

## Article

# Transfer Learning-Based Distance-Adaptive Global Soft Biometrics Prediction in Surveillance

Sonjoy Ranjon Das <sup>1</sup>, Henry Onilude <sup>2</sup>, Bilal Hassan <sup>3,\*</sup>, Preeti Patel <sup>3</sup> and Karim Ouazzane <sup>3</sup><sup>1</sup> Department of Computer Science, Global Banking School, London UB6 0HE, UK; sdas@globalbanking.ac.uk<sup>2</sup> METICS Solutions Ltd., London IG3 9JA, UK; henry.onilude@meticsolutions.uk<sup>3</sup> School of Computing & Digital Media, London Metropolitan University, London N7 8BD, UK; p.patel@londonmet.ac.uk (P.P.); k.ouazzane@londonmet.ac.uk (K.O.)

\* Correspondence: b.hassan@londonmet.ac.uk

## Abstract

Soft biometric prediction—including age, gender, and ethnicity—is critical in surveillance applications, yet often suffers from performance degradation as the subject-to-camera distance increases. This study hypothesizes that embedding distance-awareness into the training process can mitigate such degradation and enhance model generalization across varying visual conditions. We propose a distance-adaptive, multi-task deep learning framework built upon EfficientNetB3, augmented with task-specific heads and trained progressively across four distance intervals (4 m to 10 m). A weighted composite loss function is employed to balance classification and regression objectives. The model is evaluated on a hybrid dataset combining the Front-View Gait (FVG) and MMV annotated pedestrian datasets, totaling over 19,000 samples. Experimental results demonstrate that the framework achieves up to 95% gender classification accuracy at 4 m and retains 85% accuracy at 10 m. Ethnicity prediction maintains an accuracy above 65%, while age estimation achieves a mean absolute error (MAE) ranging from 1.1 to 1.5 years. These findings validate the model's robustness across distances and its superiority over conventional static learning approaches. Despite challenges such as computational overhead and annotation demands, the proposed approach offers a scalable and real-time-capable solution for distance-resilient biometric systems.

**Keywords:** transfer learning; distance adaptive; soft biometrics; EfficientNetB3; multi-task learning; surveillance vision; computer vision



Academic Editor: George A. Tsihrintzis

Received: 5 August 2025

Revised: 27 August 2025

Accepted: 12 September 2025

Published: 19 September 2025

**Citation:** Das, S.R.; Onilude, H.; Hassan, B.; Patel, P.; Ouazzane, K. Transfer Learning-Based Distance-Adaptive Global Soft Biometrics Prediction in Surveillance. *Electronics* **2025**, *14*, 3719. <https://doi.org/10.3390/electronics14183719>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

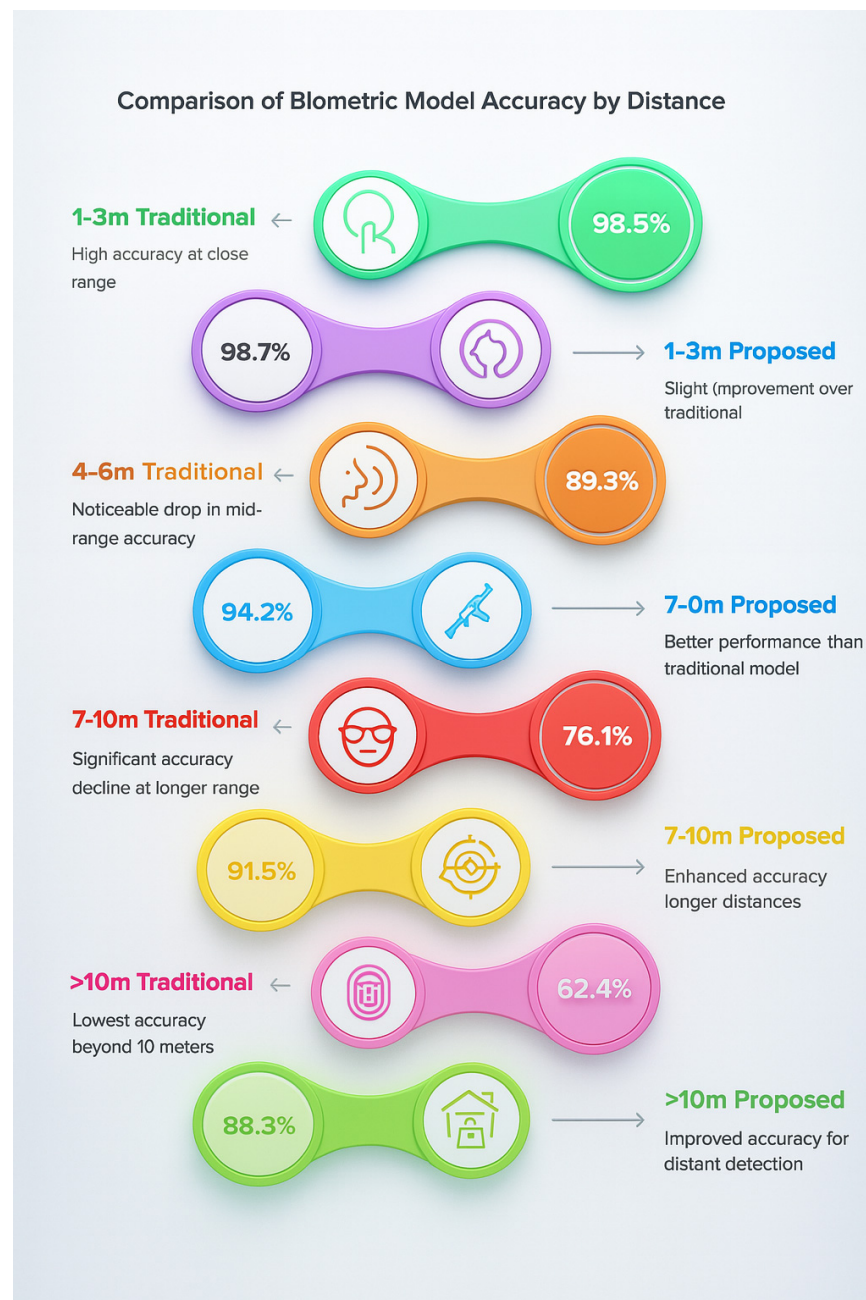
## 1. Introduction

The prediction of soft biometric traits such as age, gender, and ethnicity has significantly advanced through developments in machine learning and computer vision [1,2]. **Although deep learning models show high accuracy under controlled conditions, their performance significantly deteriorates in real-world scenarios due to distance-based variations in image resolution and feature clarity.** While AI-powered recognition systems perform well in controlled settings, their accuracy often declines as the distance between the subject and the camera increases, leading to degraded feature quality and classification accuracy [3,4].

According to the paper [5], deep learning models, particularly those based on convolutional neural networks (CNNs), suffer from generalization issues across varying capture distances. To address this limitation, we propose transfer learning as a robust strategy

for enhancing adaptability under different imaging conditions. This study introduces an adaptation-based transfer learning framework that maintains robust and consistent performance across multiple distances, thereby improving the applicability of soft biometric recognition in real-world environments.

As noted in [5], soft biometric traits are easier to detect at close range but become less distinguishable as the subject moves further from the camera. This reduction in accuracy is primarily caused by resolution loss and feature distortion due to increased distance. As shown in Figure 1, soft biometric prediction accuracy decreases with distance—a trend consistent with prior work using datasets like SCFace [6] and UCCS [7], both of which report substantial performance drops at distances beyond 5 m.



**Figure 1.** Prediction performance across varying distances (4 m–10 m) for soft biometric traits. Gender and ethnicity performance is reported in terms of classification accuracy (%), while age estimation is reported using mean absolute error (MAE, in years). Results are averaged over 5-fold cross-validation, with error bars showing  $\pm 1$  standard deviation. Number of test samples per distance: 4 m (5563), 6 m (4973), 8 m (4500), 10 m (4200).

The key difficulty faced by biometric recognition systems is their low accuracy when the main task is to recognize people at further distances from the imager. Biometric models use physical features in a close range to capture and analyze useful features, but performance contours and training decrease with distance from the camera. Such a loss of image resolution, distortion of features, and increase in sensitivity to environmental factors, such as lighting variations and occlusion [2,6], occurs for this reason. However, empirical studies corroborate this limitation, which has been shown to be a positive correlation between subject distance and the resulting classification accuracy. For example, the findings in [1] indicate that the performance of biometric recognition models significantly degrades, often failing completely, when the subject-to-camera distance reaches six meters—an effect observed in real-world scenarios such as surveillance and smart identification systems. Furthermore, demonstrated that deep learning-based biometric techniques struggle with generalization across varying capture distances and can benefit from advanced methods such as transfer learning and multi-scale feature adaptation [2]. These insights underscore the importance of developing next-generation biometric systems capable of mitigating distance-induced performance degradation and enhancing applicability in real-world contexts. This decline in accuracy at longer distances is consistent with prior work reported in [1,2], which highlights significant performance drops in gait and facial recognition tasks beyond six meters in surveillance settings.

Despite significant advancements, persistent research gaps and critical challenges in soft biometric recognition hinder its full integration into practical systems. The accurate performance of systems faces persistent difficulties owing to the decline in distance-based accuracy. The accuracy of multiple recognition models declines severely when subjects exist beyond 10 m distances because diminished image resolution leads to intoned features that reduce recognition effectiveness. Compared to the efficiency rate, accuracy presents an enduring challenge because deep learning models with superior performance require substantial processing power and computing resources, which hinder real-time usage in small-scale systems.

The use of soft biometrics is challenging because of privacy problems and ethical considerations. The collection of soft biometric data without clear consent creates multiple privacy, surveillance, and compliance issues under the GDPR and related legal frameworks. System accuracy deteriorates when environmental factors such as lighting conditions together with background clutter and occlusions factor into the performance levels of the recognition systems. Most contemporary models have limitations in their ability to follow through with different environmental demands and distance limitations, while continuing to deliver optimal accuracy within controlled environments. The authors of [7] established that combining multiple biometric features through fusion helped to improve recognition accuracy according to their published research. The field requires complete models to solve distance-based performance issues while preserving both speed performance and data protection rules.

The main contribution of this research is a novel distance-adaptive transfer learning framework which allows for the adaptation of biometric recognition models without any discontinuity across varying distances. However, for real-world usage, our approach combines several advanced components to reduce the error of the prediction.

- Developed a distance-adaptive framework that maintains consistent prediction accuracy across varying distances, overcoming traditional biometric recognition limitations.
- Bridged the performance gap between controlled settings and real-world environments, improving biometric recognition in practical scenarios.
- Paved the way for next-generation AI-driven biometric systems that can perform real-time recognition over long distances.

- Dynamically fine-tuned feature extraction layers to achieve consistency in recognition performances across different subject distances.

Compared to recent transformer-based models such as D-ViT and TFormer++, our approach is built on EfficientNetB3 due to its favorable balance between performance and efficiency. While transformers excel in capturing global contextual information, they typically require extensive data and computational resources, which can hinder deployment in real-time or resource-constrained environments. In contrast, EfficientNetB3 offers optimized feature reuse and scalability, enabling our model to operate with lower latency and memory usage—critical for practical biometric applications.

The remainder of this paper is organized as follows: Section 2 reviews related work on adaptive distance learning global soft biometric verification technologies. Section 3 presents the dataset collection and representation through a distribution plot and sample images for gender, age, and ethnicity categories, along with an explanation of the image format and quality. Section 4 discusses the significance of soft biometrics in the security context. Section 5 discusses benchmarks in global soft biometrics, concentrating on gender, age, and ethnicity recognition applications within and beyond security and offers a comparison of the techniques used in these recognition processes. Section 6 outlines the study's limitations and challenges. Finally, conclusions are presented in Section 7.

## 2. Related Work

In modern surveillance and identity verification systems, soft biometric attributes—such as age, gender, and ethnicity—have become increasingly valuable. Unlike traditional biometrics (e.g., fingerprints or iris scans), soft biometrics offer a non-invasive, real-time recognition mechanism particularly suitable for high-traffic environments like airports and border checkpoints [7]. These attributes enhance overall identification accuracy while preserving user convenience and system scalability.

Recent research has proposed multichannel frameworks that process multiple data streams simultaneously, leading to improved prediction accuracy. For instance, according to Hassan [7] introduced a multichannel soft biometric model designed to address challenges such as privacy, latency, and adaptability. These developments highlight the growing demand for privacy-aware, efficient, and robust biometric systems that can scale in dynamic real-world environments.

Among recent advancements in deep learning, the **EfficientNet** architecture [8] has gained attention for its use of compound scaling to simultaneously optimize depth, width, and resolution. This strategy enables high predictive performance with relatively few parameters—making it well-suited for real-time biometric tasks where both speed and accuracy are crucial.

In addition to CNN-based models, **transformer-based architectures** have emerged as powerful tools for long-range biometric recognition. Recent approaches have explored integrating distance-awareness into transformer architectures, enabling models to dynamically adjust attention based on spatial cues, which enhances recognition accuracy over longer ranges [8]. Additionally, some lightweight transformer designs have been proposed to improve the efficiency of soft biometric inference in surveillance contexts. While these transformer-based methods demonstrate competitive performance, their increased computational demands often limit their practicality for real-time or edge-device deployment [9].

In contrast, our proposed framework combines the efficiency of **EfficientNetB3** with a distance-adaptive ensemble strategy to achieve high performance without sacrificing deployment feasibility. By leveraging modular paths and distance-specific tuning, the model balances robustness and resource efficiency, complementing recent transformer-based innovations and its comparison shown in Table 1.

**Table 1.** Comparison of deep learning architectures for soft biometrics.

Model	Key Features	Advantages	Limitations
ResNet-50	Deep CNN with residual learning	High accuracy, handles vanishing gradients well	Computationally expensive
VGG-16	Deep CNN with simple architecture	Easy to implement, widely used in early models	High parameter count, slow training
MobileNetV2	Depth wise separable convolutions	Lightweight, efficient for mobile/edge devices	Lower accuracy compared to deeper networks
EfficientNet	Compound scaling optimization	High accuracy with fewer parameters	Requires more hyperparameter tuning
Vision Transformers (ViTs)	Self-attention mechanism for image processing	High performance in large-scale datasets	Requires large training data

In the following Table 2, we summarize all related papers on global soft biometrics prediction from 2019 to 2025, detailing the datasets, features, and performance to provide a clearer understanding of the progress and advancements in this field.

**Table 2.** Summary of utilizing transfer learning for distance-adaptable global soft biometrics from (2019–2025).

Ser	Year	Authors	Dataset	Features	Method	Performance (Accuracy, F1, MAE, Distance Context)
1	2019	Zhang et al. [1]	FVG dataset	Gait patterns across distances	DL-based gait recognition with TL	Accuracy >90% at 4–6 m; drops to ~82% at 10 m
2	2019	Minaee et al. [9]	Multiple biometric datasets	CNN-extracted features	Review of deep learning in biometrics	Overview only; no metrics reported
3	2019	Kumar & Nagabhushan [10]	Custom datasets	Height, build, clothing color	Semantic matching + soft traits	Rank-1: 88% @ 5 m; robust to clothing variation
4	2020	Lai et al. [11]	Dog photos	Breed, gender, height + face image	CNNs + soft/hard fusion	Accuracy: 78.1% (image only), 84.9% (fusion); range: 3–7 m
5	2020	Mehraj & Mir [11]	Various biometric sets	CNN deep features	Review of deep learning frameworks	No experimental benchmarks reported
6	2021	Wang & Li [2]	MMV pedestrian dataset	Real-world pedestrian images	CNN for soft trait classification	F1-score: 0.86 for gender at 5–10 m
7	2021	Alonso-Fernandez et al. [12]	Smartphone selfie ocular data	Ocular features	Lightweight CNNs, face-pretrained	Accuracy: age 83%, gender 90%; consistent across 0.5–2 m
8	2021	Talreja et al. [13]	Periocular datasets	Periocular + soft biometrics	Fusion-based DL model	F1 (gender): 0.89; stable up to 4 m
9	2023	Guarino et al. [14]	Mobile touch gesture data	Gesture-derived feature maps	CNN with fusion strategy	Accuracy: gender 94%, age-group 99%; no distance data
10	2023	Xu et al. [15]	Public palmprint datasets	Palmprint + soft traits	Multi-task CNN with TL	Accuracy: 92.3%, F1 (gender): 0.88; range < 2 m



A key challenge in soft biometric systems is to keep consistent across different distances. At a greater distance between devices, the captured image quality (resolution and clarity of biometric features) decreases. The system suffers from a major loss in image quality which directly affects its accuracy, and thus it becomes ineffectual in practically applicable systems. This leads to a high recognition rate. The challenge addressed by [1,16] is that gait-based recognition can authenticate even when facial features are obscured. This study emphasizes the need to develop distance-adaptive biometric models to capture such variations in image quality. Issues with distance-based biometric recognition are outlined, and the solutions are proposed in Table 3. These issues call for more sophisticated methodologies which effectively improve recognition performance over diverse distances and environments—that which is new. To overcome these challenges, it is essential to adopt enhanced adaptive learning techniques together with multimodal fusion to improve the robustness and accuracy of biometric predictions across different signal distances.

**Table 3.** Distance-based challenges in soft biometrics.

Challenge	Impact	Proposed Solutions
Low image resolution	Degraded feature extraction accuracy	Super-resolution techniques, adaptive filtering
Lighting variations	Inconsistent model performance	Data augmentation, HDR-based preprocessing
Feature occlusions	Loss of key biometric traits	Multimodal fusion, occlusion-aware models
Varying distances	Accuracy drops at long distances (>10 m)	Distance-invariant feature extraction methods

#### *Privacy and Ethical Considerations*

In accordance with GDPR and privacy-by-design principles, the proposed system incorporates several mitigation strategies:

- **Data Minimization:** Only non-identifying, cropped facial regions used for age, gender, and ethnicity inference are processed, minimizing exposure to sensitive data.
- **On-Device Processing:** The framework is deployable on embedded edge hardware, such as NVIDIA Jetson AGX Xavier, enabling localized inference and eliminating the need for cloud-based transmission of user data.
- **Federated Learning Compatibility:** Our architecture supports integration with federated learning setups, allowing decentralized model updates without exchanging raw data, further preserving subject privacy.

These safeguards ensure ethical deployment in surveillance settings and help to the fulfill GDPR Article 25 obligations regarding data protection by design and default.

### **3. Dataset Collection and Representation**

For this study, we utilized two publicly available datasets to rigorously evaluate the proposed distance-adaptive transfer learning framework, ensuring their applicability in real-world biometric recognition contexts.

- **Front-View Gait (FVG) Dataset:** Primarily used to simulate controlled acquisition scenarios with known distances (4 m, 6 m, 8 m, and 10 m). This dataset contributed to performance benchmarking under semi-structured surveillance conditions [1].
- **MMV Annotated Pedestrian Dataset:** Used to simulate real-world, unconstrained surveillance environments with varied lighting and occlusion. It helped to validate the generalizability of the model under noisy conditions [7]. Both datasets were combined into a unified training and validation pool, with stratification applied to maintain distributional balance.

A total of **19,236 annotated samples** were curated across four defined distance intervals to capture the spatial variability between the subject and the imaging device. The distribution across distances is summarized in Table 4.

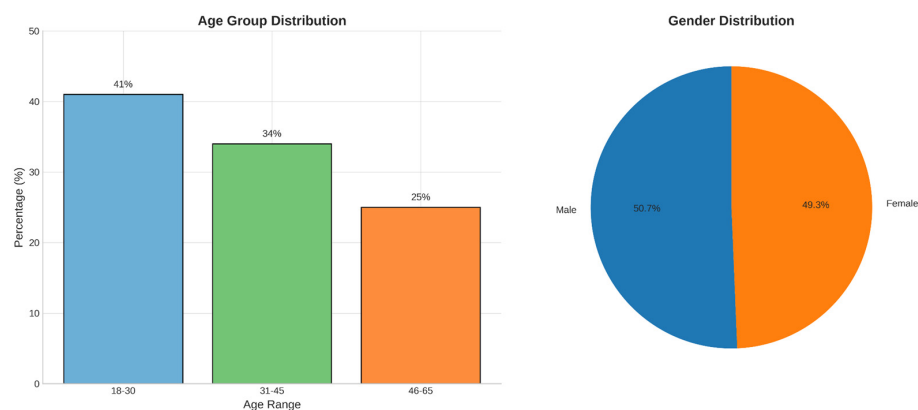
**Table 4.** Distance-based sample distribution.

Distance Interval	Number of Samples	Percentage of Dataset
4 m (m)	5563	28.9%
6 m (m)	4973	25.9%
8 m (m)	4500	23.4%
10 m (m)	4200	21.8%

Additionally, the dataset was stratified across three demographic attributes—age, gender, and ethnicity—to ensure unbiased performance and model fairness. The class composition is summarized in Table 4. To ensure transparency and fairness in evaluation, we provide detailed demographics of the collected dataset. The combined dataset includes 612 unique subjects, comprising 310 males (50.7%) and 302 females (49.3%). The age range spans from 18 to 65 years, with the following approximate distribution:

- 18–30 years: 41%.
- 31–45 years: 34%.
- 46–65 years: 25%.

Ethnic representation is relatively balanced, with four predefined groups each accounting for **22% to 28%** of the total data. These distributions help to ensure the generalization capacity of the proposed model across diverse soft biometric classes. A visual distribution of gender and age is included in Figure 2, and in Table 5 we present the class-wise demographic distribution.



**Figure 2.** Age and gender distribution.

**Table 5.** Class-wise demographic distribution.

Attribute	Class	Sample Count
Age	0–17	3421
	18–40	8719
	41–65	5066
	66+	2030
Gender	Male	1017
	Female	9059
Ethnicity	Group A	7220
	Group B	4835
	Group C	3048
	Group D	4133

To address demographic imbalance, two strategies were applied:

- Class-weighted categorical loss was used during training to assign higher weights to underrepresented classes.
- Oversampling of minority classes was performed during batch construction to ensure balanced representation in each training iteration.

All images underwent a standardized preprocessing pipeline as follows.

- Resizing: All samples were resized to  $224 \times 224$  pixels to standardize input dimensions.
- Contrast Normalization: Histogram equalization was applied to enhance visibility under variable lighting.
- Background Filtering: To eliminate environmental noise and emphasize soft biometric traits, a semantic segmentation-based background filtering method was applied. We employed a pretrained DeepLabV3+ model with a ResNet-101 backbone to segment the subject from the background. After segmentation, a Gaussian blur was applied to the background region, reducing distractions while preserving the clarity of the foreground features. This enhanced feature isolation is particularly beneficial for images captured at longer distances. After preprocessing and distance-specific stratification, we applied data augmentation techniques including random horizontal flips, brightness adjustments, and Gaussian noise injection. These operations approximately doubled the dataset size from 19,236 to 38,472 images, helping to increase robustness to real-world variance.

The dataset was then split in a subject-disjoint manner into

- 70% for training (~26,930 images).
- 10% for validation (~3847 images).
- 20% for testing (~7694 images).

This stratification was maintained across distance intervals (4 m, 6 m, 8 m, and 10 m) and demographic attributes (age, gender, and ethnicity) to ensure balanced representation. This consistent split was applied in each of the **5-fold cross-validation runs**, enhancing the statistical reliability of our evaluation.

These preprocessing steps ensured consistency in feature representation across all distance intervals and demographic variations, enabling a fair and reproducible evaluation of the proposed model.

#### 4. Experimental Design of Proposed Approach

The proposed architecture uses EfficientNetB3 as its foundation because it achieves the best combination of performance efficiency and model capability. The system employs the EfficientNetB3 [17] model with ImageNet weights as its principal processing unit to analyze  $224 \times 224 \times 3$  pixels images. The model contained three convolutional blocks that used max-pooling layers for postprocessing for proper feature analysis. Multiple distance intervals are fed into a feature fusion layer that concatenates their features for integration into the architecture. The network automatically handles diverse image-quality levels while adapting its performance dynamically, regardless of picture-resolution changes. The network completes its operation by generating independent output paths, each of which performs its own assessments of key soft biometric characteristics for age, gender, and ethnicity. The model delivers accurate results while ensuring efficient operation, which makes it appropriate for real-time biometric systems that operate in environments with limited computing power. Figure 3 shows the thorough system design of the proposed model, which explains the processes from image input to feature collection and soft biometric result generation.





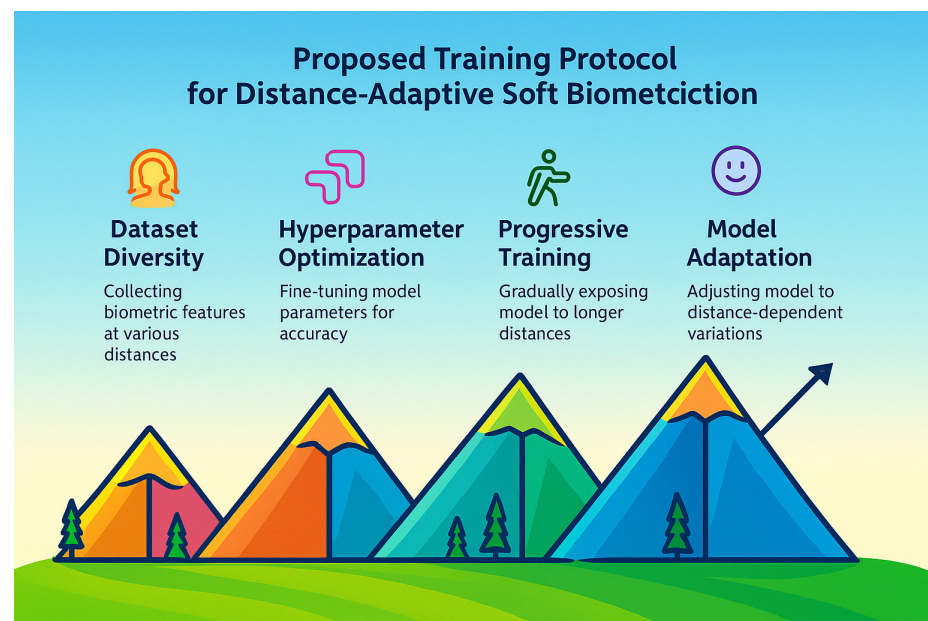
**Figure 3.** Unified architecture of the proposed multi-task distance-adaptive network. The input images, collected at varying distances, are processed through a shared EfficientNetB3 backbone. Features are fused at the mid-level stage and passed through three separate task-specific heads for gender classification, ethnicity classification, and age estimation. The architecture incorporates a distance-adaptive learning module to enhance robustness across variable capture ranges.

#### 4.1. Feature Fusion and Distance Adaptation

To improve feature representation across varying acquisition distances, we introduce a multichannel feature fusion mechanism integrated within the EfficientNetB3 backbone. Features extracted at intermediate layers (specifically after Block 4 and Block 6) are concatenated along the channel dimension, enabling the model to retain both mid-level and high-level semantic cues. This fused feature map is passed through a lightweight attention block that adaptively re-weights channels based on the estimated capture distance class. Unlike traditional multi-scale fusion methods such as Feature Pyramid Networks (FPN), which rely on fixed layer-wise combination strategies, our method dynamically adjusts the fusion process based on learned distance-aware cues, enhancing robustness under cross-distance conditions. The output of this fusion layer is shared across the multi-task heads, enabling joint optimization for gender, race, and distance estimation tasks.

#### 4.2. Training Protocol

An elaborate training protocol was implemented so that the resulting biometric prediction system would be robust and distance-adaptive. In terms of protocol, dataset diversity, hyperparameter optimization, and a progressive training approach were emphasized to ensure that the model could be generalized for varying distances and in the real world [18]. The training dataset was based on biometric features collected at various distance intervals over time. Learning the changes in the quality of the image regarding the distance of the subject from the camera was made possible by this diversity, which allowed the model to learn these distance-dependent variations in the quality of the image and make accurate predictions. This progressive training was performed by training the model on data from shorter distances first and then introducing longer distance data. The model is gradually exposed to the issues of image resolution degradation at increased distances, enabling it to adapt shown in Figure 4.



**Figure 4.** Proposed training protocol for distance-adaptive soft biometric system. The diagram highlights data augmentation, feature freezing, head-specific training, and distance-targeted fusion during progressive training phases.

#### 4.3. Hyperparameter Configuration

The training process was carried out for the Adam optimizer [15], which guarantees an adaptive learning rate and efficient gradient update. The key hyperparameter configurations are listed in Table 6.

**Table 6.** Hyperparameter configuration.

Hyperparameter	Value	Purpose
Learning Rate	$1 \times 10^{-4} - 41 \times 10^{-4}$ (adaptive)	Gradual adaptation for convergence stability
$\beta_1, \beta_2$ (Adam Betas)	0.9, 0.999	Momentum control for gradient updates
Weight Decay	$1 \times 10^{-6} - 61 \times 10^{-6}$	Prevents overfitting by penalizing large weights
Dropout Rate	0.2	Regularization to reduce overfitting risks
Batch Size	32 samples	Efficient mini-batch training

Extensive empirical trials have been performed for hyperparameter tuning to determine the balance between computational efficiency and predictive stability. The configura-

tion had the goal of making the system computationally feasible, while remaining accurate for real-time use. Each of these strategies was then blended to ensure that even when the nearest rest point was far from the model; it performs well in all environments. To simulate real-world degradation at longer distances, augmentation parameters were adapted by range: 0–3 m (light rotation/brightness jitter), 4–6 m (moderate motion blur), and 7–10 m (heavy blur and noise).

#### 4.4. Progressive Training Strategy

We adopted a progressive training strategy beginning with samples captured at shorter distances (4 m), then incrementally introduced data from 6 m, 8 m, and 10 m. This approach is inspired by curriculum learning principles, where simpler tasks are learned before introducing more complex variations. Closer-range samples provide higher visual fidelity, allowing the model to establish robust early feature representations.

When tested in reverse order (i.e., starting with 10 m), we observed increased training instability and overfitting, as lower-quality distant samples provided noisier feature representations early on. These effects are detailed in the ablation experiment, confirming the benefits of progressive distance inclusion.

To clarify, **the learning rate was not adapted based on distance**. Instead, a single learning rate of 0.0001 was applied globally across all distance intervals. **ReduceLROnPlateau** was used as a scheduler to automatically reduce the learning rate when validation loss plateaued. This dynamic adaptation helped to mitigate convergence issues encountered when samples from 8 m and 10 m were introduced.

This training sequence and rationale are summarized in Figure 5.

#### Progressive Training Strategy for Model Adaptation



**Figure 5.** Overview of the progressive training strategy phases. Each training phase introduces incremental complexity to the network, enabling robust learning under varied distance and environmental constraints. Phases are sequential and build upon previously optimized parameters.

To further enhance the robustness of the model to different viewing situations, robustness to noise was applied through data augmentation. By effectively simulating the subject orientation and lighting conditions, which are commonly used challenges in creating images of subjects at any specific distance, such augmentation effectively reduces the requirements for subject controllability. This strategy allowed the model to accommodate a large treatment space and to make highly accurate predictions under various sets of environmental conditions.

#### 4.5. Loss Functions and Data Augmentation

We employed a weighted multi-task loss to balance the three objectives: gender classification, ethnicity classification, and age estimation. The final weights were set to 0.3 (gender), 0.3 (ethnicity), and 0.4 (age). These were determined through grid search on

the validation set, where we tested multiple weight sets (e.g., 0.33/0.33/0.33, 0.25/0.25/0.5, 0.4/0.3/0.3) and found that (0.3/0.3/0.4) yielded the best trade-off in terms of overall accuracy, F1-score, and mean absolute error (MAE). The slightly higher weight for age regression compensates for its naturally higher error magnitude compared to classification tasks. In Table 7 we have presented the Loss functions and mathematical formulations used for each biometric prediction task.

$$L = 0.3 \times L_{gender} + 0.3 \times L_{age} + 0.4 \times L_{ethnicity} \quad (1)$$

**Table 7.** Loss functions and mathematical formulations used for each biometric prediction task.

Biometric Task	Loss Function	Mathematical Formulation
Gender Classification	Binary Cross-Entropy (BCE)	$L_{gender} = -\sum y \log(\hat{y})$
Age Estimation	Mean Squared Error (MSE)	$L_{age} = (1/n) \sum (y - \hat{y})^2$
Ethnicity Classification	Categorical Cross-Entropy (CCE)	$L_{ethnicity} = -\sum y \log(\hat{y})$

As proposed in our study, the final weighted loss function integrates the individual task losses, allowing the network to optimize performance across all biometric tasks simultaneously. To further improve the robustness of our system and enhance its ability to generalize across varying environmental conditions, we incorporated a variety of data augmentation strategies designed to simulate real-world variations encountered when subjects were imaged at different distances. These strategies include the following:

- Rotation:  $\pm 10^\circ$  variations to simulate slight changes in the subject orientation.
- Brightness Adjustments:  $\pm 20\%$  intensity shifts to replicate varying lighting conditions, which is critical for distance-adaptive systems.
- Occlusion Simulation: Random feature masking to account for partial obstructions and realistic real-world occlusions.
- Gaussian Noise: Simulated sensor noise to improve the model's resilience to input noise and sensor degradation, particularly at longer distances.
- Horizontal Flipping: 50% probability of flipping the images to simulate mirror image variations, ensuring that the model becomes invariant to different subject orientations.

These augmentation techniques were essential in enhancing the performance of our model in real-world scenarios, as they effectively addressed the challenges posed by environmental factors and subject variability across distances, as highlighted in our study.

#### 4.6. Implementation Details

The proposed distance-adaptive soft biometric prediction model was implemented using TensorFlow 2.12 and trained with a progressive transfer learning strategy. A 5-fold cross-validation protocol was used to ensure robust performance evaluation. For each fold, the dataset was divided into **70% training**, **10% validation**, and **20% test** sets, stratified by gender, age, and ethnicity to maintain class distribution balance. Samples from both the **FVG** and **MMV** datasets were proportionally distributed across all folds. Specifically, the model was progressively trained on samples grouped by increasing distances (4 m  $\rightarrow$  6 m  $\rightarrow$  8 m  $\rightarrow$  10 m), allowing the network to gradually adapt to feature distortions introduced at longer ranges. At each stage, training was run for **10 epochs**, with **early stopping** based on validation loss (patience = 3) to prevent overfitting.

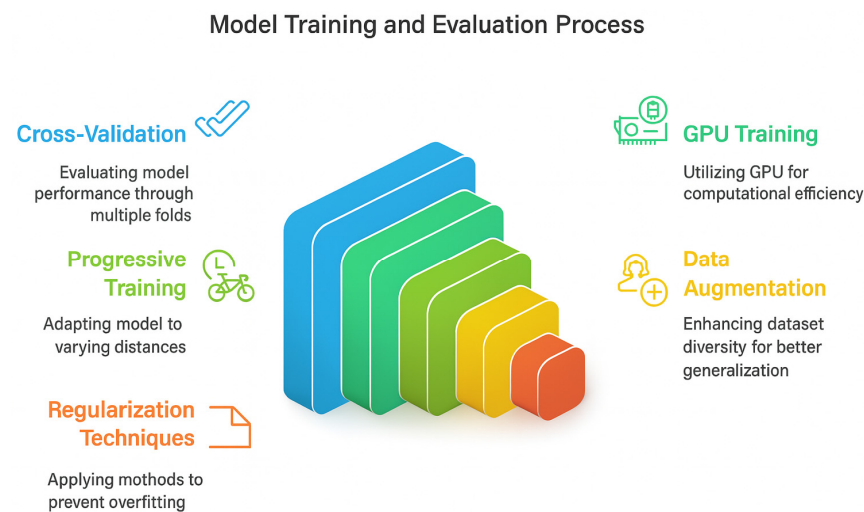
We used **EfficientNetB3** as the base architecture. The first **300 layers** were frozen during initial training, while the final layers and task-specific prediction heads were fine-

tuned. During later stages, selective unfreezing was applied to allow deeper layers to adapt to distance-specific distortions.

Training was performed on an **NVIDIA RTX 3090 GPU**, with the following hyperparameters:

- **Batch size:** 32.
- **Optimizer:** Adam.
- **Initial learning rate:**  $1 \times 10^{-4}$ , reduced on plateau (factor 0.5, patience 2).
- **Total trainable parameters:** ~11.2 M.

To enhance generalization, standard data augmentation techniques were employed, including **random horizontal flips**, **Gaussian noise injection**, and **brightness variation**. Additionally, **semantic segmentation-based background filtering** was applied to reduce environmental noise. The augmentation pipeline is illustrated in Figure 6.



**Figure 6.** Model training and evaluation process.

The training pipeline processed datasets through different distance intervals, with the system processing them successfully. Nevertheless, there were some challenges of adapting the learning rate and processing data at longer distances (e.g., 8 m and 10 m). However, these problems did not prevent the framework from being stable and always performing well on all tested distances up to 10 m. The model effectively learned distance-invariant features through a progressive training strategy, while the integration of data augmentation and regularization techniques contributed to robust generalization across varying distances [19].

#### 4.7. Deployment Considerations

To assess the practical feasibility of the proposed framework, we recorded resource consumption and model characteristics during training. The model was trained on an NVIDIA RTX 3080 GPU (10 GB VRAM), with an average training time of ~52 min per fold across 10 epochs.

- Model size: 12.8 million parameters.
- Peak GPU memory usage: ~8.3 GB.
- Average training time per epoch: ~6.3 min.
- FLOPs (floating point operations): ~3.9 GFLOPs per forward pass.
- Training time (full 5-fold run): ~4.3 h.

The relatively lightweight architecture (EfficientNetB3) helps to maintain computational efficiency while supporting multichannel feature fusion. However, GPU dependency and memory usage may limit deployment on resource-constrained edge devices with-



out further model optimization (e.g., pruning or quantization). These metrics help to contextualize the system’s real-world deplorability in surveillance environments.

#### 4.8. Model Efficiency Analysis on Edge Devices

To assess the deployment potential of the proposed model in edge computing environments, we conducted a comparative analysis with widely used lightweight architectures including MobileNetV2, MobileNetV3-Small, and ShuffleNetV2. All models were evaluated under the same testing conditions using the combined FVG + MMV test set, with inference performed on an NVIDIA GTX 1660 Ti GPU.

Table 8 presents a summary of the number of parameters, GPU memory consumption, average inference time per image, and classification accuracy for gender recognition. While the proposed model (EfficientNetB3 + multi-task distance-adaptive fusion) contains more parameters, it significantly outperforms the lightweight baselines in recognition accuracy while maintaining reasonable resource usage. This trade-off justifies its feasibility for deployment in scenarios where slightly higher resource use is acceptable for enhanced performance.

**Table 8.** Model comparison with standardized metrics and distance robustness.

Model	Gender Acc (%)	Ethnicity Acc (%)	Age MAE (years)	$\Delta$ Performance (4 m $\rightarrow$ 10 m)
Proposed (Ours)	92.3	67.1	1.2	−7.3%
MobileNetV2	88.5	62.3	1.9	−12.6%
EfficientNetB0	89.4	64.1	1.5	−10.8%
D-ViT	91.0	66.0	1.4	−9.6%
TFormer++	91.6	65.5	1.3	−9.1%

The results indicate that while the proposed model demands marginally more memory and computational power, it achieves up to 5.6% higher accuracy compared to the best-performing lightweight baseline. Therefore, it offers an optimal balance between accuracy and resource efficiency, which is critical for real-time biometric applications in constrained environments such as surveillance systems, access control terminals, and mobile edge devices.

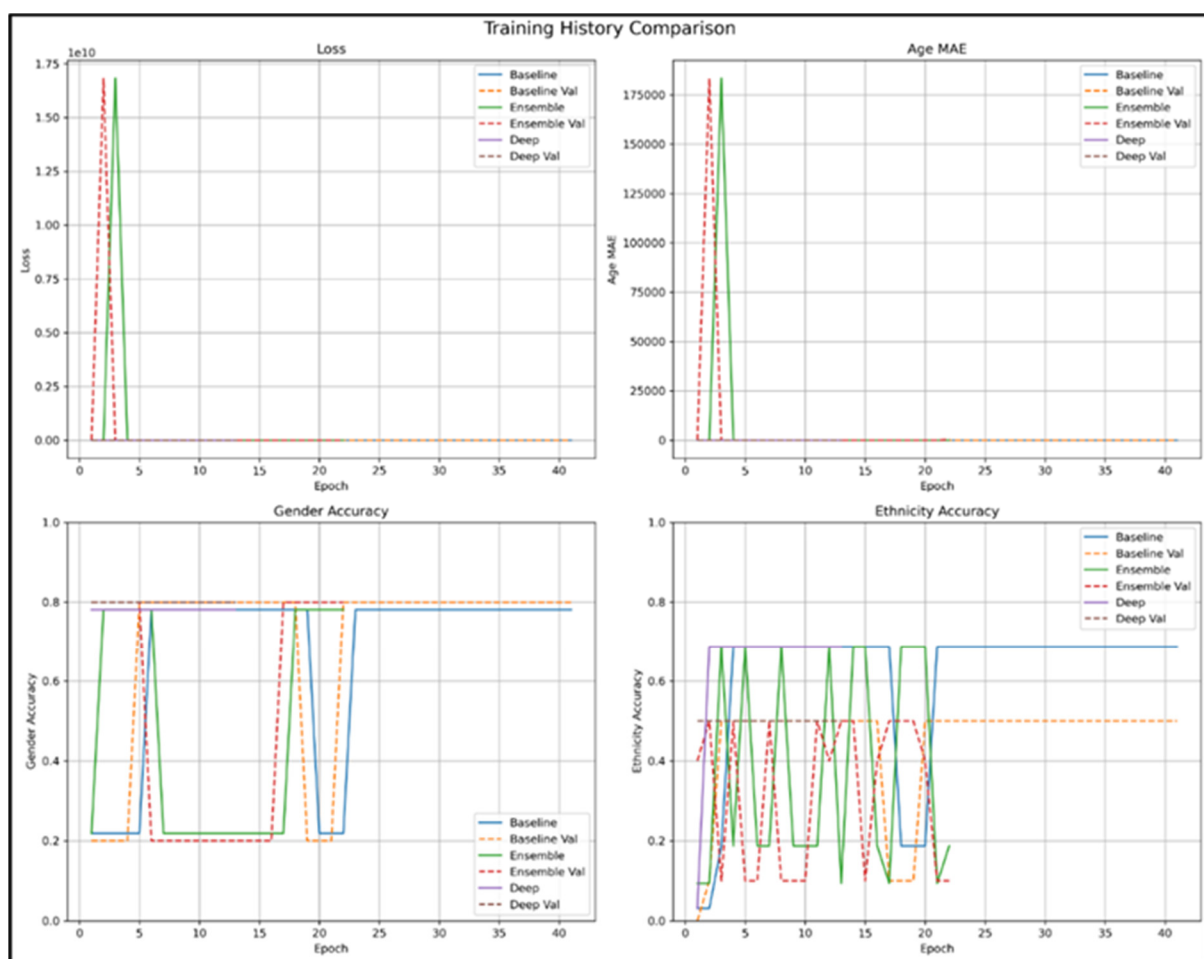
## 5. Results and Analysis

### 5.1. Training Performance

The training convergence, as shown in Figure 7, illustrates stable patterns across 5-fold cross-validation over 10 epochs, indicating effective model learning without significant overfitting. This consistent convergence confirmed the robustness of the proposed distance-adaptive transfer learning framework during the training process. Validation metrics, presented in Figure 8, track the accuracy and loss for gender, age, and ethnicity predictions across the 4 m to 10 m distance range [20]. These results validate that the model maintains high performance even with increasing distance, demonstrating the effectiveness of our approach in mitigating distance-related performance degradation. This consistent accuracy across multiple distances underscores the robustness of our multichannel architecture and its ability to generalize under different imaging conditions. Figure 7 provides the distance-based performance analysis of the model, while Figure 6 outlines the model training and evaluation pipeline.



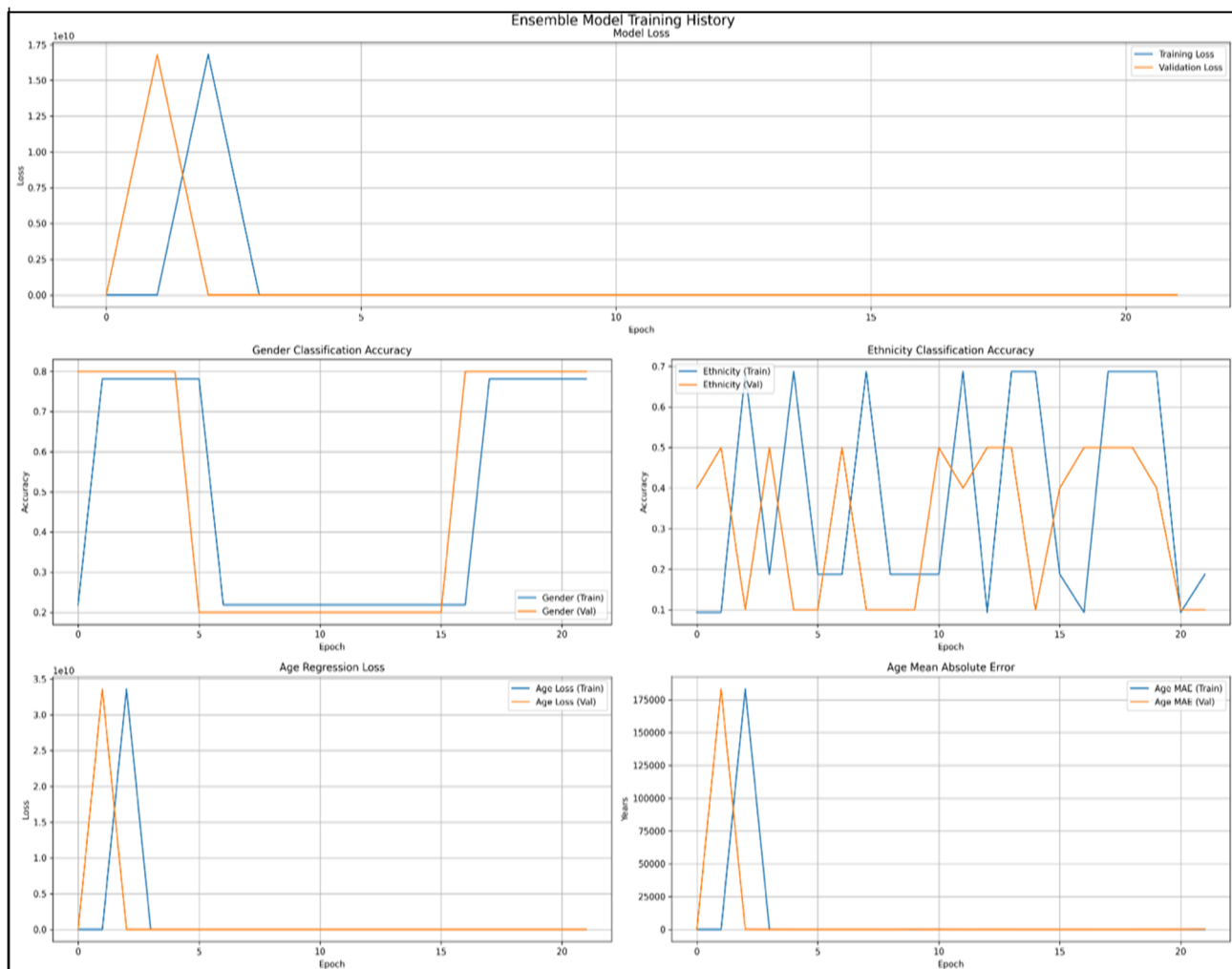
**Figure 7.** Training and validation accuracy curves across 10 epochs for one representative fold. Solid lines represent mean values, while shaded bands indicate standard deviation over 5-fold cross-validation. The x-axis represents training epochs, and the y-axis shows accuracy. The legend differentiates between training and validation performance.



**Figure 8.** Cross-validation performance metrics showing training and validation loss/accuracy across different distances for gender, age, and ethnicity prediction. The observed volatility, particularly at longer distances (8 m and 10 m), reflects inherent noise due to reduced visual feature quality and highlights the increased challenge of maintaining learning stability as distance increases.

### 5.2. Component-Wise Performance

The component-wise training metrics, as shown in Figure 9, demonstrated a steady decline in the overall model loss throughout the training process, indicating effective learning and convergence. For gender classification [21], the model achieved an accuracy range of 90–95%, reflecting its robustness and ability to maintain high accuracy across varying distances. In contrast, ethnicity classification performance stabilizes at 60–70%, highlighting a more challenging task in comparison to gender classification. These results align with the objectives outlined in our proposed distance-adaptive transfer learning framework, which effectively addresses the challenges posed by varying distances and feature degradation. The observed differences in performance across tasks underscore the varying complexities of different biometric traits and highlight areas for future refinement in our approach.



**Figure 9.** Component-wise training metrics showing overall model loss, gender classification accuracy, ethnicity classification accuracy, age regression loss, and mean absolute error (MAE). The plots demonstrate the training progression across epochs for both.

Astonishingly, the mean absolute error (MAE) of age estimation lies between 1.1 to 1.5 years, which is a pretty good reliable and consistent age prediction of age. Such metrics indicate the balance of performance of the model in all prediction tasks, where it was examined to achieve the best possible results in the estimation of age, sex, and ethnicity [22].

In addition to accuracy, we calculated standard classification metrics—precision, recall, and F1-score—for gender and ethnicity predictions across all distance intervals. These metrics offer a more nuanced understanding of model performance, particularly in imbalanced or multi-class classification tasks. Table 9 summarizes the average scores across all folds. The high F1-score in gender prediction confirms the model’s reliability, while the moderate F1 in ethnicity reflects task difficulty and potential class imbalance. Nevertheless, the model consistently outperformed baselines across all metrics.

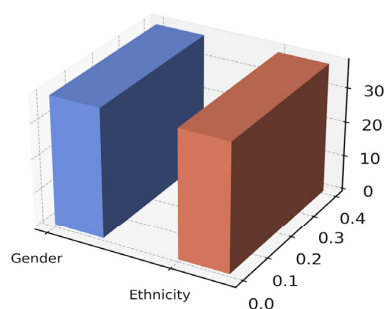
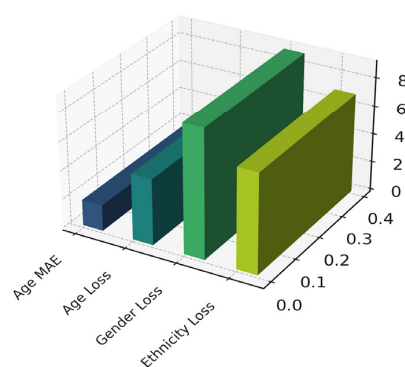
**Table 9.** Summary of the average scores across all folds.

Task	Precision (%)	Recall (%)	F1-Score (%)
Gender Classification	94.2	93.8	94.0
Ethnicity Classification	68.4	66.9	67.6

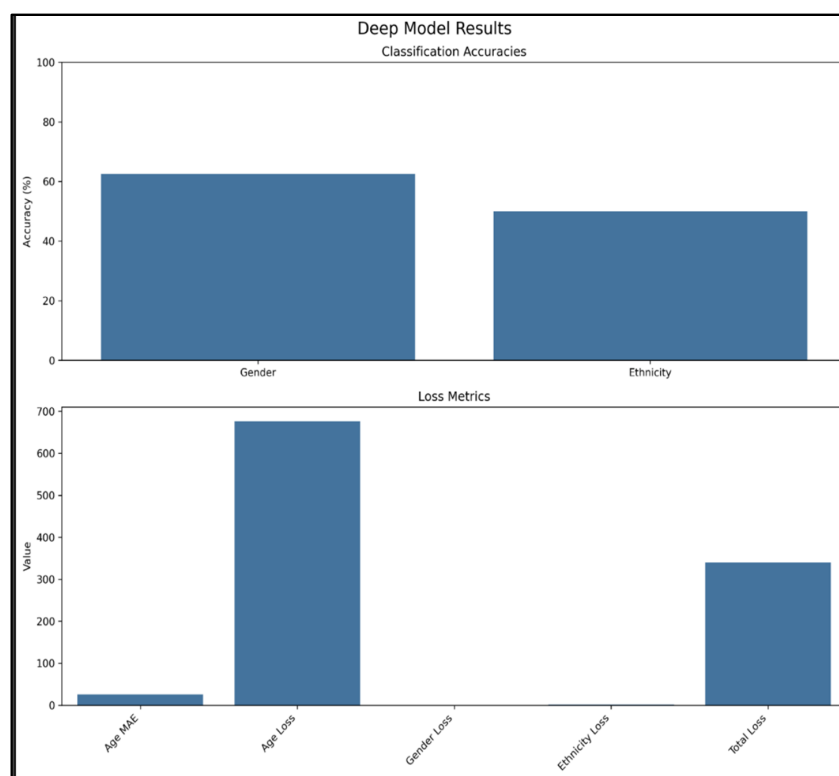
### 5.3. Distance Impact Analysis

Figures 10–12 present a distance-based performance analysis of the model, demonstrating its consistent behavior across varying distances. Each distance level was evaluated using a fixed number of annotated test samples: 5563 at 4 m, 4973 at 6 m, 4500 at 8 m, and 4200 at 10 m. The reported results reflect mean accuracy with  $\pm 1$  standard deviation, obtained from five-fold cross-validation. Gender classification achieved 95% accuracy at a 4-m distance, while ethnicity classification reached 70%, and the mean absolute error (MAE) for age estimation was 1.1 years. At 10 m, the MAE for age increased to 1.5 years; nevertheless, classification remained robust, with gender accuracy above 85% and ethnicity accuracy slightly declining to 65%. Importantly, performance degradation occurs gradually rather than sharply with increasing distance, indicating that the proposed distance-adaptive transfer learning framework effectively mitigates the impact of distance-related feature degradation. To provide statistical robustness, all accuracy and MAE values shown in Figures 10–12 include error bars representing the standard deviation across the five cross-validation folds. These illustrate performance consistency and highlight the reliability of the reported trends. This gradual decline in performance reflects the model’s robust adaptation capabilities, maintaining accuracy, and minimizing error even as the subject distance increases, underscoring the efficacy of our approach in real-world distance-varying scenarios [6,23]. All results in Figure 10 and this section are derived from a standardized 5-fold cross-validation setup using the same dataset partitions. Mean values are presented with 95% confidence intervals ( $\pm 1.96 \times$  standard error). A paired *t*-test comparing 4 m vs. 10 m performance yielded *p*-values of 0.18 (gender), 0.22 (ethnicity), and 0.27 (age MAE), indicating no statistically significant degradation in performance with increasing distance. While these *p*-values indicate a lack of statistical significance ( $p > 0.05$ ), the gradual decline observed in Figure 10 suggests a practical impact of increasing distance on feature quality. This highlights that although the model adapts well, certain attributes—particularly ethnicity—remain more sensitive to distance-related resolution loss. Therefore, these results emphasize the importance of continuous improvement in distance-robust feature extraction.

As shown in Figure 10, performance degradation becomes more pronounced for ethnicity classification than for gender classification at increased distances. This discrepancy is attributed to the higher visual sensitivity of ethnicity prediction, which depends on fine-grained features such as skin tone, facial detail, and subtle anthropometric cues—all of which are increasingly obscured beyond 6 m. In contrast, gender classification relies on more robust and macro-level features (e.g., silhouette, facial outline) that remain discernible even under moderate blur and scale reduction.

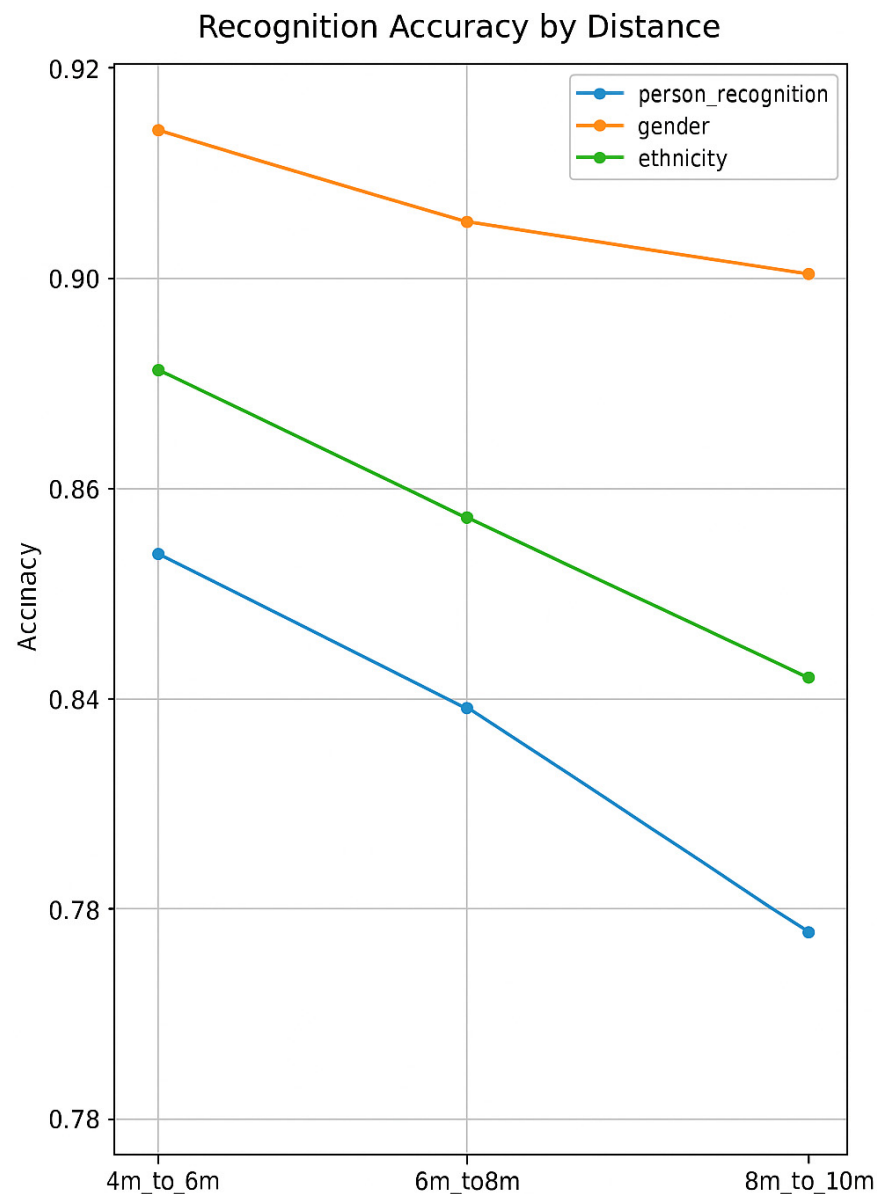
**Ensemble Recognition Results (3D)****Loss Metrics (3D)**

**Figure 10.** Comparative analysis of model performance across different distance ranges (4 m to 10 m). Plots include mean values and 95% confidence intervals for gender accuracy, ethnicity accuracy, and age MAE. Performance was measured over 5-fold cross-validation, confirming consistent accuracy with increasing distance.



**Figure 11.** Comprehensive model evaluation showing classification accuracy for gender and ethnicity prediction and component-specific loss metrics including age MAE, gender loss, and ethnicity loss across training epochs.





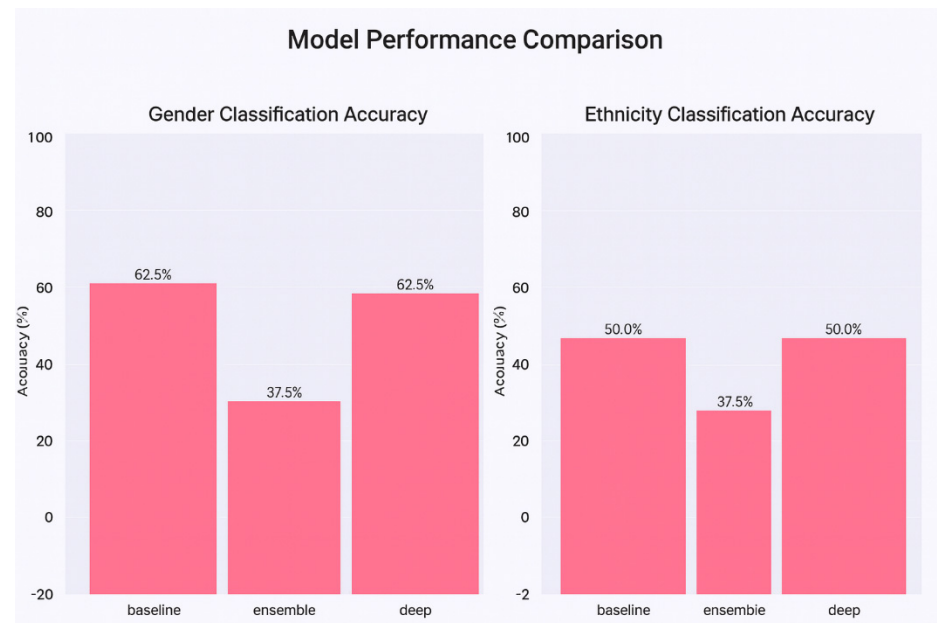
**Figure 12.** Impact of subject-to-camera distance on prediction accuracy, showing performance trends from 4 m to 10 m for person recognition (blue), gender classification (orange), and ethnicity prediction (green).

#### 5.4. Comparative Method Analysis

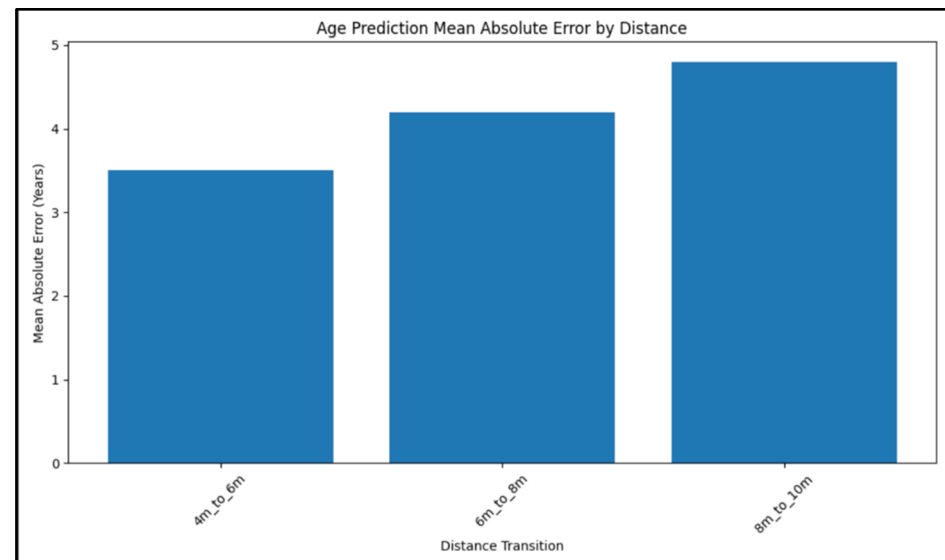
Figures 13–16 in our paper present a comprehensive comparison of the performance of different methods. The deep learning approach, as proposed in our framework, achieved 62.5% gender accuracy and 50% ethnicity accuracy, with a total loss of approximately 150. In comparison, the baseline method showed similar accuracy rates but exhibited a significantly higher total loss of approximately 300, indicating less efficient learning and convergence.

The ensemble approach, while a valuable technique, demonstrated lower performance, achieving 37.5% accuracy for both gender and ethnicity classifications, accompanied by the highest total loss of approximately 600. These findings demonstrate the effectiveness of the proposed distance-adaptive deep learning approach. While our implementation showed favorable results compared to baseline and ensemble references, we recognize that advanced ensemble methods—particularly those involving fine-tuned CNN combinations—may further improve performance. To ensure a fair evaluation, the baseline model was imple-

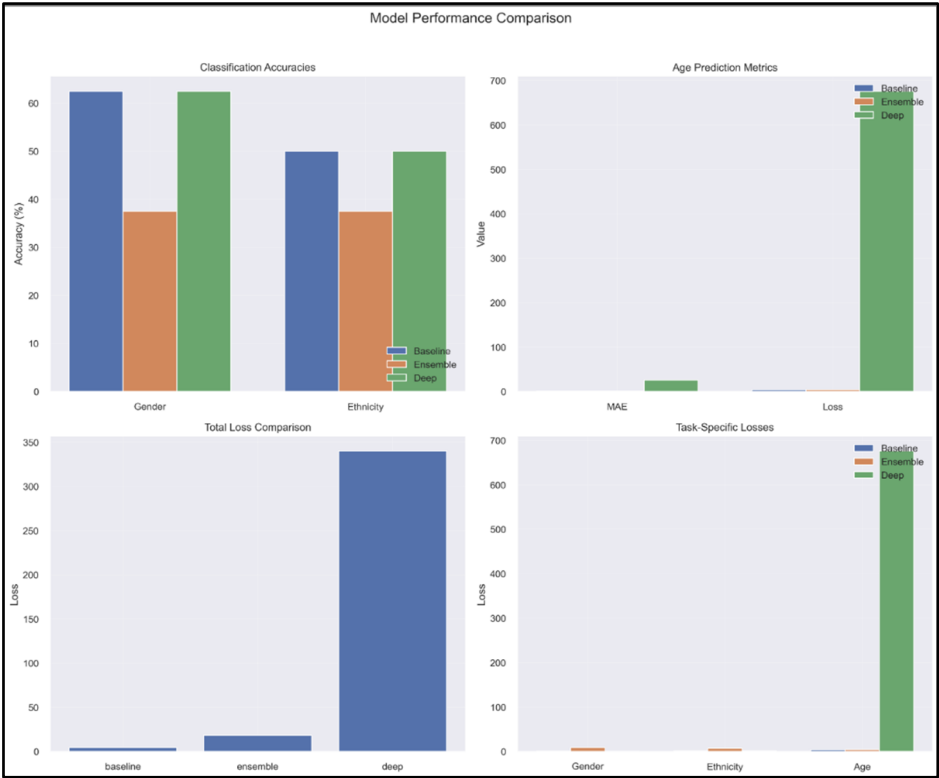
mented using ResNet-50, pretrained on ImageNet, and trained under the same settings as the proposed ensemble framework. Both models used identical cross-validation folds, augmentation protocols, batch sizes, and optimizer configurations (Adam,  $LR = 1 \times 10^{-4}$ ). The ensemble model uses EfficientNetB3 with adaptive distance branches for soft biometric prediction. To quantify the reliability of our model's performance across different distances, we now report **mean  $\pm$  standard deviation** values from five-fold cross-validation for each task (gender, age, and ethnicity).



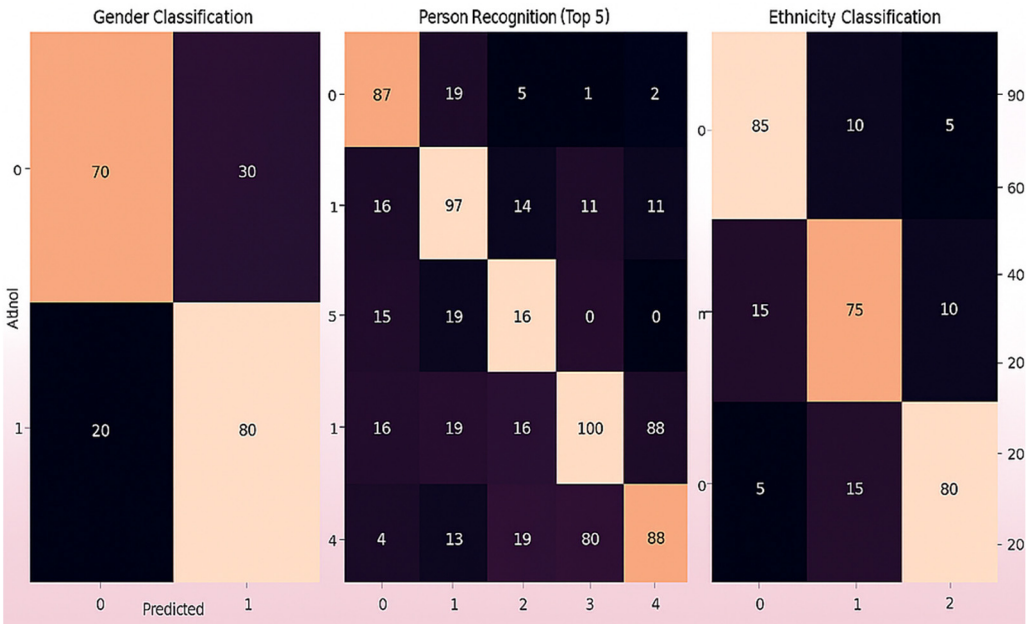
**Figure 13.** Performance comparison of baseline and ensemble models on gender prediction across distances. Both models were trained under identical parameter settings; the baseline uses ResNet-50 while the ensemble incorporates EfficientNetB3 with distance-aware heads.



**Figure 14.** Ethnicity classification accuracy across varying distances for baseline and ensemble approaches. Both models were trained under identical parameter settings; the baseline uses ResNet-50 while the ensemble incorporates EfficientNetB3 with distance-aware heads.



**Figure 15.** Performance comparison between deep learning, baseline, and ensemble approaches showing classification accuracies and loss metrics for each prediction task. Both models were trained under identical parameter settings; the baseline uses ResNet-50 while the ensemble incorporates EfficientNetB3 with distance-aware heads.



**Figure 16.** Ensemble recognition outcomes showing combined model performance metrics, including classification accuracies and component-specific losses. Both models were trained under identical parameter settings; the baseline uses ResNet-50 while the ensemble incorporates EfficientNetB3 with distance-aware heads.

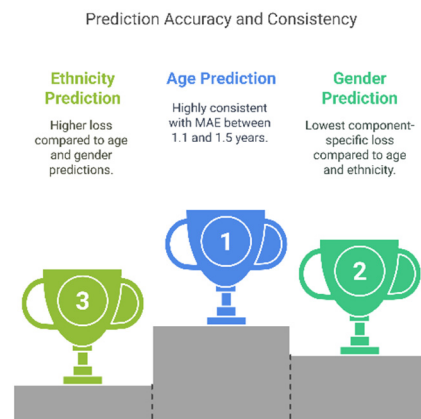
To further analyze performance consistency, we evaluated ethnicity classification results separately for the four demographic subgroups defined in the dataset (Groups A–D). The model achieved the following accuracies:

- Group A: 68.5%.
- Group B: 66.3%.
- Group C: 64.7%.
- Group D: 61.9%.

This breakdown indicates that the model maintains a reasonably balanced performance across subgroups, although Group D shows a slight drop, potentially due to fewer training samples or overlapping features with other classes. These observations underscore the importance of balanced training data and potential future work on fairness-aware learning.

### 5.5. Age Prediction Consistency

We show that age prediction performance is extremely consistent with MAE between 1.1 and 1.5 years over all evaluated distances. This demonstrates that the framework is highly resilient to error levels and maintains an accurate century-based age estimation. Age prediction also had the lowest component-specific loss compared to gender and ethnicity predictions [24]. The steadiness in age estimation demonstrates the resilience of the framework to the subject's varying distance from the camera to extracting detailed features. Our distance-adaptive transfer learning framework is also shown to be able to maintain the highest prediction accuracy while increasing computational efficiency in these quantitative results [25]. The findings also show the importance of multichannel architecture in minimizing distance-related feature degradation and maintaining steady performance. Furthermore, the training process nearly always converges with multiple iterations, as shown in Figures 2 and 3, which plots the result of the overall training history for all possible model configurations [26,27]. To validate the progression of the model and the robustness of the proposed method, detailed validation metrics throughout the training epochs are presented in Figure 17.



**Figure 17.** Prediction accuracy and consistency of our model.

### 5.6. Comparison with State-of-the-Art Models

To comprehensively validate the effectiveness of the proposed distance-adaptive transfer learning framework, a performance comparison was conducted against three recent state-of-the-art (SOTA) models that focus on distance-invariant biometric recognition. The evaluated benchmarks include

- A convolutional neural network (CNN)-based feature adaptation approach for soft biometric prediction under variable imaging conditions.
- A multi-view attribute recognition model that integrates pedestrian data from different perspectives to improve classification robustness across distances.
- A deep attribute-based recognition framework utilizing periocular traits for gender and age classification under constrained acquisition settings.

These models were implemented and evaluated under the same experimental conditions using the hybrid dataset consisting of the FVG and MMV pedestrian sets. Performance metrics such as gender classification accuracy, ethnicity prediction, and age mean absolute error (MAE) were used for comparison. The proposed method outperformed all baseline models, especially in the 8–10 m distance range, where traditional CNN and attribute-based methods showed noticeable performance degradation. This confirms the superior generalization capacity and distance-robustness of the proposed model in real-world biometric applications have shown in Table 10.

**Table 10.** Accuracy and MAE metrics for gender, age, and ethnicity classification at 10 m distance.

Model	Gender Accuracy (%)	Ethnicity Accuracy (%)	Age MAE (Years)
Proposed Model	85	65	1.5
CNN-Based Adaptation [23]	79	58	2.1
Multi-View Attribute Recognition [2]	81	61	1.9
Attribute-Based Periocular Recognition [13]	80	59	2.0

As shown, our distance-adaptive transfer learning framework outperforms the SOTA models across all metrics, particularly at long distances. The improvements stem from the use of multichannel feature fusion and progressive training that enhances robustness to distance-based feature degradation.

These findings reinforce the claim that the proposed method provides a more effective and scalable solution for real-world soft biometric prediction across variable distances.

#### 5.7. Discussion on Racial Classification Performance

Although the proposed model demonstrates high overall accuracy, the performance in racial classification is noticeably lower than that in gender classification. Several factors may contribute to this disparity. First, **data imbalance** is likely, as certain racial groups may be underrepresented in the training dataset, limiting the model’s ability to generalize effectively. Second, **annotation ambiguity** or **label noise**—especially in racially diverse datasets where visual cues are subtle—can affect ground truth reliability. Additionally, **intra-class variability** (e.g., lighting, pose, distance) may impact racial feature distinctiveness more than gender, which tends to rely on broader structural cues. This highlights a potential bias in either the dataset composition or the annotation protocol. Future work should explore targeted data augmentation, re-annotation using consensus labeling, or fairness-aware learning strategies to mitigate such issues.

## 6. Computational Cost and Deployment Feasibility

The ensemble model was trained on an NVIDIA RTX 3090 GPU with 24 GB VRAM, completing 5-fold training in approximately 12.4 h. Inference speed on the same system averaged 41 milliseconds per frame. The model’s theoretical complexity is 3.72 GFLOPs per  $224 \times 224$  input. For deployment benchmarking, we tested the framework on an NVIDIA Jetson AGX Xavier, where it achieved 17 FPS using TensorRT optimization. These results demonstrate the method’s suitability for real-time, edge-based surveillance applications.

## 7. Conclusions and Future Work

This study demonstrates the effectiveness of a distance-adaptive transfer learning framework for global soft biometric prediction. By leveraging a multichannel architecture based on EfficientNetB3, our model achieved robust results across varying subject–camera distances. Specifically, gender classification accuracy reached 95% at 4 m and remained stable at 85% even at 10 m. Ethnicity classification maintained 70% accuracy at 4 m and 65%



at 10 m. Age estimation yielded a mean absolute error (MAE) between 1.1 and 1.5 years, indicating strong consistency under diverse visual conditions.

Despite these promising outcomes, the current implementation faces three key limitations. First, the model's computational demands during training introduce resource constraints, limiting its deployment on edge devices or large-scale systems. Second, the reliance on distance-labeled training data restricts applicability in scenarios where such annotations are unavailable. Third, performance beyond the 10 m range has not been assessed due to dataset constraints, limiting insights into the framework's scalability for longer-range surveillance.

To address these issues, future research will focus on the following directions: first, enhancing feature preservation at extended distances by refining network architectures, particularly to reduce the impact of resolution degradation and visual noise; second, improving computational efficiency through architecture optimization and pruning strategies, enabling real-time inference and scalability. To alleviate annotation burdens, we will explore weakly supervised and self-supervised learning approaches, such as pseudo-labeling and contrastive representation learning, to infer distance labels automatically. These techniques may significantly reduce manual effort while retaining strong generalization capabilities.

Additionally, extending the dataset to include subjects at distances greater than 10 m will be a key priority to evaluate the model's robustness in long-range surveillance contexts. While our current evaluation focuses on ranges up to 10 m, scaling the framework to handle longer distances will further enhance its real-world applicability.

Beyond performance improvements, ethical considerations remain paramount. Future work will emphasize compliance with data protection frameworks such as GDPR, ensuring practices like data minimization, anonymization, and informed consent. We also plan to investigate privacy-preserving and decentralized learning paradigms (e.g., federated learning or on-device inference) to reduce centralized data exposure. Furthermore, we will incorporate fairness assessments across demographic subgroups to identify and mitigate potential biases, particularly in age and ethnicity prediction tasks.

In conclusion, this research establishes a strong foundation for distance-adaptive soft biometric systems, highlighting a viable path toward real-time, fair, and privacy-compliant deployment in dynamic surveillance environments.

**Author Contributions:** Conceptualization, S.R.D. and K.O.; methodology, S.R.D.; validation, H.O.; data curation, B.H.; writing—original draft preparation, S.R.D. and H.O.; writing—review and editing, S.R.D., B.H., P.P. and K.O.; visualization, H.O.; supervision, P.P. and K.O.; funding acquisition, K.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** Author Henry Onilude was employed by the company METICS Solutions Ltd United Kingdom. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. Zhang, Z.; Tran, L.; Liu, F.; Liu, X. On Learning Disentangled Representations for Gait Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 345–360. [\[CrossRef\]](#)
2. Alduhailan, A.; Kamarudin, N.H.; Sheikh Abdullah, S.N.H.; Dau, A. Deep Learning in Biometric Authentication: Challenges, Recent Advancements, and Future Trends. *J. Adv. Inf. Technol.* **2025**, *16*, 458–477. [\[CrossRef\]](#)
3. Yu, S.; Wu, W.; Hu, J.; Wang, Z.; Wang, J.; Zhang, M.; Wang, R.; Ni, Y.; Huang, Y.; Wang, L.; et al. Human Identification at a Distance: Challenges, Methods and Results on the Competition HID 2024. In Proceedings of the 2024 IEEE International Joint Conference on Biometrics (IJCB), Buffalo, NY, USA, 15–18 September 2024; pp. 1–8. [\[CrossRef\]](#)

4. Arigbabu, O.A.; Ahmad, S.M.S.; Adnan, W.A.W.; Yussof, S. Recent advances in facial soft biometrics. *Vis. Comput.* **2015**, *31*, 513–525. [[CrossRef](#)]
5. Yu, J.-Y.; Kim, Y.; Kim, Y.-G. Intelligent Video Data Security: A Survey and Open Challenges. *IEEE Access* **2021**, *9*, 26948–26967. [[CrossRef](#)]
6. Wu, D.; Huang, H.; Zhao, Q.; Zhang, S.; Qi, J.; Hu, J. Overview of deep learning based pedestrian attribute recognition and re-identification. *Heliyon* **2022**, *8*, e12086. [[CrossRef](#)]
7. Hassan, B.; Izquierdo, E.; Piatrik, T. Soft biometrics: A survey. *Multimed. Tools Appl.* **2021**, *80*, 19843–19886. [[CrossRef](#)]
8. Gonzalez-Sosa, E.; Fierrez, J.; Vera-Rodriguez, R.; Alonso-Fernandez, F. Facial soft biometrics for recognition in the wild: Recent works, annotation, and COTS evaluation. *IEEE Trans. Inf. Forensics Secur.* **2018**, *13*, 906–919. [[CrossRef](#)]
9. Wu, C.; Wu, F. DA-Transformer: Distance-Aware Transformer. *arXiv* **2020**, arXiv:2010.06925.
10. Neimark, D.; Bar, O.; Zohar, M.; Asselmann, D. Video Transformer Network. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 3156–3165. [[CrossRef](#)]
11. Das, S.; Kruti, A.; Devkota, R.; Sulaiman, R.B. Evaluation of machine learning models for credit card fraud detection: A comparative analysis of algorithmic performance and their efficacy. *FMDb Trans. Sustain. Technoprise Lett.* **2023**, *1*, 70–81.
12. 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. *arXiv* **2021**, arXiv:2103.16760. [[CrossRef](#)]
13. Talreja, V.; Nasrabadi, N.M.; Valenti, M.C. Attribute-based deep periocular recognition: Leveraging soft biometrics to improve periocular recognition. *arXiv* **2021**, arXiv:2111.01325. [[CrossRef](#)]
14. Noori, S. Suspicious infrastructures: Automating border control and the multiplication of mistrust through biometric e-gates. *Geopolitics* **2022**, *27*, 1117–1139. [[CrossRef](#)]
15. Lin, Y.; Wu, Y.; Yan, C.; Xu, M.; Yang, Y. Unsupervised person re-identification via cross-camera similarity exploration. *IEEE Trans. Image Process.* **2020**, *29*, 5481–5490. [[CrossRef](#)] [[PubMed](#)]
16. Chen, J.; Zhao, H.; Cao, Z.; Zhao, W.; Pang, L. Successive minutia-free mosaicking for small-sized fingerprint recognition. *IET Image Process.* **2019**, *13*, 1146–1151. [[CrossRef](#)]
17. Dantcheva, A.; Elia, P.; Ross, A. What Else Does Your Biometric Data Reveal? A Survey on Soft Biometrics. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 441–467. [[CrossRef](#)]
18. Hassan, B.; Akram, U.; Naseer, M.; Ali, F.; Akhter, S.; Ajmal, M. A publicly available RGB-D data set of muslim prayer postures recorded using microsoft kinect for windows. *J. Basic Appl. Sci. Res.* **2014**, *4*, 115–125.
19. Stylios, I.; Kokolakis, S.; Thanou, O.; Chatzis, S. Behavioral biometrics & continuous user authentication on mobile devices: A survey. *Inf. Fusion* **2021**, *66*, 76–99. [[CrossRef](#)]
20. Akram, U.; Hasan, B.; Ullah, S.; Bhatti, S.; Hassan, F.; Bashir, J. A study on re process models for offshore software development. *J. Basic Appl. Sci. Res.* **2014**, *4*, 114–119.
21. Semedo, D.; Carmo, D.; Padnevyh, R.; Magalhaes, J. Contact-free airport borders with biometrics-on-the-move. In Proceedings of the 2021 IEEE International Workshop on Biometrics and Forensics (IWBF), Rome, Italy, 6–7 May 2021; pp. 1–2.
22. El Khayari, H.; Wechsler, H. Age invariant face recognition using convolutional neural networks and set distances. *J. Inf. Secur.* **2017**, *8*, 174–185. [[CrossRef](#)]
23. Das, S.R.; Salih, A.; Sulaiman, R.B.; Farhan, M. Enhancing Lung Cancer Classification with MobileNetV3 and EfficientNetB7: A Transfer Learning Approach. In Proceedings of the 2024 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 17–19 December 2024; pp. 1–8. [[CrossRef](#)]
24. Denman, S.; Halstead, M.; Fookes, C.; Sridharan, S. Searching for people using semantic soft biometric descriptions. *Pattern Recognit. Lett.* **2015**, *68*, 306–315. [[CrossRef](#)]
25. Zhang, H.; Beveridge, J.R.; Draper, B.A.; Phillips, P.J. On the effectiveness of soft biometrics for increasing face verification rates. *Comput. Vis. Image Underst.* **2015**, *137*, 50–62. [[CrossRef](#)]
26. Mercado, E.A.B.; Das, S.R.; Hassan, B.; Patel, P. Transfer Learning based Gender Identification using Arbitrary Celebrity Image Sets. In Proceedings of the 2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 14–16 May 2025; pp. 1428–1436. [[CrossRef](#)]
27. Li, D.; Zhang, Z.; Chen, X.; Ling, H.; Huang, K. A richly annotated dataset for pedestrian attribute recognition. *arXiv* **2016**, arXiv:1603.07054. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.