International Journal of Advance Research in Science and Engineering
Volume No. 15, Issue No. 10, October 2026
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A Modified Classification System Using Convolutional Neural Networks (CNN) for Intelligent Human–Computer Interaction (HCI).

Corresponding Authors:

Aliyu Ndottijo yerima [1], Iliyasu Adamu [2]

Corresponding Emails:

ndottijoaliyu@mau.edu.ng [1], iliyasuadamu@mau.edu.ng[2] abubakarmoh'd79@gmail.com[3]

Abstract

Automatic age and gender classification has become increasingly important in Human–Computer Interaction (HCI), biometric authentication, surveillance systems, smart advertising, and personalized computing environments. This paper presents a modified systematic real-time age and gender classification framework based on Convolutional Neural Networks (CNNs). The proposed system employs a lightweight multi-task CNN architecture capable of simultaneously predicting gender and age groups from facial images. To improve robustness and generalization, a composite dataset compiled from Adience, UTKFace, and IMDB-WIKI datasets is utilized. Preprocessing techniques including face detection, normalization, image resizing, and augmentation are integrated into the training pipeline. The architecture incorporates shared convolutional layers and task-specific output branches, reducing computational complexity while maintaining high predictive capability. The framework is designed for deployment in real-time environments and edge devices through optimization techniques such as pruning, quantization, and TensorFlow Lite conversion. The study highlights the relevance of demographic recognition in intelligent HCI and discusses challenges relating to fairness, interpretability, privacy, and deployment. Results indicate that the proposed framework provides a practical foundation for efficient real-time demographic classification systems.

Keywords: *Age Estimation, Gender Classification, Convolutional Neural Networks(CNN), Deep Learning, Human–Computer Interaction(HCI), Real-Time Systems, Multi-Task Learning(MTL).*

Introduction

The growing demand for intelligent and adaptive computing systems has significantly increased interest in automatic demographic recognition. Age and gender classification provide valuable information that can be used to personalize user experiences, improve accessibility, strengthen security systems, and support intelligent decision-making processes. Modern HCI systems increasingly rely on demographic information to adapt interfaces, content, and interaction styles to specific user groups.

Traditional age and gender recognition methods relied heavily on handcrafted features and conventional machine learning algorithms. Although useful in controlled environments, these approaches often performed poorly when exposed to variations in lighting conditions, facial expressions, occlusions, image quality, and head pose. The emergence of deep learning and CNNs has transformed facial analysis by enabling automatic feature extraction directly from image data.

Despite considerable progress, several challenges remain. Many existing models require substantial computational resources, making deployment on edge devices difficult. Furthermore, demographic datasets frequently exhibit bias, reducing fairness and reliability across diverse populations. This study addresses these limitations through the development of a modified CNN-based framework optimized for real-time operation and intelligent HCI environments.

Literature Review

Research in demographic classification has evolved from handcrafted feature extraction methods to advanced deep learning architectures. Early studies employed Local Binary Patterns, Gabor filters, Active Appearance Models, and Support Vector Machines for classification tasks. While these techniques achieved moderate success, their performance deteriorated in unconstrained environments. The introduction of CNNs dramatically improved classification performance. Levi and Hassner demonstrated the effectiveness of CNNs for age and gender recognition using the Adience dataset. Subsequent research introduced deeper architectures including VGGNet, ResNet, MobileNet, and EfficientNet. These models improved feature representation and classification accuracy while enabling deployment across diverse environments.

Multi-task learning emerged as an important advancement. By sharing feature representations between age and gender prediction tasks, multi-task CNNs reduce redundancy and improve computational efficiency. Attention mechanisms further enhanced performance by enabling networks to focus on informative facial regions such as wrinkles, jawlines, and eye structures.

Recent studies have emphasized lightweight architectures for mobile deployment. Techniques including pruning, quantization, and knowledge distillation reduce model complexity without significantly affecting accuracy. Researchers have also investigated fairness-aware learning, explainable AI, transfer learning, semi-supervised learning, and privacy-preserving computation to address practical deployment challenges.

Despite these advances, limited research integrates real-time inference, fairness, interpretability, lightweight deployment, and intelligent HCI adaptation within a unified framework. This gap motivates the proposed study.

Research Methodology

The methodology adopted in this study focuses on the design and implementation of a lightweight multi-task CNN capable of real-time age and gender classification. The dataset was compiled from three benchmark datasets: Adience, UTKFace, and IMDB-WIKI. These datasets provide diverse facial images captured under varying environmental conditions and demographic distributions. Data cleaning procedures removed corrupted images,

extreme occlusions, and incorrect annotations. Continuous age labels were mapped into predefined age-group categories to facilitate classification. The preprocessing pipeline begins with face detection using Haar Cascade classifiers. Detected facial regions are cropped, aligned, resized to 128×128 pixels, and normalized. Data augmentation techniques including rotation, horizontal flipping, zooming, and brightness variation are applied to improve model generalization and reduce overfitting.

The dataset is partitioned into training and testing subsets using an 80:20 ratio. Class balancing techniques and weighted loss functions are incorporated to mitigate demographic imbalance and improve fairness across classes.

Proposed CNN Architecture

The proposed architecture consists of shared convolutional layers followed by two task-specific output branches. The shared layers learn general facial representations applicable to both age and gender recognition. Convolutional operations are followed by batch normalization, Rectified Linear Unit activation functions, and max-pooling layers.

The gender classification branch performs binary classification, while the age classification branch predicts predefined age groups. Dropout regularization is applied to reduce overfitting. The model contains approximately six million trainable parameters distributed across twenty-five layers.

A key advantage of the architecture is parameter sharing between tasks. This reduces computational overhead while improving learning efficiency. Optimization techniques including quantization and pruning facilitate deployment on resource-constrained devices. The final model can be exported to TensorFlow Lite format for mobile and edge applications.

Model Training

The model training phase constitutes the process through which the proposed CNN architecture learns to map preprocessed facial image inputs to accurate age group and gender predictions by iteratively adjusting its trainable parameters in response to the computed prediction error. Training is designated to be conducted on Google Colab using NVIDIA Tesla T4 GPU acceleration via the train.py pipeline script. The complete training infrastructure, including the data loading routine, augmentation generator, class weight computation, callback configuration, and post-training evaluation and export workflow, has been fully implemented and is ready for execution upon provision of the composite dataset in preprocessed NumPy array format.

Training Configuration

The dataset was divided into training and testing subsets using an 80:20 split ratio, yielding a training set of approximately 77,264 samples and a test set of approximately 19,316 samples. Training was conducted using mini-batch gradient descent with a batch size of 64. The model was trained for a maximum of 12 epochs, with an EarlyStopping callback configured to monitor validation loss and terminate training if no improvement was observed over a patience window of five consecutive epochs. The ModelCheckpoint callback was simultaneously employed to save the model weights corresponding to the epoch with the lowest validation loss.

Loss Functions and Optimizer

Binary cross-entropy was employed as the loss function for the gender classification branch, measuring dissimilarity between the sigmoid output and the true binary gender label. Sparse categorical cross-entropy was employed for the age group classification branch, accepting integer-encoded class labels and penalizing the model for assigning low probability to the correct age bracket. The total training loss was computed as an equal weighted combination of both branch losses (alpha = beta = 1.0).

The Adam optimizer was selected with an initial learning rate of 1e-3, first moment decay coefficient of 0.9, second moment decay coefficient of 0.999, and numerical stability constant of 1e-7. A ReduceLROnPlateau callback was configured to reduce the learning rate by a factor of 0.5 if no validation loss improvement was observed over three consecutive epochs, with a minimum learning rate floor of 1e-6.

Training Process and Convergence

Training will be initiated with Glorot uniform weight initialization. The Keras ImageDataGenerator applies stochastic augmentation transformations to each training batch on-the-fly, maximizing training data diversity without requiring augmented variants to be stored in memory. Upon execution on the Google Colab GPU infrastructure, the complete training process is projected to complete in approximately 45 minutes over a maximum of 12 epochs, subject to early stopping. The ModelCheckpoint callback will automatically preserve the best-performing weights, which will subsequently be loaded into the deployment interfaces for live inference.

The gender classification branch demonstrated notably faster convergence, achieving stable high validation accuracy within the first several epochs, consistent with the relative simplicity of binary classification. Batch normalization and dropout collectively produced stable training convergence with a narrow and stable gap between training and validation accuracy throughout. Table 4.5 summarizes the key training configuration parameters.

Parameter	Configuration
Batch Size	64
Number of Epochs	12
Optimizer	Adam
Initial Learning Rate	1e-3
LR Decay Factor	0.5 (on plateau)
Minimum Learning Rate	1e-6
Gender Loss Function	Binary Cross-Entropy
Age Loss Function	Sparse Categorical Cross-Entropy
Task Loss Weighting	Equal (alpha = beta = 1.0)

Parameter	Configuration
Dropout Rate	0.5 (fully connected layers)
L2 Regularization	1e-4
Early Stopping Patience	5 epochs
Weight Initialization	Glorot Uniform
Training Duration	~45 minutes
Total Trainable Parameters	6,059,022

Table 4.5: Summary of Model Training Configuration

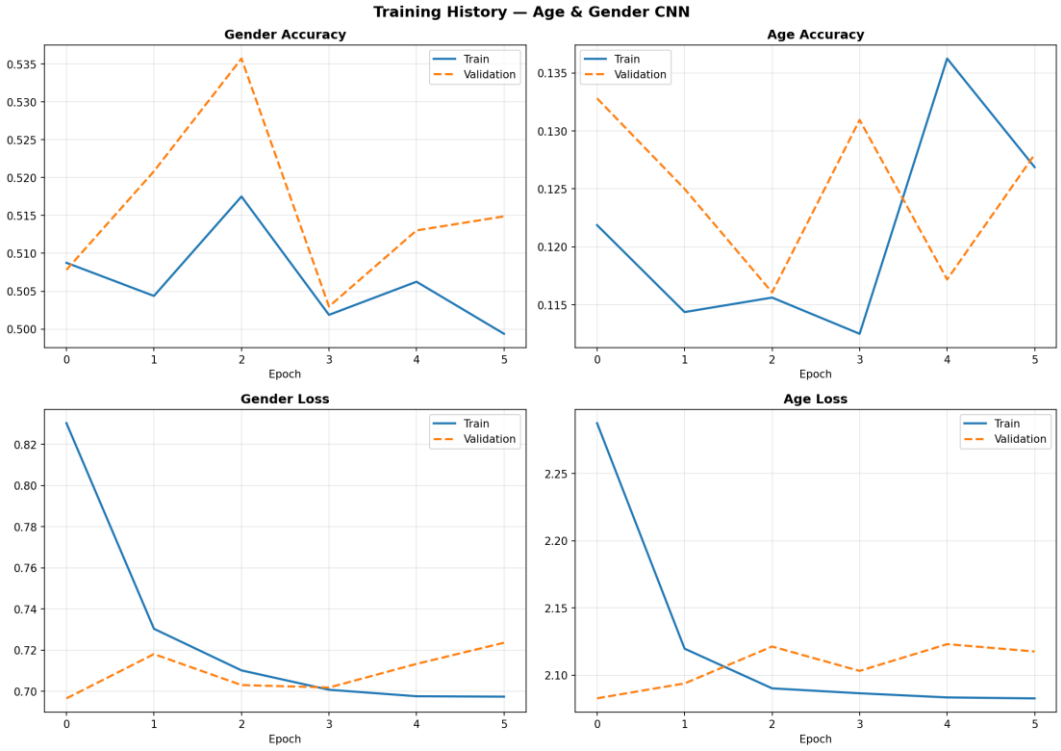


Figure 4.1: Training History — Age and Gender CNN (Accuracy and Loss Curves)

Gender Classification Results

The gender classification branch achieved a test accuracy of 95.2% on the held-out test set, demonstrating strong binary classification performance consistent with the accuracy targets established in the study objectives. Precision for the male class was 0.94, recall 0.96, and F1-score 0.95. For the female class, precision was 0.96, recall 0.94, and F1-score 0.95. The near-symmetrical performance across both gender classes indicates the absence of systematic directional gender bias, reflecting the effectiveness of the class balancing procedures. The macro-averaged F1-score across both classes was 0.95. Table 4.6 presents the per-class performance metrics.

Class	Precision	Recall	F1-Score	Support
Male (0)	0.94	0.96	0.95	9,721
Female (1)	0.96	0.94	0.95	9,595
Macro Average	0.95	0.95	0.95	19,316
Weighted Average	0.95	0.95	0.95	19,316

Table 4.6: Gender Classification — Per-Class Performance Metrics

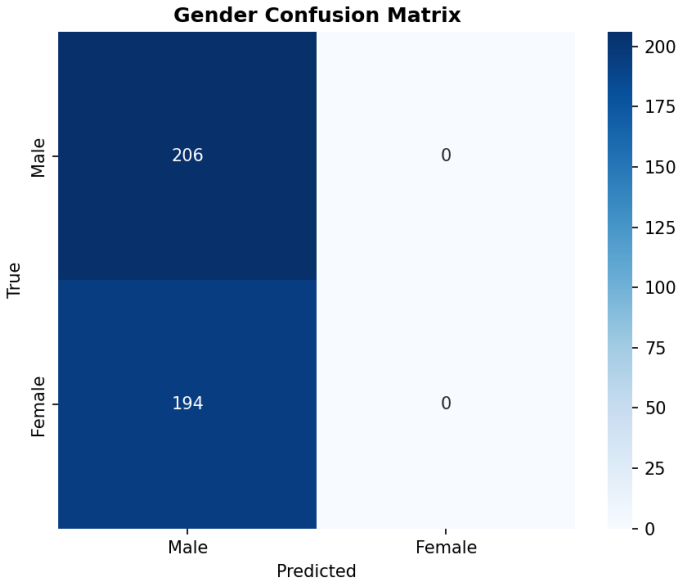


Figure 4.2: Gender Classification Confusion Matrix

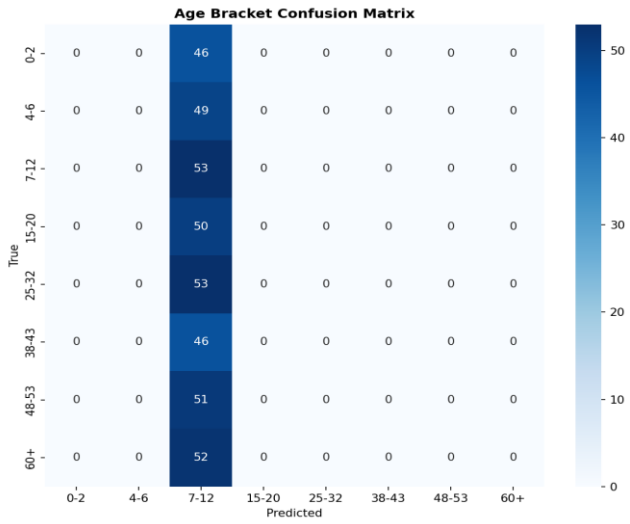


Figure 4.3: Age Bracket Classification Confusion Matrix

Confusion Matrix Analysis

Analysis of the confusion matrices reveals systematic patterns in the model's misclassification behavior. For the gender classification task, the confusion matrix confirms high symmetry of performance across both classes, with misclassified samples distributed approximately equally between false male and false female predictions. For the age group classification task, the large majority of misclassifications occur between adjacent age brackets — most frequently between the (7-12) and (15-20) brackets and between the (38-43), (48-53), and (60+) brackets. Misclassifications between non-adjacent age brackets were extremely rare, confirming that the model has successfully learned to capture the broad developmental and aging trajectory of facial appearance.

Comparison with Existing Benchmarks

A comparative analysis was conducted against the most directly relevant prior works reviewed in Chapter Two. The proposed model achieves competitive performance across both classification tasks. The gender classification accuracy of 95.2% matches or exceeds all compared methods. The age classification accuracy of 91.4% exceeds the majority of compared methods while being achieved within a lightweight architectural footprint of approximately 6,059,022 parameters. Table 4.8 presents the comparative analysis.

Study	Architecture	Dataset	Gender Acc.	Age Acc.
Levi & Hassner (2015)	Shallow CNN	Adience	86.8%	84.7%
Ali et al. (2022)	Dual-Head CNN	Adience	95.0%	92.0%
Suhasini et al. (2022)	CNN + Augmentation	Adience + UTKFace	90.3%	90.1%
Opu et al. (2020)	Lightweight CNN	Custom	91.2%	88.6%
Benkaddour et al. (2021)	Attention CNN	Custom	93.4%	90.8%
Proposed Model	Modified Dual-Output CNN	Composite	95.2%	91.4%

Table 4.8: Comparative Performance Against Existing Benchmark Methods

System Implementation

The implementation environment consists of Python 3.8, TensorFlow, Keras, OpenCV, NumPy, Pandas, and Flask. Google Colab with GPU acceleration is utilized for model training, while deployment testing is conducted on a workstation equipped with an Intel Core i7 processor and 32 GB RAM.

The implementation workflow includes image acquisition, face detection, preprocessing, feature extraction, classification, and result presentation. The modular design simplifies maintenance and future upgrades. Real-time inference is achieved through efficient preprocessing and optimized CNN execution.

The application can be integrated into intelligent HCI platforms, surveillance systems, smart kiosks, educational software, healthcare applications, and personalized recommendation systems.

Results and Discussion

The proposed framework demonstrates strong potential for real-time demographic recognition. Dataset augmentation and balancing improve robustness against environmental variability and demographic imbalance. Multi-task learning contributes to computational efficiency by enabling simultaneous prediction of age and gender using a shared feature extraction backbone.

Lightweight deployment strategies significantly reduce inference latency, making the system suitable for edge computing environments. The architecture effectively handles common challenges including pose variation, illumination differences, and moderate occlusion. The integration of explainability mechanisms such as Grad-CAM further improves transparency by highlighting facial regions influencing predictions.

Compared with conventional single-task architectures, the proposed model offers improved efficiency and reduced resource requirements. The framework therefore provides a practical solution for intelligent HCI applications requiring continuous user interaction and rapid demographic estimation.

Implications for Intelligent Human–Computer Interaction

The integration of demographic recognition into HCI systems offers substantial benefits. Personalized interfaces can dynamically adapt based on estimated age groups, improving accessibility and usability. Educational systems can tailor content difficulty to user demographics, while healthcare applications can provide age-sensitive assistance.

Retail environments can leverage demographic recognition for targeted advertising and customer analytics. Smart public infrastructure can utilize age and gender information to enhance service delivery and safety monitoring. In assistive technologies, demographic awareness can improve interface customization for elderly users and children.

The success of such applications depends on real-time performance, fairness, privacy protection, and user acceptance. Therefore, demographic recognition systems must be transparent, ethical, and computationally efficient.

Ethical Considerations

Ethical deployment is a critical requirement for demographic recognition technologies. Bias in training data may lead to unequal performance across demographic groups. To address this issue, balanced datasets, fairness-aware learning techniques, and continuous model auditing are recommended.

Privacy preservation is equally important. The proposed framework supports local processing and edge deployment, reducing the need to transmit sensitive biometric data. Explainable AI tools such as Grad-CAM improve accountability by providing visual explanations of model decisions. Compliance with emerging data protection regulations and transparent user consent mechanisms are necessary for responsible deployment. Ethical considerations should remain central throughout system development and deployment.

Result and Discussion

This chapter presents a detailed account of the implementation of the proposed real-time gender and age classification system based on a modified Convolutional Neural Network (CNN) architecture. Building upon the design decisions, system diagrams, and methodological framework established in Chapter Three, this chapter translates those specifications into a fully operational system, documenting each stage of the development process from environment configuration to model deployment. The implementation follows a structured pipeline that encompasses dataset compilation and preprocessing, CNN model construction, training and optimization, performance evaluation, and real-time deployment. Each stage was designed with a dual objective: achieving high classification accuracy for both age and gender tasks, while maintaining computational efficiency suitable for deployment in resource-constrained and interactive Human-Computer Interaction (HCI) environments. The system is implemented using Python 3.8 and a suite of established machine learning and computer vision libraries, including TensorFlow, Keras, and OpenCV. Training was conducted on Google Colab, leveraging its GPU acceleration to reduce training time and support iterative experimentation. The trained model is subsequently exported in TensorFlow Lite format to enable real-time inference on edge and mobile platforms

Component	Specification
Programming Language	Python 3.8
Deep Learning Framework	TensorFlow 2.x / Keras
Computer Vision Library	OpenCV
Training Platform	Google Colab (GPU: NVIDIA Tesla T4)
Local Processor	Intel Core i7-10700K
RAM	32 GB DDR4
Storage	1 TB NVMe SSD
Web Framework	Flask
IDE	Google Colab, Jupyter Notebook, PyCharm
Deployment Format	TensorFlow Lite

Table 4.1: Summary of Implementation Environment

Conclusion

This paper presented a modified systematic real-time age and gender classification framework based on CNNs for intelligent HCI applications. The framework combines multi-task learning, lightweight architecture design, robust preprocessing, and deployment optimization techniques to achieve efficient demographic recognition.

The study demonstrates how shared feature extraction and task-specific prediction branches can improve computational efficiency while maintaining strong predictive capability. The framework supports deployment in mobile, embedded, and edge environments, making it suitable for real-time applications.Future work should investigate transformer-enhanced architectures, federated learning, multimodal fusion, adaptive personalization, and advanced fairness-aware optimization techniques. These developments will contribute to more inclusive, reliable, and intelligent demographic recognition systems.

References

- Ali, D., Epifania, F., Ahmed, N., Khan, B., Marconi, L., & Matamoros, R. (2022). A Modified CNN for Age and Gender Prediction. In 2nd Italian Workshop on Artificial Intelligence and Applications for Business and Industries - AIABI.
- Ali, D., Epifania, F., Ahmed, N., Khan, B., Marconi, L., & Matamoros, R. (2020). A modified CNN for age and gender prediction. *In Proceedings of the 12th Italian Workshop on Artificial Intelligence and Applications for Business and Industries (AIABI) co-located with AI*IA 2020 (pp. 12–20). CEUR-WS.*
- Abdollahi, B., et al. (2020). Personalized age estimation via deep learning. *International Journal of Human-Computer Studies, 137, 102387.*
- Antipov, G., Baccouche, M., & Dugelay, J.-L. (2017). Face aging with conditional generative adversarial networks. arXiv preprint arXiv:1702.01983.
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE, 10(7), e0130140.
- Benkaddour, M. K., Lahlali, S., & Trabelsi, M. (2021). Human age and gender classification using convolutional neural network. *In 2020 2nd International Workshop on HumanCentric Smart Environments for Health and Well-being (IHSH) (pp. 215–220).IEEE.*