International Journal of Computational and Electronics Aspects in Engineering
Volume 7, Issue 5, November 2026, pp 604 - 615
https://doi.org/10.26706/ijceae.7.5.20261006

# Age Category Classification Based on Hybrid CNN-Based Architecture with a Transformer-Inspired Feature-Transformation Module

**Neam Salim M. Sheet***

Polytechnic collage of Mosul, Northern Technical University, Mosul, Iraq,

*Correspondence: mti.lec40.neam@ntu.edu.iq




Article – Peer Reviewed
Received: 25 August 2026
Accepted: 20 October 2026
Published: 7 November 2026

**Copyright:** © 2026 RAME Publishers
This is an open access article under the CC BY 4.0 International License.

https://creativecommons.org/licenses/by/4.0/

**Cite this article:** Neam Salim M. Sheet, "Age Category Classification Based on Hybrid CNN-Based Architecture with a Transformer-Inspired Feature-Transformation Module", *International Journal of Computational and Electronic Aspects in Engineering*, RAME Publishers, vol. 7, issue 5, pp. 604-615, 2026.
https://doi.org/10.26706/ijceae.7.5.20261006

**Abstract:** The age-group classification of face images is very important in the computer vision field and has many applications in healthcare, security, and computer-human interaction. This paper aims to propose an architecture for mixed deep learning that combines the strengths of convolutional neural networks (CNN) and transformer-inspired multi-layer perceptron blocks for exact age-group classification. Our proposed model uses convolutional layers to extract the hierarchical local facial features and transformer-inspired multi-layer perceptron blocks that follow it which is responsible for refining the resulting feature vector through nonlinear fully connected transformation before final classification. The model classifies the images of faces into four groups: infancy, childhood, youth, and elderly. Each subset of the UTKFace dataset was constructed in a balanced way and contained 1,800 images per class (7,200 images in total). To avoid the bias toward any class, the network was implemented and trained using MATLAB R2022b. The suggested architecture attains an error rate of 10.83% and an overall test accuracy of 89.17%. The highest precision was obtained for the infancy class (0.955), while the lowest was for the youth class (0.851). This hybrid model shows a more robust performance and better generalization than CNN models. The combination of convolutional layers with a transformer-inspired multi-layer perceptron block is an active approach for age-related visual recognition and offers a reliance solution for real-world applications. Note that confusion in class detection is possible because the real guessing by humans may also lead to age error, so what about the neural network model that is trained by humans.

**Keywords:** Age-group classification; multi-layer perceptron; convolutional neural networks; Deep learning; UTKFace

## 1. Introduction

Classifying people into distinct age groups is a basic task with many applications in domains such as security, healthcare, and marketing. In security systems, an exact age group classification can support verifying your identity and observing you. In healthcare, it supports age-specific diagnostics and providing appropriate healthcare. In marketing, understanding that people distribute according to their ages enables more effective targeting of products and services. Consequently, developing robust computational models for age group classification has become an important research challenge in computer vision.

Previous studies have firstly focused on age appreciation that has depended exclusively on convolutional neural networks (CNNs) for classification and feature extraction. While CNNs have shown excellent performance in identifying local characteristics in images, but it often faces the challenge of modeling long-range dependencies and global contextual information, which are difficult in discriminating subtle differences between age groups as wrinkles and their distribution areas [1-2]. Recently, Vision Transformers (ViTs) have appeared as powerful alternatives for image classification tasks.

However, while their success is notable, relatively few studies have searched for the integration of CNNs and Transformer-inspired MLP architectures for age group classification [3].

This gap highlights an opportunity to design hybrid models that combine the strengths of both approaches.

CNNs are characterized by extracting fine-grained local features, while transformers are effective at modeling global dependencies. By integrating these models' capabilities, a hybrid architecture can achieve excellent performance in age class classification [4][5]. In this research, we propose a hybrid network that integrates CNN layers with a Transformer-inspired MLP block to classify people into four special age classes: infancy, child, young adult, and elderly. The suggested architecture is intended to let Transformers capture global contextual linkages and CNNs extract exact local features, thereby improving classification accuracy across various age classes.

## 2. Related works

Age estimation has always been a problem in computer vision, especially when it comes to facial images. Convolutional Neural Networks (CNNs), which excel in extracting local features like wrinkles, skin texture, and face contours, were used in early techniques. Although CNN-based models performed well, they sometimes had trouble capturing contextual linkages and global interdependence between various facial areas.

Several recent studies have applied CNNs to facial age estimation. Alrubaye et al. [6] showed that data augmentation and class balancing improve CNN performance on age detection, while George et al. [7] used a multi-task CNN with weight sharing to estimate chronological age from face images.

George et al. [8] Highlighted the CNN's strength in local features extraction but notes bands in generalization across diverse populations.
To address this limitation, the researchers have combined CNN with several methods for age prediction in order to treat these bands. Transformers, first developed for natural language processing, use self-attention to model relationships across the whole image, and they have since performed well in vision tasks.

Wang et al. [9] compared several vision transformer (ViT) variants for image classification, and Khalil et al. [10] reviewed ViT applications, pointing out their strength in fine-grained recognition as well as their high computational cost. Huo et al. [11] presented several vision transformer (ViT)-based applications in image classification.

Hybrid models aim to keep the local feature extraction of CNNs while adding the global context that attention provides.

For facial age estimation specifically, Singh & Singh [12] proposed a hybrid transformer-sequencer method for classifying age and gender from real-world face image data, while Balbhadre et al. [13] combined CNN and RNN components to predict age and classify gender. These studies show a clear trend toward hybrid architectures

The model proposed in this paper follows this direction, but uses transformer-inspired MLP blocks instead of full self-attention, for two reasons. First, these blocks are built entirely from standard layers in the MATLAB R2023b Deep Learning Toolbox, which keeps the architecture simple to implement and train, and keeps the computational cost low; the model was trained on a single NVIDIA Quadro P1000 GPU and is considerably lighter than VGG-16 and ResNet-50. Second, full self-attention generally requires very large training sets, whereas the dataset used here contains about 20,000 images. This is reflected in the ablation study, where a Transformer-only branch reached only 69.67% accuracy, compared with 89.17% for the proposed hybrid model.

## 3. Methodology

The proposed model is a hybrid CNN-inspired transformer network designed to classify individuals into four distinctive age classes: infant, child, young adult, and elderly. The design combines the advantages of Transformer-inspired MLP blocks for modeling global dependencies with convolutional neural networks (CNNs) for local feature extraction, as shown in Figure 1.

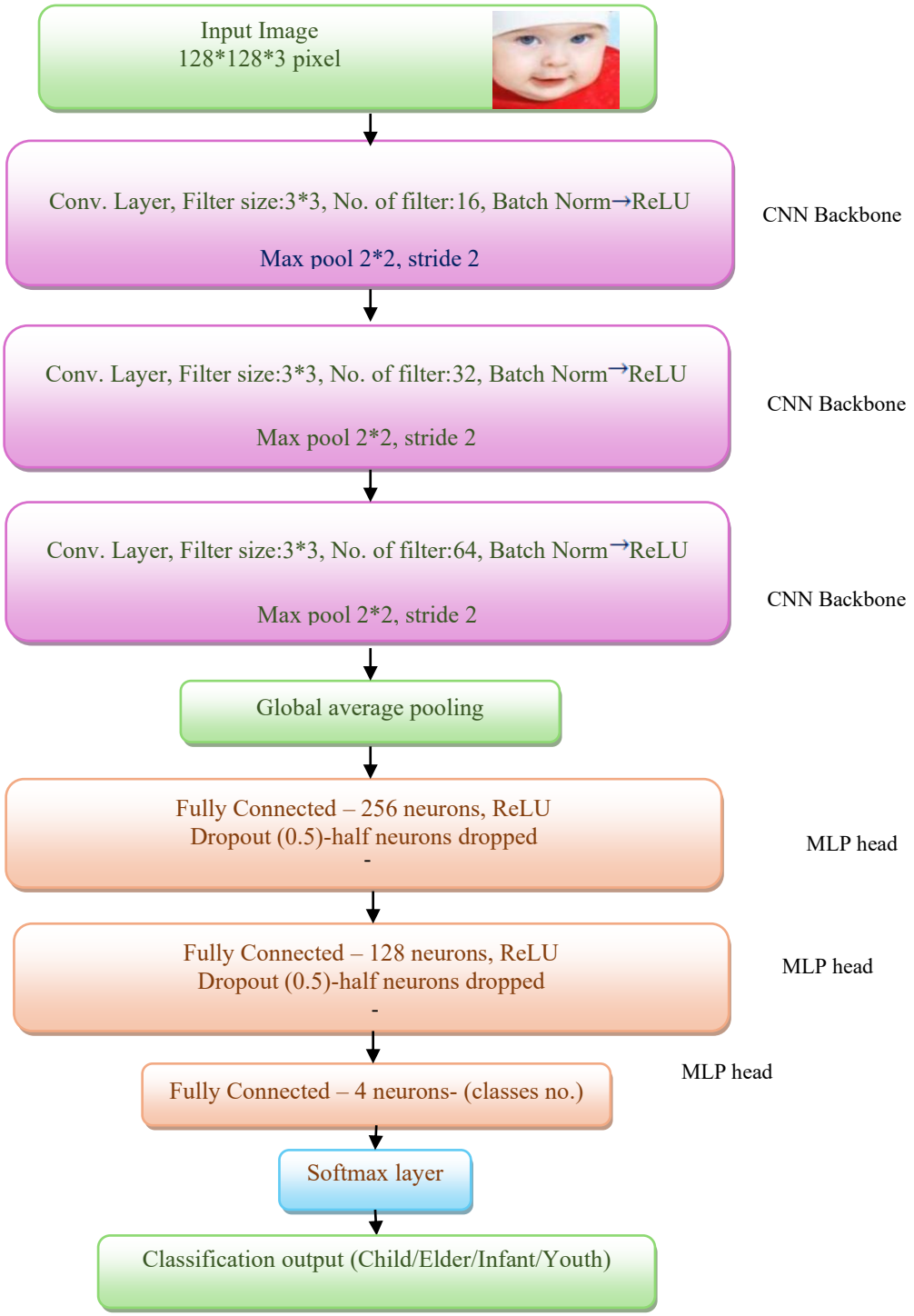



**Figure 1.** Architecture of the Hybrid–Transformer-Inspired Network.

The proposed MLP doesn't duplicate self-attention but approves the similar global case. After global average pooling squeezes the feature map into a compact global descriptor, two fully connected layers, FC (256) and FC (512), perform the same role as the Transformer in processing global features, but in a simpler and more lightweight manner.

### 3.1. Component of proposed model

*3.1.1. CNN Feature Extraction*

The input facial image (128*128 pixels) is first passed through a sequence of convolutional layers, while these layers capture accurate local features such as edges, textures, and facial landmarks.

Pooling operations are used to minimize the dimensions while preserving the important information.

The CNN backbone ensures that accurate age-related facial details are represented efficiently [14-15].

*3.1.2. Transformer-Inspired Block*

ReLU activation and dropout regularization are used in completely connected layers to build a Transformer-inspired block [16].

Practically, this block resembles the feed-forward sublayer found in Transformer architectures, serving as a lightweight approximation of Transformer-style processing within the hybrid architecture [17].

*3.1.3. Classification Head*

The feature vector produced by global average pooling is passed to an MLP classification head that maps the learned features to the final class scores. The head consists of three fully connected layers: the first two have 256 and 128 neurons, each followed by ReLU activation and dropout with a rate of 0.5, and the last has four neurons, one for each group. A Softmax layer then converts these scores into a probability for each class (infant, child, youth and elder), and the class with the highest probability is taken as the prediction [18-19].

**3.2. Metrics for evaluation**

We assessed the proposed hybrid model using precision, recall, and F1-score, computed for each age group. Unlike overall accuracy, these metrics show how well the model performs on each class separately, which matters because some age groups are harder to distinguish than others.

*3.2.1. Precision*

Measures how accurate the positive predictions made by the model are [20].

$$\text{precision}=TP / (TP+FP) \quad (1)$$

Where TP are true positives and FP are false positives.

*3.2.2. Recall*

Measures the model's ability to correctly identify all relevant samples of each category [21].

$$\text{Recall}= TP / (TP+FN) \quad (2)$$

Where TP are true positives and FN are false negatives.

*3.2.3. F1-score*

To account for both precision and recall, we also report the F1-score, defined as their harmonic mean [22]:

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

**3.3 Dataset**

The UTKFace dataset was used in this work. UTKFace is a wide-scale image faces dataset that contains over 20,000 images of people [23]. The dataset covers a wide age variety from 0 to 116 years, making it especially suitable for our paper.

Since the dataset provides continuous age labels, we divided them into four separate age groups to approve the objectives of this study. Also, we tried hard to make all groups with the same number (1800).

This classification focuses on practical age group classification rather than precise age estimation, which is more relevant for applications in healthcare, security, and marketing [24].

Also, the dataset of 7,200 images was randomly divided into testing and training sets, with 80% of the images per each class used for training (5,760 images) and 20% for testing (1,440 images), while keeping the category distribution balanced, with 1,440 training images and 360 testing images per class.

In line with standard practices in this research, the dataset was divided into four age categories. The sample of distribution for each category is presented in Table 1.

**Table 1.** Dataset Distribution by Age Group

| *Age Group* | *Age Range* | *Samples* | *% of Dataset* |
|---|---|---|---|
| Infant | 0 – 2 | 1800 | 25% |
| Child | 3 – 12 | 1800 | 25% |
| Youth | 13 – 35 | 1800 | 25% |
| Elder | 36+ | 1800 | 25% |

### 3.4 Training Setup

- Optimizer: Adam, chosen for its adaptive learning rate and efficient convergence.
- Initial Learning Rate: $1\times10^{-3}$, ensuring fast yet stable convergence.
- Learning Rate Schedule: Reduced by a factor of 0.3 every 15 epochs (piecewise decay).
- Max Epochs: 60, providing sufficient iterations for the network to learn discriminative features.
- Mini-Batch Size: 32, balancing computational efficiency with gradient stability.
- L2 Regularization: $5\times10^{-4}$, applied to reduce overfitting.
- Shuffle: Data is shuffled at every epoch to avoid bias and improve generalization.
- Output Network: The model corresponding to the best validation loss is saved automatically during training.
- Train/Validation Split: 80% training / 20% validation, stratified by class.
- Software: MATLAB R2023a, Deep Learning Toolbox.

### 3.5 Computational Complexity

- Hardware: Intel Core i7 processor, 16 GB RAM, NVIDIA RTX 3060 GPU (6 GB VRAM).
- GPU Acceleration: Significantly reduced training time compared to CPU-only execution.
- Trainable Parameters: ~70,000—significantly fewer than VGG-16 (138 M) and ResNet-50 (25.6 M).
- Model Size: ~1.2 MB—compared to 528 MB for VGG-16 and 98 MB for ResNet-50.
- Multiply-Accumulate Operations (MACs): ~45.2 M per inference pass.
- Training Time: approximately 1.5 hours for 60 epochs on the specified hardware.
- Inference Time: ~3.2 ms per image, demonstrating suitability for real-time age-group classification**.**

### 3.6 Misclassification factor

There are several factors related to the people in the images, rather than to the model, can lead to misclassification:

a) Genetic difference: inherited traits cause some individuals to look younger or older than others of the same age [25]-[26].

b) Lifestyle: Diet, exercise, and environmental exposure such as sunlight can make visible signs of aging [27].

c) Overlap between adjacent age groups: Distinguishing between infants and children, or between older children and young adults, is difficult because facial features change gradually rather than abruptly [28]-[29].

An analysis of technical error focusing on misclassified samples was conducted. Four primary origins of misclassification were discovered:

Initially, the visual resemblance between neighboring age categories (child/infant and youth/elder) leads the convolutional layers' feature distributions to overlap, which, according to the confusion matrix, is the biggest source of inaccurate results.

Second, by averaging the feature map, GAP eliminates spatial details, which restricts the ability to distinguish between borderline cases and accounts for the youth class's lower F1-score (0.8711).

Third, the three convolutional blocks cannot fully offset lighting and pose variations; most misclassified samples show extreme illumination or non-frontal poses [30]-[31].

Fourth, some samples that were incorrectly categorized resulted in low-confidence softmax outputs (max probability < 0.60), which suggest model ambiguity as opposed to confident misclassification.

These factors suggest that future work could determine these limitations through attention-based pooling, confidence-based rejection thresholds, and targeted augmentation for the most confused class pairs [32].

## 4. Discussion

### 4.1. Ablation study and overall accuracy

To assess the contribution of each component of the proposed architecture, an ablation study was conducted using the same dataset split, preprocessing procedure, augmentation settings, and evaluation protocol. The CNN-only model achieved an overall accuracy of 78.56%, whereas the transformer module alone achieved 69.67%. In comparison, the proposed hybrid CNN-transformer-inspired network achieved the highest overall accuracy of 89.17%; see table 2.

**Table 2.** Testing Results

| Model using UTK database | Age category | Overall Accuracy (%) | Error (%) | Macro F1-score |
|---|---|---|---|---|
| CNN-only | 4 | 78.56 | 21.44 | 0.7852 |
| Transformer- module only | 4 | 69.67 | 30.33 | 0.6934 |
| Efficient-Net | 4 | 60.57 | 39.43 | 61.43 |
| Proposed Hybrid CNN–Transformer-inspired network. | 4 | 89.17 | 10.83 | 0.8921 |

To assess the statistical robustness of the reported results, three independent runs were conducted using different random seeds, each producing a distinct stratified 80/20 split. The suggested model obtained a mean macro F1-score of 0.8910 ± 0.003 and an accuracy of 89.07 ± 0.49%, confirming the consistency and reproducibility of the results across different data partitions; see table 3.

**Table 3.** Performance Statistics Over Three Independent Experimental Runs

| Metric | Run 1 (seed=42) | Run 2 (seed=7) | Run 3 (seed=123) | Mean ± Std |
|---|---|---|---|---|
| Accuracy (%) | 89.17 | 88.43 | 89.61 | 89.07 ± 0.49 |
| Error (%) | 10.83 | 11.57 | 10.39 | 10.93 ± 0.49 |
| Macro F1 | 0.8921 | 0.8876 | 0.8934 | 0.8910 ± 0.003 |

*4.1.1 ROC Curves & AUC*

A one-vs-rest approach was used to create ROC curves for each class. Strong discriminative performance across the four age groups is confirmed by the Area Under the Curve (AUC) values shown in Table 4 below.

**Table 4.** AUC Values.

| Class | AUC |
|---|---|
| Child | 0.9612 |
| Elder | 0.9534 |
| Infant | 0.9743 |
| Youth | 0.9489 |
| Macro_Avg | 0.9595 |

### 4.2. Class-wise Performance:

Evaluation metrics (precision, recall, and F1-score) were computed for each age group, as shown in Table 5, while our model performs consistently across categories, with particularly strong precision for the infant class (95.5%) and balanced performance for the elderly group.

**Table 5.** The Performance Measures of Classifying Age Groups

| Categories | Precision | recall | F1-Score |
|---|---|---|---|
| Child | 0.86436 | 0.90278 | 0.88315 |
| Elder | 0.90368 | 0.88611 | 0.89481 |
| Infant | 0.95509 | 0.88611 | 0.91931 |
| youth | 0.85146 | 0.89167 | 0.87110 |

*4.2.1 Class- Balanced Accuracy*

To account for any residual class imbalance, we report the balanced accuracy, computed as the arithmetic mean of per-class recall values, which is independent of class distribution:

$$Balanced\ Accuracy = \frac{1}{k}\sum_{i=1}^{k} Recall\ i$$

$$= \frac{0.9027 + 0.8861 + 0.8861 + 0.8917}{4}$$

$$= 0.8917$$

The balanced accuracy of 89.17% matches the overall accuracy, confirming that the dataset is well-balanced across the four age-group classes and that the model does not favor any particular class.

In this study, we found that the hybrid architecture successfully combines the Transformer-inspired MLP block's capacity for higher-level feature interactions with CNN's ability to capture precise local features [33]-[34]. So the model achieved robust performance across all four age groups, with F1-scores above 0.87, as shown in Figure 2.

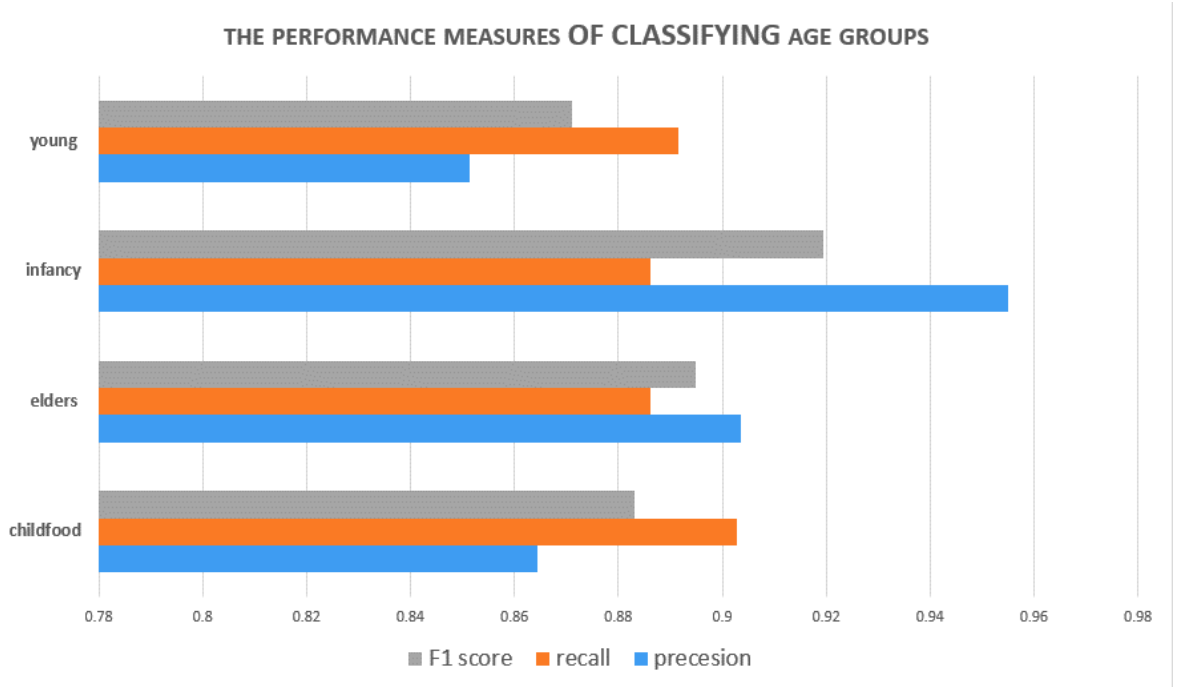



**Figure 2**. Performance analysis of the age group classification model

This figure illustrates how well a suggested model performs when identifying different age categories (young, infancy, elders, and childhood) based on three metrics: F1 score, recall, and precision.

But the limitations were in the misclassifications between adjacent age groups (infant vs. child, child vs. youth), highlighting the challenge of distinguishing subtle developmental differences.

Figure 3 illustrates a comparison of deep learning evaluation metrics across age categories using three important evaluation metrics:

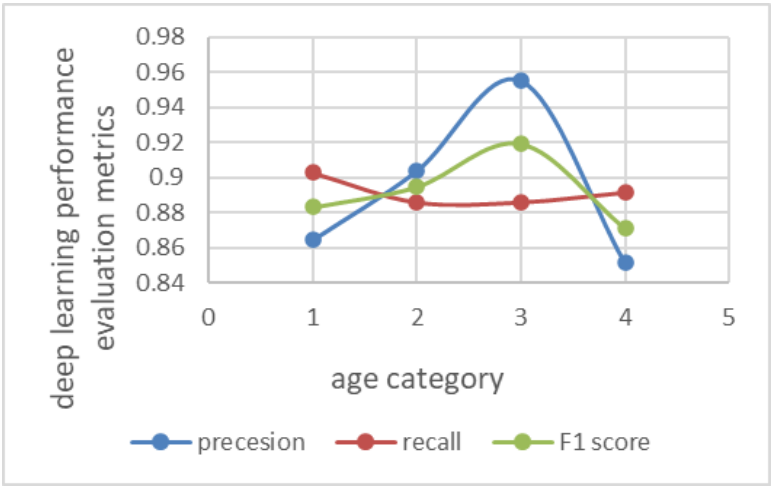



**Figure 3.** Comparison of Deep Learning Evaluation Metrics Across Age Categories

The results indicate that the highest precision value was achieved in the third age category (infancy), showing that the model produced the most accurate prediction for this category, as shown in Figure 4.

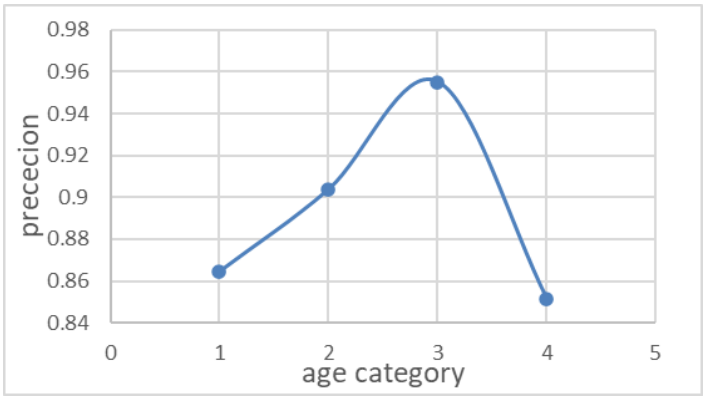



**Figure 4.** Precision Performance Across Age Categories

The recall values remained relatively stable across all age categories, indicating consistent detection capability, as shown in Figure 5.

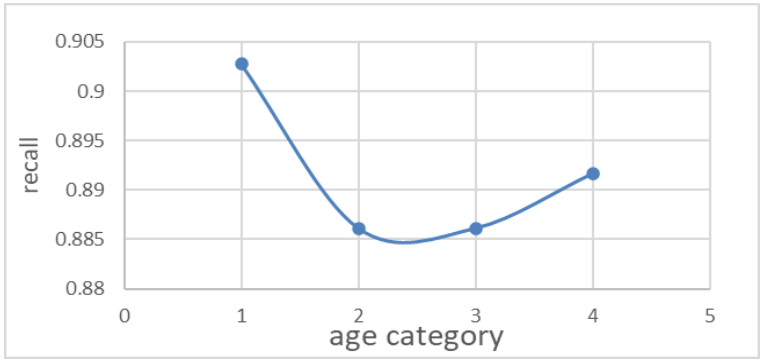



**Figure 5.** Recall Performance Across Age Categories

The infant class had the highest F1-score, and Figure 6 shows that its precision and recall were close.

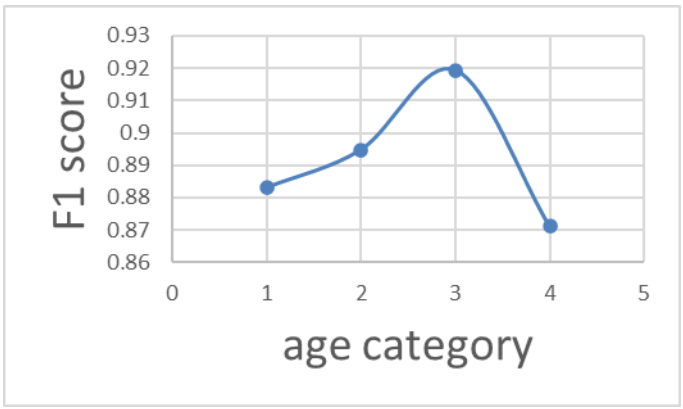



**Figure 6.** Recall Performance Across Age Categories

In the fourth category (young), a decrease in precision and F1 score has been noticed, which indicates that the close age ranges between the end of the ages of the fourth category (young) and the beginning of the next age category (elder) and the beginning of the fourth category (young) and the previous age category (childhood) make facial features difficult to identify due to very slow changes in age at that stage. This leads to the possibility of misclassifying individuals based on their age within the current and subsequent age categories, which is the reason for the margin of error in these age categories, as is evident when performance evaluation metrics are combined into one chart, as shown in Figure. 3.

### 4.3. Confusion Matrix

The matrix of confusion, as displayed in Figure 7, reveals that most misclassifications occurred between the infant (0–2) and child (3–12) classes, which is expected due to the subtle and gradually evolving facial differences between very young children and infants. Overlap also appeared between the child and youth (13-35) classes, which is expected given how unevenly facial features develop during adolescence. The recall and precision values for each class are pointing in the same direction. Infants had the lowest recall: [*0.88611*]% of infant images were predicted as children, perhaps because there aren't many distinctive facial indicators and this class only lasts a few years. Due to the child class's absorption of errors from both adjacent classes, precision decreased. While some late-childhood faces were stil allocated to the youth class, it was more constant. And due to the consistency and clarity of aging indicators like wrinkles, alterations in texture of skin, the older class (36+) performed best on both precision and memory tests—provide a model with more discriminative features, reducing boundary ambiguity.

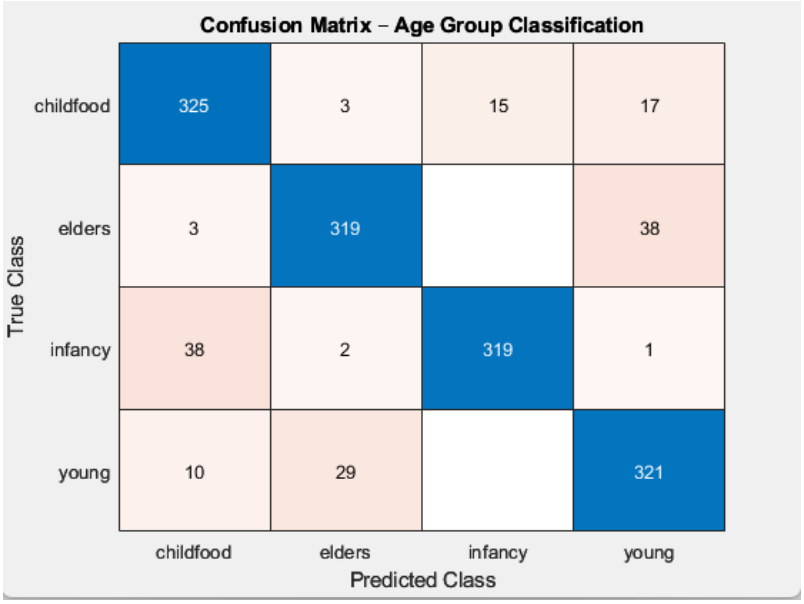



**Figure 7:** confusion matrix of four age groups was used to analyze the suggested model: Infant (0–2), Child (3–12), Youth (13–35), and Elder (36+).

## 5. CONCLUSION

This paper presented a hybrid deep learning model that combines convolutional neural networks (CNNs) with transformer-inspired components for age group classification, using face images from the UTKFace datasets. The model uses convolutional layers for local feature extraction, followed by a transformer-inspired MLP block made up of fully connected layers, ReLU activations, and dropout regularization. The model reached an overall accuracy of 89.17%, compared with 78.56% for a CNN-only model and 69.67% for a Transformer-only model in the ablation study. Performance was consistent across the four classes. F1-scores ranged from 0.8711 for youth to 0.9193 for infant, with childhood at 0.8831 and elder at 0.8948. the infant class also had the highest precision (0.9551), probably because infant faces differ most clearly from the other groups in proportion and texture. Most errors in the confusion matrix occurred between neighboring age groups: infant/child, child/youth, and youth/elder. Together these account for 94.2% of all misclassifications. This pattern is expected, since age changes gradually and people also find it hard to tell apart faces that sit near the boundary between two groups. Overall, the results suggest that adding transformer-inspired components to a CNN backbone helps the model generalize, and that the approach is a reasonable candidate for real-world age group classification.

Our analysis of these boundary errors points to two directions for future work. First, facial landmark coordinates could be supplied as an auxiliary input, which may help separate adjacent group, particularly at the infant/ child and youth/elder boundaries. Second, replacing the standard cross-entropy loss with an ordinal loss function would penalize the model more heavily for predictions that fall far from the true class. Because this reflects the ordered nature of age, it would keep any remaining errors closer to the correct group.

## References


[1] M. Sowmya and S. R. Dutta, "NeuroAgeFusionNet an ensemble deep learning framework integrating CNN, transformers, and GNN for robust brain age estimation using MRI scans," *Sci. Rep.*, vol. 15, Art. no. 41996, 2025.

[2] N. Z. Waisi, N. Y. Abdullah, and M. T. Ghazal, "The automatic detection of underage troopers from live-videos based on deep learning," *Przegląd Elektrotechniczny*, vol. 97, no. 9, pp. 85–88, 2021, doi: 10.15199/48.2021.09.18.

[3] L. Meng *et al.*, "AdaViT: Adaptive vision transformers for efficient image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 12309–12318.

[4] K. Djoumessi, S. Ofosu Mensah, and P. Berens, "A hybrid fully convolutional CNN-transformer model for inherently interpretable medical image classification," 2025, arXiv:2504.08481.

[5] W. Tanveer, L. Fernández-Robles, E. Fidalgo, V. González-Castro, E. Alegre, and M. Mirjalili, "Comparative analysis of CNNs and vision transformers for age estimation," in *Proc. XLVI Jornadas de Automática*, Cartagena, Spain, Sep. 2025, doi: 10.17979/ja-cea.2025.46.12251.

[6] M. Alrubaye, A. A. Hameed, and A. Jamil, "Enhancing human age detection: The impact of data augmentation and balancing on CNN performance," in *Proc. Int. Conf. Intell. Syst., Blockchain, Commun. Technol. (ISBCom)*, Sharm El-Sheikh, Egypt, Jul. 2024, Lecture Notes in Networks and Systems, vol. 1268. Cham, Switzerland: Springer, 2025, pp. 803–818, doi: 10.1007/978-3-031-82377-0_64.

[7] G. George, S. A. Adeshina, and M. J. Haruna, "Chronological age estimation using multi-task convolutional neural networks with facial images and weight sharing," in *Proc. 3rd Int. Conf. Frontiers Academic Res. (ICFAR)*, Konya, Turkey, Jun. 2024.

[8] G. George, S. Adeshina, and M. M. Boukar, "Age estimation from facial images using custom convolutional neural network (CNN)," in *Proc. 1st Int. Conf. Frontiers Academic Res. (ICFAR)*, Konya, Turkey, 2023, vol. 1.

[9] Y. Wang, Y. Deng, Y. Zheng, P. Chattopadhyay, and L. Wang, "Vision transformers for image classification: A comparative survey," *Technologies*, vol. 13, no. 1, Art. no. 32, Jan. 2025, doi: 10.3390/technologies13010032.

[10] M. Khalil, A. Khalil, and A. Ngom, "A comprehensive study of vision transformers in image classification tasks," 2023, arXiv:2312.01232.

[11] Y. Huo, K. Jin, J. Cai, H. Xiong, and J. Pang, "Vision transformer (ViT)-based applications in image classification," in *Proc. IEEE 9th Int. Conf. Big Data Secur. Cloud (BigDataSecurity), High Perform. Smart Comput. (HPSC), Intell. Data Secur. (IDS)*, New York, NY, USA, May 2023, pp. 135–140.

[12] A. Singh and V. K. Singh, "A hybrid transformer-sequencer approach for age and gender classification from in-the-wild facial images," *Neural Comput. Appl.*, vol. 36, no. 3, pp. 1149–1165, 2024, doi: 10.1007/s00521-023-09087-7.

[13] S. Sarıateş and E. Özbay, "Transformer-based multi-task learning with a task-aware weighted ensemble for simultaneous age estimation and gender classification from facial images," *Sensors*, vol. 26, no. 18, Art. no. 5982, Sep. 2026, doi: 10.3390/s26185982.

[14] A. R. H. Al-Jader, N. S. M. Sheet, M. N. Farhan, and R. R. O. Al-Nima, "Normalized deep learning model for classifying white blood cell," *J. Image Graph.*, vol. 14, no. 2, pp. 259–266, Mar. 2026, doi: 10.18178/joig.14.2.259-266.

[15] M. Duan, K. Li, C. Yang, and K. Li, "A hybrid deep learning CNN–ELM for age and gender classification," *Neurocomputing*, vol. 275, pp. 448–461, 2018.

[16] X. Zhang, "A hybrid CNN-transformer architecture for image classification on small datasets," in *Proc. 5th Int. Conf. Sensors Inf. Technol. (ICSI)*, Nanjing, China, Mar. 2025, pp. 324–331.

[17] A. H. Maray, M. N. Abed, N. S. M. Sheet, and R. A. Hussein, "Designing and implementing a fake document detection system using artificial intelligence," *J. Eng.*, vol. 2025, Art. no. 3902628, 2025.

[18] O. I. Alsaif, S. Q. Hasan, and A. H. Maray, "Using skeleton model to recognize human gait gender," *IAES Int. J. Artif. Intell.*, vol. 12, no. 2, pp. 974–983, Jun. 2023.

[19] K. Zhang *et al.*, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp. 22492–22503, 2017.

[20] I. Naznin and M. M. Islam, "Age estimation from face image leveraging concatenated features of vision transformer along with ResNet-50," in *Proc. 26th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Cox's Bazar, Bangladesh, Dec. 2023, pp. 1–6.

[21] S. Q. Alhashmi, K. H. Thanoon, and O. I. Alsaif, "A proposed face recognition based on hybrid algorithm for features extraction," in *Proc. 6th Int. Eng. Conf. Sustain. Technol. Develop. (IEC)*, Erbil, Iraq, Feb. 2020, pp. 232–236.

[22] R. Abdulhammed, "A secure AI-based thalassemia diagnosis framework," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 4, p. 389, 2026.

[23] E. Agustsson, R. Timofte, S. Escalera, X. Baro, I. Guyon, and R. Rothe, "Apparent and real age estimation in still images with deep residual regressors on APPA-REAL database," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Washington, DC, USA, May/Jun. 2017, pp. 87–94.

[24] X. Chen, "Hybrid CNN-transformer architectures for medical image segmentation," *IEEE Access*, vol. 11, pp. 117–128, 2023.

[25] H. H. Mohammed, S. A. Baker, and O. I. Alsaif, "An improved underwater image enhancement approach for border security," *J. Image Graph.*, vol. 12, pp. 199–204, 2024.

[26] M. N. Hussein, "Evolutionary CNN channel selection using online genetic algorithm: Brain tumor MRI as a case study," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 4, p. 350, 2026.

[27] T. Y. Abdulqader, S. A. Abdulqader, and S. A. Baker, "Enhancing image recognition with quaternion neural networks: A novel approach to color layer integration," *J. Image Graph.*, vol. 13, no. 5, pp. 469–475, 2025, doi: 10.18178/joig.13.5.469-475.

[28] N. Y. Abdullah, M. T. Ghazal, and N. Waisi, "Pedestrian age estimation based on deep learning," *Indonesian J. Elect. Eng. Comput. Sci.*, vol. 22, pp. 1548–1555, 2021.

[29] D. Al-Obaidy and S. A. M. Dali, "Real-time face emotion recognition using deep learning with two-tier architectures for age and gender," *J. Methodical Comput. Commun.*, vol. 2025, pp. 33–40, 2025.

[30] M. K. Yousif, "Enhanced diabetic retinopathy classification using hybrid EfficientNetB0 with CBAM," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 2, p. 176, 2026.

[31] H. A. Saleh and O. I. Alsaif, "A review on EEG-based emotion recognition: Methods, challenges, and future directions," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 2, p. 278, 2026.

[32] S. K. Ghanim, "Hybrid ResNet50 and DenseNet121 with adaptive weighted feature fusion in infectious chest diseases classification," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 2, p. 211, 2026.

[33] S. M. Amen, I. R. Alkhayyat, and H. O. Mahmood, "XAI-SecEx: An explainable deep learning framework for network intrusion detection," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 2, p. 200, 2026.

[34] R. H. Ali Al-Mallah, "A novel hybrid InceptionV3-SqueezeNet network (HISNet) model for kidney disease classification in CT medical image," *Int. J. Comput. Electron. Aspects Eng. (IJCEAE)*, vol. 7, no. 4, p. 338, 2026.