International Journal of Data Science and Artificial Intelligence (IJDSAI)
Volume 04, Issue 01, January – February (2026)

RESEARCH ARTICLE

FACE CNN YOLO-SA: FACIAL ATTENDANCE USING CNN-BASED ENHANCED FUSION OF YOLO AND SPATIAL ATTENTION FOR ACCURATE REAL-TIME RECOGNITION SYSTEM

Mathumitha M S[1,*], and Pushpalatha C[2]

[1]Arunachala College of Engineering for Women, Manavilai, Nagercoil, India.
[2]Assistant Professor, Department of Computer Science and Engineering, Arunachala College of Engineering for Women, Manavilai, Nagercoil, India.

[*]*Corresponding e-mail: mmathumitha643@gmail.com*

Abstract – Face recognition system is based on combining conventional methods and the computerized system that uses deep learning methods to automate the process of attendance recording. The current method does not give precision in the real-life educational context as the brightness, direction, blockage and image quality vary. To overcomes this problem, a novel model proposes employing a Convolutional Neural Network (CNN) with spatial attention to identify essential facial characteristics and detect students, with an objective of boosting accuracy, minimizing time consumption and enhancing overall efficiency and reliability of attendance management in educational settings. A Wiener and Median (WieMed) Filter are utilized for preprocessing for improving the acquired image from the camera which becomes the input data for the CNN. The face is detected using You Only Look Once (YOLO) and it is extracted by CNN for efficient attendance marking of the student. The extracted features are of the face uses the PCA for feature reduction then Euclidian Distance Clustering recognize the face of the student the attendance is stored in the Excel sheet. The proposed model achieved an accuracy of 97.50% for recognizing the face. The proposed model improves overall accuracy by 25.07%, 7.82%, 1.15%, and 7.82% compared to the Eigen Face Recognizer algorithm, CNN, MTCNN, FaceNet, and Deep learning approach, correspondingly.

Keywords *– Face recognition, Convolution Neural Network, You Only Look Once, Euclidian Distance Clustering, Deep Learning.*

1. INTRODUCTION

Facial recognition (FR)-based attendance systems automatically recognize and authenticate students by evaluating their facial features via online images and video streams [1]. FR technology has been used to build and implement automated attendance systems in a variety of settings, including businesses, educational institutions, and event sites [2]. In college attendance is required for every class. Roll-call or manually signing a document is the most basic ways to take attendance [3]. Particularly in the workplace and academic institutions around the world, attendance has been a part of many institutions. It serves as a gauge of a person's dedication to their work [4]. Facial recognition is the most widely used identification method and continues to advance for accurate individual verification [5]. In large classrooms attendance tracking becomes time-consuming, inefficient and error-prone [6].

Recent research uses a CNN based single sample per person facial recognition algorithm to rapidly and accurately validate student attendance in real-world classroom and online learning situations [7]. Face recognition recognizes people using image data with LBPH ensuring low cost and CNNs delivering high accuracy [8]. This research describes a DL-based face recognition system for accurate contactless student attendance under a variety of scenarios [9]. The paper describes a handwritten digit recognition system that uses preprocessing techniques with OpenCV and is implemented as a Flask-based web application for practical use [10].

In a variety of situations attendance systems are prone to inaccuracies, spoofing hazards and decreased accuracy [11]. There are several issues with attendance systems including time-consuming, security issues, high costs and limited flexibility [12]. Although automated attendance systems can be used in practice, they still face threats of spoofing, high price tag and decreasing accuracy in the real world [13]. Traditional attendance systems are inefficient unreliable and prone to fraud [14] which leads to poor accuracy in attendance and administrative difficulties. There are variations in image quality, lighting, and motion affecting recognition accuracy and there is a time lag in comparing images which restricts real-time performance [15]. Previous methods are unsuccessful in terms of illumination and sharpness and sometimes cannot recognize all students in large classrooms. As a result, a cost-effective hybrid model is required to increase accuracy and identify all pupils. The primary contribution of the research has been outlined here:

- The proposed model integrates CNN with Spatial Attention to enhance the accuracy of recognizing the student's face.
- WieMed filters are used for enhancing the quality of the images through minimizing noise and upgrading the images so that the student face can be recognized easily.
- The VGG Face dataset compares student images within the educational setting and provides their current position.
- Recognized students have been assigned as present while others are assigned as absent. A database records every student's status.

The remaining work had been already organized as follows: Section 2 discusses recent research on face recognition-based attendance systems. Section 3 concisely outlines the proposed CNN with Spatial Attention model for a facial recognition-based attendance system. Section 4 evaluates the final result of the proposed CNN with Spatial attention model and analyzes them to the proposed and current approaches. The conclusion has been included in Section 5.

2. LITERATURE SURVEY

Several research have been conducted to classify human stress using a range of ML and DL methodologies. This system provides a comparative review of existing methodologies including suggested techniques and achieved accuracies.

In 2022 Rajagopalan., [16] employed a CNN-based face recognition network for real-time identification. It's used in automatic attendance and surveillance systems. The model handles position, lighting and emotion alterations well. It permits constant monitoring and quick processing. The system has 97.2% accuracy.

In Alameen., et al., [17] combined the LBPH algorithm which is connected with GSM technology. It is utilized for automated attendance and communication support. The system reduces manual effort and allows for remote alerts. It performs well under mild settings. It has 94.6% accuracy.

In Temiz., [18] used PCA Eigenfaces technique for face recognition which decreases dimensionality while effectively extracting crucial facial features. It is used to create a quick and automated classroom attendance system. It works well in regulated situations with low processing delay. The system achieves 92.3% accuracy indicating that it is reliable for attendance marking.

In 2022 Cheng., [19] This work uses a CNN-based face recognition algorithm to automate sign-in in smart classrooms. It is intended to replace manual attendance by providing precise and contactless student identification. It decreases human mistake and increases efficiency in attendance tracking. The model achieves 96.5% accuracy indicating dependable and consistent performance.

In 2022 Cao., et al., [20] developed a pose-aware face recognition model in conjunction with quality assessment methodologies. It is used to improve accuracy in different head postures and image quality circumstances. To reduce errors the algorithm examines the quality of facial images before recognizing them. The technique obtains an accuracy of 97.4% indicating increased robustness.

The literature survey indicates that many methods use Face CNN YOLO-SA for face recognition-based attendance system and achieve high accuracy but they require complex models, big datasets and expensive processing resources restricting their application in real time. Many systems focus only on small dataset for face recognition-based attendance system. To address these constraints, we present a more efficient and integrated methodology for accurate Face CNN YOLO-SA.

3. PROPOSED FACE CNN YOLO-SA METHODOLOGY

In this section, the proposed system utilizes CNN with Spatial attention and YOLO to provide an advanced and intelligent framework for accurate face recognition-based attendance system. Figure 1 represents the Schematic representation of the proposed model.

3.1. WieMed Filter

The face recognition image restoration and wireless communication systems frequently pollute received data with various types of noise including Gaussian noise and impulsive salt-and-pepper noise. Traditional filtering strategies are often intended to handle only one sort of noise adequately. For Gaussian noise the Wiener filter performs best in terms of minimal mean square error (MMSE) however the Median filter is particularly successful at reducing impulse noise while keeping edges. Thus, combining these two filters yields a Hybrid Wiener-Median Filter which performs better in complex noisy settings. The mathematical representation of the Median filter is given as:

$$\widehat{D}(n) = Median(n_u) \tag{1}$$

After median filtering the Wiener filter is used to achieve optimal linear estimates. The Wiener filter minimizes the mean square error between estimated and true signals by utilizing statistical aspects such as signal autocorrelation and cross-correlation. In a two-order Wiener filter the estimated signal is represented as:

$$\widehat{H}_{2D} = k_{2D,0}\widehat{H}_{m1} + k_{2D,1}\widehat{H}_{m2} \tag{2}$$

The Wiener filter adapts based on signal correlation and works well with Gaussian noise. When there is little noise, it acts similarly to linear interpolation but with superior noise suppression. The Hybrid Wiener-Median filter combines both techniques first removing impulse noise and then applying optimal estimation:

$$\widehat{H}_{Hybrid} = k_{2D,0}\widehat{D}_{p1} + k_{2D,1}\widehat{D}_{p2} \tag{3}$$

In this hybrid technique the Median filter reduces impulsive noise during the preprocessing step while the Wiener filter optimizes MMSE-based estimation on the cleaned signal. This combination increases robustness, decreases noise and improves accuracy in applications including face recognition, image denoising and wireless channel estimation. As a result, the Hybrid Wiener-Median filter is an effective and dependable solution for real-world signal processing situations involving several noise types.

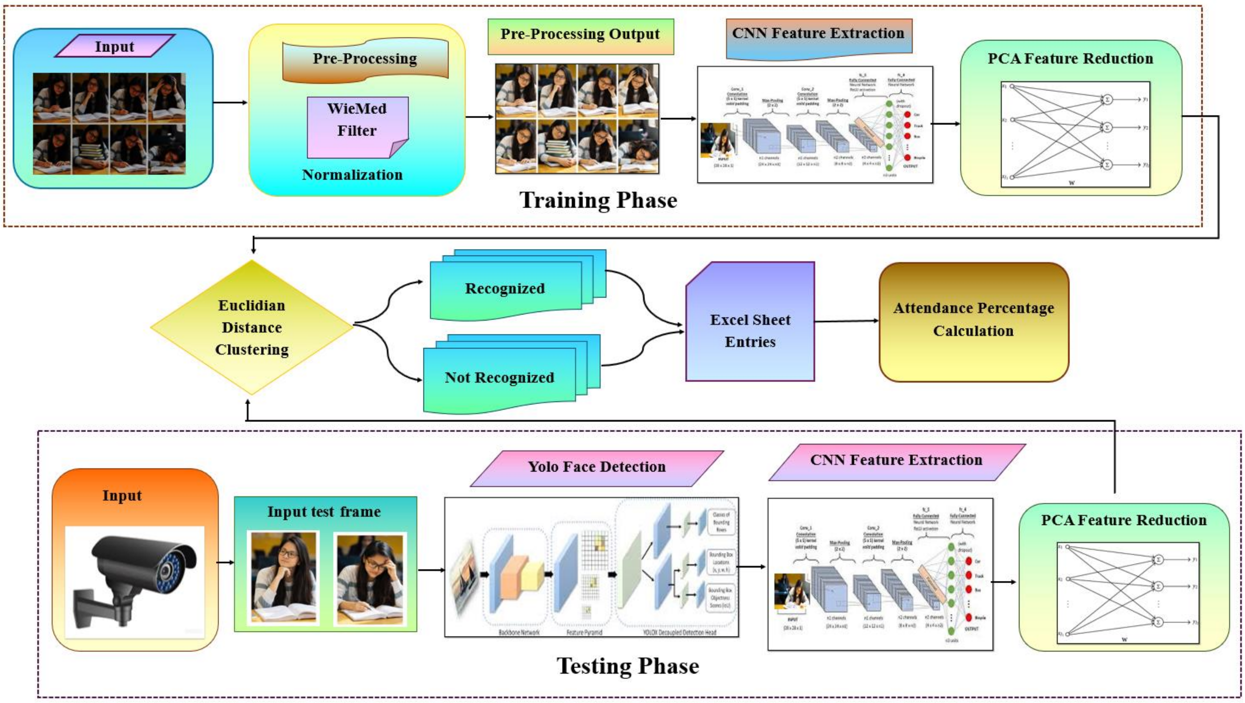


Figure 1. Schematic representation of the proposed Face CNN YOLO-SA model

3.2. CNNSpa

In the proposed system, a CNN with a spatial attention (CNNSpa) mechanism is used to extract relevant patterns from sequential input. The CNN uses convolutional, activation, pooling and fully connected layers to learn local dependencies along a single dimension in Fig 2. The convolution operation extracts features using kernels stated as follows:

$$d_u = p(g_u * b_u + y_u) \tag{4}$$

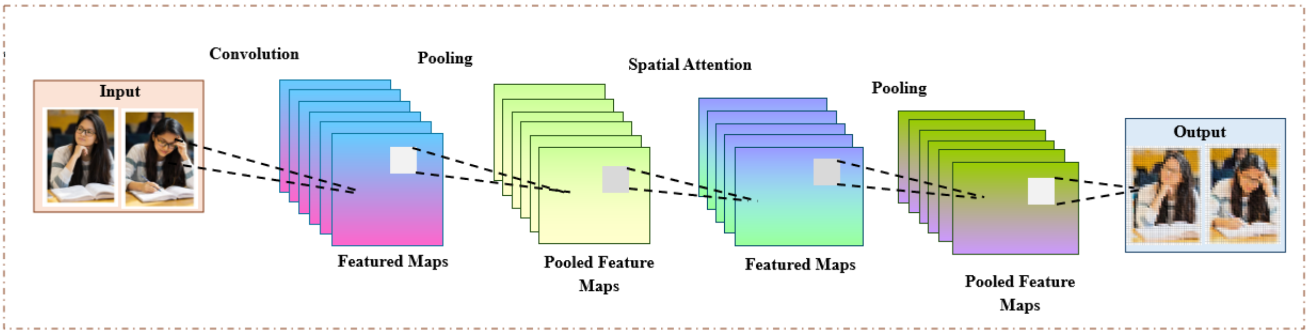


Figure 2. Architecture of CNN with Spatial Attention

where b_u represents the input to the convolution layer, d_u denotes the u^{th}output feature map, $e_i$is the weight matrix kernel, $*$ denotes the convolution operation, y_u is the bias, and $p(\cdot)$is the activation function. To introduce nonlinearity and improve learning capability, the Rectified Linear Unit (ReLU) activation function is commonly used which is defined as:

$$d_u = p(f_u) = \max(0, f_u) \tag{5}$$

where f_u is the intermediate feature obtained after convolution. This activation helps in suppressing negative values and retaining important positive activations. Max pooling reduces dimensionality while retaining dominating features:

$$m_u = \gamma(d_u, d_{u-1}) + \beta_u \tag{6}$$

A spatial attention method is then used to highlight critical regions in the feature maps hence boosting feature quality by focusing on relevant data. Finally, the fully connected layer generates the following output

$$b_u = p(t_u m_i + \delta_u) \tag{7}$$

where b_u represents the final output, t_u is the weight matrix, δ_u is the bias, and $p(\cdot)$is the activation function.

The hybrid CNNSpa allows for effective feature extraction by capturing local trends decreasing redundancy and stressing key spatial regions. This method considerably enhances system performance in applications requiring important feature identification such as signal analysis, face recognition and classification tasks.

3.3. YOLO v8

YOLOv8 is used in a face recognition-based attendance system to accurately and quickly recognize faces from images or video feeds. It uses convolution layers, SiLU activation and batch normalization to extract deep facial features as well as C2f modules to further improve feature flow and capture fine features. The SPPF module enhances multi-scale feature extraction enabling accurate feature detection from different angles and distances.

Normalized bounding boxes (a, b, w, h) are used to localize faces where the center coordinates and sizes are normalized to be uniform across images sizes. The normalization is calculated as:

$$d_w = \frac{1}{w} \quad , d_h = \frac{1}{h} \tag{8}$$

This is a normalization of the horizontal position of the identified face. The middle point on the x axis is calculated as follows:

$$a = \frac{a_1 + a_2}{2} \times d_w \tag{9}$$

Once the faces are detected YOLOv8 extracts important facial features which are then passed to the recognition block to recognize people. Comprehensive loss function is used for learning to guarantee detection and classification accuracy. The localization loss ensures that the face area detected is close to the actual face area.

Model performance is evaluated using:

$$mAP = \int_0^1 r(o)do \tag{10}$$

where $r(o)$represents the precision at different recall levels.

Finally, YOLOv8 is efficient for feature extraction with the small convolutional process improved feature flow with C2f module and multi-scale representation with SPPF. The normalized bounding box formulation with comprehensive loss function enhances the model ability of learning accurate and stable features which is suitable for the task of real-time object identification.

3.4. PCA Reduction

In this section, Principal Component Analysis (PCA) is used in a facial recognition-based attendance system to reduce dimensionality of image data keeping the important information. It displays faces using Eigenfaces and shows the most variance in the data set. The data should be normalized using the mean face. The normalized data are then used to calculate the covariance matrix as follows:

$$Z = \frac{1}{s} \sum bb^T \tag{11}$$

is is a covariance matrix of the pixel values. Then eigen values and eigen vectors of the covariance matrix are calculated as follows

$$ZU = U\Lambda \tag{12}$$

The Mean Squared Error (MSE) is used to measure the quality of the reconstruction after dimensionality reduction.

$$MSE = \frac{1}{M \times N} \sum (U - U_p)^2 \tag{16}$$

Finally, an attendance system based on face recognition is significantly improved by using PCA to reduce the dimensionality of the data and retain important features of the face. The system forms its eigenfaces from the most important eigenface thus reducing the storage space and increasing recognition accuracy as well as computation speed. This is why PCA is a helpful and widely-adopted method for feature reduction in real-time attendance systems.

3.5. Euclidean Distance Clustering

In this section, the attendance system compares and clusters the facial feature vectors with Euclidean Distance Clustering. Each face after PCA is a feature vector and similarity is computed based on the Euclidean distance between the two vectors. Faces within the same person form clusters and the system finds a test image by finding the closest cluster. If the likeness is within a certain threshold, then the person is identified otherwise, they are considered unknown. Once detected attendance is automatically recorded and kept for future purpose.

4. RESULT AND DISCUSSION

The proposed method is software and hardware hybrid based to perform fast stress analysis and model execution which is implemented in python using the following software libraries: TensorFlow, Keras, NumPy, Pandas and Scikit-learn. Google Colab is used for training faster with GPU support and Visual Studio Code for coding and debugging. It requires a minimum of 8GB of RAM and a strong processor and it is recommended that it be supported by a GPU for optimal performance. Data is split into training data and testing data at a typical 80:20 ratio to prevent overfitting and to give a reliable evaluation.

4.1. Performance Analysis

The proposed system performance is evaluated using measures such as accuracy (AY), precision (PN), specificity (SY), recall (RL) and F1-score (F1-S) all of which indicate dependable recognition results. The combination of YOLOv8, PCA and Euclidean distance accelerates detection, decreases complexity and increases overall efficiency. The technology functions consistently under different settings such as illumination and position thanks to sophisticated feature extraction techniques. Overall, it delivers excellent accuracy while requiring minimal processing time making it ideal for real-time attendance applications.

$$\mathrm{AY} = \frac{\mathrm{T_P} + \mathrm{T_N}}{\mathrm{T_P} + \mathrm{T_N} + \mathrm{F_P} + \mathrm{F_N}} \tag{17}$$

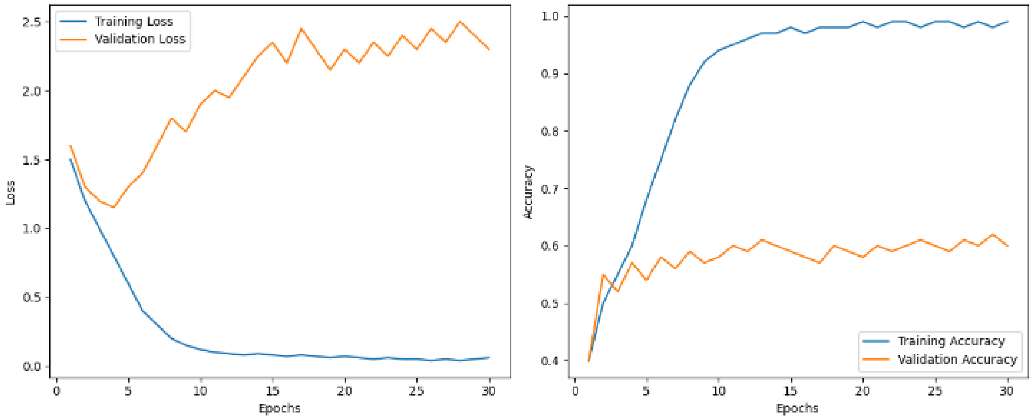
$$\mathrm{RL} = \frac{\mathrm{T_P}}{\mathrm{T_P} + \mathrm{F_N}} \tag{18}$$

$$PN = \frac{T_P}{T_P + F_P} \tag{19}$$

$$SY = \frac{T_N}{T_N + F_P} \tag{20}$$

$$F1 - S = 2 * \frac{PN \times RL}{PN + RL} \tag{21}$$

From the above equation (17-21) T_P represents True positive, T_N represents True negative F_P represents False positive and F_N represents False Negative in analysing the performance metric of face recognition-based attendance system.



(a) **(b)**

Figure 3. Training and testing of the proposed Face CNN YOLO-SA model (a) Accuracy Curve (b) Loss Curve

Figure 3 depicts the proposed model's training and testing performance across 30 epochs. The accuracy graph shows a consistent increase in both training and testing accuracy indicating robust convergence and generalization with little overfitting. Figure 3 (a), (b) The loss graph demonstrates a constant decrease in training and testing loss indicating higher learning efficiency. Overall, the curves show the model's robustness and stability in face recognition-based attendance system.

4.2. Comparison Analysis

Table 1. Comparison of DL network with proposed Face CNN YOLO-SA model

Networks	AY (%)	SY (%)	PN (%)	RL (%)	F1-S (%)
CNN	95.3	94.4	94.3	93.1	93.3
DenseNet	95.9	94.9	94.6	93.6	93.8
DNN	96.2	95.3	95.1	94.5	94.7
ResNet	96.7	95.8	95.6	94.8	95.1
CNNSpa (OURS)	**97.5**	**96.7**	**96.5**	**95.7**	**96.2**

Table 1 evaluates the overall efficacy of deep learning models following the suggested methods namely CNN, DenseNet, DNN and ResNet. Performance has been assessed using a variety of metrics included the DL technique's AY, SY, PN, RL and F1-S. CNNSpa networks have performed better than other networks. It increases accuracy by 2.26 %, 1.67%, 1.33% and 0.82% beyond CNN, DenseNet, DNN and ResNet respectively.

Table 2. Performance comparison: Existing models vs Proposed model

AUTHORS	APPROACHES	AY (%)
Rajagopalan., [16]	CNN-based deep face recognition	97.2
Alameen., et al., [17]	LBPH + GSM-based attendance system	94.6
Temiz., [18]	PCA (Eigenfaces) for face recognition	92.3%
Cheng., [24]	CNN-based face recognition for smart classroom	96.5%
PROPOSED	**Face CNN YOLO-SA**	**97.5%**

The experimental settings are described in Table 2, and included test samples which were gathered to test the effectiveness of the former frameworks. The proposed system has a higher overall accuracy over the CNN, LBPH+GSM, PCA and CNN technique parameter by 0.31%, 2.97%, 5.33% and 1.02% respectively.

5. CONCLUSION

This research proposed a novel Face CNN YOLO-SA Model for facial recognition-based attendance management system that employs image preprocessing and a CNN with spatial attention to address the drawbacks of traditional attendance approaches in real-world educational environments. To enhance the quality of the image WieMed filters were employed followed by deep feature extraction with CNNSpa to identify critical facial regions. The technology automate attendance recording, reduces human error and raises dependability by effectively dealing with fluctuations in brightness, posture and image quality. YOLO v8 is used to detect the face Experimental results indicates that the proposed approach has a high recognition accuracy of 97.5% and an F1-score of 96.2% suggesting that it is flexible and reliable. CNN, LBPH+GSM, PCA and CNN technique parameter by 0.31%, 2.97%, 5.33% and 1.02%

respectively. Although the proposed approach succeeds well it requires adequate computing power for real-time implementation and may have issues with excessive obstacles or low lighting scenarios. Future studies are expected to focus towards strengthening robustness via minimalist architectures more focus towards techniques and video-based temporal analysis in additional to adopting combating spoofing and safeguarding confidentiality strategies to guarantee trustworthy and sustainable implementation in smart educational environments.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the supervisor and faculty members for their valuable guidance and continuous support throughout this research work.

REFERENCES

[1] G. Al-Muhaidhri and J. Hussain, "Smart attendance system using face recognition," *International Journal of Engineering Research and*, vol. 8, pp. 10–17577, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[2] P. B. Jha, A. Basnet, B. Pokhrel, B. Pokhrel, G. K. Thakur and S. Chhetri, "An automated attendance system using facial detection and recognition technology," *Apex Journal of Business and Management*, vol. 1, no. 1, pp. 103–120, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] A. Rao, "AttenFace: A real time attendance system using face recognition," in *2022 IEEE 6th Conference on Information and Communication Technology (CICT)*, pp. 1–5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] R. H. A. Jr, J. G. Bernadas, A. F. A. Pangandaman and L. C. Borja, "Automated Attendance System using RFID, Face Recognition, and SMS." [CrossRef] [Google Scholar] [Publisher Link]

[5] S. Chowdhury, S. Nath, A. Dey and A. Das, "Development of an Automatic Class Attendance System using CNN-based Face Recognition," 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] M. W. H. Alameen, Y. M. Hussein, A. M. Kadhim and F. M. Mustafa, "Attendance System using Face Recognition Techniques and GSM." [CrossRef] [Google Scholar] [Publisher Link]

[7] F. P. Filippidou and G. A. Papakostas, "Single sample face recognition using convolutional neural networks for automated attendance systems," in *2020 Fourth International Conference on Intelligent Computing in Data Sciences (ICDS)*, pp. 1–6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] A. Budiman, R. A. Yaputera, S. Achmad and A. Kurniawan, "Student attendance with face recognition (LBPH or CNN): Systematic literature review," *Procedia Computer Science*, vol. 216, pp. 31–38, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] E. Kagona and S. Usman, "Students' Attendance Management in Higher Institutions Using Azure Cognitive Service and OpenCV Face Detection & Recognition Attendance System," *Technology*, vol. 5, no. 1, pp. 43–55, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] H. Huang, Z. Chen, T. Bai and Z. Zhao, "Research on Handwritten Digital Image Recognition Model Based On Deep Learning and Construction of Browser Service Platform," in *Proceedings of the 2022 5th International Conference on Image and Graphics Processing*, pp. 102–108, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] U. Essien and G. Ansa, "A deep learning-based face recognition attendance system," *Global Journal of Engineering and Technology Advances*, vol. 17, no. 1, pp. 009–022, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] W. Kuang and A. Baul, "A real-time attendance system using deep-learning face recognition," 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] A. S. Lateef and M. Y. Kamil, "Face recognition-based automatic attendance system in a smart classroom," *Iraqi Journal for Electrical and Electronic Engineering*, vol. 20, no. 1, pp. 37–47, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] M. Hemanth, V. P. Sastry, M. S. S. Tej, N. Vishnu and R. V. L. S. N. Sastry, "Smart Attendance and Payroll Management Using Face Recognition," [CrossRef] [Google Scholar] [Publisher Link]

[15] M. B. Hossain, M. J. Zobair, R. N. Mohalder, H. Al Mortuza, M. S. Alam, R. Ara and M. A. Rahim, "Image Recognition System for Security in Public Examination," [CrossRef] [Google Scholar] [Publisher Link]

[16] A. Rajagopalan, "Real-Time Deep Learning-Based Face Recognition System," 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] M. W. H. Alameen, Y. M. Hussein, A. M. Kadhim and F. M. Mustafa, "Attendance System using Face Recognition Techniques and GSM." [CrossRef] [Google Scholar] [Publisher Link]

[18] H. Temiz, "Rapid marking attendance with face recognition," *Avrupa Bilim ve Teknoloji Dergisi*, no. 36, pp. 78–86, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] L. Cheng, "A Study of a Student-Designed Face Recognition Sign-in System in an English Smart Class," in *Proceedings of the 14th International Conference on Education Technology and Computers*, pp. 261–265, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Z. Cao, L. Shi, H. Ju, L. Yang and L. Pang, "Improving face recognition by pose-aware quality assessment and judgement," in *Pattern Recognition and Tracking XXXIII*, vol. 12101, pp. 149–155, 2022. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

Mathumitha M S born in Kanyakumari district, Tamil Nadu. She completed her Bachelor of Mathematics degree from women's Christian college, Nagercoil, affiliated with Manonmaniam Sundaranar University. Currently, she is pursuing M.C.A degree at Arunachala College of Engineering for Women. She is currently working at Research point India. Her interested research domain area is Image Processing.

Pushpalatha C working as an Assistant Professor in the Department of Computer Science and Engineering at Arunachala College of Engineering for Women, Manavilai and I have eighteen years of teaching experience. My area of interest includes Aerial Image Processing, Programming Languages and Networking. I have attended Practical workshops, Hands-on programming in MATLAB and Python at various technical institutions around India and contributed to numerous publications.

Arrived: 21. 02. 2026

Accepted: 17.03. 2026