

# Machine Learning - Final Project

Michael Dadush 206917908

Shay Gali 315202242

March 2025

## 1 Introduction

## 2 Description of the Database

We chose to work with the Adience Benchmark dataset for Gender and Age Classification[1]. This dataset has real-world images of faces, which makes it good for testing different machine learning methods. The dataset includes:

### Labels in Dataset

The dataset provides many labels for each data point, we used the following labels for our analysis:

- **gender:** Male / Female
- **age:** The dataset categorizes images into the following age ranges: 0–2 years, 4–6 years, 8–12 years, 15–20 years, 25–32 years, 38–43 years, 48–53 years, 60 and above
- **user\_id, face\_id, original\_image:** These labels are used to identify the image and the person in the image.
- The dataset includes more labels that we did not use in our analysis, but we explain what each label means in the dataset in the data exploration notebook[2].

## 3 Analysis of the Database

### 3.1 Data Exploration and Preprocessing

Our exploratory analysis of the Adience Benchmark dataset revealed several key characteristics and pre-processing requirements:

1. **Duplicate Directories:** An initial examination uncovered two directories named "faces" containing image files. Further investigation showed that one directory (containing 40 files) was a subset of the other (containing 168 files). We selected the comprehensive directory containing all 168 unique user folders for our analysis.
2. **Data Organization:** The dataset follows a 5-fold cross-validation structure, with data distributed across files named "fold\_0\_data.txt" through "fold\_4\_data.txt". Folds 0-3 (containing 15,554 entries) were allocated for training, while fold 4 (3,816 entries) was reserved for testing, resulting in a total of 19,370 entries before preprocessing.
3. **Missing Values:** The raw dataset contained some entries with missing values. After removing these entries, we retained 18,551 complete records, indicating a relatively small proportion (approximately 4.2%) of incomplete data.

### 3.2 Age Classification Standardization

One of the significant challenges in working with the Adience dataset was the inconsistent representation of age information:

1. **Overlapping Age Ranges:** The dataset contained multiple overlapping age ranges that required consolidation:
  - The ranges (38, 42), (38, 43), and (38, 48) were merged into (38, 43) based on frequency analysis (46, 2293, and 6 instances respectively)
  - An anomalous (8, 23) age range with only one instance was removed
  - The less common (27, 32) range (77 instances) was merged into the more frequent (25, 32) range (4,953 instances)
2. **Single Age Values:** The dataset contained 1,094 entries with single age values rather than ranges. We mapped these to appropriate age ranges based on proximity:
  - Ages 2, 3  $\rightarrow$  (0, 2), (4, 6) respectively
  - Age 13  $\rightarrow$  (8, 12)
  - Ages 22, 23, 29, 34  $\rightarrow$  (25, 32)
  - Ages 35, 36, 42  $\rightarrow$  (38, 43)
  - Ages 45, 46, 55  $\rightarrow$  (48, 53)
  - Ages 57, 58  $\rightarrow$  (60, 100)

This standardization resulted in a consistent set of eight age ranges: (0, 2), (4, 6), (8, 12), (15, 20), (25, 32), (38, 43), (48, 53), and (60, 100).

### 3.3 Gender Distribution Analysis

- Female: 9,332 samples (50.3%)
- Male: 9,216 samples (49.7%)

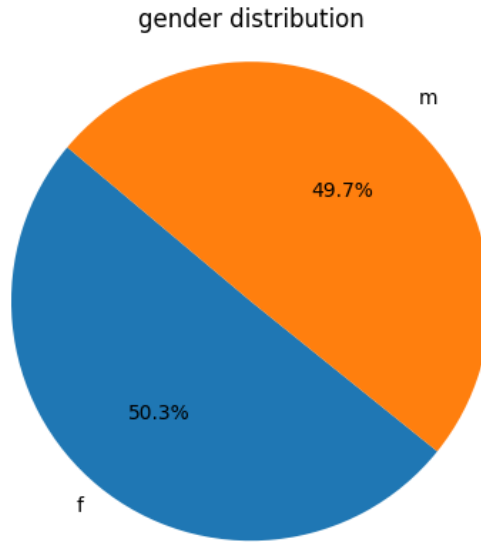


Figure 1: Gender Distribution in the Adience Benchmark Dataset

### 3.4 Age Distribution Analysis

After standardization, the age distribution across the eight defined ranges was:

- (0, 2): 2,491 samples (13.4%)
- (4, 6): 2,158 samples (11.6%)
- (8, 12): 2,287 samples (12.3%)
- (15, 20): 1,642 samples (8.9%)
- (25, 32): 5,391 samples (29.1%)
- (38, 43): 2,695 samples (14.5%)
- (48, 53): 990 samples (5.3%)
- (60, 100): 896 samples (4.8%)

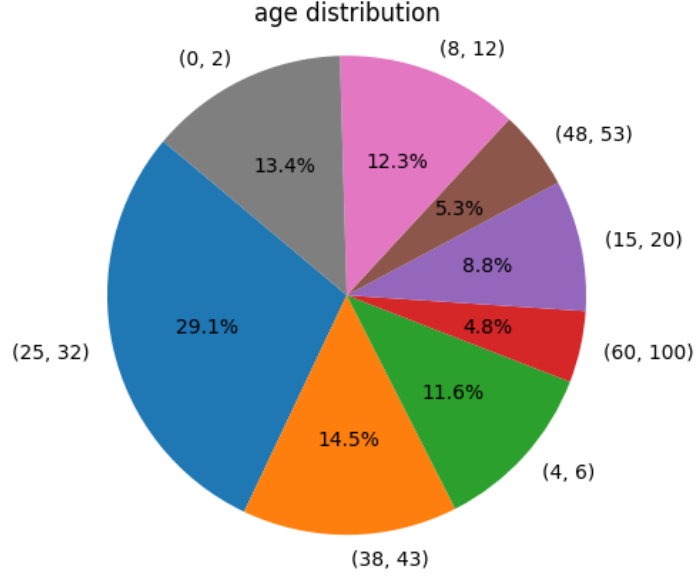


Figure 2: Age Distribution in the Adience Benchmark Dataset

### 3.5 Combined Age-Gender Analysis

The intersection of age and gender categories revealed interesting patterns:

- The (25, 32) age range showed the highest representation for both genders, with 3,026 female and 2,363 male samples.
- Gender distribution within age groups varied significantly:
  - (0, 2): 684 female vs. 1,807 male
  - (8, 12): 1,353 female vs. 932 male
  - (25, 32): 3,026 female vs. 2,363 male
  - (48, 53): 458 female vs. 532 male
  - (60, 100): 428 female vs. 467 male

This combined analysis provides valuable insights into potential intersectional biases that might emerge during model training, particularly in the youngest age group where gender assignment was artificial rather than determined from image features.

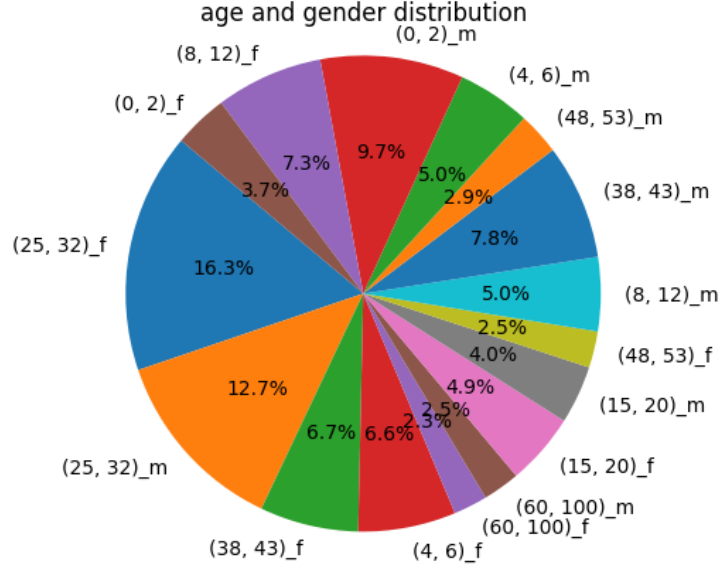


Figure 3: Combined Age and Gender Distribution in the Adience Benchmark Dataset

### 3.6 Data Partitioning and Integrity

To ensure proper evaluation methodology, we split the dataset into three partitions:

- Training set: 11,850 samples (63.9%)
- Validation set: 2,963 samples (16.0%)
- Test set: 3,729 samples (20.1%)

We rigorously verified that no overlap existed between these sets by checking the intersection between training and testing data based on image identifiers and face IDs. This verification confirmed complete separation, which is crucial for unbiased performance evaluation.

### Tools used in the project

The following tools and techniques were used in this project :

- Softmax
- SVM
- Random Forest

- Adaboost
- KNN
- Multi-layer Perceptron (MLP)
- PCA
- ResNet50- as feature extractor and as a model

## 4 Training Methodology

### 4.1 Feature Extraction

For most of our models, we used pre-trained networks to extract meaningful features from face images rather than training directly on raw pixel values. This approach, known as transfer learning, leverages knowledge from models pre-trained on large datasets.

- **ResNet50 Feature Extraction:** We utilized the ResNet50 architecture pre-trained on ImageNet to transform raw images into 2048-dimensional feature vectors. This was implemented for both RGB and grayscale processing pipelines.
- **Image Preprocessing:** Before feature extraction, images were resized to  $224 \times 224$  pixels and normalized according to ImageNet statistics. For grayscale processing, single-channel images were duplicated across three channels to match ResNet50’s input requirements.
- **Batch Processing:** To manage memory efficiently, images were processed in batches of 64, with separate feature sets created for training, validation, and test partitions.

We also try PCA as a feature extraction method, but the results were not as good as the ResNet50 feature extraction method, so we decided to use the ResNet50 feature extraction method for our models.

### 4.2 Data Preprocessing

After feature extraction, several preprocessing steps were applied to prepare the data for model training:

- **Feature Standardization:** Features were standardized to have zero mean and unit variance using scikit-learn’s StandardScaler, fitted on the training set and applied to both validation and test sets.
- **Label Encoding:** The combined age and gender labels (e.g., ”(25, 32)\_f” for a female in the 25-32 age range) were encoded into numeric indices using LabelEncoder.
- **Class Weights:** To address class imbalance in the dataset, we computed balanced class weights inversely proportional to class frequencies, ensuring that minority classes received appropriate attention during training.

### 4.3 Model Training Approaches

We trained two parallel sets of models—one using RGB features and another using grayscale features—to compare their effectiveness:

- **Traditional Machine Learning:** For Softmax (Logistic Regression), SVM, Random Forest, AdaBoost, and KNN models, we used scikit-learn implementations.
- **Neural Network Models:** For MLP model, we implemented architectures in TensorFlow with dropout and batch normalization to prevent overfitting. Training used the Adam optimizer with categorical cross-entropy loss and early stopping based on validation set performance.
- **CNN Training:** These models were trained directly on image data rather than on extracted features.

#### 4.4 Evaluation Methodology

To ensure robust evaluation of model performance:

- **Metrics:** We primarily used accuracy as our evaluation metric, calculated on both training and test sets to assess overfitting.
- **Confusion Matrix Analysis:** We analyzed confusion matrices to understand the types of errors made by each model, particularly focusing on whether errors occurred between adjacent age groups or across gender boundaries.
- **Cross-Comparison:** Performance was compared across RGB and grayscale processing pipelines to determine whether color information provided meaningful advantages.

All models were saved after training for later deployment and testing in real-world scenarios, which informed our analysis of practical applications detailed in the research questions section.



## 5 Model Performance

### 5.1 Grayscale scores

#### 5.1.1 Grayscale Model Training and Testing Accuracy

Table 1 summarizes the accuracy results of different models on the dataset using grayscale image.

Model	Train Accuracy	Test Accuracy	Notes
Softmax	0.7095	0.3135	most of the error in the confusion matrix was in the adjacent age group.
SVM	0.2046	0.1781	the model predict for the most of the data to 25-32 age group, and the gender was female.
Random Forest	0.7730	0.3167	the model predict for the most of the data to 25-32 age group, but he got the gender right.
AdaBoost	0.3467	0.2650	the have berrly predict classes that he dont have a lot of data.
K-NN (k=1)	0.9998	0.2202	we can see the overfitting - the train accuracy is very high, but the test accuracy is very low.
K-NN (k=3)	0.7525	0.2258	we can see start of overfitting - most of the error in the confusion matrix was near the (25,32) age group - the majority class.
K-NN (k=5)	0.7001	0.2314	the over fitting is less significant
K-NN (k=7)	0.6631	0.2422	the over fitting is less significant
K-NN (k=9)	0.6331	0.2440	the over fitting is less significant
MLP	0.7521	0.3229	the model with the less overfitting, and when the model got wrong predictions, it was not by much (the most of the error in the confusion matrix was in the adjacent age group).
CNN	0.4371	0.3153	this model have overffiting - it predict most of the data to the (25,32) age group.

Table 1: Train, Test, and Validation Accuracy of Various Models with grayscale method

### 5.1.2 Grayscale Model Confusion Matrices

The following confusion matrices illustrate the classification patterns for each grayscale model, showing where models tend to make correct predictions (diagonal) versus misclassifications (off-diagonal).

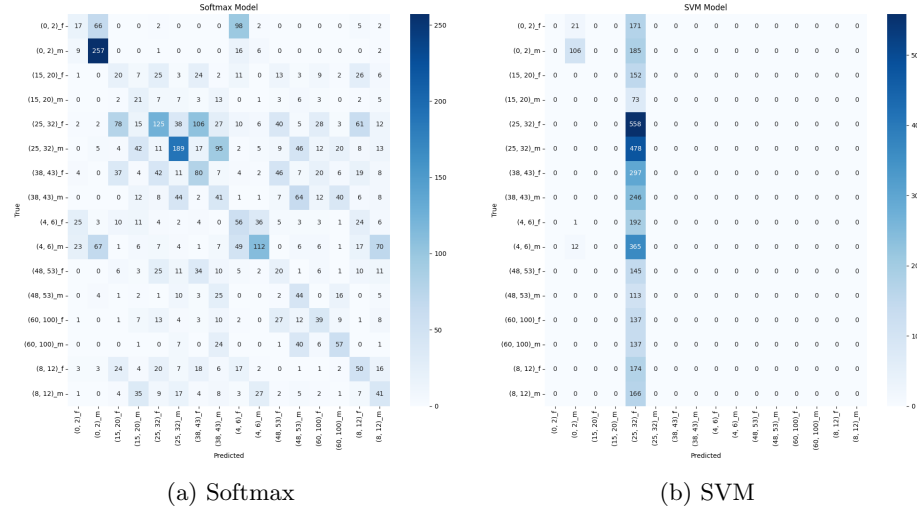


Figure 4: Confusion matrices for Softmax and SVM models (grayscale)

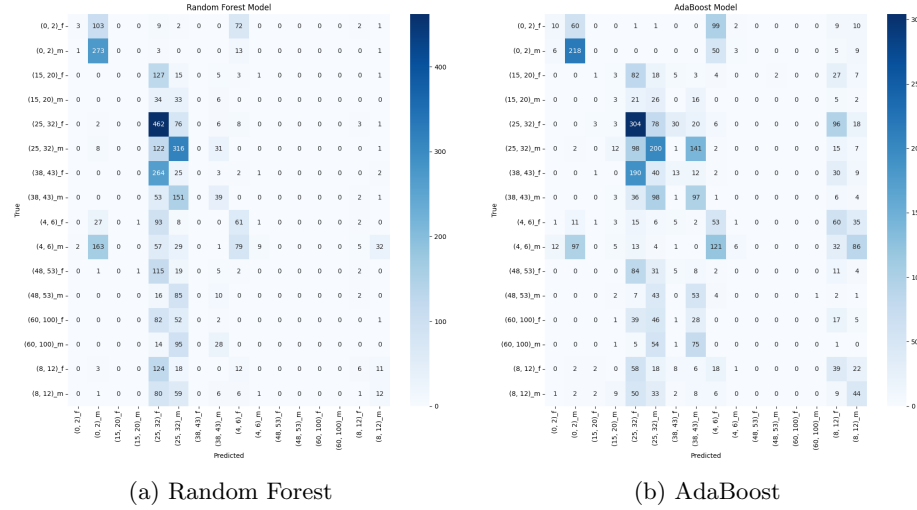


Figure 5: Confusion matrices for Random Forest and AdaBoost models (grayscale)

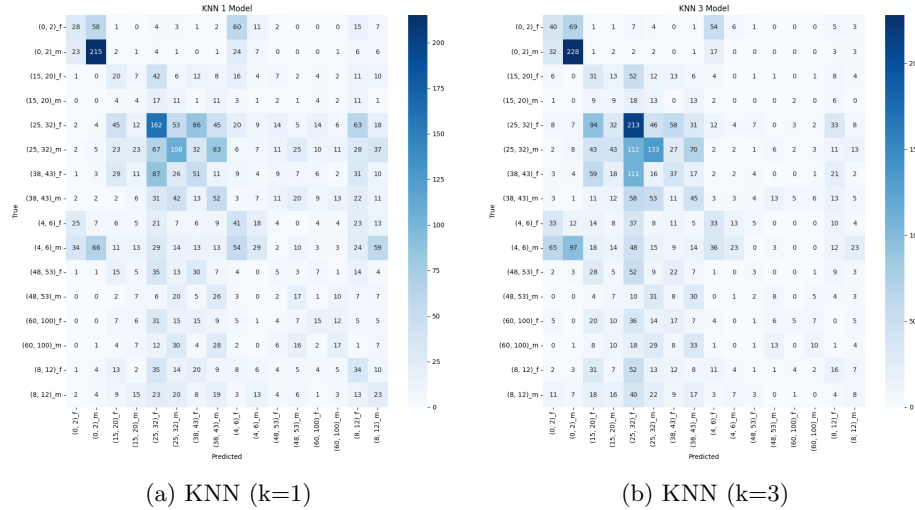


Figure 6: Confusion matrices for KNN models with k=1 and k=3 (grayscale)

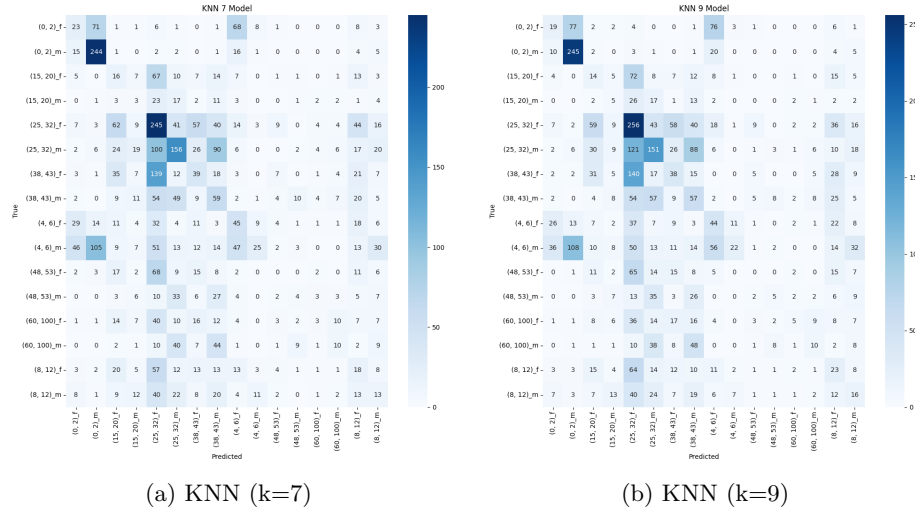


Figure 7: Confusion matrices for KNN models with k=7 and k=9 (grayscale)

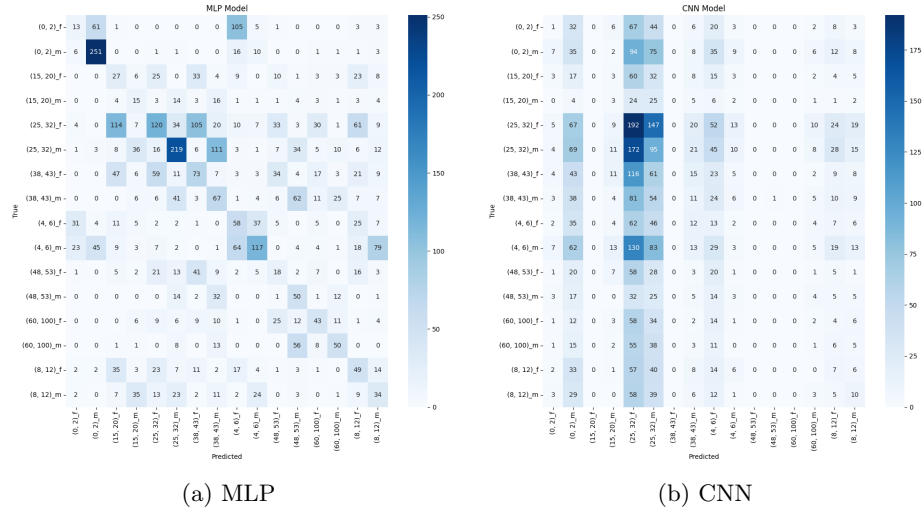


Figure 8: Confusion matrices for MLP and CNN models (grayscale)

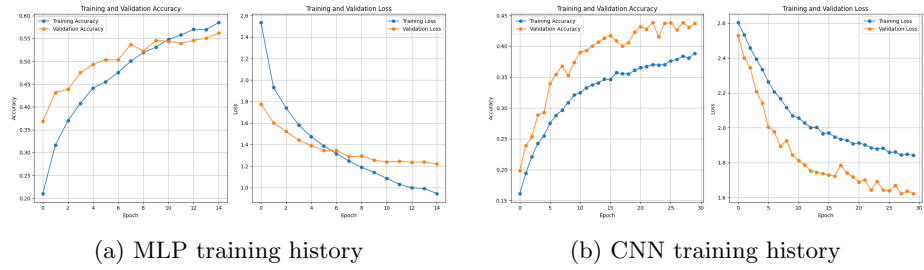


Figure 9: Taining history for MLP and CNN models (grayscale)

## 5.2 RGB scores

### 5.2.1 RGB Model Training and Testing Accuracy

Table 2 summarizes the accuracy results of different models on the dataset using RGB image. (we dont add a Note column because the results are similar to the grayscale image)

Model	Train Accuracy	Test Accuracy
Softmax	0.7331	0.3237
SVM	0.2045	0.1751
Random Forest	0.7728	0.3218
AdaBoost	0.3716	0.3162
K-NN (k=1)	0.9998	0.1917
K-NN (k=3)	0.7778	0.2092
K-NN (k=7)	0.6886	0.2279
K-NN (k=9)	0.6658	0.2333
MLP	0.7950	0.3478
CNN	0.4718	0.3298

Table 2: Train and Test Accuracy of Various Models with RGB image

### 5.2.2 RGB Model Confusion Matrices

The following confusion matrices illustrate the classification patterns for each rgb model, showing where models tend to make correct predictions (diagonal) versus misclassifications (off-diagonal).

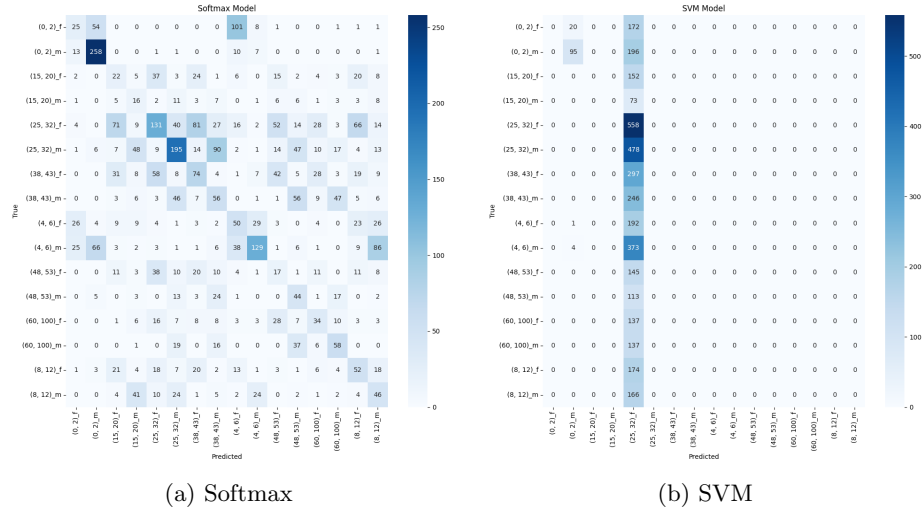


Figure 10: Confusion matrices for Softmax and SVM models (rgb)

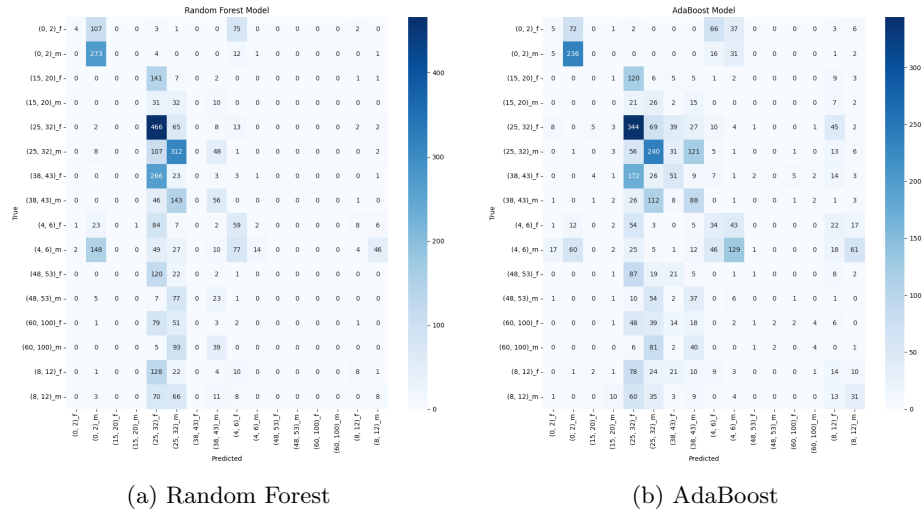


Figure 11: Confusion matrices for Random Forest and AdaBoost models (rgb)

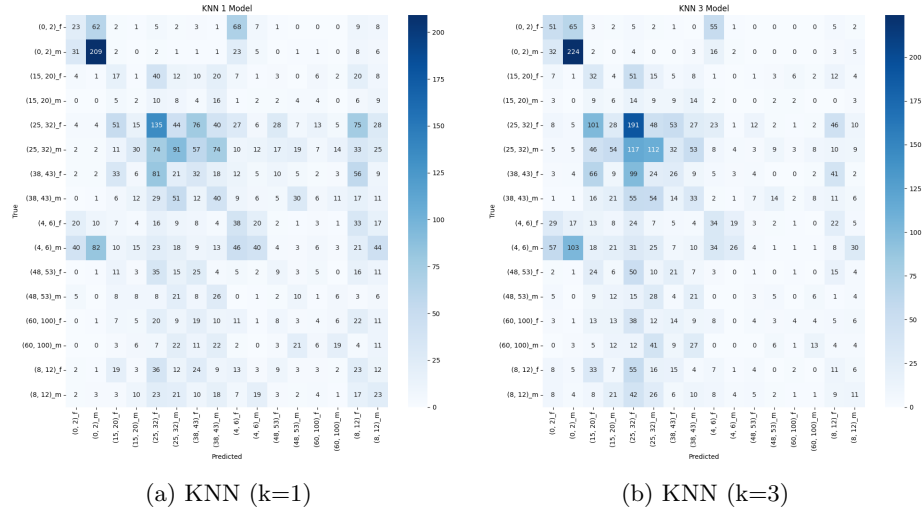


Figure 12: Confusion matrices for KNN models with k=1 and k=3 (rgb)

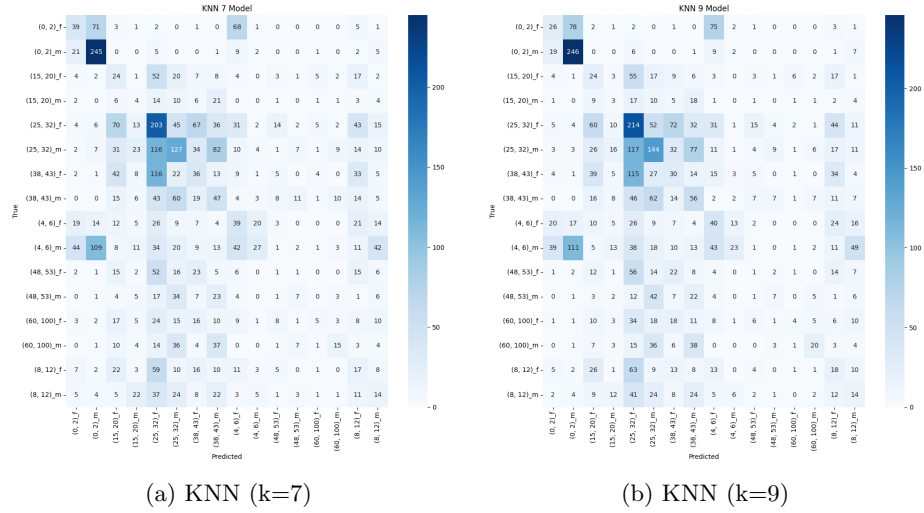


Figure 13: Confusion matrices for KNN models with k=7 and k=9 (rgb)

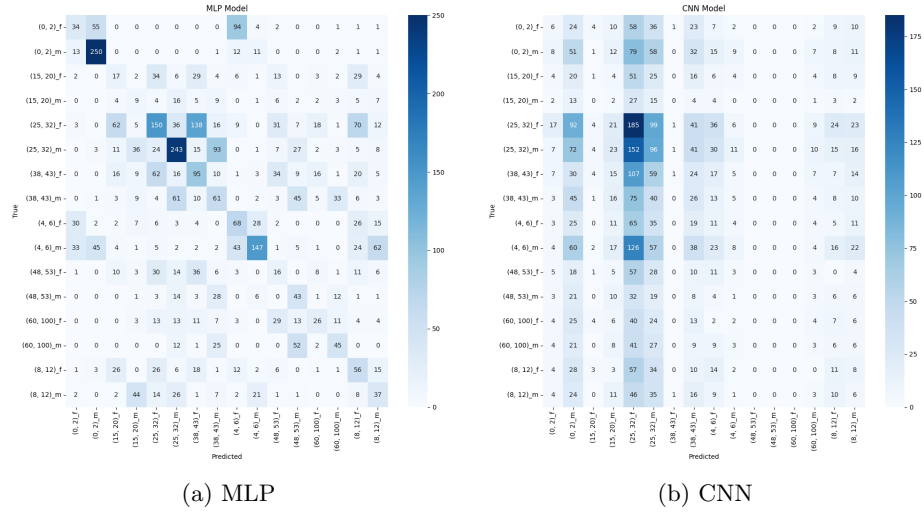


Figure 14: Confusion matrices for MLP and CNN models (rgb)

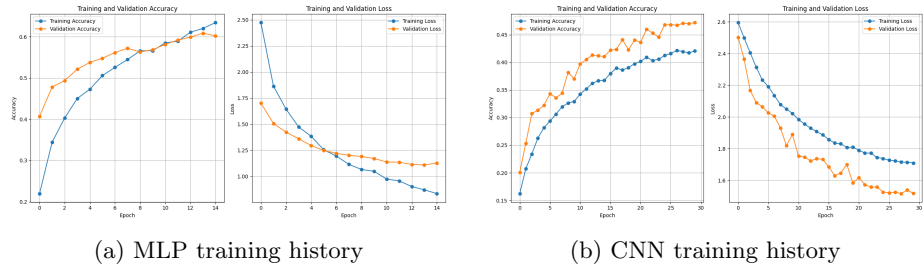


Figure 15: Taining history for MLP and CNN models (rgb)



### 5.3 Analysis of Results

Our experimental results reveal several important patterns across different model architectures and input processing methods. This section provides a detailed analysis of these findings.

#### 5.3.1 General Performance Trends

Overall, models trained on RGB features slightly outperformed their grayscale counterparts, suggesting that color information provides valuable signals for age and gender classification. The best-performing model was the MLP with RGB features (34.78% test accuracy), while the poorest performer was the SVM (17.51% test accuracy on RGB features).

The relatively modest test accuracies (17.51% to 34.78%) reflect the inherent difficulty of the 16-class joint age-gender classification task. For context, random guessing would yield approximately 6.25% accuracy, so our models provide 3-5 times better performance than chance.

#### 5.3.2 Model-Specific Analysis

**Softmax Regression:** Softmax achieved reasonable performance (32.37% test accuracy with RGB features) despite being one of the simplest models. This suggests that the decision boundaries in our feature space have some linear separability. The model’s confusion matrix showed it performed best on majority classes like (0, 2)<sub>m</sub> and (25, 32)<sub>m</sub>, with most errors occurring between adjacent age groups rather than gender misclassification.

**Support Vector Machine (SVM):** SVM performed notably poorly (17.51% test accuracy), suggesting that the complex, high-dimensional feature space was not amenable to the approach of finding optimal hyperplanes. The SVM predominantly predicted the (25, 32)<sub>f</sub> class—the most common in our dataset—indicating a strong bias toward majority classes despite our class weighting efforts.

**Random Forest:** Random Forest models performed well (32.18% test accuracy with RGB), showing their ability to capture complex patterns through ensemble decision trees. The model effectively learned the feature importance patterns from ResNet50 embeddings, with depth-limited trees (max\_depth=10) helping to control overfitting. However, like most models, it struggled with underrepresented classes.

**AdaBoost:** AdaBoost showed moderate performance (31.62% with RGB), focusing its learning on difficult-to-classify examples. This approach was particularly effective for our dataset with imbalanced classes, as evidenced by its relatively smaller gap between training and test accuracy compared to Random Forest.

**K-Nearest Neighbors (KNN):** KNN demonstrated clear patterns of overfitting, especially at lower  $k$  values. With  $k=1$ , the model achieved near-perfect training accuracy (99.98%) but poor test performance (19.17%), exemplifying "memorization" rather than generalization. As  $k$  increased, both overfitting and overall accuracy decreased, with  $k=9$  showing the best test performance among KNN models. This confirms that allowing more neighbors to vote helps smooth decision boundaries in high-dimensional spaces.

**Multi-Layer Perceptron (MLP):** The MLP achieved the best overall performance (34.78% test accuracy with RGB), demonstrating the advantage of deep learning architectures in capturing complex patterns. The network's combination of dense layers with dropout (0.5) and batch normalization effectively controlled overfitting while maintaining representational power. Early stopping further enhanced generalization by preventing overtraining.

**Convolutional Neural Network (CNN):** The CNN performed well (32.98% with RGB) despite being trained directly on raw image data rather than pre-extracted features. This indicates that end-to-end learning can match transfer learning approaches on this task. However, the narrower gap between training (47.18%) and testing (32.98%) accuracy suggests that CNN architecture inherently provides some regularization compared to other models.

### 5.3.3 Overfitting Analysis

All models exhibited some degree of overfitting, visible in the gap between training and test accuracy. This can be attributed to several factors:

- **Limited data diversity:** Despite having thousands of images, the number of samples per class was insufficient for reliable generalization across all 16 classes.
- **Class imbalance:** The uneven distribution of samples across age and gender categories made it difficult for models to learn accurate representations of minority classes.
- **High-dimensional feature space:** The 2048-dimensional ResNet50 features created a vast feature space relative to the number of training examples, making overfitting more likely.

### 5.3.4 RGB vs. Grayscale Performance

RGB models consistently outperformed their grayscale counterparts, though the margin was modest (typically 1-3 percentage points). This suggests that color information provides useful but not essential signals for age-gender classification. The color information may help with subtle cues like skin tone variations associated with aging or cosmetic differences between genders.

### 5.3.5 Error Pattern Analysis

Confusion matrices revealed consistent error patterns across models:

- Most misclassifications occurred between adjacent age groups (e.g., (25, 32) and (38, 43)), highlighting the difficulty in precisely categorizing ages into distinct groups based on visual features alone.
- Gender misclassification was less common than age group errors, indicating that gender features are more distinct than age features in facial images.
- Certain age-gender combinations, particularly those with fewer training examples, showed consistently poorer classification results across all models.

This analysis suggests that while current models provide valuable age-gender classification capabilities, there remains significant room for improvement through more sophisticated architectures, data augmentation, and training strategies tailored to address the specific challenges of this multi-class classification problem.

## 6 Research Questions

### Question 1: How Models Work with Different Images

- What happens when images are not clear or blurry?
- Do facial expressions (such as smiling or not smiling) affect the results?

#### Answer Question 1

To investigate the impact of facial expressions on the predictions of the model, we examined four different emotional states: smiling, serious, mad and regular. Our analysis revealed that all models produced consistent predictions in these expressions, indicating that facial expressions do not significantly influence the results.

### Question 2: Can We Use This in Real Life?

- Will it work fast enough for apps or security cameras?
- What about when there is bad lighting or people wear glasses?

#### Answer Question 2

To evaluate real-time performance, we built a face recognition system that detects age and gender when people appear in front of the camera. Our results show that most models perform well in real-time scenarios, even with poor lighting or glasses, making it suitable for applications such as mobile applications and security cameras.

Just one model - the SVM - showed a significant decrease in performance in real-time scenarios, indicating that it may not be suitable for such applications.

### Question 3: About Age and gender prediction

- Which model is best at dealing with different age and gender?

#### Answer to Question 3

From the analysis of the models, we can see that the model with the less overfitting, and when the model got wrong predictions, it was not by much (the most of the error in the confusion matrix was in the adjacent age group), it was the MLP model.

### Question 4: Different Types of People

- Does it work the same for people from different places?
- Can we use the same model everywhere or do we need different ones?

### **Answer to Question 4**

To explore these questions, we used the UTKFace dataset, which contains images of people from diverse ethnic backgrounds with labeled age, gender, and ethnicity information. Our analysis aimed to determine whether the model performs consistently between different groups.

## 7 Challenges in Gender and Age Classification using Adience Benchmark Dataset

### 7.1 Data Structure and Organization Challenges

The initial exploration of the Adience Benchmark dataset revealed several structural challenges:

- **Redundant directory structure:** The dataset contained duplicate `faces` folders with overlapping content, requiring careful analysis to determine which directory to use without losing data.
- **Inconsistent file organization:** Files were distributed across multiple fold directories, necessitating custom parsing logic to consolidate the dataset properly.
- **Maintaining subject independence:** Ensuring that images from the same user didn't appear in both training and testing sets required tracking the `user_id` field across fold splits.

### 7.2 Data Quality and Preprocessing Challenges

The dataset presented significant quality and preprocessing challenges that required careful handling:

- **Inconsistent age group labeling:** The age labels included both specific ages (e.g., '22', '35', '58') and age ranges (e.g., '(0, 2)', '(25, 32)'), requiring a standardization process to map individual ages to appropriate ranges.
- **Overlapping age ranges:** Several overlapping age groups existed in the data:
  - Multiple variants of the same general range: (38, 42), (38, 43), and (38, 48)
  - Similar ranges with different boundaries: (25, 32) and (27, 32)
  - Conflicting categorizations: (8, 12), (8, 23), and (15, 20)
- **Undefined gender categories:** A significant portion of images (particularly in the (0, 2) age range) had undefined gender labels ('u'), requiring either removal or inferential assignment.
- **Missing data:** Several records contained NaN values that needed to be handled through appropriate imputation or removal strategies.

### 7.3 Class Imbalance Challenges

Analysis of the dataset revealed significant imbalances across different dimensions:

- **Age distribution skew:** The (25, 32) age range was heavily overrepresented (comprising 5391 samples), while other ranges like (60, 100) were underrepresented (only 896 samples).
- **Gender distribution variability:** Within specific age groups, gender distribution was uneven, with certain age-gender combinations having significantly fewer samples.
- **Disproportionate unknown gender labels:** The vast majority of unspecified gender labels ('u') appeared in the (0, 2) age range, creating a potential bias when handling these samples.

### 7.4 Feature Engineering and Extraction Challenges

Converting the raw image data into meaningful feature representations presented several challenges:

- **Computational complexity of ResNet50:** Using a deep architecture like ResNet50 required significant computational resources and careful optimization for the specific classification task.
- **Face localization and normalization:** The dataset provided bounding box coordinates (`x`, `y`, `dx`, `dy`) and orientation data (`tilt_ang`, `fiducial_yaw_angle`), but transforming these into properly aligned face images required complex preprocessing.
- **Balancing feature richness and model efficiency:** Determining the optimal dimensionality and representation of features extracted from ResNet50 to balance discriminative power with computational efficiency.

### 7.5 Overfitting and Generalization Challenges

Working with deep neural networks on the Adience dataset presented significant overfitting challenges:

- **Limited sample diversity:** Despite having thousands of images, the diversity of faces, particularly in underrepresented age groups, was insufficient to prevent the model from memorizing specific examples rather than learning generalizable patterns.
- **High model capacity versus dataset size:** The ResNet50 architecture contains millions of parameters, creating a high risk of overfitting when trained on a relatively small dataset like Adience.

- **Domain-specific feature learning:** Facial features relevant for age and gender classification needed to be learned despite variations in lighting, pose, and image quality, requiring the model to ignore irrelevant patterns that might lead to memorization.
- **Monitoring overfitting signals:** Determining reliable indicators of overfitting required careful observation of:
  - Diverging training and validation loss curves
  - Stagnating or decreasing validation accuracy despite improving training accuracy
  - Performance discrepancies across different demographic subgroups

## 7.6 Early Stopping Implementation Challenges

Implementing effective early stopping strategies to prevent overfitting introduced additional complexities:

- **Metric selection complexity:** Determining the most appropriate validation metric to monitor posed challenges:
  - Single metric for multi-task learning: Balancing the importance of age versus gender accuracy
  - Validation loss versus accuracy trade-offs: Loss might continue decreasing while accuracy plateaus
  - Class-weighted metrics: Considering whether to weight metrics to account for class imbalance
- **Patience parameter tuning:** Determining the optimal patience value (number of epochs to wait for improvement before stopping) required experimentation:
  - Too low: Risk of stopping before finding optimal weights
  - Too high: Computational inefficiency and potential overfitting
- **Learning rate scheduling interaction:** Coordinating early stopping with learning rate reduction strategies added complexity:
  - Distinguishing between temporary plateaus that could be overcome by learning rate adjustments versus genuine convergence
  - Determining whether to reset patience after learning rate changes
- **Checkpoint management:** Implementing model checkpointing alongside early stopping required decisions about:
  - Storage requirements for saving multiple model states



- Criteria for determining the "best" model (lowest validation loss versus highest validation accuracy)
- Restoration strategy when training terminates
- **Resource allocation planning:** Early stopping directly impacted computational resource planning:
  - Difficulty in estimating total training time due to variable convergence rates
  - Balancing between thorough hyperparameter exploration and efficient resource utilization

## 7.7 Multi-Label Classification Complexity

Handling the joint prediction of age and gender introduced additional complexity:

- **Decision on problem formulation:** Determining whether to treat the task as:
  - Two separate classification problems (independent age and gender models)
  - A single multi-class problem with combined age-gender categories (as suggested by the `age_gender` column creation)
  - A multi-task learning problem with shared representations
- **Loss function design:** Balancing the contribution of age and gender classification errors when using a joint model.
- **Performance evaluation complexity:** Defining appropriate metrics that reflect both age and gender classification performance.

## 7.8 Training and Validation Strategy Challenges

Establishing a robust training and evaluation framework presented several challenges:

- **Split methodology selection:** The dataset came with predefined folds, but required additional splitting for validation, raising questions about the most appropriate approach.
- **Ensuring no data leakage:** Verification was needed to ensure no overlap between train, validation, and test sets, particularly concerning images from the same subjects.
- **Managing fold information:** The original dataset structure included fold-specific information that needed to be preserved or appropriately transformed when creating the final training splits.

## 7.9 Ethical and Representational Challenges

Working with demographic data raised important ethical considerations:

- **Inferential gender assignment:** The decision to assign all unknown genders in the (0, 2) age range as male introduces potential bias in the model.
- **Age boundary ambiguity:** The overlap and inconsistency in age ranges raises questions about the certainty of predictions, particularly near boundary cases.
- **Demographic representation concerns:** The dataset's imbalances may lead to models that perform better for certain demographic groups than others, raising fairness concerns.

## 8 Real time application

As mentioned in the research questions, we built a face recognition system that detects faces in real-time and predicts the age and gender of the person. The system uses the best model we found in our analysis - the MLP model with RGB features.

The app was build using the OpenCV library for face detection and the TensorFlow library for the model prediction, and Tkinter for the GUI.

### 8.1 How the app works

- The app opens a window with a live video feed from the camera.
- For each face detected in the video feed, the app predicts the age and gender of the person. the prediction happen for each 3 frames.
- The app displays the age and gender prediction on the above the face of the person, and the confidence of the prediction.

### 8.2 Image of the app

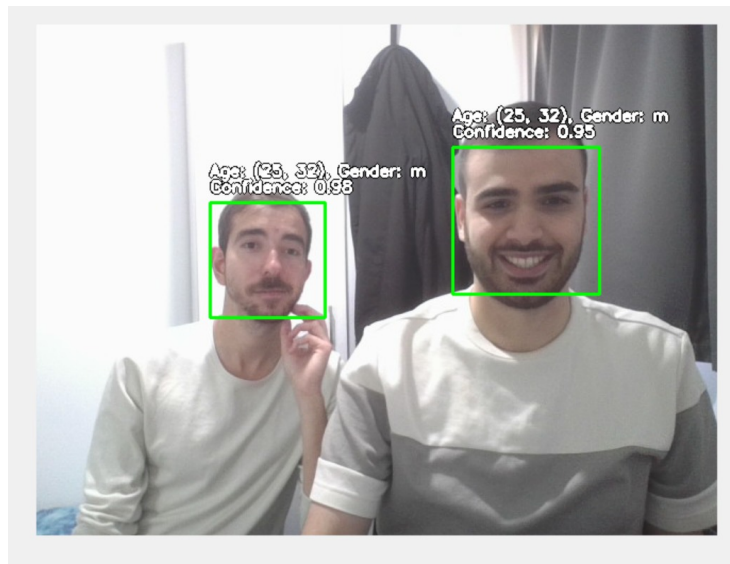


Figure 16: Real-time Age and Gender Prediction App

## References

- [1] Tung Le. Adience benchmark gender and age classification dataset. Kaggle, 2024. Available at: <https://www.kaggle.com/datasets/ttungl/adience-benchmark-gender-and-age-classification>.
- [2] Michael Dadush Shay Gali. Github repository. GitHub, 2025. Available at: <https://github.com/ShayGali/VisAge>.