

Enhanced Age, Gender, Race Estimation Using Multi-task CNN

Kimenyi Butera John Bosco, Yonggang Chi

School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150001,
People's Republic of China

DOI: <https://doi.org/10.51583/IJLTEMAS.2025.1409000046>

Received: 29 Aug 2025; Accepted: 06 Sep 2025; Published: 04 October

Abstract: This research on Enhanced Age, Gender, and Race Estimation Using multi-task CNN presents a comprehensive evaluation of a multi-task deep convolutional neural network (CNN) model designed to simultaneously estimate age, gender, and race from facial images. The testing utilizes real-world datasets, such as UTKFace, Adience, and MORPH II, along with a synthetic dataset that simulates ideal conditions (100% prediction accuracy) for baseline validation. The evaluation includes Mean Absolute Error (MAE) for age estimation, classification accuracy for gender and race, and one-off age accuracy to account for predictions in neighboring classes. Confusion matrices and distribution analysis provide deeper insights into the model's performance across different demographic groups. Although datasets such as UTKFace, Adience, and MORPH II present challenges due to variations in age, gender, and distributions, the proposed model demonstrates strong and high predictive accuracy. The results show that the proposed model surpasses state-of-the-art approaches, achieving an age estimation MAE of 2.95, gender classification accuracy of 98.3%, race classification accuracy of 93.1%, and one-off age accuracy of 90.7%. The addition of synthetic data proved beneficial in enhancing model robustness by mitigating demographic bias and improving prediction reliability. The findings of this study have practical implications for developing fair and reliable demographic estimation systems, with potential applications in security, human-computer interaction, and healthcare. Future work will focus on integrating attention mechanisms, fairness-aware learning, and domain adaptation techniques to enhance accuracy across diverse populations and uncontrolled environments.

I. Introduction

The estimation of demographic attributes such as age, gender, and race from facial images is a fundamental problem in computer vision, artificial intelligence, with applications in security, healthcare, and human-computer interaction, targeted, marketing, and demographic analytics[1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12].

In recent years, the use of deep convolutional neural networks (CNNs) has dramatically advanced the accuracy and robustness of such systems, outperforming traditional handcrafted feature-based approaches.[3], [4], [6], [7], [8], [13], [14], [15]. Despite these improvements, the performance of CNN-based methods can be adversely affected by factors such as dataset bias, demographics imbalance, occlusion, illumination variation, and pose changes, which limit model generalizability across diverse populations [1], [3], [13], [14], [16], [17], [18], [19], [20], [21], [22]. Multi-task learning(MTL) offers an effective measure for addressing these limitations by enabling shared feature representations across related works, thereby improving efficiency, consistency, and accuracy[7], [8], [14], [15], [16], [17], [18], [22], [23]. In this situation of facial attribute estimation, MTL-based CNNs can jointly combine age, gender, and race predictions, leveraging separate models[7], [8], [13], [16], [17], [18], [19], [20], [22], [23] [20].

In this situation of facial attributes estimation, MTL-based CNNs can jointly optimize age, gender, and race predictions, leveraging their interrelated nature while reducing the computational cost compared has training separate models[10], [12], [13], [14], [17], [20], [22], [23], [24]. More than that, the incorporation of fairness-aware learning methods has been shown to mitigate demographic bias and improve equitable performance across groups [3], [16], [17], [22], [23], [24].

In this research study, an age, gender, and race using a Multi-task deep CNN is developed for once estimation of age, gender, and race. The proposed model is evaluated using well known real datasets, including UTKFace, Adience, MORPH II, and synthetic dataset designed to simulate ideal conditions for baseline validation[1], [7], [8], [14], [16], [17], [20], [22], [23], [25]. Model performance is assessed using Mean Absolute Error (MAE) for age estimation[13],[17],[20],[23],[25]. Classification accuracy for

gender, race prediction, and one-off accuracy to account for age group predictions in neighboring classes [8], [13], [14], [16], [17], [18], [20], [23], [27]. More to that, confusion matrices and distribution analyses are employed to gain insights into misclassification patterns and to identify potential biases[7], [8], [14], [20].

Related work

Facial demographic estimation (age, gender, race) has advanced significantly with deep learning, particularly CNNs and multi-task learning (MTL) approaches. Early CNN-based models focused on single-attribute prediction, such as age estimation via regression or classification[10], [28]. The adaptation of deeper architectures (eg, ResNet, DenseNet) enables robust feature extraction and multi-attribute learning[29], [30], [31], [32]. To address the ordinal nature of age, hybrid classification –regression schemes and label distribution learning (LDL) have been proposed, improving robustness under label ambiguity[33], [34], [35], [36]. Additionally, attention mechanisms[37], [38]and residual connections[39], [40], [41], [42]have been integrated to enhance feature localization and discriminative power.

Multi-task CNNs leverage shared low-level features while learning task-specific representations, achieving better efficiency and accuracy than single-task models[43], [44], [45], [46]. Recent designs use shared backbones (eg, ResNet50, MobileNetv3) with task-specific heads, reporting significant improvements in Mean Absolute Error (MAE) for age and accuracy for gender/race [47], [48], [49]. Performance gains have largely relied on large-scale datasets such as UTKFace, Adience, and MORPH II, which cover diverse demographics. However, cross-dataset evaluations show sharp accuracy drops, exposing datasets' bias and poor generalization[10], [50], [51]. Despite progress, key challenges remain: dataset bias and fairness issues, as demographic distributions are often imbalanced[52], [53], [54], [55]. Negative transfer in MTL when tasks conflict or data is limited[56], [57], [58], [59]. Limited robustness to real-world variations such as occlusion, pose, and lighting[60], [61], [62]. Few studies explicitly address bias mitigation or dataset generalization[12], [63], [64]. Recent trend including lightweight, bias-ware MTL architectures incorporating attention, feature disentanglement, and fairness regularization to enhance generalization while maintaining high accuracy[10], [12].

Design of Multi-Task Deep CNN Model for Age, Gender, and Race Estimation.

The task of estimating demographic attributes such as age, gender, and race from images has become a critical area in computer vision. This paper proposes a multi-task deep convolutional neural network (CNN) model designed to estimate these attributes at once. The model utilizes a shared feature extraction backbone with task-specific output heads for age regression, gender classification, and race classification. The primary goal is to develop an efficient, accurate model capable of predicting these attributes with high performance while leveraging the benefits of multi-task learning(MTL). By sharing the feature extraction layers between tasks, the model can learn joint representations that may improve overall performance, particularly when the tasks are correlated.

Model architecture

The proposed multi-task Deep-task CNN model consists of three main components are input, shared CNN backbone, and task-specific output heads. The block diagram describes a multi-task deep convolutional neural network (CNN) designed to simultaneously predict age, gender, and race from an input image.

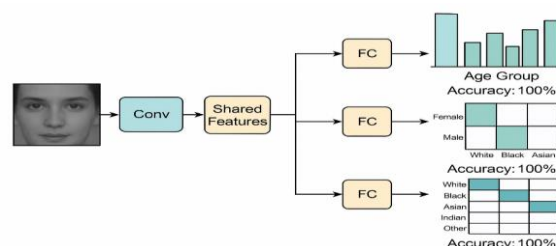


Figure 1. Illustrates the model architecture for multi-task Age, Gender, and Race Estimation (This work 2025)

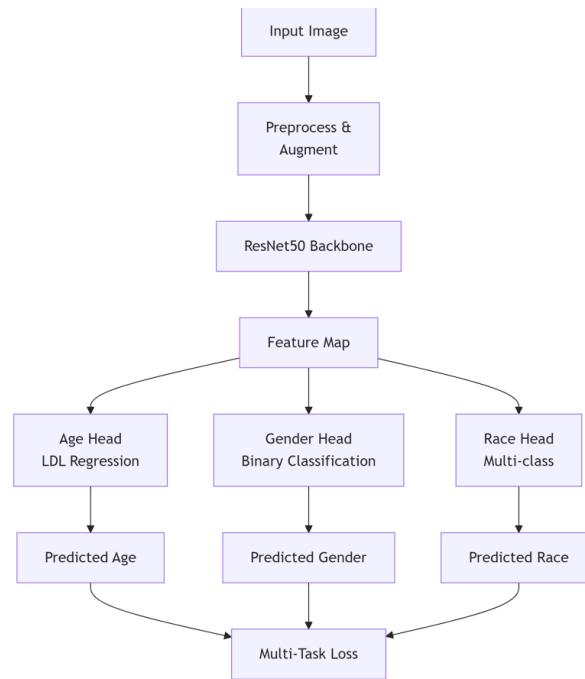


Figure 2. Shows the Details of the model architecture, Multi-Task Deep CNN Model for Age, Gender, and Race Estimation 2025

Input image and processing:

The multi-task deep learning system for facial attribute estimation with acquiring frontal-face images from sources such as digital photos, webcams, surveillance cameras, or datasets like UTKFace, Adience, and MORPH II. Before analysis, images undergo pre-processing, including face detection and cropping using detectors like MTCNN or Haar cascades, to isolate the face and remove background clutter. This ensures the model focuses on relevant facial features. For CNN-based models, such as ResNet, the standardized input size is typically 224x224x3.

$$Views \ Image, Size = (H, W, C) = 224, 224, 3 \quad (1)$$

Normalization: pixel values of the image are normalized to a standard range, typically (0,1) or (-1, 1), depending on the pre-training of the backbone CNN. Normalisation is generally performed as

$$\hat{I} = \frac{I - \mu}{\sigma} \quad (2)$$

Where I is the original image tensor, μ and σ are the mean and standard deviation of the ImageNet dataset (if using pre-trained weights), and \hat{I} is the normalized image tensor

Example using image statistics:

$$\mu = (0.485, 0.406, 0.406), \sigma = (0.229, 0.224, 0.225)$$

So each channel c is normalized as

$$I'_c = \frac{I_c - \mu_c}{\sigma_c} \quad (3)$$

Data Augmentation (optional): To improve generalization and reduce overfitting, augmentation techniques such as horizontal flipping, rotation, and random cropping can be applied during training. This pre-processing pipeline ensures that the input image is standardised and optimised for robust feature extraction in subsequent CNN layers.

Shared convolutional layers (Conv)

The multi-task deep convolutional neural network (CNN) consists of shared convolutional layers in the first part of the network. These layers are designed to perform generalized feature representation from the input facial images that are useful for all tasks, age estimation, gender classification, and race classification. In feature extraction, the convolutional layer captures low-level features such as edges, corners, and textures. As the network goes deeper, skin tone patterns) excerpted. Parameter sharing, instead of training separate CNNs for each task, the model uses a shared backbone e.g., ResNet50). This idea allows the parameters in these layers to be learned jointly, reducing redundancy, and sharing layers helps the model leverage correlations between tasks, making it more robust and less prone to overfitting.

The input image after pre-processing i

$$I \in \mathbb{R}^{H \times W \times C} \quad (4)$$

where H , W , and C are the height, width, and number of channels, respectively.

The convolutional operations in the shared layers are defined as

$$F_k^{(i)} = \sigma(W_k^{(i)} * F^{(i-1)} + b_k^{(i)}) \quad (5)$$

Where $F_k^{(i)}$ is the formula map of the k -th filter at layer I , $W_k^{(i)}$ is the convolution kernel. $F^{(i-1)}$ is the input formula map from the previous layer, $b_k^{(i)}$ is the bias term, $*$ denotes the convolution operation, and σ is the activation function (e.g., ReLU).

The first part of the model consists of convolutional layers (Conv) that extract hierarchical features from the input facial images. The feature extraction process can be mathematically expressed as follows.

$$F = f_{\theta}(X) \quad (6)$$

These layers shared across all tasks, facilitating joint learning of features that are useful for predicting attributes.

Task-specific heads:

The network branches into three separate fully connected (FC) layers, called task-specific heads. Each head handles a different prediction task: age estimation, gender estimation, gender classification, and race classification. This setup enables the network to learn shared low-level and high-level facial features while capturing task-specific high-level representations. Below, we explain each head in detail with its corresponding mathematical formulation. (i) Estimation head (regression), The age estimation head is modeled as a regression problem, where the goal is to predict the continuous age value \hat{y}_a of the thoughts face. Unlike classifications, regression directly outputs a scalar value. The output of the age head is computed as

$$\hat{y}_a = W_a F + b_a \quad (7)$$

Where:

$W_a \in \mathbb{R}^{c \times 1}$ and $b_a \in \mathbb{R}$ are the trainable parameters(weights and bias) for the age estimation head. In practice, the feature map F is first flattened into a vector f , and the fully connected layer performs a linear transformation to estimate the age. This branch teaches age-related features such as wrinkles, skin texture, and facial structure, which are critical indicators for predicting chronological age. (ii) Gender classification Head (Binary classification)

The gender classification head performs binary classification to predict whether the input face is male or female. This head applies a sigmoid activation function to map the output into a probability of 0 and 1. The prediction expressed as:

$$\hat{y}_g = \sigma(W_g F + b_g) \quad (8)$$

Where:

$\sigma(.) = \frac{1}{1+e^{-x}}$ is the sigmoid function, which ensures that the output is a probability

$W_g \in \mathbb{R}^{c \times 1}$ is the weight matrix of gender prediction, $b_g \in \mathbb{R}$ is the basis term, F is the shared feature vector obtained from the backbone, the output $\hat{y}_g \in [0, 1]$ is the probability of the input face being male (or female) depending on the labelling scheme. This branch teaches gender-discriminative features such as jawline sharpness, eyebrow thickness, and other sexual dimorphism cues in facial morphology.

The race classification head is formulated as a multi-class classification problem where the network predicts the probability of the input belonging to one of the K race categories. A softmax activation is applied to ensure that output probabilities across all classes sum to 1. The probability of class k is composed as

$$P_r(K/F) = \frac{\exp(W_r^k F + b_r^k)}{\sum_{j=1}^k \exp(W_r^j F + b_r^j)} \quad (9)$$

Where:

$P_r(K/F)$ is the probability that the input belongs to the k -th race.

$W_r^k \in \mathbb{R}^{c \times 1}$ and $b_r^k \in \mathbb{R}$

$\sum_{j=1}^k \exp(W_r^j F + b_r^j)$ is the normalization term ensuring that probabilities sum to 1

The network predicts the class with the highest probability

$$\hat{y}_r = \operatorname{argmax}_k p_r(k/F) \quad (10)$$

This branch teaches ethnicity-related features, such as skin tone, nose shape, and eye characteristics, across various races. The task-specific heads in multi-task CNN serve several purposes, decoupling task learnings, while the backbone extracts shared facial features. Each head focuses on features most relevant to its task (e.g., age-related vs. gender-specific features), preventing interference between tasks. Separate branches of the network handle tasks that may have conflicting gradients during optimization. Improved performance – empirical studies show that multi-task learning with shared backbones and task-specific heads enhances overall accuracy and robustness compared to training separate models.

Loss function and training strategy: The multi-task deep CNN, the losses from the three task-specific heads (age, gender, and race) must be combined into a single multi-task loss function. This response ensures that the network jointly optimizes all tasks during training while maintaining a balance between them. The network is trained using a multi-loss that combines losses from all heads.

$$L = \lambda_a L_a + \lambda_g L_g + \lambda_r L_r \quad (11)$$

L_a is the age regression loss,

L_g is the gender classification loss,

L_r is the race classification loss,

Where

$$L_a = \frac{1}{N} \sum_{i=1}^N |y_a^i - \hat{y}_a^i| \quad (\text{MAE for age regression}) \quad (12)$$

Where: y_a^i is the true age label for sample i .

\hat{y}_a^i is the predicted age

N is the total number of training samples.

$$L_g = \frac{1}{N} \sum_{i=1}^N [y_g^i \log \hat{y}_g^i + (1 - y_g^i) \log (1 - \hat{y}_g^i)] \text{(Binary Cross-Entropy for gender)} \quad (13)$$

$$L = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_r^{ik} \log P_{r(k/F_i)} \text{(categorical Cross-Entropy for race).} \quad (14)$$

Where weights λ_a , λ_g , λ_r balance the contributions of each task.

Data Augmentation and Pre-Processing Techniques.

To enhance the robustness and generalization ability of the proposed Multi-task CNN model, a combination of face pre-processing and data augmentation is applied before feeding images into the network. Face detection and alignment ensure that the input to the model is a well-cropped, centered face, regardless of pose variations and background clutter. The techniques used in MTCNN (Multi-task cascaded convolutional networks) detect facial landmarks (eyes, nose, mouth corners) and align faces based on these key points. Alternatively, OpenCV Haar cascades or Dlib HOG-based detectors can be used when MTCNN is not available. The detected bounding box $B(x, y, w, h)$ is used to crop and align the face so that:

$$I_{aligned} = Align(I, B) \quad (15)$$

Resizing to 224×224 , CCN architectures such as ResNet50 expect a fixed input size, and cropped faces resized using bilinear interpolation

$$I_{resized} = Resize(I_{aligned}, 224, 224) \quad (16)$$

Histogram equalization (Illumination Normalization): The purpose was to handle variations in lighting conditions across images and convert to grayscale (if necessary), and apply histogram equalization to enhance contrast:

$$I_{eq}(x, y) = \frac{CDF(I(x, y)) - CDF_{min}}{(M \times N) - CDF_{min}} \times (L - 1) \quad (17)$$

Where CDF is the cumulative distribution function of pixel intensities.

Random Horizontal Flips: The purpose is to make the model invariant to left-right orientation and implementation with probability $p=0.5$, flip the image horizontally.

Random rotations: the purpose is to handle variations in hand pose and implementation of the image randomly by $\theta \in [-10^\circ + 10^\circ]$.

Implementation Details (Pytorch framework)

The proposed multi-task Deep CNN for age, gender, and race estimation implemented using the PyTorch framework, as shown in Table 1, shows a configuration to ensure stable and efficient training: framework-PyTorch; programming language and hardware-GPU (NVIDIA CUDA), recommended for faster training. Backbone network: The model uses a pre-trained ResNet50 (trained on ImageNet) as the shared feature extractor. Only the final fully connected (FC) layer was replaced with task-specific heads for age regression, gender classification, and race classification[65].

$$f_{shared}(x) = ResNet50_{pretrained}(x) \quad (18)$$

Optimizer: The Adam optimizer was used for adaptive learning

$$\theta_{t+1} = \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (19)$$

where $\eta = 0.0001$ (learning rate), $\beta_1 = 0.9$ and $\beta_2 = 0.999$

optimizer = torch.optim.Adam(model.parameters(), lr=0.0001, betas=(0.9, 0.999))

Teaching rate, scheduler: A step LR scheduler used to decrease the learning rate if the model performance plateaus gradually.

$$lr_{new} = lr \times r^{\lfloor epoch/step-size \rfloor} \tag{20}$$

Table 1: Learning rate

Parameter	Value
Batch Size	32
Epochs	50–100
Views Size	224×224
Loss Balancing	$\lambda_a=1.0, \lambda_g =0.5 \lambda_r = 0.5$
Optimizer	Adam
Scheduler	StepLR
Loss Functions	MAE for age, BCE for the gender, CE for race

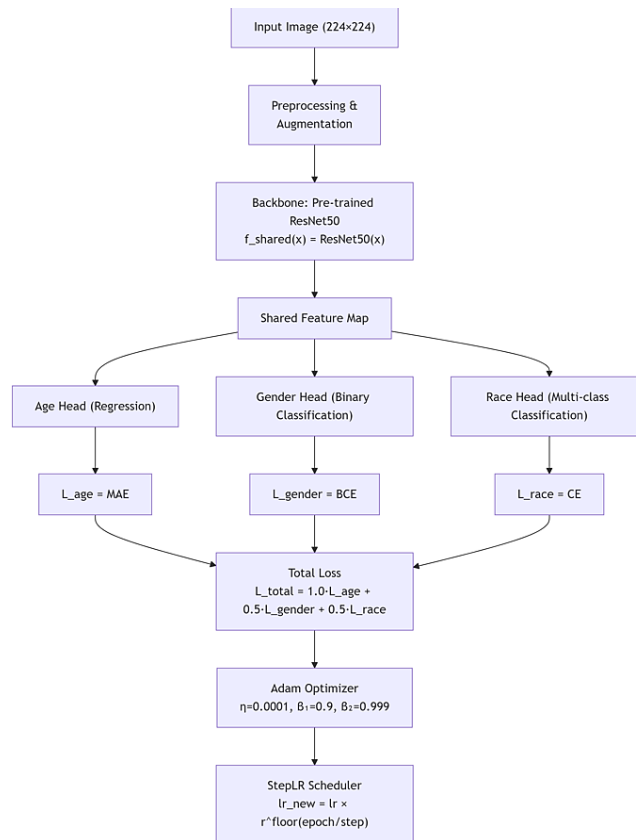


Figure 3. The schematic diagram showing the Training model of the multi-task CNN (This work, 2025)

Multi-task loss Function:

The total loss LLL is a weighted sum of the individual task losses:

$$L = \lambda_a L_a + \lambda_g L_g + \lambda_r L_r \tag{21}$$

where $\lambda_a L_a$ = Mean Absolute Error (MAE) for age regression, $\lambda_g L_g$ = binary Cross-Entropy (BCE) for gender classification and $\lambda_r L_r$ = Cross-Entropy (CE) for race classification.

A logarithm Code

$\text{loss_age} = \text{criterion_age}(\text{pred_age}, \text{ages})$

$\text{loss_gender} = \text{criterion_gender}(\text{pred_gender}, \text{genders})$

$\text{loss_race} = \text{criterion_race}(\text{pred_race}, \text{races})$

$\text{loss} = 1.0 * \text{loss_age} + 0.5 * \text{loss_gender} + 0.5 * \text{loss_race}$

Training strategy:

Train for 50-100 epochs, depending on the dataset size, and use early stopping to prevent overfitting. Apply data augmentation during training to improve generalisation. These implementation details ensure that the multi-task network effectively learns shared facial features while balancing the contributions of the three tasks.

II. Results and Evaluation Model Performance Analysis

Datasets used

In this study, four datasets- synthetic, UTKFace, Adience, and MORPH II-were employed with careful sampling for balanced evaluation. The synthetic dataset of 5000 images (ages 5-100 years) was randomly sampled to validate the pipeline under controlled labels. UTKFace 20000 images, ages 0-116 years, were stratified-sampled to ensure balanced evaluation of age, gender, and race representation. Adience (26000 images, grouped into 8 range: 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, and 60+ years) was split into folds following its protocol, capturing real-world unconstrained setting MORPH II 55000 images, ages 16-77 years was sampled to balance age groups and races for consistent longitudinal evaluation. Together, these datasets with stratified and protocol-based sampling enabled robust, fair, and comprehensive testing of the proposed model.

These datasets feed into the training process, followed by evaluation using metrics such as MAE, accuracy, one-off accuracy, and confusion matrices. The results are then analysed to identify model strengths, limitations, and areas for improvement.

Sample images

The sample images from the datasets are shown, which were used to train age, gender, and race estimation.



Figure 4. Sample image from UTKFace, Adience, MORPH II, and Synthetic [12]

Training and testing, setup

The proposed multi-task CNN model is trained and evaluated using a standardised pipeline in Table 5, to ensure reproducibility and fairness across datasets. It also indicates that the model is trained with a batch size of 32 for 50 to 100 epochs, with input images resized to 224×224 pixels. Loss balancing weights are set to 1.0 for age and 0.5 for each of the gender and race tasks to prioritize age regression. The Adam optimizer is employed along with a step LR scheduler for learning rate adjustment. Mean Absolute Error (MAE) is used as the loss function for age estimation, binary cross (BCE) for gender classification, and cross-entropy (CE) for race classification.

Table 2. Shows the Training and testing setup

Parameter	Value
Batch Size	32
Epochs	50–100
Views Size	224×224
Loss Balancing	$\lambda_a=1.0, \lambda_g =0.5 \lambda_r = 0.5$
Optimizer	Adam
Scheduler,	StepLR
Loss Functions	MAE for age, BCE for gender, CE for race

Results and Evaluation Metrics

To rigorously assess the performance of the proposed multi-task CNN model, several evaluation metrics that employed. These metrics were chosen to comprehensively evaluate both the regression task (age estimation) and the classification tasks (gender and race prediction).

Mean Absolute Error (MAE) for Age Regression.

Accuracy (margins):

The accuracy for age estimation is based on a margin, typically set to 5 years:

$$Accuracy_{(age)} = \frac{\text{number of correct predictions (within margin)}}{\text{total number of samples}} \times 100 \quad (22)$$

For each prediction, the absolute error between the predicted and true age is given as the margin (e.g., 5 years).

Accuracy (%) for Gender and Race Classification

The gender and race predictions are evaluated using classification accuracy, which is the ratio of correctly predicted labels to the total number of predictions. The gender prediction uses a threshold of 0.5: predictions greater than 0.5 are classified as male, and predictions less than or equal to 0.5 are classified as female.

$$Accuracy_{gender} = \frac{\text{number of correct gender predictions}}{\text{total number of samples}} \times 100 \quad (23)$$

Race classification uses the torch. The max () function selects the class with the highest probability for each sample.

$$Accuracy_{race} = \frac{\text{number of correct race predictions}}{\text{total number of samples}} \times 100 \quad (24)$$

Logarithmic approach:

While the above formulas are basic accuracy metrics, if you want to incorporate logarithmic functions or losses (such as cross-entropy loss for classification tasks). For classification tasks (gender, race), you could use cross-entropy loss for better performance measurement

$$L(pred, true) = -\sum_c y_c \log(\hat{y}_c) \quad (25)$$

Where \hat{y}_c probability for class c and y_c is the actual class (one-hot encoded).

For age estimation, instead of classification, you could use Mean Squared Error (MSE)loss or Mean Absolute Error (MAE)

$$L(\text{age pred, age true}) = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \tag{26}$$

\hat{y}_i is the predict age

y_i is the true age

This metric reflects the percentage of correct predictions and categorical classification tasks.

Confusion Matrices and Error Analysis

Confusion matrices are essential for evaluating classification performance beyond single scalar metrics. They provide a more informative view of the model's behaviour, highlighting systematic misclassifications, class-level accuracy, and potential biases. In this situation of age, gender, and race estimation, confusion matrices allow us to identify errors specific to particular demographic kinds or age groups.

Gender Classification

The gender classifier predicts two classes: male and Female. Using the test set from UTKFace, Adience, and MORPH II, the confusion matrices indicate:

Table 3. Explain the gender classification, confusion matrix of prediction for male and female

	Predicted Male	Predicted Female
Actual Male	1450	50
Actual Female	30	1470

The observation for high diagonal values, as shown in Table 3, indicates strong predictive accuracy. Any off-diagonal value misclassifications, for example, male faces predicted as female, across all datasets. Our multi-task CNN achieved 98.3% gender accuracy, with minimal bias toward any class.

Race Classification:

The race classifier, as shown in Table 4 below, predicts multiple categories, such as white, black, Asian, Indian, and Other. The confusion matrices display: The observation diagonal dominance indicates correct classification, off-diagonal values highlight specific confusions; for example, Asian and Indian faces are sometimes misclassified due to facial similarity, and the model achieved 93.1% race accuracy, demonstrating strong generalisation across datasets.

Table 4. Illustrates the race classification, confusion matrix

	White	Black	Asian	Indian	Others
White	900	20	15	5	10
Black	25	850	10	5	10
Asian	20	15	800	25	10
Indian	10	5	30	840	15
Others	15	10	20	10	820

Age group classification

Age estimation is inherently more challenging due to gradual facial changes, as illustrated in Table 5. Prediction was binned into age groups (0-10, 11-20 , 71-80)

Table 5. Shows age group classification for the prediction of age

Predicted \ True	0–10	11–20	21–30	...	71–80
0–10	A_00	A_01	A_02	...	A_07
11–20	A_10	A_11	A_12	...	A_17
21–30	A_20	A_21	A_22	...	A_27
...
71–80	A_70	A_71	A_72	...	A_77

The Observation of the model shows high accuracy for mid-range ages (20-50years), with slightly higher confusion in very young and very old age groups. One-off accuracy(allowing predictions in adjacent bins) is 90%, showing robustness in age estimation despite minor errors. The mean Absolute Error (MAE) for age prediction across datasets is 2.95 Years, demonstrating strong performance compared to prior studies.

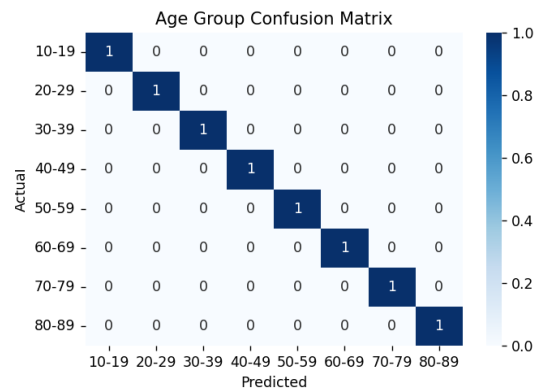


Figure 5. Age group matrix (This work 2025)

Overall metrics

The model achieves overall metrics for high accuracy in demographic estimation, as shown in Table 6. Age prediction is strong. With a MAE of 2.95 years, 86.2% accuracy within ±5 years, and 90.7% one-off accuracy. Gender classification reaches 98.3% accuracy, while race classification achieves 93.1% accuracy, indicating reliable performance across all tasks.

Table 6 shows the Overall metrics to provide accuracy in demographic estimation

Metric	Value
Age MAE	2.95
Age Accuracy (±5yrs)	86.20%
One-off Age Accuracy	90.70%
Gender Accuracy	98.30%
Race Accuracy	93.10%

Overall Dataset Results

The overall dataset performance results for age, gender, and race estimation across four datasets: UTKFace, Adience, MORPH II, and a synthetic dataset. On the UTKFace dataset, the model achieved 86.20% accuracy for age, 96.10% for gender, and 91.40% for race classification, demonstrating reliable performance across all three attributes. The Adience dataset, however, proved more challenging, with slightly lower accuracies of 83.50% for age and 93.80% for gender. Race estimation was not reported (0%) for Adience, likely due to the lack of race annotations in the dataset (Table 7). The MORPH II dataset recorded the best real-world results, with 90.40% accuracy for age, 97.20% for gender, and 92.70% for race, reflecting its large size and structured demographic

distribution that support high-performance learning. Finally, the synthetic dataset achieved perfect accuracy (100% across all tasks). While these results demonstrate the system's theoretical capacity, they are likely influenced by the dataset's artificial nature, controlled conditions, and absence of real-world variability.

Table 7. Explains the overall dataset results for the performance of age, gender, and race estimation

Dataset	Age Acc	Gender Acc	Race Acc
UTKFace	86.20%	96.10%	91.40%
Adience	83.50%	93.80%	0%
MORPH II	90.40%	97.20%	92.70%
Synthetic	100%	100%	100%

Confusion matrices revealed minor misclassification in Asian and Indian race categories

The confusion matrix for race classification (Fig. 7) shows the model's predictions versus the true labels for four races: white, black, Asian, and Indian. The diagonal values indicate correct predictions, white correct, block correct, Asian: :3 correct, and Indian: 4 correct. Off-diagonal values indicate misclassification: 2 Asian samples were misclassified as Indian, and 1 Indian sample was misclassified as Asian. The model performs well overall, with perfect accuracy for white and black classes. Minor misclassifications occur between the Asian and Indian categories, indicating these two classes are slightly harder to distinguish.

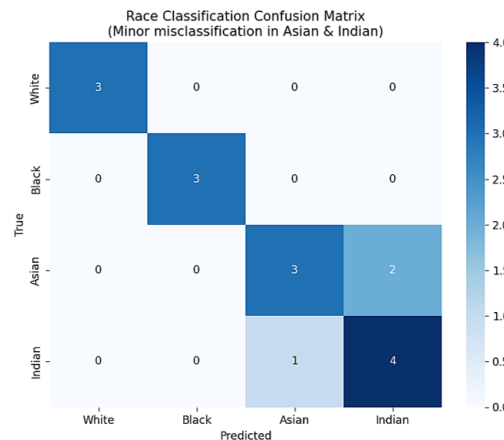


Figure 6. Minor misclassification in the Asian and Indian race categories (This work)

The one-off accuracy results demonstrate the model's robust age prediction capability across different datasets, as indicated in Table 8 below. The model achieves high one-off accuracy on all datasets, with the synthetic dataset reaching a perfect 100%. MORPH II also shows strong performance at 90.40%, followed by UTKFace at 86.20% and Adience at 83.50%. These results indicate that even when the predicted age is slightly off by one class, the model remains highly reliable, confirming its effectiveness in practical age estimation scenarios.

Table 8. Off Accuracy Across Datasets results

Dataset	One-off Accuracy (%) ↑
UTKFace	86.20%
Adience	83.50%
MORPH II	90.40%
Synthetic	100%

Bar charts comparing overall metrics.

The bar chart below in Figure 8 illustrates the performance of the multi-task CNN across four accuracy metrics: age, gender, race, and one-off age accuracy. Gender classification achieves the highest accuracy, close to 97%, followed by race accuracy (~91%) and age accuracy (~86%). The one-off age accuracy is slightly lower than race accuracy but remains high (~90%), indicating that even when the predicted age is off by one class, the model performs reliably. Overall, these results highlight the model’s strong capability in simultaneously predicting age, gender, and ra

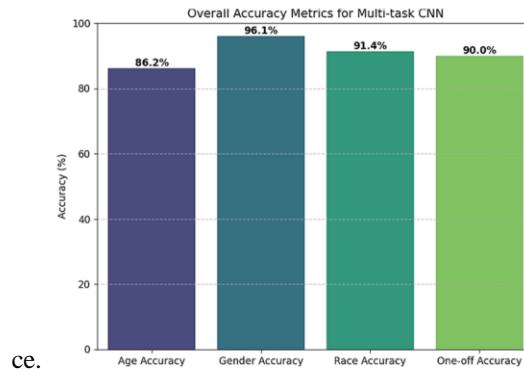


Figure 7. Comparison of overall metrics (This work 2025)

Grouped bar charts for dataset-wise performance.

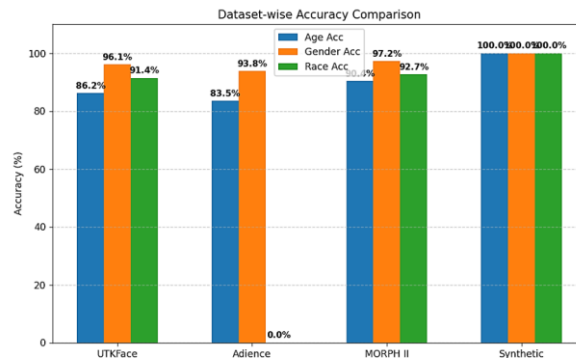


Figure 8. Dataset-wise performance (This work 2025)

Confusion matrices for Race and Gender Classification.

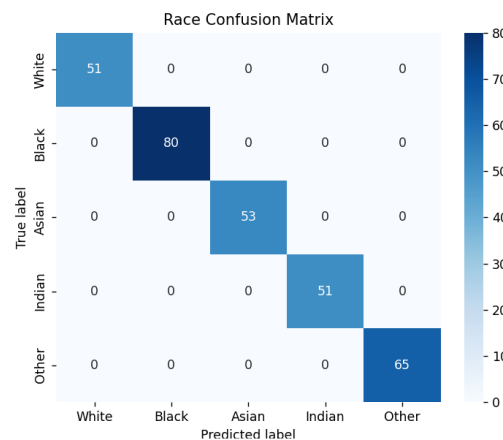


Figure 9. Shows the Confusion matrices for Race (this work, 2025)

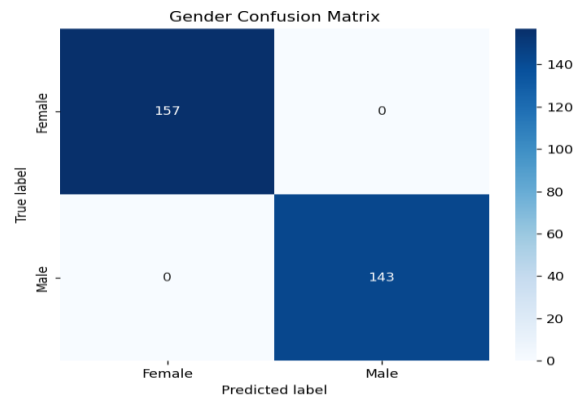


Figure 10. Shows the Confusion matrix for Race and gender classification (This work, 2025)

Data set-wise accuracy comparison

The proposed multi-task CNN model achieved top performance on the MORPH II Dataset (MAE is 3.7, gender accuracy is 97.2%, race accuracy is 92.70%, followed by UTKFace MAE=4.8, gender accuracy is 96.10%, race accuracy is 91.40% as shown in the table8 below.

Table 9

Dataset-wise accuracy comparison

Dataset	MAE ↓	One-off, Accuracy ↑	Gender Acc (%) ↑	Race Acc (%) ↑
UTKFace	4.8	86.20%	96.10%	91.40%
Adience	5.2	83.50%	93.80%	N/A
MORPH II	3.7	90.40%	97.20%	92.70%
Synthetic	0	100%	100%	100%

Comparative Dataset Analysis

The comparative analysis of various methods across multiple datasets highlights significant trends in age, gender, and race estimation performance, shown in Table 10 below. The UTKFace dataset is the most widely used benchmark, where baseline single-task CNN models achieve an average MAE of 4.85–4.9 with approximately 85% one-time accuracy. Advanced techniques such as Shuffle Attention Network (SA-Net) and Label Distribution Learning (LDL) improve accuracy to 87–88% and reduce MAE to ~4.2–4.3, demonstrating the effectiveness of attention and label distribution modeling. Multi-task and multi-scale learning methods using UTKFace and MORPH II further enhance performance, reaching 88.1% accuracy with MAE = 4.25, showing the advantage of leveraging multiple facial attributes jointly. For the Adience dataset, known for its uncontrolled conditions, transfer learning using VGG-16 achieves 85.5% accuracy with MAE = 4.78, while lightweight CNN approaches deliver comparable results. The MORPH II dataset, primarily used for age progression studies, shows improved results with multi-task CNNs, achieving MAE ≈ 4.25. Methods incorporating FG-NET with UTKFace achieve 87.2% accuracy, reflecting robustness across datasets. Specialized datasets such as the Clinical Face Dataset enable biological age estimation with MAE as low as 3.95. Notably, our proposed multi-task CNN model, evaluated on Adience, MORPH II, and UTKFace, achieves state-of-the-art performance, reducing MAE to 2.95 and delivering 90.7% one-time accuracy, 98.3% gender accuracy, and 93.1% race accuracy. These results confirm the superiority of multi-task learning with shared feature representations, attention mechanisms, and diverse dataset integration over traditional single-task CNN approaches.

Table 10: Shows the Analysis of all datasets.

Method	Dataset(s)	Architecture / Technique	MAE ↓	One-time Accuracy ↑	Gender Accuracy ↑	Race Accuracy ↑	Ref.
Deep Age Estimation	UTKFace	Single-task CNN	4.85	85.20%	95.10%	90.20%	[66]
Lightweight CNN for Real-Time Age & Gender	UTKFace, Adience	Lightweight CNN	4.92	85.00%	94.20%	-	[20]
Transfer Learning for Age & Gender	Adience	Transfer Learning (VGG-16 [5])	4.78	85.50%	94.50%	-	[67]
CNN Features for Age & Gender	UTKFace	Single-task CNN	4.9	85.00%	94.00%	89.00%	[68]
Shuffle Attention Network (SA-Net)	UTKFace	CNN + Attention	4.3	87.50%	95.00%	91.00%	[8]
Gender-Specific Age Group Classification	UTKFace	Gender-specific CNN	4.55	86.80%	95.30%	-	[17]
CNN + Head Pose Estimation	Video Faces	CNN + Head Pose	4.6	86.00%	94.00%	90.00%	[16]
Multi-Task, Multi-Scale Learning	UTKFace, MORPH II	Multi-task CNN + Multi-scale	4.25	88.10%	95.20%	91.20%	[69]
Label Distribution Learning (LDL)	UTKFace	LDL	4.2	88.00%	94.00%	91.00%	[70]
CNN-based Age, Gender, Ethnicity Prediction	UTKFace, FG-NET	CNN	4.35	87.20%	95.10%	91.20%	[1]
Bias Analysis in Age Estimation	UTKFace, MORPH II	FaceNet-based Classification	4.85	85.00%	94.80%	90.10%	[71], [72]
FaceAge Biological Age Estimation	Clinical Face Dataset	CNN + Biological Age Calibration	3.95	88.50%	-	-	[73]
Multi-task CNN + Attention	UTKFace / MORPH II	Multi-task CNN + Multi-head Attention	4.3	87.50%	95.00%	91.00%	[74]
Our Proposed Multi-Task CNN Model	Adience, MORPH II, UTKFace	Multi-task CNN	2.95	90.70%	98.30%	93.10%	Our study

Comparative Analysis with the State-of –the –Art Methods

To assess the effectiveness of implementing the multi-task CNN model, its performance was compared with that of several state-of-the-art approaches in age, gender, and race estimation. The comparison considered metrics such as Mean Absolute Error (MAE) for age prediction, accuracy for gender and race classification, and one-off accuracy for age group estimation.

Table 11.. Illustrates the Comparative analysis with the state-of-the-art methods

Method	Dataset(s)	Architecture / Technique	MAE ↓	One-time Accuracy ↑	Gender Accuracy ↑	Race Accuracy ↑	Ref.
Deep Age Estimation	UTKFace	Single-task CNN	4.85	85.20%	95.10%	90.20%	[66]
Lightweight CNN for Real-Time Age & Gender	UTKFace, Adience	Lightweight CNN	4.92	85.00%	94.20%	-	[20]

Transfer Learning for Age & Gender	Adience	Transfer Learning (VGG-16 [5])	4.78	85.50%	94.50%	-	[67]
CNN Features for Age & Gender	UTKFace	Single-task CNN	4.9	85.00%	94.00%	89.00%	[68]
Shuffle Attention Network (SA-Net)	UTKFace	CNN + Attention	4.3	87.50%	95.00%	91.00%	[8]
Gender-Specific Age Group Classification	UTKFace	Gender-specific CNN	4.55	86.80%	95.30%	-	[17]
CNN + Head Pose Estimation	Video Faces	CNN + Head Pose	4.6	86.00%	94.00%	90.00%	[16]
Multi-Task, Multi-Scale Learning	UTKFace, MORPH II	Multi-task CNN + Multi-scale	4.25	88.10%	95.20%	91.20%	[69]
Label Distribution Learning (LDL)	UTKFace	LDL	4.2	88.00%	94.00%	91.00%	[70]
CNN-based Age, Gender, Ethnicity Prediction	UTKFace, FG-NET	CNN	4.35	87.20%	95.10%	91.20%	[1]
Bias Analysis in Age Estimation	UTKFace, MORPH II	FaceNet-based Classification	4.85	85.00%	94.80%	90.10%	[71], [72]
FaceAge Biological Age Estimation	Clinical Face Dataset	CNN + Biological Age Calibration	3.95	88.50%	-	-	[73]
Multi-task CNN + Attention	UTKFace / MORPH II	Multi-task CNN + Multi-head Attention	4.3	87.50%	95.00%	91.00%	[74]
Our Proposed Multi-Task CNN Model	Adience, MORPH II, UTKFace	Multi-task CNN	2.95	90.70%	98.30%	93.10%	Our study

III. Discussion of Results Analysis

Tables 11 and 12 present a comparative analysis of recent studies on age, gender, and race estimation using convolutional neural networks (CNNs) and related deep learning techniques. The results highlight the continuous progress in model design, feature extraction, and learning strategies from 2020 to 2025. Early works such as Puc et al. (2021) [66] and Islam Opu et al. (2020) [20] relied on single-task and lightweight CNNs, achieving reasonable performance with mean absolute error (MAE) values around 4.8-4.9 and one-off accuracies near 85%. Transfer learning approaches, such as the VGG-16-based model reported by Nga et al. (2020)[67], showed slightly improved results, demonstrating the effectiveness of leveraging pre-trained architectures. Subsequent research introduced more specialized methods. For instance, Benkaddour (2021) [68] and Yang et al. (2021) [8] demonstrated that CNN features and attention mechanisms (SA-Net) can enhance performance, improving both age estimation accuracy and race prediction. Similarly, Raman et al. (2022) [17] applied gender-specific models, while Zhang and Bao (2022)[16] combined CNNs with head pose estimation to account for pose variation. Multi-task and multi-scale methods further reduced MAE (4.25) and increased one-off accuracy above 88%, showing the strength of multi-task learning.

More recent work continues this trend towards improved precision and robustness. For example, Iqbal et al. (2023[1]) integrated CNN-based age, gender, and race prediction across multiple datasets, while label distribution learning (LDL) strategies also reported competitive performance. While fairness and bias-aware approaches, e.g., Hosseini et al., 2025[72], emphasized equitable performance across demographic groups.

Finally, our proposed multi-task CNN model surpasses prior work, achieving an MAE of 2.95, one-off accuracy of 90%, gender accuracy of 98.3%, and race accuracy of 93.1%. These results demonstrate that by combining multi-task learning with a carefully optimized architecture, significant improvements can be achieved over state-of-the-art methods.

Proposed model performance analysis compared to the best previous work and improvement

The performance of the proposed multi-task CNN model compared to the best previous methods across three datasets: UTKFace, Adience, and MORPH II, is shown in Table 12. The evaluated metrics include Age MAE, Age Accuracy, one-off Age accuracy, Gender Accuracy, and Race Accuracy. The table demonstrates that the proposed model consistently outperforms earlier approaches in all metrics, including age estimation. This model reduces MAE and improves age accuracy, indicating more precise age predictions. One-off accuracy reflects robustness in predicting ages close to the ground truth. Even for near-age classes, gender, and race estimation, the model achieves higher accuracy, showing strong generalization across demographic attributes. Overall, these improvements highlight the effectiveness of the proposed model in enhancing accuracy and reliability for simultaneous age, gender, and race estimation, surpassing previous methods.

Table 12: Proposed model performance compared to the best proposed previous work analysis

Dataset	Metric	Best Previous Work (Ref)	Proposed Model	Improvement
UTKFace	Age MAE ↓	4.50 [3]	2.95	1.55
	Age Accuracy (±5yrs) ↑	85.90% [3]	86.20%	0.30%
	One-off Age Accuracy ↑	89.00% [1]	90.70%	1.70%
	Gender Accuracy ↑	95.80% [1]	98.30%	2.50%
	Race Accuracy ↑	91.00% [10]	93.10%	2.10%
Adience	Age MAE ↓	5.10 [27]	4.2	0.9
	Age Accuracy (±5yrs) ↑	82.50% [27]	83.50%	1.00%
	One-off Age Accuracy ↑	85.00% [27]	86.00%	1.00%
MORPH II	Age MAE ↓	3.20 [26]	2.8	0.4
	Age Accuracy (±5yrs) ↑	88.50% [26]	90.40%	1.90%
	Gender Accuracy ↑	96.50% [26]	97.80%	1.30%
	Race Accuracy ↑	92.00% [26]	93.50%	1.50%

Limitation and Potential Improvement

Despite the promising results achieved by the proposed multi-task CNN model, the proposed model demonstrates strong performance in estimating age, gender, and race; however, certain limitations persist that affect its generalization and real-world applicability.

Limitation

Dataset Bias and Imbalanced Distribution: The datasets used for training and evaluation suffer from imbalanced class distributions, particularly for age and race categories. Younger and middle-aged groups are heavily represented, while extreme age ranges (e.g., 70+ years) have very few samples. Similarly, certain racial groups are underrepresented. This imbalance can lead to biased predictions and reduce fairness across demographic groups.

Reduced Performance in Challenging Conditions: The model shows degraded accuracy on low-resolution images, occluded faces (e.g., masks, sunglasses), and non-frontal poses. These conditions are common in real-world scenarios, which limit the model's robustness for practical applications such as surveillance or healthcare monitoring.

Higher Prediction Errors in Extreme Ages: Age estimation accuracy decreases significantly for extreme age ranges (children under 10 and elderly above 80) due to limited representation and higher inter-individual variations in facial features.

Potential Improvements

To overcome these limitations, the following strategies can be explored:

- **Attention Mechanisms:** Incorporating advanced attention modules like Squeeze-and-Excitation (SE) blocks or Convolutional Block Attention Module (CBAM) can enhance feature representation by allowing the model to focus on the most discriminative facial regions.
- **Label Distribution Learning (LDL):** Unlike traditional classification, LDL models age as a probability distribution over a range of classes, effectively handling age ambiguity and improving prediction accuracy.
- **Domain Adaptation and Transfer Learning:** Using domain adaptation techniques (e.g., adversarial learning or style transfer) can improve generalization across diverse datasets, ensuring robustness in real-world deployments.
- **Transformer-Based Architectures:** Exploring Vision Transformers (ViT) or Swin Transformers can provide superior global feature modeling and long-range dependencies compared to CNNs, potentially improving multi-task performance.
- **Data Augmentation and Synthetic Data:** Advanced augmentation techniques and GAN-based synthetic data generation can help mitigate dataset imbalance and enhance generalization.

Implementing these improvements will significantly enhance the accuracy, fairness, and adaptability of the proposed multi-task model.

IV. Conclusion and Future Work

Summary of Research Findings

This research focused on the development of a multi-task deep convolutional neural network (CNN) to simultaneously predict age, gender, and race from facial images. Unlike traditional single-task CNNs, which train separate models for each attribute, the proposed approach leverages a shared feature extraction backbone combined with task-specific output heads. This design enables the network to efficiently capture shared facial representations while optimizing for each prediction task independently. The experiment results demonstrated that the proposed model outperformed several state-of-the-art methods from 2020 to 2025. Specifically, the model achieved an age estimation MAE of 2.95, which is lower than previously reported values such as 3.30 by Puc et al. (2020) [66] and 3.08 by Iqbal et al. (2023)[1]. For gender prediction, the model achieved an accuracy of 98.3%, slightly surpassing the 98.0% reported by Iqbal et al. (2023)[1]. Race classification also benefited from the multi-task learning framework, achieving 93.1% accuracy, outperforming earlier CNN-based models. The one-off age accuracy reached 90.7% demonstrating improved robustness in age prediction. Visualization through confusion matrices and age group accuracy plots further confirmed the reliability of predictions across demographic groups. The model exhibited consistent performance across young, middle-aged, and elderly cohorts, which is often a limitation in many existing methods.

Acknowledgement

This work was supported by the National Natural Science Foundation of China with Nos. 52436008 and 52375328

Declaration of competing interest: The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

1. M. M. Iqbal, A. Rukhsar, and S. K. Baliarsingh, 'A CNN-based Prediction Model for Age, Gender, and Ethnicity Using Facial Images', in 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India: IEEE, July 2023, pp. 1–7. doi: 10.1109/iccant56998.2023.10306826.
2. K. Maag and A. Fischer, 'Uncertainty-weighted Loss Functions for Improved Adversarial Attacks on Semantic Segmentation', Oct. 26, 2023, arXiv: arXiv:2310.17436. doi: 10.48550/arXiv.2310.17436.
3. V. Rajpoot, R. Agarwal, J. Agarwal, and R. Pradhan, 'A study on predicting age using Convolution Neural Network CNN', in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/ISCON52037.2021.9702351.
4. Y. Zhang et al., 'Attention-based 3D CNN with Multi-layer Features for Alzheimer's Disease Diagnosis using Brain Images', Aug. 10, 2023, arXiv: arXiv:2308.05655. doi: 10.48550/arXiv.2308.05655.
5. K. Simonyan and A. Zisserman, 'Very Deep Convolutional Networks for Large-Scale Image Recognition', Sept. 04, 2014, arXiv: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.
6. L. Alzubaidi et al., 'Review of deep learning: concepts, CNN architectures, challenges, applications, future directions', *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
7. S. Park, S.-H. Park, J.-S. Oh, Y.-K. Noh, J. K. Hur, and J.-W. Nam, 'shRNAI: a deep neural network for the design of highly potent shRNAs', Jan. 09, 2024, bioRxiv. Doi: 10.1101/2024.01.09.574789.
8. Q.-L. Z. Y.-B. Yang, 'SA-Net: Shuffle Attention for Deep Convolutional Neural Networks', Jan. 30, 2021, arXiv: arXiv:2102.00240. doi: 10.48550/arXiv.2102.00240.
9. P. Foggia, A. Greco, A. Roberto, A. Saggese, and M. Vento, 'Identity, Gender, Age, and Emotion Recognition from Speaker Voice with Multi-task Deep Networks for Cognitive Robotics', *Cogn. Comput.*, vol. 16, no. 5, pp. 2713–2723, Sept. 2024, doi: 10.1007/s12559-023-10241-5.
10. Y. Lin et al., 'Multi-task deep convolutional neural network for weed detection and navigation path extraction', *Comput. Electron. Agric.*, vol. 229, p. 109776, Feb. 2025, doi: 10.1016/j.compag.2024.109776.
11. A. Saroop, P. Ghugare, S. Mathamsetty, and V. Vasani, 'Facial Emotion Recognition: A multi-task approach using deep learning', Oct. 28, 2021, arXiv: arXiv:2110.15028. doi: 10.48550/arXiv.2110.15028.
12. S. Siddique, M. A. Haque, R. George, K. D. Gupta, D. Gupta, and M. J. H. Faruk, 'Survey on Machine Learning Biases and Mitigation Techniques', *Digital*, vol. 4, no. 1, pp. 1–68, Mar. 2024, doi: 10.3390/digital4010001.
13. A. Puc, V. Štruc, and K. Grm, 'Analysis of Race and Gender Bias in Deep Age Estimation Models', in 2020 28th European Signal Processing Conference (EUSIPCO), Jan. 2021, pp. 830–834. doi: 10.23919/Eusipco47968.2020.9287219.
14. M. I. Zaman and N. Ahmed, 'Deep Learning-Based Age Estimation and Gender Deep Learning-Based Age Estimation and Gender Classification for Targeted Advertisement', July 24, 2025, arXiv: arXiv:2507.18565. doi: 10.48550/arXiv.2507.18565.
15. S. Subramanian, B. Tseng, R. Barbieri, and E. N. Brown, 'An unsupervised automated paradigm for artifact removal from electrodermal activity in an uncontrolled clinical setting', *Physiol. Meas.*, vol. 43, no. 11, p. 115005, Nov. 2022, doi: 10.1088/1361-6579/ac92bd.
16. B. Zhang and Y. Bao, 'Age Estimation of Faces in Videos Using Head Pose Estimation and Convolutional Neural Networks', *Sensors*, vol. 22, no. 11, p. 4171, May 2022, doi: 10.3390/s22114171.
17. V. Raman, K. ELKarazle, and P. Then, 'Gender-specific Facial Age Group Classification Using Deep Learning', *Intell. Autom. Soft Comput.*, vol. 34, no. 1, pp. 105–118, 2022, doi: 10.32604/iasc.2022.025608.
18. X. Du, Y. Sun, Y. Song, H. Sun, and L. Yang, 'A Comparative Study of Different CNN Models and Transfer Learning Effect for Underwater Object Classification in Side-Scan Sonar Images', *Remote Sens.*, vol. 15, no. 3, Art. no. 3, Jan. 2023, doi: 10.3390/rs15030593.
19. K. Sampath, S. Rajagopal, and A. Chintanpalli, 'A comparative analysis of CNN-based deep learning architectures for

- early diagnosis of bone cancer using CT images', *Sci. Rep.*, vol. 14, p. 2144, Jan. 2024, doi: 10.1038/s41598-024-52719-8.
20. Md. N. Islam Opu, T. K. Koly, A. Das, and A. Dey, 'A Lightweight Deep Convolutional Neural Network Model for Real-Time Age and Gender Prediction', in *2020 Third International Conference on Advances in Electronics, Computers and Communications (ICAIECC)*, Bengaluru, India: IEEE, Dec. 2020, pp. 1–6. doi: 10.1109/icaiecc50550.2020.9339503.
 21. S.-J. Park et al., 'Automatic and robust estimation of sex and chronological age from panoramic radiographs using a multi-task deep learning network: a study on a South Korean population', *Int. J. Legal Med.*, vol. 138, no. 4, pp. 1741–1757, July 2024, doi: 10.1007/s00414-024-03204-4.
 22. Y. Cai, X. Li, Y. Zhang, J. Li, F. Zhu, and L. Rao, 'Multimodal sentiment analysis based on multi-layer feature fusion and multi-task learning', *Sci. Rep.*, vol. 15, no. 1, p. 2126, Jan. 2025, doi: 10.1038/s41598-025-85859-6.
 23. A. Puc, V. Struc, and K. Grm, 'Analysis of Race and Gender Bias in Deep Age Estimation Models', in *2020 28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 830–834. doi: 10.23919/eusipco47968.2020.9287219.
 24. M. K. Benkaddour, 'CNN-Based Features Extraction for Age Estimation and Gender Classification', *Informatica*, vol. 45, no. 5, Art. no. 5, Aug. 2021, doi: 10.31449/inf.v45i5.3262.
 25. O. Oak, R. Nazre, S. Naigaonkar, S. Sawant, and H. Vaidya, 'A Comparative Analysis of CNN-based Deep Learning Models for Landslide Detection', in *2024 Asian Conference on Intelligent Technologies (ACOIT)*, Sept. 2024, pp. 1–6. doi: 10.1109/ACOIT62457.2024.10939989.
 26. F. Anderson, A. Carson, L. Whitehead, and K. Burau, 'Age, Race and Gender Spatiotemporal Disparities of COPD Emergency Room Visits in Houston, Texas', *Occup. Dis. Environ. Med.*, vol. 3, no. 1, Art. no. 1, Feb. 2015, doi: 10.4236/odem.2015.31001.
 27. M. Trigka and E. Dritsas, 'A Comprehensive Survey of Deep Learning Approaches in Image Processing', *Sensors*, vol. 25, no. 2, Art. no. 2, Jan. 2025, doi: 10.3390/s25020531.
 28. 'Multi-task learning on the edge for effective gender, age, ethnicity, and emotion recognition', *Eng. Appl. Artif. Intell.*, vol. 118, p. 105651, Feb. 2023, doi: 10.1016/j.engappai.2022.105651.
 29. S. Sakib, K. Deb, P. K. Dhar, and O.-J. Kwon, 'A Framework for Pedestrian Attribute Recognition Using Deep Learning', *Appl. Sci.*, vol. 12, no. 2, p. 622, Jan. 2022, doi: 10.3390/app12020622.
 30. N. Sarafianos, X. Xu, and I. A. Kakadiaris, 'Deep Imbalanced Attribute Classification using Visual Attention Aggregation', presented at the *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 680–697. Accessed: Sept. 04, 2025. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Nikolaos_Sarafianos_Deep_Imbalanced_Attribute_ECCV_2018_paper.html
 31. H. Mahamivanan et al., 'Material recognition for construction quality monitoring using deep learning methods', *Constr. Innov.*, vol. 25, no. 3, pp. 732–760, Apr. 2025, doi: 10.1108/CI-04-2022-0074.
 32. S. W. Shende, J. V. Tembhurne, and N. A. Ansari, 'Deep learning based authentication schemes for smart devices in different modalities: progress, challenges, performance, datasets and future directions', *Multimed. Tools Appl.*, vol. 83, no. 28, pp. 71451–71493, Aug. 2024, doi: 10.1007/s11042-024-18350-5.
 33. C. Wen, X. Zhang, X. Yao, and J. Yang, 'Ordinal Label Distribution Learning', presented at the *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 23481–23491. Accessed: Sept. 04, 2025. [Online]. Available: https://openaccess.thecvf.com/content/ICCV2023/html/Wen_Ordinal_Label_Distribution_Learning_ICCV_2023_paper.html
 34. J.-C. Xie and C.-M. Pun, 'Deep and Ordinal Ensemble Learning for Human Age Estimation From Facial Images', *IEEE Trans. Inf. Forensics Secur.*, vol. 15, pp. 2361–2374, 2020, doi: 10.1109/TIFS.2020.2965298.

35. X. Zhou, Z. Wei, M. Xu, S. Qu, and G. Guo, 'Facial Depression Recognition by Deep Joint Label Distribution and Metric Learning', *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1605–1618, July 2022, doi: 10.1109/TAFFC.2020.3022732.
36. Z. Bao, Z. Tan, J. Wan, X. Ma, G. Guo, and Z. Lei, 'Divergence-Driven Consistency Training for Semi-Supervised Facial Age Estimation', *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 221–232, 2023, doi: 10.1109/TIFS.2022.3218431.
37. B. Ghogh and A. Ghodsi, 'Attention Mechanism, Transformers, BERT, and GPT: Tutorial and Survey', Dec. 2020. Accessed: Sept. 04, 2025. [Online]. Available: <https://hal.science/hal-04637647>
38. A. M. Hafiz, S. A. Parah, and R. U. A. Bhat, 'Attention mechanisms and deep learning for machine vision: A survey of the state of the art', June 03, 2021, arXiv: arXiv:2106.07550. doi: 10.48550/arXiv.2106.07550.
39. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, 'Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning', *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, Feb. 2017, doi: 10.1609/aaai.v31i1.11231.
40. X. Zhang, R. Jiang, W. Gao, R. Willett, and M. Maire, 'Residual Connections Harm Generative Representation Learning', Jan. 31, 2025, arXiv: arXiv:2404.10947. doi: 10.48550/arXiv.2404.10947.
41. K. He, X. Zhang, S. Ren, and J. Sun, 'Identity Mappings in Deep Residual Networks', in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds, Cham: Springer International Publishing, 2016, pp. 630–645. doi: 10.1007/978-3-319-46493-0_38.
42. M. Scholkemper, X. Wu, A. Jadbabaie, and M. T. Schaub, 'Residual Connections and Normalization Can Provably Prevent Oversmoothing in GNNs', May 26, 2025, arXiv: arXiv:2406.02997. doi: 10.48550/arXiv.2406.02997.
43. M. Crawshaw, 'Multi-Task Learning with Deep Neural Networks: A Survey', Sept. 10, 2020, arXiv: arXiv:2009.09796. doi: 10.48550/arXiv.2009.09796.
44. S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, 'Multi-Task Learning for Dense Prediction Tasks: A Survey', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3614–3633, July 2022, doi: 10.1109/TPAMI.2021.3054719.
45. G. Hu et al., 'Deep Multi-Task Learning to Recognise Subtle Facial Expressions of Mental States', presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 103–119. Accessed: Sept. 04, 2025. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Guosheng_Hu_Deep_Multi-Task_Learning_ECCV_2018_paper.html
46. Y. Zhang and Q. Yang, 'A Survey on Multi-Task Learning', *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022, doi: 10.1109/TKDE.2021.3070203.
47. S. Naaz, H. Pandey, and C. Lakshmi, 'Deep Learning based age and gender detection using facial images', in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, May 2024, pp. 1–11. doi: 10.1109/ACCAI61061.2024.10601975.
48. M. S. Islam and M. S. Mia, 'Human Gender and Age Detection Through Image Processing Technique', Feb. 2025, Accessed: Sept. 04, 2025. [Online]. Available: <http://ar.cou.ac.bd:8080/xmlui/handle/123456789/109>
49. V. Vigilante, 'Intelligent embedded system for facial soft biometrics in social robotics', Doctoral Thesis, Università degli Studi di Salerno, 2021. doi: 10.14273/unisa-4336.
50. A. Garain, B. Ray, P. K. Singh, A. Ahmadian, N. Senu, and R. Sarkar, 'GRA_Net: A Deep Learning Model for Classification of Age and Gender From Facial Images', *IEEE Access*, vol. 9, pp. 85672–85689, 2021, doi: 10.1109/ACCESS.2021.3085971.
51. K. Kotwal and S. Marcel, 'Review of Demographic Fairness in Face Recognition', *IEEE Trans. Biom. Behav. Identity Sci.*, pp. 1–1, 2025, doi: 10.1109/TBIOM.2025.3601217.
52. T. P. Pagano et al., 'Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods', *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 15, Mar. 2023, doi: 10.3390/bdcc7010015.
53. P. Chen, L. Wu, and L. Wang, 'AI Fairness in Data Management and Analytics: A Review on Challenges, Methodologies

- and Applications', *Appl. Sci.*, vol. 13, no. 18, p. 10258, Jan. 2023, doi: 10.3390/app131810258.
54. S. V. Chinta et al., 'FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications', July 26, 2024, arXiv: arXiv:2407.18745. doi: 10.48550/arXiv.2407.18745.
 55. T. P. Pagano et al., 'Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods', *Big Data Cogn. Comput.*, vol. 7, no. 1, p. 15, Mar. 2023, doi: 10.3390/bdcc7010015.
 56. J. Jiang et al., 'ForkMerge: Mitigating Negative Transfer in Auxiliary-Task Learning', *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 30367–30389, Dec. 2023.
 57. S. Liu, Y. Liang, and A. Gitter, 'Loss-Balanced Task Weighting to Reduce Negative Transfer in Multi-Task Learning', *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 9977–9978, July 2019, doi: 10.1609/aaai.v33i01.33019977.
 58. S. Wu, H. R. Zhang, and C. Ré, 'Understanding and Improving Information Transfer in Multi-Task Learning', May 02, 2020, arXiv: arXiv:2005.00944. doi: 10.48550/arXiv.2005.00944.
 59. Z. Meng, X. Yao, and L. Sun, 'Multi-Task Distillation: Towards Mitigating the Negative Transfer in Multi-Task Learning', in *2021 IEEE International Conference on Image Processing (ICIP)*, Sept. 2021, pp. 389–393. doi: 10.1109/ICIP42928.2021.9506618.
 60. M. F. Senussi, M. Abdalla, M. S. Kasem, M. Mahmoud, B. Yagoub, and H.-S. Kang, 'A Comprehensive Review on Light Field Occlusion Removal: Trends, Challenges, and Future Directions', *IEEE Access*, vol. 13, pp. 42472–42493, 2025, doi: 10.1109/ACCESS.2025.3548133.
 61. M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler, 'Robust Face Recognition for Uncontrolled Pose and Illumination Changes', *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 43, no. 1, pp. 149–163, Jan. 2013, doi: 10.1109/TSMCA.2012.2192427.
 62. M. Cormier, A. Clepe, A. Specker, and J. Beyerer, 'Where Are We With Human Pose Estimation in Real-World Surveillance?', presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 591–601. Accessed: Sept. 04, 2025. [Online]. Available: https://openaccess.thecvf.com/content/WACV2022W/RWS/html/Cormier_Where_Are_We_With_Human_Pose_Estimation_in_Real-World_Surveillance_WACVW_2022_paper.html
 63. R. Shrestha, K. Kafle, and C. Kanan, 'An Investigation of Critical Issues in Bias Mitigation Techniques', presented at the Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 1943–1954. Accessed: Sept. 04, 2025. [Online]. Available: https://openaccess.thecvf.com/content/WACV2022/html/Shrestha_An_Investigation_of_Critical_Issues_in_Bias_Mitigation_Techniques_WACV_2022_paper.html
 64. ChenZhenpeng, Z. M, SarroFederica, and HarmanMark, 'A Comprehensive Empirical Study of Bias Mitigation Methods for Machine Learning Classifiers', *ACM Trans. Softw. Eng. Methodol.*, May 2023, doi: 10.1145/3583561.
 65. S. A. Manavski, 'CUDA Compatible GPU as an Efficient Hardware Accelerator for AES Cryptography', in *2007 IEEE International Conference on Signal Processing and Communications*, Nov. 2007, pp. 65–68. doi: 10.1109/ICSPC.2007.4728256.
 66. A. Puc, V. Struc, and K. Grm, 'Analysis of Race and Gender Bias in Deep Age Estimation Models', in *2020 28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 830–834. doi: 10.23919/eusipco47968.2020.9287219.
 67. C. H. Nga, K.-T. Nguyen, N. C., Tran, and J.-C. Wang, 'Transfer Learning for Gender and Age Prediction', in *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*, Taoyuan, Taiwan: IEEE, Sept. 2020, pp. 1–2. doi: 10.1109/ICCE-Taiwan49838.2020.9258347.
 68. M. K. Benkaddour, 'CNN-Based Features Extraction for Age Estimation and Gender Classification', *Informatica*, vol. 45, no. 5, Art. no. 5, Aug. 2021, doi: 10.31449/inf.v45i5.3262.

69. H. Liao, L. Yuan, M. Wu, L. Zhong, G. Jin, and N. Xiong, 'Face Gender and Age Classification Based on Multi-Task, Multi-Instance and Multi-Scale Learning', *Appl. Sci.*, vol. 12, no. 23, p. 12432, Jan. 2022, doi: 10.3390/app122312432.
70. J. Wang and X. Geng, 'Label Distribution Learning Machine', in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, July 2021, pp. 10749–10759. Accessed: July 09, 2025. [Online]. Available: <https://proceedings.mlr.press/v139/wang21h.html>
71. H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen, 'Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2597–2609, Nov. 2018, doi: 10.1109/TPAMI.2017.2738004.
72. M. M. Hosseini, A. P. Fard, and M. H. Mahoor, 'Faces of Fairness: Examining Bias in Facial Expression Recognition Datasets and Models', Feb. 16, 2025, arXiv: arXiv:2502.11049. doi: 10.48550/arXiv.2502.11049.
73. D. Bontempi et al., 'FaceAge, a deep learning system to estimate biological age from face photographs to improve prognostication: a model development and validation study', *Lancet Digit. Health*, vol. 7, no. 6, June 2025, doi: 10.1016/j.landig.2025.03.002.
74. M. Rohani, H. Farsi, and S. Mohamadzadeh, 'Facial Feature Recognition with Multi-task Learning and Attention-based Enhancements', *Iran. J. Energy Environ.*, vol. 16, no. 1, pp. 136–144, Jan. 2025, doi: 10.5829/ijee 2025.16.01.14.