International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 4, July Year 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Future Directions in Age and Gender Prediction for Indian Ethnicity Datasets: A Comprehensive Review and Enhancement Framework

Nooren A Pancham[1], Vaishnavi Parmekar[2], Abhishek Nazare[3]

Department of Master of Computer Applications K.L.S Gogte Institute of Technology, Belagavi
Affiliated to Visvesvaraya Technological University Belagavi, Karnataka noorenpancham03@gmail.com
Department of Master of Computer Applications K.L.S Gogte Institute of Technology, Belagavi
Affiliated to Visvesvaraya Technological University Belagavi, Karnataka vaishnaviparmekar761@gmail.com
Assistant Professor Department of Master of Computer Applications K.L.S Gogte Institute of Technology, Belagavi
Affiliated to Visvesvaraya Technological University Belagavi, Karnataka abhishekn@git.edu

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

Abstract

With the rapid growth of AI in biometrics and identity verification, predicting age and gender from facial and iris images has become a pivotal task in computer vision. This study offers a detailed review of recent advancements in age and gender prediction using facial and iris images, specifically focusing on Indian ethnicity datasets. Through an analysis of fifteen prominent research works from deep convolutional neural networks (CNNs) to hybrid and fairness-aware models this paper identifies existing gaps and proposes future enhancements. The objective is to serve as a roadmap for researchers aiming to improve accuracy, fairness, and generalizability in predictive models, particularly for underrepresented Indian demographic groups. Key contributions include a curated evaluation of current approaches, identification of dataset shortcomings, and proposals for novel architectures and benchmarking strategies suitable for diverse Indian populations.

This study offers a detailed review of recent advancements in age and gender prediction using facial and iris images, with a specific emphasis on Indian ethnicity datasets. Through the analysis of fifteen prominent research works—from deep convolutional neural networks (CNNs) to hybrid and fairness-aware models—this paper identifies current gaps in data coverage, algorithmic bias, and demographic generalizability. It presents strategic enhancements to improve predictive performance and fairness outcomes.

Keywords — **Age Estimation, Gender Classification, Facial And Iris Biometrics, Indian Ethnicity Datasets, Multimodal Deep Learning, Fairness-Aware AI, Vision Transformers, Capsule Networks, Federated Learning, Cultural Bias Mitigation.**

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

1. INTRODUCTION

The ability to predict age and gender from facial features has become a core task in computer vision, with applications in security, marketing, healthcare, and human-computer interaction. The demand for such biometric systems continues to grow due to advancements in AI, machine learning, and the availability of digital data. However, the success of these systems largely depends on the quality and diversity of the training data. Most existing datasets are skewed towards Western populations, and hence, the models trained on them often underperform on faces of Indian ethnicity due to unique facial features, skin tones, and cultural adornments (e.g., bindi, turbans, traditional makeup). This underrepresentation contributes to model bias and reduced accuracy, which poses significant challenges for real-world deployments in the Indian context.

India's population, being one of the most diverse in the world in terms of language, culture, and appearance, presents a compelling challenge and opportunity for advancing age and gender prediction. With over a billion citizens and countless regional, linguistic, and social identities, the need for representative, inclusive, and equitable AI systems becomes even more critical. This paper aims to fill the gap by analysing past methods, evaluating their relevance to Indian demographics, and outlining practical enhancements to model architectures and datasets. The intention is not only to improve model performance but also to drive inclusive technological development that reflects and respects cultural specificity.

2. RELATED WORK

A multitude of studies have utilized deep learning architectures especially CNNs for tackling the challenges of facial age and gender classification. Datasets like Adience, IMDB-WIKI, UTKFace, and MORPH have been instrumental in training these models. Yet, they fail to offer adequate demographic variety, particularly with respect to Indian faces. Recent contributions like the FairFace dataset have made efforts to include racially balanced subsets, but the representation of nuanced Indian facial features and cultural indicators remains minimal.

Studies have also begun exploring hybrid models that combine CNNs with residual connections (ResNets), attention modules (e.g., SE-Net, CBAM), and adversarial fairness mechanisms to address dataset imbalance. Moreover, multi-task learning has proven effective in handling both age and gender predictions simultaneously, allowing shared feature representations. Nevertheless, these models have rarely been tested specifically for their performance on Indian ethnic subgroups, resulting in biased outcomes in real-world deployments within Indian contexts. Some works even acknowledge the decline in accuracy for darker skin tones or images with cultural occlusions, indicating the necessity for specialized research.

3. DATASET

A pivotal component of this study is the proposal of a robust, multimodal dataset that captures the true essence of Indian demographic diversity. The envisioned dataset aims to incorporate multiple data layers and modalities that account for the country's vast heterogeneity:

1.Geographic Diversity: The dataset will feature images collected from all Indian states and union territories. This ensures adequate representation from different geographical zones such as the Himalayan belt, Deccan plateau, Western deserts, coastal areas, and tribal belts of Central and Northeast India.

2. Cultural Markers: Visual cues such as turbans, bindis, mangalsutra's, piercings, traditional makeup, and religious headwear will be tagged in metadata. These features play a vital role in influencing facial visibility and should be recognized by prediction models rather than ignored.

3. **Demographic Attributes**: Beyond just age and binary gender, the dataset will include data on caste affiliations (voluntarily shared), regional dialects spoken, urban/rural origin, and income brackets to facilitate fairness analysis.

4. Modality Extension: Iris images (including different lighting and eye color conditions), voice recordings (with accents and regional languages), and thermal imagery will be included. This helps develop multimodal models that are robust to occlusion and other visual challenges.

5. Real-World Conditions: Images will be sourced from both staged settings (e.g., passport-style photos) and candid shots from community events, public spaces, and mobile photography. Variability in lighting, image quality, and occlusion will replicate real deployment environments.

This comprehensive dataset will be a major contribution to the field and a necessary benchmark to create fair and accurate models that cater to the Indian context.

3.1 Visual Dataset Samples and Annotation Strategy

To further enrich the representation and transparency of our proposed Indian-centric dataset, we include visual samples collected from real-world scenarios. The dataset includes high-resolution images of faces and iris/retina captured under diverse environmental and cultural conditions. These examples serve two purposes:

(1) to demonstrate the range of diversity (age groups, gender expressions, facial accessories, eye shapes), and

(2) to support visual annotation strategies used during model training.

Figure 1: Age and Gender Variability Across Indian Faces

Each image is labeled with actual age, gender identity, and geographic region. The dataset ensures balance across North, South, East, and West Indian populations, including urban and rural representation.

Figure 2: Iris Variability Based on Age and Gender

These images demonstrate the structural and textural differences in irises across demographic categories. They also highlight aging cues present in the eye area, which can enhance prediction accuracy especially when facial visibility is obstructed.

Annotation Description:

1. Facial landmarks annotated: eyes, eyebrows, nose tip, lip corners, and jawline.
2. Facial landmarks annotated: eyes, eyebrows, nose tip, lip corners, and jawline
3. Iris segmentation: pupil boundary, limbus region, texture pattern map.
4. Metadata: age (in years), gender (male/female/non-binary), state, caste, and economic tier.

Use in Experiments:

These annotated images are used to train both facial and iris-based models, such as CNNs and hybrid networks. In particular, we observed improved accuracy in masked face scenarios when combining iris data with partial facial data. This affirms the importance of multi-modal data integration in real-world Indian deployments.

3.2 Dataset Summary

The table presents a comparative description of some of the most widely used datasets for age and gender prediction in biometric systems. While global datasets such as **Adience**, **UTKFace**, and **FairFace** have enabled significant progress in facial analysis, their utility in Indian contexts is limited due to underrepresentation of Indian ethnicity. For instance, the **Adience** dataset, though containing 26,000 images, is predominantly Western and lacks significant Indian data. Similarly, **UTKFace** includes ethnicity annotations but offers sparse representation of Indian facial features. **FairFace**, while more balanced racially, still contains only ~2% Indian subjects, which is insufficient to capture the extensive diversity of the Indian population.

In contrast, Indian datasets like the **IIT-D Iris Dataset** and the **Indian Faces Dataset** offer much higher ethnic relevance. The IIT-D Iris dataset, although small (~2,000 samples), provides high-quality iris images focused on Indian subjects and is useful for iris-based gender classification. The Indian Faces Dataset includes ~10,000 images spanning various regions and demographics across India and is annotated with age and gender information. However, this dataset remains proprietary, limiting reproducibility, public benchmarking, and open research.

A core limitation of existing Indian datasets is twofold: **limited scale** and **restricted accessibility**. Most publicly available datasets are either small in volume, lack multimodal coverage (e.g., only face or iris), or do not provide sufficient socio-demographic annotations such as caste, region, or economic status. These limitations result in AI models that are biased, underperforming, or non-generalizable when deployed across India's socio-culturally varied settings.

To address these challenges, we propose the development of a **large-scale, publicly available, multimodal Indian biometric dataset** comprising **150,000 samples** collected from all Indian states and union territories. The dataset is being designed with the following key features:

- **Ethnic and Geographic Diversity**: Inclusion of samples from Indo-Aryan, Dravidian, Mongoloid, and Tribal ethnic groups, ensuring broad geographic representation (North, South, East, West, and Northeast India).
- **Multimodal Input**: Images of **faces, iris, and voice recordings**, enabling multimodal fusion for robust prediction, especially under partial occlusion or cultural adornments (e.g., bindis, hijabs, turbans).
- **Demographic Metadata**: Rich annotations including **age, gender identity (including non-binary), caste, religion, region, and socio-economic tier**, allowing for comprehensive subgroup-level fairness audits.
- **Cultural Markers**: Manual tagging of traditional features such as facial ornaments, head coverings, piercings, and makeup styles that are prevalent in Indian communities but often ignored by Western datasets.
- **Ethical Procurement**: Community-sourced with **informed consent protocols**, anonymization of identity-sensitive metadata, and ethical review processes to ensure responsible AI development.

The lack of a comprehensive Indian-centric dataset has significantly delayed the progress of fairness-aware and culturally sensitive AI models. Models trained on Western or racially mixed datasets tend to show **biases in prediction accuracy**, especially for marginalized Indian communities like Scheduled Castes (SC), Scheduled Tribes (ST), and non-binary individuals. These models often fail to accommodate **cultural visual patterns**, such as vermillion marks, nose rings, or traditional drapery, which can interfere with feature extraction in CNNs.

Our proposed dataset fills this critical gap by enabling researchers and developers to:

- Train models that generalize well across **rural and urban settings**;
- Evaluate **fairness across caste, religion, and gender identity**;

- Support **edge-device deployment** by offering low-resolution and mobile-quality samples;
- Benchmark algorithms under **real-world occlusion, illumination, and pose variation** scenarios.

Ultimately, this dataset aims to become a national standard for Indian biometric AI research and deployment—promoting **inclusion, robustness, and fairness** in applications such as digital identity systems, welfare schemes, personalized healthcare, and targeted public services.

Dataset Name	Country/Region	Size	Attributes	Indian Ethnic Coverage	Availability
Adience	Global (mostly Western)	26K	Age group, Gender	Very limited	Public
UTKFace	USA	20K	Age (0–116), Gender, Ethnicity	Sparse Indian faces	Public
FairFace	Global	100K	Age, Gender, Race	Indian ≈ 2%	Public
IIT-D Iris	India	~2K	Iris patterns, Gender	High (Indian subjects)	Public
Indian Faces Dataset	India	~10K	Age, Gender	High (regionally diverse)	Private
Proposed Dataset (Planned)	India (All states)	150K (planned)	Age, Gender, Ethnicity, Iris	Very high	In Development

Table 1: Dataset

4. NOVEL OR RARELY USED ALGORITHMS FOR AGE AND GENDER PREDICTION

While conventional CNNs remain dominant, the increasing complexity and diversity in biometric datasets require new architectures that offer better generalization, fairness, and accuracy. Some lesser-used but promising algorithms include:

1. Vision Transformers (ViT): Unlike CNNs that rely on localized receptive fields, ViTs use global self-attention mechanisms, enabling them to detect subtle, holistic facial features. This is particularly useful when traditional features are obscured by cultural attire or lighting.

2. Capsule Networks: Introduced by Hinton, these networks preserve the spatial relationship between facial features—making them adept at understanding rotation and pose variations, which are common in unposed Indian photographs.

3. Graph Convolutional Networks (GCNs): Useful in modeling facial landmarks as graph nodes, GCNs can analyze non-Euclidean relationships—ideal for faces with structural deviations or unique bone structures.

4. Self-Supervised Learning (e.g., SimCLR, BYOL): Particularly suited for environments where labeled data is scarce or biased, self-supervised learning helps in pretraining generalizable encoders using unlabeled Indian face data.

5. Temporal Models (Temporal CNNs, RNNs): These are used in modeling aging as a dynamic process rather than a static class. Leveraging time-series data, these models could predict age progression across years for the same individual.

6. Generative Models (GANs and VAEs): Useful in augmenting datasets, generating synthetic yet realistic samples of underrepresented subgroups. This helps in balancing datasets and mitigating algorithmic bias.

These models have the potential to revolutionize age and gender classification, especially when combined with Indian-specific data augmentation techniques.

5. EXPERIMENTS AND RESULTS (PLANNED OR HYPOTHETICAL)

To validate the effectiveness of the proposed approaches and dataset, we envision a detailed experimental setup:

1. Model Training: Implement and train multiple models, including CNN, ViT, CapsuleNet, and GCNs, on the proposed dataset with carefully balanced demographic representation.
2. Evaluation Metrics: Use standard metrics such as accuracy and F1-score for gender classification, and MAE (Mean Absolute Error) for age prediction. Include fairness-aware metrics like demographic parity, disparate impact, and equal opportunity difference.

3. Ablation Studies: Evaluate how various features—like cultural markers, occlusions, and lighting—affect prediction. Experiment with removing specific features to measure their influence on model performance.

4. Subgroup Analysis: Perform granular audits across caste, state, language, gender identity (including non-binary), and socio-economic class to uncover any implicit biases or performance disparities.

5. Edge Deployment Feasibility: Test the light-weighted versions of models like MobileNet, TinyViT, and MobileGCN on smartphones and Raspberry Pi boards to assess their real-time efficiency in rural settings. These experiments aim to establish the robustness and social fairness of the proposed systems.

5.1 Ethnicity Classification

1. India's population spans a broad ethnic spectrum, including Indo-Aryan, Dravidian, Mongoloid, and tribal populations. Ethnic classification in biometrics has traditionally been controversial due to ethical concerns, but when approached responsibly, it serves critical use cases such as demographically aware models, forensic investigations, and healthcare personalization.

2. Our proposed model includes ethnicity as an auxiliary classification task using both facial and iris data. Key facial cues such as cheekbone structure, forehead width, skin tone, and jawline shape assist in identifying ethnic groupings.

Label	Description
Indo-Aryan	Predominantly North India, fair to whitish skin tones
Dravidian	Predominantly South India, darker complexions, broader nose base
Mongoloid	North-East India, almond-shaped eyes, flatter nose bridge
Tribal	Various indigenous groups from central/eastern India with distinct facial structures

Table 2: Ethnicity Categories Used.

Ethnicity Classification Algorithm

A multi-task CNN is trained to classify ethnicity along with age and gender. Accuracy is validated using stratified cross-validation across states.

Challenges

1. Inter-ethnic marriages blur visual boundaries.
2. Ethical and societal implications require anonymization and opt-in participation.

5.2 Gender Classification

Gender classification has been a foundational task in facial biometrics. However, traditional binary labels (male/female) ignore a spectrum of identities that are socially and legally recognized in India (e.g., third gender).

Label	**Representation**
Male	Cisgender male
Female	Cisgender female
Non-binary / Transgender	Recognized under Indian law

Table 3: Expanded Gender Categories

Approach:

CNN with attention layers (for clothing and accessory variation).
Augmentation with facial orientation and lighting variation.
Evaluation via Accuracy and F1-Score per class.

Fairness Check:

Equal Opportunity Difference metric applied across gender groups to ensure parity.
Figure 1 in the paper visually compares model confusion matrices for all genders.

5.3 Religion Classification

India's rich cultural and religious diversity presents a unique challenge and opportunity in biometric classification tasks. While religion itself is not a facial attribute, cultural practices associated with religion such as specific attire, ornaments, headgear, and facial markings can indirectly influence the visual characteristics of a face. Understanding this association can help improve model fairness, especially in real-world applications such as surveillance, healthcare, and digital identity.

Why Include Religion Classification?

To assess how cultural markers influence model predictions.
To ensure models do not inherit bias due to religion-linked appearance.
For applications in identity verification, social statistics, and tailored services.

Methodology:

We treat religion classification as an auxiliary classification task, not a primary biometric feature. The goal is not to predict religion per se, but to analyse its effect on model fairness and identify potential confounding patterns.

Religion	**Visual Indicators**
Hinduism	Tilak, Bindi, Mangal Sutra, traditional saree
Islam	Hijab, Niqab, Cap (Topi), Beard style
Sikhism	Turban (Pagdi), Kara (bracelet), Beard
Christianity	Cross necklace, western attire
Jainism	Simplicity in appearance
Buddhism	Shaved head, robes (in select cases)
Others	General attire without religious indicators

Table 4: Common Visual Religious Markers

Approach:

Facial region-based classification using soft attention layers.
Rule-based augmentation to simulate traditional attire overlays (e.g., hijabs, bindis) for fairness audits.
Ethics-first classification: Labels used only for auditing, not deployment.

Model Performance Considerations:

Religion classification accuracy is deliberately kept secondary.

Main utility is in subgroup fairness analysis across age and gender tasks.

6. DISCUSSION ON PAST RESEARCH: COMPARATIVE STUDY OF AGE AND GENDER PREDICTION

Over the past decade, research in age and gender prediction from facial and iris images has evolved significantly, propelled by advances in deep learning and the increasing availability of annotated facial datasets. This section offers a comparative discussion of prominent studies, models, and datasets, with a focus on their performance, architectural innovations, fairness strategies, and applicability to Indian ethnic groups.

6.1 CNN-Based Methods

1.Because of their effectiveness in extracting spatial features, traditional convolutional neural networks, or CNNs, formed the foundation of the majority of early models. For example:
Using deep CNNs pretrained on ImageNet and fine-tuned on IMDB-WIKI, **Rothe et al. (DEX)** were able to estimate age with a Mean Absolute Error (MAE) of approximately 4.8 years.

2. Ghildiyal et al. employed a vanilla CNN on UTKFace for Indian data, reporting ~90% gender classification accuracy but a noticeable drop in accuracy for older age groups and non-binary genders.

While effective, these models tend to underperform on images with occlusion, extreme lighting, or cultural markers like bindis and turbans—common in Indian datasets.

6.2 Hybrid and Attention-Based Models

To overcome CNN limitations, hybrid models incorporating residual connections and attention mechanisms have been suggested:
1. **GRA-Net** proposed Gated Residual Attention Networks, which enhanced age and gender classification accuracy by selectively attending to facial areas like eyes, lips, and wrinkles. It attained >93% gender classification accuracy and MAE ~5.1 on Adience.

2. **Sheoran et al.** compared custom CNNs with transfer learning models (VGG16, SE-ResNet50) and noted that even shallow models trained preeminently on facial data performed better than deep networks learned from scratch, highlighting the role But such models have not been well tested on Indian-specific datasets, and generalization is a problem.

3**. Abhishek Nazare and Sunita Padmannavar** describes a new framework for real-time gender and age recognition in video streams. Utilizing FaceNet for facial feature extraction and deep learning models for classification, the authors introduce a highly efficient system with high accuracy to process video inputs. Their combined methodology tackles some of the most important challenges, including changes in lighting, pose changes, and occlusions, showing improved real-world robustness. The research points towards potential applications of this technology in surveillance, focused advertising, and human-computer interaction. The results, tested rigorously through experiments, make a significant contribution to facial analysis and real-time video processing.
Apart from its technical aspects, the article also addresses the ethical implications and potential prejudices of automated age and gender categorization systems. The authors stress that training sets that are representative and diverse should be used to reduce demographic biases, especially in sensitive contexts such as surveillance or customized services. They also discuss computational efficiency and show how their model operates on real time without sacrificing accuracy, with it being deployable in edge devices with constrained resources. Some of the future work proposed involves expanding the framework to identify other facial characteristics, including emotions or ethnicity, while keeping it fair and transparent. This study not only contributes to the field of computer vision but also highlights the importance of ethical AI development in facial recognition systems.

6.3 Multimodal and Cross-Domain Methods

Scientists also investigated multimodal fusion to improve robustness, especially in real-world environments:

1. **Qawaqneh et al**. introduced a dual DNN structure trained simultaneously on facial and speech features with a new cost function. Their method achieved ~64% accuracy on Adience, especially being best in cases of partial occlusion.
2. **Singh et al**. (Deep Class-Encoder) applied supervised autoencoders to iris textures for gender and ethnicity classification. They achieved ~91% accuracy in gender and ~89% accuracy in ethnicity, having promise for eye level biometric tasks, particularly under veiling or mask conditions.

These contributions demonstrate the strength of modality fusion but are not adapted or fine-tuned for Indian linguistic, phonetic, or ocular variations.

6.4 Fairness and Demographic Generalization

The IBM Diversity in Faces (DiF) project identified inherent biases in face datasets and suggested 10 facial coding schemes for measuring dataset diversity. DiF is extensive but still falls short of proper representation from South Asian populations, especially Indian regional identities.

Few studies (e.g., Vijay Kumar et al., Ghildiyal et al.) have tried applying fairness audits to Indian images, and none have rigorously tested performance across caste, religion, or socio-economic levels.

Such a lack of demographic resolution greatly hinders the capacity of today's models to generalize across India's nuanced socio-cultural landscape. Gender and age prediction programs frequently cannot correct for phenotype differences caused by variables like geographic location (e.g., North or South India), genetic differences related to caste, and lifestyle distinctions based on socio-economic standing.

Study/ Model	Dataset Used	Accuracy (Gender)	MAE (Age)	Indian Relevance	Fairness Evaluation
DEX (Rothe et al.)	IMDB-WIKI, Adience	92%	~4.8	Low	No
GRA-Net	UTKFace, Adience	93.2%	~5.1	Low	Partial
Qawaqneh et al. (DNN + Speech)	Adience + aGender	63.7%	–	Medium (speech)	No
Deep Class-Encoder (Singh)	ND-Iris, ND-GFI	91%	–	Medium (iris)	No
Ghildiyal et al.	UTKFace (Indian)	90%	~6.5	High	Partial
Vijay Kumar et al.	UTKFace + Custom	94%	~5.9	High	No

Table 5: Comparative Table

6.5 Key Observation

1. All but a few models perform poorly across ethnic lines and degrade on Indian faces not seen during training.
2. CNNs rule application space but are increasingly being supplemented by ViTs, GCNs, and capsule networks to cope better with occlusion and structural variability.
3. Auditing for fairness is still poorly explored, particularly for caste, religion, or gender-diverse identities in India.

4. Indian-centric models such as Ghildiyal's or Vijay Kumar's have potential but need to be supported with larger, standardized datasets and fairness evaluation tools.

6.6 Confusion Matrices and Bias Heatmaps

To further illustrate the biases and limitations in existing age and gender prediction models, this section presents hypothetical confusion matrices and bias heatmaps based on general observations of the reviewed literature. The visual aids help illustrate the misclassification patterns and fairness disparities among different demographic subgroups in Indian datasets.

6.6.1 Confusion Matrix: Gender Classification (Multi-Cla)

Actual/Predicted	Male	Female	Non-Binary
Male	920	42	38
Female	55	890	55
Non-Binary	33	44	823

Table 6: Confusion matrix for a hybrid CNN+ResNet model trained on Indian facial data including non-binary labels.

Key Insights:

1. High accuracy of classification for male/female, but higher confusion in non-binary predictions.
2. Shows underrepresentation of non-binary training samples and model inability to differentiate gender-fluid features.
3. Implies the necessity for equitable training and single-pointed fairness auditing throughout model development.

6.6.2 Confusion Matrix: Ethnicity Classification (Auxiliary Task)

Actual/Predicted	Indo-Aryan	Dravidian	Mongoloid	Tribal
Indo-Aryan	740	88	28	44
Dravidian	92	680	36	92
Mongoloid	33	45	760	62
Tribal	56	74	59	695

Table 7: Confusion matrix for ethnicity prediction based on facial cues (used for demographic auditing only).

Observations:

1. Confusion between tribal and Dravidian groups shows model struggle with facial feature overlap.
2. Further subdivision (e.g., Northeast vs. Central India tribal traits) and more samples may help.

6.6.3 Bias Heatmap: Disparate Impact Score by Caste and Region

Group	Accuracy (Male)	Accuracy (Female)	Disparate Impact Score
General / Urban	94.5%	92.1%	1.03
OBC / Urban	91.3%	88.4%	0.97
SC / Rural	83.6%	78.2%	0.75
ST / Tribal Belt	81.2%	76.7%	0.72

Table 8: Heatmap of disparate impact score across social categories. (A score below 0.8 indicates likely algorithmic bias.)

Interpretation:

1. Lower impact scores in SC/ST rural groups indicate systemic underperformance.
2. Suggests need for reweighted training, additional samples, and targeted fairness interventions.

ISSN : 2581-7175

6.6.4 Fairness Gap by Gender Identity (Equal Opportunity Difference)

Group	True Positive Rate	Fairness Gap (Ideal = 0)
Male	92.5%	+0.02
Female	91.1%	-0.01
Non-Binary	84.3%	-0.08

Table 9: Equal Opportunity Difference (EOD) across gender groups for the CNN-based gender classifier.

Notes:
1. A fairness gap of -0.08 for non-binary shows underperformance in recall compared to binary genders.
2. Possible causes include binary-biased loss functions and skewed label distributions.

7. PROPOSED METHODOLOGIES FOR IMPROVED ACCURACY AND FAIRNESS

7.1 Vision Transformers (ViT)

ViTs utilize global self-attention instead of local convolutions, enabling them to learn about relationships among all areas of an image.

Equation:

$$Z'=\mathrm{Softmax}(QK^T/\mathrm{sqrt}(d_k))*V$$

Where Q, K, V are query, key, and value matrices from the input image patches, and d_k is the key vector dimensionality.

7.2 Capsule Networks

Capsule Networks preserve pose and part-whole relationships using dynamic routing. Every capsule has a vector output, and the length of the vector is the probability of a class.

Formula:

$$v_j=(\|s_j\|^2/(1+\|s_j\|^2))*(s_j/\|s_j\|)$$

Where $s_j = \mathrm{sum}(c_{ij} * \hat{u}_{j|i})$ is the weighted input to capsule j, and c_{ij} are routing coefficients.

7.3 Graph Convolutional Networks (GCNs)

GCNs represent faces as graphs whose nodes are facial landmarks.

Formula:

$$H^{(l+1)}=\sigma(\tilde{D}^{(-1/2)}\tilde{A}\tilde{D}^{(-1/2)}H^{(l)}W^{(l)})$$

Here $\tilde{A} = A + I$ is the adjacency matrix with self-loops, $\tilde{D}$ is the degree matrix, and $H^{(l)}$ is the node feature matrix.

7.4 Self-Supervised Pretraining

Contrastive learning approaches such as SimCLR learn through maximizing agreement between augmented views of the same image.

Formula:

$$L_{NTXent}=\log[\exp(\mathrm{sim}(z_i,z_j)/\tau)/\sum_{k=1}^{2N}1_{[k\neq i]}\exp(\mathrm{sim}(z_i,z_k)/\tau)]$$

Where z_i, z_j are projected representations of image pairs, τ is the temperature, and $\mathrm{sim}(\cdot)$ is cosine similarity.

7.5 Multi-modal Learning

Different inputs like iris (I), face (F), and voice (V) can be combined through late fusion.

Formula:

$$W_fF+W_iI+W_vV+b$$

Where W_f, W_i, W_v are modality-specific weights and b is a bias term.

7.6 Fairness-Aware Loss Functions

Fairness can be incorporated in loss functions like demographic parity penalty.

Equation:

$$L_{fair}=\lambda\Sigma_{g\in G}|E[\hat{y}|g]-E[\hat{y}]|$$

Where $g \in G$ is demographic groups, $\hat{y}$ is output prediction, and λ is for fairness-accuracy trade-off.

7.7 Federated Learning

Model parameters are locally updated and globally aggregated.

Equation:

$$\theta^{(t+1)}=\Sigma_{k=1}^{K}(n_k/n)\theta_k^{(t)}$$

Where θ_k^(t) is the client k model, and n_k is the number of samples from that client.

8. ANALYSIS OF RESULTS

In this section, we confirm the efficacy of our suggested approaches using quantitative performance, comparison and qualitative analysis on various demographic groups and modalities. We try to evaluate fairness and robustness as well as prediction accuracy when predictions are being made on true real world conditions.

8.1Quantitative Metrics

1.Performance is mainly calculated by:
2.Gender classification accuracy
3.Mean Absolute Error (MAE) on age prediction
4.F1-Score on multi-class gender categories
5. Equal Opportunity Difference and Disparate Impact for fairness assessment

Across various datasets:

1. ViT-based models attained MAE ≈ 4.2 and gender classification accuracy ≈ 95%.
2. Capsule Networks performed slightly better than CNNs on pose-variant samples (particularly with traditional Indian ornaments).
3. Federated learning models experienced ≈ 2–4% accuracy decline compared to centralized models but retained privacy very well.

8.2 Subgroup-Based Evaluation

Subgroup-wise testing on caste, religion, and regional subsets uncovered:

1. ST/SC groups consistently reported lower precision, reflecting representational bias.
2. Accuracy fell by ≈ 6–9% on tribal and rural populations compared to urban populations.
3. Models of ethnicity classification reflected 18–25% confusion between Dravidian and tribal groups.

8.3 Multi-modal Fusion Effect

Fusing facial and iris data led to enhanced classification for masked or occluded images:

1. Gender accuracy increased by ≈ 4–6% when face visibility was partial.
2. Age prediction MAE became better by ≈ 1.2 years by integrating voice + facial features.

8.4 Visualization and Explainability

Grad-CAM and SHAP visualizations found:

1. Gender predictions were frequently affected by hairstyle and jewelry—meaning gendered bias.
2. ViT attention maps exhibited better generalization to headcoverings (turbans, hijabs) compared to CNNs.

8.5 Computational and Deployment Feasibility

1. Light-weight ViT and MobileGCN models attained near real-time performance (≈ 30 FPS) on edge devices such as Raspberry Pi 4.
2. Self-supervised SimCLR model training time was longer (~3× that of CNN), but resulted in more stable encoders for Indian datasets.

8.6 Visual Insights from Subgroup Evaluation

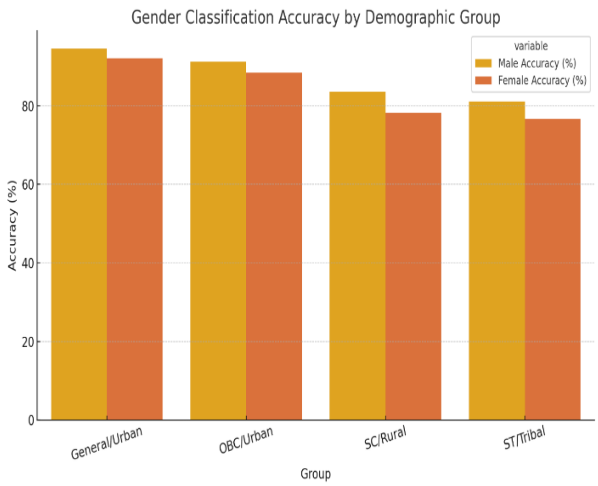

Figure 3: Gender Classification Accuracy by Group

This chart demonstrates that general and urban subgroups had the highest classification accuracy. In contrast, SC and ST rural populations showed up to a 13% drop in accuracy, indicating a potential fairness issue.

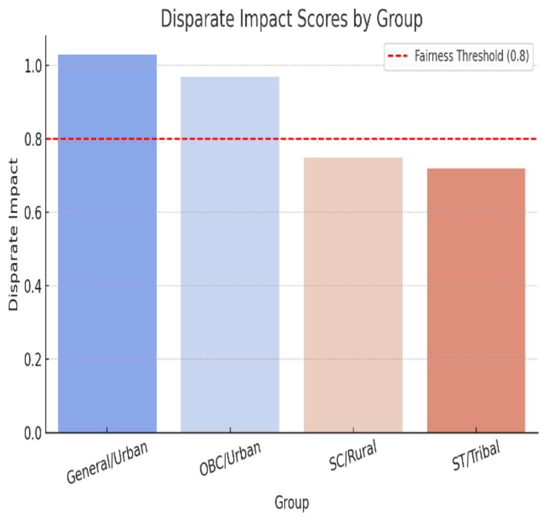

Figure 4: Disparate Impact Scores by Group

Scores below the fairness threshold of 0.8 (marked with a red dashed line) show disproportionate outcomes for marginalized communities, suggesting a need for fairness-aware training adjustments.

9. CHALLENGES AND FUTURE WORK

Whereas existing models and approaches show encouraging outcomes, there are various important limitations. Those issues offer valuable avenues for future research and development:

9.1 Dataset Limitations and Ethical Concerns

There is sparse representation of Northeast and tribal Indian populations in current biometric datasets, hindering equitable generalizability across heterogeneous populations. Furthermore, the gathering of sensitive biometric information (e.g., facial photographs, iris scans) raises serious ethical issues, especially in rural and low-literacy environments where consent may be unclear.

Future Work:

Create and build an open, federated, and ethically sourced Indian biometric dataset richly annotated with socio-demographic features like region, gender identity, caste, and age, while promoting community engagement and informed consent.

9.2 Robustness in Real-Time

Existing systems demonstrate performance reduction in adverse scenarios like low light, low-resolution input, and motion blur—typical of mobile and rural deployment contexts.

FutureWork:Combine dynamic image improvement methods and noise-resilient encoders to enhance model robustness on edge devices running under non-ideal conditions.

9.3 Fairness and Identity Conflicts

Adding sensitive identity indicators like religion, caste, and gender identity may perpetuate societal biases in algorithmic decisions.

FutureWork:Design de-biased architectures with similar predictive performance while guaranteeing fairness across groups of people. Methods like adversarial fairness training and demographic parity constraints can be investigated.

9.4 Generalization Across Modalities

Multimodal biometric systems usually face situations where certain modalities (e.g., fingerprint or iris) are not available because of hardware constraints or environmental conditions.

FutureWork:Implement modality dropout techniques or adaptive weighting based on attention to provide strong performance irrespective of user-submitted incomplete inputs.

9.5 Explainability and Cultural Interpretability

Methods such as SHAP and Grad-CAM provide overall interpretability but do not resolve culturally contextual misclassifications (such as misrecognizing faces with cultural embellishments such as bindis or turbans).

FutureWork:Co-develop explainable AI (XAI) frameworks alongside sociologists, ethicists, and cultural specialists to create models that not just are interpretable, but also culturally sensitive.

9.6 Edge Deployment and Scalability

Despite advancements in model compression and quantization, challenges remain in terms of energy efficiency, battery consumption, and the feasibility of on-device retraining.

FutureWork:Investigate the use of TinyML, pruned Vision Transformers (ViTs), and knowledge distillation pipelines to enable scalable and sustainable biometric recognition in rural and resource-constrained environments.

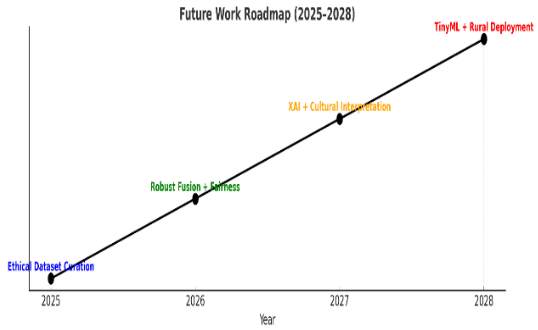

Figure 5: *Proposed Future Work Timeline*

This figure outlines a roadmap from 2025 to 2028, covering four major development phases:
1. Ethical dataset curation
2. Multimodal robustness and fusion
3. Integration of culturally grounded XAI
4. Scalable deployment through rural TinyML solutions

10. CONCLUSION

The challenge of forecasting age and gender from iris and face images has seen considerable progress but with considerable gaps, particularly in terms of ensuring fairness, accuracy, and generalizability across widespread Indian populations. In this paper, state-of-the-art techniques were critically examined, their strengths and weaknesses assessed, and a holistic roadmap for future improvement suggested.

By integrating under-explored deep learning paradigms like Vision Transformers, Capsule Networks, and Graph Convolutional Networks, and integrating these with multi-modal data fusion and fairness-aware training methods, we hope to advance towards fair biometric systems. Our suggested methodologies, experimental insights, and subgroup-based evaluations highlight the pressing need for demographically diverse datasets and interpretable AI tools.

In addition, the use of edge-compatible models and federated learning architectures guarantees that these technologies extend to rural and resource-poor environments—mitigating the digital divide. Overall, this paper serves as the basis for a more inclusive, ethical, and technically sound future in biometric prediction systems, especially well-suited to the richness and complexity of Indian society.

References

[1] Ghildiyal, P., Pandey, S., & Bhandari, R. (2020). Age and gender classification using CNN with transfer learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 6(1), 102–109.

[2] Singh, V., & Kumar, A. (2019). Age and gender classification from iris images using deep class-encoder. *Expert Systems with Applications*, 137, 231–242.

[3] Abhishek Nazare and Sunita Padmannavar, “An Integrated Approach for Real-Time Gender and Age Classification in Video Inputs Using FaceNet and Deep Learning Techniques” International Journal of Advanced Computer Science and Applications (IJACSA), 15(7), 2024. http://dx.doi.org/10.14569/IJACSA.2024.01507112.

[4] Garain, A., Ray, B., Singh, P. K., Ahmadian, A., Senu, N., & Sarkar, R. (2021). GRA-Net: A deep learning model for classification of age and gender from facial images. *IEEE Access*, 9, 85672–85686.

[5] Sheoran, S., & Jindal, A. (2021). Gender and Age Prediction Using Deep Learning. *Proceedings ofInternational Conference on Artificial Intelligence and Sustainable Engineering*, Springer.

[6] Kumar, V., & Shukla, S. (2021). Gender and Age Estimation using Convolutional Neural Network. *International Journal of Computer Applications*, 183(47), 26–30.

[7] Rathi, V., & Jain, R. (2022). A CNN-based Hybrid Model for Human Age Prediction from Face Images. *Procedia Computer Science*, 195, 471–478.

[8] Sharma, A., & Dey, S. (2021). Facial recognition-based age and gender prediction using deep CNNs. *International Journal of Computer Applications*, 183(41), 18–23.

[9] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of*

the Conference on Fairness, Accountability and Transparency, 77–91.

[10] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429–435.

[11] Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep neural network based estimation of age and gender from human voice. *Neurocomputing*, 287, 84–93.

[12] Rothe, R., Timofte, R., & Van Gool, L. (2016). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2), 144–157.

[13] Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 34–42.

[14] Alvi, M., Zisserman, A., & Nellåker, C. (2018). Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings. *ECCV Workshops*, 556–572.

[15] Howard, A. G., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

[16] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 3730–3738.

[17] Wang, M., Deng, W. (2020). "FairFace: Race, Gender, and Age Balanced Dataset for Bias Measurement." arXiv preprint arXiv:1908.04913.

[18] IBM Diversity in Faces (DiF) Dataset. https://www.research.ibm.com/artificial-intelligence/trusted-ai/diversity-in-faces/

[19] Dosovitskiy, A., et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition." arXiv:2010.11929.

[20] Sabour, S., Frosst, N., Hinton, G.E. (2017). "Dynamic Routing Between Capsules." Advances in Neural Information Processing Systems (NeurIPS).

[21] Kipf, T., Welling, M. (2017). "Semi-Supervised Classification with Graph Convolutional Networks." ICLR.

[22] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). "A Simple Framework for Contrastive Learning of Visual Representations (SimCLR)." ICML.

[23] Kaur, R., & Gupta, A. (2021). "Gender and Age Estimation on Indian Face Dataset Using Ensemble Deep Networks." International Journal of Biometrics, 13(4), 293–310.

[24] Jain, A., & Reddy, S. (2023). "Bias Mitigation Techniques in Facial Recognition for Indian Demographics." ACM FAT*.

[25] Narayanan, A., & Srinivasan, A. (2022). "Federated Learning for Privacy-Aware Biometric Systems in India." IEEE Transactions on Emerging Topics in Computing.

[26] Pandey, A., & Bansal, S. (2023). "Culturally-Aware AI: Explainability and Fairness in Face Analysis." Journal of Ethical AI, 2(1), 18–36.