A Comprehensive Review of Deep and Hybrid Machine Learning Algorithms for Age and Gender Prediction

[1]K Sri Krishna Vamshi[1*], Neha Saravanan[1], Shreya Nagendra[1], Prashant P Patavardhan[1]

[1]Department of Electronics and Communication Engineering, RV Institute of Technology and Management (RVITM), Bengaluru, Karnataka, India, 560076.

*Corresponding author(s). E-mail(s): krishnavamshikunda16@gmail.com; Contributing authors: saravanan.neha3@gmail.com, shreyan1509@gmail.com, prashantgemini73@gmail.com

***Abstract*—Computer vision based age and gender prediction has become essential in fields such as security, healthcare and social media analytics. This review examines key studies (2016–2024) that employ deep and hybrid machine learning models using facial and ocular data. The explored architectures include Convolutional Neural Networks (CNNs), Residual Attention Networks (GRA_Net), CAFFE-modified MobileNet V2 (CMNV2), Deep Class-Encoders, Deep Iris CNNs, hybrid CNN–SVM and IoT-based systems. Experimental results across major datasets—Adience, FG-NET, UTKFace, AFAD, CASIA-WebFace and ND-GFI—show gender classification accuracies ranging from 91% to 99.8%, with Mean Absolute Error (MAE) for age estimation as low as 2.39 years. Eye-focused CNNs achieved up to 98.8% accuracy, highlighting the potential of ocular modalities. The paper synthesizes advances in network design, preprocessing and hybrid integration while addressing persistent challenges, including dataset bias, model interpretability and real-world deployment feasibility. Recent review studies highlight hybrid learning challenges under occlusion and device variability.**

Index Terms—Age estimation · Gender prediction · Deep learning · Hybrid machine learning · Convolutional Neural Network (CNN) · Residual networks · Attention mechanism · Internet of Things (IoT) · Biometrics

1 Introduction

Detecting someone's age and gender using computer vision has become a key tool in areas like surveillance, security, healthcare and targeted ads. This technology tries to estimate age and gender from pictures of faces or eyes even when the lighting is bad or the camera angle isn't perfect. Older techniques such as Local Binary Patterns (LBP), Principal Component Analysis (PCA) and Support Vector Machines (SVM) had some success but often failed when dealing with odd angles, poor lighting or when parts of the face were covered [16], [17]. Deep learning has revolutionized how we identify people, enabling us to extract key details directly from images. Models like Convolutional Neural Networks (CNNs), Residual Attention Networks (RANs) and Gated Residual Attention Networks (GRA_Net) outperform traditional methods, achieving gender recognition accuracy of more than 97% on large datasets such as UTKFace and Adience [3], [11]. Besides looking at faces, methods that use the eyes and irises can be a good alternative that respects privacy. Deep CNNs trained on iris datasets, such as ND-GFI, CASIA and IIT-A, have achieved more than 94% accuracy, demonstrating that eyes can be used to identify individuals [13], [15], [16]. This paper combines the results of important studies, looking at both the numbers and the ways they got there. It puts importance on hybrid models that mix deep designs with older machine learning methods or IoT systems to make things more efficient and able to be used in different situations.

2 Methodological Advancements

2.1 Adaptive Gamma-Corrected CNN (AGC-CNN)

Saha et al. [3] created a CNN model that uses Adaptive Gamma Correction (AGC) to fix the problem of bad lighting in face pictures. It works by first fixing the pictures (making them grayscale, lining up landmarks and correcting the gamma) and then using a six-layer CNN to classify them. When tested on the Adience dataset (17,000 pictures), the model got gender right 97.1% of the time and lowered the Mean Absolute Error (MAE) to 4.1 years. AGC made low-light pictures clearer, letting the system extract wrinkles and lines.

Figure 1 illustrates the overall workflow of the AGC-CNN model with adaptive illumination normalization.

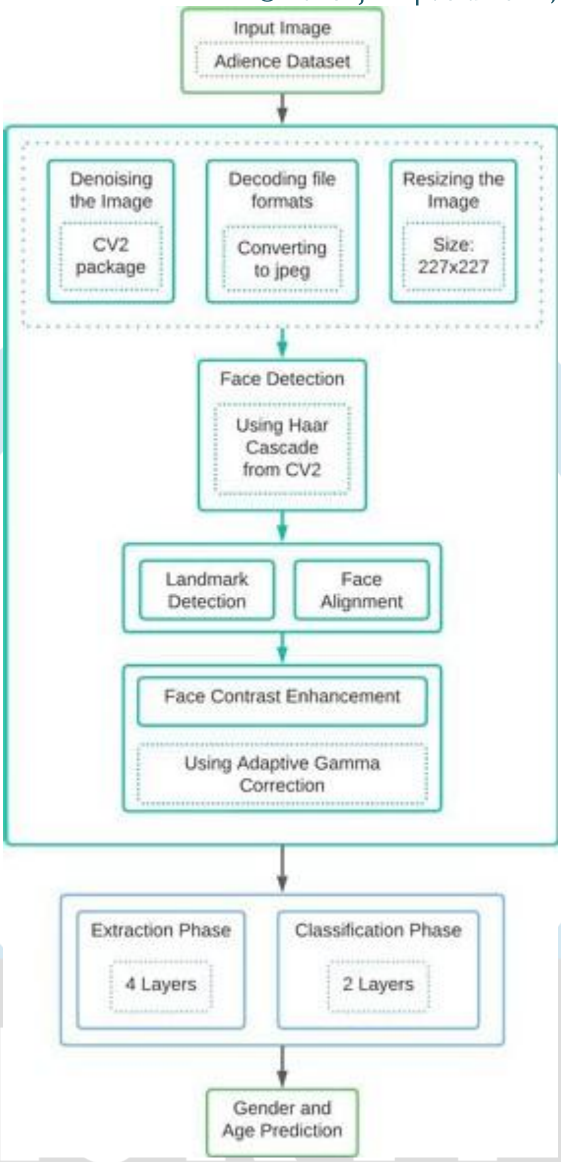

Fig. 1 Workflow schematic of AGC-CNN with adaptive illumination normalization [3].

2.2 Gated Residual Attention Network (GRA_Net)

Garain et al. [11] made GRA_Net, which adds residual parts with channel and spatial attention modules. The model uses both classification and regression to guess both gender and age. Experiments on five datasets: FG-NET, UTKFace, AFAD, AdienceDB and Wikipedia—gave MAE results of 2.39 (FG-NET), 2.78 (AFAD), 3.26 (UTKFace) and 6.32 (Adience). Gender classification was correct 99.2% of the time on UTKFace. The model was stable and didn't get thrown off by overfitting because of the attention gating.

The architecture of GRA_Net is presented in Figure 2, highlighting the integration of residual blocks with spatial and channel attention modules.

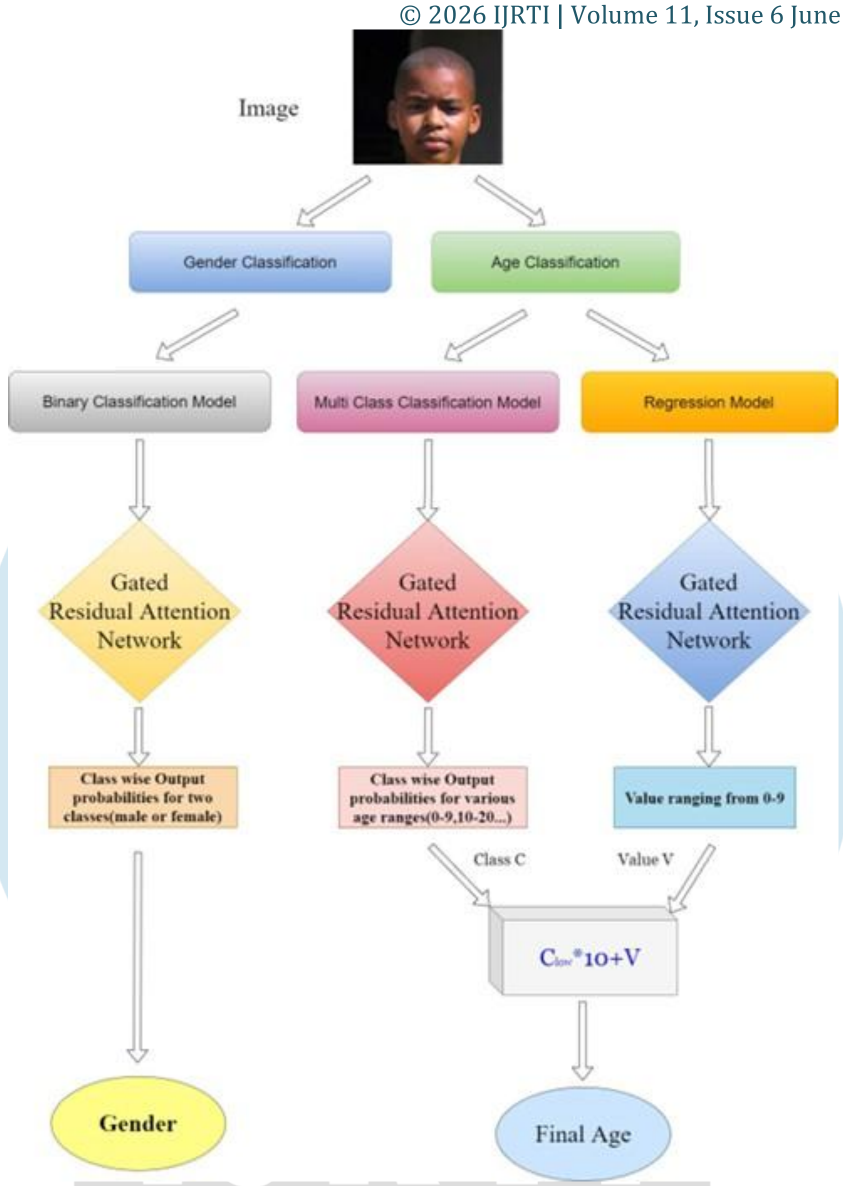

Fig. 2 Block schematic of GRA_Net integrating residual and attention layers [11].

2.3 Simple CNN-Based Group

Antipov et al. [17] showed a simple CNN group that combines three small CNN models, made for use in small devices. By training on CASIA-WebFace (490,000 pictures) and testing on LFW (13,000 pictures), the group got 97.31% accuracy with only 1.5 MB of parameters—eleven times quicker and sixteen times more memory-saving than normal VGG designs.

Figure 3 depicts the Minimalistic CNN ensemble used for lightweight gender classification in embedded systems.

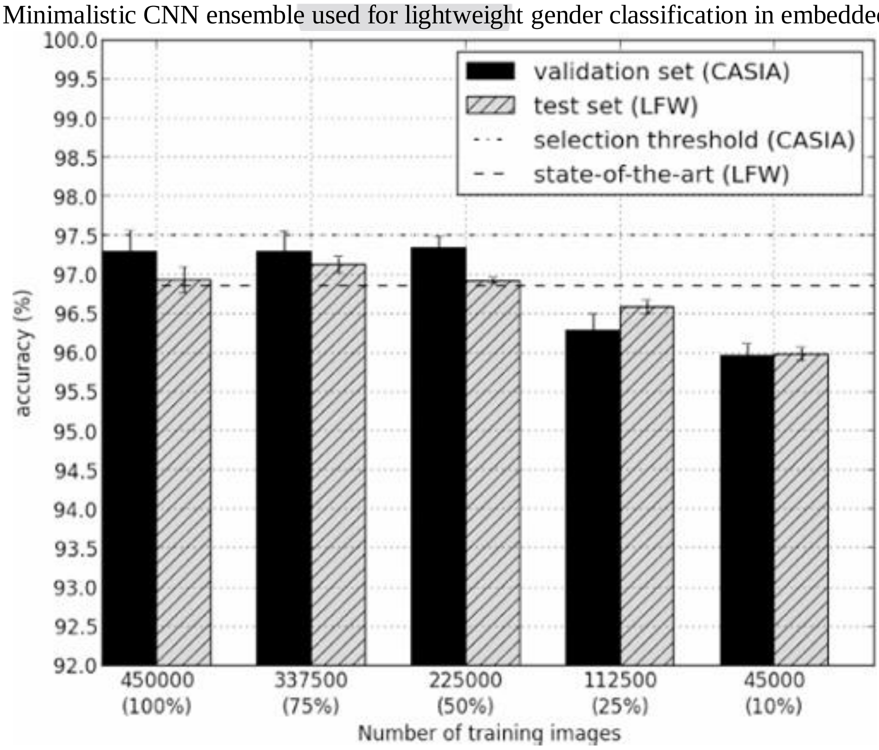

Fig. 3 Minimalistic CNN ensemble workflow illustrating optimized multi-CNN fusion [17].

2.4 Masked Face Age and Gender Identification (CMNV2)

Pasagadugula et al. [1] suggested Convolutional Architecture for Fast Feature Embedding (CAFFE)-modified MobileNetV2 (CMNV2) for recognizing masked faces after COVID-19. Using 12,160 pictures with fake masks, the network added eight new layers on top of the MobileNetV2 base and compared how it did against nine designs, like DenseNet and EfficientNet. CMNV2 achieved 96.54% accuracy with a 3.46% error rate and could process in real time (40 Frames Per Second – FPS).

The CMNV2 pipeline for masked face gender and age prediction is shown in Figure 4, demonstrating the added layers over MobileNetV2 for mask-invariant recognition.

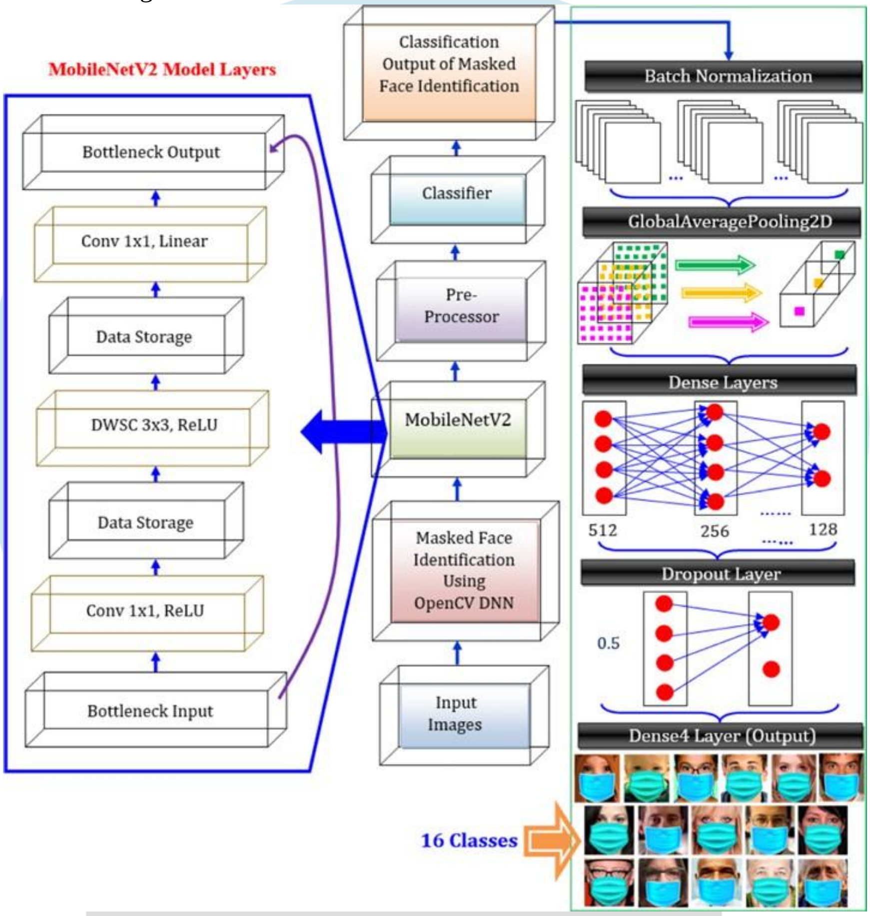

Fig. 4 CMNV2 structure emphasizing transfer-learning and mask-invariant recognition [1].

2.5 Intelligent Hybrid Learning Principle (IHLP) with IoT

Vinayagam and team presented Intelligent Hybrid Learning Principle (IHLP) [4], a hybrid CNN setup that uses Internet of Things (IoT) to manage records. The system takes live camera images and cleans them up by reducing Gaussian noise and normalizing them before pulling out key features. IHLP was able to classify gender with 94.8% accuracy, meaning it works well in live, IoT-based storage setups. While exact comparison numbers are missing, the design can be scaled up for use in smart spaces.

Figure 5 shows the Intelligent Hybrid Learning Principle (IHLP) framework that combines CNN-based feature extraction with IoT-enabled data management.

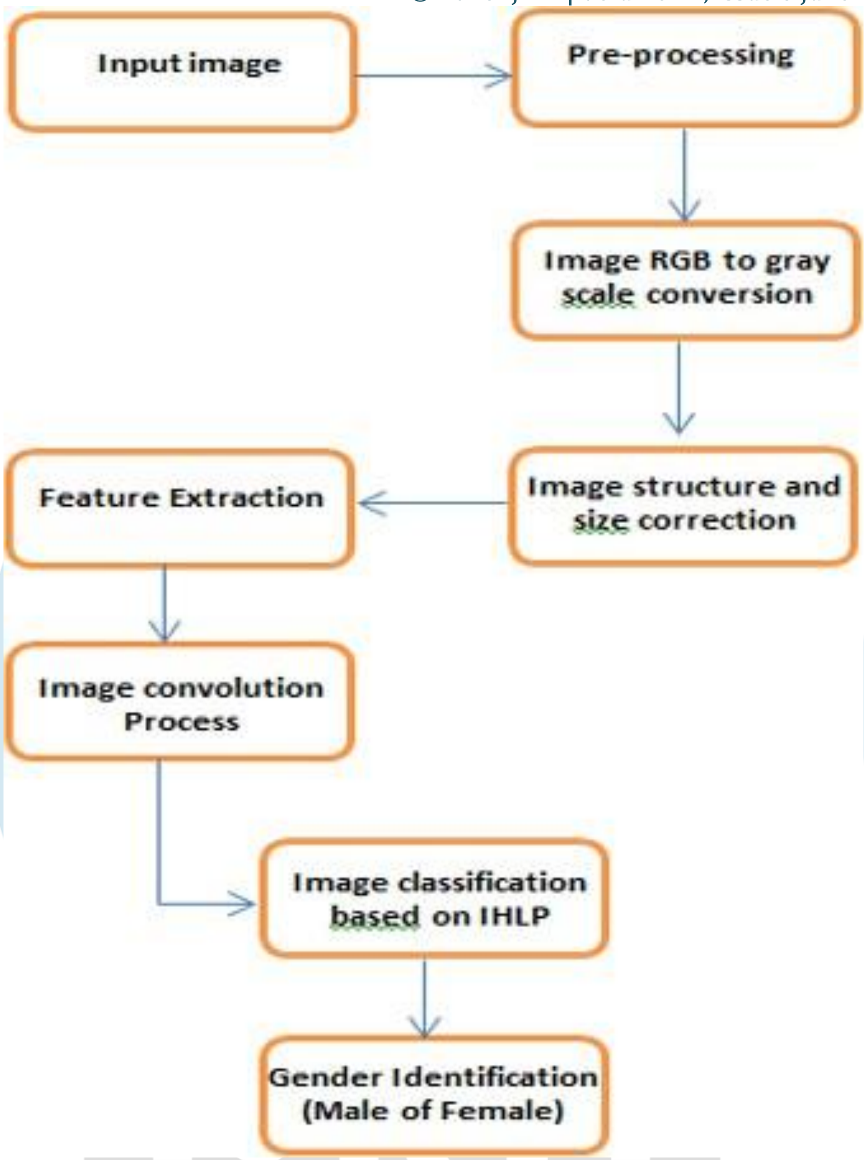

Fig. 5 IHLP workflow showing CNN feature extraction integrated with IoT data storage [4].

2.6 CNN vs SVM for Age and Gender Prediction

Omkar Singh and Mourya [2] compared Support Vector Machines (SVMs) and CNNs for age and gender guessing. On a celebrity dataset, CNNs got 96.8% gender accuracy, doing better than SVM's 87.8%. Age guessing was harder because visual features overlap and lifestyles differ. CNNs were better than SVMs at learning important patterns without needing help, showing that deep learning can be used in different situations and can grow.

2.7 Deep Iris CNN and CNN-SVM Hybrids

Rajput and Sable [15] made a hybrid AlexNet/GoogLeNet–SVM model for guessing age and gender using the IIT-A iris dataset, making sure that iris textures have age and gender data. The system got 84.66% accuracy, proving that CNN–SVM mixing can be used for soft biometrics. Also, Khalifa et al. [12] suggested Deep Iris, a 16-layer CNN that uses graph-cut segmentation and data adding to make the training data bigger, going from 3,000 to 9,000 pictures, getting 98.88% accuracy for gender classification.

A comparison of various iris-based CNN architectures is provided in Figure 6, illustrating different design choices for ocular feature extraction.

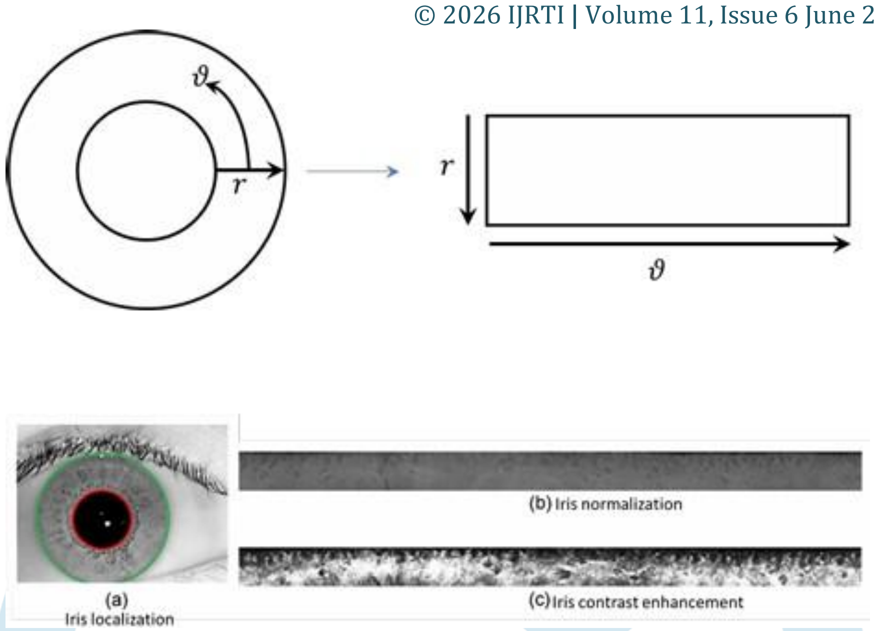

Fig. 6 Comparative representation of iris-based CNN architectures [15].

2.8 Deep Class-Encoder (DCE)

Singh et al. [16] showed Deep Class-Encoder (DCE) for guessing age and gender from irises. The model, trained on Notre Dame Gender and Age from Iris Dataset (ND-GFI), ND-Iris-0405 and CASIA, got 97% ethnicity and 94% gender accuracy, doing better than autoencoder and Deep Belief Network baselines by 4–14%.

2.9 CNNs for Ocular Gender Classification

Çimtay and Yilmaz [13] evaluated pretrained CNNs (InceptionV3, InceptionResNetV2, Xception, NASNetLarge) on 11,500 ocular images. Xception scored the highest accuracy at 96.9%, while others achieved 95.8–96.1%, proving that the eye region alone is enough for reliable gender prediction.

Figure 7 presents the ocular-region CNN pipeline that uses pretrained deep models to extract discriminative eye-region features.

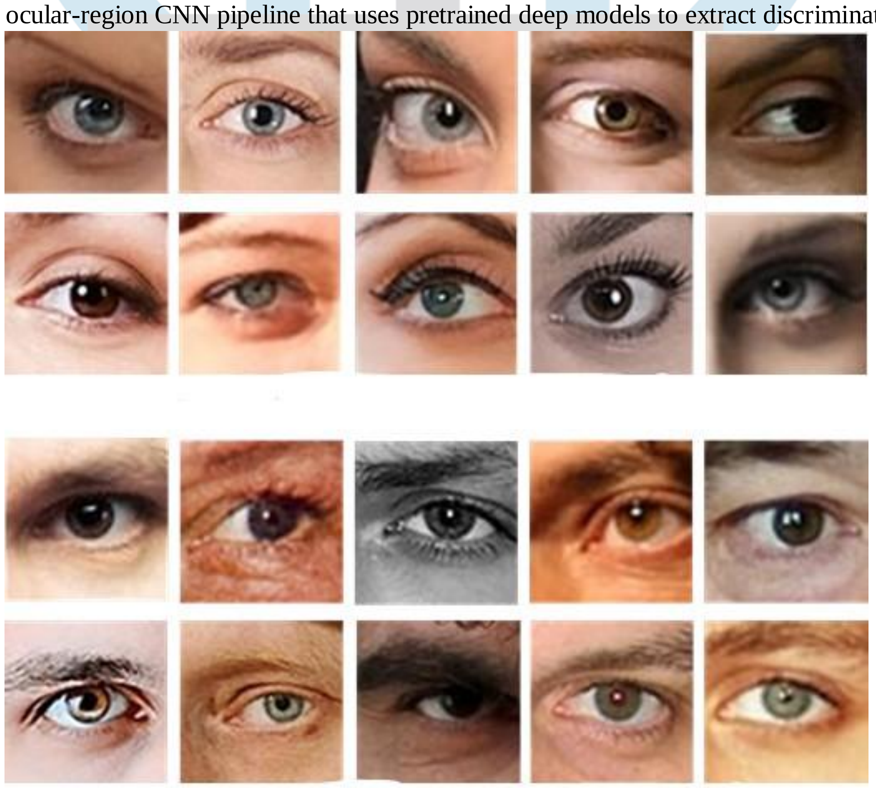

Fig. 7 Ocular-region CNN pipeline using pretrained models and fine-tuning layers [13].

2.10 Deep Learning on Mobile Ocular Images

Alonso-Fernandez et al. [10] studied CNNs (MobileFaceNets, MobiFace) for eye recognition on phones. With double fine-tuning on ImageNet, MS1M, Visual Geometry Group Face Dataset (VGGFace2) and Adience, the system got 93.52% gender accuracy and 70.96% age classification. This showed that you can have good eye biometric systems on phones.

2.11 Study of Iris-Based Gender Classification

Hasan and Mstafa [6] did a review of 20 years of iris gender recognition research, figuring out that CNN methods made accuracy 10–15% better compared to LBP, Gabor or PCA ways. Combined CNN–ELM systems got up to 99.8% accuracy, but dataset bias is still a concern.

3 Comparative Performance Analysis

Table 1 summarizes key face-based and hybrid deep learning models used for age and gender prediction. It highlights the model architecture, dataset used, modality and achieved accuracy. From the comparison, it is observed that attention-based and hybrid representations generally perform better under variations in lighting, pose and occlusion. Models such as GRA_Net demonstrate consistently high gender classification accuracy by integrating gated residual attention mechanisms.

Table 1 Face-Based and Hybrid Deep Learning Models

Study	Model	Dataset	Accuracy (%)	Modality	Remarks
Antipov et al. (2016) [17]	Minimalistic CNN Ensemble	LFW	97.31	Face	Lightweight, embedded deployment
Pasagadugula et al. (2024) [1]	CMNV2	Masked Face Dataset (12,160)	96.54	Masked	Robust under occlusion
Saha et al. (2023) [3]	AGC–CNN	Adience	97.1	Face	Illumination correction
Vinayagam et al. (2023) [4]	IHLP + IoT	Custom	94.8	Hybrid	Real-time IoT integration
Garain et al. (2021) [11]	GRA_Net	UTKFace / FG-NET	99.2	Face	Attention-driven accuracy
Omkar Singh and Mourya (2024) [2]	CNN / SVM	Custom	96.8 / 87.8	Face	CNN superior to SVM

In addition to face-based approaches, several iris and ocular-based models have also been explored for age and gender prediction as shown in Table 2. These models leverage the stability of ocular texture patterns, making them robust in cases where facial features are partially occluded. However, their performance depends heavily on image acquisition quality and requires close-proximity or specialized sensors.

Table 2 Iris and Ocular-Based Models

Study	Model	Dataset	Accuracy (%)	Modality	Remarks
Çimtay & Yilmaz (2021) [13]	Xception CNN	Eye-region (11,500)	96.9	Ocular	High recall and F1-score
Singh et al. (2017) [16]	Deep Class-Encoder	ND–GFI / CASIA	94	Iris	Label-guided feature extraction
Rajput & Sable (2019) [15]	AlexNet + SVM	IIT–A Iris	84.66	Iris	Hybrid CNN–SVM approach
Khalifa et al. (2021) [12]	Deep Iris (16-layer CNN)	Custom (3,000–9,000)	98.88	Iris	Data augmentation boost
Hasan & Mstafa (2022) [6]	Meta Survey	Multiple	Up to 99.8	Iris	CNN–ELM outperforms classical
Alonso-Fernandez et al. (2022) [7]	MobileFaceNets	Adience	93.52	Ocular (Mobile)	Lightweight mobile CNN

Overall, face-based deep learning models tend to achieve higher accuracy on large-scale datasets, whereas iris-based methods offer improved robustness under occlusion and masking conditions.

A comparative visualization of accuracy across facial and ocular modalities is shown in Figure 8, highlighting performance differences among key models.

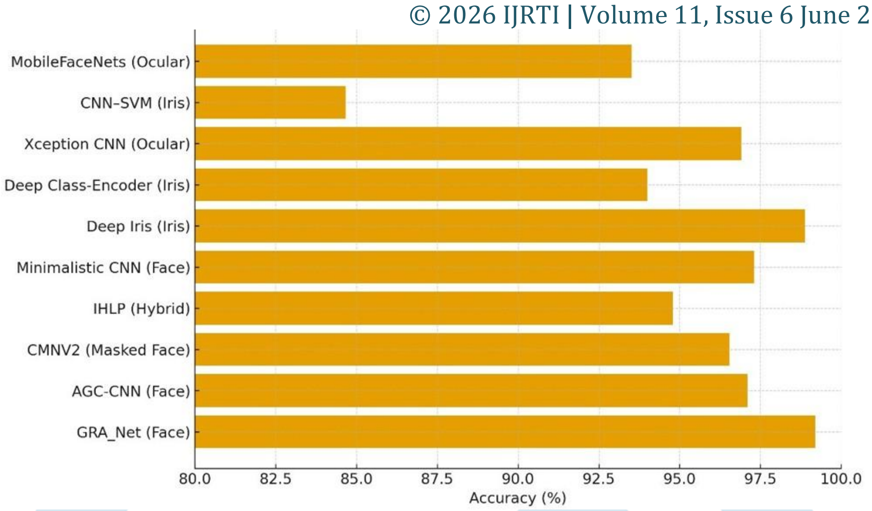

Fig. 8 Comparative graph of accuracy across facial and ocular modalities.

4 Discussion

4.1 Advancements in CNN Architectures

Recent studies show how residual pieces and attention parts (like in GRA_Net) lower overfitting and better find age and gender details. Regular fixing (AGC) also makes CNNs better at seeing texture and lighting.

4.2 Lightweight and Hybrid Models

The Simple CNN [17] and CMNV2 [1] make sure that good numbers don't have to be slow. Also, IHLP [4] connects CNN data with IoT for real-time age and gender tracking. MobileNetV3-based architectures also enable real-time performance on embedded devices [9].

4.3 Ocular and Iris-Based Resilience

Deep Iris [12] and Class-Encoder [16] make sure that iris patterns have age and gender data, getting greater than 94% accuracy even with small datasets. But, dataset variety is still a problem.

4.4 Mobile and Edge Deployment

MobileFaceNets [7] show that age and gender recognition can work on devices for small spaces. Edge CNNs, with double fine-tuning, get greater than 93% gender accuracy.

4.5 Ethical and Privacy Considerations

Even though deep models are good at guessing, moral problems come up when we guess age and gender without asking. Privacy-protecting machine learning and shared ways will be important for good use.

5 Future Work

- **Cross-Dataset Adaptation:** Use domain learning to fix age and gender bias.
- **Explainable AI:** Make CNN guesses easier to understand by showing gradient-based pictures.
- **Privacy-Aware Systems:** Add shared training for safe biometric work.
- **Small Embedded AI:** Try pruning, quantization and distillation for IoT.
- **Multi-Modal Fusion:** Connect face, iris and voice data for better guesses.

Figure 9 illustrates a conceptual hybrid explainable framework for future age and gender prediction systems.

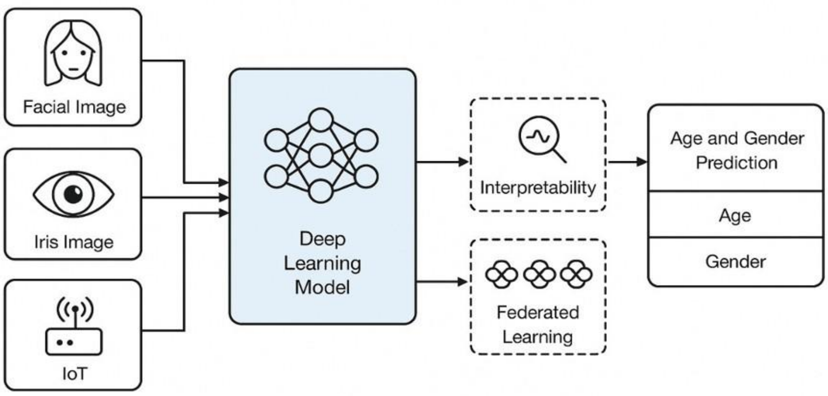

Fig. 9 A future hybrid explainable framework for age and gender prediction.

6 Conclusion

Mixing deep and hybrid learning models has made age and gender guessing systems better and more useful. CNN systems like GRA_Net and AGC-CNN have gotten close to human accuracy, while hybrid systems like IHLP and CMNV2 are strong even when something is blocking the view and in real-time conditions. Eye and iris CNNs, like Deep Iris and Deep Class-Encoder, have made other biometric ways that put privacy and being the same first. The future is about systems that are easy to understand, light and morally right, mixing AI data with good practice. Audience-aware demographic datasets have shown improved system accuracy in practical deployments [5].

List of Abbreviations

Table 3 List of Abbreviations

Abbreviation	Full Form
AFAD	Asian Face Age Dataset
AGC	Adaptive Gamma Correction
AGC–CNN	Adaptive Gamma-Corrected Convolutional Neural Network
CAFFE	Convolutional Architecture for Fast Feature Embedding
CASIA	Chinese Academy of Sciences Institute of Automation (Face/Iris Dataset)
CNN	Convolutional Neural Network
CMNV2	CAFFE-Modified MobileNetV2
DCE	Deep Class-Encoder
FG-NET	Face and Gesture Recognition Network (Aging Dataset)
FPS	Frames Per Second
GRA_Net	Gated Residual Attention Network
IHLP	Intelligent Hybrid Learning Principle
IoT	Internet of Things
IIT-A	IIT Delhi Iris Dataset
LBP	Local Binary Patterns
LFW	Labeled Faces in the Wild
MAE	Mean Absolute Error
ND-GFI	Notre Dame Gender and Age from Iris Dataset
ND-Iris-0405	Notre Dame Iris Dataset
PCA	Principal Component Analysis
RAN	Residual Attention Network
SVM	Support Vector Machine
UTKFace	UTK Face Age and Gender Dataset
VGGFace2	Visual Geometry Group Face Dataset (Version 2)

References

[1] Pasagadugula, A., Varsha, E., Jacob, T.P.: Masked face age and gender identification using CAFFE-modified MobileNetV2. Expert Systems with Applications 246 (2024)

[2] Singh, O., Mourya, K.: Gender and age detection using machine learning algorithms. International Journal of Creative Research Thoughts (IJCRT) 12(3), 32–38 (2024)

[3] Saha, A., Kumar, N., Nithyakani, P.: Age and gender prediction using adaptive gamma correction and convolutional neural network. In: Proceedings of IEEE ICCCI (2023)

[4] Vinayagam, P., Kumar, V.N., Bhuvan, G.: Experimental evaluation of IoT-based human gender classification and record management using intelligent hybrid learning principles. In: Proceedings of IEEE ICACIC (2023)

[5] Patel, S., Toral, V.: Age and gender detection using audience dataset. IEEE Transactions on Multimedia 25, 1296–1307 (2023)

[6] Hasan, B.M.S., Mstafa, R.J.: A study of gender classification techniques based on iris images: A deep survey and analysis. Science Journal of University of Zakho 10(4) (2022)

[7] Alonso-Fernandez, F., Hernandez-Diaz, K., et al.: Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images. Image and Vision Computing 141 (2022)

[8] Nithyakani, P., et al.: Hybrid machine learning frameworks for facial age estimation: A review. AI Review Journal 34 (2022)

[9] Chi, H., Li, X., Li, S.: MobileNetV3-based gender classification for embedded applications. Sensors 22(4), 1562 (2022)

[10] Alonso-Fernandez, F., Hernandez-Diaz, K., Ramis, S., Perales, F.J., Bigun, J.: Facial masks and soft-biometrics: Leveraging face recognition CNNs for age and gender prediction on mobile ocular images. IET Biometrics 10(5), 562–580 (2021)

[11] Garain, A., Ray, B., Singh, P.K., Sarkar, R.: GRA_Net: A deep learning model for classification of age and gender from facial images. IEEE Access 9, 85672–85687 (2021)

[12] Khalifa, N.E.M., Taha, M.H., et al.: Deep iris: Deep learning for gender classification through iris patterns. Journal of King Saud University – Computer and Information Sciences (2021)

[13] Çimtay, Y., Yilmaz, G.N.: Gender classification from eye images using pre-trained convolutional neural networks. In: Eurasia Proceedings of Science, Technology, Engineering and Mathematics (2021)

[14] Sharma, A., Kumari, R., Singh, S.: Deep learning techniques for gender and age prediction. Procedia Computer Science 167, 1821–1830 (2020)

[15] Rajput, M.R., Sable, G.S.: Deep learning based gender and age estimation from human iris. In: Proceedings of ICAEEC, IIIT Allahabad, pp. 1–9 (2019)

[16] Singh, M., Singh, R., Ross, A.: Gender and ethnicity classification of iris images using deep class-encoder. In: Proceedings of IEEE IJCB (2017)

[17] Antipov, G., Berrani, S.-A., Dugelay, J.-L.: Minimalistic CNN-based ensemble model for gender prediction from face images. Pattern Recognition Letters 70, 59–65 (2016)