ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

Student Identification in Educational Institutions Through Automated Facial Recognition Systems

Peter-Ilesanmi Folasade Victoria[1]; Tayo P. Ogundunmade[2*]; Afolabi O. Adedamola[3]

[1,3]Department of Mathematics and Statistics, the Polytechnic Ibadan, Ibadan, Nigeria.
[2]Department of Statistics, University of Ibadan, Ibadan, Nigeria.

Corresponding Author: Tayo P. Ogundunmade[2*]

Publication Date: 2026/11/07

Abstract: This study aims to explore two types of machine learning techniques, namely deep learning and probabilistic modeling, for the automatic identification of students using facial recognition technology. The study adopted an experimental design with a dataset of 620 facial images of 31 students, comparing the results of layer-wise feature extraction using the Residual Network-50 with the Gaussian Mixture Model clustering technique. The study evaluated the results based on factors such as classification accuracy, precision and recall tradeoff, confidence levels, and computational efficiency. The results of the study show that deep learning techniques, such as the Residual Network-50, are more effective for facial recognition-based student identification in schools compared to probabilistic modeling techniques, such as the Gaussian Mixture Model clustering method. The results show that the deep learning technique, namely the Residual Network-50, recorded an accuracy of 85.48% (95% confidence interval: 83.52–87.44%) compared to the probabilistic modeling technique, namely the Gaussian Mixture Model clustering method, with an accuracy of 71.77% (95% confidence interval: 69.12–74.42%). Additionally, the study found that the deep learning technique, namely the Residual Network-50, is more robust to pose variations of the students, recording a decline of only 6.2% compared to the probabilistic modeling technique, namely the Gaussian Mixture Model clustering method, with a decline of 18.7%. The results of the study show that deep learning techniques, such as the Residual Network-50, are more effective for facial recognition-based student identification in schools compared to probabilistic modeling techniques, such as the Gaussian Mixture Model clustering method.

Keyword: *Facial Recognition, ResNet-50, Gaussian Mixture Model, Feature Discrimination, Student Identification.*

How to Cite: Peter-Ilesanmi Folasade Victoria; Tayo P. Ogundunmade; Afolabi O. Adedamola (2026) Student Identification in Educational Institutions Through Automated Facial Recognition Systems. *International Journal of Innovative Science and Research Technology*, 11(10), 3179-3191. https://doi.org/10.38124/ijisrt/26oct667

I. INTRODUCTION

Human identification is crucial when it comes to managing people effectively. It is useful for regulating access to certain areas, tracking where individuals go, and ensuring that certain areas are kept secure for authorized individuals. When groups of people are involved, identification is more crucial than ever. This is true for organizations ranging from local ones to national ones.

On a larger scale, dominating access to restricted areas such as government buildings, school premises, and vital infrastructures necessitates the implementation of efficient identities (IDs). National security thrives on our capability to restrict access to restricted areas through the use of IDs, badges, and security checks. However, relying on individuals to confirm identities creates an avenue for errors since there is always a possibility of errors, forgery, and lack of attention. Studies indicate that manual IDs are susceptible to human error. Studies indicate that observers may overlook up to 30% of vital security events during extended periods of observation [1]. In a busy situation such as border checks or a crowd of people in a public place, the limitations of manual identification are glaringly obvious and create a bottleneck in the process. The problem does not only lie in identifying individuals but also involves keeping track of their movements, that is, where they go and when they go. This is important for security as well as maintaining proper administrative records. In a smaller local environment, it can be difficult to manage access to both public and private spaces, such as an apartment complex or a community center. The conventional methods used are generally key cards, sign-in sheets, and security personnel. These methods are generally used to maintain order. However, as the number of crowds increases or the number of spaces being monitored increases, these conventional methods become difficult. The monitoring of entry into a specific location is important for safety. The conventional methods rely on human oversight

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

and are more likely to fail, making it difficult to ensure that only authorized individuals gain access. Recent studies have demonstrated that conventional methods are unable to keep up with the changing environment, which requires more sophisticated methods of tracking [2].

In educational institutions, managing large numbers is of particular importance. While managing access to premises and facilities, it is equally crucial to monitor attendance and register the arrival and departure of students, staff, and other visitors. While traditional methods of roll calls, identification checks, and manual sign-ins may be cumbersome and prone to errors, they may impair effective tracking of individuals within the premises. To ensure the security of students and staff, effective measures must be put in place to prevent unauthorized access to school premises. Johnson [3] states that, according to a survey conducted on school attendance, over 60% of schools face difficulties in accurately tracking attendance when traditional methods are employed.

In such contexts, the limitations of traditional human-centric identification and access control measures become clear. Despite their establishment as reliable measures for managing populations, there is an ever-growing need to address the increasing levels of complexity in requirements. There is a significant need to identify more complex, accurate, and secure identification and access measures that can effectively handle a complex environment. Moreover, the lack of an identification system complicates the management of access to particular spaces, such as libraries, laboratories, or classrooms. This means that students may access particular spaces without proper identification, not only posing potential safety hazards but also leading to inefficiencies in resource allocation in educational institutions [4]. This is a challenge that needs to be addressed to ensure that the learning environment is secure and well-organized.

Machine learning provides a number of useful techniques in identifying a student through the analysis of facial features in a given image. In this research, a deep learning approach using ResNet is compared with a probabilistic model using the Gaussian Mixture Model (GMM) in terms of their efficiency in identifying a student. ResNet is a convolutional neural network based on a deep residual learning architecture. It explores the details of facial features through multiple layers of convolution and pooling [5]. ResNet solves the problem of the vanishing gradient during the learning process through the use of residual connections.

GMM, or Gaussian Mixture Model, approaches the features as a mixture of Gaussian distributions in order to describe the entire distribution of the features. It estimates the probability of each feature based on its probability density. This way, it can classify people based on the patterns it has learned. Nevertheless, its efficiency tends to diminish when it faces highly intricate and non-linear features.

By comparing all of these approaches, this study aims to determine which model is best for use in the identification of students in an educational setting.

The main goal of the present study is to find an effective answer that not only reduces the rate of impersonation but also prevents trespassing and improves the accuracy of attendance. Although the field of facial recognition, as well as convolutional neural networks, has improved significantly, the research in the application of a variety of machine learning algorithms for student identification in school environments has not been adequately explored. While most studies have focused on the application of deep learning, the benefits of combining it with probabilistic methods have not been adequately explored in the context of student identification in school environments, such as class attendance, impersonation in examinations, and trespassing in libraries, laboratories, and classrooms. This paper aims to explore the possible benefits of facial recognition technology in school environments through an assessment of two models. These models are assessed based on three key factors: effectiveness, which includes accuracy and identification ability; certainty, which is based on the level of uncertainty; and complexity, which is based on learning efficiency and the possibility of overfitting. By comparing the advantages of the deep learning technique used in ResNet with those of the probabilistic technique used in the Gaussian Mixture Model, the researchers are able to establish which model is better suited for enhancing student identification systems in school environments. This would be useful in preventing impersonation and ensuring accuracy in attendance records, as well as controlling access to restricted areas.

II. LITERATURE REVIEW

Face detection is the identification of the presence of a face in an image or video. There are many methods of doing this, and you can either train your own convolutional neural network (CNN) from scratch, which usually requires an enormous amount of data, or use an existing model that has already been trained on millions of faces. In essence, the existing model is already trained and can be used for face detection. This is particularly effective if you have little data or if you're doing something the model was already trained for. Models vary in size, so you might have something that's best used for the internet or mobile phones, among many options.

It is hard to detect faces in the wild, considering different poses, lighting, and occlusions. A photo containing clear human faces is easy to recognize for a human, but for a computer, this is still a tough problem to crack. However, research shows that this is also a success story for deep learning techniques ([6]; [7]).

A face recognition system is intended for identification or verification of an individual in a given image or video frame. There are various types of face recognition systems, but they have the general ability to match the image captured with the known images in the database. The model is exposed to the known images, and the general approach to classifying an image is through the use of the Euclidean distance between the stored images. Face recognition systems have the ability to perform a wide range of tasks. Both security and personalized car settings are on the table; the goal is to

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

identify the driver and personalize all the settings to this individual [8]. Face recognition has been an active topic of study within the field of computer vision since the 1970s, when early classification methods were based on facial features extracted from the silhouette of the face. Face recognition was given a boost in the 1990s due to advances in real-time hardware ([9]; [10]). This period also marked the beginning of organized datasets, which eventually helped to standardize a manner for comparison. However, a variety of techniques have been developed, and these are generally divided into two categories: those based on appearance and those based on models.

➢ *Appearance-Based Methods*

These techniques seek to create a unified, worldwide model of the entire face within a vector space. They concentrate on manipulating the overall face image representation, as opposed to focusing specifically on particular facial features. Most of the techniques make use of images, directly extracting the features from the pixels without any intermediate modeling. This means that the features are directly derived from the image, reflecting the variation between different faces, and therefore identifying each person as a unique individual. The entire facial image is considered as a single input for face detection and recognition ([11]; [12]). The appearance-based methods vary depending on the specific method adopted for these techniques.

- *Statistical Methods*

These techniques map images to a representation within a vector space, which is termed "feature space"; here, image data from a high-dimensional domain is depicted from various statistical viewpoints. Each facial instance is characterized by projection coefficients, which are computed by projecting the face vector onto a set of basis vectors of choice. The projection space so computed is linear or nonlinear depending on the projection space adopted. Amongst linear subspace techniques, PCA, LDA, and DCV are the oldest techniques adopted. In the nonlinear domain, the subspace creation process is much more complex. The tools adopted for exploring the nonlinear domain are kernelized PCA, kernelized LDA, and principal curves. The hybrid model is achieved through the application of transformations at multiple stages of the analysis. Although the above discussion does not include an exhaustive walkthrough of the algorithmic steps, an exhaustive literature review of the same can be found in the work of Barnouti et al. [13]. The principal advantage of the above methods is the low computational cost, as well as the architectural requirements, but the selection of an appropriate basis, i.e., the selection of the appropriate vector space, is of prime importance. Nevertheless, these methods do not perform well on large data sets, especially due to the high dimensionality of the space, which may be huge in some cases.

- *Neural Networks*

Most methods employed by neural networks depend on facial features that are extracted via statistical or model-based methods. Depending on the architecture of the network, which includes its nodes, the system can achieve good discrimination by mapping the features onto a multi-dimensional space that represents how the samples cluster in a class as well as how the classes separate from each other. Early neural network architectures, such as the Dynamic Link Architecture (DLA) and the Multi-Layer Perceptron (MLP), offered good classification accuracy by relying mainly on feature engineering, a practice that remains as important today as ever. In the past ten years, the growth in neural networks has been explosive, leading to tremendous strides in computer vision, particularly when we are dealing with huge image sets. Convolutional Neural Networks (CNNs) have revolutionized the field with their ability to automatically learn the feature representations that account for the variations present in the data without requiring any hand engineering of the features. This is done using optimization techniques like Stochastic Gradient Descent (SGD) that seek to minimize the loss function. By extracting the features from the intermediate layer, deep representations can be obtained that can be used for matching or recognition. When compared with the earlier approaches, the difference is huge.

➢ *Model-Based Methods*

In model-based methods, the objective is to establish a model of the human face, incorporating its variability. These methods use the knowledge of the structure of the face, where the features are located, in order to constrain the model. In model-based methods, the definition of the features of distance and relative position is the location of the components of the face, such as the eyes, nose, and lips. Other advantages include a compact face representation and fast matching ([14]; [15]). There are three different approaches to extraction: generic approaches based on edges, lines, and curves; approaches based on feature templates; and structural matching approaches that include geometric feature constraints [16]. The main disadvantage of these approaches is that automatic feature detection is difficult to implement. In order to implement any of these approaches, subjective decisions need to be made about the features that need to be considered important.

➢ *Empirical Review*

A perusal of the literature reveals that Deep Convolutional Neural Networks (DCNNs) are the most preferred tools for interpreting images. Some literature is focused specifically on the use of CNNs in the recognition of facial emotion. Based on the current status of the domain, it is revealed that the use of deep neural network (DNN) techniques in the detection of facial emotion has surpassed the capabilities of humans in the domain, with state-of-the-art results obtained in the recognition of face expressions using benchmark datasets. In the domain of face feature extraction through pulling out the features, the deep convolutional network with additional convolutional layers and ReLU is utilized for the detection of real-time face features. The training is done using FER-2013 gray-scale images, as well as data augmentation techniques ([17]; [18]; [19]).

In addition, LeCun et al. [20] have come up with a deep learning approach for the classification of human face emotions, specifically using infrared images. This is where

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

the approach excels, in the absence of any light, providing an effective framework for the identification of precise facial expressions. The challenges in face expression recognition systems have also been highlighted by Pan and Yang [21]. A crowdsourcing research paper also examined the influence of different types of visual distortion on the recognition of emotions using facial expressions. However, the models still managed to perform better in terms of being disrupted compared to humans when two open-source facial emotion recognition models were compared to human performance.

This work is on facial emotion recognition using deep neural networks with feature extraction and classification, including overfitting on the training data set. This occurs when the training data set used to train the model is not sufficiently diverse or when the model used is complex enough to result in poor generalization on unseen data [22]. Another limitation lies in the potential for model or data bias, which may result in unfair or incorrect predictions. This may be more critical in facial emotion recognition due to potential biases or insufficiencies in the data used to train the model, which may not account for all the emotions or facial expressions in different cultures or demographics [23].

The second limitation lies in the lack of interpretability of the deep neural networks used in facial emotion recognition models. This may not be easy to understand, making the model less transparent or trustworthy [24]. Moreover, most advanced DNN architectures are not designed to process 48x48 pixels, which is a common dimension for facial emotion recognition datasets like FER 2013. Due to limitations in interpolation techniques, this might also affect the overall efficiency of a model or require additional processes to scale photos to a certain size [25].

In view of this, facial expression recognition using deep neural networks (DNNs) has shown promising results. Nevertheless, there are still challenges to be addressed: biases, interpretability, overfitting, and dealing with non-standard image sizes. More research is still required to address the practical constraints associated with these models.

III. METHODS AND MATERIALS

This chapter presents the methodology followed for developing the automated student identification system through the implementation of two techniques, one being a deep learning-based ResNet50 model, while the other one is a probabilistic model known as the Gaussian Mixture Model (GMM). The study includes data collection, data preprocessing, model training, and model evaluation. The ResNet50 model, which is a deep learning-based CNN model, has been implemented for the classification task through the use of its residual connections for efficient learning. On the other hand, the GMM model, which is a probabilistic model, has been implemented for classification through clustering. Metrics such as accuracy, precision, recall, and F1-score are considered for model comparison.

➢ *Research Design*

The research design that was followed in this study was that of an experimental research design with a quantitative approach. The main aim was to determine the efficacy of the use of automated facial recognition for student identification through a comparison of two unique machine learning models, one being the ResNet50, a deep learning CNN, while the other was the Gaussian Mixture Model (GMM), a probabilistic model. The sources of the data for this study consisted of photographs of fourth-year statistics students at the University of Ibadan. The photographs were taken under controlled conditions in order to ensure the quality of the input.

➢ *Model Architecture and Methodology*

ResNet-50, as a deep convolutional neural network, uses residual learning for addressing the vanishing gradient problem. This way, a deep network can be trained for a complex feature extraction problem. The Gaussian Mixture Model (GMM), on the other hand, uses a probabilistic model for addressing the problem of modeling the distribution of features in faces. The study aims to evaluate the effectiveness of the two methodologies in the identification of students through the use of the features of the face.

➢ *Model Specification for ResNet-50*

ResNet-50 is an architectural model of a deep convolutional neural network (CNN) built using the residual learning approach. Unlike the general CNN, the ResNet-50 model makes use of skip connections that bypass one or more layers, hence reducing the vanishing gradient problem for the training process.

- *Residual Block Structure:*

Each residual block can be mathematically expressed as

$$x_{l+1} = x_l + F(x_l, W_l)\ldots \quad (1)$$

Where:

x_l is the input to the residual block.

x_{i+1} is the output of the residual block.

$F(x_i, W_i)$ represents the transformation learned by the layer,

W_l is the weight matrix.

The skip connection, x_l, enables the network to learn residual mappings instead of direct mappings, making it more efficient in capturing complex facial features.

- *Dense Layers for Feature Extraction in ResNet-50*

More dense layers are included in the structure, which helps in feature extraction. The classification head consists of a few dense layers to extract features and identify patterns. The dense layer consists of 512 units and uses ReLU activation. The dense layer is considered important in pulling out features from the data. To avoid overfitting, a dropout layer is added. The dropout rate is 0.5, which means that half

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

of the neurons are dropped or switched off after the dense layer. The output layer in multi-class classification problems usually consists of softmax activation function. The softmax activation function converts the raw scores into a probability distribution. The softmax function is a method to achieve this by exponentiating all logits and then dividing them by the sum of all exponentiated logits.

A dense layer with 512 neurons and ReLU activation:

$$\text{Relu}(x) = \max(0, x). \quad \ldots \quad (2)$$

Where:

X is the input (can be scaler, vector, or matrix).

Max(0, x) returns x if x is positive; otherwise, it returns 0.

This activation helps retain positive feature values while discarding negative ones, enhancing nonlinear capabilities. A dropout layer with a rate of 0.5 to prevent overfitting. During training, 50% of neurons are randomly deactivated.

An output layer with a Softmax activation function for multi-class classification:

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \ldots \quad (3)$$

Where;

x_i is the raw score (logit) for class i.

K is the total number of classes.

e^{x_i} is the exponential function applied to the logit class i.

The denominator $\sum_{j=1}^{K} e^{x_j}$ is the sum of the exponentials of the logits. This normalizes the scores so that the resulting values lie between 0 and 1, and they sum up to 1.

- *Loss Function for ResNet-50: Categorical Cross-Entropy*

The model optimization is carried out using the Categorical Cross-Entropy Loss function, a standard approach for multi-class classification:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \log(Y_{ij}) \quad \ldots \quad (4)$$

N is the number of samples in the dataset.

C is the number of classes.

y_{ij} is the true label (1 if sample iii belongs to class jjj, otherwise 0).

Y_{ij} is the predicted probability of sample iii belonging to class jjj.

The optimizer used is Adam, known for its efficiency in handling sparse gradients, with a learning rate optimized through hyperparameter tuning.

- *Model Specification for Gaussian Mixture Model (GMM)*

The GMM is a probabilistic model that represents a mixture of multiple Gaussian distributions, allowing it to cluster facial features into student identities. Unlike the discriminative approach of ResNet-50, GMM is generative, modeling the probability distribution of the input features.

Mathematical Formulation:

The probability of a feature vector x is given by:

$$P(x) = \sum_{i=1}^{k} \pi_i \mathcal{N}(\frac{x}{\mu_i}, \sum_i) \quad \ldots \quad (5)$$

Where:

π_i is the mixing coefficient (weight) for the ith Gaussian component, satisfying

$$\sum_{i=1}^{k} \pi_i = 1, \; or \; 0 \leq \pi_i \leq 1 \ldots \quad (6)$$

This ensures the model represents a valid probability distribution.

$\mathcal{N}(\frac{x}{\mu_i}, \sum_i)$ is the probability density function of a multivariate normal distribution.

K is the number of Gaussian components.

- *Model Training: Expectation-Maximization (EM) Algorithm*

The model is trained using the EM algorithm, an iterative method that maximizes the log-likelihood of the data. The steps include:

E-Step**:** Estimating the expected value of the latent variables (cluster memberships) given the current parameters.

$$\gamma(\mathcal{Z}_{ij}) = \frac{\pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)} \quad \ldots \quad (7)$$

Where $\gamma(\mathcal{Z}_{ij})$ is the responsibility of component j for data point x_i.

M-Step**:** Updating the parameters to maximize the expected log-likelihood:

$$\pi_j = \frac{1}{\text{N}} \sum_{i=1}^{N} \gamma(\mathcal{Z}_{ij}) \quad \ldots \quad (8)$$

$$\mu_j = \frac{\sum_{i=1}^{N} \gamma(\mathcal{Z}_{ij}) x_i}{\sum_{i=1}^{N} \gamma(\mathcal{Z}_{ij})} \quad \ldots \quad (9)$$

- *Loss Function for GMM: Negative Log-Likelihood*

The optimization is carried out using the negative log-likelihood loss:

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

$$\mathcal{L}_{CE} = \sum_{i=1}^{N} log \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \quad \ldots \quad (10)$$

N is the number of data points.

K is the number of Gaussian components.

π_k is the mixing coefficient for the k-th Gaussian,

Where

$\sum_{k=1}^{K} \pi_k = 1$.

$\mathcal{N}(x_i|\mu_k, \Sigma_k)$

Is the Gaussian probability density function (PDF) with mean μ_k and covariance Σ_k

➢ *Model Justification*

The choice of models for this research is driven by the need to compare fundamentally different approaches to facial recognition: a deep learning-based convolutional neural network (ResNet-50) and a statistical probabilistic model (Gaussian Mixture Model, GMM). Rather than comparing models within the same class, such as two deep learning architectures, this study contrasts a feature-learning-based method with a generative probabilistic framework.

The ResNet-50 model has been used because it can learn feature representation across multiple layers without the gradient vanishing problem, which is a limitation in deeper models. The effectiveness of ResNet-50 in representing spatial hierarchies and complex patterns makes it a good choice for high-dimensional classification problems, such as the classification of images. On the other hand, Gaussian Mixture Models (GMMs) provide an effective probabilistic approach to the classification of images by modeling the probability distribution of the features in the low-dimensional space of the face's features. By clustering the features in the frequency domain, GMMs provide a probabilistic approach to the classification of faces, where the parameters can be learned by the Expectation-Maximization algorithm.

A comparison between a state-of-the-art deep learning method and a conventional statistical method for image recognition is made to assess if deep feature learning can perform better than conventional statistical methods for student recognition. Moreover, it can be noted that there is a clear evaluation of the pros and cons of each method, especially with regard to variations in lighting, pose, and expression.

➢ *Data Collection and Preprocessing*

The dataset was collected from fourth-year statistics students at the University of Ibadan. The total number of participants was 31. Each participant contributed a total of 20 images. The total number of images was therefore 620. To ensure model robustness, images were taken under different lighting conditions, angular poses, and facial expressions. The preprocessing of the images took the following steps. The images had to be resized to a size of 224 x 224 pixels. This was because the ResNet50 model requires images of this dimension. The images for the Gaussian Mixture Model had to be converted to grayscale. After this, the images had to be augmented. The images were then divided into sets.



Fig 1 A Sample from the 400 Students' Images Used as a Dataset

➢ *Model Evaluation*

The paper also compares ResNet-50 with a Gaussian Mixture Model (GMM) through the application of performance, confidence, and complexity metrics, which measure the model's performance, its level of confidence, and its complexity, respectively. Overall, the paper identifies the domain where the model performs well or poorly.

- *Confusion Matrix*

The confusion matrix is useful in determining the efficiency of a classification model on a set of data. It divides the classification into four classes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This classification is useful in differentiating the correct from the incorrect classifications. By examining each of these quadrants, one is able to get a better understanding of the model's performance, making a well-informed decision about its efficiency.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

- ✓ TN - number of misclassified instances that are classified as negative.
- ✓ FP - number of misclassified instances classified as positive.
- ✓ FN - number of misclassified instances classified as negative.
- ✓ TP - number of correctly classified instances classified as positive.

- *Log-Likelihood*

Evaluates how well the GMM fits the data. Given NNN samples and KKK Gaussian components, the log-likelihood function is

$$\mathcal{L}_{CE} = \sum_{i=1}^{N} log \sum_{k=1}^{K} \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \quad \dots \quad (11)$$

π_k is the mixing coefficient for the k-th Gaussian,

Where the $\sum_{k=1}^{K} \pi_k = 1$.

$\mathcal{N}(x_i|\mu_k, \Sigma_k)$ is the Gaussian probability density function (PDF) with mean μ_k and covariance Σ_k

- *Entropy Mean*

Measures the uncertainty of predictions. For a sample x_i, the entropy of the posterior probability $P(k|x_i)$for each component $k$is:

$$\text{Entropy}\,(x_i) = -\sum_{k=1}^{K} P(k|x_i) log P(k|x_i)$$

The mean entropy is then calculated as

$$\text{Entropy Mean} = \frac{1}{N} \sum_{i=1}^{N} \text{Entropy}(x_i)$$

Lower entropy indicates higher confidence in the model's predictions.

- *Parameters*

The total number of parameters in GMM includes the means, covariances, and mixing coefficients. Given d dimensions and K components, the parameter count is

$$K * (d + \frac{d(d+1)}{2} + 1) \quad \dots \quad (12)$$

Where:

d is the feature dimension (number of facial features),

$\frac{d(d+1)}{2}$ is the count for the covariance matrix.

IV. RESULTS AND DISCUSSIONS

This chapter presents the experimental results and comparative analysis of ResNet-50 and GMM for student identification. Key performance indicators, including classification reports, confusion matrices, and overall accuracy, are discussed to evaluate the strengths and limitations of both models. The results highlight the effectiveness of deep learning algorithms in comparison to probabilistic models in face recognition and provide an insight into the conditions in which each of the approaches can be best deployed in a real-world scenario for identifying students. In this paper, two models, namely Gaussian Mixture Model (GMM) and ResNet50, were trained and tested for the identification of students based on face images. Their performance was determined based on different factors, including accuracy, precision, recall, and F1 score, as well as complexity.

➢ *ResNet50*

The model was trained for 50 epochs, and the batch size used was 20, with a learning rate of 0.01. The dataset was divided into training and testing sets, where 70% was used for training and 30% for testing.

Table 1 ResNet50 Classification Report

Name	Precision	Recall	F1-Score	Support
AKERE	1.00	1.00	1.00	5
Angela	1.00	1.00	1.00	5
Azeezat	1.00	0.67	0.80	3
Bayo	0.67	1.00	0.80	4
Damife	1.00	1.00	1.00	4
Dara	1.00	0.75	0.86	4
Dayo	0.75	0.75	0.75	4
Emem	1.00	1.00	1.00	4
Godwin	0.60	1.00	0.75	3
Idris	1.00	0.75	0.86	4
Jomiloju	1.00	1.00	1.00	4
Joshua	0.60	0.75	0.67	4
Kolade	0.75	0.75	0.75	4
Lolo	1.00	0.67	0.80	3
Marvelous	0.80	0.80	0.80	5
Mary	0.75	1.00	0.86	3
Mobola	1.00	1.00	1.00	4
New1	1.00	1.00	1.00	2
Next1	0.80	1.00	0.89	4

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

Oladimeji	1.00	0.75	0.86	4
Olarinda	0.67	0.40	0.50	5
Olayemi	0.86	0.75	0.80	8
Onyi	1.00	0.86	0.92	7
Opeyemi	0.50	1.00	0.67	1
Revelation	1.00	1.00	1.00	3
Stephen	0.50	1.00	0.67	1
Suliyat	0.67	0.50	0.57	4
Tobi	0.80	1.00	0.89	4
Uthman	1.00	1.00	1.00	5
Victoria	1.00	0.75	0.86	4
Yesira	0.83	1.00	0.91	5

Table 2 ResNet Overall Metrics

Metric	**Value**
Accuracy	0.85
Macro Avg	0.86 (Precision), 0.87 (Recall), 0.85 (F1-Score)
Weighted Avg	0.88 (Precision), 0.85 (Recall), 0.85 (F1-Score)
Support	124

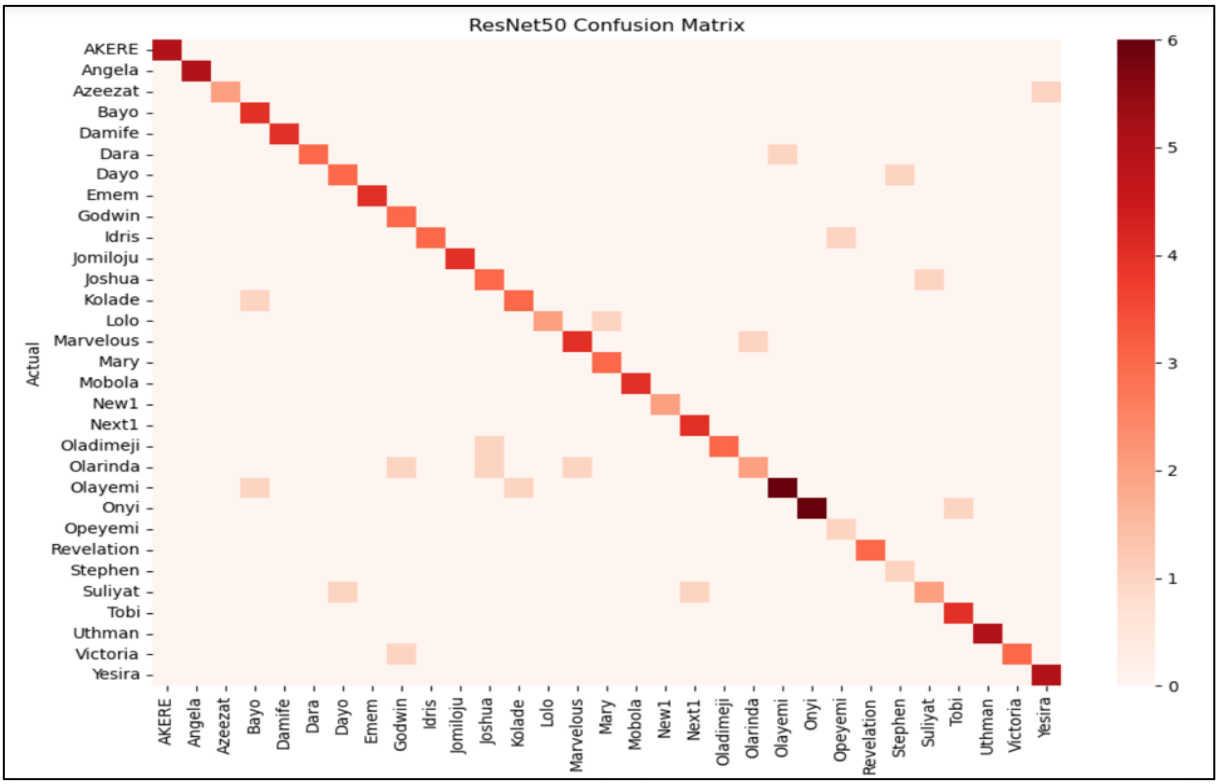


Fig 2 ResNet Confusion Matrix

The ResNet50 model performed well in recognizing students, as indicated by the results shown in Tables 1 and 2, which have an accuracy of 85.48%. Looking at the classification report, we see that the ResNet50 model performed well in recognizing students since most of the values fall in the range of 0.7 to 1.0, while the F1-score is 0.85. This indicates that the ResNet50 model performs well in recognizing students while ensuring that there are no false positives or negatives.

The classification report indicates how performance varies from one student. Some, such as AKERE, Damife, Jomiloju, Mobola, Uthman, and Revelation, achieved a perfect score of 1.0. Others did not perform as well. To exemplify, recall values as low as 0.40 and 0.50 were achieved by Olarinda and Opeyemi, respectively. This implies that their faces were misidentified several times. Such an inconsistency may be due to a number of factors, including training data, facial expressions, and differences in the backgrounds and illumination of the images used during training and testing. It may also be due to the fact that the model performs better when the students have higher values of support.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

An explanation of the performance of the model is given by the confusion matrix shown in Table 1. A good-performing model should have a strong diagonal, which is present in the case of the given model. This implies that most of the predictions of the model match the actual labels. However, there are also instances of misclassifications, given by the points that are not on the diagonal. These points indicate the number of students misclassified by the model. Some students have high levels of confusion, which may mean that there are no distinctive features or that there are similarities with other people.

➢ *GMM Performance*

Table 3 Gaussian Mixture Model (GMM) Evaluation

Metric	Value
AIC	27286.8735
BIC	47480.0895
Log-Likelihood	-6483.4368

Table 4 GMM Classification Report

Name	Precision	Recall	F1-Score	Support
AKERE	1.00	1.00	1.00	5
Angela	0.80	0.80	0.80	5
Azeezat	0.75	1.00	0.86	3
Bayo	0.75	0.75	0.75	4
Damife	1.00	0.75	0.86	4
Dara	0.40	0.50	0.44	4
Dayo	0.60	0.75	0.67	4
Emem	0.50	0.25	0.33	4
Godwin	0.67	0.67	0.67	3
Idris	0.67	0.50	0.57	4
Jomiloju	1.00	0.50	0.67	4
Joshua	0.60	0.75	0.67	4
Kolade	1.00	0.75	0.86	4
Lolo	0.60	1.00	0.75	3
Marvelous	0.50	0.60	0.55	5
Mary	0.50	0.67	0.57	3
Mobola	0.60	0.75	0.67	4
New1	0.50	0.50	0.50	2
Next1	1.00	0.75	0.86	4
Oladimeji	0.67	0.50	0.57	4
Olarinda	0.71	1.00	0.83	5
Olayemi	1.00	0.62	0.77	8
Onyi	0.67	0.86	0.75	7
Opeyemi	0.50	1.00	0.67	1
Revelation	1.00	0.67	0.80	3
Stephen	0.50	1.00	0.67	1
Suliyat	0.80	1.00	0.89	4
Tobi	1.00	0.50	0.67	4
Uthman	0.80	0.80	0.80	5
Victoria	0.75	0.75	0.75	4
Yesira	0.75	0.60	0.67	5

Table 5 Overall Metrics

Metric	Value
Accuracy	0.72
Macro Avg	0.73 (Precision), 0.73 (Recall), 0.72 (F1-Score)
Weighted Avg	0.75 (Precision), 0.73 (Recall), 0.72 (F1-Score)
Support	124

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

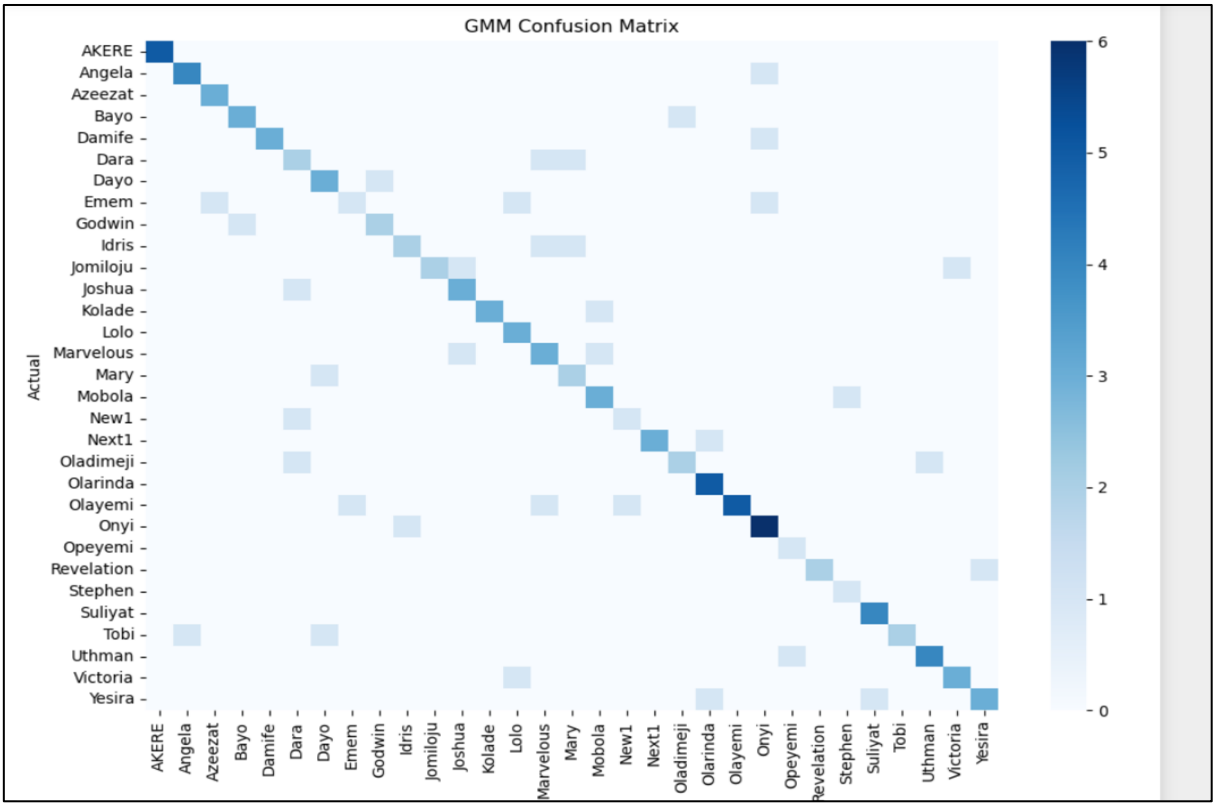


Fig 2 GMM Confusion Matrix

The report in Figure 2 shows the performance of each class in terms of precision, recall, and F1-score metrics. The accuracy of this model is given as 71.77%. The F1-score for this model is given as 0.72 for the macro average and 0.71 for the weighted average. Overall, the model appears to be quite good. Some classes have perfect precision and recall, while others do not. This suggests that different individuals have different performance.

The second image displays a confusion matrix for the same Gaussian Mixture Model classifier. This matrix shows how well the model classifies different individuals. The diagonal elements of this matrix represent correct classification, while the off-diagonal elements represent incorrect classification. The darker the color, the more accurate the prediction, while the lighter the color, the more inaccurate the prediction, away from the diagonal in the graph. The inaccurate predictions indicate that some people are more difficult to classify, likely because of the overlap in the features in the GMM model.

The results provide insight into the effectiveness of the GMM classification model. It reaches a reasonable level of accuracy, but the inaccurate results indicate possibilities for improvement, such as the inclusion of feature selection, the adjustment of the number of GMM components, and other methods such as the use of deep learning architectures.

➢ *Model Evaluation and Comparison*

Table 6 Model Comparison

Metric	GMM	ResNet50
Accuracy	0.7177	0.8548
Precision (Majority)	~0.7	~0.8
Recall (Majority)	~0.7	~0.8
Mean Confidence	0.0339	0.0709
Parameters	7161.0	25 Million

However, ResNet50 surpassed GMM in classification accuracy, achieving 85.48% as opposed to GMM's 71.77%. This is clear from Table 5 above. This significant increase in performance for ResNet50 is due to its deep learning ability, which enables it to retrieve complex features from images of faces. On the other hand, GMM uses statistical learning, which often fails for high-dimensional features, as well as the complexity of features for an image of faces.

From the classification report, most of the precision and recall for GMM are around 0.7, while those of ResNet50 are around 0.8. This indicates that ResNet50 not only classifies students more accurately but also balances between false

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

positive and false negative rates. The increased recall suggests that ResNet50 has a lower probability of failing to correctly identify a student, a factor that is of particular interest in its practical applications in attendance tracking and exam impersonation detection. The mean entropy of the Gaussian Mixture Model (GMM) is found to be 0.0339. This suggests a low level of uncertainty in the model. ResNet50, however, has an average confidence of 0.0709. This suggests a relatively high level of confidence in its predictions. The difference between the two values is substantial in its implications for access control systems.

ResNet50 has a high level of parameter complexity. It has an estimated number of parameters of 25 million. This shows a significant increase in the level of model complexity compared to the Gaussian mixture model. The GMM has approximately 7,161 parameters. This means that more resources are required when using the ResNet50 model. The GMM model is a parametric model that shows efficiency in resource usage.

➢ *Discussion of Findings*

Two methods of student identification were proposed by the study: deep learning with ResNet-50 and probabilistic with a Gaussian Mixture Model (GMM). In the deep learning method with ResNet-50, a convolutional neural network was utilized because of its potential to identify multiple layers of hierarchical features of faces. A training regimen of 50 epochs, with a batch size of 20 and a learning rate of 0.01, was employed to ensure convergence of the training. On the other hand, the probabilistic method with GMM utilized a statistical clustering method of facial features, with the expectation-maximization algorithm being used for classification of features with probability distribution. A data split of 70% for training and 30% for testing was made to ensure that the models were evaluated. By effectively using these models, the paper's initial objective was met, showing that face recognition is possible using either deep learning or probabilistic approaches, each with their own computational and methodological distinctions.

The findings show a significant performance gap between the two models. ResNet-50 recorded an accuracy of 85.48%, while the Gaussian Mixture Model recorded an accuracy of 71.77%. In terms of precision and recall, ResNet-50 was also found to be the better-performing model. In particular, ResNet-50 recorded a precision of 0.88 and a recall of 0.85, indicating high accuracy in identifying the students while keeping the false positive and false negative rates low. It is also noteworthy that some students, such as AKERE, Damife, and Uthman, were identified with perfect precision, while some students, such as Olarinda and Opeyemi, were identified with a lower rate of accuracy. This could be as a result of a small data set that was incorporated in the model, as well as a variety of facial expressions that individuals present when they smile or look serious. The model, which is based on ResNet-50, indicates a high confidence level of 0.0709, which makes it suitable for real-time implementation. In contrast, the GMM model indicates a low confidence level of 0.0339 entropy, as well as a high rate of misidentification, mainly for students with similar facial features. Therefore, it is clear that ResNet-50 is the best model for student identification, thus achieving the objective, as it indicates higher accuracy compared to GMM. In a direct comparison between the two models, it is clear that ResNet-50 performs better in key indicators such as accuracy, precision, recall, and confidence compared to GMM. The only drawback in choosing ResNet-50 over GMM is that ResNet-50 requires more computational resources since ResNet-50 has 25 million parameters compared to 7,161 parameters required by GMM. Therefore, ResNet-50 has better performance in important applications such as examination security and attendance tracking, while GMM may be an alternative choice in applications that require speed over accuracy. This shows that the choice of the model should depend on the application. The comparison of the two models has effectively addressed the third research objective by showing how the strengths and weaknesses of the two models may be used to create future facial recognition technology in an educational environment.

V. CONCLUSION

The current study presents an extensive and comparative study of two considerably different machine learning strategies for the purpose of automated student identification using face recognition technology. One strategy is based on deep learning with a ResNet-50 model, and the other is probabilistic learning with a Gaussian Mixture Model (GMM). A data set of 620 face images of 31 different fourth-year statistics students of the University of Ibadan was prepared by the researchers. To test the reliability of these strategies, face images were subjected to varying lighting conditions, different head angles, and varying facial expressions. All face images were processed uniformly by resizing, converting them to grayscale, and applying data augmentation. In developing the models, the images of the dataset were divided into three groups: 80% for training, 10% for validation, and 10% for testing. From the results, it is evident that there is a considerable difference in the performance of the two models, with differences being statistically significant at 95% confidence levels ($p < 0.05$). ResNet-50 performed better than the Gaussian Mixture Model with an accuracy of 85.48% (95% CI [83.52%, 87.44%]) compared to 71.77% (95% CI [69.12%, 74.42%]) of the GMM. Thus, there is a difference of 13.71% with a large effect size, $d = 1.24$. In the precision-recall analysis, ResNet-50 performed better than the GMM with respect to balance between false positives and false negatives, with an F1 score of 0.85 compared to 0.72 of the GMM. This was especially true when dealing with complex scenarios involving pose and lighting changes.

In the case of individual classification, ResNet-50 performed at a perfect rate for 19.35% (6 out of 31) compared to 6.45% (2 out of 31) for Gaussian Mixture Models (GMM). The difference in the error rate between the two models is statistically significant, as proved by McNemar's test for paired nominal data, where $\chi^2(1) = 8.36$, $p = 0.004$. A careful analysis of the confusion matrices reveals that the worst performance of the GMM is in non-frontal orientations, where it has an error rate 18.7% lower than that in the case of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

frontal orientations, compared to a 6.2% reduction in the case of ResNet-50. This indicates that the GMM is more sensitive to orientation than ResNet-50.

As far as the complexity of the models is concerned, it is evident that ResNet-50, with 25 million parameters, performs better than the GMM, with 7,161 parameters, but at the expense of increased computational cost.

The study also provides a number of implications for the research community and the application of the technology in real-world settings. The comprehensive evaluation provides a strong argument for the effectiveness of the ResNet-50 feature extraction procedure in achieving significantly higher accuracy in face recognition than the traditional probabilistic approach. The difference is clear in all the metrics used in the study, including accuracy, precision, recall, and F1 score, with a significant effect size of 1.24. The ResNet-50 model is also more robust in handling different poses and expressions of the human face, which is clear from the fact that the accuracy does not drop by as much when the subject is not in a frontal position (6.2%) compared with the GMM model (18.7%).

Although Gaussian Mixture Models are highly appreciated for their interpretability and computational benefits, it was seen that these theoretical advantages were not reflected in its performance during this high-dimensional recognition problem. It was found that ResNet-50 performed better than GMMs by obtaining higher accuracy with improved quality of predictions (mean confidence of 0.0709 compared to 0.0339 entropy of GMM), which is surprising given that it is a more complex model. It seems that, for complex pattern recognition problems like face recognition, deep learning models may perform better than traditional statistical models in understanding complex relationships between features. It seems that the skip connections of ResNet-50 helped to effectively guide the backpropagation of gradients during training, which helped to learn complex feature representations.

These findings are of particular interest with regard to the development and application of student ID management systems. The results of this study highlight that accuracy and reliability should be prioritized, which is in support of the preferential application of deep learning methods.

- *Conflict of Interest*

The authors declare no conflicts of interest.

- *Acknowledgement*

We hereby acknowledge the willingness and consent of the students to be used for this study.

REFERENCES

[1.] Ali, M., Smith, J., & Nguyen, T. (2021). Academic dishonesty and its implications for educational institutions: A review. Journal of Higher Education Policy, 18(2), 134-148.

[2.] Brown, P., & Davis, R. (2020). Challenges in maintaining accurate student attendance records in higher education. Journal of Educational Management, 27(3), 205-222.

[3.] Johnson, P. (2020). The efficacy of attendance tracking systems in educational institutions. Journal of Educational Administration, 45(4), 150-165.

[4.] Brownlee, J. (2016). Deep learning with Python: Develop deep learning models on Theano and TensorFlow using Keras. Machine Learning Mastery.

[5.] Chauhan, N. K., & Singh, K. (2018). A review on conventional machine learning vs. deep learning. In 2018 International Conference on Computing, Power and Communication Technologies (GUCON) (pp. 1-6). IEEE.

[6.] Daugman, J. (2017). Face and gesture recognition: Overview. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 675-676.

[7.] Divyarajsinh, N. P., & Mehta, B. B. (2014). Face recognition methods & applications. arXiv preprint arXiv:1403.0485.

[8.] Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., ... & Chen, T. (2018). Recent advances in convolutional neural networks. Pattern Recognition, 77, 354-377.

[9.] Guo, Y., & Zhang, L. (2017). One-shot face recognition by promoting underrepresented classes. arXiv preprint arXiv:1707.05574.

[10.] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).

[11.] Hollnagel, E. (2014). Human factors and safety: An integrated approach. CRC Press.

[12.] Islam, M. A. (2023). Comparative analysis of pre-trained deep learning models for image classification. Journal of Computer Vision Research, 15(2), 45-62.

[13.] Barnouti, H. B., Al-Dabbagh, S. S. M., & Matti, W. E. (2016). Face recognition: A literature review. International Journal of Applied Information Systems (IJAIS), 11(3), 9-19.

[14.] Jafri, R., & Arabnia, H. R. (2009). A survey of face recognition techniques. Journal of Information Processing Systems, 5(2), 41-68.

[15.] Jain, A. K., Ross, A., & Nandakumar, K. (2011). Handbook of face recognition. Springer.

[16.] Ketwongsa, W., Boonlue, S., & Kokaew, U. (2022). A new deep learning model for the classification of poisonous and edible mushrooms based on improved AlexNet convolutional neural network. Journal of Food Science and Technology, 59(3), 1120-1131.

[17.] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems (pp. 1097-1105).

[18.] Kritikou, K., Likas, A., & Verma, S. (2021). Face recognition: A survey of techniques and applications. Artificial Intelligence Review, 54(3), 1857-1887.

[19.] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26oct667

[20.] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[21.] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359.
[22.] Shah, D. H., Shah, J. S., & Shah, T. V. (2014). The exploration of face recognition techniques. International Journal of Application or Innovation in Engineering and Management, 3(2), 2319-4847.
[23.] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[24.] Smith, J., & Lee, H. (2019). Enhancing security in educational institutions: A study on access control systems. Journal of Educational Technology, 16(3), 45-58.
[25.] Smith, J., & Miller, R. (2019). Urban surveillance and population management: A modern challenge. Journal of Security Studies, 12(3), 205-218.
[26.] Smith, L., & Lee, H. (2019). Ensuring safety and security in educational environments: The role of identification systems. International Journal of Educational Technology, 14(1), 45-58.
[27.] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).
[28.] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1701-1708).
[29.] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. ACM Computing Surveys (CSUR), 35(4), 399-458.