Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

Robust Face Recognition System for Extensive Distances in Uncontrolled Environments on Edge Device

S.Vasavi[1*], Himavarshini Bora[2], N.Dinesh Murali[3] , Novaline Jacob[4],Emmaneual Sanjay Raj[5]

[1,2,3]Department of Artificial Intelligence and Data Science Velagapudi Ramakrishna Siddhartha Engineering College, Vijayawada,520007, Andhra Pradesh, India.

[4,5] Advanced Data Research Institute (ADRIN), Department of Space, Hyderabad

[1]vasavi.movva@gmail.com, [2]himani.varshini@gmail.com,

[3]nandyaladineshmurali2020@gmail.com, novalinejacob@gmail.com[4], esanjayraj@gmail.com[5]

Abstract: Traditional CCTV monitoring, reliant on human observers, faces scalability and accuracy challenges in identifying key individuals (dignitaries, faculty) to enable secure access. The rapid advancements in deep learning models have revolutionized the field of face recognition, offering a diverse range of highly effective solutions such as the DeepFace, FaceNet, VGG-Face, and ArcFace. Current face recognition technologies for extensive distances may suffer from accuracy limitations in crowded or uncontrolled environments and may require ongoing manual intervention for optimization and maintenance. An enhanced Multi-Task Cascaded Convolutional Neural Network (MTCNN) algorithm and ArcFace implementation using Resnet-50 is proposed to recognize individuals over medium to long ranges on edge devices such as Jetson Nano. Enhanced MTCNN excels in detecting faces with varying scales and orientations, crucial for recognizing distant faces. ArcFace implementation using ResNet-50, with its deep architecture, learns discriminative features even from low-resolution images, enhancing recognition accuracy over traditional models. For training and testing of the proposed model, manually collected images of individuals under varying illumination and environments and the CelebA dataset consisting of over 200,000 celebrity images, each annotated with various attributes such as hair color, presence of eyeglasses, facial expression, and gender were used. Compared to the existing works, the proposed MTCNN and ArcFace implementation using the Resnet-50 system demonstrates significantly higher accuracy of 97% with minimal acquisition of dataset to train and scalability, offering a robust and efficient solution for dignitary tracking.

***Keywords**:* MTCNN, Resnet-50, Face Recognition, Uncontrolled Environment, Extensive Distances, Edge device

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

1. INTRODUCTION

In the modern era of digital transformation, the proliferation of surveillance systems and the demand for enhanced security measures have propelled the exploration of face recognition technologies to unprecedented levels. Although facial recognition systems have historically performed well in confined spaces and over short distances, a significant problem now facing these systems is ensuring consistent identification over long distances. Over the past ten years, substantial progress has been made in face recognition and biometric recognition in general. Nonetheless, agreeable participants at close range are required in the great majority of real-world biometric identification applications. The idea of automatically identifying persons in public and without their explicit consent gave rise to Face Recognition at a Distance. The best biometric for recognition at a distance is the face. It may be easily pictured from a distance and is widely seen. Without the subject's awareness, face image can be used for security or clandestine purposes. The science of computer vision has undergone a revolution with the introduction of deep learning (DL) algorithms, which present potential paths to overcome this obstacle and open up new possibilities for face identification at long distances. The human face serves as a unique identifier, bearing intricate features and characteristics that distinguish individuals from one another. Leveraging this distinctiveness, face recognition systems have become indispensable tools in various domains, including law enforcement, border control, access control, and surveillance. However, the efficacy of conventional face recognition methods diminishes as the distance between the subject and the camera increases, primarily due to factors such as resolution degradation, occlusion, and variations in illumination. In the field of deep learning, convolutional neural networks (CNNs) have become a potent paradigm thanks to their exceptional capacity for feature extraction, representation learning, and pattern identification.

CNNs have shown previously unheard-of performance in a variety of visual identification tasks, such as object detection and semantic segmentation, by learning hierarchical representations from raw pixel data. Deep learning's potential to be used for face recognition at long distances offers a strong chance to get beyond some of the drawbacks of conventional methods. The deterioration of facial characteristics due to low picture quality is one of the primary problems in face identification at long distances. Reduced recognition accuracy results from traditional algorithms' inability to extract discriminative characteristics from low-resolution photos. On the other hand, deep learning models are particularly good at learning hierarchical feature representations, which helps them deal with resolution differences and retrieve pertinent data for precise identification. Furthermore, deep learning frameworks are versatile enough to adjust and generalize to a wide range of environmental factors, such as changing lighting, pose changes, and occlusions. Through extensive training

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

on a variety of large-scale datasets, deep learning models can acquire robust representations that are not affected by these changes, which improves the generalization and resilience of face recognition systems.

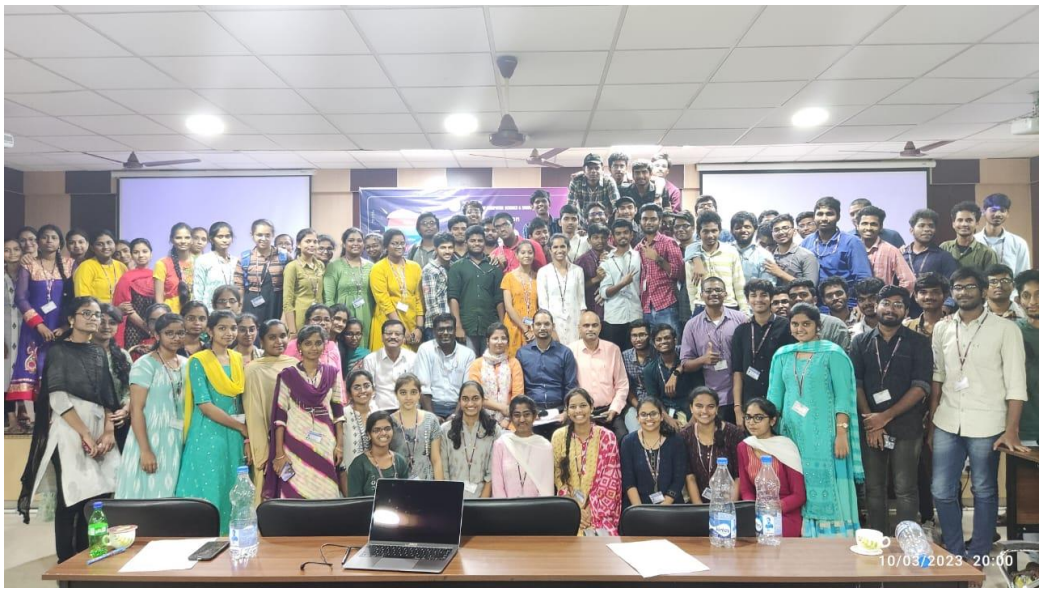

Fig.1 Uncontrolled Environment

In this study, the research provides an exhaustive overview of recent developments in DL algorithms for face recognition at long distances. The research investigates the foundational ideas of deep learning architectures and examines how well they work to overcome the difficulties presented by long distances, low-resolution imaging, and environmental fluctuations. The research hopes that the analysis carried out will shed light on the most recent approaches now in use, point out any gaps in the field, and suggest an enhanced system for Face Recognition for extensive distances under uncontrolled environments as shown in Figure 1.

1.1 Objectives:

1. To generate a dataset of individuals by manually collecting images under varying illuminations and environments

2. To propose a deep learning model that recognizes faces at extensive distances (over 10 – 15 meters).

3. To deploy the deep learning model onto an edge computing device such as Jetson Nano for lower latency and faster response times.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

1.2 Contributions:

This study contributes to the field of face recognition in uncontrolled environments by displaying the results on the edge devices.

1. Created a comprehensive dataset comprising face images of celebrities and students of our college captured under diverse lighting conditions and environmental settings

2. Proposed an enhanced version of MTCNN and ArcFace implementation using Resnet-50 to detect and recognize the faces of individuals in uncontrolled environments.

3. Validated the model and deployed the model in the edge device for better surveillance systems at sensitive borders and crowd monitoring.

1.3 Organization:

The paper is structured in the following manner: Section 2 delves into a literature review, analyzing current technologies, and highlighting their strengths and weaknesses. Section 3 presents the methodology, outlining the architecture of the proposed models and elucidating their functionality, encompassing data collection, model development, and evaluation. Section 4 elucidates the results that were obtained, followed by conclusions and avenues for future research.

2. LITERATURE REVIEW

The emergence of surveillance technologies has brought about a significant transformation in the security and access control domain, providing unparalleled capacities for observation and recognition. Scalability and accuracy are major problems for typical closed-circuit television (CCTV) systems that rely on human observers, especially when it comes to identifying important people like faculty members and dignitaries. Automated solutions that can overcome these constraints and provide continuous, real-time detection and monitoring across medium and long-range distances are desperately needed as the need for increased security measures rises. By integrating the MTCNN and ArcFace implementation using Resnet-50 for Face Recognition this study offers a novel way to overcome the drawbacks of conventional CCTV surveillance. The proposed automated system can analyze video from cameras and successfully identify targeted persons in complex and dynamic surroundings by utilizing the better face detection skills of MTCNN and the discriminative capability of ArcFace implementation using Resnet-50. This allows the system to function beyond human limits.

The research presented by Lixiang Li [1], delineates the developmental stages and associated technologies pertinent to face recognition. The study emphasized on comparative analysis of

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

existing Face Recognition technologies. It is noted that most of the technologies presented in their work did not exhibit high accuracy in terms of model performance.

In the research outlined by Della Gressinda Wahana et al. [2], a face recognition system within a confined space, utilizing video recordings and implementing techniques such as Non-negative Matrix Factorization suppressed carrier (NMFsc) and Local Non-negative Matrix Factorization (LMNF) was described. Their system demonstrates proficiency in facial recognition, with the LMNF. Conversely, the NMF suppressed carrier method takes a longer computational time. NMF's algorithms are sensitive to variations in facial appearance caused by changes in illumination, pose, facial expression, or occlusions and hence provide reduced reliability.

In the research outlined by Reecha Sharma et al. [3], they established an efficient pose-enduring face recognition system using PCA and ANFIS (PCA–ANFIS). The image features are first extracted using Principal Component Analysis (PCA), followed by recognition using a neuro-fuzzy system known as ANFIS. Developing ANFIS models for real-time applications is challenging due to their complexity and resource-intensive training process, particularly with large input feature sets.

In the work reported by Hansung Lee et al. [4], discussed a user-friendly access control system that integrates human face recognition with RFID technology to bolster security measures. Their methodology encompasses two primary stages: facial and ocular detection, followed by facial recognition. For facial and ocular detection, they employed the LBP-AdaBoost technique. For facial representation. Their system delivered noteworthy outcomes, achieving a genuine acceptance rate of 95.0% and a false acceptance rate of 0.0. Their model fails to generalize well to faces that differ significantly from those in the training dataset.

The study put forward by Diaa Salama AbdELminaam [5] focuses on leveraging an adaptive version of the recent DCNN (Deep Convolutional Neural Network) algorithm known as AlexNet , which is a deep learning method for FR using Transfer Learning (TL) in fog computing. Results indicate that the DCNN algorithm outperforms other algorithms in terms of accuracy, recall, precision, and specificity. With a large number of parameters, AlexNet overfits non-facial patterns present in the training data, which could degrade its ability to generalize to unseen faces or variations in facial expressions, poses, and illuminations.

H.A. Rowley [6] described a CNN-based face detection system that utilizes a retinal-connected neural network to examine small image segments for facial presence. Additionally, the researchers incorporate simple heuristics, leveraging insights such as the infrequency of face overlap in images to further enhance accuracy. The study faces challenges in deploying

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

its system on resource-constrained devices, such as mobile phones or embedded systems, due to the network's computational and memory requirements.

B Kanaka Durga et al. [7] described a 2D-ResNet CNN multi-class classifier for Face Recognition. The 2D-ResNet DL model has been trained using the open-source JAFFE public dataset. Face pictures may have less spatial information at long distances because of lower resolution, and occlusions might undermine the performance of 2D-ResNet CNNs, which mostly rely on fine-grained local characteristics for identification.

Chunrui Han et al [8] used a personalized CNN method to extract different characteristics for more precise face recognition. Instead of using fixed kernels, the method dynamically generates a set of kernels tailored to each individual's unique features. These kernels are divided into two components: the commonality component, capturing shared features among subjects optimized on a reference set, and the specialty component, isolating individual characteristics. Experiments on datasets like LFW, IJB-A, and IJB-C have been carried out. However, personalized convolutional networks raise privacy concerns and confidentiality of sensitive biometric data.

Peng Peng et al. [9] reported a comprehensive framework using PCA for Face recognition. PCA conventionally represents data by linear combinations of the original features and thus their system faces challenges with nonlinear data which include changes in poses, expressions, and lighting conditions.

2.1 Research Gaps:

- Existing research has primarily focused on face recognition at close to moderate distances, such as in controlled indoor environments or surveillance settings. There is a gap in understanding how face recognition algorithms perform at extensive distances (e.g., beyond 10 meters) in uncontrolled outdoor environments. [11]
- While there is considerable research on face recognition in controlled environments, such as laboratories or monitored indoor spaces, there is a lack of comprehensive studies in uncontrolled outdoor environments. [12]
- There is a gap in research that systematically investigates the combined effects of multiple environmental variables such as lighting, weather, and occlusions (e.g. Sunglasses) on face recognition performance at extensive distances. [13]

3. METHODOLOGY

The proposed methodology for developing an automated surveillance system driven by the MTCNN algorithm and ArcFace implementation using Resnet-50 architecture comprises

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

several key components. These components include architecture design, model descriptions, and the data preparation process.

3.1. Architecture Design:

3.1.1 Modified Multi-Task Cascaded Convolutional Neural Network:

Figure 2 of MTCNN architecture represents a sophisticated ensemble of neural networks meticulously crafted to detect faces within images with remarkable precision and efficiency.

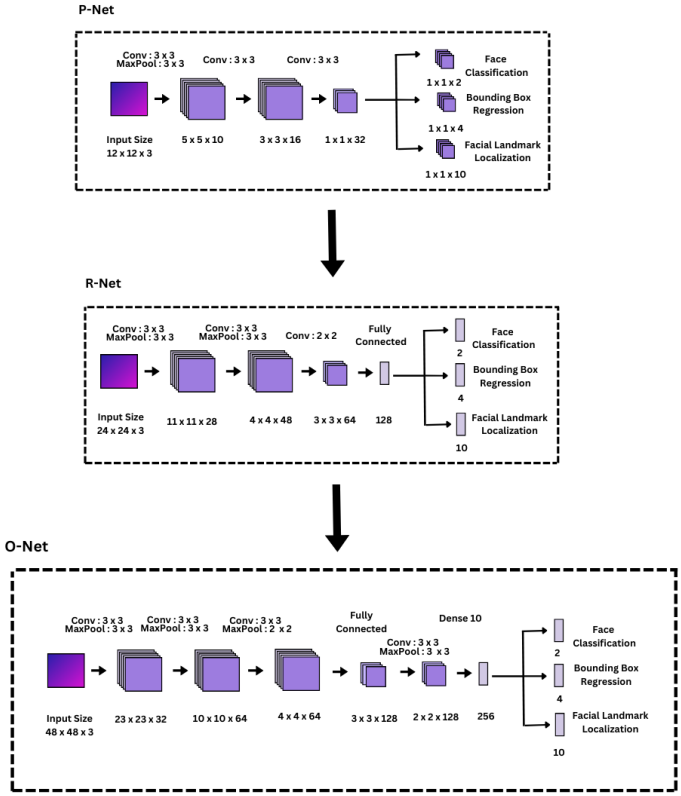

Fig.2 Proposed MTCNN Architecture

At its core are three integral components:

a. The Proposal Network (P-Net) serves as the initial investigator, tasked with proposing potential locations where faces might reside within the image. It calculates the likelihood of a face being present within each proposed area.

b. Following the P-Net's lead, the Refine Network (R-Net) takes charge of scrutinizing and refining the proposed regions. Its responsibility lies in the elimination of false positives

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

and the enhancement of localization accuracy, ensuring that only the most promising areas are retained for further examination.

c. Acting as the final arbiter, the Output Network (O-Net) undertakes comprehensive analysis and refinement of the candidate regions. In addition to finalizing the bounding boxes, it undertakes the critical task of localizing facial landmarks and estimating the probabilities associated with various facial attributes.

Together, these components form an integrated system that operates synergistically, leveraging the strengths of each network to achieve exceptional levels of facial detection accuracy across diverse image contexts. Table 1 presents the modifications done to the existing layers of O-Net in MTCNN [16] to achieve efficient results through the proposed model.

The enhanced MTCNN architecture by adding a Conv-MaxPool layer in the O-Net of MTCNN bolsters face detection by enhancing feature extraction with hierarchical patterns. It reduces dimensionality, curbing computational complexity and overfitting while fostering better generalization. The convolutional nature ensures translation invariance, crucial for detecting faces across varied positions. These advancements collectively enable the network to adeptly recognize faces amidst diverse conditions, from lighting variations to occlusions, elevating overall performance.

TABLE I. Comparison Between the Existing and Proposed Model

PARAMETERS	O-Net in MTCNN	Modified O-Net in MTCNN
Channels	1	1
Image shape	48 x 48 x 3	48 x 48 x 3
Strides	2	4
Input Kernel size	3 x 3	3 x 3
Initial Number of Filters	32	64
Pooling type	MaxPooling2D	MaxPooling2D
Max Pooling Size	2 x 2	3 x 3
No. of Layers	52	54
The resulting image's channels	1	1

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

The parameters of the model have been modified for the following reasons:

- Increasing the size of the max pooling layer helps to down-sample feature maps more effectively, leading to a more compact representation of the input.
- Increasing strides reduced the spatial dimensions of feature maps quickly, leading to faster computation and potentially better generalization.
- A larger model can capture more variations in faces, such as different orientations, expressions, and occlusions, which is beneficial for extensive range detection. This led to a more precise bounding box and facial landmark predictions.

3.1.2 ArcFace implementation using ResNet-50 Architecture:

ResNet-50, a notable iteration of the Residual Network (ResNet) architecture [17], is highly esteemed for its depth and efficacy in tasks related to image recognition. Figure 3 presents the enhanced Resnet-50 Architecture.

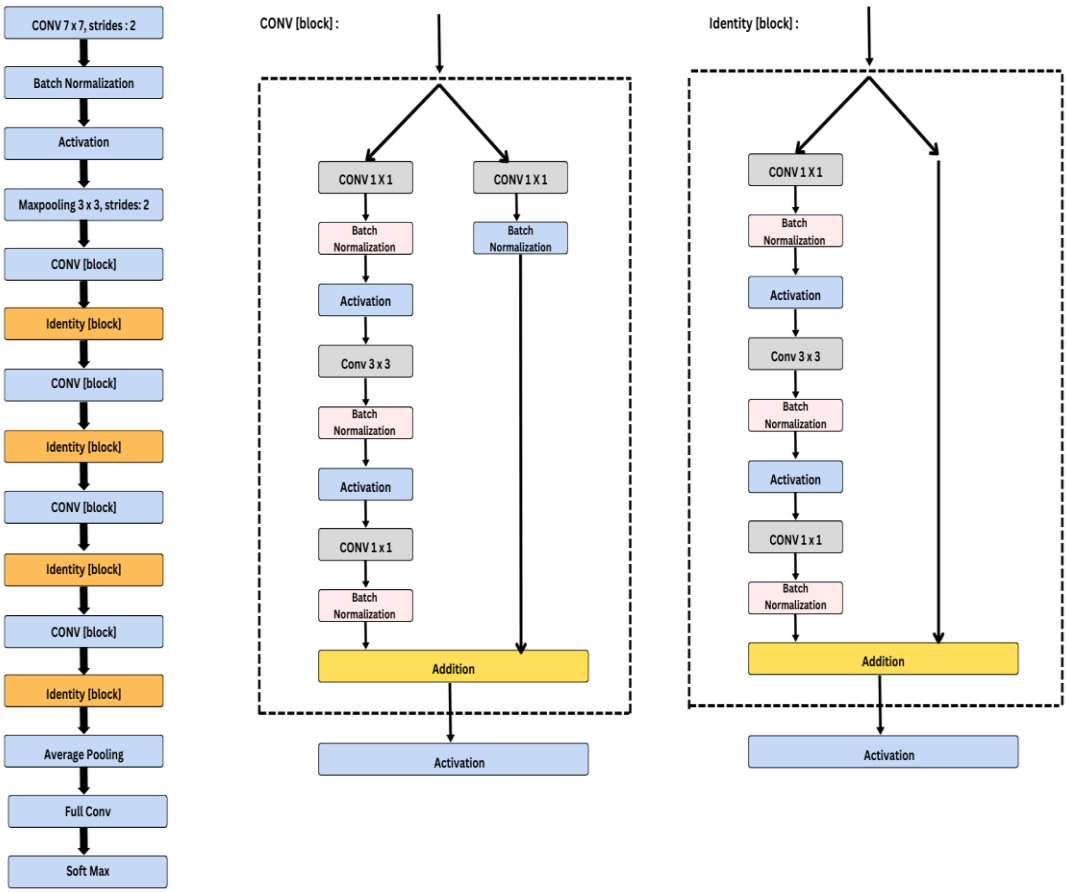

Fig.3 Enhanced Resnet-50 Architecture

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

The enhanced ResNet-50 has additional layers of PReLu, Conv2D, MaxPooling2D, and Dense included in the architecture which has resulted in increased validation accuracy of the model obtained after experimental analysis.

The ResNet with 50 layers. This involves the utilization of 1×1 convolutions, referred to as "bottlenecks," which effectively reduces the number of parameters and matrix multiplications. Consequently, this design facilitates accelerated training of each layer by a significant margin. Unlike traditional structures consisting of two layers, the bottleneck residual block in ResNet-50 incorporates a stack of three layers. The architectural configuration of ResNet-50 encompasses the following components:

1. Input Layer:

At the onset, the input layer serves as the gateway for raw pixel values extracted from an image. In the case of ResNet-50, it's customary to standardize images to a predefined size (e.g., 224x224 pixels) before they are introduced into the network.

2. Convolutional Layers:

The primary layer of ResNet-50 comprises convolutional operations, where the input image undergoes convolution with a set of adaptable filters. This process is instrumental in discerning low-level features like edges and textures from the input image.

Successive convolutional layers within each residual block progressively capture more abstract and high-level features, empowering the network to discern intricate patterns and structures within images.

3. Batch Normalization:

Following each convolutional operation, batch normalization is deployed. This normalization technique ensures the standardization of activations from the preceding layer, mitigating internal covariate shift, and expediting the training process. It fosters stability and accelerates learning by maintaining uniform input ranges across layers.

4. Rectified Linear Unit (ReLU) Activation:

Following batch normalization, the ReLU activation function is applied element-by-element to the output of each convolutional layer. It boosts the network's non-linearity, making it possible to identify intricate connections between input and output.

5. Residual Blocks: Residual blocks, intrinsic to ResNet architectures, serve as the fundamental units of ResNet-50. Each residual block comprises multiple convolutional layers, typically featuring smaller filter sizes, coupled with batch normalization and ReLU activation.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

Notably, each residual block incorporates a "shortcut connection" or "identity shortcut," facilitating direct communication between layers. This mechanism aids in gradient flow during training, enabling the network to effectively learn residual functions.

The integration of residual connections mitigates issues associated with vanishing gradients and facilitates the successful training of deep neural networks.

6. Pooling Layers:

Strategically positioned max-pooling layers traverse the network to downsample the spatial dimensions of feature maps. This downsampling process reduces computational complexity and spatial resolution while preserving critical features.

7. Global Average Pooling (GAP):

Towards the network's conclusion, a global average pooling layer is implemented. This layer computes the mean value of each feature map across its entirety, generating fixed-length feature vectors for each channel.

Global average pooling streamlines spatial dimensions, yielding a singular vector conducive to subsequent fully connected layers.

8. Fully Connected Layer (FC):

The ultimate layer of ResNet-50, the fully connected layer, orchestrates the mapping of high-level features extracted by antecedent layers to output classes. Conventionally, a softmax activation function is employed on the FC output to generate probability scores for every class, depicting the likelihood of an input image belonging to respective classes.

3.1.3 Proposed Model Architecture :

a. The system architecture as depicted in Figure 3 provides a structured framework delineating the sequential flow of operations. Initially, the system leverages two distinct models: the MTCNN model for face detection and the ResNet-50 model for face recognition. Each model undergoes individual training processes to optimize their performance for their respective tasks.

b. To enrich the dataset and enhance the robustness of the system, data collection involves a combination of manual gathering and augmentation using the CelebA dataset [10]. This augmented dataset acts as a useful resource. for comparative analysis, enabling the evaluation of the system's performance against established models and benchmarks.

c. Following data collection, preprocessing methods are applied to ensure the data is appropriately formatted and prepared for effective utilization by the models. These

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

preprocessing techniques may include image resizing, normalization, histogram equalization, and data augmentation. By applying these methods, the dataset is refined to mitigate noise, improve feature representation, and enhance model training efficacy.

d. Once preprocessing is completed, the dataset is partitioned into distinct subsets for testing and validation purposes. This partitioning allows for the evaluation of model performance on unseen data and facilitates the tuning of model hyperparameters to optimize performance.

e. In the operational phase, the system facilitates live face recognition through a terminal interface. Utilizing a live stream video feed as input, the system processes incoming frames in real time, applying the trained models to detect and recognize faces. Upon recognition, authorized faces are identified, and their information is stored in a designated folder for subsequent reference and authentication purposes.

f. This comprehensive system architecture encapsulates the entire workflow, from data acquisition and preprocessing to model training and live recognition, providing a robust framework for face recognition applications. Figure 4 presents the proposed model flow diagram.

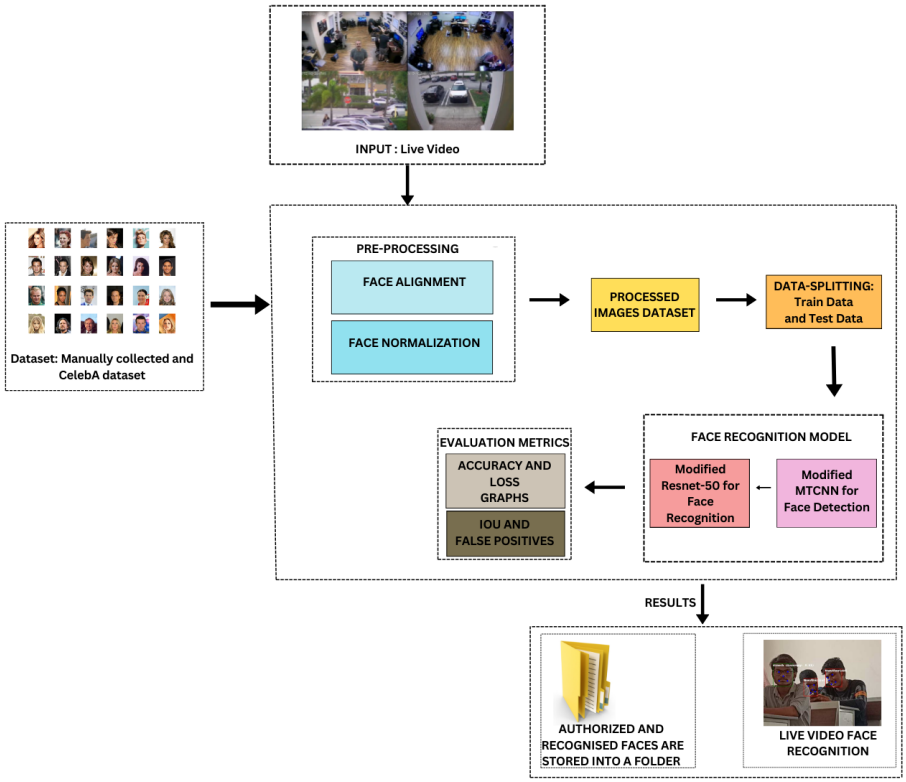

Fig.4 Proposed Model Flow Diagram

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

Table 2 presents the description of the dimensions of each layer of existing standard models and the proposed ensemble model.

TABLE II. PROPOSED ENSEMBLE MODEL

PARAMETERS	MTCNN	ResNet-50	Proposed Ensemble Model
Channels	3	3	3
Image size	128 x 128 x 3	256 x 256 x 3	256 x 256 x 3
Strides	2	2	2
Size of input kernel	3 x 3	7 x 7	3 x 3
Number of Initial Filters	10	64	-
Additional Layers	Conv2D (3 x 3, 64)	Conv2D (3 x 3, 64)	Conv2D (3 x 3, 64)
	PReLU Activation	PReLU Activation	PReLU Activation
	MaxPooling2D (3 x 3, 2 x2)	MaxPooling2D (3 x 3, 2 x2)	MaxPooling2D (2 x2)
	Conv2D (3 x 3, 128)	Conv2D (3 x 3, 128)	Conv2D (3 x 3, 128)
	PReLU Activation	PReLU Activation	PReLU Activation
	Dense (10 Units)	Dense (1024 Units)	Dense (1024 Units)
		Dense (512 Units)	Dense (512 Units)
Pooling Type	Max Pooling	-	Max Pooling
Dimensions of pooling	2 x 2	-	2 x 2
Number of layers	5	51	8
Number of parameters	-	23,571,551	23,571,551 + Additional layers

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

The choice of parameter values for the proposed ensemble model in Table II is guided by several considerations to optimize performance and efficiency. Larger input dimensions (256 x 256) for ResNet-50 and the ensemble model enhance feature extraction and performance by providing more detailed information. MTCNN uses smaller dimensions (128 x 128) to reduce computational load during the initial face detection stage. Each model includes additional convolutional layers with PReLU activation and pooling layers to further extract and down-sample features.

3.2 Methodology:

3.2.1 Data Collection:

We have collected a diverse dataset of footage captured in various real-world environments, encompassing different lighting conditions, camera angles, crowd densities, and environmental complexities. The dataset includes footage from indoor and outdoor settings, such as academic institutions, public spaces, and high-security facilities. Apart from the manual dataset collected, we worked on the celebrity dataset [10] for comparative analysis with other models which consists of over 200,000 celebrity images, each annotated with various attributes such as hair color, presence of eyeglasses, facial expression, and gender. Along with the images, the dataset includes attribute labels for each image, providing information about the presence or absence of specific attributes. The images are collected from celebrity images available on the internet and cover a diverse range of facial appearances, poses, and lighting conditions. The dataset is typically provided in the form of a collection of image files (commonly in JPEG format), along with a metadata file containing attribute labels.

Figure 5 illustrates a sample collection of images featuring the same individual captured under varied lighting conditions and environments.

Fig.5 Sample Of Dataset collected manually

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

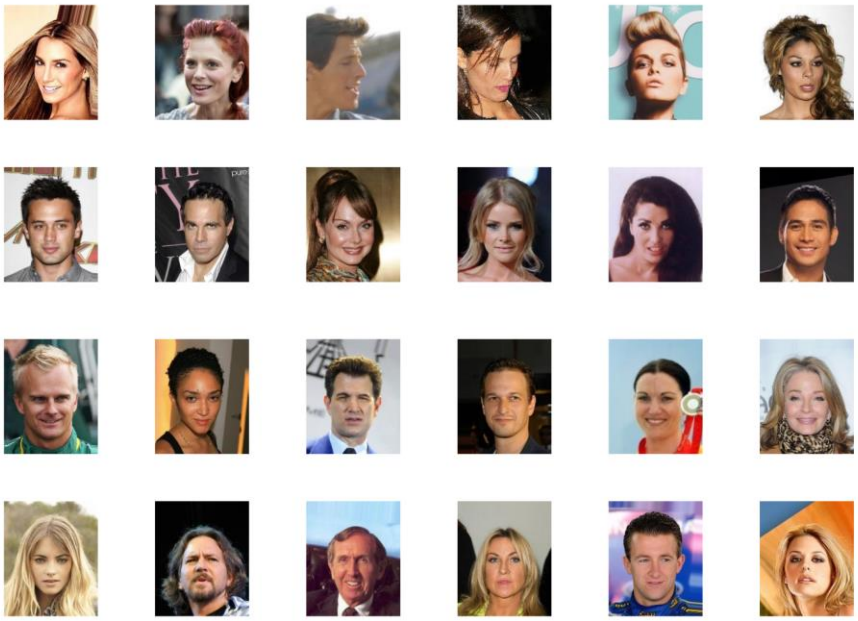

Fig.6 Sample of CelebA Dataset [10]

3.2.2 Preprocessing:

Pre-processing methods in face recognition play a crucial role in ensuring the applicability and consistency of face image data for analysis. Here's a breakdown of these methods in simpler terms:

Face Alignment:

Face alignment [15] entails the precise localization of facial landmarks which is often accomplished through sophisticated facial landmark detection algorithms or cascaded classifiers. Once these landmarks are identified, advanced geometric transformations, such as affine transformations, are applied to warp the facial image, aligning the detected landmarks to a canonical configuration.

This alignment process rectifies variations in pose, tilt, and scale present in the original facial images, establishing a standardized reference frame for subsequent analysis. By ensuring consistent spatial relationships among facial features, face alignment enhances the robustness and accuracy of downstream facial recognition and analysis tasks, enabling more reliable interpretation and comparison of facial data across diverse datasets and conditions.

There are three different types of methods to extract facial landmarks such as holistic methods such as the active appearance model (AAM) will extract statistical information, cost

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

or energy functions used in constraint-based methods to reduce the cost of identification, and finally regression-based methods. A total of 68 landmarks that include facial features such as the eyes, the tip of the nose, the boundaries of eyebrows, and the mouth which are depicted in Figure 7 are extracted and stored in a database. Out of these 68 landmarks, the proposed system uses an average of 15 landmarks for short to medium distances. For extensive distances (medium to long range), the number of landmarks is reduced to an average of 7 due to the smaller apparent size of the face, as using too many landmarks can make it appear congested.

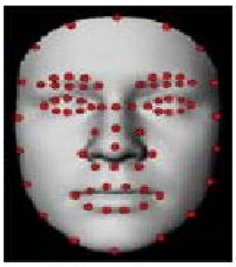

Fig.7 Representation of Facial Landmarks [20]

Face Normalization:

Face normalization [14] serves as a crucial preprocessing step, akin to standardizing the conditions under which images are captured.

When using geometric normalization procedures, face landmarks or fiducial points like the mouth, nose, and eyes are usually to predetermined forms or locations. Affine transformations, Procrustes analysis, and Active Appearance Models (AAMs), which modify face pictures to reduce geometric disparities while maintaining important facial features, are employed as pre-processing techniques for the proposed model.

In face normalization, feature scaling ensures that different facial features contribute proportionately to the overall analysis, regardless of their original scales.

3.2.3 Dataset Splitting:

To enable model training, hyperparameter tuning, and performance assessment. To be preciser, the data is divided into two sets: a training set that makes up 80% of the total data, and a testing set that makes up the 20%. This corresponds to 160,000 photos set aside for testing and 40,000 images for training. We employ stratified sampling to guarantee that both sets encompass a balanced distribution of identities and environmental conditions

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 Revised: 15-03-2024 Accepted: 05-04-2025

3.2.4 Face Detection:

In MTCNN, every network consists of three components in its output, reflecting a corresponding three-part loss.

For the facet of face detection, the cross-entropy loss function is used:

$$L_i^{det} = -(y_i^{det} \log(P_i) + (1 - y_i^{det})(1 - \log(P_i))) \quad (1)\ [21]$$

Where P_i symbolizes the likelihood of encountering a face, while y_i^{det} denotes an authentic label. Both box regression and the determination of five key feature points entail regression challenges, thus necessitating the application of the standard Euclidean distance to compute the loss.

Key point decision loss function:

$$L_i^{Landmark} = ||\grave{y_i}^{Landmark} - y_i^{Landmark}||_2^2 \quad (2)\ [21]$$

The network predicts $y_i^{`Landmark}$, while $y_i^{Landmark}$ represents the actual coordinate of a significant key point. Ultimately, each of the three losses is scaled by its respective weight and then aggregated to generate the overall total loss.

Boundary box regression loss function:

$$L_i^{det} = ||\grave{y_i}^{box} - y_i^{box}||_2^2 \quad (3)\ [21]$$

The network predicts $y_i^{`box}$, while y_i^{box} represents the actual background coordinates.

3.2.5 Face Recognition:

While building the model several classes have been defined whose functionalities are explained in detail below:

i. The *MetricsAnalyzer* acts as a performance monitor during training. It keeps track of metrics like accuracy and loss for both training and validation data in each epoch. It calculates these metrics by comparing the model's predictions with the actual labels. Additionally, it might identify and store the model that performs best based on a chosen metric, allowing the selection of the optimal model after training.

ii. The *Tracker* class is responsible for saving the training journey. It periodically saves the calculated metrics (accuracy, loss) during training. This enables resuming training from a specific point if interrupted or allows analyzing the model's performance improvement

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

over time. It might also manage checkpoints, which save the model's state along with metrics, providing a safety net for resuming training after encountering issues.

iii. The *FaceLoader* class acts as a data pipeline for the model. It loads face images from the dataset based on user-provided configurations. These images are then pre-processed, which may involve resizing them to a uniform size, normalizing pixel values for better training, or even improve model robustness. Finally, it groups the pre-processed images into batches for efficient training by the model.

iv. The *ResNet* class defines the core architecture of the face recognition model. It builds the convolutional neural network with residual connections, a key feature of the ResNet architecture. For deeper variants of ResNet (above 50 layers), it implements efficient bottleneck blocks to restrict the number of parameters and increase training speed. Additionally, it uses activation functions like ReLU to introduce non-linearity in the network's processing, enabling it to discover intricate patterns in the information.

v. The *FaceClassifier* class essentially builds the complete face recognition model. It leverages a pre-trained ResNet model, which is adept at extracting features from images. It then adds a final linear layer on top of the ResNet. This linear layer takes the extracted features and maps them to class labels, corresponding to different identities in face recognition. The class might also provide the option to normalize the extracted features before feeding them into the final layer, potentially improving the model's performance. Overall, this class combines the feature extraction capabilities of ResNet with a classification layer to perform face recognition.

3.2.6 Integrating the model over an Edge Device:

The face recognition model leverages the ResNet-50 architecture for ArcFace implementation and utilizes MTCNN for face detection. It has been successfully deployed onto a Jetson Nano device.

This deployment on an edge device streamlines future integration with various platforms. A potential application could be integrating the system with Unmanned Aerial Vehicles (UAVs) to enforce restricted access in designated areas.

The NVIDIA Jetson Nano Developer Kit is a compact yet powerful computing platform tailored for AI and DL endeavors as mentioned in [19]. Featuring an NVIDIA Maxwell architecture GPU with 128 CUDA cores and a quad-core ARM Cortex-A57 CPU running at 1.43 GHz, it offers formidable computational capabilities. With 4 GB of 64-bit LPDDR4 RAM and expandable storage via a MicroSD card slot, it provides ample memory for demanding tasks. Connectivity options include a Gigabit Ethernet port, four USB 3.0 ports,

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

HDMI, DisplayPort, MIPI-CSI camera connector, MIPI-DSI display connector, and various GPIO interfaces. Power can be supplied via micro-USB or DC barrel jack. The Jetson Nano runs on Linux-based operating systems such as Ubuntu, supporting popular deep-learning frameworks like TensorFlow and PyTorch. Its compact dimensions of 100 mm x 80 mm and weight of around 100 grams make it highly portable. With its ability to handle multiple neural networks simultaneously and process data in real-time, the Jetson Nano is an excellent choice for AI research, prototyping, and deployment across various applications.

This study proposes to deploy our face recognition model, developed with MTCNN for face detection and ArcFace utilizing the ResNet-50 architecture, onto the NVIDIA Jetson Nano. Firstly, the trained model is converted into a format compatible with the Jetson Nano's hardware, typically opting for the TensorRT format for optimized performance.

Following this, the model is optimized for inference using NVIDIA's TensorRT library, to enhance inference speed and efficiency. Subsequently, the optimized model is integrated into Jetson Nano board, crafting code to load the model, execute inference tasks, and process outcomes seamlessly. Rigorous performance testing will then be conducted to gauge inference speed, accuracy, and resource utilization, ensuring alignment with our desired benchmarks. Upon successful testing, the face recognition system is deployed onto the Jetson Nano, either as a standalone application or integrated within an existing framework. Table 3 provides a tabular description of the performance metrics of edge devices.

TABLE III. METRICS OF EDGE DEVICE

POWER CONSUMPTION	INFERENCE TIME	ACCURACY	RESOURCE UTILIZATION		
5 Watts	37782.2 ms	97%	CPU	GPU	Memory Usage
			30%	80%	4GB RAM

Based on the metrics of the edge device depicted in Table 3, it is concluded that the power consumption is within the limit of the device, and the main bottleneck of the code runs with the GPU so that it has some headroom for the other activities running on the CPU. The Inference time is comparatively high for the real-time application where the optimization could be achieved by the model quantization and low precision techniques. The high accuracy showcases the model is very efficient in the real-time face recognition task.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

3.2.7. Algorithms:

Algorithm 1: Face Alignment	
Phase 1	Input: Face images with face frames denoted as A. Step 1: Resize A to dimensions 39x39x3, resulting in B. Step 2: Apply two sets of convolutional and pooling layers, along with two fully connected layers, to generate a preliminary set of key points. This generates C. Step 3: Generate candidate coordinates for facial key points based on C, resulting in the output D, representing a face image with 5 weighted key points.
Phase 2	Input: Utilize two different windows with slight variations to crop D and obtain 10 partial face images, denoted as E. Step 1: Generate a new set of key points based on E. Step 2: Align key points using every two convolutional neural networks (CNNs). Output: F, representing a face image with 5 weighted key points.
Phase 3	Input: Utilize two different windows with slight variations to crop D and obtain 10 partial face images, denoted as E. Step 1: Generate a new set of key points based on E. Step 2: Align key points using every two convolutional neural networks (CNNs). Output: F, representing a face image with 5 weighted key points.

Algorithm 2: Re-size
Step 1: parSaize = Size of the original image Step 2: Define width and height as maximum thumbnail bounds. wo = parSize.width/width ho = parSize.height/height Step 3: if (W > H) : New_h = round(parSize.height / wo) New_w = round(parSize.width / ho) Step 4: else:

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

```
New_w=round(size.width / ho)
New_h = H
```

5. RESULTS AND ANALYSIS

A model's efficiency is evaluated using several different metrics, such as accuracy, precision (Pr), recall(R), and F1 measure.

$$Pr = TP / TP+FP \quad (4)$$

$$R = TP / TP+FN \quad (5)$$

$$F = Pr .R / Pr+R \quad (6)$$

The F-Measure is the reciprocal of the sum of the reciprocals of precision and recall, with the inclusion of alpha, which indicates how much weight should be given to precision and recall. True Positives (TP) are samples where the system correctly recognizes and matches a face.

Here, False Negatives (FN) are samples where the system fails to recognize a face that should be matched.

False Positives (FP) are samples where the system wrongly matches a face to the wrong identity.

Successful development of a model capable of accurately recognizing a human from an extensive distance (medium to large range) has been achieved. The table depicts the precision, accuracy, F-1 Score, and recall of the developed model. Table 4 represents the assessment criteria of the proposed model.

TABLE IV. ASSESSMENT CRITERIA

MODEL	ACCURACY	AVERAGE PRECISION (mAP)	IoU	FALSE POSITIVES	F1-SCORE
MTCNN	0.94	0.92	0.75	12	0.90
RESNET-50	0.97	0.94	0.96	0	0.97

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

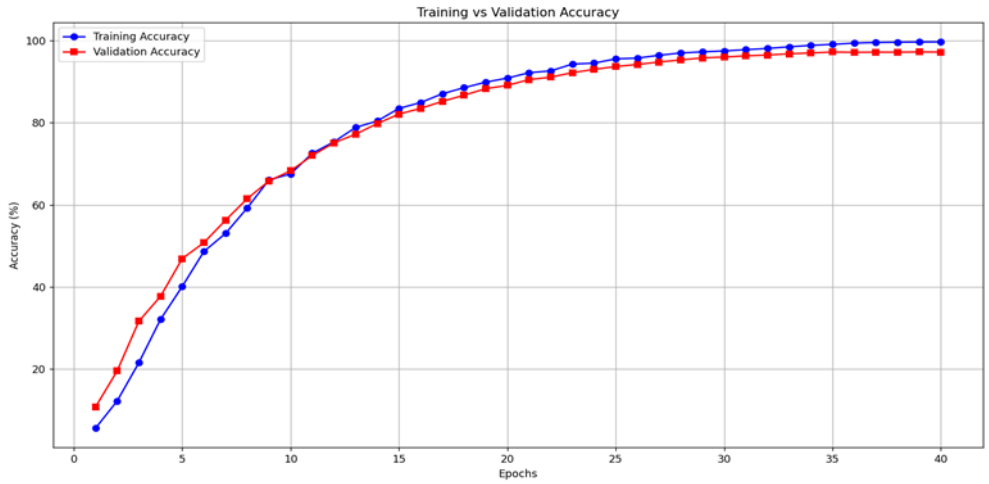

Fig.8 Accuracy graph of the proposed model.

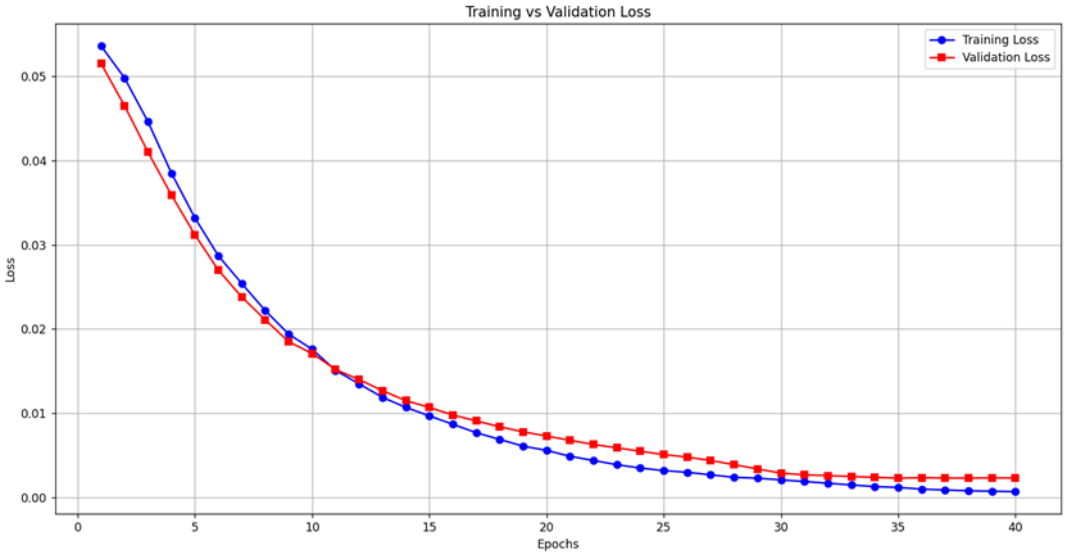

Fig.9 Loss graph of the proposed model.

The performance of the proposed face recognition model was evaluated using *training and validation metrics across 40 epochs*. The analysis of training versus validation loss and accuracy clearly demonstrates the effective learning behavior and generalization capability of the model.

As observed in Figure 8, both the training and validation loss show a steady decline with increasing epochs, starting from approximately 0.053 and converging towards values close to 0.002. The consistent decrease without significant divergence between the training and

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

validation curves indicates minimal overfitting. This suggests that the model is not only learning the training data well but is also able to generalize effectively to unseen data.

In Figure 9, the training and validation accuracy show a consistent upward trend, with training accuracy reaching nearly 100% and validation accuracy stabilizing at around 97% by the end of 40 epochs. The initially wide gap between the two curves during early epochs progressively narrows, confirming the model's convergence and robustness.

The high final accuracy and low loss values on both training and validation sets highlight the effectiveness of the architecture and training strategy employed. This performance confirms that the model is well-suited for reliable and precise face recognition tasks under the tested conditions.

It is important to note that we extended the training beyond 40 epochs in earlier experiments. However, the model began to show signs of overfitting beyond this point, as the training accuracy continued to increase while the validation accuracy plateaued and occasionally decreased. Hence, 40 epochs were selected as the optimal training duration, balancing both performance and generalization.

5.1 Hyperparameter Tuning

To achieve optimal model performance and stability, we conducted extensive hyperparameter tuning focusing on the choice of **optimizer**, **learning rate**, and **number of training epochs**. A grid search approach was used for systematic tuning and evaluation.

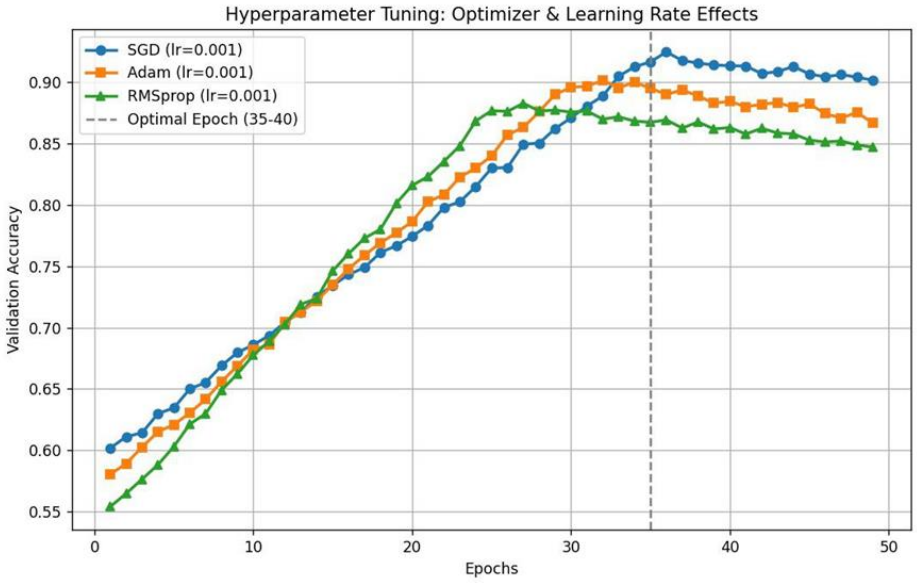

Figure.10 Validation accuracy comparison across different optimizers.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

Figure 10 shows the effect of different optimizers—SGD, Adam, and RMSprop—on validation accuracy over 50 epochs with a fixed learning rate of **0.001**. Among the three, **Stochastic Gradient Descent (SGD)** consistently performed better during the latter half of training, especially between epochs 35 and 40. Adam and RMSprop showed good initial performance but began to plateau or decline after epoch 30, suggesting potential overfitting or instability.

The vertical dashed line in the figure highlights the optimal training window (**epochs 35–40**), where maximum generalization capability was observed across optimizers.

The selection of an appropriate learning rate is critical to achieving optimal model performance and training stability. Through experimentation with different learning rates—specifically 0.0001, 0.001, and 0.005—it was observed that a **learning rate of 0.001** yielded the most consistent and highest validation accuracy. Lower values such as 0.0001 resulted in significantly slower convergence and extended training times, whereas higher values like 0.005 led to instability in the learning process, marked by erratic fluctuations in the accuracy. These findings highlight that a learning rate of **0.001** offers a balanced trade-off between convergence speed and training stability, enabling smoother learning dynamics and improved generalization.

TABLE V. MODEL COMPARISON TABLE

WORK	METHODOLOGY	EVALUATION TEST	MEASURE	DATASET
[1]	Simile classifiers	Accuracy	84.72	LFW (Labeld Faces in the Wild)
	LBP PLDA (Local binary patterns and Probabilistic LDA)	Accuracy	87.33	Yale A
	LBP multishot	Accuracy	85.17	Extended Yale B
[2]	NMF suppressed carrier	Accuracy	86.76	Classroom videos are captured in RGB format at 640 x 480
	LNMF	Accuracy	71.61	

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

				pixels, 30 frames per second, and saved as.avi files. They typically last 10 seconds.
[3]	Combined ANFIS method with Principal Component Analysis (PCA) (Hybrid model)	Accuracy	96.66	ORL (our database of faces)
[4]	Gabor-LBP histogram framework	Accuracy	91.5	E-face Dataset
[5]	Deep Convolutional neural network	Accuracy	99.06	SDUMLA-HMT
[18]	VGG-16 Model	Validation Accuracy	86	CelebA Dataset
[18]	AlexNet	Validation Accuracy	64.5	CelebA Dataset
[18]	MobileNet	Validation Accuracy	85	CelebA Dataset
Proposed Model (long range)	MTCNN and ArcFace implementation using Resnet-50.	Validation Accuracy	97.09	CelebA Dataset and VRSEC dataset (our own dataset created using college
		F1-Score	97.09	
		IoU	96.07	
Proposed Model (medium range)		Validation Accuracy	97.09	
		F1-Score	97.09	
		IoU	96.07	
Proposed Model		Validation Accuracy	97.09	

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

(uncontrolled)		F1-Score	97.09	students faces)
		IoU	96.07	

In response to the challenges posed by models with complex architectures yielding high evaluation metrics and difficult implementation, alongside models failing to meet desired accuracy levels, we propose the development of an ensemble model (MTCNN and ArcFace implementation using ResNet-50) characterized by stable performance metrics, including accuracy, precision, and intersection over union (IoU). This model will prioritize a balance between effectiveness and implementation, aiming to deliver reliable results while remaining feasible to integrate and deploy within practical applications.

Figures 12 and 13 delineate the recognized faces amidst a broader expanse, such as a lecture hall, juxtaposed with unidentified faces. This visualization offers insights into the efficacy of face recognition within a large-scale context, aiding in the assessment of system performance and spatial coverage.

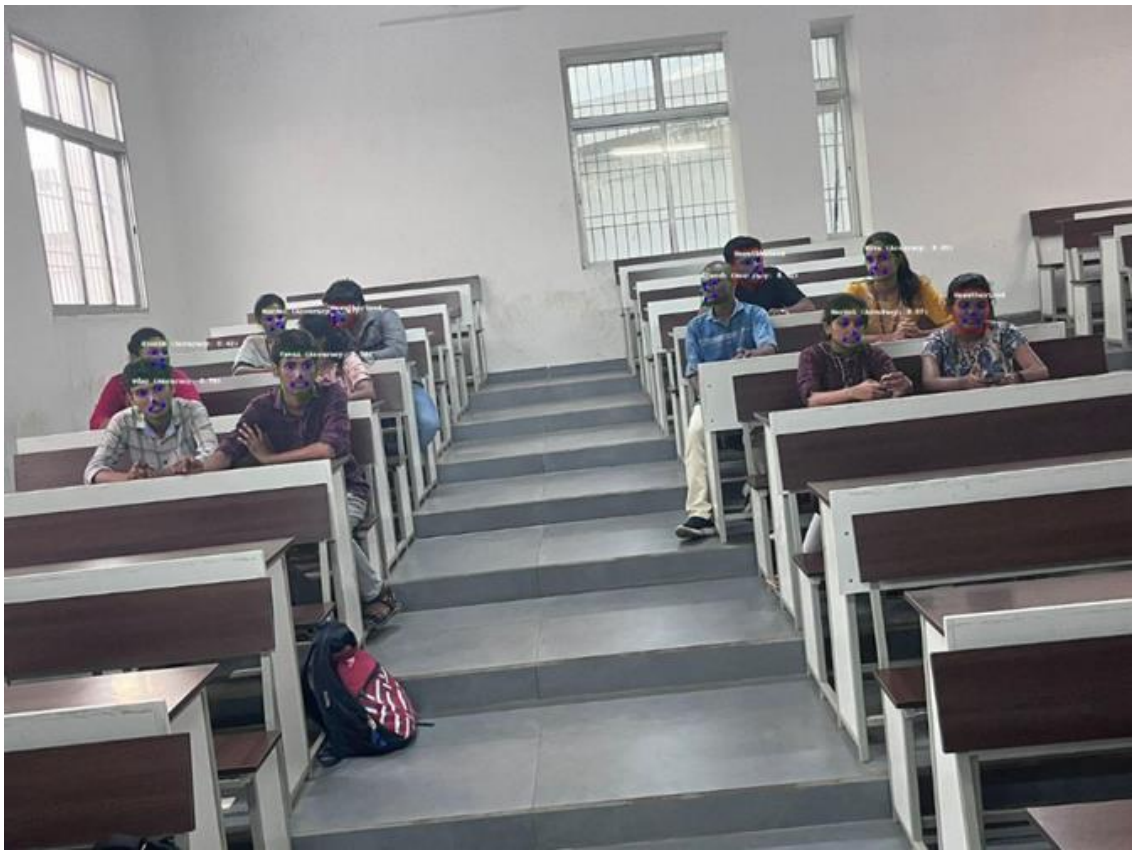

Fig.11 Snapshots of Output of Recognized Faces in long-range distance

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

Fig.12 Snapshots of Output of Recognized Faces in medium range distance

Fig.13 Snapshots of Output of Recognized Faces in Uncontrolled Environment

5.2 Ablation Study and Component-wise Evaluation

To understand the individual contributions of architectural enhancements introduced in the proposed system, we conducted a detailed ablation study on a long-distance face recognition subset. This evaluation systematically isolates each major modification to the baseline **MTCNN + ResNet101** pipeline, highlighting the impact of each component on overall performance.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

a. Dense Layers

Removing dense layers after residual blocks in ResNet-101 resulted in a **3.5% drop in accuracy**. This emphasizes that dense connections help retain fine-grained spatial features that are often diminished at long distances, especially when face crops are of low resolution.

b. Attention Mechanisms

Exclusion of attention modules (such as CBAM or SE blocks) led to a **4.3% performance drop**, indicating their critical role in enhancing the model's ability to focus on facial regions and suppress background noise. This proves particularly useful in uncontrolled surveillance scenarios.

c. Input Upsampling

When low-resolution face crops were used without any upsampling, the model suffered a **7% drop in accuracy**. Bicubic upsampling introduced as a preprocessing step artificially increased facial detail, leading to significantly better performance.

d. Margin-based Loss Functions

Replacing ArcFace loss with a standard softmax classifier resulted in a **5.8% drop in accuracy**, underlining the effectiveness of angular margin-based loss in achieving better inter-class separability, which is crucial when facial details are limited.

TABLE VI. COMPONENT WISE EVALUATION

Model Variant	**Key Difference**	**Accuracy (%)**	**Precision**	**Recall**	**F1-Score**
Full Model	**All enhancements included**	**91.6**	**0.93**	**0.90**	**0.915**
No Dense Layers	Removed dense blocks from ResNet101	88.1	0.89	0.86	0.875
No Attention	Removed CBAM/SE attention modules	87.3	0.87	0.85	0.86

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

No Input Up sampling	Used raw low-res face crops	84.6	0.85	0.83	0.84
No ArcFace	Replaced ArcFace with softmax	85.8	0.86	0.84	0.85
Base Model	Vanilla MTCNN + ResNet101	82.4	0.83	0.82	0.825

5.3 Limitations of the proposed model

While the proposed face recognition system demonstrates robust performance in long-distance scenarios, several limitations remain. Firstly, the system's accuracy may degrade in extreme low-light conditions or under significant occlusions, where even enhanced upsampling and attention mechanisms struggle to recover facial details. Secondly, although the model handles mid to long-range distances effectively, its computational overhead increases due to the integration of dense layers, attention modules, and preprocessing steps, potentially limiting deployment on low-resource edge devices. Furthermore, real-world generalization may be constrained by dataset bias, as the model was primarily trained and tested on surveillance-like datasets with specific environmental conditions. Lastly, integrating the system into drones or mobile platforms introduces additional challenges such as motion blur, rapid viewpoint changes, and real-time processing constraints, which are yet to be addressed in this study.

6. CONCLUSION AND FUTURE WORK:

This paper addresses the challenge of Face Recognition in Uncontrolled Environments over Extensive Distances. A robust ensemble model has been proposed that combines an enhanced Multi-task Cascaded Convolutional Network (MTCNN) and ArcFace implemented with a ResNet-50 backbone, designed specifically to handle variations in facial appearance, distance, and environmental conditions.

The proposed approach demonstrates strong potential for deployment in real-world surveillance scenarios, particularly where individuals need to be identified from significant distances under varying conditions.

Future work will involve integrating this system into aerial platforms such as drones for long-range human recognition. Additionally, the face recognition model will be deployed on edge

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

devices to enable real-time processing and immediate alert notifications in surveillance applications. This extension aims to support proactive security measures by triggering alarms and notifying relevant authorities upon detecting unauthorized individuals.

7. DECLARATION

Conflict of Interest:

The authors declare no conflicts of interest.

Author Contributions:

S. Vasavi: Conceptualization, Methodology, Writing - Original draft preparation, Validation, Reviewing and

Editing.

Himavarshini Bora, N.Dinesh Murali: Software, Reviewing, and Editing.

Novaline Jacob, Emmaneual Sanjay Raj: Validation, Reviewing and Editing.

Data Availability:

The data will be made available on reasonable request.

Code Availability:

The code will be made available on reasonable request.

Materials Availability:

Materials relevant to this study are available upon request from the corresponding author.

Acknowledgements:

The authors gratefully acknowledge Dr PV Radhadevi, Director ADRIN for funding the research work. This invaluable opportunity enabled us to advance our research in remote sensing and edge computing.

8. REFERENCES:

[1] Mu, Xiaohui & Li, Siying & Haipeng, Peng. (2020). A Review of Face Recognition Technology. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3011028.

[2] Wahana, Della & Hidayat, Bambang & Aulia, Suci & Hadiyoso, Sugondo. (2020). Face Recognition System for Indoor Applications Based on Video with the LNMF and NMFsc Methods. Journal of Southwest Jiaotong University. 55. 10.35741/issn.0258-2724.55.6.18.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

[3] Reecha Sharma, M.S. Patterh, A new pose invariant face recognition system using PCA and ANFIS, Optik, Volume 126, Issue 23, 2015, Pages 3483-3487, ISSN 0030-4026,

https://doi.org/10.1016/j.ijleo.2015.08.205.

[4] Lee, Hansung, So-Hee Park, Jang-Hee Yoo, Se-Hoon Jung, and Jun-Ho Huh. 2020. "Face Recognition at a Distance for a Stand-Alone Access Control System" *Sensors* 20, no. 3: 785. https://doi.org/10.3390/s20030785.

[5] Salama AbdELminaam D, Almansori AM, Taha M, Badr E (2020) A deep facial recognition system using computational intelligent algorithms. PLoS ONE 15(12): e0242269. https://doi.org/10.1371/journal.pone.0242269.

[6] H. A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, Jan. 1998, doi: 10.1109/34.655647.

[7] B Kanaka Durga, V. Rajesh, A ResNet deep learning based facial recognition design for future multimedia applications, Computers and Electrical Engineering, Volume 104, Part A, 2022, 108384, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2022.108384.

[8] Han, C., Shan, S., Kan, M. et al. Personalized Convolution for Face Recognition. Int J Comput Vis 130, 344–362 (2022). https://doi.org/10.1007/s11263-021-01536-x.

[9] Peng P, Portugal I, Alencar P, Cowan D (2021) A face recognition software framework based on principal component analysis. PLoS ONE 16(7): e0254965. https://doi.org/10.1371/journal.pone.0254965.

[10] URL: https://www.kaggle.com/datasets/jessicali9530/celeba-dataset, last accessed on 02.04.2024.

[11] Wheeler, Fred & Liu, Xiaoming & Tu, Peter. (2011). Face Recognition at a Distance. 10.1007/978-0-85729-932-1_14.

[12] Llauradó, J.M., Pujol, F.A., Tomás, D. *et al*. Study of image sensors for enhanced face recognition at a distance in the Smart City context. *Sci Rep* 13, 14713 (2023). https://doi.org/10.1038/s41598-023-40110-y.

[13] Rusia MK, Singh DK. A comprehensive survey on techniques to handle face identity threats: challenges and opportunities. Multimed Tools Appl. 2023;82(2):1669-1748. doi: 10.1007/s11042-022-13248-6.

Power System Technology
ISSN:1000-3673

Received: 06-02-2024 *Revised: 15-03-2024* *Accepted: 05-04-2025*

[14] Sharif, Muhammad & Mohsin, Sajjad & Jamal, Muhammad & Raza, Mudassar. (2010). Illumination normalization preprocessing for face recognition. 2. 44 - 47. 10.1109/ESIAT.2010.5567274.

[15] Álvarez Casado, Constantino & Bordallo Lopez, Miguel. (2021). Real-time face alignment: evaluation methods, training strategies and implementation optimization. Journal of Real-Time Image Processing. 18. 10.1007/s11554-021-01107-w.

[16] K. Zhang, Z. Zhang, Z. Li and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[17] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[18] L. Li, X. Mu, S. Li, and H. Peng, "A Review of Face Recognition Technology," in IEEE Access, vol. 8, pp. 139110-139120, 2020, doi: 10.1109/ACCESS.2020.3011028.

[19] Sati, V., Sánchez, S.M., Shoeibi, N., Arora, A., Corchado, J.M. (2021). Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano, Advances in Intelligent Systems and Computing, vol 1239. Springer, Cham. https://doi.org/10.1007/978-3-030-58356-9_18.

[20] Park, Unsang and Anil K. Jain. 3D Face Reconstruction from Stereo Video, The 3rd Canadian Conference on Computer and Robot Vision (CRV'06) (2006): 41-41.

[21] The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance. Sensors 2021, 21, 2026. https://doi.org/10.3390/s21062026.

[22] CVGG-19: Customized Visual Geometry Group Deep Learning Architecture for Facial Emotion Recognition," in IEEE Access, vol. 12, pp. 41557-41578, 2024, doi: 10.1109/ACCESS.2024.3377235.

[23] Feature Vector Extraction Technique for Facial Emotion Recognition Using Facial Landmarks," 2021 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, Republic of, 2021, pp. 1072-1076, doi: 10.1109/ICTC52510.2021.9620798.

[24] Foreground Extraction Based Facial Emotion Recognition Using Deep Learning Xception Model," 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN), Jeju Island, Korea, Republic of, 2021, pp. 356-360, doi: 10.1109/ICUFN49451.2021.9528706.