# AGE ESTIMATION USING DEEP LEARNING

*Heemin Kang, Suchet Mangat, Ragib Sina, Lucien Somorai*

*Github Repository: https://github.com/HK-Kang1/ENEL-645-Final-Project*

## ABSTRACT

This study explores the challenge of age estimation from facial images, a significant challenge in the field of computer vision with applications in security, healthcare, and marketing. Employing a comprehensive approach, deep learning models were applied to delve into the core aspects of the issue, utilizing preprocessing techniques to ensure robust data collection and analysis. The findings reveal that ResNet50 achieved the highest test accuracy among the evaluated models, indicating strong performance in age classification. However, MobileNetV2 demonstrated the most promising Grad-CAM visualizations, effectively focusing on relevant facial features, which suggests its strong potential for interpretability and practical application. Conclusively, the study proposes improvements to automated age estimation, emphasizing the implications of the research in contributing to identify verification, targeted advertising and medical diagnostics. This abstract encapsulates the essence and scope of the research, providing a clear overview of its objectives, methodology, results, and the subsequent impact on the field.

## 1. INTRODUCTION

The human face conveys numerous information such as identity, emotions, and demographic characteristics such as age and gender. Facial features including the eyes, nose, mouth, skin texture and wrinkles change over time making age estimation a complex task. Humans can often approximate age instinctively based on someone's facial features but the process to teach machines to do this task presents a challenge due to variations of facial structure, lighting conditions, and image quality.

Raising the ability of a machine to recognize and interpret facial characteristics such as age has important applications in numerous fields. In security and law enforcement, it can aid in identity verification and fraud detection [1]. In healthcare, age estimation can assist in diagnosing age-related diseases and tailoring medical treatments. In addition, it can yield insights into biological age and physiological health which can translate to a clinically useful measure [2].

This study aims to build a model for classifying age from facial images using the UTKFace dataset, which contains over 20,000 labeled images [3]. We explore the effectiveness of different deep learning models namely MobileNetV2, ResNet50, EfficientNet-B0 by experimenting with different hyperparameters to optimize performance.

## 2. RELATED WORK

Facial aging features such as wrinkles and texture changes provide significant cues for age estimation, which has substantial applications across various domains, including security, healthcare, and commercial sectors. Studies leveraging the UTKFace dataset demonstrate advanced methodologies in machine learning that enhance the precision of age classification.

One such study investigates the influence of gender on age estimation accuracy, employing gender-specific models to improve predictions, suggesting that gender plays a crucial role in age-related facial features [4]. Another research focuses on a deep convolutional neural network's ability to discern age from facial images, showcasing the potential of deep learning architectures in capturing intricate age indicators from a comprehensive dataset [5]. Additionally, the integration of age and gender classification highlights the convolutional layers' effectiveness in identifying key aging features, which could be pivotal in refining model accuracy for age-specific applications [6]. Further exploration into the fusion of facial geometric features and texture patterns has revealed that combining these different types of data can significantly boost the accuracy of age estimation systems.

Advanced machine learning techniques, such as ensemble learning, have been employed to integrate these features effectively, allowing models to leverage both macro and micro-level details for more precise age prediction [7]. Additionally, the adoption of semi-supervised learning frameworks has shown promise in utilizing unlabeled data alongside labeled examples, enhancing the training process and enabling models to generalize better on unseen data. This approach is particularly beneficial given the intrinsic variability of facial aging, which can vary widely across different ethnicities and individual lifestyles [8].

## 3. MATERIALS AND METHODS

In this section, we will discuss the materials and methods used in the study, starting with the dataset. The dataset comprises face images categorized by age into four groups (0-25, 26-50, 51-75, and 76-116) and divided into training, validation, and testing subsets. Data augmentation techniques are applied to the training data to improve model generalization. Next, we will cover the image preprocessing steps, including resizing the images to 224x224 pixels, normalizing them based on ImageNet standards, and applying data augmentation (such as random horizontal flips) to ensure the data is prepared effectively for model training.

The section will also detail the use of three pre-trained deep learning models—MobileNetV2, ResNet50, and EfficientNet-B0. These models have their feature extraction layers frozen, and the classifier layers are fine-tuned to classify the images into age categories. The training process, which involves using the Adam optimizer and CrossEntropyLoss over 10 epochs, will be explained. Finally, we will describe the custom fully connected classifier added to each model to further enhance performance, including its structure and activation functions.

### 3.1. Dataset

The dataset used in this study consists of face images stored in different subdirectories. The images are labeled based on the age information extracted from the filenames [3]. The age labels are grouped into four categories: 0-25, 26-50, 51-75, 76-116. The dataset is divided into three subsets: training (70%), validation (15%), and testing (15%). To improve model generalization, data augmentation techniques were applied to the training dataset.

### 3.2. Image Preprocessing

The preprocessing steps involve loading image files, extracting age labels, and applying specific transformations. All images are resized to 224x224 pixels to ensure consistency across models. Each image undergoes a normalization process that aligns with the well-established mean and standard deviation values used for ImageNet models, specifically [0.485, 0.456, 0.406] for the mean and [0.229, 0.224, 0.225] for the standard deviation [10, 11, 12]. This approach ensures that each image is treated consistently and effectively across various deep learning models, helping to stabilize the training process by aligning the input data distribution with the distribution expected by models pretrained on the ImageNet dataset [9]. In addition, a random horizontal flip is applied to the training dataset as a common form of data augmentation. This technique helps make the model robust to variations in the data by simulating different orientations.

### 3.3. Model Architectures and Training

Three pre-trained deep learning models were used for the classification task: MobileNetV2 [10] ResNet50 [11] EfficientNet-B0 [12]. The feature extraction layers of each model were frozen, and only the classifier layers were fine-tuned for the five age categories. The models were trained using the Adam optimizer with a learning rate of 0.001 and CrossEntropyLoss as the loss function. Training was conducted for 10 epochs, and the best model was saved based on validation accuracy.

To enhance classification performance, a custom fully connected classifer was added to each model. The classifier consists of:

- First layer: 512 neurons, ReLU acitvation, and a dropout of 0.3
- Second layer: 256 neurons, ReLU acitvation, and a dropout of 0.3
- Output layer: 4 neurons (corresponding to the four age categories) with softmax activation

### 3.4. Environmental Settings

The training and evaluation processes were conducted on the TALC cluster, which provides GPU support to accelerate computation. The cluster was configured to allocate 1 GPU, 2 CPU cores, and 16GB of memory for each training job. The SLURM workload manager was used to submit training jobs, ensuring efficient resource utilization. Training runs were limited to a maximum execution time of 24 hours per job to optimize resource scheduling and availability.

## 4. RESULTS AND DISCUSSION

This section presents the performance of the three deep learning models, MobileNetV2, ResNet50, and EfficientNet-B0, trained for age classification using the UTKFace dataset. We evaluate each model based on training and validation loss, accuracy, and test accuracy. Additionally, we provide Grad-CAM visualizations to analyze which facial regions the models focus on during classification.

### 4.1. Performance of MobileNetV2

MobileNetV2 showed steady improvement over ten training epochs. The training accuracy increased from 64.85% to 72.07%, while the validation accuracy improved from 69.03% to 71.49%. The training loss decreased from 0.8144 to 0.6276, while the validation loss reduced from 0.7413 to 0.6697.

The best validation and test accuracy for the MobileNetV2 model were 71.49% and 70.58% respectively.

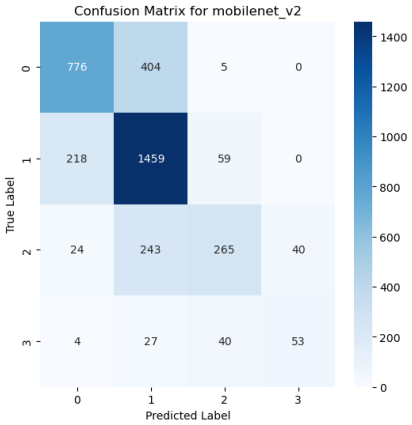


**Fig. 1.** Confusion matrix for MobileNetV2.

### 4.2. Performance of ResNet50

ResNet50 demonstrated slightly better performance than MobileNetV2. The training accuracy improved from 64.85% to 71.60%, while the validation accuracy increased from 69.88% to 71.16%. The training loss reduced from 0.8091 to 0.6571, while validation loss dropped from 0.6907 to 0.6460.

The best validation and test accuracy for the Resnet50 model were 71.16% and 72.38% respectively.

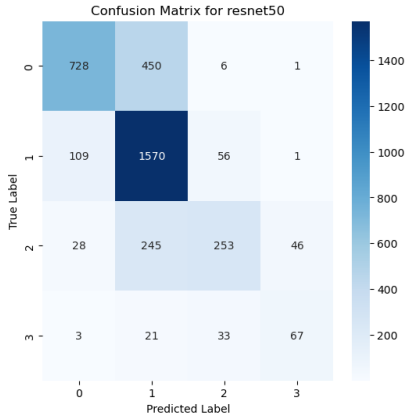


**Fig. 2.** Confusion matrix for Resnet50.

### 4.3. Performance of EfficientNet-B0

EfficientNet-B0 achieved the highest training accuracy among the models, increasing from **66.54%** to **80.25%**. The validation accuracy peaked at **71.21%**, and the test accuracy reached **72.27%**. The training loss dropped significantly from **0.7751** to **0.4589**, while the validation loss fluctuated, ending at **0.8159**.

The best validation and test accuracy for the EfficientNet-B0 model were 71.21% and 72.27% respectively.

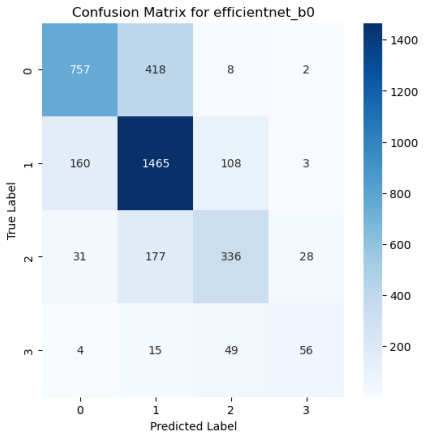


**Fig. 3.** Confusion matrix for EfficientNet-B0.

### 4.4. Grad-CAM Visualization

Grad-CAM visualizations were used to interpret the models' decision-making by highlighting the most influential facial regions. Among the three models, MobileNetV2 exhibited the most accurate attention to facial features, focusing consistently on the central face regions. However, despite this advantage in interpretability, its test accuracy was slightly lower than the other models. ResNet50 and EfficientNet-B0, while achieving higher test accuracy, occasionally exhibited attention shifts toward background elements or non-facial regions, which might have contributed to some misclassifications.

MobileNetV2

ResNet50


**Fig. 4.** Grad-CAM visualization for all models

Overall, while all three models performed similarly in terms of accuracy, their Grad-CAM visualizations suggest differences in feature extraction and attention mechanisms. This insight can be useful for future improvements in age classification models by refining how attention is distributed across facial features.

### 4.5. Model Comparison

From the results, ResNet50 achieved the highest test accuracy, followed closely by EfficientNet-B0, while MobileNetV2 had the lowest performance. The results suggest that deeper models, such as ResNet50, are more effective for age classification due to their superior feature extraction capabilities. However, EfficientNet-B0, despite being more compact than ResNet50, performed comparably, indicating its efficiency in balancing computational cost and accuracy.

Although MobileNetV2 had the lowest accuracy, its lightweight nature and computational efficiency make it a viable option for real-time or mobile applications where inference speed is a critical factor [13]. Additionally, MobileNetV2's Grad-CAM visualizations demonstrated the most accurate focus on facial features, suggesting that while it may struggle with classification accuracy, it excels in correctly identifying the relevant image regions for age classification. In contrast, ResNet50 and EfficientNet-B0 may be better suited for applications requiring higher classification accuracy, such as medical or forensic age estimation. Future work could explore hybrid approaches that combine the strengths of these architectures to improve both performance and efficiency.

Table 1. Test accuracies of different models.

| **Model** | **Test Accuracy** |
|---|---|
| MobileNetV2 | 70.58% |
| ResNet50 | 72.38% |
| EfficientNet-B0 | 72.27% |

## 5. CONCLUSION

The task of age estimation using facial imagery is one of growing significance in security, healthcare and commercial applications. To aid that effort, this study presents a comprehensive analysis of deep learning CNN frameworks fine-tuned using the UTKFace dataset. Throughout this study the MobileNetV2, ResNet50 and EfficientNet-B0 were all evaluated to best gauge the effectiveness, application and interpretability of each implementation. Among the three models ResNet50 achieved the highest test accuracy at 72.38% demonstrating the effectiveness of deeper neural networks and their ability to capture the complexities of age-related facial features. EfficientNet-B0 while being more compact, followed closely behind Resnet50 boasting a 72.27% test accuracy. MobileNetV2, although marginally lower in test accuracy at 70.58%, provided more interpretable Grad-CAM visualizations while exhibiting superior spatial attention to facial regions, reinforcing its potential for lightweight, real-time applications.

These findings underscore the nuanced trade-offs between accuracy, efficiency, and interpretability in the context of age estimation. While ResNet50 and EfficientNet-B0 achieved marginally higher test accuracy, MobileNetV2 demonstrated superior attention to relevant facial features as revealed by Grad-CAM visualizations. This highlights the importance of incorporating attention-based interpretability methods, which offer insights beyond standard performance metrics. Such methods enable researchers to assess whether a model is making decisions based on meaningful facial cues or irrelevant regions, thereby supporting the development of more trustworthy and robust systems. This is particularly valuable in safety-critical or user-facing applications, where understanding a model's decision-making process is essential for ensuring fairness, transparency, and accountability.

Ultimately, this work contributes to the growing body of research on automated age estimation, providing empirical benchmarks and interpretability insights that may guide the development of more accurate, efficient, and explainable facial analysis systems.

## 6. REFERENCES

[1] T. L. Johnson, N. N. Johnson, V. Topalli, D. McCurdy, and A. Wallace, "Police facial recognition applications and violent crime control in U.S. cities," *Cities*, vol. 155, p. 105472, Oct. 2024, doi: https://doi.org/10.1016/j.cities.2024.105472.

[2] [1]O. C. Zalay *et al.*, "Decoding biological age from face photographs using deep learning," Sep. 2023, doi: https://doi.org/10.1101/2023.09.12.23295132.

[3] "UTKFace," *UTKFace*. https://susanqq.github.io/UTKFace/

[4] M. ACA, A. Rovere, and G. Pirk, “Germination of Gutierrezia solbrigii and Senecio subulatus, endemic Asteraceae from Argentina,” *Phyton*, vol. 85, no. 1, pp. 314–323, Jan. 2016, doi: https://doi.org/10.32604/phyton.2016.85.314.

[5] İ. Akgül, “Deep convolutional neural networks for age and gender estimation using an imbalanced dataset of human face images,” *Neural Computing and Applications*, Sep. 2024, doi: https://doi.org/10.1007/s00521-024-10390-0.

[6] C. Yan, C. Lang, T. Wang, Du Xuetao, and C. Zhang, “Age Estimation Based on Convolutional Neural Network,” pp. 211–220, Jan. 2014, doi: https://doi.org/10.1007/978-3-319-13168-9_22.

[7] R. Kumar, K. Singh, D. P. Mahato, and U. Gupta, “Face-based age and gender classification using deep learning model,” *Procedia Computer Science*, vol. 235, pp. 2985–2995, 2024, doi: https://doi.org/10.1016/j.procs.2024.04.282.

[8] C. Li, Q. Liu, W. Dong, X. Zhu, J. Liu, and H. Lu, “Human Age Estimation Based on Locality and Ordinal Information,” vol. 45, no.11, pp.2522–2534, Nov.2015, doi: https://doi.org/10.1109/tcyb.2014.2376517.

[9] [1]“Image Normalization - an overview | ScienceDirect Topics,”*www.sciencedirect.com*.https://www.sciencedirect.com/topics/engineering/image-normalization

[10]“PyTorch,”*www.pytorch.org*.https://pytorch.org/hub/pytorch_vision_mobilenet_v2/

[11] “resnet50 — Torchvision main documentation,” *pytorch.org*. https://pytorch.org/vision/main/models/generated/torchvision.models.resnet50.html

[12] “efficientnet_b0 — Torchvisionmain documentation,” *Pytorch.org*, 2023. https://pytorch.org/vision/main/models/generated/torchvision.models.efficientnet_b0.html

[13] A. Howard *et al.*, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” Apr. 2017. Available: https://arxiv.org/pdf/1704.04861