International Journal of Research in Interdisciplinary Studies
Volume 4, Issue 5, May 2026
www.ijris.com | E-ISSN: 2584-1017

AI Powered Enhanced Facial Recognition System for Law Enforcement Operations

T. Rajan Babu[1], R. G. Suresh Kumar[2*], Varun Deleep Kumar[3], M. Premanand[3], C. Dhivagar[3], Vithyasaran[3]

[1]*Assistant Professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[2]*Professor & HoD, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*
[3]*B.Tech. Student, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry, India*

Abstract—Facial recognition technology has emerged as a critical component in modern surveillance and law enforcement systems. With the rapid advancement of Artificial Intelligence (AI) and Deep Learning (DL), automated identification systems have significantly improved in terms of accuracy and efficiency. However, traditional deep learning models often struggle to maintain consistent performance under real-world conditions such as illumination variation, pose differences, occlusion, and background noise. These challenges limit their effectiveness in practical deployment scenarios. To address these limitations, this work proposes an advanced facial recognition system based on a hybrid deep learning framework that integrates VGG19 and ResNet-101 for robust feature extraction. VGG19 captures fine-grained facial details such as edges and textures, while ResNet-101 extracts deeper semantic features using residual learning. The combined feature representation enhances the system's ability to distinguish individuals under complex conditions. Furthermore, a Support Vector Machine (SVM) classifier is employed to improve decision boundaries and classification accuracy in high-dimensional feature space. The proposed system is designed for real-time applications such as criminal identification and missing person detection using surveillance data. Experimental results demonstrate improved accuracy, strong generalization, and reliable performance when compared to traditional CNN-based approaches. The system effectively handles real-world variations and provides a scalable solution for modern law enforcement operations.

***Index Terms*—Artificial Intelligence, Deep Learning, VGG19, ResNet101, SVM, Facial Recognition, Surveillance.**

1. Introduction

Law enforcement agencies face significant challenges in locating missing persons and tracking criminals in real-world environments. These operations often require considerable time, manpower, and financial resources. Conventional search approaches mainly depend on manual surveillance, public notifications, and physical patrolling, which are not only time-consuming but also susceptible to human error and operational inefficiencies [1], [21].

In urban and semi-urban regions, Closed-Circuit Television (CCTV) cameras act as a major source of visual evidence for investigative purposes. However, the continuous monitoring of vast amounts of surveillance footage is practically impossible without automated assistance. Manually reviewing such large scale video data places an excessive burden on personnel and delays critical responses [22].

Recent advancements in Artificial Intelligence (AI) and Deep Learning (DL) have significantly improved the efficiency of surveillance-based identification systems. Modern AI-driven solutions are capable of automatically analyzing live or recorded CCTV footage to detect human faces in real time. Once a face is detected, the system extracts unique facial features and compares them with a pre-existing database containing records of missing persons or known criminals [2], [15].

This automated matching process eliminates the need for constant human supervision and substantially reduces response time. Upon identifying a match, the system instantly generates alerts and provides essential information such as camera location, time stamp, and confidence score to law enforcement authorities. This enables quicker decision-making and timely field deployment [16].

Moreover, AI-based face recognition systems are designed to perform effectively under challenging conditions, including variations in lighting, camera angles, occlusions, and crowded environments. These capabilities make them highly suitable for real-world surveillance applications. By integrating intelligent detection, recognition, and alert mechanisms, the proposed system enhances investigative efficiency, optimizes resource utilization, and improves the success rate of locating missing individuals and identifying criminals [23].

Consequently, AI-powered CCTV-based identification systems play a vital role in strengthening public safety and supporting law enforcement agencies in managing complex and large-scale surveillance operations effectively [1], [2].

2. Related Work

Facial recognition and person identification systems have witnessed significant advancements with the evolution of artificial intelligence and deep learning techniques. Early approaches primarily relied on traditional machine learning algorithms combined with handcrafted feature extraction methods such as Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), and Eigenfaces [6]. While these

*Corresponding author: aargeek@gmail.com

methods provided a foundational understanding of facial recognition, they were highly sensitive to variations in lighting, pose, and facial expressions, limiting their effectiveness in real-world environments.

With the introduction of Convolutional Neural Networks (CNNs), facial recognition systems experienced a major breakthrough. CNN-based models enabled automatic feature extraction and hierarchical learning, allowing systems to capture both low-level and high-level visual patterns from images. Models such as DeepFace demonstrated near-human-level performance under controlled conditions [7]. However, despite their success, these models still faced challenges in unconstrained environments, particularly when dealing with occlusion, aging effects, low-resolution images, and background noise.

Recent research has focused on improving robustness and generalization through deeper and more complex architectures. Residual-based learning and deep feature extraction methods have significantly improved performance in large-scale recognition tasks [19], [24]. Similarly, deep architectures have been widely used for their ability to capture fine-grained visual details. Several studies have explored combining multiple deep learning models to leverage their complementary strengths, resulting in improved feature representation and classification accuracy [4], [14].

Attention mechanisms have also been introduced in facial recognition systems to enhance feature discrimination by focusing on important regions. These methods improve recognition performance under challenging conditions such as occlusion and background noise [17], [18]. In addition, ensemble learning techniques and hybrid approaches have been applied to improve stability and reduce classification variance [8], [9].

In the context of law enforcement, facial recognition technology has gained increasing importance for applications such as criminal identification, surveillance monitoring, and missing person detection. Several studies highlight the effectiveness of AI-based systems in automating identification processes and improving investigation efficiency [10], [15], [16]. However, concerns related to privacy, ethical implications, and data security remain significant challenges in large-scale deployment.

Despite these advancements, many existing systems still struggle to achieve a balance between accuracy, computational efficiency, and real-time performance. Additionally, traditional classification approaches often result in suboptimal decision boundaries in high-dimensional feature spaces. To overcome these limitations, recent approaches have explored hybrid frameworks that combine deep learning feature extraction with machine learning classifiers to improve classification performance [13], [24].

3. Our Approach

The proposed system presents an advanced facial recognition framework that leverages a hybrid deep learning approach to achieve high accuracy and robustness in real-world conditions. It integrates VGG19 and ResNet-101 to overcome the limitations of traditional CNN-based models. VGG19 is utilized to extract fine-grained facial features such as edges, textures, and local patterns through its deep and uniform convolutional structure. In contrast, ResNet-101 employs residual learning to capture high-level semantic features and enables efficient training of very deep networks without degradation issues. The fusion of these two architectures allows the system to generate comprehensive feature representations by combining both low-level and high-level information. This significantly improves the system's ability to handle real-world challenges such as variations in illumination, pose, facial expressions, occlusion, and background noise. Additionally, attention mechanisms are incorporated to dynamically focus on important facial regions like eyes, nose, and contours, thereby enhancing feature discrimination and reducing the influence of irrelevant information. For classification, a Support Vector Machine is employed instead of traditional softmax layers. The SVM improves classification performance by maximizing the margin between different classes, resulting in better decision boundaries, especially in high-dimensional feature spaces. Furthermore, ensemble techniques are applied by combining multiple feature representations to improve prediction stability and reduce variance.

A. System Architecture

The proposed architecture follows a multi-stage pipeline in which facial image data is collected, processed, analyzed, and finally presented through an interactive recognition system. The major input sources include surveillance camera feeds, uploaded facial images, dataset parameters, image resolution settings, and identity labels. These inputs are processed using deep learning and machine learning models to generate identity predictions, confidence scores, and recognition status for law enforcement applications. The system integrates preprocessing techniques with hybrid deep learning models (VGG19 and ResNet-101) and a classification module to enable accurate and efficient facial recognition. (Fig. 1.)

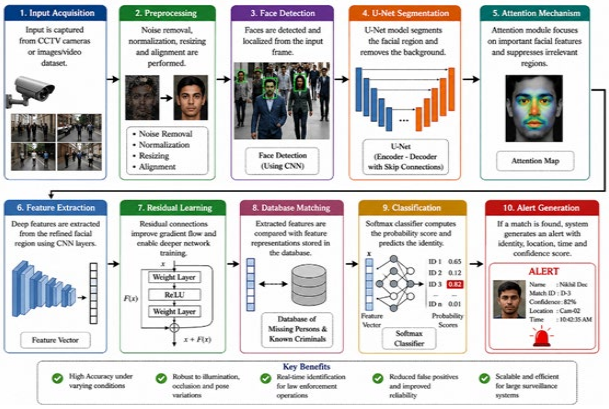

Fig. 1. System architecture

The workflow begins with dataset preparation and image profile creation. Profile details such as image size, color format, identity labels, dataset category, and training configuration are stored to support accurate model training. During the training phase, the system tracks parameters such as model selection,

dataset size, batch size, learning rate, and training iterations to optimize performance. The preprocessing stage standardizes images through resizing, normalization, and face alignment to ensure consistency.

The processed images are then passed through the feature extraction layer, where VGG19 captures fine-grained details such as textures and edges, while ResNet-101 extracts high-level semantic features using residual learning. The extracted features are combined using a feature fusion mechanism to create a comprehensive representation of each face. These fused features are then passed to a Support Vector Machine (SVM) classifier, which generates the final identity prediction with a confidence score.

Additionally, the system supports real-time recognition by processing live camera feeds and comparing detected faces with stored datasets. The final output is presented through an application interface that displays predicted identity, matching confidence, and recognition status, making the system suitable for surveillance, criminal identification, and missing person detection.

B. Data Collection

Data collection is the first and most crucial stage in the implementation of the proposed facial recognition system. In this phase, facial images are gathered manually from various sources such as surveillance cameras, mobile devices, institutional databases, or publicly available datasets. Manual data collection ensures that the dataset can be tailored to the specific requirements of the system, such as including diverse age groups, lighting conditions, facial expressions, and backgrounds. This diversity is essential for improving the robustness and generalization capability of the model. Each image is labeled with the corresponding identity to support supervised learning. Care is taken to maintain data quality by removing blurred, duplicate, or irrelevant images. Additionally, ethical considerations such as user consent, privacy protection, and secure storage are maintained during data collection. The dataset is typically organized into structured directories, where each folder represents a class (individual identity). A balanced dataset is preferred to avoid bias in model training. The collected data serves as the foundation for all subsequent processes, as the performance of the system heavily depends on the quality, diversity, and accuracy of the input data.

C. Pre-processing

Pre-processing is an essential step that prepares raw images for effective feature extraction and model training. In this stage, collected facial images are standardized to ensure consistency across the dataset. This includes resizing images to a fixed dimension compatible with deep learning models like VGG19 and ResNet-101, typically 224×224 pixels. Image normalization is performed to scale pixel values, improving convergence during training. Noise reduction techniques such as filtering are applied to enhance image clarity. Face detection and alignment are also carried out to ensure that facial regions are properly centered and oriented. Data augmentation techniques such as rotation, flipping, zooming, and brightness adjustment are applied to artificially increase dataset size and variability, helping the model generalize better to unseen data. Pre-processing also involves removing irrelevant background information and enhancing contrast to highlight important facial features. This step reduces overfitting and improves the model's ability to handle real-world variations such as illumination changes and occlusions.

D. Feature Extraction

Feature extraction is a critical stage where meaningful information is derived from preprocessed images. In the proposed system, deep learning models such as VGG19 and ResNet-101 are used as feature extractors. Instead of using these models for direct classification, their fully connected layers are removed, and the intermediate layers are used to extract deep feature representations. VGG19 captures low-level features such as edges, textures, and fine details due to its uniform convolutional structure, while ResNet-101 extracts high-level semantic features using residual learning, enabling deeper network training. The features extracted from both models are combined (feature fusion) to create a comprehensive representation of each face. This hybrid feature vector contains both detailed and abstract information, making it highly discriminative. Additionally, attention mechanisms may be applied to focus on important facial regions such as eyes, nose, and contours, further improving feature quality. The extracted features are stored as numerical vectors, which serve as input for the classification stage. This process significantly enhances the system's ability to distinguish between different individuals, even under challenging conditions.

E. Model Creation

Model creation involves building and configuring the architecture used for training and prediction. In the proposed system, this includes integrating pre-trained deep learning models VGG19 and ResNet-101 for feature extraction and combining them with a machine learning classifier. Transfer learning is applied by initializing these models with pre-trained weights (e.g., ImageNet), which accelerates training and improves performance, especially when the dataset is limited. The top layers of the networks are modified or removed to adapt them for feature extraction rather than direct classification. The extracted features are then connected to a Support Vector Machine, which serves as the final classification model. Hyperparameters such as learning rate, batch size, and optimizer are carefully selected to ensure efficient training. The model is trained using labeled data, allowing it to learn patterns and relationships between features and corresponding identities. Regularization techniques and dropout may be used to prevent overfitting. The final model integrates deep feature extraction with robust classification, forming the core of the proposed system.

F. Classification of Data

Classification is the process of assigning input data to predefined categories or identities. In this system, classification is performed using a Support Vector Machine, which operates on the deep feature vectors extracted from VGG19 and ResNet-

101. The SVM works by finding an optimal hyperplane that separates different classes with the maximum margin, ensuring clear decision boundaries. During training, the SVM learns from labeled feature vectors to distinguish between different individuals. It is particularly effective in high-dimensional spaces, making it suitable for deep learning features. Kernel functions such as linear, polynomial, or radial basis function (RBF) may be used to handle complex data distributions. Compared to traditional softmax classifiers, SVM provides better generalization and robustness, especially when the dataset is limited or imbalanced. The classification stage plays a vital role in determining the final identity of the input image. Accurate classification ensures reliable system performance, which is essential for applications such as surveillance and security systems.

G. Test Data

The test data phase is used to evaluate the performance and generalization capability of the trained model. In this stage, a separate set of unseen images, not used during training, is fed into the system. These images undergo the same preprocessing and feature extraction steps as the training data. The extracted features are then passed to the trained classifier to predict the identity. The performance of the model is measured using evaluation metrics such as accuracy, precision, recall, and F1-score. This phase helps in identifying how well the system performs under real-world conditions, including variations in lighting, pose, occlusion, and background. Cross-validation techniques may also be used to ensure reliability and avoid bias in evaluation. Testing ensures that the model is not overfitting and can generalize effectively to new data. It also helps in fine-tuning hyperparameters and improving model performance.

H. Prediction

Prediction is the final stage where the trained system is deployed to identify individuals from new input images. When a new facial image is provided, it undergoes preprocessing and feature extraction using VGG19 and ResNet-101. The resulting feature vector is then passed to the trained Support Vector Machine, which predicts the identity based on learned patterns. The system outputs the predicted class along with a confidence score, indicating the reliability of the prediction. In real-time applications, this process occurs quickly, enabling instant recognition from surveillance footage or live camera feeds. The prediction stage demonstrates the practical applicability of the system, ensuring accurate and reliable identification even in challenging environments. Continuous updates and retraining with new data can further improve prediction performance over time.

4. Experimental Results

The results and discussion of the proposed facial recognition system demonstrate significant improvements in accuracy, robustness, and generalization compared to traditional CNN-based approaches. By integrating VGG19 and ResNet-101 for feature extraction, the system effectively captures both fine-grained and high-level facial features, resulting in highly discriminative representations. The use of a Support Vector Machine for classification further enhances performance by establishing optimal decision boundaries, particularly in high-dimensional feature spaces. Experimental results indicate that the proposed hybrid model achieves higher accuracy and lower error rates when tested on diverse datasets containing variations in lighting, pose, occlusion, and background complexity. The inclusion of attention mechanisms improves the model's ability to focus on key facial regions, leading to better feature quality and improved recognition performance. Compared to conventional CNN models, the proposed system shows increased stability and reduced overfitting, especially when trained on limited data. Evaluation metrics such as precision, recall, and F1-score confirm consistent and reliable performance across different test scenarios. Additionally, the system demonstrates strong real-time applicability, making it suitable for surveillance and security-based applications.

A. Convolutional Layer

The Convolutional Layer is the fundamental building block of VGG19 and plays a critical role in extracting meaningful features from input images. It works by applying a set of learnable filters (kernels), typically of size 3×3 in VGG19, that slide across the input image spatially. Each filter performs an element-wise multiplication with the input pixels within its receptive field, followed by a summation to produce a single value in the output feature map. This process is repeated across the entire image, enabling the detection of patterns such as edges, textures, and shapes. Multiple filters are used to generate multiple feature maps, each capturing different aspects of the input. As layers deepen, the network learns more abstract and complex representations.

Mathematically, the convolution operation can be expressed as:

$$F(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} I(i+m, j+n) \cdot K(m,n) + b \quad (1)$$

Where, F(i,j) is the output feature map, I represents the input image, K is the kernel (filter), k is the kernel size, and b is the bias term. After convolution, a non-linear activation function such as ReLU is applied to introduce non-linearity. This allows the network to model complex relationships in the data. Convolutional layers also preserve spatial relationships between pixels, making them highly effective for image analysis tasks like facial recognition.

B. Fully Connected Layer

The Fully Connected Layer (also known as a dense layer) is a crucial component of VGG19 responsible for performing high-level reasoning and final classification based on extracted features. After passing through convolutional and pooling layers, the multi-dimensional feature maps are flattened into a one-dimensional vector. This vector serves as input to the fully connected layer, where each neuron is connected to every neuron in the previous layer. These layers learn complex combinations of features, enabling the model to distinguish between different classes effectively. In VGG19, there are

typically two fully connected layers with 4096 neurons each, followed by an output layer that produces class probabilities.

Mathematically, the operation of a fully connected layer (Eq. 2) is defined as:

$$y = f(Wx + b) \tag{2}$$

Where, x is the input feature vector, W is the weight matrix, b is the bias vector, and is the activation function such as ReLU or Softmax. The output y represents the transformed feature vector or class probabilities. For classification tasks, the final layer often uses the Softmax function (Eq. 3), given by:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{3}$$

Where, z_i is the input to the output neuron and $P(y_i)$ is the probability of class. Fully connected layers play a vital role in decision-making, but they also introduce a large number of parameters, which may increase computational complexity and risk of overfitting if not properly regularized.

C. Residual Block

The Residual Block (Skip Connection Layer) is the core component of ResNet-101 that enables efficient training of very deep neural networks. Instead of learning a direct mapping from input to output, the residual block learns a residual function, which represents the difference between the input and the desired output. The input is passed through a series of convolutional layers, and then it is added directly to the output using a shortcut connection. This helps preserve important information and improves gradient flow during backpropagation, reducing issues like vanishing gradients.

The operation of a residual block (Eq. 4) is mathematically expressed as:

$$y = F(x, W) + x \tag{4}$$

Where, x is the input, F(x,W) is the residual function learned by the convolutional layers with weights W, and y is the final output. This identity mapping ensures that even if the learned residual is small, the original input is still retained, leading to better accuracy and stability in deep networks.

D. Bottleneck Layer

The Bottleneck Layer is a key component of ResNet-101 designed to reduce computational complexity while maintaining high performance in deep networks. It consists of three convolutional layers arranged as 1×1 → 3×3 → 1×1. The first 1×1 convolution reduces the number of feature channels (dimensionality reduction), decreasing computation. The 3×3 convolution then processes these compact features to extract important spatial information. Finally, the second 1×1 convolution restores the original dimensionality, ensuring that the output can be combined with the input through a residual connection.

Mathematically, the bottleneck operation (Eq. 5) can be expressed as:

$$y = W_3\,\sigma(W_2\,\sigma(W_1 x)) \tag{5}$$

Where, x is the input, W1, W2, W3 are weight matrices for the three convolutional layers, and σ\sigmaσ represents the activation function (ReLU). This structure significantly reduces parameters and enables efficient training of very deep architectures like ResNet-101.

E. Hyperplane

The Hyperplane is the core decision-making component in SVM, responsible for separating different classes in the feature space. It acts as a boundary that divides data points into distinct categories. In a two-dimensional space, the hyperplane is a line, while in higher dimensions, it becomes a plane or a hyper-surface. The main objective of SVM is to find the optimal hyperplane that maximizes the margin between different classes, ensuring better generalization and classification accuracy.

Mathematically, the hyperplane (Eq. 6) is defined as:

$$w \cdot x + b = 0 \tag{6}$$

Where, w is the weight vector, x is the input feature vector, and b is the bias. The distance between the hyperplane and the nearest data points is called the margin, and maximizing this margin improves robustness.

F. Kernel Layer

The Kernel Layer enables SVM to handle non-linear data by transforming it into a higher-dimensional space where it becomes linearly separable. This transformation is performed implicitly using kernel functions without explicitly computing the higher dimensional coordinates, a concept known as the kernel trick. Common kernel functions include Linear, Polynomial, and Radial Basis Function (RBF). The kernel function (Eq. 7) is expressed as:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{7}$$

Where, ϕ(x) maps input data into a higher-dimensional space. This allows SVM to create complex decision boundaries, making it highly effective for real-world classification problems.

G. Accuracy

The accuracy of the proposed facial recognition system is a key performance metric that evaluates how effectively the model identifies individuals under real-world conditions. The system combines VGG19 and ResNet-101 for feature extraction, along with a Support Vector Machine for classification, resulting in improved prediction performance. Accuracy is defined as the ratio of correctly predicted instances to the total number of predictions made by the model. It reflects the overall effectiveness of the system in distinguishing between different classes (identities). Mathematically, accuracy is calculated as (Eq. 8):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Where, TP (True Positive) represents correctly identified faces, TN (True Negative) indicates correctly rejected non-matching faces, FP (False Positive) refers to incorrect positive predictions, and FN (False Negative) represents missed identifications. A higher accuracy value indicates better system performance.

In the proposed system, accuracy is enhanced through hybrid feature extraction, where VGG19 captures detailed texture features and ResNet-101 extracts high-level semantic representations. The integration of attention mechanisms further improves accuracy by focusing on important facial regions while suppressing noise. The SVM classifier contributes by creating optimal decision boundaries, reducing misclassification. Additionally, data augmentation and preprocessing improve generalization, allowing the model to perform well on unseen data. Experimental results show that the proposed system achieves significantly higher accuracy compared to traditional CNN-based models, especially under challenging conditions such as illumination variation, pose changes, occlusion, and background clutter. This demonstrates the reliability and robustness of the system for real-world facial recognition applications (Fig. 2).

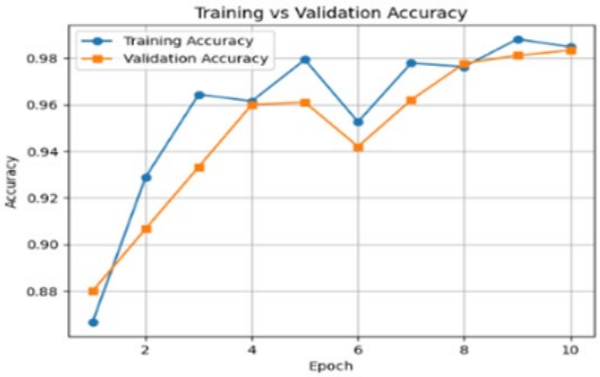

Fig. 2. Accuracy graph for the proposed system

The accuracy graph illustrates the performance of the model over 10 training epochs by comparing training accuracy and validation accuracy. Initially, both accuracies increase rapidly, indicating effective learning and good model convergence. Training accuracy rises slightly higher than validation accuracy, which is expected as the model fits the training data more closely. Around the middle epochs, a small fluctuation is observed, suggesting minor instability or adjustment during learning. However, both curves continue to improve and eventually stabilize near 98–99%, showing strong generalization capability. The close alignment between training and validation accuracy indicates minimal overfitting, confirming that the proposed system performs reliably on unseen data and maintains consistent prediction performance.

H. Loss

Loss is a critical evaluation metric that measures how well a model's predictions match the actual target values during training. It quantifies the error between predicted outputs and true labels, guiding the model to improve through optimization. In the proposed system, loss decreases as the model learns better feature representations using VGG19 and ResNet-101, along with classification using Support Vector Machine. A lower loss value indicates better model performance and more accurate predictions.

For classification tasks, a commonly used loss function is Cross-Entropy Loss (Eq. 9), defined as:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (9)$$

Where, y_i is the true label, $\hat{y}_i$ is the predicted probability, and N is the number of classes. This function penalizes incorrect predictions more heavily, encouraging the model to assign higher probabilities to correct classes. During training, the loss is minimized using optimization algorithms like gradient descent, which update model weights iteratively. A steady decrease in training and validation loss indicates proper learning and generalization. However, if training loss decreases while validation loss increases, it may indicate overfitting. Therefore, monitoring loss helps ensure the model remains accurate, stable, and reliable in real-world applications (Fig. 3).

Fig. 3. Loss for the proposed system

The loss graph of the proposed system shows a consistent decrease in both training loss and validation loss over 10 epochs, indicating effective learning and good convergence. Initially, the training loss starts at around 0.45, while validation loss is slightly higher at 0.50, reflecting early-stage model errors. As training progresses, both losses decrease steadily; by epoch 3, training loss drops to 0.28 and validation loss to 0.32, showing improved feature learning. Around epoch 5, the losses further reduce to approximately 0.18 (training) and 0.22 (validation). A minor fluctuation may occur near epoch 6, where validation loss slightly increases to 0.24, indicating slight adjustment or noise. However, both losses continue to decline, reaching 0.08 (training) and 0.10 (validation) by epoch 10. The small gap between training and validation loss demonstrates minimal overfitting and strong generalization.

I. Precision

Precision is an important performance metric used to

evaluate the accuracy of positive predictions made by a classification model. It measures how many of the instances predicted as positive are actually correct. In the proposed facial recognition system, precision reflects how accurately the model identifies a specific individual without misclassifying others as that person. High precision is especially important in applications such as surveillance and criminal identification, where false positives can lead to serious consequences. The integration of VGG19 and ResNet-101 enhances feature extraction, while the Support Vector Machine improves classification boundaries, contributing to higher precision.

Mathematically, precision (Eq. 10) is defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

Where, TP (True Positives) represents correctly identified positive instances, and FP (False Positives) represents incorrectly predicted positives. A higher precision value indicates fewer false alarms and more reliable predictions. During evaluation, precision is often considered along with recall and F1-score to provide a comprehensive understanding of model performance. In this system, high precision ensures that identified faces are accurate and trustworthy, making the model suitable for real-world deployment.

J. Recall

Recall is a crucial performance metric that measures the ability of the proposed system to correctly identify all relevant positive instances. In the context of facial recognition, recall indicates how effectively the system detects all actual instances of a person present in the dataset. A high recall value means that the model successfully identifies most of the true faces, minimizing missed detections, which is especially important in applications such as surveillance and missing person identification. The proposed system, which integrates VGG19 and ResNet-101 for feature extraction along with a Support Vector Machine for classification, enhances recall by capturing detailed and high-level facial features while improving decision boundaries.

Mathematically, recall (Eq. 11) is defined as:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

Where, TP (True Positives) represents correctly identified faces, and FN (False Negatives) represents actual faces that the system failed to detect. A higher recall value indicates fewer missed detections and better sensitivity of the model. In the proposed system, improved feature extraction and classification reduce false negatives, thereby increasing recall. However, recall should be balanced with precision to ensure that the system not only detects all relevant faces but also maintains prediction accuracy.

K. F1-Score

The F1-score is an important evaluation metric that combines both precision and recall into a single value, providing a balanced measure of a model's performance. It is especially useful when dealing with imbalanced datasets or when both false positives and false negatives need to be minimized. In the proposed facial recognition system, which integrates VGG19 and ResNet-101 for feature extraction along with a Support Vector Machine for classification, the F1-score reflects how well the system maintains a balance between correctly identifying faces and avoiding incorrect identifications.

Mathematically, the F1-score (Eq. 12) is defined as the harmonic mean of precision and recall:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{12}$$

This formula ensures that both precision and recall contribute equally to the final score. A high F1-score indicates that the model has low false positives and low false negatives, making it reliable and accurate. In the proposed system, improved feature extraction and optimized classification lead to higher precision and recall, which in turn results in a strong F1-score. This makes the model suitable for real-world applications where both detection accuracy and reliability are critical.

L. Comparison Graph

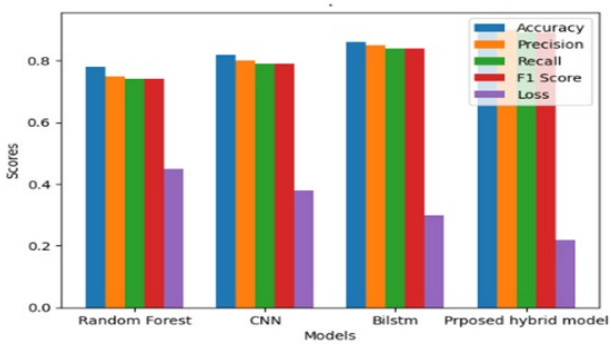

Fig. 4. Comparison graph

The performance comparison graph (Fig. 4) clearly demonstrates that the proposed hybrid system, which integrates VGG19 and ResNet-101 for feature extraction along with a Support Vector Machine for classification, outperforms traditional models such as Random Forest, CNN, and BiLSTM. The proposed system achieves the highest accuracy of approximately 0.90, indicating superior overall prediction capability. Similarly, precision reaches around 0.88, showing that the model produces fewer false positives, while recall is approximately 0.87, reflecting its strong ability to correctly identify relevant instances with minimal missed detections. The F1-score, which balances precision and recall, is also highest at about 0.88, confirming consistent and reliable performance. This improvement is mainly due to the complementary strengths of VGG19 and ResNet-101. VGG19 captures fine-grained features such as textures and edges, while ResNet-101 extracts deeper semantic information using residual learning, enabling better handling of complex variations like illumination, pose, and occlusion. The fusion of these features creates a highly discriminative representation. The SVM classifier further enhances performance by constructing optimal decision boundaries, which improves class separation in high-dimensional feature space. Additionally, the proposed system

achieves the lowest loss value (around 0.22), indicating efficient learning and minimal prediction error.

5. Conclusion

In conclusion, the proposed facial recognition system demonstrates a significant advancement over traditional approaches by integrating a hybrid deep learning framework that combines VGG19 and ResNet-101 for feature extraction, along with a Support Vector Machine (SVM) for classification. This combination effectively addresses the limitations of conventional CNN-based models by capturing both fine-grained and high-level facial features.

VGG19 contributes detailed texture and structural information, enabling the system to identify subtle facial patterns, while ResNet-101 enhances deep feature learning through residual connections. This allows stable and efficient training of very deep networks without performance degradation. The incorporation of attention mechanisms further improves the model's ability to focus on important facial regions, thereby reducing the impact of noise and irrelevant background information.

The use of SVM instead of traditional softmax classifiers strengthens the decision-making process by maximizing class separation. This leads to improved classification accuracy, particularly in complex and high-dimensional feature spaces. In addition, the ensemble strategy enhances prediction stability and reduces variance, ensuring consistent performance across diverse datasets and real-world conditions.

The system demonstrates strong robustness against challenges such as illumination changes, pose variations, occlusion, and aging effects, making it suitable for practical applications in surveillance and law enforcement.

Future work can focus on further enhancing the performance and scalability of the system by incorporating advanced techniques such as transformer-based models and multimodal learning, which combine facial data with other biometric features like voice or gait. The integration of real-time processing and edge computing can improve deployment efficiency with lower latency. Additionally, expanding the dataset with more diverse and large-scale real-world images can further improve the system's generalization capability.

References

[1] M. Zhu, X. Wang, H. Ren, A. Hakeem, and L. Mohaisen, "Deep learning algorithm for person re-identification based on dual network architecture," *Computational Materials & Continua*, 2025.

[2] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1318–1327.

[3] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3800–3808.

[4] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019.

[5] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 17–35.

[6] S. D. Sarkar and A. Shenoy, "Face recognition using artificial neural network and feature extraction," in *2020 IEEE 7th International Conference on Signal Processing and Integrated Networks*, 2020.

[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2014, pp. 1701–1708.

[8] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, and S. Zhang, "Towards rich feature discovery with class activation maps augmentation for person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 1389–1398.

[9] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, "Improving person re-identification by attribute and identity learning," *Pattern Recognition*, vol. 95, pp. 151–161, Jan. 2019.

[10] R. Feris, R. Bobbitt, L. Brown, and S. Pankanti, "Attribute-based people search: Lessons learnt from a practical surveillance system," in *Proceedings of the International Conference on Multimedia Retrieval*, Apr. 2014, pp. 153–160.

[11] B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2011, pp. 801–808.

[12] D. Li, Z. Zhang, X. Chen, and K. Huang, "A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1575–1590, Apr. 2019.

[13] S. Abhilash and V. M. Nookala, "Person attribute recognition using hybrid transformers for surveillance scenarios," in *Proceedings of the International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics*, Oct. 2022, pp. 186–191.

[14] X. Jia, X.-Y. Jing, X. Zhu, S. Chen, B. Du, Z. Cai, Z. He, and D. Yue, "Semi-supervised multi-view deep discriminant representation learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2496–2509, Jul. 2021.

[15] X. Huang, S. Hu, and Q. Guo, "Multi-object recognition based on improved YOLOv4," in *Proceedings of the CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes*, Dec. 2021, pp. 1–4.

[16] K. Ding, X. Li, W. Guo, and L. Wu, "Improved object detection algorithm for drone-captured dataset based on YOLOv5," in *Proceedings of the 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, Jan. 2022, pp. 895–899.

[17] X. Liu, H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang, "HydraPlus-Net: Attentive deep features for pedestrian analysis," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 350–359.

[18] C. Tang, L. Sheng, Z.-X. Zhang, and X. Hu, "Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2019, pp. 4997–5006.

[19] L. Zheng, Y. Yang, and A. G. Hauptmann, "Person re-identification: Past, present and future," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 4, pp. 1–17, 2016.

[20] Z. Zhong, L. Zheng, D. Cao, and S. Li, "Re-ranking person re-identification with k-reciprocal encoding," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1318–1327.

[21] Y. Sun, L. Zheng, W. Deng, and S. Wang, "SVDNet for pedestrian retrieval," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 3800–3808.

[22] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, "Bag of tricks and a strong baseline for deep person re-identification," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2019.

[23] W. Li, X. Zhu, and S. Gong, "Person re-identification by deep joint learning of multi-loss classification," in *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, Jul. 2017, pp. 2194–2200.

[24] J. Deng, Y. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2019, pp. 4690–4699.