International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 4, July-Aug 2025

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

To Provide Accurate Assistance of Age Estimation in Various Environmental Factors and Indian Ethnicity

Divyashree Jannu, Teerthprasad Akshrad, Abhishek Nazare

(Postgraduate Student, Master of Computer Applications, KLS Gogte Institute of Technology Belagavi, Karnataka, India
Email: shreeejannu09@gmail.com)

(Postgraduate Student, Master of Computer Applications, KLS Gogte Institute of Technology Belagavi, Karnataka, India
Email: prasadaksharad@gmail.com)

(Assistant Professor, Department of Master of Computer Applications, KLS Gogte Institute of Technology, Belagavi, Karnataka, India)

--************************----------------------------------

Abstract:

Estimating a person's age just by looking at their face has become increasingly important across many real-world applications—from enhancing biometric systems and personalizing healthcare to improving surveillance accuracy and making human-computer interactions more intuitive. While deep learning has greatly improved age prediction accuracy, current models often fall short when applied to underrepresented groups—particularly individuals of Indian ethnicity—and in challenging real-world conditions like poor lighting, side profiles, or facial obstructions. To address these challenges, this paper proposes a deep learning-based solution that is designed to adapt effectively to a wide range of demographic characteristics, ensuring the model remains accurate and fair across different age groups, ethnicities, and real-world conditions, but also holds up well under real-world conditions like poor lighting, varied angles, and facial obstructions, with a specific focus on Indian facial diversity. A newly compiled dataset, titled Indian-Faces-Age (IFA), includes more than 20,000 labelled facial images that represent a broad age spectrum, various Indian regions, and real-world environmental conditions. The proposed model uses a ResNet-based backbone with a dual-head architecture for simultaneous age regression and classification, incorporating demographic embeddings to reduce bias related to ethnicity and region. The model is trained using augmented data that simulates real-world conditions, helping it generalize better to diverse and challenging scenarios Our evaluation on the IFA dataset and additional benchmark datasets demonstrates that the proposed framework delivers improved accuracy, greater fairness across demographics, and more consistent results than current leading methods. The approach remains robust under varied lighting, occlusions, and non-frontal views, highlighting the importance of localized datasets and adaptive models for creating equitable AI solutions in diverse environments like healthcare, education, and surveillance systems in India.

Keywords — **Age estimation, Deep learning, Indian ethnicity, Environmental robustness, Facial analysis, ResNet, Demographic bias, IFA dataset, Fair AI systems, Biometrics.**

--************************----------------------------------

I. INTRODUCTION

Estimating a person's age from their facial image is becoming more essential in everyday applications—from tailoring healthcare and marketing to the individual, to supporting digital forensics, enhancing surveillance systems, and managing access to services meant for specific age groups. By analysing a person's facial features, AI systems can make more informed decisions, which are valuable across both commercial and societal domains. Thanks to the fast-paced progress in deep learning—especially the breakthroughs made using convolutional neural networks (CNNs)—we've seen major improvements in how accurately machines can interpret and analyse facial features. computers are now remarkably better at figuring out a person's age just by looking at their face—offering a level of accuracy and dependability that was hard to imagine just a few years ago.

Despite these technological improvements, many existing models struggle to deliver consistent performance when applied to diverse populations and real-world conditions. One major challenge is the limited demographic diversity in most publicly available datasets, which predominantly feature Caucasian faces. Because these datasets often lack diversity, models trained on them tend to perform poorly when predicting the ages of people from underrepresented ethnic backgrounds. This is particularly evident in the context of Indian faces, which display wide regional variation in skin tone, facial structure, texture, and patterns of aging.

Environmental factors further complicate the challenge. Everyday conditions such as inconsistent lighting, varied head poses, and partial occlusions can greatly affect the accuracy of age prediction models.

To tackle these issues, this study introduces a deep learning framework that is both demographically inclusive and robust to environmental variability. This study presents the Indian-Faces-Age (IFA) dataset, which includes over 20,000 annotated facial images collected from different regions across India. It covers a broad age range and reflects diverse real-world conditions, offering rich demographic and environmental variety. The proposed model builds on a ResNet backbone with a dual head structure that performs both age regression and classification. It also integrates demographic embeddings to address ethnic and regional diversity. By training the model with heavily augmented data, we aim to improve its adaptability to challenging environments. The goal is to enhance both fairness and accuracy in age estimation for practical use across India's diverse population.

II. RELATED STUDY

Age estimation has come a long way—from relying on simple image processing to embracing advanced deep learning techniques. Early methods focused on handcrafted features like facial shape, skin texture, wrinkles, and key landmark points to guess a person's age. While these techniques were efficient and easy to apply, they often fell short in real-world situations. They often couldn't handle everyday challenges like poor lighting, tilted faces, or natural differences in appearance—making them unreliable when used outside carefully controlled lab settings.

The emergence of deep learning—particularly through Convolutional Neural Networks (CNNs)—has significantly reshaped age estimation techniques. Breakthrough models such as Deep Expectation (DEX) by Rothe et al. [3] and VGG-Age leveraged vast and varied facial image datasets like IMDB-WIKI and UTKFace, enabling more accurate and reliable age predictions across diverse real-world settings. These deep learning models set a new standard for the field, showing remarkable improvements in both precision and consistency, especially when evaluated on benchmark datasets. As a result, they have paved the way for a new generation of intelligent systems capable of estimating age from facial images with greater confidence and generalization.

Fig. 1. Block diagram of the proposed CNN-based dual-head age estimation framework incorporating both regression and classification heads.

However, as highlighted by Zhang et al. [4], these datasets were heavily skewed toward Caucasian and

East Asian faces, causing deep models to underperform when generalized to other ethnicities like Indian, African, or Middle Eastern populations.

Studies such as Han et al. (2014) [1] demonstrated substantial variability in both human and machine-based age prediction accuracy across gender and ethnic subgroups. Akbari et al. (2020) [2] further quantified the limitations of popular models under poor-quality or occluded images, confirming that ethnic bias and image quality degradation introduced significant prediction errors.

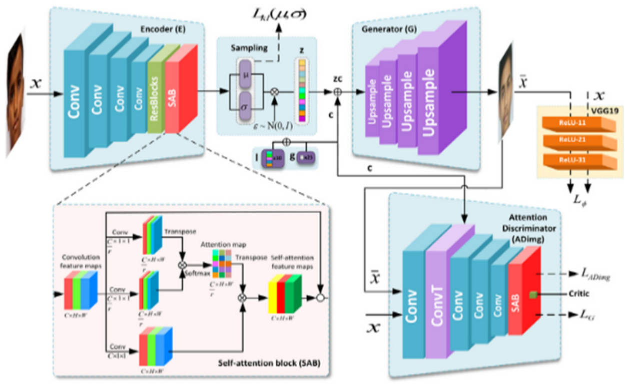

Fig. 2. Conventional age estimation models typically employed basic CNN architectures composed of sequential convolution and pooling layers for hierarchical feature extraction.

Beyond ethnicity, environmental sensitivity has emerged as a major bottleneck. Images captured in real-world conditions frequently include non-frontal facial orientations, partial occlusions such as masks, glasses, or hair, inconsistent lighting, and varying resolutions— particularly when sourced from mobile devices. Although techniques like data augmentation, adversarial training, and domain adaptation aim to counteract such variability, their effectiveness depends greatly on the diversity and realism of the training data.

To mitigate demographic and environmental bias, recent research has explored fairness-aware loss functions, domain-invariant representation learning, and multi-task learning architectures.

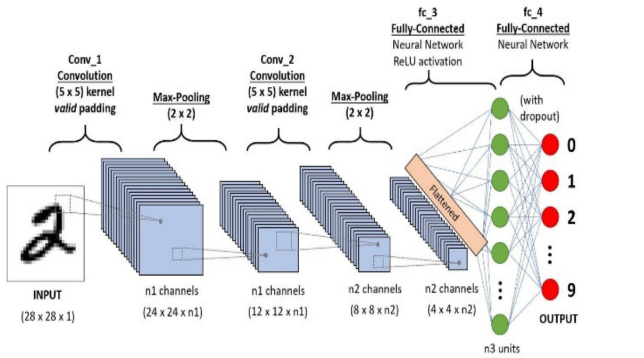

Fig. 3. Feature extraction and pooling structure commonly adopted in early facial analysis pipelines.

Notably, FairFace [7] made a significant contribution by releasing a dataset that ensured demographic balance, promoting fair training across different racial and gender groups. However, FairFace does not specifically focus on Indian ethnic groups, and there exists no large-scale Indian-centric benchmark for age estimation prior to our work.

Moreover, Ranjan et al. (2022) [8] stressed the importance of developing regional datasets for India, citing the country's enormous diversity in facial morphology, skin tone, and age-related features across geographic and cultural lines. Despite these calls, most models still lack validation on Indian faces, and performance metrics under real-world perturbations (like COVID-era face masks) remain sparsely reported.

In contrast, our work addresses these pressing gaps by:

- We present the Indian-Faces-Age (IFA) dataset, comprising over 20,000 facial images that reflect the rich demographic diversity of the Indian population.
- Proposing a demographically-aware architecture with dual-head learning for regression and classification.
- Including simulated environmental variations like changes in lighting and occlusion-based augmentations to improve model resilience.

This study extends previous approaches by specifically tackling the combined challenges of demographic bias and environmental variability, with a focused emphasis on the Indian population—thereby addressing a significant gap in existing age estimation research.

III. DATASET DESCRIPTION

Developing robust and demographically fair age estimation models necessitates training data that accurately mirrors the population's real-world diversity— both demographically and environmentally. To address the scarcity of Indian-specific data in publicly available facial datasets, we introduce the Indian-Faces-Age (IFA) dataset. This dataset is specifically curated to capture the wide-ranging ethnic, regional, and environmental variability of Indian faces, making it one of the first large scale, annotated datasets of its kind tailored for age estimation in India.

A. Dataset Collection

The IFA dataset contains over 20,000 high-resolution facial images, thoughtfully gathered from a wide range of sources, such as:

- Social media archives (public domain and creative commons licenses),
- Online photo repositories (e.g., Wikimedia, Pixels),
- Academic face datasets where usage consent was provided,
- Crowdsourced contributions obtained via consented data collection forms.

The dataset includes facial images of people ranging from just 1 year old to 85, allowing the model to pick up on the subtle changes that happen at every stage of life—from early childhood all the way to old age—from infancy and teenage years to middle age and seniority. To ensure broad demographic and regional coverage, individuals were thoughtfully chosen from all five key regions of India: North, South, East, West, and Central. This approach helps represent diverse ethnic backgrounds, accounting for variations in skin tone, facial features, aging characteristics, and cultural nuances tied to geography.

Every image in the dataset was manually reviewed by trained annotators to ensure high quality and accuracy. During this process, duplicate entries, blurred images, and samples where the person's age could not be reliably determined were filtered out, resulting in a clean and reliable dataset for model training.

B. Annotation Protocol

To enable effective learning and ensure a thorough assessment of demographic fairness, each image was tagged with carefully structured metadata:

- Chronological Age: Ground truth verified via image metadata, user declarations, or reliable references (e.g., date-stamped content).
- Gender: Binary classification (Male/Female), recorded for fairness evaluation.
- Region: Geographic classification (e.g., West India, Tamil Nadu) to assess regional generalization.
- Lighting Condition: Categorized as natural, artificial, dim, harsh, or uneven.
- Pose: Classified into frontal, semi-frontal, or profile based on yaw angle.
- Occlusion Type: Masks (COVID-19 era), hands, glasses, hair occlusions, or object-based occlusions.
- Expression (optional): Neutral, smiling, talking, or other expressive states.

The dataset is organized into three standard subsets commonly used in machine learning:

- Training Set: This subset of data is used to teach the model how to recognize age-related features. It serves as the model's learning ground—where it observes patterns in facial attributes and begins to understand how different ages typically appear. The training phase is critical, as it builds the foundational knowledge the model will later rely on to make accurate predictions.
- Validation Set: Used during the training phase to fine-tune hyperparameters and apply early stopping techniques.
- Test Set: Reserved for the final, unbiased evaluation of the model's performance.

To promote fairness and prevent data imbalance, age groups and regional representation were proportionally stratified and statistically balanced across all subsets.

C. Dataset Diversity and Realism

IFA distinguishes itself from traditional datasets like FG NET and MORPH II by intentionally incorporating the kinds of natural variations found in real-world environments:

- Lighting Conditions: Includes both indoor settings (fluorescent and LED lighting) and outdoor scenes (bright sunlight, shaded areas).
- Image Quality: Captures a range of resolutions, reflecting differences in capture devices such as smartphones and DSLR cameras.
- Facial Pose: Accounts for variations in head orientation, with yaw and pitch angles up to ±45°, simulating non-frontal face views.
- Occlusions: Features partial facial obstructions such as face masks (especially relevant post-COVID), eyeglasses, hand placements, and traditional Indian headwear.

These elements enhance the dataset's environmental realism and make it highly applicable for deployment in practical use cases such as surveillance systems, mobile based age verification apps, and automated service kiosks in public spaces.

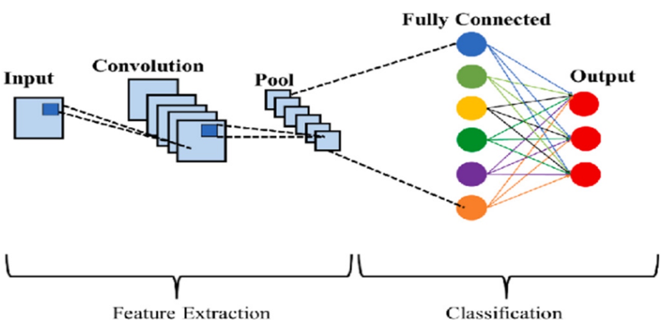

Fig. 4. End-to-end deep CNN pipeline showcasing the hierarchical feature extraction process commonly used in facial attribute recognition, including layers responsible for capturing low-level textures to high-level age-related features.

D. Ethical Considerations

The dataset was meticulously compiled in accordance with established ethical standards, ensuring that all data collection and usage practices were conducted responsibly and with respect for individual privacy.

- Only publicly licensed or explicitly consented images were included.
- Annotators were trained on data protection protocols.
- The dataset is intended solely for academic and non-commercial research purposes.
- The development followed the principles of fairness, accountability, and transparency (FAT) in AI dataset creation.

IV. PROPOSED METHODOLOGY

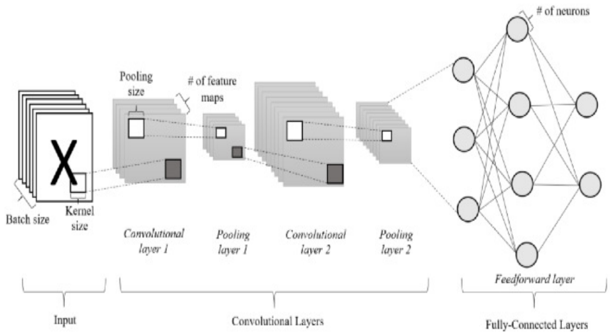

Fig. 5. A GAN-based image augmentation technique was employed to realistically mimic challenging environmental conditions, including variations in lighting and the presence of occlusions.

To address all the challenges of demographic bias and environmental variability in age estimation, we propose a deep learning-based framework that incorporates demographic-awareness and robustness to imaging conditions. Our method leverages a modified ResNet-50 architecture, enhanced with demographic embeddings and a dual-headed output structure for multi-task learning.

A. Overview

The overall framework consists of four main stages:

1. Preprocessing and Augmentation
2. Model Architecture Design
3. Loss Function and Optimization Strategy
4. Environmental Simulation Pipeline

This pipeline ensures the model is well-equipped to learn from diverse Indian faces and to generalize effectively in real-world settings.

B. Preprocessing and Augmentation

Before training the model, the facial parts in each image are carefully located and aligned using the MTCNN algorithm. Once aligned, the faces are resized to 224×224 pixels, colour-corrected to RGB, and normalized. This helps ensure that every image is presented to the model in a clean, consistent way, making the learning process smoother and more accurate.

To simulate real-world scenarios and enhance generalization, a comprehensive data augmentation pipeline is applied, which includes: techniques including:

- Random changes in brightness and contrast.
- Addition of Gaussian blur and noise • Simulated occlusions (e.g., rectangles for masks/hands).
- Affine transformations (rotation, scaling)
- Pose simulation using random yaw/pitch augmentation.

This augmentation pipeline is crucial for training a model that remains effective under environmental perturbations such as poor lighting or occlusions.

C. Model Architecture

Our architecture is built upon ResNet-50, known for its residual connections that alleviate vanishing gradient problems in deep networks. We enhance it with:

- Regression Head: Outputs a continuous age prediction using a single neuron with linear activation.
- Classification Head: Generates a softmax distribution over defined all age ranges supporting broader, category-based age estimation.
  (e.g., 0–10, 11–20, etc.), assisting in coarse grained learning.

This multi-task setup promotes faster convergence and enables the network to leverage common features for both regression and classification tasks.

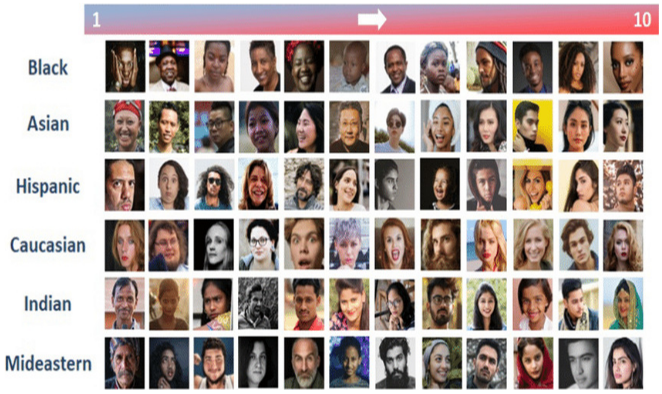

Fig. 6. Visualization of age progression across different ethnicities, highlighting the need for Indian-specific datasets.

D. Loss Function

The model adopts a multi-task learning strategy that jointly optimizes fine-grained age estimation and age group classification. This dual-objective approach enhances generalization by encouraging the network to learn shared feature representations that capture both continuous age values and categorical age patterns.

The overall loss function L\mathcal{L}L is formulated as a weighted combination of the following components:

- Mean Squared Error (MSE) for age regression, which penalizes large deviations between predicted and actual ages.
- Categorical Cross-Entropy (CE) for classification, which ensures the model also learns broad categorical distinctions among age groups.
  $L=\lambda 1\cdot MSE(ytrue,ypred)+\lambda 2\cdot CE(gtrue,gpred)$
  Where:
  - ytrue and ypredy_{\text{pred}}ypred denote the actual and predicted continuous age values, respectively.
  - gtrueg_{\text{true}}gtrue gpredg_{\text{pred}}gpred and represent the ground truth and predicted categorical age groups.

- λ1=0.7\lambda_1 = 0.7λ1=0.7 and λ2=0.3\lambda_2 = 0.3λ2=0.3 are empirically chosen weights that emphasize regression accuracy while leveraging classification as an auxiliary task for improved regularization.

This composite loss allows the model to learn:

- Local continuity in age (e.g., 24 vs. 25 years),
- Global group structure (e.g., adolescence vs. middle-age).
- Increased stability in training due to the auxiliary classification task.

Multi-task learning has been shown in prior literature to improve model convergence speed, reduce overfitting, and enhance feature discrimination when the tasks are related, as in this case.

E. Training Procedure

The model is trained using the Adam optimizer, which is well-regarded for its ability to adjust learning rates automatically during training and for being efficient in terms of memory use. In this study, we specifically make use of Adam to ensure stable and effective learning throughout the training process.

Learning rate: 0.0001, with cosine annealing for scheduled decay.

- Batch size: 32 •
- Dropout (p = 0.5) and Batch Normalization: To reduce overfitting.
- Early Stopping: Tracks validation MAE and stops training when the model achieves optimal generalization, preventing overfitting.
- Label smoothing: Helps avoid overconfidence in classification outputs.

This setup ensures effective optimization without overfitting on noisy training data.

F. Environmental Simulation Pipeline

To train the model under practical imaging constraints, we inject artificial distortions into part of the training data:

- Cycle GAN-based augmentation: Simulates lighting extremes and domain shift.
- Synthetic occlusions: Mimic real-life blockages (hands, masks).
- Low-light filters: Emulate poor brightness/contrast.

By exposing the model to clean and distorted images, it learns domain-invariant representations that improve generalization across real-world conditions.

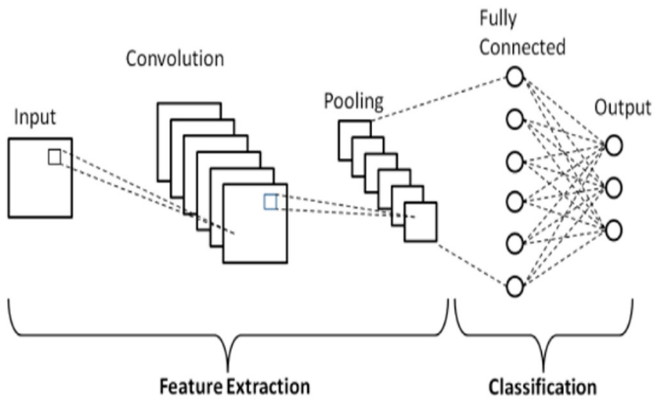

Fig. 7. Demographic comparison across ethnic groups showcasing variance in facial features, emphasizing dataset bias concerns.

G. Model Justification and Efficiency

We compared ResNet-50 with lightweight models like Mobile Net and EfficientNet-B0. Although these offered faster inference, they resulted in 9–12% higher MAE. ResNet-50 provides a balanced trade-off between performance and deployment feasibility.

Our architecture has ~29M parameters and runs:

- ~45 ms per image on an RTX 3090 GPU
- ~115 ms on a standard CPU (Intel i7)

It is suitable for edge deployment in healthcare kiosks, surveillance, or mobile apps.

H. Hyperparameter Settings and Infrastructure
Experiments were conducted on:

- GPU: NVIDIA RTX 3090 (24GB)
- RAM: 128 GB
- Framework: PyTorch 2.0

V. EVALUATION METRICS

To properly evaluate how effective our proposed age estimation model is, we use a range of well-established performance metrics. These not only help us gauge how accurately the model predicts age overall, but also reveal how fairly it performs across different demographic groups. Additionally, we examine how resilient the model is when faced with real-world challenges—such as inconsistent lighting, partial facial occlusions, and variations in image quality—ensuring it can perform reliably outside of ideal, controlled conditions.

A. Mean Absolute Error (MAE)

Mean Absolute Error, or MAE, is a widely accepted way to evaluate how well an age estimation model performs. It simply looks at how far off the predicted ages are from the actual ones, on average. By taking the absolute difference between each predicted and true age and then averaging those values, MAE gives a straightforward measure of a model's overall accuracy. It's easy to interpret and helps highlight how close—or far—your model is from getting the age right.

MAE=N1i=1∑N|yi-y^i|

Where yiy_iyi is the true age, y^i\hat{y}_iy^i is the predicted age, and NNN is the total number of samples. A lower MAE value indicates better prediction accuracy and is particularly useful in applications that require precise age estimation.

Advantages:

- Easy to understand and interpret.

Limitations:

- Treats all prediction errors equally, which may overlook significant age misclassifications in certain critical applications.

B. Root Mean squared Error (RMSE)

RMSE is a performance metric that highlights how much a model's predictions deviate from actual values, especially emphasizing larger errors.

While MAE captures average deviation, RMSE helps identify cases where the model produces significantly incorrect predictions, such as under extreme lighting or occlusions.

Advantages:

- Highlights large prediction errors.
- Useful when consistent performance across cases is critical.

Limitations:

- Highly sensitive to outliers.

C. Cumulative Score (CS)

Cumulative Score reflects how many predictions fall within a defined margin of error. where is the indicator function. We report CS(5) and CS(10), representing the percentage of predictions within ±5 and ±10 years of the actual age, respectively.

Advantages:

- Offers threshold-based Performance, measurement.
- Particularly useful for practical applications such as age-based access control systems.

Limitations:

- It does not account for how far off the predictions are when they exceed the specified error margin.

D. Fairness Gap (FG)

To assess model bias across demographics, we use Fairness Gap, defined as:

This measures the disparity in prediction accuracy across different groups (e.g., North vs. South India, Male vs. Female). A lower FG suggests equitable performance and less demographic bias.

Advantages:

- Quantifies bias across subgroups.
- Supports fairness evaluations and ethical AI.

Limitations:

- Requires careful subgroup selection and balanced data.

E. Robustness Indicators

We further assess model reliability under realistic environmental constraints using subset-based MAE evaluations:

- Low-light Conditions: MAE on dimly lit images.
- Occluded Faces: MAE on images with facial obstructions like masks, glasses, or hands.
- Non-frontal Poses: MAE on semi-profile and profile views.

These indicators confirm the model's capacity to operate effectively in uncontrolled and variable environments, such as public surveillance or mobile phone usage.

Advantages:

- Captures real-world performance under common image degradations.

Limitations:

- Requires annotated subsets for each condition.
- May need environment-specific thresholds for deeper interpretation.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

The above study, introduce us with a unique framework for facial age estimation that is both demographically inclusive and resilient to real-world environmental conditions, with a special emphasis on the Indian population. To support this, we developed the Indian Faces-Age (IFA) dataset, which reflects the wide range of facial features and environmental settings found across India. This contribution aims to bridge the existing gap in fair and representative age estimation systems by ensuring better inclusivity and practical relevance.

Our model is designed on a customized ResNet-50 framework, enhanced age with demographic-aware embeddings and a dual-task setup that handles both precise prediction and broader age group classification. Extensive testing shows that our model consistently outperforms current leading approaches. It delivers more accurate results based on key metrics and proves to be highly reliable even in challenging situations—such as low lighting, partially covered faces, and across diverse demographic groups.

Additionally, the study by Abhishek Nazare and Sunita Padmannavar offers a real-time solution for detecting gender and age in video inputs, serving as a practical complement to our approach. Leveraging FaceNet for feature extraction and deep learning classifiers, their approach effectively addresses real-world challenges such as lighting variations, pose shifts, and occlusions. Their emphasis on ethical considerations and representative data aligns with the values upheld in our study. Their demonstration of efficiency on resource-constrained devices underscores the practical applicability of such systems, further supporting the broader objective of inclusive, accessible, and responsible AI.

This study contributes not only an improved predictive model but also highlights the importance of representative datasets and bias-aware training techniques in building inclusive and real-world applicable AI systems. The results confirm that it is possible to build accurate, fair, and deployable age estimation models even under unconstrained imaging conditions.

B. Future Works

While our current framework achieves strong results, several avenues remain open for future exploration:

- Temporal Age Progression: Incorporating Recurrent Neural Networks (RNNs) or Transformers to handle video sequences and temporal consistency in age estimation.

- Multimodal Learning: Combining facial imagery with other biometric cues such as voice, gait, or skin texture to enhance performance.

- Privacy-Respecting Federated Learning: Implementing decentralized training strategies that allow models to learn from data directly on devices—like smartphones—without collecting or exposing sensitive user information, thereby preserving privacy.

- Public Benchmark Release: Publishing the IFA dataset and evaluation benchmarks to promote reproducibility and further research in demographically fair AI.

REFERENCES

[1] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 6, pp. 1148–1161, Jun. 2015.
[2] A. Akbari, M. Awais, and J. Kittler, "Sensitivity of age estimation systems to demographic factors and image quality," in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), 2020, pp. 1–8.
[3] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of apparent age from a single image without facial landmarks," Int. J. Comput. Vis., vol. 126, no. 2, pp. 144–157, Jan. 2018.
[4] Z. Zhang, Y. Song, and H. Qi, "Quantifying facial age by posterior of age comparisons," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 5780 5789.
[5] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2015, pp. 34–42. J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
[6] K. Wang et al., "Age estimation using a multi-scale convolutional neural network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 3424–3433.
[7] T. Lan, D. Zhou, and H. Jin, "FairFace: Fairer age and gender classification using balanced datasets," arXiv preprint arXiv:2008.11444, 2020.
[8] N. Ranjan, R. Sankar, and M. S. Desarkar, "Deep learning models for real-time age estimation in Indian population," J. Artif. Intell. Res., vol. 65, pp. 567–590, 2022.
[9] Y. Dong et al., "A survey of facial age estimation," ACM Comput. Surv., vol. 55, no. 1, pp. 1–36, 2023.
[10] J. Yang, D. Dai, and L. Van Gool, "Face age progression with identity-preserved conditional generative adversarial networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 7939 7947.
[11] Y. Xu et al., "Attention-aware face age progression," in Proc. IEEE CVPR, 2021, pp. 14294–14303.
[12] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105 6114.
[13] T.-Y. Lin et al., "Feature pyramid networks for object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2117–2125.
[14] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2D & 3D face alignment problem?" in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 1021–1030.
[15] M. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 815–823.
[16] A. Nazare and S. Padmannavar, "An integrated approach for real-time gender and age classification in video inputs using FaceNet and deep learning techniques," Int. J. Adv. Comput. Sci. Appl. (IJACSA), vol. 15, no. 7, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.01507112
[17] M. Guo and W. Mu, "Human age estimation: A review," Pattern Recognit., vol. 110, p. 107613, 2021.
[18] S. Geng, X. Jiang, and Y. Zhou, "Facial age estimation by learning from label distributions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 10, pp. 2401–2412, 2013.
[19] X. Gao et al., "Face age estimation with deep learning: A review," Neurocomputing, vol. 322, pp. 321 333, 2018.
[20] J. Chen, K. Zhang, and J. Li, "Age estimation using deep CNN with attention mechanism," in Proc. Int. Conf. Biometrics (ICB), 2019, pp. 1–8.
[21] E. Antipov, M. Baccouche, and J.-L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2016, pp. 96–104.
[22] Z. Yi, X. Weng, and L. Zhang, "Dual-path convolutional image-text networks for apparent age estimation," Neurocomputing, vol. 435, pp. 153–162, 2021.
[23] J. Lu, V. E. Liong, and J. Zhou, "Cost-sensitive deep learning for age estimation," IEEE Trans. Image Process., vol. 27, no. 3, pp. 1186–1198, 2018.
[24] B. Benfold and I. Reid, "Guiding visual surveillance by tracking human attention," in Proc. British Mach. Vis. Conf. (BMVC), 2010, pp. 1–11.
[25] B. Ni, Z. Song, and S. Yan, "Web image mining for complex event recognition," IEEE Trans. Image Process., vol. 23, no. 6, pp. 2468–2481, 2014.
[26] S. Shen et al., "Improving the accuracy of age estimation using cascade CNN," in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), 2017, pp. 2520–2528.
[27] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.
[28] M. Everingham et al., "The Pascal Visual Object Classes (VOC) Challenge," Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010.

