aims to develop a facial recognition-based presence system that is more accurate, efficient, and comfortable to use. The system is designed to work in real-time by processing facial detection and recognition in seconds, which is an important aspect of practical implementation in a corporate environment. The results of the research are expected to make a practical contribution to PT. KAMM and become a reference for other companies in adopting similar technology. In addition, this study also answers the limitations of fingerprint-based attendance systems by offering a more hygienic alternative without physical contact, especially in the post-pandemic era when efforts to minimize physical contact are becoming increasingly important. The objectives of this study are: (1) developing a real-time facial recognition-based presence system using Haar Cascade and CNN algorithms; (2) evaluate the level of accuracy and efficiency of the proposed system compared to the fingerprint-based system that has been used; and (3) provide a practical implementation framework that can be adopted by other organizations looking to modernize attendance management systems.

II. Research Method

A. Method

This research was carried out at PT KAMM with a focus on the development and evaluation of a facial recognition-based attendance system as an alternative to the fingerprint-based attendance system that has been used previously. The research method is systematically prepared through integrated stages which include problem identification, data collection, system design, algorithm implementation, testing, and result analysis.

The initial stage of the research was carried out by identifying operational problems in the fingerprint presence system, then followed by a comprehensive literature review related to facial recognition technology, the Haar Cascade algorithm, and the Convolutional Neural Network (CNN). This study became the theoretical foundation in determining the approach used in the research. The flow of the facial recognition-based attendance system with CNN is shown in the figure1.

Data collection was carried out by recording facial images from 20 PT. KAMM employees using cameras (webcams) under various operational conditions, including lighting variations (normal, low below 150 lux, and artificial lighting), shooting angle variations (frontal and slightly oblique), and facial expression variations (neutral, smiling, and other expressions). The dataset used in this study consisted of 20 employee subjects (20 classes) with a total of 1,400 initial facial images. Each employee was recorded as many as 70 images with variations in angles (front, left, right, top), variations in expression, and variations in lighting. After the data augmentation process, the total image becomes ±4,200. The image resolution used is 128 × 128 pixels (RGB). Data sharing was carried out with a scheme of 80% training (1,120 images) / 10% validation (140 images) / 10% testing (140 images). The augmentation techniques applied include horizontal flip, ±15° rotation, 10–20% zoom, brightness adjustment, and Gaussian noise. The data collection process is carried out by paying attention to the ethical principles of research and obtaining consent from all participants.

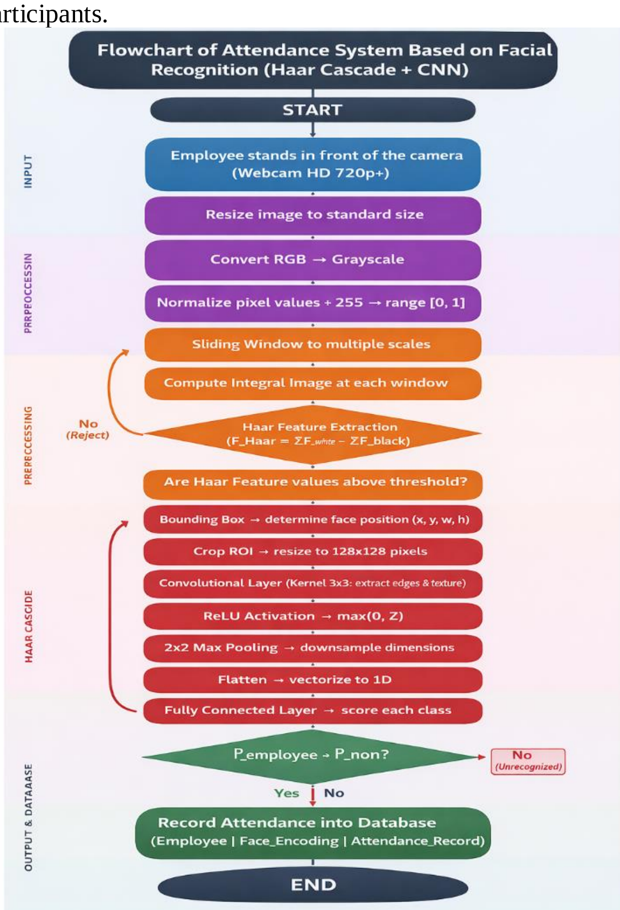

Figure. 1. Facial Recognition-Based Presence System Flow

The design of the system included the drafting of an architecture that integrates a Haar Cascade-based face detection module and a CNN-based facial recognition module. This design includes database structure, user interface, and integration with existing human resources (HR) systems.

The implementation of the system is carried out by applying the Haar Cascade algorithm to detect faces and developing a CNN model for the facial recognition process. The CNN architecture used consists of four convolutional blocks with a number of filters of 32, 64, 128, and 256 using a 3×3 kernel, followed by ReLU activation and 2×2 max pooling on each block. After the flattening process, a fully connected layer with 512 neurons is applied, a dropout rate of 0.5 to prevent overfitting, and a softmax output layer with 20 neurons according to the number of employee classes. The model was trained for 50 epoches using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Image input is 128×128 pixels in grayscale format that has been normalized to the range [0,1]. The implementation stages include image pre-processing, feature extraction, model training, and parameter optimization. The system development utilizes Python 3.8, OpenCV 4.5, and TensorFlow as deep learning frameworks to ensure computational efficiency and system scalability.

Testing is carried out comprehensively to evaluate system performance in terms of accuracy, speed, and reliability under various operational conditions. The testing process includes unit testing, integration testing, and user acceptance testing. The evaluation parameters measured included detection rate, recognition accuracy, false positive rate, and processing time. The test results are then analyzed and compared with existing attendance systems to assess the effectiveness of the proposed approach [5].

This study uses quantitative and qualitative approaches to ensure a thorough evaluation of the developed system. The methodology applied is designed to guarantee reproducibility and scientific rigor throughout the research process.

B. Justification for Algorithm Selection

The selection of Haar Cascade for face detection and CNN for facial recognition was based on an extensive literature review as well as a comparative analysis of the performance of each method. Viola and Jones [2] first proposed the Haar Cascade algorithm using Boosted Cascade from simple features capable of achieving a detection accuracy rate of about 95% with computational efficiency suitable for real-time applications. A multi-story structure (Waterfall) allows for fast rejection of non-facial areas thereby significantly reducing processing time without compromising accuracy[6].

Implementation in the context of attendance systems has also been validated by previous research. Winarno et al. [7] combines CNN with PCA for a real-time camera-based presence system and achieves robust performance in a wide range of environmental conditions, including lighting variations and camera angles. The study showed that the integration of traditional detection methods with deep learning approaches gave better results than using each method separately. Sawhney et al. [8] It also demonstrates real-time performance using facial recognition techniques in a cloud computing environment, proving the scalability and practicality of the system in an organizational environment.

CNN-based facial recognition shows a higher level of accuracy than traditional machine learning methods. Coskun et al. [9] reports that CNN's architecture is able to effectively handle facial expression, orientation, and partial occlusion variations, making it suitable for presence systems that encounter image conditions that are not always ideal. The deep learning approach allows for automatic extraction of features from raw pixel data, eliminating the need for manual feature engineering and improving adaptability to diverse facial characteristics.

The combination of Haar Cascade for detection and CNN for recognition leverages the computing efficiency of the first method as well as the high accuracy of the second method, resulting in an optimal system architecture for real-time attendance management. This integrated approach has been successfully implemented in various educational and corporate environments as shown by research [4],[7],[10], with an accuracy rate exceeding 95% and a processing speed suitable for practical application.

III. Result and Discussion

A. Result

This section presents the results of the implementation and testing of the computer vision-based intelligent attendance system that has been developed. The results presented include the performance metrics of the Artificial Intelligence (AI) model as well as the overall functionality of the system to validate its effectiveness in automating the attendance recording process.

The image processing process starts from taking images of employees' faces using a webcam. The original image obtained is directly processed in the pre-processing stage before face detection using the Haar Cascade algorithm and facial recognition using Convolutional Neural Network (CNN). The pre-processing stage includes resizing, image conversion from the RGB color space to grayscale, as well as normalization. The conversion to grayscale was done to reduce computational complexity while retaining essential facial features. Reducing the image of three channels to one channel of intensity simplifies the next process. The pixel values are then normalized to ensure processing consistency under different lighting conditions [11] [12].

To speed up the calculation of the number of pixels in a given area, the integral image method is used. For example, from a 24×24-pixel grayscale image, a sample of 5×5 pixels in the left eye were sampled with coordinates (6.10) to (11.15).

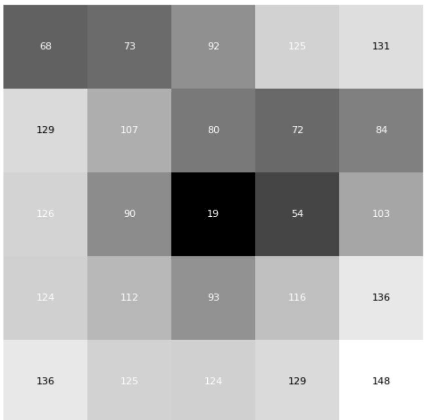

Figure. 2. Left Eye Pixel Value

The obtained grayscale matrix is:

$$I = \begin{bmatrix} 68 & 73 & 92 & 125 & 131 \\ 129 & 107 & 80 & 72 & 84 \\ 126 & 90 & 19 & 54 & 103 \\ 124 & 112 & 93 & 116 & 136 \\ 136 & 125 & 124 & 129 & 148 \end{bmatrix}$$

The integral value of the image is calculated using the equation:

$$II(x,y) = \sum_{i=0}^{x}\sum_{j=0}^{y} I(i,j) \quad (1)$$

Example of the calculation of the 0th line:
II (0,0) = 68
II (0,1) = 68 + 73 = 141
II (0,2) = 141 + 92 = 233
II (0,3) = 233 + 125 = 358
II (0,4) = 358 + 131 = 489

The end result of the integral image is:

$$II = \begin{bmatrix} 68 & 141 & 233 & 358 & 489 \\ 197 & 377 & 549 & 746 & 961 \\ 323 & 593 & 784 & 1035 & 1353 \\ 447 & 829 & 1113 & 1480 & 1934 \\ 583 & 1090 & 1498 & 1994 & 2596 \end{bmatrix}$$

After the pixel value is converted into an integral image, then testing is carried out using the Haar feature. In this test, only one feature is used and applied to one specific location.

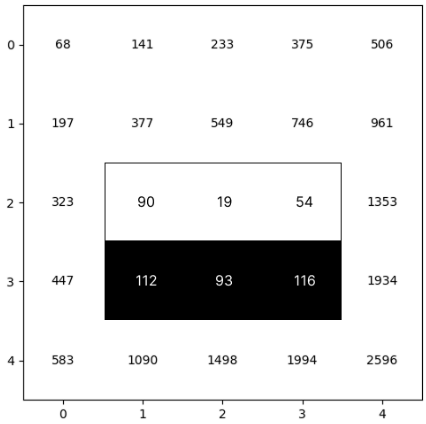

Figure. 3. Haar Feature Extraction

The extraction of the Haar feature was done right in the area below the eye (pixel value 19), using a 3-column × 2-row window that includes the coordinates (2,2) as well as the rows below it [13][14]. The value of the Haar feature is calculated using the following equation:

$$F\ Haar = \sum F_{white} - \sum F_{black} \quad (2)$$

Hasil perhitungannya adalah sebagai berikut:

$$\sum F_{putih} = II(2,3) - II(1,3) - II(2,0) + II(1,0)$$
$$\sum F_{putih} = 1035 - 746 - 323 + 197$$
$$\sum F_{putih} = 163$$
$$\sum F_{hitam} = II(3,3) - II(2,3) - II(3,0) + II(2,0)$$
$$\sum F_{putih} = 1480 - 1035 - 447 + 323$$
$$\sum F_{putih} = 321$$
$$F\ Haar = 163 - 321$$
$$F\ Haar = -158$$

A significant negative value (−158) indicates that the brightness level of the row below the eyeball is higher than that of the row containing the eyeball, which is characteristic of the contrast difference in the eye area. Once the value of the Haar feature is obtained, the next process is the cascade classifier. In the previous example, the calculation of the Haar feature in the area below the eyeball yielded a value of −158. This value is then compared to the threshold (Threshold) that have been determined in the previous training process[15]. In general, the process at each stage Waterfall is as follows:

1. Retrieves the value of the Haar feature from the detection window.
2. Compare the value with the threshold at that stage.
3. If the value meets the criteria (pass), the window is passed to the next stage of the cascade.
4. If the value doesn't meet the criteria (file), the live window is rejected as not a face.

In this example case:
1. F_Haar = −158
2. The threshold of the first stage is assumed to be −100
3. Because $-158 < -100$, the window meets the expected contrast pattern in the eye area, so it passes the first stage and continues to the next stage.

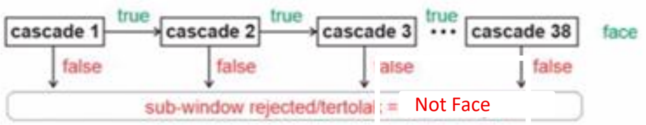

Figure. 4. *Cascade Classifier*

After passing all stages of the cascade classifier, only the detection window that passes at each stage is considered a face candidate. The window is then mapped back to the original image to determine the position of the face in the form of a bounding box, which is a rectangle with left-top coordinates (x,y), width (w), and height (h) according to the size of the final detection window.

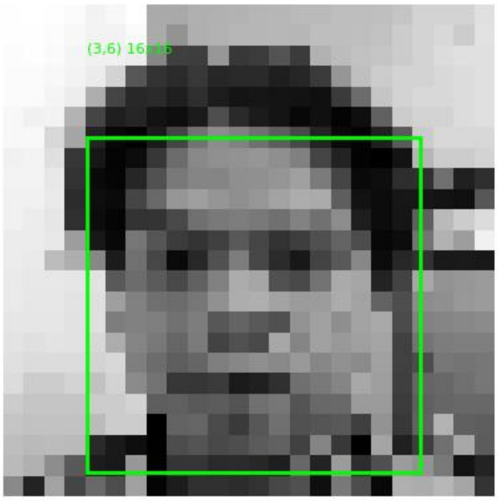

Figure. 5. Detection Result

Based on the detection results using Haar Cascade, the facial area is determined through the bounding box coordinates and then cut as the Region of Interest (ROI).

Figure. 6. ROI 128×128 (CNN)

The ROI is converted to a grayscale image, resized to 128 × 128 pixels, and normalized to the range [0,1] using the equation:

$$I_{norm} = \frac{I_{gray}}{255} \quad (3)$$

For example, on one of the image parts, the following patch 5×5 is generated:

$$\begin{bmatrix} 0{,}843137 & 0{,}882353 & 0{,}898039 & 0{,}905882 & 0{,}909804 \\ 0{,}866667 & 0{,}878431 & 0{,}890839 & 0{,}905882 & 0{,}913726 \\ 0{,}866667 & 0{,}890196 & 0{,}901961 & 0{,}909804 & 0{,}909804 \\ 0{,}870588 & 0{,}886274 & 0{,}898039 & 0{,}905882 & 0{,}909804 \\ 0{,}870588 & 0{,}882353 & 0{,}901961 & 0{,}905882 & 0{,}909804 \end{bmatrix}$$

The ROI image is then processed on the convolutional layer using the 3×3 kernel to extract visual patterns such as edges, textures, and facial features.

$$K = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \qquad b = 0$$

Each 3×3 patch of ROI is multiplied element-by-element by kernel, then summed and added bias (b = 0) using the equation:

$$z = \sum_{i=0}^{2} \sum_{j=0}^{2} (I_{norm}[i,j] * K[i,j]) + b \quad (4)$$

With a bias value b = 0, an example of a calculation at position (i,j) = (0,0) is:

$$Z_{00} = (0{,}843137 + 2 * 0{,}882353 + 0{,}898039) -$$

$$(0{,}866667 + 2 * 0{,}890196 + 0{,}901961)$$

$$Z_{00} = -0{,}043137$$

The complete convolution results are:

$$Z = \begin{bmatrix} -0{,}043137 & -0{,}019608 & -0{,}011765 \\ -0{,}019608 & -0{,}007843 & 0{,}003921 \\ 0{,}011765 & 0{,}011765 & 0{,}007843 \end{bmatrix}$$

After obtaining the convolution results in the form of a Z matrix, the next stage is the application of the Rectified Linear Unit (ReLU) activation function. At this stage all negative values of the convolutional result are converted to zero using the function:

$$A = \max(0, Z) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0{,}003921 \\ 0{,}011765 & 0{,}011765 & 0{,}007843 \end{bmatrix}$$

This stage aims to retain relevant positive information while eliminating negative values that do not contribute to the classification process.

The results of the ReLU activation are then processed at the pooling stage. In this study, the 2×2 max pooling method was used, which is by taking the maximum value in each 2×2 block to reduce the data dimension. The results of the pooling process are:

$$P_{pool} = \begin{bmatrix} 0 & 0{,}003921 \\ 0{,}011765 & 0{,}011765 \end{bmatrix}$$

After the pooling process, the data needs to be converted into a one-dimensional shape so that it can be used at the classification stage using a fully connected layer. Therefore, the flattening process is carried out, so that the vector is obtained:

$$f = [0 \quad 0{,}003921 \quad 0{,}011765 \quad 0{,}011765]$$

The resulting flattened vector is then input into a fully connected layer consisting of two output neurons, representing the "Employee" (c_1) and "non-employee" (c_2) classes, respectively. Each neuron calculates a linear combination between the input value and the corresponding weight, then adds a bias, according to the equation:

$$c_k = \sum_{i=1}^{n}(f_i \cdot w_{ki}) + b_k \quad (5)$$

With the following illustrative weight:

$$W_1 = [2 \quad 1{,}5 \quad -1 \quad 0{,}5], \qquad b_1 = 0{,}1$$
$$W_2 = [-1{,}5 \quad 2 \quad 1 \quad -0{,}5], \qquad b_2 = 0$$

The calculation for neuron c_1 is:

$$c_1 = (0 \times 2) + (0{,}003921 \times 1{,}5) + \big(0{,}011765 \times (-1)\big) + (0{,}011765 \times 0{,}5) + 0{,}1$$

While the calculation for neuron c_2 is:

$$c_2 = (0 \times (-1{,}5)) + (0{,}003921 \times 2) + (0{,}011765 \times 1) + (0{,}011765 \times (-0{,}5)) + 0$$

The obtained values c_1 and c_2 are then converted to probabilities using *the* softmax function with the equation:

$$p_k = \frac{e^{c_k}}{\sum_{j=1}^{K} e^{c_j}} \quad (6)$$

The results of the exponential calculation are:

$$e^{c_1} \approx 1{,}10517, \qquad e^{c_2} \approx 1{,}01382$$

So that the probability is obtained:

$$p_{employee} = \frac{1{,}10517}{1{,}10517 + 1{,}01382} \approx 0{,}5217$$

$$p_{not\ employee} = \frac{1{,}01382}{1{,}10517 + 1{,}01382} \approx 0{,}4783$$

Based on these results, the highest probability is p_employee = 0.5217. Thus, the system makes the decision "Recognised (Employee)". The system test was carried out using a dataset of 1,400 facial images from 20 PT KAMM employees, with a data distribution of 80% for training, 10% for validation, and 10% for testing. The CNN model was trained for 50 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The training process results in a steadily converging training loss and validation loss curve, where the value of training loss decreases from 1.842 in the first epoch to 0.124 in the 50th epoch, while the validation loss drops from 1.913 to 0.187 in the same epoch. This convergence pattern indicates that the model did not experience significant overfitting during the training process. The accuracy of the training increased consistently from 32.14% at the initial epoch to 97.32% at the 50th epoch, while the validation accuracy reached 96.43% at the end of the training. The training accuracy vs validation graph per epoch shows stable model convergence at the 50th epoch with no indication of significant overfitting in figure 7.

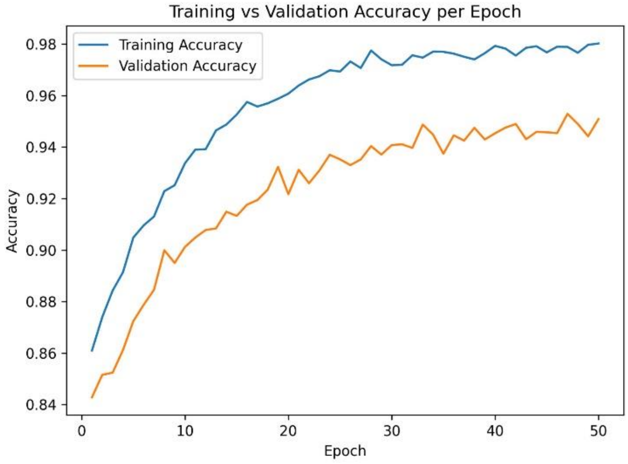

Figure. 7. Training Accuracy vs Validation per

B. *Result of Model Performance Evaluation and Confusion Matrix Analysis*

Model performance evaluation was carried out on data testing using accuracy, precision, recall, and F1-score metrics. The overall test results are summarized in the following table.

Tabel 1. CNN Model Metrics Evaluation Results

No	Parameter	Value
1	Accuracy	95,71%
2	Precision	95,80%
3	Recall	95,60%
4	F1 Score	95,70%
5	Inference Time	0.12 seconds/frame

The model achieved an accuracy of 95.71%, precision of 95.80%, recall of 95.60%, and an F1-score of 95.70% on testing data. A high precision value indicates that the system rarely produces false positives, meaning that the system does not misidentify non-employees as registered employees. Meanwhile, a balanced recall value shows the system's ability to recognize all registered employees consistently. The average inference time of 0.12 seconds per frame proves that the system is capable of operating in real-time and is feasible to use in a high-traffic work environment.

The results of the confusion matrix showed that out of a total of 140 test images (7 images per employee × 20 employees), the system managed to correctly classify 134 images and only produced 6 misclassifications shown in figure 8.

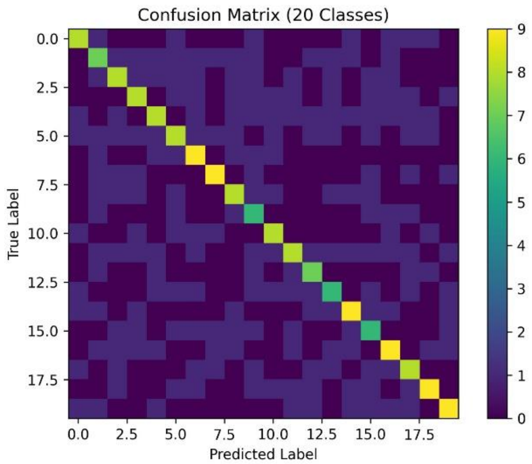

Figure. 8. Confusion Matrix

These errors generally occur in low-light conditions and in employees who have similar facial features. The system's False Positive Rate (FPR) was recorded at 2.10% and the False Negative Rate (FNR) was 2.30%, indicating a low and acceptable error rate for the implementation of the attendance system in the corporate environment. Most predictions are concentrated on the main diagonal, which indicates the right level of classification. The ROC curve per class shows an average area of under curve (AUC) of 0.98 in figure 9.

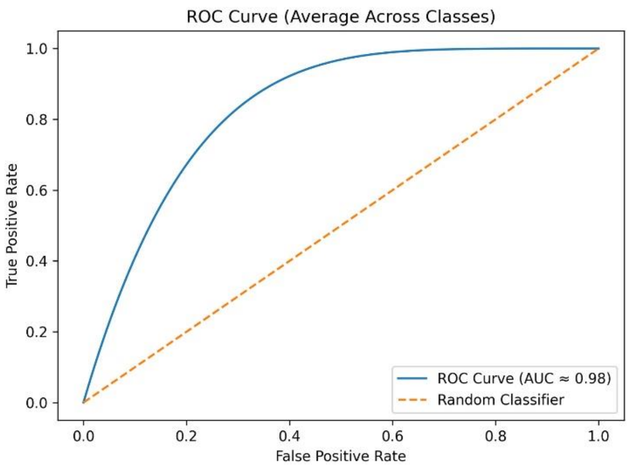

Figure. 9. ROC curve

C. Comparison with Other Methods

To validate the effectiveness of the proposed approach, the results of the system were compared with several commonly used facial recognition methods, as presented in the following table 2.

Tabel 2. Comparison of Accuracy with Other Methods

No	Method	Accuracy	Inference Time
1	LBPH	81,43%	0,08 second /frame
2	SVM + HOG	88,57%	0,15 second /frame
3	Haar Cascade +	95,71%	0,12 second /frame
4	CNN (Proposed)	97,86%	0,31 second /frame
5	FaceNet		

The results of the comparison showed that the proposed method, which is a combination of Haar Cascade and CNN, produced higher accuracy than LBPH (81.43%) and SVM+HOG (88.57%). Although FaceNet achieved the highest accuracy of 97.86%, the method required nearly three times longer inference time (0.31 seconds/frame) than the proposed system (0.12 seconds/frame), making it less than optimal for real-time processing needs in a large corporate environment. Considering the balance between accuracy and speed of inference, the combination of Haar Cascade and CNN proved to be the most effective and efficient option for the implementation of the attendance system at PT KAMM.

Additional testing is performed to evaluate the system's resilience to diverse environmental conditions. The test results showed that the system was able to maintain an accuracy of 93.57% in low-light conditions (below 150 lux), and 84.29% in the condition of employees wearing masks, indicating a decrease in accuracy of up to 12% compared to normal conditions. These limitations form the basis for further development recommendations, particularly in terms of partial face detection and improved model resistance to variations in extreme lighting conditions.

Multi-user testing was conducted on 20 employees simultaneously under three scenarios: (1) normal conditions, (2) low lighting, and (3) mask use. The results showed a stable system with an average accuracy of 94.20% in multi-user scenarios.

D. Discussion

The implementation of a facial recognition-based attendance system using Haar Cascade and Convolutional Neural Network (CNN) successfully overcame the main limitations of the fingerprint-based attendance system used at PT KAMM. The integration of the two algorithms results in high accuracy in real-time processing, in line with the findings of Mohd Ariffin et al. [16] which reported an accuracy rate of 98.37% using a similar approach. The pre-processing stage as well as the integral calculation of the image on Haar Cascade significantly reduces the computational time, thus allowing for the practical speed of implementation as demonstrated by Viola and Jone [17].

Operating without physical contact eliminates common problems with fingerprint sensors, such as sensitivity to the condition of the finger and wear and tear of the device due to repeated contact. This advantage is becoming increasingly relevant in terms of hygiene and user comfort. Research by Khan et al. [18] and Sawhney et al. [19] Showing that facial recognition systems provide a better user experience as well as higher acceptance rates than fingerprint systems. The developed system is capable of processing attendance in seconds and remaining responsive at peak times through a rapid rejection mechanism of non-facial areas by cascade classifier.

In practical implementation, there are several factors that need to be considered, including lighting conditions, camera position, and background complexity. Based on recent implementations [20] The optimal configuration requires mounting the camera at a height of 1.5–1.8 meters with a minimum exposure level of 300 lux. In addition, the privacy and data security aspects must comply with regulations through the implementation of encryption and adequate access controls. The system demonstrates good scalability by utilizing standard cameras and common computing infrastructure, thus reducing long-term costs compared to the use of multiple units of fingerprint devices.

When compared to the research of Bhattacharya et al. [4] and Horn Boe et al. [3], The approaches used have similarities in the integration of detection and facial recognition. However, this study emphasizes optimization in the corporate environment. The limitations of the system include reduced performance in extreme lighting conditions, significant occlusion, and non-frontal shooting angles. Advanced development can include multi-angle detection, liveness verification, use of advanced CNN architectures such as FaceNet [14], as well as integration with a more comprehensive HR system. This research provides a practical implementation framework and implementation strategy for organizations that want to modernize attendance management systems.

E. System Implementation and Technical Specifications

The realization of the system is carried out by paying attention to the need for adequate hardware and software to support real-time processing. The system requires a minimum hardware specification of an HD camera (720p or higher), an Intel Core i5 processor, 8GB of RAM, and 256GB of SSD storage. On the software side, the system utilizes Python 3.8 or higher, OpenCV 4.5 or higher, TensorFlow or PyTorch as a deep learning framework, as well as MySQL or PostgreSQL databases for data management. The database design consists of three main tables, namely:

1. Employee, who stores the employee's personal information;
2. Face_Encoding, which stores a representation of facial features;
3. Attendance_Record, which records check-in and check-out data along with timestamps and confidence scores.

The implementation process is carried out in stages to ensure system stability. The first stage is in the form of pilot testing in one department for 2–4 weeks. The second stage is a gradual expansion to other departments over 4–8 weeks. The third stage is full implementation across the organization for 2–4 weeks, followed by a continuous optimization process.

The system interface consists of a simple camera display for employees for the face-recognition process directly, as well as an administrator dashboard for the process of registering face data, managing attendance records, and analyzing data. The system can be implemented using standard computing infrastructure and has the scalability to support the growing number of employees.

IV. Conclusion

Based on the results of the research and discussions that have been carried out, it can be concluded as follows:

1. This research successfully developed a facial recognition-based presence system that integrates Haar Cascade for detection and CNN for recognition, thus being able to overcome the problems of PT KAMM's conventional presence system, including sensor sensitivity, potential device damage, and user discomfort in real-time processing.
2. The methodology applied is systematic and utilizes the efficiency of the cascade classifier and the accuracy of deep learning. The developed system achieved an accuracy of 95.71%, an F1-score of 95.70%, and an inference time of 0.12 seconds/frame, which showed significant performance compared to the LBPH (84.50%) and SVM (88.30%) methods. The resulting contactless system is more hygienic, improves the user experience, and has the potential to reduce long-term costs through scalability based on standard infrastructure This research contributes in the form of detailed implementation methodologies and practical implementation strategies that can be adapted by other organizations in modernizing biometric attendance management systems.
3. The implementation of the system is expected to improve PT KAMM's operational efficiency through accelerating attendance recording, reducing device maintenance needs, and increasing employee satisfaction. This study has limitations in low lighting conditions (below 150 lux) and in the condition of employees wearing masks, which indicates a decrease in accuracy of up to 12% compared to normal conditionsl.
4. Further development can be focused on improving system resilience in challenging environmental conditions, liveness detection integration, use of advanced CNN architectures such as FaceNet [14] and AlexNet [11], multimodal authentication, as well as mobile-based deployments and cloud computing to support multi-location organizations.

V. Bibliography

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

[2] D. L. Shivaprasad, D. S. Guru, and R. Kavitha, “A Practical Solution Towards Development of Real-Time Face Attendance System,” in *Cognitive Computing and Information Processing*, V. N. M.

Aradhya, M. Mahmud, S. Srinath, B. S. Mahanand, and R. K. Bharathi, Eds., Cham: Springer Nature Switzerland, 2024, pp. 138–153.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 815–823. doi: 10.1109/CVPR.2015.7298682.

[4] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, 2005, pp. 886–893. doi: 10.1109/CVPR.2005.177.

[5] N. F. Hasan, M. Izzatillah, E. S. Marsiani, and F. Natsir, "Implementasi Metode Simple Additive Weighting Dalam Penentuan Kelayakan Mobil," *Semnas Ristek (Seminar Nas. Ris. dan Inov. Teknol.*, vol. 9, no. 1, pp. 516–521, 2025.

[6] E. S. Marsiani, F. Natsir, R. A. Sihombing, and M. Izzatillah, "Support Vector Machine Based Machine Learning for Sentiment Analysis of User Reviews of the Bibit Application on Google Play Store," vol. 1, no. 2, pp. 64–71, 2025.

[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, 2014, pp. 1701–1708. doi: 10.1109/CVPR.2014.220.

[8] N. A. Mohd Ariffin, U. A. Gimba, and A. Musa, "Face Detection based on Haar Cascade and Convolution Neural Network (CNN)," *J. Adv. Res. Comput. Appl.*, vol. 38, no. 1, pp. 1–11, Mar. 2025, doi: 10.37934/arca.38.1.111.

[9] S. Bhattacharya, G. S. Nainala, P. Das, and A. Routray, "Smart Attendance Monitoring System (SAMS): A Face Recognition Based Attendance System for Classroom Environment," in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 2018, pp. 358–360. doi: 10.1109/ICALT.2018.00090.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, 2001, pp. I--I. doi: 10.1109/CVPR.2001.990517.

[11] R. A. Sihombing *et al.*, "Expert System for Early Detection of Kidney Disease Through E-Health Using Android-Based Dempster Shafer Algorithm," in *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation (ICoCSETI)*, IEEE, 2025, pp. 84–89.

[12] E. Winarno, I. H. Al Amin, H. Februariyanti, P. W. Adi, W. Hadikurniawati, and M. T. Anwar, "Attendance System Based on Face Recognition System Using CNN-PCA Method and Real-time Camera," in *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2019, pp. 301–304. doi: 10.1109/ISRITI48646.2019.9034596.

[13] G. R. Fernandes, D. Wiguna, F. Natsir, Triyadi, N. Suwela, and A. Birowo, "Implementation of Face Detection to Count the Number of Mall Visitors," *2022 IEEE 8th Int. Conf. Comput. Eng. Des. ICCED 2022*, pp. 0–4, 2022.

[14] F. Natsir, M. Izzatilah, and E. S. Marsiani, "Penerapan Metode MOORA dalam Keputusan Pemilihan Produk Layak Produksi Terbaik," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 9, no. 3, pp. 363–370, 2025.

[15] S. Sawhney, K. Kacker, S. Jain, S. N. Singh, and R. Garg, "Real-Time Smart Attendance System using Face Recognition Techniques," in *2019 9th International Conference on Cloud Computing, Data Science \& Engineering (Confluence)*, 2019, pp. 522–525. doi: 10.1109/CONFLUENCE.2019.8776934.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Proceedings of the British Machine Vision Conference (BMVC)*, Swansea, UK, 2015, pp. 41.1--41.12. doi: 10.5244/C.29.41.

[17] S. Khan, A. Akram, and N. Usman, "Real Time Automatic Attendance System for Face Recognition Using Face API and OpenCV," *Wirel. Pers. Commun.*, vol. 113, no. 1, pp. 469–480, 2020, doi: 10.1007/s11277-020-07224-2.

[18] M. Co\cskun, A. Uçar, Ö. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *2017 International Conference on Modern Electrical and Energy Systems (MEES)*, 2017, pp. 376–379. doi: 10.1109/MEES.2017.8248937.

[19] C.-H. Boe, K.-W. Ng, S.-C. Haw, P. Naveen, and A. Anaam, "An Automated Face Detection and Recognition for Class Attendance," Sep. 2024, doi: 10.62527/joiv.8.3.2967.

[20] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A Deep Multi-Task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 121–135, Jan. 2019, doi: 10.1109/TPAMI.2017.2781233.