Journal of Information Technology Management
ISSN-OnLine: 2980-7972
University of Tehran

An Intelligent Hybrid Feature-Engineering Approach for Covariant Face Recognition Using Deep Learning

Pavitha U S*

*Corresponding author, Research Scholar (Part Time), Research Centre - Department of Electronics and Communication Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, India; Assistant Professor, Department of Electronics and Communication Engineering - M. S. Ramaiah Institute of Technology, Bengaluru, India. E-mail: Pavitha@msrit.edu

Suma K V

Associate Professor, Department of Electronics and Communication Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, India; Affiliated to Visvesvaraya Technological University, Belagavi-590018, Karnataka, India. E-mail: Sumakv@msrit.edu

Journal of Information Technology Management, 2025, Vol. 18, Issue 1, pp. 102-122
Published by the University of Tehran, College of Management
doi: https://doi.org/10.22059/jitm.2026.106256
Article Type: Research Paper

Received: August 11, 2025
Received in revised form: September 06, 2025
Accepted: December 28, 2025
Published online: January 20, 2026

Abstract

Face recognition (FR) is a non-contact biometric method integral to national and social security. It is pivotal in sectors like security, healthcare, banking, and criminal identification. Various techniques are being developed, including appearance and hybrid approaches, which either target specific facial features or consider the whole face for effective image recognition. This study explores hybrid machine learning techniques and enhanced covariates related to FR. The suggested approach is examined from a number of input viewpoints, including illumination, position variation, facial emotions, occlusions, and aging, which resulted in the widespread use of FR systems. The Generative Adversarial Network (GAN) is used to multiply the image on the dataset for image augmentation purposes. The facial covariants are extracted with the help of a hybrid feature engineering technique, such as a voting classifier that includes algorithms like K Nearest Neighbour, Support Vector Machine, and Random Forest Algorithm (KNN-SVM-RF). The dimension redundancy is achieved with the help of a combination of Principal Component Analysis and Independent Component Analysis algorithms. The modified VGG-16 algorithm is used to predict the image with the covariant similarity percentage of a person. Experiments conducted on the CelebFaces Attributes (CelebA) Dataset and Celebrity Face Image Dataset demonstrate that the hybrid strategy

yields superior accuracy and robustness compared with CNN-only or classical approaches, achieving notable improvements in Precision, Recall, F1 score, and overall Accuracy.

Keywords: Face Recognition, Face Covariant, Voting Classifier, CelebA Dataset, Modified VGG-16 Algorithm, True Successive Rate, False Acceptance Rate, False Rejection Rate.

Introduction

Face Recognition (FR) is a non-contact biometric identification technology that recognizes or verifies individuals by analyzing unique facial characteristics extracted from digital images or video streams. Due to its reliability, convenience, and user-friendly nature, FR has become a fundamental component in national security, surveillance systems, healthcare monitoring, banking authentication, border control, and criminal identification. Unlike traditional biometric methods such as fingerprint or iris scanning, face recognition does not require physical contact, making it highly suitable for real-time and large-scale applications. Recent research trends emphasize hybrid and intelligent frameworks to improve facial detection and classification performance under complex environmental conditions. In this context, enhanced hybrid facial emotion detection and classification techniques have demonstrated that integrating multiple learning strategies can significantly improve recognition robustness and accuracy in unconstrained environments (Ajlouni et al., 2025). The working mechanism of a typical FR system consists of sequential stages, including face detection, preprocessing, feature extraction, dimensionality reduction, and classification. Initially, face detection algorithms identify and localize facial regions within an image using computer vision sensors and image processing techniques. After detection, preprocessing operations such as normalization, illumination correction, and alignment are applied to standardize facial inputs. Feature extraction methods then generate discriminative representations of facial landmarks, textures, or deep embeddings. Traditional machine learning models and deep convolutional neural networks (CNNs) are widely used for this purpose. Vision-based detection and recognition algorithms have significantly improved automated identification accuracy in digital image environments (Lu et al., 2021). Moreover, advanced architectures such as hyper-attentive multimodal transformers have further enhanced robustness by capturing contextual and spatial dependencies across facial regions, enabling real-time and reliable facial expression recognition even under pose variations and occlusions (Tagmatova et al., 2025).

Despite substantial progress, FR systems still face critical challenges that limit their effectiveness in real-world deployments. Variations in illumination, facial expressions, aging, occlusion (e.g., masks or glasses), and pose changes significantly affect recognition performance. Additionally, identity-related threats, including spoofing attacks, adversarial manipulation, and presentation attacks, compromise system reliability and security. Comprehensive surveys indicate that handling face identity threats remains a major research

challenge, requiring robust feature representation and secure classification mechanisms (Rusia et al., 2023). Another limitation is the dependency of deep learning models on large-scale labeled datasets. Insufficient data diversity can lead to overfitting and poor generalization. To address data scarcity, Generative Adversarial Networks (GANs) have been introduced as powerful data augmentation tools capable of generating realistic synthetic facial images. GAN-based models have shown promising results in enhancing dataset variability; however, challenges such as training instability, mode collapse, and convergence complexity still exist (Gui et al., 2021). Motivated by these challenges, the present study proposes a hybrid machine learning framework that integrates enhanced facial covariates to improve recognition accuracy and robustness. The proposed system incorporates GAN-based image augmentation to increase dataset diversity and strengthen generalization capability. For feature engineering, a voting classifier is employed that combines K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Random Forest (RF) algorithms to leverage the strengths of multiple classifiers. To reduce dimensional redundancy and enhance computational efficiency, Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are integrated for optimal feature selection and transformation. Furthermore, a modified VGG-16 deep convolutional architecture is utilized to predict facial similarity percentages and perform final classification. By combining classical machine learning and deep learning paradigms, the proposed hybrid approach aims to achieve superior precision, recall, F1-score, and overall accuracy compared to standalone CNN or traditional models, thereby providing a more reliable and secure face recognition framework suitable for real-world applications.

- To put in place a robust face recognition system that can adapt to variations in lighting, posture, facial emotions, occlusions, and ageing.
- To integrate hybrid machine learning techniques that blend classical techniques (SVM, RF, KNN) and deep learning (CNN, Modified VGG-16).
- To improve recognition accuracy by extracting the facial covariants using the sophisticated feature engineering technique.
- To develop a feature selection method using Independent Component and Principal Component Analysis (ICA-PCA).
- To verify the improvements in the performance metrics, such as TSR, FAR, FRR, and accuracy, by testing the novelty system on the CelebA dataset.

The dataset and method used in the work to support the suggested work using performance measures are mentioned below.

- Hybrid Algorithm Development: Suggested a hybrid approach of integrating CNN, SVM, and the Random Forest to achieve higher levels of face recognition accuracy.

- Improvement of Feature Engineering: Applied voting classifier strategy (SVM + RF + KNN) to enhance covariant-based feature engineering.
- Dimensionality Reduction: A joint PCA + ICA model has also been introduced to reduce the number of redundant features, but still retain discriminative details.
- Adjusted VGG-16 Model: The VGG-16 architecture has been adjusted to display faces based on covariant similarity percentage to provide better classification.
- Strong Analysis: Carried out massive testing on the CelebA data and proved to be more robust and accurate than CNN-only and conventional FR.
- Performance Enhancement: It has obtained dramatic success in terms of TSR, FAR, FRR, and the overall Accuracy, showing the effectiveness of the hybrid strategy.

The remaining paper is organized in the following way: Section 2 presents the literature review of the studies related to the area of face recognition, including the current methodology, the algorithms, and the outcomes obtained. Section 3 describes the methodology that has been used in this project, including the development environment and data augmentation techniques. Section 4 provides the results achieved on the system and performance, accuracy discussion. Section 5 ends the paper with a summary of the findings and the description of the potential future work directions, including model optimization, real-time performance improvement, and system scalability. Lastly, the References section lists all sources and research materials mentioned throughout the paper.

Literature Review

The latest developments in Face Recognition (FR) studies show that there has been a major shift in the traditional handcrafted feature-based to deep learning-based and hybrid intelligent systems. Conventional early deep learning-based detection and recognition systems focused on the convolutional neural networks (CNN) due to the robustness of the feature extraction. As an example, face recognition detection models based on deep learning were proposed to improve the accuracy of automated identification by using layered convolutional structures; nevertheless, these models are commonly sensitive to changes in illumination and pose (Tao et al., 2024). Metaheuristic-based neural frameworks like the Grey Wolf Optimization Cuckoo (GWO-Cuckoo) optimized neural networks along with Minimum Redundancy Maximum Relevance (MRMR) feature selection have been proposed to achieve better optimization efficiency, which involves compressed hybrid domain fusion using multi-stage optimization at the cost of higher computational cost (Shanmugam et al., 2024).

Density estimation methods with manifold learning have been studied to refine face recognition in preserving the intrinsic geometric structure of the face data distribution;

however, these approaches encounter the problem of scalability in large-scale real-life datasets (Ge et al., 2024). Deep learning-based hybrid image augmentation methods have been suggested to enhance user/environment-free recognition systems, and they have shown to be more robust but less generalized with highly unconstrained face databases (Awaluddin et al., 2024). Wide use of artificial Neural Networks (ANN), machine learning algorithms has also been employed to utilize spatial information to better feature discrimination, but these models require quality feature engineering and large labeled datasets (Goel et al., 2023). Transformer-based architectures have just become strong competitors to CNN models. Convolutional transformer designs Efficient Convolutional Transformer (ECTFormer: Efficient Convolutional Transformer) designs are convolutional-layer-based image recognition designs that utilize attention mechanisms to win the race on contextual representation and demand high computational resource consumption (Sa et al., 2025). The wider scans of the algorithms in Machine Learning (ML) emphasize their versatility and their ability to be used in various areas; the problem of overfitting, bias, and imbalance are listed among such problems (Sarker et al., 2021). Hybrid offline-and-online reinforcement learning models, including dynamics-aware hybrid reinforcement learning, are more adaptable to dynamic settings, but are only applicable to simpler tasks that can be trained in face recognition (Niu et al., 2022). A hybrid face recognition model combining Haar Cascade (HC), Softmax classifier, and Convolutional Neural Network (CNN) has been suggested to facilitate high detection and classification performance, yet they frequently suffer because of the constraints of occlusion and real-time system application (Singh et al., 2024). More advanced Principal Component Analysis (PCA: Principal Component Analysis) has been used in access control systems to reduce dimensionality and represent the facial features efficiently, although PCA-based systems can be discriminative to nonlinear data (Lin et al., 2025). Gender-based age-invariant face recognition methods have enhanced resistance to aging variations, but their effectiveness is likely to decline when gender classification is faulty (Nayak and Indiramma, 2022).

Other indirect links to the state of the art in multi-object tracking come through object detection-based modeling frameworks originally designed to deal with pedestrian behavior and social force analysis, but not directly optimized to achieve fine-grained facial identity recognition (Yang et al., 2025). The capability to infer subtle hereditary similarities of faces among siblings and the use of CNN architectures have proven to be effective deep learning-based sibling identification models, yet they require a large amount of training data and high computing power (Goel et al., 2021). Higher-order pooling representation. Attention-based neural networks have improved the spatial feature discrimination in image recognition tasks, but attention-based mechanisms increase the cost of training and parameter size (Kong et al., 2022). Facial emotion recognition Benchmark experiments comparing deep networks to facial emotion recognition in unconstrained conditions indicate that although CNN models in this instance are highly accurate, the accuracy decreases significantly in cases of occlusion, varying illumination, and pose. Face Recognition (XAI-FR: Explainable Artificial

Intelligence Face Recognition) frameworks that enhance the transparency and trustworthiness of deep neural models are based on explainable artificial intelligence (explainable AI) frameworks, but they add computational cost and slow inference time (Rajpal et al., 2023). Better robustness of enhanced face recognition in crowded scenes with 2D/3D features with Parallel Hybrid Convolutional Neural Network-Recurrent Neural Network (CNN-RNN) and Stacked Autoencoder (SAE: Stacked Autoencoder) has demonstrated enhanced robustness, although multimodal features can be integrated into the system, which makes the system more complex (Elloumi et al., 2025).

Deep Convolutional Generative Adversarial Networks (DCGAN: Deep Convolutional Generative Adversarial Network) are data augmentation methods that have shown to be effective in enhancing the accuracy of recognition by generating synthetic training samples, but GAN-based models are unstable in adversarial training (Wu et al., 2020). The invention of Very Deep Convolutional Networks (Visual Geometry Group 16-layer Network VGG-16: Visual Geometry Group 16-layer Network) created a basic architecture of large-scale image recognition that offers deep hierarchical feature extractions, yet needs vast amounts of computational power and memory (Simonyan and Zisserman, 2014). The later studies on CNN models of face recognition established that deep learning is much more effective than conventional techniques; however, the generalization of models in cross-dataset conditions is still an issue (Hu et al., 2015). Deep-learning-based methodologies of face recognition on the basis of texture have enhanced the representation of texture, but hand-crafted descriptors can be vulnerable to noise and variability in the environment (Sekhar et al., 2025). Taken together, the literature shows that, despite the fact that deep learning and hybrid structures have a profound improvement on face recognition performance, crucial issues are still present, such as illumination, occlusion, aging, adversarial, computational, and scalability. These shortcomings drive the upscale design of superior hybrid solutions combining streamlined feature engineering, dimensionality reduction, data augmentation, and deep convolutional architectures to gain a better level of accuracy, reliability, and real-world implementation.

Methodology

A deep learning and ensemble-based classification model's sequential workflow is presented in the scheme. To improve model resilience, this is accomplished by first enriching the dataset by expanding and diversifying the training data. The important features are then chosen using Principal Component Analysis (PCA), which eliminates unnecessary information and highlights only the important aspects to improve accuracy and computation efficiency. After that, an ensemble method that combines KNN-SVM-RF is used to extract features. These algorithms complement one another and increase the dependability of the extracted features. The collected and processed features are then fed into a modified VGG-16

CNN model for classification, where the deep learning framework analyzes the data to find patterns and improve accuracy.

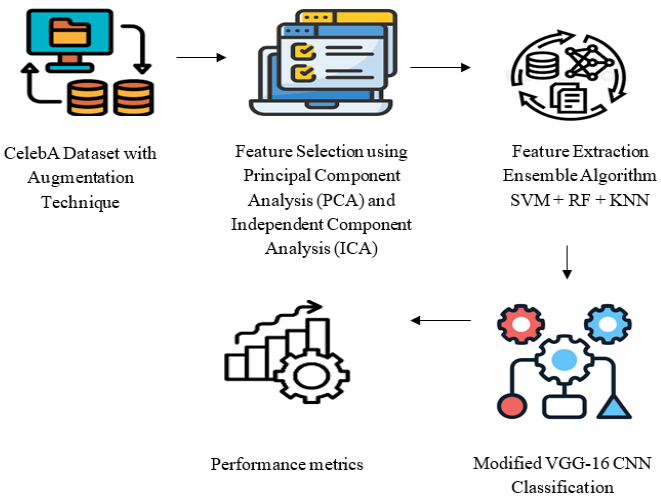

Figure 1. Proposed Work's overall flow diagram

Lastly, the overall performance of the framework is measured in terms of performance metrics, meaning that the accuracy, precision, recall, and other important measures in the system are measured accordingly. The combined strategy is useful in the optimization of features, ensemble learning, and deep CNN classification to obtain high-performance results.

Dataset details

This dataset is outstanding for training and evaluating face detection models, particularly for recognizing facial characteristics like brown hair, smiles, and glasses. Large position variations, background clutter, and a multiplicity of persons are all included in the images, which are backed by numerous photos and detailed annotations. This dataset, a carefully selected subset of the CelebFaces Attributes (CelebA) Dataset, is meant for deep learning applications such as picture synthesis and facial identification. There are 50,000 well-known face images from different identities that display a range of facial characteristics, environments, and positions.

Data Augmentation technique using GAN

Even though GAN can make realistic and distinctive images, it needs a huge amount of training data. A common technique for increasing data is data augmentation, which involves a variety of augmentation techniques. However, augmentation leaks, in which unwanted distortion effects from enhanced photos show in images produced by the GAN, can result from using augmented data to train GANs. The discriminator and the generator make up a GAN [Wu, Q., Chen, Y., 2020]. The discriminator measures the variance among the distribution of produced samples and genuine samples to decide if the sample is real or

fraudulent. The generator creates pictures from a given noise z. An initial GAN's objective function for generator G, discriminator D, random noise z, and actual data x is:

$$min_G\ max_D\ V\ (D,G) = E_{x\sim Pdata(x)}[logD\ (x)] + E_{z\sim p(z)}[\log(1 - D\ G(Z))] \quad (1)$$

Here

$Pdata(x)$ - depicts the distribution of actual samples

$p(z)$ - shows how the produced samples are distributed.

The chance that D's inputs are real samples is represented by $D(x)$. The GAN training procedure makes use of the max-min. Throughout the training phase, G is optimized after D is fixed, and D is optimized after G is fixed.

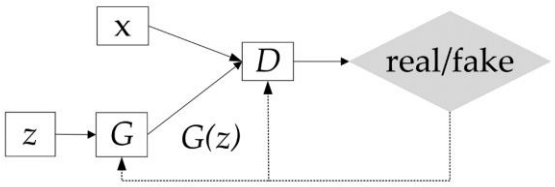

Figure 2. Architecture of Deep Convolutional GAN

The generator (G) and discriminator (D) make up the majority of the Deep Convolutional GAN. To get to the equilibrium, the discriminator and generator compete with one another. The DCGAN structure is seen in Figure 2. Putting the hidden variable z into the generator yields $G(z)$, which is typically a random noise focus to a Gaussian distribution. Once the discriminator has $G(z)$, it compares it to the actual data, determines whether it is true or not, and then feeds the result back to the generator. The two-classification issue is comparable to the discriminator-optimized procedure.

Feature Selection using Principal Component Analysis (PCA)

Facial recognition with feature selection based on PCA is a good method of simplifying the data by eliminating irrelevant data and maintaining the key attributes of the faces. Raw face images are also thousands of pixels, and thus, directly processing them is computationally intense and can cause redundancy. PCA is used to reduce the high-dimensional data to a smaller set of mutually independent features, known as the principal components, that maximize the variance in the data. These elements are some of the key facial patterns, including the eyes, nose, and mouth structure. PCA can help to maximize the accuracy of recognition, minimize noise, and boost processing efficiency by choosing the most important components.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (2)$$

In face recognition, equation (2) is the mean vector in PCA. In this case, x_i represents every sample of an image translated into a vector, and n is the number of training pictures. Adding all image vectors and dividing them by N, come to the mean or average appearance. This average face is a mean of all the images with some common features, which can be used as a reference. By removing this mean from each image (centering), differences in the data set are to be evaluated in reference to the mean structure of faces.

$$\emptyset_i = x_i - \mu \quad (3)$$

Equation (3) represents the centering step in PCA for face recognition. Here, x_i is the vector form of the x_i face image, and i^{th} is the mean face vector calculated from all training samples. By subtracting the mean vector from each image, we obtain the deviation vector $\emptyset_i$, which highlights the unique variations of that face compared to the average. This step removes common background information and ensures that the analysis focuses on differences among faces, making PCA more effective for feature extraction.

$$C = \frac{1}{N}\sum_{i=1}^{N} \emptyset_i \emptyset_i^T \quad (4)$$

Equation (4) defines the covariance matrix in PCA for face recognition. Here, $\emptyset_i$ is the centered image vector (obtained by subtracting the mean face). Multiplying $\emptyset_i \emptyset_i^T$ gives the outer product, which measures how features vary together. Summing these products over all samples and dividing by N produces the covariance matrix C. This matrix captures the relationships and variance patterns among facial features. It is essential because the next step in PCA uses this covariance matrix to compute eigenvalues and eigenvectors, identifying the most informative facial components.

Independent Component Analysis

ICA is a feature selection method used in datasets to select the most statistically independent features to enhance the efficiency and accuracy of the model. ICA converts the original correlated variables to a new set of independent components, which depict underlying hidden factors in the data. Through these elements, one will be able to eliminate irrelevant and redundant elements, and preserve the most informative elements. This minimizes dimensions, increases interpretability, and minimizes computation complexity.

$$X = A.S \quad (5)$$

The ICA, equation (5), explains that observed data X is formed by mixing hidden independent source signals S through a mixing matrix A. Each column of X represents an observed feature that is actually a combination of several independent sources. Since both A

and S are unknown, ICA works by estimating an unmixing matrix W, recovering the original independent components. This process allows the separation of mixed signals into meaningful and independent features, which can then be used for dimensionality reduction and feature selection in machine learning.

$$X_c = X - \mu \tag{6}$$

Equation (6) represents the centering step in ICA or PCA. Here, X is the unique dataset, μ\muμ is the mean of each feature, and X_c is the centered data. Subtracting the mean ensures that each feature has a zero mean, which is important for simplifying covariance calculations and improving accuracy in extracting independent or principal components. By removing the mean, the data becomes normalized around the origin, allowing ICA to focus only on variations and dependencies between features rather than being biased by absolute values.

$$X_\omega = VD^{-\frac{1}{2}}V^T X_c \tag{7}$$

Equation (7) represents the whitening (sphering) step in ICA or PCA. Here, X_c is the mean-centered data, V is the matrix of eigenvectors of the covariance matrix, and DDD is the diagonal matrix of corresponding eigenvalues. The operation $D^{-\frac{1}{2}}$ scales the data so that each transformed feature has unit variance. Multiplying by V and V^T ensures that the features are uncorrelated. Thus, whitening transforms correlated variables into a new set of orthogonal features with zero mean and unit variance, preparing the data for extracting independent components in ICA.

Feature extraction using a voting classifier

The voting classifier for covariant feature extraction is used in the proposed work. The Ensemble methods can also be used to combine SVM with other classifiers, such as Random RF and KNN, to enhance predictive accuracy. First, the data is transformed into important features, which is usually done via PCA, which decreases the number of dimensions and preserves the key information. They are then input into classifiers, SVM, RF, and KNN, which each make their own predictions. The last class in a voting ensemble is determined either by majority voting or predicted average. A meta-model is used in stacking, whereby the outputs of base classifiers are given to a meta-model, like logistic regression, to produce the final prediction.

$$\hat{y} = mode\{h_1(x), h_2(x), \dots h_M(x)\} \tag{8}$$

Equation (8) describes the hard voting ensemble approach to machine learning. Here, $h_1(x), h_2(x), \dots h_M(x)$ are the predictions of M, an SVM-based, a random Forest-based, or a K-Nearest Neighbors-based base classifier, on a given input x_i. Voting is done by each classifier for the class, and the final predicted class y is the one that has the majority vote i.e.,

a class that has most of the classifiers. The method combines the strengths of several models, removes the bias of single models, and tends to enhance the general classification accuracy as the various model predictions are pooled together to produce a single robust output.

$$\hat{y} = arg_c^{max} \frac{1}{M} \sum_{i=1}^{M} P_i\,(y = c|x) \tag{9}$$

Equation (9) represents the soft voting ensemble method in machine learning. Here, $P_i(y = c|x)$ denotes the probability that the i^{th} base classifier assigns input x to class c, and M is the total number of classifiers, such as KNN-SVM-RF. The term $\frac{1}{M}\sum_{i=1}^{M} P_i\,(y = c|x)$ calculates the average predicted probability of class ccc across all classifiers. The final predicted class $\hat{y}$ is the one with the highest average probability. This method is more flexible than hard voting, as it considers confidence levels of classifiers, often improving overall accuracy.

$$z = g(h_1(x), h_2(x), \ldots h_M(x)) \tag{10}$$

Equation (10) represents the stacking ensemble method in machine learning. Here, $h_1(x), h_2(x), \ldots h_M$ are the predictions of M base classifiers, such as SVM, RF, and KNN, for an input x. The outputs of these classifiers are combined as an input vector z for a meta-model g, often a logistic regression or another classifier, which learns how to best weight and combine the base predictions. This approach leverages the strengths of multiple classifiers, capturing complex relationships between their predictions, and typically results in more accurate and robust final predictions than individual models or simple voting.

$$\hat{y} = MetaModel(z) \tag{11}$$

Equation (11) represents the final prediction step in a stacking ensemble. Here, z is the combined output vector from multiple base classifiers, such as SVM, Random Forest, and K-Nearest Neighbors, for a given input x. The meta-model, often a logistic regression or another classifier, learns how to optimally weight and combine these base predictions to improve overall accuracy. By feeding z into the meta-model, the ensemble generates a single robust prediction. This approach captures the strengths of individual models, reduces errors from weaker classifiers, and typically outperforms single-model or simple voting methods.

Modified VGG-16 Convolutional Neural Network (CNN)

The 16-layer CNN deep learning model is called modified VGG16 (Visual Geometry Group), or VGG Net. The VGG-16, which was introduced by [Simonyan, K. & Zisserman, A, 2015], has garnered a lot of interest due to its remarkable performance in image classification tasks. The VGG-16 architecture consists of 16 layers in total, including 3 fully connected layers and 13 convolutional layers.

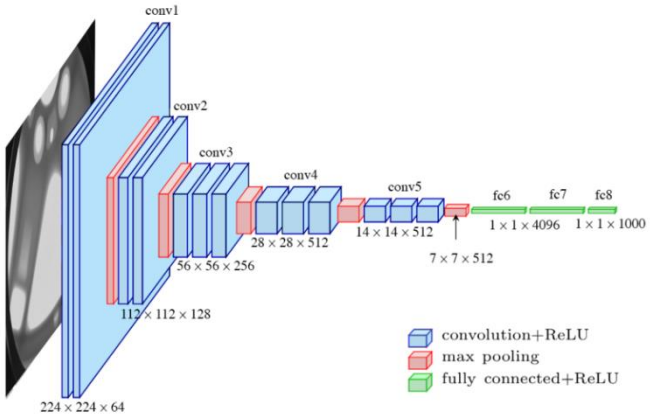

Figure 3. Proposed architecture of the Modified VGG-16 classifier algorithm

A well-liked deep learning approach for image categorisation is the Modified VGG-16 Convolutional Neural Network (CNN). 13 convolutional layers and 3 fully connected layers make up the default VGG-16 network, which aims to have hierarchical picture characteristics. In order to tailor the network to a given classification goal, the modified version permits changing or tweaking certain layers. Examples of such changes include lowering the number of completely connected layers, changing the number of output neurons to the target classes, or adding dropout layers to prevent overfitting.

$$F_{i,j}^{k} = \sum_{m} \sum_{n} I_i + m, j + n \,.\, W_{m,n}^{k} + b^{k} \tag{12}$$

Equation (12) represents the convolution operation in a CNN, specifically in VGG-16. Here, III is the input image or feature map, $W^k$is the convolution kernel (filter) for the k^{th}feature map, and b^k is the bias term. The double summation creates a single value $F_{i,j}^{k}$ at position (i,j) in the feature map of output by sliding the filter over the input, multiplying related elements, and adding the results. Local patterns like edges and textures are captured by this method. The network captures hierarchical characteristics from low-level to high-level by using several filters, which is crucial for precise image classification.

$$A_{i,j}^{k} = \max(0, F_{i,j}^{k}) \tag{13}$$

Equation (13) represents the ReLU activation function in a CNN. Here, $F_{i,j}^{k}$is the value at position (i,j) of the k^{th} feature map obtained from the convolution layer. By producing the maximum of zero and $F_{i,j}^{k}$, ReLU converts this value, essentially setting all negative values to zero while maintaining positive values. This gives the network non-linearity, which enables it to extract intricate patterns from the input. The network can better simulate complex structures in images for precise classification by capturing non-linear correlations between pixels by using ReLU over all feature maps.

$$P_{i,j}^{k} = \max\left\{A_{m,n}^{k}\right\}, (m,n) \in pooling\ window \quad (14)$$

The CNN's max pooling process is represented by equation (14). In this case, the pooling window specifies a tiny area of the feature map, while $A_{m,n}^{k}$are the activated values from the ReLU layer. Max pooling creates $P_{i,j}^{k}$ at position (i,j) in the pooled map by sliding this window across the input feature map and choosing the maximum value inside each region. This preserves the most important properties while reducing the spatial dimensions, which lowers the computational cost and memory requirements. Max pooling also offers translation invariance by concentrating on dominating activations, enabling the network to identify features even if they slightly change in the image.

Results

The Face recognition project was created and run on Windows 11 on the platform of Thonny IDE. The machine, which was used to run the system, had a processor of Intel Core i5 13th generation, RAM memory of 16 GB, as well as a graphics card of 16 GB, which was made by Nvidia. This hardware implementation allowed a high enough number of processing units to run face recognition tasks without any difficulties. The selection of Thonny IDE provided an easy and convenient code writing, debugging, and testing system and was appropriate in terms of application and testing face recognition algorithms.

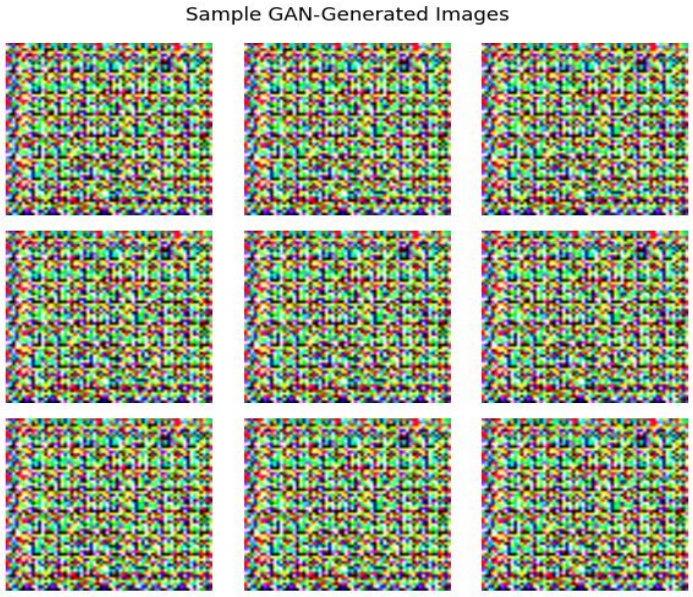

Figure 4. Sample output of DCGAN applied in celebA dataset

The DCGAN trained on the CelebA dataset is shown in Figure 4. DCGAN is a potent design that uses massive quantities of data to understand features that underlie facial anatomy in order to produce realistic images. In order to improve diversity and the network's overall capacity for generalisation, the formed fake images are enhanced using a variety of augmentation techniques, such as rotation, flipping, and scaling operations. The DCGAN is gradually improving its ability to produce visually comprehensible images by reducing the gap between actual and synthetic samples, even when the outputs are still noisy in the early stages of training or a small number of epochs. This experiment shows that artificial samples of the CelebA dataset can be produced using both augmentation and adversarial learning techniques.

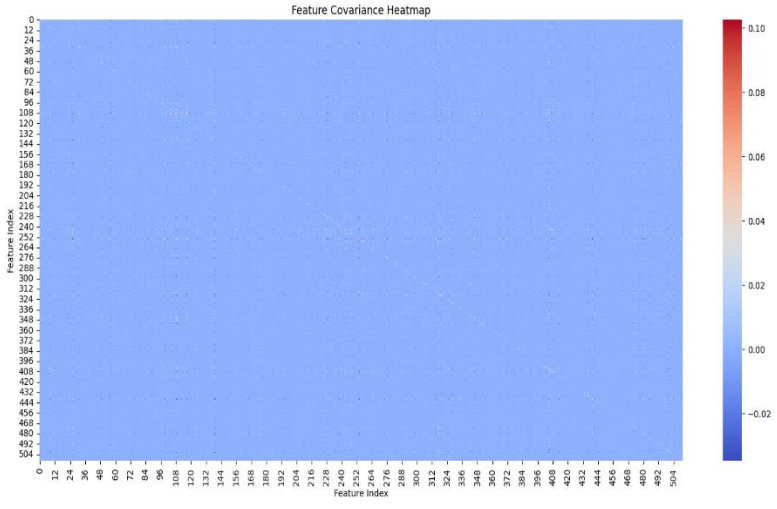

Figure 5. Feature covariance heatmap generated by the PCA algorithm

Figure 5 above shows the Feature Covariance Heatmap produced using the CelebA dataset, which is widely used for facial attribute analysis and recognition tasks. Here, the dataset's covariant facial traits were identified and determined using Principal Component Analysis (PCA). An image that shows how different face traits or main components relate to the covariance is called a heatmap. The diagonal dominance of the plot illustrates the dimensionality reduction and feature decorrelation effects of PCA by demonstrating that most features are not connected with one another after the PCA transformation. The range of the color scale between blue and red indicates the level of covariance, with closer to blue (high level of covariance) indicating less linear relationship and the more towards red (strong linear relationship). Generally, this visualization supports the idea that PCA is a great way of converting the high-dimensional face data of CelebA into a series of statistically independent features that can be further used to analyze or classify faces.

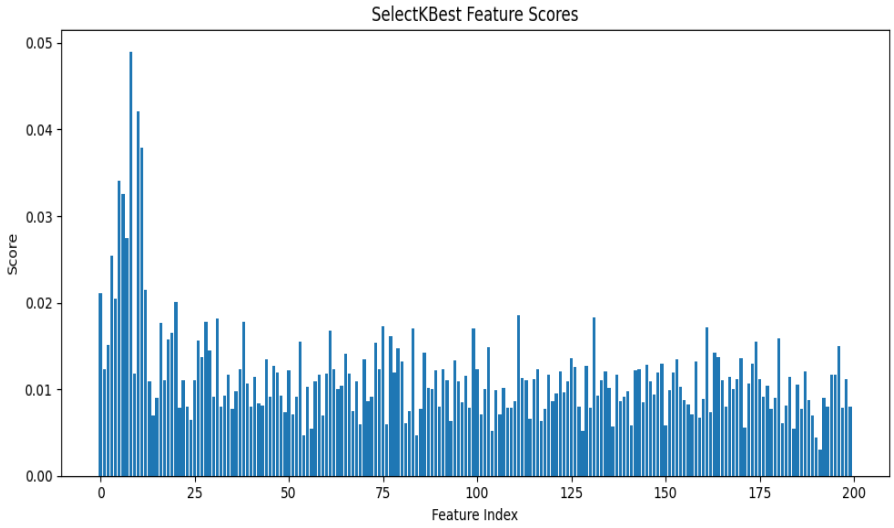

Figure 6. SelectKBest Feature scores using a hybrid feature selection technique

The SelectKBest- feature selection scores for the CelebA dataset using a hybrid feature selection methodology that combines PCA and ICA are displayed in Figure 6. The height of a bar indicates the statistical significance of a feature's contribution to the model's performance; the bars are linked to the feature indices. Because it helps to extract the most informative components that are independent of one another from the high-dimensional data of face attributes in CelebA, the hybrid PCAICA approach is effective in dimensionality reduction. The fact that the early feature indices have higher scores indicates that the initially extracted components still have the greatest discriminative potential, with the latter features having less of an impact. This form of selective filtering maintains only the most pertinent visual and statistical regularities to be used in future learning and to optimize computational efficiency and classification accuracy in face attribute recognition tasks.

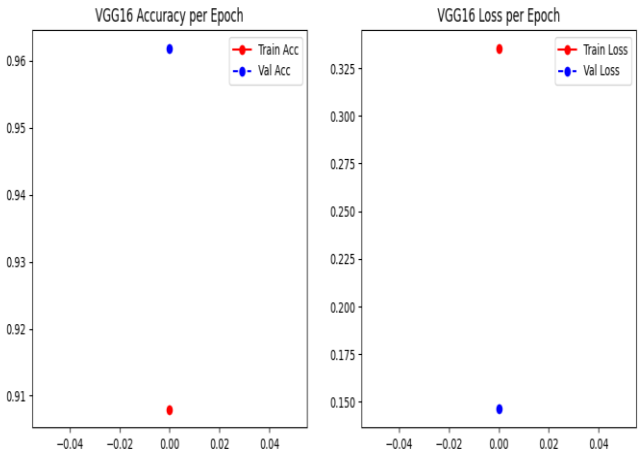

Figure 7. Accuracy of Training and Validation of the proposed VGG16 model

Figure 7 shows the training and validation performance of a modified VGG16 model with respect to epochs. The VGG16 Accuracy per Epoch graph on the left displays the model's training accuracy and validation accuracy, while the VGG16 Loss per Epoch graph on the right displays the model's training loss and validation loss. With a training accuracy of roughly 0.91 and a validation accuracy of roughly 0.96, the modified VGG16 architecture demonstrates steady learning features and high generalisation abilities. In line with this, training loss is close to 0.32, and validation loss is close to 0.15, indicating efficient optimization and less overfitting. These findings verify that the updated VGG16 model can utilize discriminative features with only a few epochs and that it will guarantee a higher degree of convergence and higher quality of validation than the training period.

The confusion matrix represented in Figure 8 belongs to a VGG16 model, used in a face recognition task, with two classes, namely, "Celebrity Faces Dataset" and "img_align_celeba." The matrix indicates the count of the images of each of the true classes that were predicted as that specific class. In the true class of the "Celebrity Faces Dataset," the model made the right prediction on 328 images and wrongly predicted 45 images as the image "img_align_celeba." The algorithm properly identified 2,981 photos in the true class (img align celeba), misclassifying only 7 images as belonging to the "Celebrity Faces Dataset." Except for the "img_align_celeba" class, which records a great degree of accuracy and dependability in making predictions regarding the said classification, this demonstrates that the model is highly effective in differentiating between the two classes with minimal misclassifications. Overall, the confusion matrix suggests that the VGG16 model is helpful in the given binary classification task.

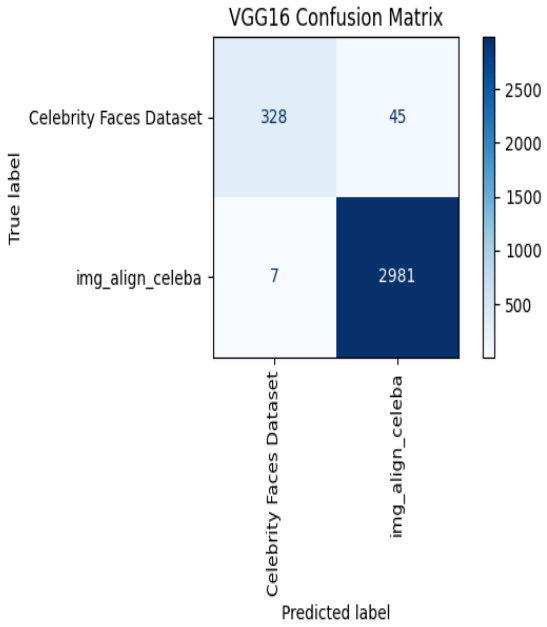

Figure 8. Confusion matrix of the proposed modified VGG-16 algorithm for face recognition

Figure 9. True and Prediction of the test image randomly selected from the celebA dataset

A modified VGG-16 model's ability to improve binary label prediction in the CelebA dataset's images is demonstrated in Figure 9. The challenge in this instance is to classify each image of a celebrity's face into one of two groups, which are represented by the binary labels. The model's predictions are shown in the figures, where each image is represented by either T:img_align_celeba or P:img_align_celeba, which stand for the true and expected labels, respectively. The model can be successfully trained using the CelebA dataset and categorize the photos into these two categories with high accuracy using the updated VGG-16 framework, as seen by the similarity of the actual and estimated labels of the instances.

Figure 10. Model's similarity prediction depends in face covariant

Figure 10 will illustrate how similarity prediction is done in the CelebA dataset with a binary label classification model, in which a given test image is labeled and then compared with other images in the dataset to locate the most similar image. The real label (T) and the expected label (P) of the test image are both the same and equal to the image of "img_align_celeba," and the system retrieves the closest label face. This similarity prediction is largely functional due to covariant properties of the face, including geometric structure, facial characteristics, and patterns, which are the inputs of the model in recognizing likeness. Covariant descriptors provide a good representation of these variations, and this has allowed the model to match faces not only by identity but also by overall facial similarity, which includes both the visible characteristics and intrinsic covariance-based characteristics of face recognition, which are critical in performing robust face matching.

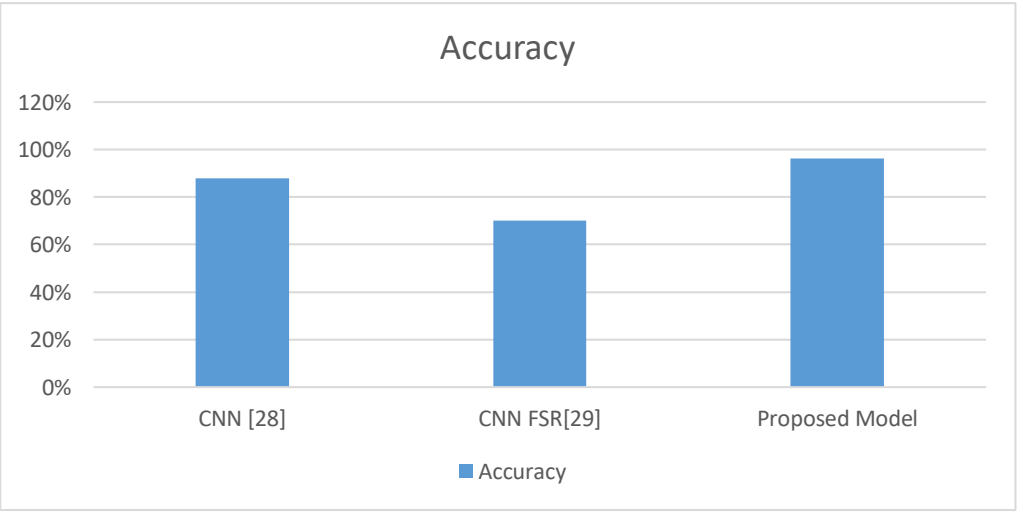

Figure 11. Accuracy comparison of the proposed modified VGG16 model with the conventional model

Based on the accuracy percentages, Figure 11's "Accuracy" section compares the three models' accuracy. CNN [Hu, Guosheng, et al, 2015], CNN FSR [Sekhar, J. C., et al, 2025], and the Proposed Model are these three. The Proposed Model is the most accurate, with an accuracy of roughly 100%, indicating that it is superior to the others. Additionally, effective is the CNN [28] model is effective, which achieves an accuracy of about 90%. Conversely, with an accuracy of about 70%, the CNN FSR [29] model performs the worst out of the three. This comparison demonstrates unequivocally that the Proposed Model is superior to CNN [Hu, Guosheng, et al, 2015] and CNN FSR [29], making it the most accurate technique in this examination.

Conclusion

This paper proposed a powerful and effective face recognition (FR) framework based on a modified VGG-16 model with a combination of hybrid feature extraction and classification methods. Using the advantages of CNN and classical voting classifier that comprised of SVM, RF, and KNN, the system overcame the typical facial covariates such as posing, light, covering, facial expression, and age. In order to expand performance and lower computational costs, the feature space was further optimised through dimensionality reduction based on a mix of PCA and ICA. In comparison to conventional CNN-based models and classical models, the proposed method achieved 96.19 of the CelebA dataset with such high accuracy. The model will be enhanced in the future by using transformer-based architectures and attention techniques to generalise to different and unrestricted contexts. Additionally, investigating adversarial resistance and privacy-saving techniques, as well as extending the model to enable real-time face recognition with a lightweight deployment on edge devices, would increase the system's relevance in high-stakes security challenges. To increase the accuracy and dependability of recognition, a combination of multi-modal biometrics, such as facial traits with voice or gait, might be taken into consideration.

Acknowledgments

The authors extend their sincere gratitude to the Research Centre, Department of Electronics and Communication Engineering, M. S. Ramaiah Institute of Technology, Bengaluru, for their constant support and research facilities. The affiliation with Visvesvaraya Technological University, Belagavi, has further enabled a conducive environment for completing this work successfully.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ajlouni, N., Özyavaş, A., Ajlouni, F., Takaoğlu, F., & Takaoğlu, M. (2025). Enhanced hybrid facial emotion detection & classification. *Franklin Open*, *10*, 100200.

Awaluddin, B. A., Chao, C. T., & Chiou, J. S. (2024). A hybrid image augmentation technique for user-and environment-independent hand gesture recognition based on deep learning. *Mathematics*, *12*(9), 1393.

Elloumi, S., Bahroun, S., Yahia, S. B., & Kaddes, M. (2025). Enhanced Face Recognition in Crowded Environments with 2D/3D Features and Parallel Hybrid CNN-RNN Architecture with Stacked Auto-Encoder. *Big Data and Cognitive Computing*, *9*(8), 191.

Ge, H., Zhu, Z., Ouyang, J., Ashraf, M. A., Qiu, Z., & Ibrahim, U. M. (2024). Integration of manifold learning and density estimation for fine-tuned face recognition. *Symmetry*, *16*(6), 765.

Goel, A., Goel, A. K., & Kumar, A. (2023). The role of artificial neural network and machine learning in utilizing spatial information. *Spatial Information Research*, *31*(3), 275-285.

Goel, R., Mehmood, I., & Ugail, H. (2021). A study of deep learning-based face recognition models for sibling identification. *Sensors*, *21*(15), 5068.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE transactions on knowledge and data engineering*, *35*(4), 3313-3332.

Hu, G., Yang, Y., Yi, D., Kittler, J., Christmas, W., Li, S. Z., & Hospedales, T. (2015). When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition. In *Proceedings of the IEEE international conference on computer vision workshops* (pp. 142-150).

Kong, J., Wang, H., Yang, C., Jin, X., Zuo, M., & Zhang, X. (2022). A spatial feature-enhanced attention neural network with high-order pooling representation for application in pest and disease recognition. *Agriculture*, *12*(4), 500.

Lin, N., Ding, Y., & Tan, Y. (2025). Optimization design and application of library face recognition access control system based on improved PCA. *Plos one*, *20*(1), e0313415.

Lu, D., & Yan, L. (2021). [Retracted] face detection and recognition algorithm in digital image based on computer vision sensor. *Journal of Sensors*, *2021*(1), 4796768.

Nayak, J. S., & Indiramma, M. (2022). An approach to enhance age invariant face recognition performance based on gender classification. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 5183-5191.

Niu, H., Qiu, Y., Li, M., Zhou, G., Hu, J., & Zhan, X. (2022). When to trust your simulator: Dynamics-aware hybrid offline-and-online reinforcement learning. *Advances in Neural Information Processing Systems*, *35*, 36599-36612.

Rajpal, A., Sehra, K., Bagri, R., & Sikka, P. (2023). Xai-fr: explainable ai-based face recognition using deep neural networks. *Wireless Personal Communications*, *129*(1), 663-680.

Rusia, M. K., & Singh, D. K. (2023). A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. *Multimedia Tools and Applications*, *82*(2), 1669-1748.

Sa, J., Ryu, J., & Kim, H. (2025). ECTFormer: An efficient Conv-Transformer model design for image recognition. *Pattern Recognition*, *159*, 111092.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 1-21.

Sekhar, J. C., Josephson, P. J., Chinnasamy, A., Maheswari, M., Sankar, S., & Kalangi, R. R. (2025). Automated face recognition using deep learning technique and center symmetric multivariant local binary pattern. *Neural Computing and Applications*, *37*(1), 263-281.

Shanmugam, M., Viswanatha, V. M., & Raja, K. B. (2024). Precise Face Recognition through GWO-Cuckoo Optimized Neural Networks and MRMR Feature Selection from Compressed Hybrid Domain Fusion. *International Journal of Intelligent Engineering & Systems*, *17*(1).

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, P., Kansal, M., Singh, R., Kumar, S., & Sen, C. (2024). A hybrid approach based on Haar Cascade, Softmax, and CNN for human face recognition: a hybrid approach for human face recognition. *Journal of Scientific & Industrial Research (JSIR)*, *83*(4), 414-423.

Tagmatova, Z., Umirzakova, S., Kutlimuratov, A., Abdusalomov, A., & Im Cho, Y. (2025). A hyper-attentive multimodal transformer for real-time and robust facial expression recognition. *Applied Sciences*, *15*(13), 7100.

Tao, Z. (2024). Face recognition detection based on deep learning. *Journal of Combinatorial Mathematics and Combinatorial Computing*, *127*, 2651-2658.

Wu, Q., Chen, Y., & Meng, J. (2020). DCGAN-based data augmentation for tomato leaf disease identification. *IEEE access*, *8*, 98716-98728.

Yang, F., Liu, R., & Zhu, D. (2025). Pedestrian dynamics modeling and social force analysis based on object detection. *Frontiers in Physics*, *13*, 1579280.

Bibliographic information of this paper for citing:

U S, Pavitha & K V, Suma (2026). An Intelligent Hybrid Feature-Engineering Approach for Covariant Face Recognition Using Deep Learning. *Journal of Information Technology Management,* 18 (1), 102-122. https://doi.org/10.22059/jitm.2026.1062556