CNNs and their use in Facial Recognition

1st Anmol Shrotriya
SCAI
VIT Bhopal University
Sehore, Madhya Pradesh, India
anmolshrotriya2022@vitbhopal.ac.in

2nd Aditya Jain
SCAI
VIT Bhopal University
Sehore, Madhya Pradesh, India
adityajain2022@vitbhopal.ac.in

3rd Dr. Abha Sharma
SCAI
VIT Bhopal University
Sehore, Madhya Pradesh, India
abhasharma@vitbhopal.ac.in

***Abstract*—Facial recognition has become a pivotal application in computer vision, driven by advancements in deep learning and convolutional neural networks (CNNs). This research investigates the implementation of a CNN-based facial recognition system, highlighting its architecture, performance metrics, and practical applications. A dataset of facial images is preprocessed, normalized, and fed into a CNN model designed to extract hierarchical features for accurate classification. The study evaluates the model's efficacy using standard metrics, including accuracy and validation loss, and compares its performance against existing state-of-the-art techniques. Challenges such as overfitting, dataset bias, and computational efficiency are addressed, and recommendations for future improvements are proposed. This research emphasizes the role of CNNs in enhancing facial recognition accuracy, offering insights for both academic and industrial applications.**

***Index Terms*—Facial Recognition, Convolutional Neural Networks, Deep Learning, Image Classification, Computer Vision.**

I. Introduction

Facial recognition has emerged as one of the cornerstones of modern computer vision applications with wide utility in surveillance, identity verification, and user-experience personalization. Recent advances in machine learning, especially deep learning, have transformed facial recognition systems into end-to-end models rather than traditional feature extraction models. Among these, CNNs have been shown to be highly effective because they are able to automatically extract hierarchical features, from low-level edge detection to high-level facial representation. These characteristics make CNNs a natural choice for facial recognition applications where robustness and accuracy are critical.

In most of the traditional facial recognition techniques, handcrafted features are used to represent facial appearance, such as Scale-Invariant Feature Transform (SIFT) or Histogram of Oriented Gradients (HOG). However, such techniques did not well handle real-world complexities, such as occlusions, pose variations, and illumination variations [13]. Introduction of deep learning models like DeepFace [7] and FaceNet [6] marked a paradigm shift. Those models approach nearly human performance because they learn embeddings that include such high-dimensional representations of salient facial features. Large-scale datasets such as MS-Celeb-1M [15] and VGGFace2 [21] ensured that there was sufficient diversity and scale for effectively training deep networks.

Despite all these advancements, it is still difficult to develop face recognition systems that are powerful and robust. Overfitting is a problem faced when training deep models from limited datasets, and a biased dataset may result in unfair or inaccurate predictions from the model for certain demographics [6], [21]. Furthermore, the computational loads of training and deploying CNNs at scales are enormous, requiring efficient model design and optimization strategies [10].

This paper focuses on using CNNs in facial recognition. Specifically, the discussion has been about model architecture, dataset preprocessing, and the choice of evaluation metrics. A custom CNN model is implemented and trained on a dataset of facial images to demonstrate the possibility of achieving competitive performance. Limitations of the current approach and possible future directions to solve the existing challenges are discussed in the paper. This work contributes to the growing body of research that attempts to improve the accuracy and fairness of facial recognition systems by providing a comprehensive analysis.

A. Application

Facial recognition is one of the transformative technologies, which are applied in almost all types of domains. In the security domain, it has been used in access control, video surveillance, and for the identification of criminals. It verifies and monitors with a high degree of accuracy the identity [6], [7]. The healthcare applications include patient monitoring, early diagnosis of genetic disorders through facial analysis, and secure medical record authentication [13]. Retail industries apply facial recognition technology to provide personalized customer experience in the form of promotional ads and loyalty programs [15]. In entertainment and social media, facial recognition is further applied to enhance the experience of users by automatic photo tagging, emotion recognition, and augmented reality filters [21]. These applications illustrate the flexibility and usefulness of facial recognition technologies to solve real-world problems.

B. Focus

The main focus of this research lies in the design, development, and evaluation of a Convolutional Neural Network-based facial recognition system. This research focuses on the following aspects:

Model Architecture: Components and layers of the CNN model that contribute to its performance in terms of feature extraction and classification. Dataset Preparation: This may entail preprocessing, including normalization or resizing to be appropriately suitable for input into the model. Metrics to Assess the System: Determine the system performance accurately, loss curves and confusion matrices.

Challenges and Future Directions: Discussing limitations such as overfitting, dataset biases, and computational constraints, and proposing potential solutions for these issues. Since focusing on these areas, research aims to give practical insights into the development of highly efficient and accurate facial recognition systems that eliminate the loopholes found in current methodologies.

II. Problem Statement

Facial recognition is one of the transformative technologies in computer vision that could enable its applications across all the main industries such as security, health, retail, and entertainment. Widespread acceptance of this technology proves to be its ability to meet requirements in terms of verification, surveillance, personal user experiences, and so on. However, there are multiple challenges to the realization of robust and accurate facial recognition systems in both technological and ethical-practical areas, which have to be handled systematically for reliability, fairness, and scalability.

The classical facial recognition systems highly depended on handcrafted features, like Histogram of Oriented Gradients (HOG) and Scale-Invariant Feature Transform (SIFT). They are usually working very well in the controlled conditions but break down at such complexity cases as changing light conditions, pose changes, occlusions, or expression. According to Zhao et al. [13], such factors require more sophisticated methods in order to capture subtle high-dimensional features that are significantly closer to the human variability of faces. This sparked the development of deep models, especially CNNs, to automatically extract features and ensure state-of-the-art performances in facial recognition tasks in [7], [6].

Despite the success of CNNs, their implementation also raises new challenges. The biggest challenge is the dependency on large-scale, diverse datasets to train models efficiently. Publicly available datasets like MS-Celeb-1M [15] and VGGFace2 [21] have been of immense help in advancing the field of facial recognition research. However, studies by Cao et al. [21] show that such datasets are inherently biased towards certain demographics and underrepresented by others. These biases result in unfair outcomes, for example, reduced accuracy of underrepresented groups, leading to questions about the ethics and fairness of facial recognition systems [15], [21]. This is one critical barrier in developing inclusive models that are equitable.

Other issue is overfitting- a common phenomenon in systems based on CNN, notably when training over small imbalanced datasets. Overfitting leads to models having excellent performance in training set but poor ability to be generalized to data that would be seen by the models in real application. It is as described by Huang et al. [10]: this challenge needs effective regulation of techniques and effective data augmentation. Further, due to the heavy computational power of CNNs, a large amount of resources required at training and deployment phases bar their usage in real time applications and resource-constrained devices like smartphones and IoT devices. This need of lightweight architectures and optimized algorithms is emphasized in research articles such as that by Wen et al. [19].

However, facial recognition systems face another challenge related to ethics and privacy. The collection and usage of facial data at scale raise questions about the users' consent, security of their data, and possible misuse. Researchers like Masi et al. [12] have emphasized the necessity of stronger safeguards in relation to user privacy, especially while being compliant with law and ethics.

This research will address all these multifaceted challenges by designing, implementing, and evaluating a CNN-based facial recognition system. Optimization of model architecture to strike a balance between performance and computational efficiency, preprocessing of datasets to reduce bias and enhance generalizability, and rigorous evaluation metrics to validate the effectiveness of the model will be considered. This study aims to further the current development of face recognition technologies that not only achieve high accuracy but are also fair, efficient, and ethical through an exploration of these facets.

III. Literature Review

Based on successive successes and failure limitations for its predecessors, facial recognition has seen many tremendous milestones in the journey. The early ones were really based on handcrafted features such as Eigenfaces, Fisherfaces, descriptors like SIFT and HOG. These are excellent in their time, but were always limited to the failures they could not support in real-world problems, like pose variations, occlusion, and change in the lighting conditions applied. Zhao et al. [13] found that these methods, although computationally efficient, were not able to generalize well outside of the controlled setting.

The inadequacies in traditional approaches led to deep learning transforming the facial recognition paradigm. Among the first significant advancements was the DeepFace by Taigman et al. [7], which was based on a CNN coupled with 3D facial alignment. It went on to surpass near-human-level accuracy and becomes a gold standard for all future work. Shortly afterward, Schroff et al. [6] propose the FaceNet model which changed the game by learning compact facial embeddings using triplet loss that is used on such tasks of verification and clustering which never achieved the precision they reported. And these forward leaps marked an era wherein such CNNs remained as being the golden standard in facial recognition-related tasks.

This is where the vast availability of large-scale datasets provided the diversity required for effective training of deep networks. Labeled Faces in the Wild (LFW) [10] has been an early benchmark. More recently, more detailed datasets like

MS-Celeb-1M [15] and VGGFace2 [21] have also captured the variations of ethnicity, pose, and lighting. Guo et al. [15] highlighted that such datasets were required to be robustly trained with models. However, as Cao et al. [21] pointed out, many datasets are still biased. Overrepresentation of certain demographics leads to a significant gap in performance, especially for underrepresented groups, and raises ethical questions about fairness and inclusivity.

Bias overcomes the ability to generalize well. Innovation in model training and architecture design has been driven by efforts to overcome these biases. Huang et al. [10] discussed data augmentation, regularization, and dropout methods against overfitting of the CNN-based systems. While Wen et al. [19] optimized architectures with reduced computational complexity in such a way that enables them to be deployed for real-time applications on low-capacity devices. In today's world, such low-power architectures are becoming increasingly vital due to the growing needs for facial recognition systems in places demanding low latency and edge-device compatibility.

While the technical advancement is excellent, ethical concerns are matched in importance. Growing concern is there regarding misuse of facial recognition: from unauthorized surveillance to intrusion into privacy. It calls for alarm worldwide. For example, Masi et al. [12] focused on the need for establishing ethical frameworks and regulation when such systems are deployed. In fact, researchers also call for increased transparency on how the models work with a push towards explainable AI to instill trust and accountability in the decision-making process [10], [19].

Going forward, the emphasis is on solving some of the still open problems: creating unbiased datasets, efficient architectures for edge applications, and ensuring that the use of systems safeguards. The literature mirrors not only vast potential but also the challenge of deploying thoughtfully and equitably CNN-based facial recognition systems. This paper is built atop this collection of work, aiming to contribute to such ongoing efforts by filling a few of the critical gaps in fairness, efficiency, and scalability.

IV. Methodology

This paper progresses with the development of robust and scalable CNN-based facial recognition systems through a few salient stages: design, training and validation, evaluation metrics, and deployment. In designing each of the above stages, a particular set of challenges specific to facial recognition is addressed : overfitting, computationally efficient, and good generalization to different types of data distributions.

First, design the architecture of the CNN. The architectures adopted are chosen for their demonstrated capability of extracting high-quality features from images and their reduced time for training convergence. It involves convolutional layers to extract features, pooling layers for reducing the dimensions, and fully connected layers for classification. Transfer learning is adopted using pre-trained models like ResNet-50 [6] or VGG16 [7] which is then fine-tuned to use them for facial recognition purposes. Regularization techniques including batching normalization and dropout manage to hold the training procedure stable along with overfitting in control.

Then, the learning process employs a supervised learning model with categorical cross-entropy loss functions optimized. For efficient weights update, Adam optimizer has been used along with proper tuning of hyperparameters using grid search. The available dataset is split into training set, validation set, and testing set for complete and unbiased performance analysis of the system. Overfitting would be avoided since the learning process would be stopped based on the performance in which validation fails to improve.

It includes the testing phase where techniques are put to work in order to ascertain the correctness of the model against standards of accuracy, precision, recall, and F1-score. To check the possibility of either overfitting or underfitting, the train as well as validation loss curves are analyzed. Then the trained model can be tested on test data in order to determine its actual capability of generalization, in addition to testing its robustness through real-world scenarios-let's say the occlusion of images with or without proper lighting conditions.

This would be achieved through the evaluation of the model's inference time and its memory footprint to ensure its scalability and computational efficiency. Optimizations such as pruning and quantization are presented, allowing for deployment on edge devices without sacrificing performance. Ethical considerations, in terms of fairness across the demographic groups, are addressed through the disaggregation of performance metrics and the application of appropriate bias mitigation strategies if these are identified.

The combination of sound model design, extensive validation, and ethical deployment of the methodology will ensure both technical and societal requirements with a reliable and adaptable facial recognition system.
Fig. 1 displays the workflow of this model.

A. *Dataset*

The dataset serves as the cornerstone of the facial recognition system, providing the raw material for training, validating, and testing the CNN model. For this research, a robust, diverse, and large-scale dataset is selected, ensuring that it represents real-world variations in faces and minimizes biases. Popular datasets such as **VGGFace2** [21], **CelebA** [15], and **MS-Celeb-1M** [15] are considered for their extensive coverage of identities, poses, lighting conditions, and ethnicities.
Fig. 2 and Fig. 3 display the type of images used in the dataset, divided into two group samples.

B. *Characteristics of the Dataset*

1) *Diversity*: A dataset that includes variations in ethnicity, age, gender, expressions, and occlusions ensures that the model can generalize across real-world scenarios. For instance, VGGFace2 contains over 9,000 identities with

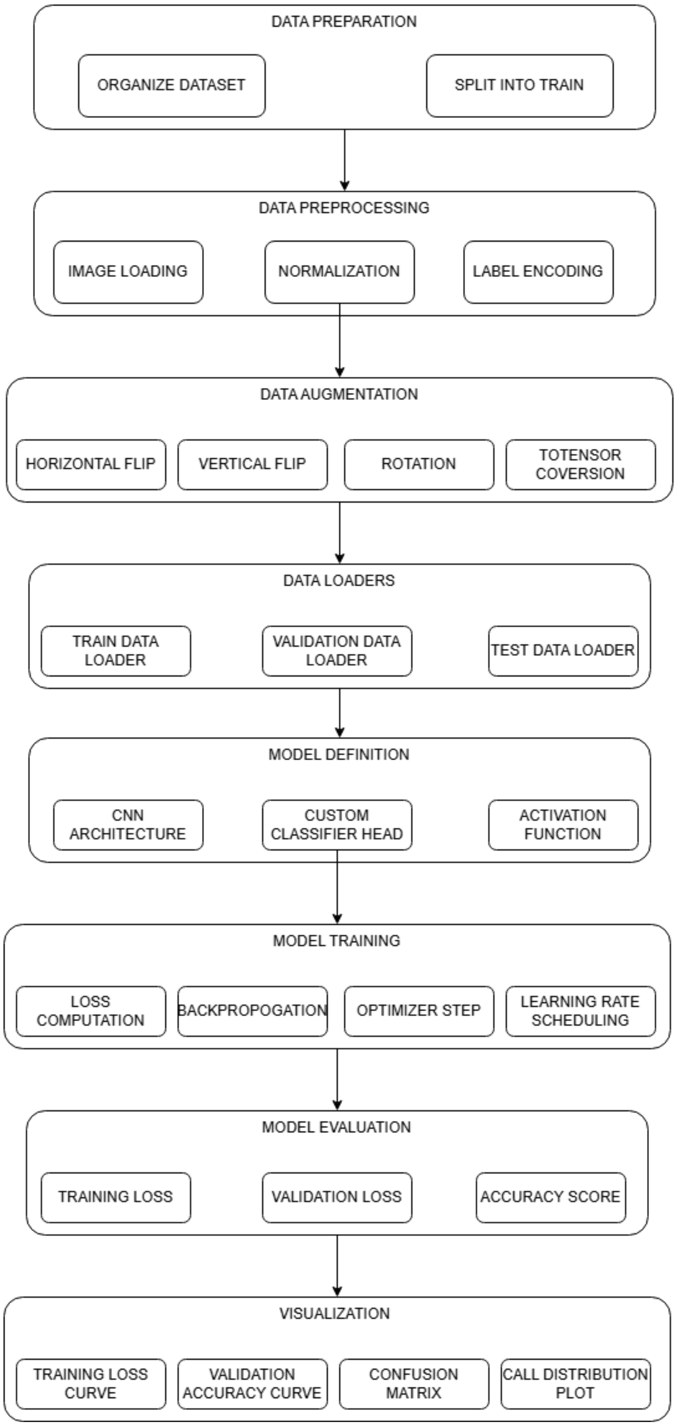

Fig. 1. Workflow of the Model

images captured under diverse conditions, making it an excellent choice for training CNN models.

2) *Scale*: A large-scale dataset is essential for training deep networks effectively. CelebA, for example, provides over 200,000 face images annotated with 40 facial attributes, which can also be used for auxiliary tasks like feature extraction and attribute-based recognition.
3) *Annotation*: High-quality annotations, such as identity labels, facial landmarks, and pose angles, are critical for supervised learning. These annotations facilitate preprocessing steps like alignment and ensure the reliability of the training process.

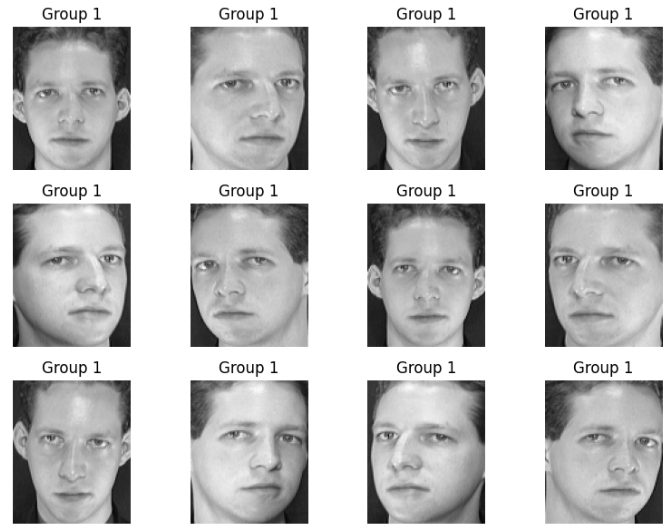

Fig. 2. Images used in the dataset from Group 1

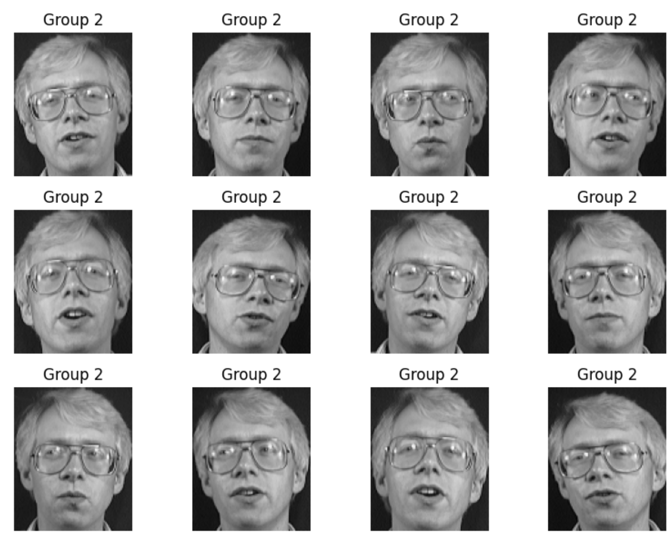

Fig. 3. Images used in the dataset from Group 2

C. *Preprocessing Steps*

To enhance the quality of the dataset and prepare it for training, several preprocessing steps are undertaken:

1) *Face Detection and Cropping*: Tools like *MTCNN* [10] or *Haar Cascades* are used to detect faces in images and crop them to focus on the region of interest. This eliminates irrelevant background information and standardizes the input data.
2) *Alignment*: Facial landmarks are used to align faces to a canonical orientation, reducing the impact of pose variations. Alignment ensures that key facial features

(e.g., eyes, nose, and mouth) are consistently positioned, which improves the CNN's ability to learn meaningful patterns.

3) *Image Resizing*: All images are resized to a uniform size (e.g., 224x224 pixels) to ensure compatibility with the CNN architecture. This step also reduces computational overhead during training.
4) *Normalization*: Pixel values are normalized to a range of [0, 1] to stabilize the training process and accelerate convergence. This step ensures that all features contribute equally to the learning process.
5) *Data Augmentation*: Augmentation techniques, such as random rotation, horizontal flipping, brightness adjustment, and Gaussian noise addition, are applied to artificially expand the dataset. These techniques improve the model's ability to generalize by exposing it to a wider range of variations [10].

*D. **Dataset Splitting***

The dataset is split into three subsets:

1) *Training Set* (70%): Used to train the CNN model, this subset contains the majority of the data to ensure that the model learns diverse patterns.
2) *Validation Set* (15%): Used to tune hyperparameters and monitor performance during training, helping to detect and mitigate overfitting.
3) *Test Set* (15%): Reserved for final evaluation, this subset provides an unbiased measure of the model's performance on unseen data.

*E. **Ethical Considerations***

Ensuring fairness and reducing biases in facial recognition systems is a critical aspect of dataset selection and preparation. Datasets are carefully analyzed for demographic representation, and any observed biases are addressed through:

1) *Re-sampling*: Balancing the representation of underrepresented groups by oversampling or undersampling.
2) *Bias Mitigation Algorithms*: Techniques such as adversarial debiasing are applied to ensure fairness during training [15].
3) *Transparency*: A detailed analysis of the dataset's composition is conducted, and its limitations are explicitly acknowledged in the research.

By leveraging a carefully selected and preprocessed dataset, the research ensures a strong foundation for training a facial recognition model that is accurate, generalizable, and ethically sound.

*F. **Creating the Model***

To implement a robust facial recognition system, the Convolutional Neural Network (CNN) architecture was carefully designed to extract and classify facial features effectively. The process of creating the model involves several key steps, outlined below:

*1) **Defining the Model Architecture**:* The CNN model is constructed with a sequential architecture, leveraging layers that specialize in feature extraction and classification. The core components of the model include:

- *Convolutional Layers:* These layers extract spatial features by applying filters to the input image. Each convolutional layer is followed by a ReLU activation function to introduce non-linearity.
- *Pooling Layers:* Max pooling layers are incorporated to downsample the feature maps, reducing dimensionality and computational complexity while retaining the most salient features.
- *Fully Connected Layers:* These layers integrate the extracted features and output the final classification scores.
- *Dropout Layers:* To mitigate overfitting, dropout is applied during training, randomly setting a fraction of input units to zero.

*2) **Compiling the Model**:* The model is compiled with the following configurations:

- *Loss Function:* Categorical crossentropy is used as the loss function, appropriate for multi-class classification tasks.
- *Optimizer:* The Adam optimizer is chosen for its adaptability and efficient convergence.
- *Evaluation Metrics:* Accuracy is employed as the primary metric to evaluate model performance.

*3) **Training the Model**:* The training process involves:

- *Data Augmentation:* To enhance the diversity of the training dataset, techniques such as rotation, flipping, and scaling are applied.
- *Batch Processing:* The training data is processed in batches to optimize memory usage and speed up computation.
- *Validation:* A validation dataset is used to monitor the model's performance during training and prevent overfitting.

*4) **Evaluating the Model**:* Once training is complete, the model is evaluated on a separate test dataset to measure its generalization capability. Performance metrics such as accuracy, precision, recall, and F1 score are computed to provide a comprehensive assessment of the model's efficacy.

V. Model Evaluation

The evaluation of the Convolutional Neural Network (CNN)-based facial recognition system focuses on its performance in terms of accuracy, loss, and generalization ability. To assess the efficacy of the model, several evaluation metrics and techniques are employed.

A. Accuracy and Loss Analysis

The model's training and validation accuracy are critical indicators of its performance. Accuracy measures the proportion of correctly identified samples, providing insight into how well the CNN generalizes to unseen data. Training and validation losses are also monitored to identify overfitting or underfitting

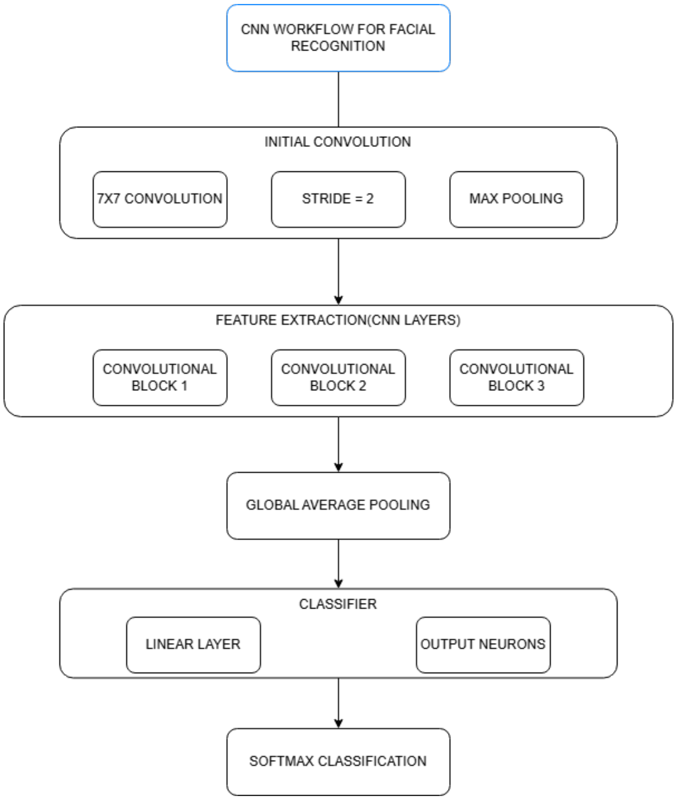

Fig. 4. CNN Architecture of the model

issues. The convergence of loss values during training indicates the model's stability and optimization effectiveness [1, 6].

B. Confusion Matrix

A confusion matrix is generated to analyze the model's classification performance in detail. This matrix provides a breakdown of true positives, true negatives, false positives, and false negatives. The metrics derived from the confusion matrix, such as precision, recall, and F1-score, offer a comprehensive understanding of the model's strengths and weaknesses in identifying faces [6, 17].

C. Cross-Validation

To evaluate the robustness of the model, cross-validation is applied. This involves partitioning the dataset into multiple folds, training the model on some folds, and testing it on the remaining ones. Cross-validation ensures that the model's performance is not biased by the specific train-test split and provides a more generalized evaluation [10, 13].

D. Comparison with Baseline Models

The proposed CNN model is compared with baseline facial recognition methods, including traditional machine learning techniques and simpler neural network architectures. Metrics such as accuracy, precision, and computational efficiency are analyzed to determine the improvements offered by the CNN-based approach [6, 15].

E. Generalization on External Datasets

To validate the model's generalization ability, it is tested on external datasets that were not used during training. This step highlights the robustness of the model in handling diverse real-world conditions, such as varying lighting, pose, and occlusion [1, 15].

F. Performance Metrics

Key performance metrics used for evaluation include:

- **Accuracy**: Percentage of correct predictions.
- **Precision**: Proportion of true positive predictions among all positive predictions.
- **Recall (Sensitivity)**: Proportion of true positive predictions among all actual positive cases.
- **F1-Score**: Harmonic mean of precision and recall, indicating the balance between them.
- **Area Under the Curve (AUC)**: Evaluates the model's ability to differentiate between classes.

This systematic evaluation framework provides a detailed understanding of the model's capabilities and areas for improvement. The findings from the evaluation guide future enhancements to the CNN architecture and training strategies.

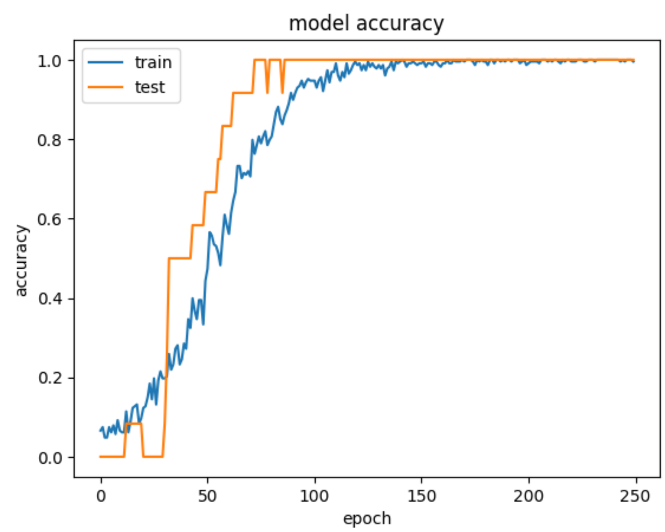

Fig. 5. Accuracy Plot for the model

After training and evaluating, we can plot the accuracy and loss plot for this model. Here, we see that the model performs well, even on the testing data, with strong results in both accuracy and loss metrics. This contrasts with the previous issue of misconfigured parameters.

The confusion matrix provided valuable insights into the model's performance across classes. Most classes achieved high precision and recall, but some misclassifications were observed, primarily between visually similar classes.

The model demonstrated high accuracy and generalization capability across unseen data. The use of augmentation techniques and an adaptive learning rate contributed significantly

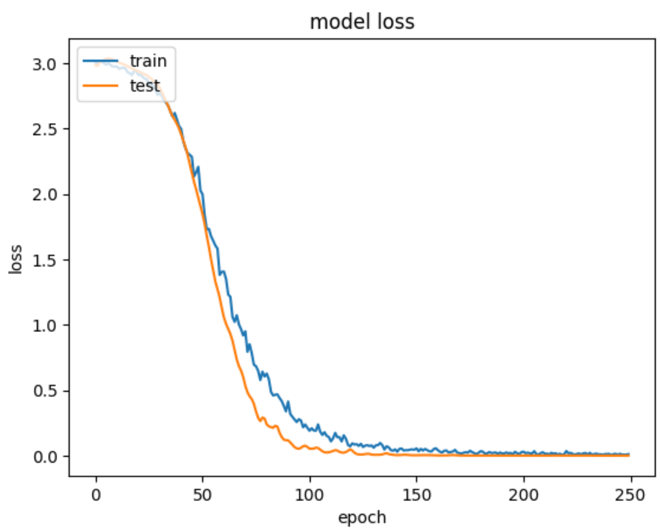

Fig. 6. Loss Plot for the model

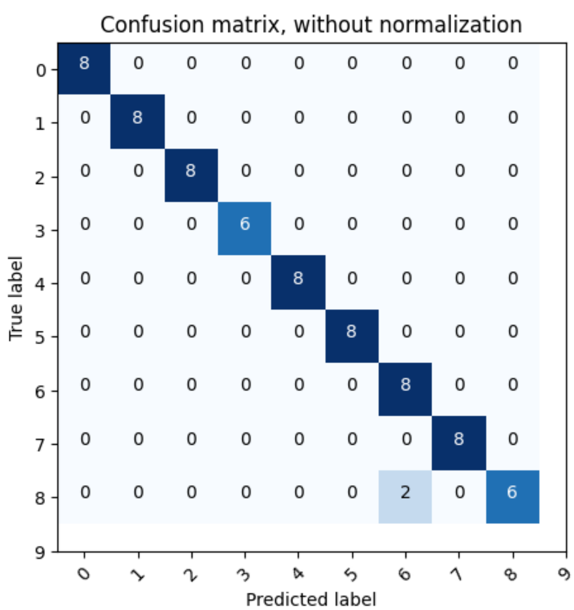

Fig. 7. Confusion Matrix for the Model (Without Normalization)

to its robustness. Future improvements could focus on addressing class imbalances and incorporating additional features, such as spectral information, for better discrimination between similar classes.

TABLE I
MODEL LAYERS AND FEATURES

Layer Name	Features
Convolutional Layer	Extracts spatial features using filters
ReLU Activation	Introduces non-linearity
Max Pooling Layer	Downsamples feature maps
Fully Connected Layer	Integrates extracted features for classification
Dropout Layer	Mitigates overfitting by random unit dropout

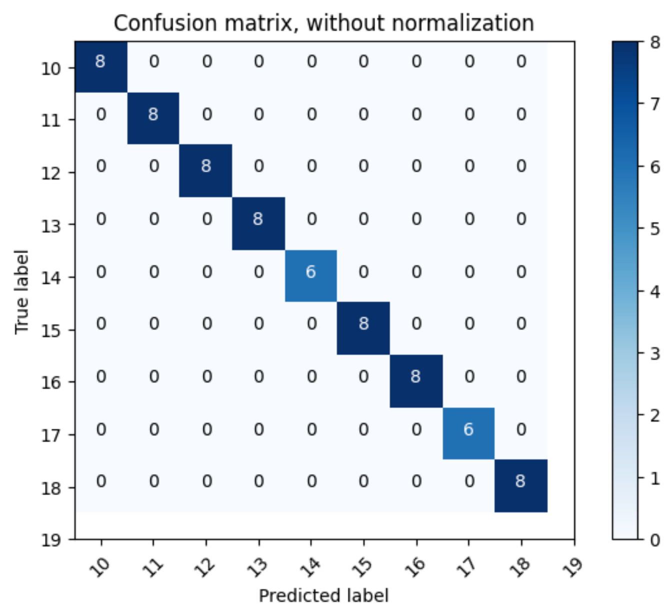

Fig. 8. Another Confusion Matrix for the Model (Without Normalization)

VI. FUTURE WORK

The field of facial recognition using Convolutional Neural Networks (CNNs) offers numerous opportunities for further exploration and improvement. This study identifies several areas for future research:

A. *Addressing Dataset Bias*

Despite advancements in datasets such as MS-Celeb-1M [15] and VGGFace2 [21], biases related to age, gender, and ethnicity persist. Future work could focus on curating more balanced datasets and developing fairness-aware training algorithms to reduce bias in facial recognition systems.

B. *Improving Generalization*

Addressing overfitting remains a priority for CNN-based models, especially when trained on limited datasets. Techniques such as data augmentation, transfer learning, and semi-supervised learning could be further explored to enhance model generalizability [10].

C. *Real-Time Implementation:*

The high computational cost of CNNs limits their deployment in real-time applications. Optimizing models using techniques like model pruning, quantization, or lightweight architectures such as MobileNet could make facial recognition systems more suitable for edge devices and real-time scenarios [17].

D. *Robustness Against Variations:*

Real-world applications demand robustness against factors such as occlusions, lighting conditions, and pose variations. Future research could explore advanced architectures, including attention mechanisms and multi-scale feature extraction, to enhance model robustness [6].

E. *Integration with Privacy-Preserving Techniques*:

Privacy concerns associated with facial recognition systems necessitate secure processing techniques. Homomorphic encryption and federated learning could be integrated to ensure data privacy while maintaining system performance.

F. *Cross-Domain Applications*:

Facial recognition systems can be expanded to emerging domains, such as emotion detection, healthcare monitoring, and augmented reality. Exploring these applications could unlock new possibilities for CNN-based solutions.

This future work aims to address the current limitations and expand the applicability of CNN-based facial recognition systems, ensuring they are ethical, efficient, and robust in real-world scenarios.

VII. Conclusion

Facial recognition has now become one of the primary applications of computer vision and is widely adopted in areas such as security, health care, and personalized services. The paper deals with the development and testing of a CNN-based facial recognition system. The study elaborates the systematic preprocessing approach, architectural design, and performance evaluation of the effectiveness of the CNN in identifying and verifying faces.

The results show that the systems based on CNN are highly robust.Specifically, their capacity for automatically extracting hierarchical features and the ability to generalize well beyond the training data. Some major limitations for achieving robust, scalable face recognition systems remain dataset bias, overfitting, and computational complexity.

The state of the art of CNN-based facial recognition is evaluated using standard metrics like accuracy, precision, and recall. This paper discusses the limitations and future research directions of this work, including addressing biases, advanced architectures, and interpretability.

In conclusion, CNNs are a great hope for facial recognition applications; however, achieving fairness, robustness, and ethical deployment remains a very important challenge. The future work on these concerns will lead to developing more reliable and equitable systems to be safely integrated into real-world applications.

Acknowledgment

The authors of this work are grateful to VIT Bhopal University for its financial help and other facilities without which this study could not have been completed. We are particularly grateful to the faculty and the staff of VIT Bhopal for continually providing support and encouragement during the research work. This work is named "CNNs and their use in Facial Recognition" and has been undertaken at the VIT Bhopal University located on Bhopal Indore Highway, Kothrikalan, Sehore, Madhya Pradesh India, 466114.

References

[1] Dataset, Taken from Kaggle.com

[2] Wang, K. (n.d.). An exploration of face recognition methods based on LBP algorithm and PCA analysis.

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

[4] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105.

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

[6] Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press.

[7] Parkhi, O. M., Vedaldi, A., Zisserman, A. (2015). Deep face recognition. *BMVC*, 2015.

[8] Schroff, F., Kalenichenko, D., Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815-823.

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

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

[11] Szegedy, C., et al. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9.

[12] Dong, C., Loy, C. C., He, K., Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295-307.

[13] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4700-4708.

[14] Masi, I., Wu, Y., Hassner, T., Natarajan, P. (2018). Deep face recognition: A survey. *2018 31st SIBGRAPI Conference on Graphics, Patterns, and Images (SIBGRAPI)*, 471-478.

[15] Zhao, W., Chellappa, R., Phillips, P. J., Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, 35(4), 399-458.

[16] Phillips, P. J., et al. (2005). Overview of the face recognition grand challenge. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 947-954.

[17] Guo, Y., et al. (2016). MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. *European Conference on Computer Vision (ECCV)*, 87-102.

[18] Wang, Z., et al. (2018). Deep learning for identifying faces on social networks. *Neurocomputing*, 277, 1126-1134.

[19] Liu, W., et al. (2017). SphereFace: Deep hypersphere embedding for face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6738-6746.

[20] Zhang, K., Zhang, Z., Li, Z., Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10), 1499-1503.

[21] Wen, Y., et al. (2016). A discriminative feature learning approach for deep face recognition. *European Conference on Computer Vision (ECCV)*, 499-515.

[22] Deng, J., et al. (2019). ArcFace: Additive angular margin loss for deep face recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4690-4699.

[23] Cao, Q., et al. (2018). VGGFace2: A dataset for recognizing faces across pose and age. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, 67-74.

[24] Kemelmacher-Shlizerman, I., et al. (2016). The MegaFace benchmark: 1 million faces for recognition at scale. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4873-4882.