

# A Deep CNN technique for Age Estimations from Facial Images

**Gilbert George<sup>1</sup>, Steve A. Adeshina<sup>2</sup>, Moussa Mahammat Boukar<sup>3</sup>**

<sup>1</sup>Department of Computer Science, BAZE University, Abuja, Nigeria

<sup>2</sup>Department of Engineering, Nile University, Abuja, Nigeria

<sup>3</sup>Department of Computer Science, Nile University, Abuja, Nigeria

Corresponding Author: [gilbert.george@bazeuniversity.edu.ng](mailto:gilbert.george@bazeuniversity.edu.ng)

**ABSTRACT** Facial age estimation plays a vital role in various applications within human-computer interaction, such as personalized user experiences, age-specific content filtering, and demographic studies. The process involves analyzing facial images to predict a person's age or categorize them into an age group, and it requires sophisticated methodologies to achieve high accuracy. One of the main challenges in this domain is the need for extensive datasets and long training times to develop accurate predictive models. In this paper, we propose a novel deep learning-based age estimation model using Convolutional Neural Networks (CNN). Our approach is designed to predict ages with high precision, addressing both regression and classification problems in age estimation. Notably, the proposed model demonstrates superior performance with a significantly reduced training data requirement, achieving a low Mean Absolute Error (MAE) of 2.1 years and an impressive accuracy rate of 93%. These results underscore the model's ability to generalize effectively across varying datasets. We implemented and evaluated two custom CNN models tailored for facial age prediction. The first model approaches age estimation as a regression problem, directly predicting the numerical age, while the second model frames it as a classification task, categorizing images into predefined age groups. The performance of our system was benchmarked against existing state-of-the-art methods, and the results highlight its enhanced efficiency and accuracy. Our findings emphasize the practicality and reliability of the proposed method, paving the way for advancements in applications requiring age estimation. This work contributes to reducing computational overhead and improving accuracy, making it an essential step forward in human-computer interaction systems.

**KEYWORDS:** Artificial Neural Network, Computer Vision, Convolutional Neural Network, Facial Age Prediction, Facial Age Group Classification.

## I. INTRODUCTION

Determining a person's age from a photograph has long been one of the most challenging tasks in facial analysis, due to the subtle and diverse factors that influence facial aging. However, advancements in computer vision and pattern recognition have propelled this research area, making automatic age estimation from facial images a growing focus within the field of artificial intelligence. Human facial images contain a wealth of information, such as identity, emotional state, ethnicity, gender, and age—key attributes that play a critical role in social and face-to-face interactions. While humans are naturally skilled at interpreting such cues, replicating this ability in computers presents a unique set of challenges and opportunities. Automated age estimation holds immense importance in real-world applications. It can be used to enhance user experiences by tailoring computing environments to different age groups, enabling more intuitive and adaptive human-computer interactions. In addition, it can regulate access to age-sensitive content online, organize digital photo albums more effectively, and even assist law enforcement by narrowing down suspects in investigations. The retail sector can benefit from age estimation to control access to age-restricted products, such as alcohol or tobacco, through vending machines or self-checkout systems. Moreover, age estimation can support medical diagnostics, particularly in analyzing age-related conditions or predicting biological age. In this study, we propose a novel method to train a Convolutional Neural Network (CNN) for precise age estimation from facial images. Our approach categorizes detected faces into three distinct groups: Child, Teenager, and Adult and predicts the person's chronological age.

Section 2 reviews prior efforts and their limitations, Section 3 outlines our methodology, Section 4 discusses experimental results, and Section 5 concludes with a summary of findings and future work directions.

Current research on age estimation from photographs primarily focuses on classification and regression approaches. In classification [5], age is grouped into discrete classes, whereas regression seeks to predict a precise age[3], although humans are generally better at estimating age ranges rather than exact ages. Traditional learning methods, such as regression, classification, ranking, and label distribution-based approaches, have yielded promising results[6], [7], [8]. The regression method, for instance, models facial aging as a continuous variable, with Euclidean loss as an optimization function, as seen in [9] use of a multi-scale CNN trained with mean squared error. Classification, meanwhile, treats age prediction as a multi-class problem but lacks interclass relation awareness. Ranking CNNs[10] create binary classifiers and aggregate predictions for approximate age, while Label Distribution Learning (LDL) interprets age prediction as a probability distribution, offering high accuracy on benchmarks by leveraging the expected value as the predicted age. However, these models have high complexity and lengthy training times. CNN architectures, known for their computational efficiency, are widely used in applications such as facial analysis, pose measurement, and speech recognition, thanks to their ability to learn critical features automatically. Recent studies emphasize CNN-based models for age and gender prediction with significant accuracy. For example, the wide-ResNet model achieved 96.26% accuracy for age and gender classification, and a CNN-based approach applied on real-world, unfiltered faces managed to classify age groups and gender effectively. Pretrained on IMDB-WIKI, refined on MORPH-II, and validated on OIU-Audience data, this model addressed real-world variability with robust preprocessing.

Historically, facial age detection relied on handcrafted machine learning approaches to manually extract age-related features. Anthropometric, Active Appearance Model (AAM), and Bio-Inspired Features (BIF) techniques were widely used for this purpose. However, CNNs now enable automated high-level feature extraction, enhancing age prediction systems' performance. The COnsistent RANk Logits (CORAL)[1] framework, for instance, supports rank-monotonicity and accurate ratings, achieving a Mean Absolute Error (MAE) of 2.64 across various datasets. Similarly, a ResNet-50 model, applying age estimation as a regression problem, demonstrated superior accuracy with an MAE of 4.49, even with reduced training data. Another advanced CNN model showed 94.01% accuracy for age and 99.86% for gender on the UTKFace dataset[6], surpassing traditional algorithms and proving effective across multiple datasets. This shift to deep learning has propelled age and gender estimation[6], making CNN-based approaches indispensable for automated facial analysis in diverse real-world applications.

## II. METHODOLOGY

To classify age groups accurately in age estimation, we developed a model based on a CNN algorithm. The techniques used are described as follows:

### A. DATASET

This study utilizes the UTKFace dataset [2], which consists of around 20,000 aligned and cropped face images, annotated for age, gender, and ethnicity. We specifically used 11700 facial images. The photos exhibit a range of variations in position, resolution, occlusion, lighting, and facial expressions. We chose this dataset due to its relatively uniform distribution and its diversity in image characteristics, such as brightness, occlusion, and positioning, as well as its inclusion of a broad public demographic. Figure 1 provides sample images from the UTKFace collection. Each image is tagged with a 3-element tuple representing age (in years), gender (Male-0, Female-1), and race (White-0, Black-1, Asian-2, Indian-3, and Others-4).

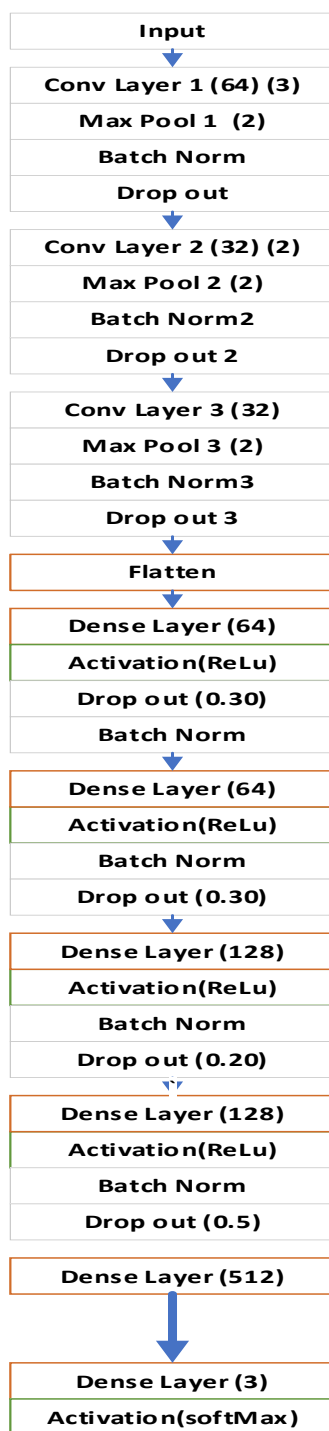


*Figure 1 sample of the UTK Face Collections*

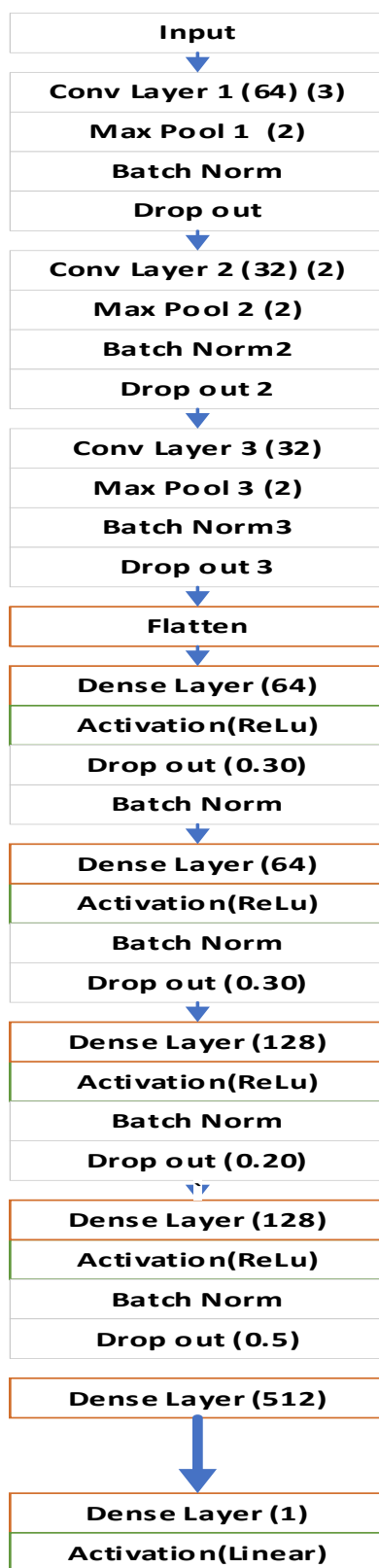
We focused primarily on the age label in this research, as gender and ethnicity were beyond our scope. For training, validation and testing, we used the same collection of images with custom CNN models, dividing the dataset into 70% for training and 20% for validation and 10% for training. To prevent distribution mismatches, we ensured that both training, validation and testing sets retained similar distributions.

#### **B. DEEP CNN**

This model begins with an input layer for images of size (imagesize, imagesize, 3) (RGB format). The initial convolutional block applies 64 filters of size 3x3, followed by max pooling (2x2), batch normalization, and dropout (0.3) to reduce overfitting. Two more convolutional blocks are added, each with 32 filters of size 2x2, followed by max pooling, batch normalization, and dropout (0.3) in each block. After the convolutional layers, the output is flattened and passed through a series of fully connected (dense) layers. The first two dense layers have 64 units each with ReLU activation, followed by dropout (0.3) and batch normalization to further mitigate overfitting. The next dense layer has 128 units, using ReLU activation, batch normalization, and dropout (0.2). Another 128-unit dense layer with ReLU activation and a final dropout (0.5) layer follows. The final dense layer before the output has 512 units with ReLU activation. The regression model outputs a single unit with a linear activation function, making it suitable for regression tasks, specifically predicting a continuous value like age. While the structure of the classification model is similar but having different last layer than the regression model, The output layer is designed for classification with three units and softmax activation, which will predict one of three possible age groups.



**Figure 1:** Age-group Classification Model Structure



**Figure 2:** Age Regression Model Structure

The defined Convolutional Neural Network (CNN) architecture as seen in figure1 is designed for multi-class classification, accepting input images of shape (224, 224, 3). It consists of three convolutional blocks, each comprising a Conv2D layer with filters (64, 32, 32) and kernel sizes (3x3, 2x2), followed by MaxPooling2D for spatial down-sampling, BatchNormalization to stabilize training, and Dropout (rate 0.3) to prevent overfitting. After feature extraction, the output is flattened using a Flatten layer and passed through four fully connected dense layers with 64, 64, 128, and 128 neurons, each employing ReLU activation. Additional BatchNormalization and Dropout layers (with rates of 0.3, 0.2, and 0.5) are applied to improve generalization. The final dense layer, containing 3 neurons and using a softmax activation function, outputs probabilities for three classes. The model is defined using TensorFlow/Keras with the input layer as inputs and the classification output as `class_cnn`. Regularization through dropout and batch normalization ensures robustness, while the combination of convolutional and dense layers facilitates effective feature extraction and classification. This architecture is suitable for multi-class classification tasks, and potential improvements could include experimenting with filter sizes, adding more convolutional blocks, or leveraging transfer learning with pre-trained models for enhanced performance and reduced training time. While the second model as seen in figure 2 is similar to model 1 but has the output modified for single continuous value(regression), The primary difference lies in the purpose of the models: one predicts class probabilities for categorical tasks (classification), while the other predicts a single continuous value for regression tasks. The shared architecture ensures effective feature extraction, but the final layer and activation functions are tailored to their specific goals. the regression model in figure 2 uses a linear activation function in the final layer to allow unbounded continuous output, suitable for regression problems.

### C. EVALUATION METRICS

Mean Average Error (MAE): MAE is the average of the absolute difference between predicted and actual output, which is represented as:

$$MAE = \frac{\sum_{i=1}^N |y - \hat{y}|}{N} \quad (1)$$

For the classification we made use of accuracy to calculate the performance of the model represented as:

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Where TP = True Positive, TN=True Negative, FP=False Positive, FN= False Negative

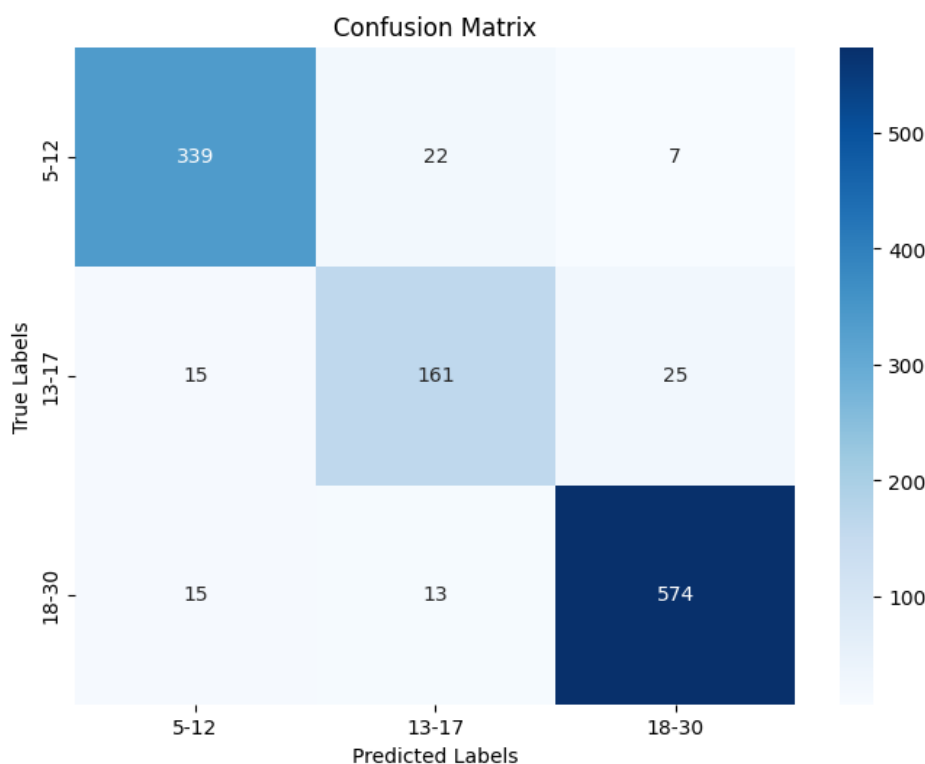
## III. EXPERIMENTAL RESULTS

From the experiments we obtained an accuracy of 93% while the regression model obtained an MAE of 2.15 in years

	precision	recall	f1-score	support
0	0.93	0.96	0.94	359
1	0.84	0.87	0.85	216
2	0.97	0.93	0.95	596
accuracy			0.93	1171
macro avg	0.91	0.92	0.92	1171
weighted avg	0.93	0.93	0.93	1171

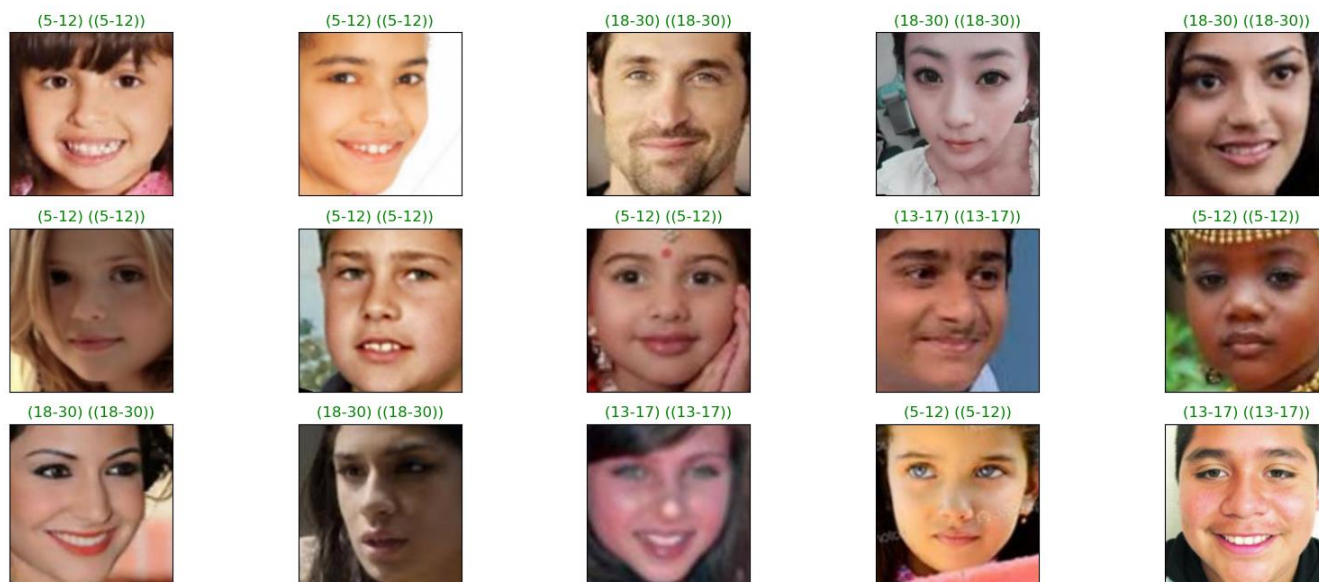
**Figure 3:** Age Classification Results of the Model

The classification model as seen in Fig 3. demonstrates strong performance across the three classes, as reflected by the precision, recall, and F1-score metrics. For Class 0 (child), the model achieves a precision of 93%, a recall of 96%, and an F1-score of 94%, indicating accurate and consistent identification of instances. Class 2(adult) shows the highest precision at 97%, with a recall of 93% and an F1-score of 95%, highlighting excellent performance. However, Class 1(teenager) has slightly lower metrics, with a precision of 84%, a recall of 87%, and an F1-score of 85%, suggesting this class is more challenging to classify, likely due to overlap with other classes even as humans we find it difficult to distinguish teenagers from 13-14 and teenagers from 17 as border cases, may also be closely categorized as either adults or children. The model achieves an overall accuracy of 93%, correctly classifying the majority of instances. The macro averages for precision, recall, and F1-score are 91%, 92%, and 92%, respectively, indicating balanced performance across all classes. Weighted averages for these metrics are 93% each, reflecting the influence of class proportions and reinforcing the model's reliability even with some class imbalance. With 359 instances for Class 0, 216 for Class 1, and 596 for Class 2, the lower metrics for Class 1 may be attributed to its smaller representation. Overall, the results suggest that the model is effective, with particularly strong performance for Classes 0 and 2, and slightly reduced but still satisfactory performance for Class 1.

**Figure 4:** Confusion Matrix for the Classification Model

From figure 4. it shows the model was able to classify 339 instances under the age -group (5-12) correctly, 22 as teenagers (13-17) and 7 as adults (18-30). For the second class (13-17) the model, classified 161 instances correctly while classifying 15 instances under the ages (5-12) and 25, instances under the ages (18-30). In the last class (18-30) we can see the model classified 574 instances correctly and classifying 13 under the ages (13-17) and 15 under the ages (10-12).





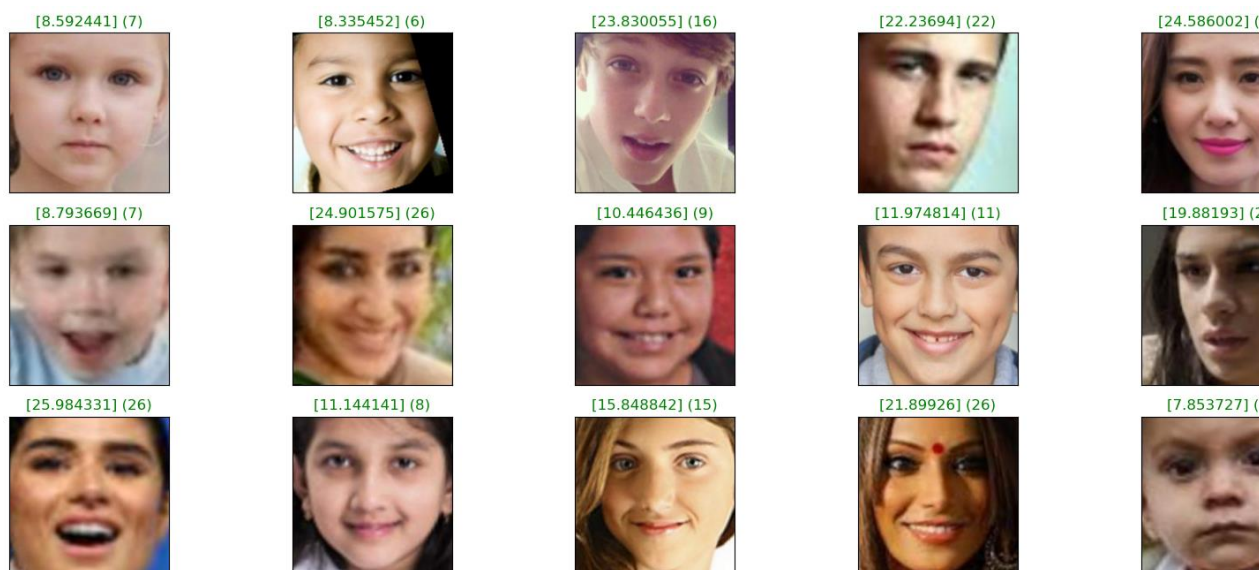
**Figure 5:** The Classification Model Performance on Test Images.

From Fig.5. we can see the model accurately classified all test samples. While for our regression model we obtained an MAE of 2.15 years.

```
37/37 [=====] - 26s 507ms/step
The MSE 8.525042155526505
The MAE 2.154280667015877
this is the rmse 2.9197674831271248
Test R^2 Score: 0.85252
```

**Figure 6:** The MAE,  $R^2$ , RMSE and MSE for the Regression Model





**Figure 7:** The Regression Model Performance on Test Images.

From the regression model we can see the model accurately classified the predicted ages to the actual ages of the test sets. With no more than 3–6-year difference.

**Table 1:** The Past Work Done and their Results on Age Classification

Work	Accuracy	Dataset
<b>Our Model</b>	93%	UTKFace
[1]	85%	All-Age Face (AAF)
[2]	94%	FG-Net
[3]	84%	Aging Faces in the Wild (AGFW)
[25]	81.2	
[24]	94%	UTKFace

**Table 2:** The Past Work Done and their Results on Age Regression

Work	Dataset	MAE
[22]	APPA-Real and UTK -Faces	4.49
[21]	MORPH II, FG-NET and CACD	2.54
[20]	DPR images	3.13
[25]	UTK-Face,FG-NET and	9.12
[19]	MORPH-2 dataset	2.64
<b>Our Model</b>	<b>UTK-Faces</b>	<b>2.15</b>

## IV. CONCLUSION

The proposed classification model demonstrates strong overall performance, achieving an accuracy of 93% and excelling in precision, recall, and F1-scores for most classes. Class 0 and Class 2 showed particularly robust results, with F1-scores of 94% and 95%, respectively. While Class 1 exhibited slightly lower metrics (F1-score of 85%), the results remain satisfactory, considering its smaller representation in the dataset. These outcomes highlight the effectiveness of the model in handling multi-class classification tasks, with reliable performance across varying class distributions. The regression model performs well, with low MAE, MSE, and RMSE values, indicating that the errors between predicted and actual values are small on average. The high  $R^2$  score further confirms that the model explains most of the variability in the target variable. However, improvements can still be explored to reduce errors further, especially if more precision is required for the specific application.

To further enhance the model's performance, especially for the underperforming Class 1 (teenager), several recommendations can be made. First, balancing the dataset by increasing the representation of Class 1 through techniques like oversampling, synthetic data generation, or targeted data collection can improve the model's ability to generalize for this class. Second, applying advanced methods such as weighted loss functions or focal loss during training could help address class imbalance by penalizing misclassifications of minority classes more heavily. Third, experimenting with hyperparameter tuning, additional network architectures, or ensemble methods may yield further improvements. Additionally, augmenting the dataset with diverse and high-quality samples could help the model better capture variations in facial features across all classes. Future work could explore incorporating transfer learning by leveraging pre-trained models on large datasets, which may reduce the dependency on extensive labeled data and improve accuracy. Finally, deploying the model in real-world applications should include regular performance monitoring and periodic retraining to adapt to new data, ensuring continued reliability and effectiveness in practical scenarios.

## V. REFERENCES

- [1] W. Cao, V. Mirjalili, and S. Raschka, "Rank consistent ordinal regression for neural networks with application to age estimation," *Pattern Recognit. Lett.*, vol. 140, pp. 325–331, Dec. 2020, doi: 10.1016/j.patrec.2020.11.008.
- [2] G. George, C. Uppin, and U. A. Bello, "Human Age Estimation from Face Images with Deep Convolutional Neural Networks Using Transfer Learning," 2024, Accessed: Nov. 07, 2024. [Online]. Available: <https://www.preprints.org/manuscript/202401.0350>
- [3] G. George, S. Adeshina, and M. M. Boukar, "The Influence of Activation Functions in Deep Learning Models Using Transfer Learning for Facial Age Prediction," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 2, Art. no. 2, 2024.
- [4] M. V. Christopher, A. Wahid, G. Z. Nabiilah, and Rojali, "Comparing Age Estimation with CNN and EfficientNetV2B1," *Procedia Comput. Sci.*, vol. 227, pp. 415–421, Jan. 2023, doi: 10.1016/j.procs.2023.10.541.
- [5] S. Sathyavathi and K. R. Baskaran, "An Intelligent Human Age Prediction from Face Image Framework Based on Deep Learning Algorithms," *Inf. Technol. Control*, vol. 52, no. 1, Art. no. 1, Mar. 2023, doi: 10.5755/j01.itc.52.1.32323.
- [6] N. Sharma, R. Sharma, and N. Jindal, "Face-Based Age and Gender Estimation Using Improved Convolutional Neural Network Approach," *Wirel. Pers. Commun.*, vol. 124, no. 4, pp. 3035–3054, Jun. 2022, doi: 10.1007/s11277-022-09501-8.
- [7] M. Ahmed and S. Viriri, "Facial Age Estimation using Transfer Learning and Bayesian Optimization based on Gender Information," 2020, *Social Science Research Network, Rochester, NY*: 3774805. Accessed: Oct. 09, 2024. [Online]. Available: <https://papers.ssrn.com/abstract=3774805>

- [8] M. Irhebhude, A. Kolawole, and F. Abdullahi, "Northern Nigeria Human Age Estimation From Facial Images Using Rotation Invariant Local Binary Pattern Features with Principal Component Analysis," vol. 45, pp. 12–28, Jan. 2021.
- [9] C. Shi, S. Zhao, K. Zhang, and X. Feng, "Multi-task multi-scale attention learning-based facial age estimation," *IET Signal Process.*, vol. 17, no. 2, Feb. 2023, doi: 10.1049/sil2.12190.
- [10] S. Sathyavathi and K. R. Baskaran, "An intelligent human age prediction from face image framework based on deep learning algorithms," *Inf. Technol. Control*, vol. 52, no. 1, pp. 245–257, 2023.