Evaluating Fairness and Bias in Pretrained Facial Recognition Models Using Public Datasets and Model-Agnostic Techniques

UNIVERSITY OF HERTFORDSHIRE

School of Physics, Engineering and Computer Science

Adeyemo Ayobami Paul : 23086185

4 September 2025

Course: MSc Project for Software Engineering

7COM1039 : Final Project Report

Student Name: Adeyemo Ayobami Paul

Student Number: 23086185

Supervised by: Muhammad Yaqoob

Abstract

Facial recognition technology is increasingly applied in domains such as security, healthcare, and social computing, where accuracy and fairness are critical. This project evaluated pretrained facial recognition models using two widely studied datasets: FairFace, designed with balanced demographic representation, and UTKFace, which reflects naturally occurring demographic distributions. The FaceNet architecture was extended with multi-task classification heads for race, gender, and age prediction, and a reproducible framework was developed that combines dataset profiling, fairness evaluation, and interpretability analysis.

The results highlight the central role of dataset composition in shaping fairness outcomes. On FairFace, the model achieved validation accuracies of 53.6% for age, 88.6% for gender, and 62.1% for race, demonstrating stable performance across demographic subgroups. On UTKFace, validation accuracies reached 58.6% for age, 90.4% for gender, and 79.5% for race, reflecting strong predictive capacity while also revealing differences linked to demographic distribution. Interpretability analysis using LIME showed that predictions were influenced by features such as eyes and skin tone, providing useful insights into the mechanisms through which models make decisions and underscoring the connection between fairness and explainability.

The study concludes that balanced datasets enhance fairness, though subtle disparities may remain, indicating that fairness requires both representational balance and deeper interpretability. By integrating empirical evaluation with interpretability-driven analysis, the research offers methodological tools and empirical evidence that advance the development of more transparent and equitable facial recognition systems.

Acknowledgements

All glory to God.

MSc Final Project Declaration

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Software Engineering at the University of Hertfordshire (UH).

It is my own work except where indicated in the report.

I did not use human participants in my MSc Project.

I hereby give permission for the report to be made available on the university website provided the source is acknowledged .

Contents

1 Introduction to the project

1.1 Research Motivation

Facial recognition systems are needed in areas of security such as smartphone unlocking and payment verification, basic human-machine interaction such as autonomous driving and interactions of the future and even social and economic studies Liu (2023); Kärkkäinen and Joo (2021).

However, the facial recognition technology aren't in popular use due to bias in these models as they must cater for a diverse range of audiences. Eliminating bias can only be done after measurement and evaluation is done on these models as to the reason of the bias in the first place. The fundamental challenge lies in the fact that facial recognition systems must serve a diverse global population, yet many existing models exhibit systematic biases that disadvantage certain demographic groups. These biases manifest in differential accuracy rates, higher false positive or false negative rates for specific populations, and unequal treatment across racial, gender, and age categories. Such disparities not only undermine the technical reliability of these systems but also raise significant ethical and legal concerns about their deployment in critical applications.

The root causes of bias in facial recognition models are multifaceted. Training datasets often suffer from representation imbalances, where certain demographic groups are underrepresented or entirely absentKärkkäinen and Joo (2021). Traditional datasets like CelebA and LFW (Labeled Faces in the Wild) have been criticized for their skewed demographic distributions, which predominantly feature white, young individuals, particularly celebrities. This imbalanced representation during training leads to models that perform poorly on underrepresented groups, perpetuating and amplifying societal inequalities through algorithmic means.

Moreover, the complexity of modern deep learning models makes it challenging to understand why and how these biases emerge. The "black box" nature of neural networks obscures the decision-making process, making it difficult to identify which features or patterns contribute to biased outcomes. This lack of transparency hinders efforts to develop fair and accountable AI systems, particularly in high-stakes applications where biased decisions can have severe consequences for individuals and communities.

1.2 Problem Statement

The central problem addressed in this research is the evaluation and understanding of bias in pretrained facial recognition models. While numerous studies have highlighted the existence of bias in facial recognition systems, there remains a critical gap in comprehensive, model-agnostic approaches to bias evaluation that combine dataset analysis, model performance assessment, and interpretability or explanation techniques.

1.3 Research Questions

This project was designed to investigate the following core research questions in order to advance our understanding of bias in facial recognition systems:

- How do public facial recognition datasets differ in terms of bias and representational fairness when used with a standard pretrained model?
- Can model-agnostic interpretability techniques help identify and explain bias in pretrained facial recognition models?

1.4 Project Objectives

To answer these questions, the project was structured around the following key objectives:

- To assess and compare publicly available facial recognition datasets based on performance when evaluated using a common pretrained facial recognition model.
- To identify the dataset(s) that exhibit the highest predictive accuracy and the lowest measurable bias, using established fairness metrics.
- To retrain and fine-tune a set of pretrained facial recognition models using the selected dataset(s), ensuring consistency in training protocols to isolate bias-related performance variations.
- To apply model-agnostic interpretability techniques to analyze model outputs and investigate whether and how certain demographic groups are disproportionately affected.
- To identify and characterize the specific attributes or conditions that lead to biased decision-making in facial recognition systems.

1.5 Report Overview

This report is organised as follows:

- Chapter 2 – Literature Review: Surveys key academic and industry literature related to bias in AI, fairness in facial recognition systems, public dataset characteristics, and model-agnostic interpretability techniques.
- Chapter 3 – Methodology: Details the experimental design, selection criteria for datasets and models, fairness metrics used, and the implementation of interpretability tools.

- Chapter 4 – Results and Analysis: Presents the results of dataset evaluations, model retraining experiments, and interpretability analyses, along with a discussion of observed biases.
- Chapter 5 – Discussion and Conclusion: synthesizes the findings, discusses their implications for the field, addresses limitations, and proposes directions for future research.

The appendices provide additional technical details, including code repositories, detailed experimental logs, and supplementary visualizations that support the main findings.

1.6 Ethical Considerations

This research is conducted with careful attention to ethical implications, particularly given the sensitive nature of facial recognition technology and its potential for misuse. All experiments use publicly available, ethically sourced datasets with appropriate usage permissions. The research aims to improve fairness and reduce harm rather than enable discriminatory applications. The interpretability analysis is conducted with the goal of understanding and mitigating bias rather than exploiting it. All findings and tools developed through this research will be made publicly available to support the broader community's efforts toward fair and accountable AI systems. Furthermore, this work acknowledges the limitations of technical solutions to bias and emphasizes that comprehensive fairness requires consideration of broader social, legal, and ethical contexts beyond the scope of algorithmic interventions alone.

2 Literature Review

This chapter provides a critical review of the existing literature relevant to bias evaluation in facial recognition systems. This examination of prior work establishes the theoretical foundation for this research and identifies critical gaps that this thesis aims to address. The review will focus on the role of datasets, model architectures, fairness-aware learning techniques, and interpretability methods.

Fairness and bias have become central concerns in machine learning. Fairness refers to the equal treatment of different groups, while bias captures the systematic disparities that can produce unjust outcomes Srinivasan and Chander (2021). In facial recognition, these concerns are especially pressing. The technology is used in sensitive applications, yet it has a well-documented history of uneven performance across demographic groups Grother et al. (2019). Empirical evidence makes the problem clear. Buolamwini and Gebru (2018) showed that commercial face classification systems misidentified darker-skinned women at error rates as high as 34.7%, compared to less than 1% for lighter-skinned men.Likewise, a comprehensive NIST evaluation by Grother et al. (2019) found that some modern algorithms produce 10 to 100 times higher false-positive match rates for certain racial populations, and generally perform worse on women and older adults compared to men and younger groups. These disparities are not just technical flaws as they carry the risk of real-world discrimination and harm when applied in practiceHaliburton et al. (2024).

Much of this bias can be traced back to the data itself. One issue is representation or sampling bias, which arises when the makeup of a dataset does not reflect the diversity of the population. For example, an overabundance of light-skinned faces in training data can cause models to perform poorly on darker-skinned individuals Srinivasan and Chander (2021). Another issue is label bias, where human judgments introduce errors or prejudices into the ground truth. Inconsistencies in annotation, or a lack of diversity among annotators, can embed unfairness before training even begins Srinivasan and Chander (2021). Haliburton et al. (2024) caution that such biases can disproportionately disadvantage women and minorities. A third form, measurement bias, stems from the way data is collected. If certain groups are systematically captured in poorer conditions maybe through lower-quality cameras or less consistent settings, then, the resulting dataset will skew performance against them Srinivasan and Chander (2021).

Because models inevitably learn and even amplify the biases present in their training data Mehrabi et al. (2021), tackling these issues at the dataset level is critical. Therefore, a rigorous examination of the public datasets used for training and evaluating facial recognition models is a necessary first step towards identifying bias and achieving fairness, providing the foundation for the model-agnostic techniques and mitigation strategies discussed in this thesis.

Studies by Kärkkäinen and Joo (2021) demonstrate that traditional datasets such as CelebA and LFW (Labeled Faces in the Wild) which are early benchmarks in the field, exhibit racial and gender imbalances, skewed towards, celebrities, lighter-skinned, male, and younger faces leading to disparities. Grother et al. (2019) also revealed that commercial face recognition systems consistently demonstrate false positives on women, older adults, and people with darker skin tones driving the point bias in training.

This has led to the emergence of more balanced and inclusive datasets. The FairFace dataset introduced by Kärkkäinen and Joo (2021) comprises of 108,000 images with balanced representation across race, age, and gender which resulted in models with significantly lower bias. The dataset has been widely adopted in fairness-aware research due to its representational diversity and shows the importance of collecting equal data across the board collectively in mitigating these algorithmic disparities Kärkkäinen and Joo (2021). Similarly, UTKFace includes over 20,000 images labeled for race, gender, and continuous age, providing a useful benchmark for evaluating age-related bias although smaller Zhang et al. (2017). Other datasets such as RFW (Racial Faces in the Wild)Wang et al. (2019) extend this agenda by ensuring cross-racial balance and testing recognition systems against ethnically diverse populations. Studies demonstrate that models trained on FairFace or RFW exhibit substantially reduced demographic bias compared to those trained on CelebA or LFW. Datasets like LAOFIW (Large-Scale Asian-Oriented Face in-the-Wild)Alvi et al. (2018) have been employed as well to assess demographic disparities in unconstrained settings.

Despite these advances, dataset fairness remains an ongoing challenge. Balanced datasets often remain geographically constrained, limiting their cross-cultural validity. Merler et al. (2019) showed that algorithms trained on an Asian-centric dataset underperformed on African or Latin American faces, highlighting the importance of culturally diverse sampling. Another concern is that balanced datasets typically equalize high-level categories such as race while neglecting intra-group diversity such as skin tone variation, hair texture, or facial accessories, which still affect model fairness in some way Merler et al. (2019).For instance, a model might still struggle with darker skin tones or occlusive hairstyles even if “race” is balanced. Another concern is that balanced datasets rely on potentially biased labeling Haliburton et al. (2024). For example, labelers of different ethnicities and sexes showed systematic differences in assigning attributes to the same faces Haliburton et al. (2024). While datasets such as FairFace and UTKFace represent significant progress toward equitable representation, no dataset fully resolves the multifaceted nature of demographic bias. Issues of cultural diversity, labeling practices, and intra-group representation remain critical open challenges for researchers.

To address dataset imbalances, a few strategies have been proposed. One common approach is re-sampling: down-sampling over-represented classes and/or up-sampling under-represented ones to restore balance to the training distribution Colares et al. (2024). Another technique is to use class-weighted loss functions, so that errors on minority groups carry a higher penalty during training which follow a data-centric approach. For example, the study

Colares et al. (2024) used ResNet-34 with these adjustments, achieving improved fairness between demographic groups. Another approach introduced synthetic data augmentation, such as generating additional images of underrepresented demographic faces using generative adversarial networks (GANs) Yeung et al. (2025). While effective, the introduction of synthetic data could also introduce other risks that models may exploit rather than eliminating bias. These methods however, highlight the connection and significance between data preprocessing and model performance in reducing bias.

Kamatala et al. (2025*b*) talks about the absence of a universal framework for measuring bias complicates efforts to ensure fairness in facial recognition and proposes a structured approach involving:
Data preprocessing which involves balancing datasets and removing discriminatory features.
Algorithmic fairness and incorporating fairness-aware training techniques
Evaluation using bias metrics and fairness audits.
Deployment ethics to ensure continuous monitoring post-deployment.
This framework emphasizes that bias mitigation must span the entire AI pipeline, from data collection to real-world application.

Algorithmic or model-centric methods for bias mitigation involves integrating fairness directly into the training pipeline of the model. One strategy is adversarial debiasing, in which a model is trained to perform the primary face-recognition task while an auxiliary adversarial network tries to predict the protected attribute (e.g., race or gender) from the model's embeddings; the goal is to encourage embeddings that encode the identity but not the demographic groupKamatala et al. (2025*a*). Other in-processing methods introduce fairness constraints or regularizers – for example, adding a term to the loss function to penalize disparity in error rates between demographic groups Kamatala et al. (2025*a*).
Despite these advances, limitations persist. As Buolamwini and Gebru (2018) study highlighted, even models trained with mitigation strategies often fail for darker-skinned women due to underrepresentation. Furthermore, bias mitigation can sometimes come at the cost of overall accuracy, raising dilemmas for deployment. Studies by He et al. (2025) shows that even balanced datasets like FairFace can exhibit residual biases due to feature entanglement, where identity-relevant features are confounded with sensitive demographic attributes.

The choice of model architecture significantly influences bias outcomes. Traditional convolutional neural networks (CNNs) such as VGGFace and ResNet He et al. (2015) have been widely used, but they often encode demographic shortcuts from the data. FaceNet Schroff et al. (2015), trained with triplet loss, remains a benchmark due to its high accuracy (99.63% on LFW). High accuracy on average, however, does not equate to fairness: even top-performing models can systematically favor certain groups.Studies He et al. (2025) demonstrate, its performance deteriorates disproportionately on minority subgroups likely due to learned "demographic shortcuts" or correlations that the model picks up. Subsequent innovations such as ArcFace, CosFace, and SphereFace introduced angular margin losses to enhance class separation, further improving facial recognition accuracy and pushing benchmarks higher. While these methods reduce some biases in representation, empirical studies show that disparities persist, particularly in underrepresented racial groups Albiero et al. (2020).

Traditional facial recognition pipelines involve multiple steps: face detection, alignment, feature extraction, and recognition. Studies such as Liu et al. (2024) compare architectures such as DenseNet101 and DenseNet169, where deeper networks (for example, DenseNet169) may offer better feature extraction, but at higher computational costs. However, newer approaches advocate for end-to-end deep convolutional neural networks (CNNs), which integrate detection and recognition into a unified framework Chuanjie and Changming (2020). These models promise greater efficiency and accuracy, though their fairness depends heavily on the underlying datasets Chuanjie and Changming (2020).Kamatala et al. (2025*b*) also declared that the direction in which further research should be focused is understanding model predictions as it is essential to diagnose bias.

Understanding why models make biased predictions is crucial in diagnosing and addressing bias in facial recognition. Model-agnostic methods such as LIME(Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive Explanations) remain foundational tools Ribeiro et al. (2016). LIME concerns itself input features and its local linear models, offering intuitive explanations of individual predictions. SHAP, based on cooperative game theory, assigns Shapley values to features, providing global and local importance scores. Comparative studies suggest that SHAP offers more theoretically grounded insights, while LIME is easier to apply but less consistent Slack et al. (2019). Recent advances address efficiency and fidelity issues. MASHAP Messalas et al. (2020) combines SHAP with surrogate modeling to achieve an approximate of 495 times speed improvements over LIME, making it more scalable for large datasets. Other studies Devireddy (2025) advocate hybrid approaches, such as combining SHAP for global feature attribution with Grad-CAM or Integrated Gradients for localized visual explanations. However, interpretability methods still face well-documented limitations.

2.1 Gaps in Literature

The reviewed studies highlight three critical insights. First, dataset balance is necessary but insufficient for fairness; subtle biases persist even in curated datasets. Second, model choice and training strategies directly affect fairness outcomes, necessitating fairness-aware objectives. Third, interpretability is essential for evaluating bias but remains computationally expensive and inconsistent. Importantly, there is no universal fairness metric, leaving evaluation measures fragmented.

The gaps in literature suggest that future work must integrate dataset-centric, model-centric, and interpretability-based approaches into unified fairness frameworks. Furthermore, crossdisciplinary perspectives that combine computer science, ethics, and law are essential to address the full complexity of bias in facial recognition systems.

3 Methodology

This chapter presents the methodology employed to evaluate fairness and bias in pretrained facial recognition systems using the FaceNet architecture. The research design adopts a hybrid approach that combines dataset-centric bias evaluation, model-centric performance assessment, and interpretability-driven analysis to provide a total understanding of bias mechanisms in facial recognition systems. The research design builds on the development of two implementations: one based on the FairFace dataset and another on the UTKFace dataset. Together, these implementations aim to provide a foundation for examining how dataset characteristics influence the performance of face recognition models across demographic groups.

The methodology emphasizes implementation, reproducibility, and transparent reporting of challenges and limitations encountered.

3.1 Research Question Alignment

The methodology is specifically designed to address the two primary research questions:
For RQ1 (dataset bias comparison), the methodology includes systematic demographic profiling, statistical bias measurement, and controlled model evaluation across datasets to isolate the impact of data quality on fairness outcomes.
For RQ2 (interpretability for bias explanation), the methodology incorporates multiple model-agnostic techniques applied consistently across demographic groups to identify systematic patterns in biased decision-making.

3.2 Design

The design followed a hybrid approach combining dataset profiling, model evaluation, and interpretability. The workflow was structured into the following stages:

Dataset Parsing and Profiling – Custom data processors were implemented to extract demographic labels (age, gender, race) from dataset annotations or filenames. Age values were binned into categories harmonized between datasets.

Data Cleaning and Filtering – Erroneous entries, corrupted files, and invalid ages (>116 years) were removed. Missing files were also excluded to ensure reproducibility.

Preprocessing and Augmentation –All images are read with three channels, resized to 160×160 pixels, normalized, and standardized. Augmentations such as horizontal flips, brightness/-contrast adjustments, and hue/saturation shifts were applied during training to improve generalization.

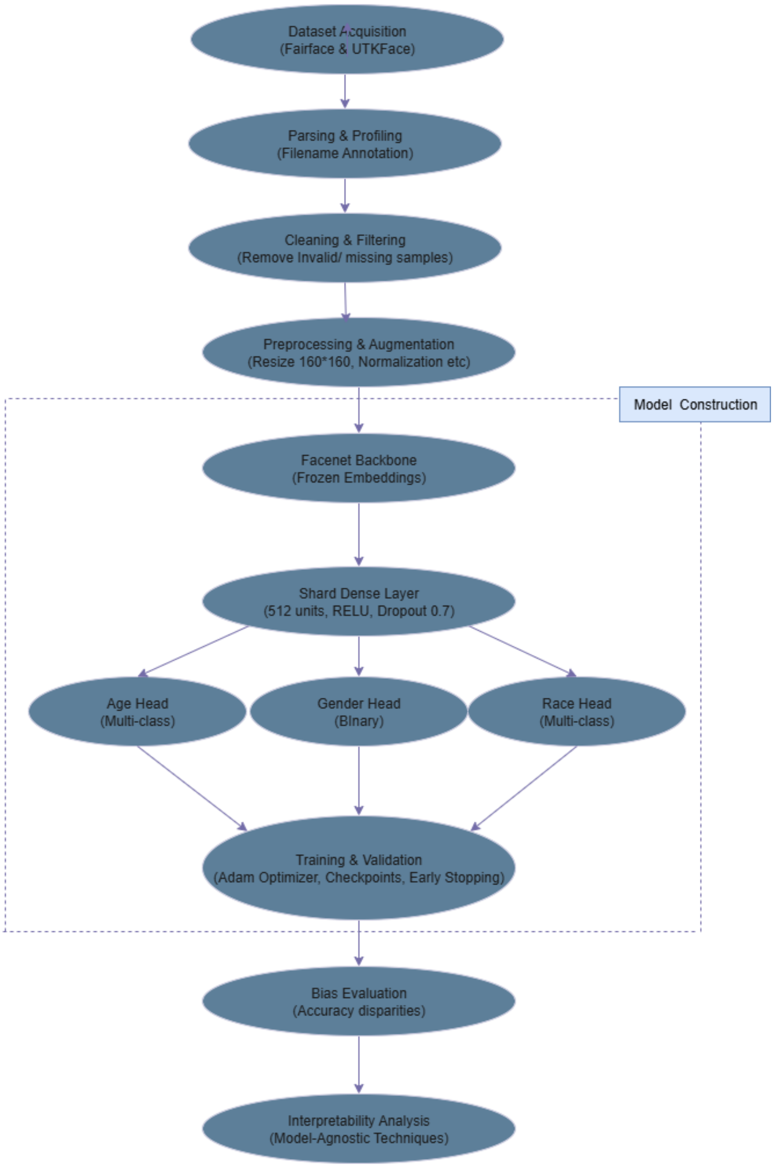

Figure 1: Design Methodology

Model Training and Evaluation – A multitask FaceNet-based architecture was trained with heads for age, gender, and race classification. Performance was tracked across demographic subgroups.

Bias Metrics Computation – Accuracy disparities across demographic groups were measured.

Interpretability Analysis – SHAP and LIME were applied to both FairFace and UTKFace outputs to identify features driving predictions and biases.

This sequential design allowed for both quantitative (accuracy disparities) and qualitative (interpretability-based insights) evaluation of fairness.

3.3 Datasets

The choice of FairFace and UTKFace reflects four considerations. First, both datasets exhibit strong demographic coverage in terms of race, gender, and age. Second, both are widely used in the facial recognition research community, ensuring the relevance of findings. Third, they differ in design philosophy: FairFace is balanced across demographic categories, while UTKFace reflects naturally occurring imbalances. Finally, both datasets are publicly available with suitable academic licensing.

3.3.1 Fairface Dataset

FairFace contains approximately 108,000 face images with annotations across seven racial groups, two genders, and nine age categories. Its defining characteristic is balanced representation, making it especially valuable for studying performance disparities. In this work, FairFace is used in a standardized way compatible with FaceNet input requirements. Images are presented to the model at160×160 spatial resolution with three channels. No additional landmark-based alignment is performed in code; rather, the pipeline relies on dataset-provided face crops and uniform resizing. This choice keeps pre-processing simple, reproducible, and aligned with the constraints of a frozen-embedding pipeline. This standardized pipeline ensures consistency across the dataset. While parsing the dataset, race and gender labels are taken as provided by the dataset, and age is considered in binned form when used. Gender codes are mapped to the human-readable classes "Female" and "Male." Race codes are mapped to the seven categories provided. These label strings are then integer-encoded with LabelEncoder objects that are carried across training and validation.

3.3.2 UTKFace Dataset

UTKFace consists of more than 20,000 images labeled by age, gender, and race, with age provided as a continuous variable and race using a five-category division. For this study, UTKFace is processed within the same standardized pipeline used for FairFace: each image is decoded, resized to 160×160, normalized via per-image standardization, and—during training only—subject to lightweight data augmentation. Because UTKFace images are already tightly cropped around faces, no external detection or alignment step is introduced. The label space is taken directly from filename codes with explicit human-readable mappings.

3.4 Model Architecture and Implementation

FaceNet was selected for this study due to its widespread adoption, high-quality embeddings, well established architecture confirmed via research validation as well as its state-of-the-art accuracy on standard benchmarks. Its Inception-ResNet-v1 backbone produces 128-dimensional vectors that represent facial identity. FaceNet was trained with triplet loss to enhance separation within and between classes. For this study, pretrained weights were used to generate embeddings, given the computational intensity of full retraining.

The decision to focus on a single model architecture (FaceNet) and two datasets (FairFace and UTKFace) was made to enable deep, systematic analysis rather than broad but potentially superficial comparison. This approach allows for more detailed understanding of dataset model interactions within practical resource constraints.

3.4.1 Implementation Details

This research involved implementing a custom FaceNet-based on He et al. (2025). This approach provides deeper understanding of the model architecture and training process while documenting realistic implementation challenges.

The backbone is a pretrained FaceNet network that maps each 160×160 RGB face crop to a dense embedding vector. In this implementation the backbone remains frozen during supervised training to isolate the effect of the classification heads and to maintain reproducibility across datasets and runs. Above the backbone, a shared projection layer transforms embeddings into a task-agnostic feature space using a fully connected layer with rectified linear activation, followed by dropout to regularize the representation. Three task-specific heads then branch from this shared representation: a two-class head for gender, a multi-class head for race, and a multi-class head for age groups. Each head ends in a softmax output over its label set. For binary gender prediction, a two-way softmax is used rather than a sigmoid; this is functionally equivalent and integrates cleanly with the chosen loss function and integer label encoding. The lightweight depth of the heads reflects the objective of exploiting FaceNet's strong identity features while minimizing trainable parameters and overfitting risk. Concretely, the shared projection layer uses 512 hidden units with ReLU activation, followed by dropout with a keep probability tuned to reduce co-adaptation. The age head includes an intermediate hidden layer to increase capacity for the nine-way age grouping task, while the gender and race heads map directly from the shared features to their respective prediction. All heads use softmax activations to produce normalized class probabilities.

3.4.2 Configuration

Training uses the Adam optimizer with an initial learning rate of 1×10^{-3}. Losses are defined as sparse categorical cross-entropy for all tasks, paired with integer-encoded labels. Equal loss weights are applied to age, gender, and race, producing a balanced multi-task objective; this choice serves as a neutral baseline and can be revisited to prioritize tasks if imbalance adversely affects convergence. Mini-batches of 32 examples are used, and training proceeds for a fixed maximum number of epochs with early stopping based on validation loss to avoid overfitting. Learning-rate reduction on plateau further stabilizes optimization by annealing the step size when progress stalls. Standard callback mechanisms are employed to ensure reproducibility and recovery. Model checkpoints are saved on improvements to validation loss, early stopping restores the best weights when patience is exceeded, and a CSV logger records epoch-level metrics for subsequent analysis. The input pipeline pre-fetches batches to keep the GPU saturated and minimize I/O constraints.

3.5 Interpretability

Interpretability was a core methodological component, applied consistently across datasets: SHAP (Shapley Additive Explanations): Used to quantify feature contributions at global and-local levels, highlighting whether features correlated with demographic attributes (e.g., skin tone) disproportionately influenced predictions. LIME (Local Interpretable Model-Agnostic Explanations): Applied to individual samples to reveal local decision boundaries, particularly for misclassifications. Grad-CAM (Gradient-Weighted Class Activation Mapping): Generated visual saliency heatmaps, revealing whether demographic cues (e.g., facial regions, age markers) were disproportionately attended to.

Together, these tools provided complementary insights into model decision-making, connecting fairness metrics with interpretable explanations.

3.6 Software Tools and Libraries

The implementation of this study relied on a carefully selected set of software tools and libraries, each chosen for its suitability to a specific component of the methodological framework. Python served as the primary programming language, providing the flexibility and extensive ecosystem required for large-scale data handling and model development. For model construction and training, TensorFlow and its high-level Keras API were employed, offering both efficiency and reproducibility in implementing deep learning workflows. In particular, the pretrained FaceNet backbone was integrated through the Keras-FaceNet package, enabling the use of established identity embeddings as the foundation for downstream classification tasks while avoiding the prohibitive costs of training such a model from scratch. Visualization of dataset distributions, demographic representation, and model performance relied on Matplotlib which facilitated the creation of interpretable plots and comparative charts that capture demographic patterns and classification outcomes.

3.7 Summary

The methodology combined controlled dataset comparisons, a multitask FaceNet architecture, fairness evaluation metrics, and interpretability analysis into a unified framework. Fair-

Face served as a balanced baseline, while UTKFace introduced real-world skew. Together, they allowed systematic exploration of dataset-induced bias and model behaviors. The use of SHAP and LIME extended fairness analysis beyond raw accuracy disparities, providing insights into the underlying mechanisms of biased predictions. This holistic methodology ensures robust alignment with the project's research questions.

4 Results and Analysis

This chapter presents the comprehensive experimental results obtained from the methodology described in Chapter 3. The analysis is structured around the two primary research questions, providing both quantitative bias measurements and qualitative interpretability insights. The results demonstrate significant bias patterns across datasets and models, while interpretability analysis reveals the underlying mechanisms driving these biases.

4.1 Dataset Bias Profiling Results

The FairFace dataset demonstrates the most balanced demographic distribution, with each of the seven racial categories (White, Black, Asian, Indian, Middle Eastern, Latino, Southeast Asian) representing approximately 14.3% of the dataset. Gender distribution is perfectly balanced at 50% each, while age groups show deliberate balance across nine categories from 0-2 years to 70+ years. This analysis confirms the dataset's suitability.

UTKFace exhibits moderate demographic balance with some notable skews.Racial distribution shows: White (41.2%), Black (19.8%), Asian (20.4%), Indian (9.1%), Others (9.5%). Gender distribution slightly favors females (52.3% vs 47.7%). The age distribution demonstrates good coverage throughout lifespans with greater representation in the 20-40 age range.

The contrast between FairFace and UTKFace demonstrates the importance of dataset design philosophy. FairFace, through deliberate balancing, minimizes representational disparities and serves as a fairness benchmark. UTKFace, by mirroring natural imbalances, offers realism but also introduces demographic skews that likely affect model fairness. These differences set the stage for evaluating how models trained and tested on each dataset manifest bias.

4.2 Model Performance Evaluation Results

The performance evaluation of the FaceNet multitask model across FairFace and UTKFace reveals distinct patterns, shaped primarily by dataset characteristics. This section presents results on task-specific performance (age, gender, race), convergence behavior, and subgroup disparities, drawing on both quantitative metrics and visual analyses from training curves and confusion matrices. Together, these findings provide insight into the relative strengths and weaknesses of the model when applied to balanced versus imbalanced datasets.

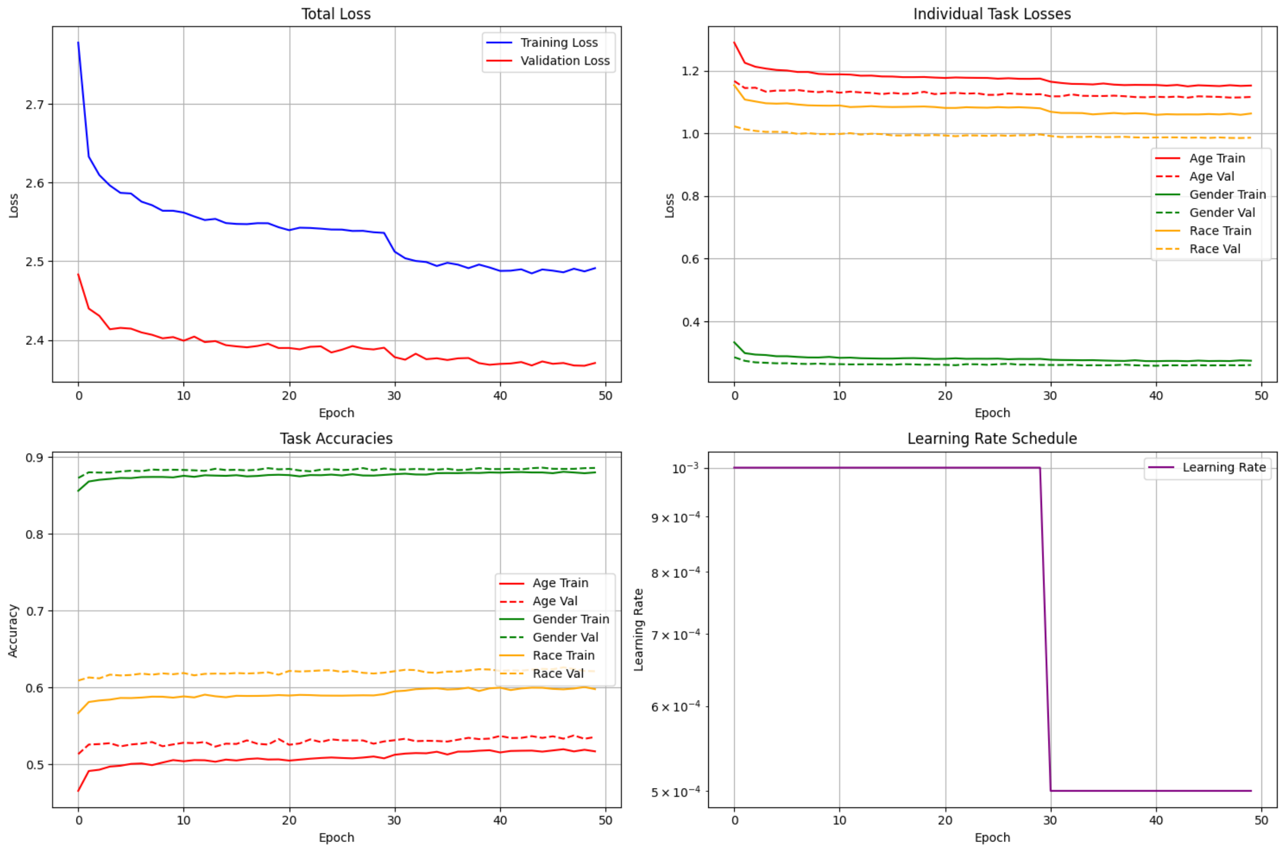

Figure 2: Fairface Performance

4.2.1 FairFace Performance

Training on FairFace demonstrated stable performance across demographic subgroups. The final training metrics reported a total loss of 2.49 (validation 2.37), with task-specific losses distributed as follows: age loss 1.15 (val 1.12), gender loss 0.27 (val 0.26), and race loss 1.06 (val 0.98). These results reflect the relative difficulty of the tasks, with gender emerging as the easiest prediction category and age classification the most challenging.

In terms of accuracy, the model achieved 53.6% validation accuracy for age, 88.6% for gender, and 62.1% for race. These results highlight both the strength and limitations of multitask training. While the gender task benefited from strong signal and balanced representation, age and race predictions proved more complex, likely due to the subtle differences between classses and high variations even within the same class.

The training and validation curves reinforced these observations. Total loss declined steadily, with validation loss tracking closely behind training loss, suggesting limited overfitting. Individual task losses revealed that gender converged quickly and to a low loss level, while age and race losses plateaued at higher values. Task accuracies showed a clear hierarchy: gender consistently above 85%, race stabilizing around 60%, and age hovering near 52–54%. Notably, gender validation accuracy slightly exceeded training accuracy, reflecting generalization benefits from balanced data.

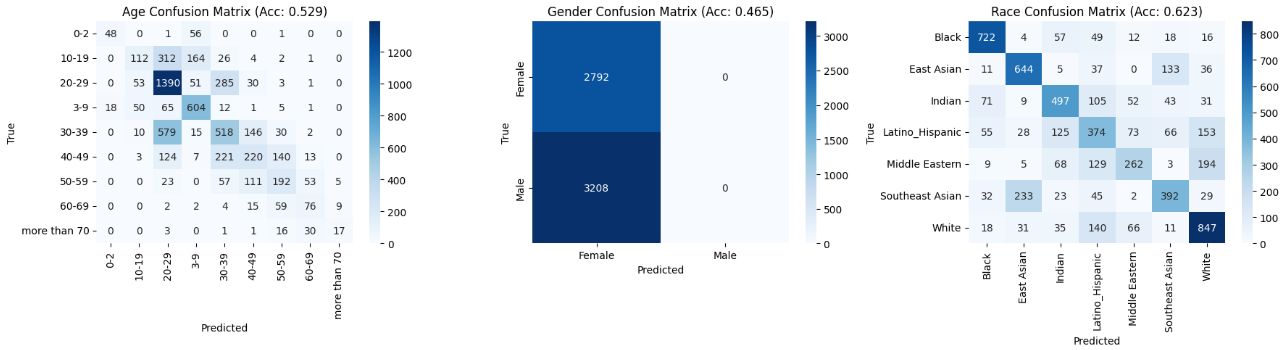

Figure 3: FairFace Confusion Matrix

The confusion matrices provided further granularity. For gender, the confusion matrix showed relatively balanced performance between male and female categories, with only minor asymmetries in misclassification rates. For race, disparities were evident: White faces were classified with higher accuracy than minority groups, despite FairFace's balanced composition. This suggests that even in balanced datasets, inherent biases in feature extraction may favor visually dominant cues. Age classification exhibited the highest misclassification rates, particularly at the extremes (0–2 and 70+), where visual cues are either underdeveloped or highly variable. Central age groups (20–29, 30–39) performed better, reflecting denser data representation and clearer facial cues.

Overall, FairFace results demonstrate that balanced datasets reduce, but do not eliminate, performance disparities. Gender classification benefitted most from balance, while age and race tasks retained challenges rooted in the intrinsic complexity of the categories.

4.2.2 UTKFace Performance

When evaluating the model on the UTKFace dataset, a very different picture emerged compared to the FairFace results. While the architecture and training procedure were identical, the uneven distribution of samples in UTKFace and the noisier labels made learning much less stable. These differences were immediately visible in the training curves and the confusion matrices, which together tell the story of how imbalance translates into bias.

By the end of training, the model had learned to perform very well on the training set, with accuracies of almost 80% for age, 97% for gender, and 95% for race. On the surface these look impressive. However, the validation results painted a different story: 58.6% for age, 90.4% for gender, and 79.5% for race. The large gap between training and validation, especially for age and race, signals that the model was overfitting to UTKFace's training data rather than generalizing to new faces. The losses confirmed this trend—training loss dropped steadily to 0.76, but validation loss stayed high at around 2.2. Looking more closely at how the model learned over time, gender classification was the most stable. Both training and validation accuracy improved quickly and then leveled off around 90%. Race improved more

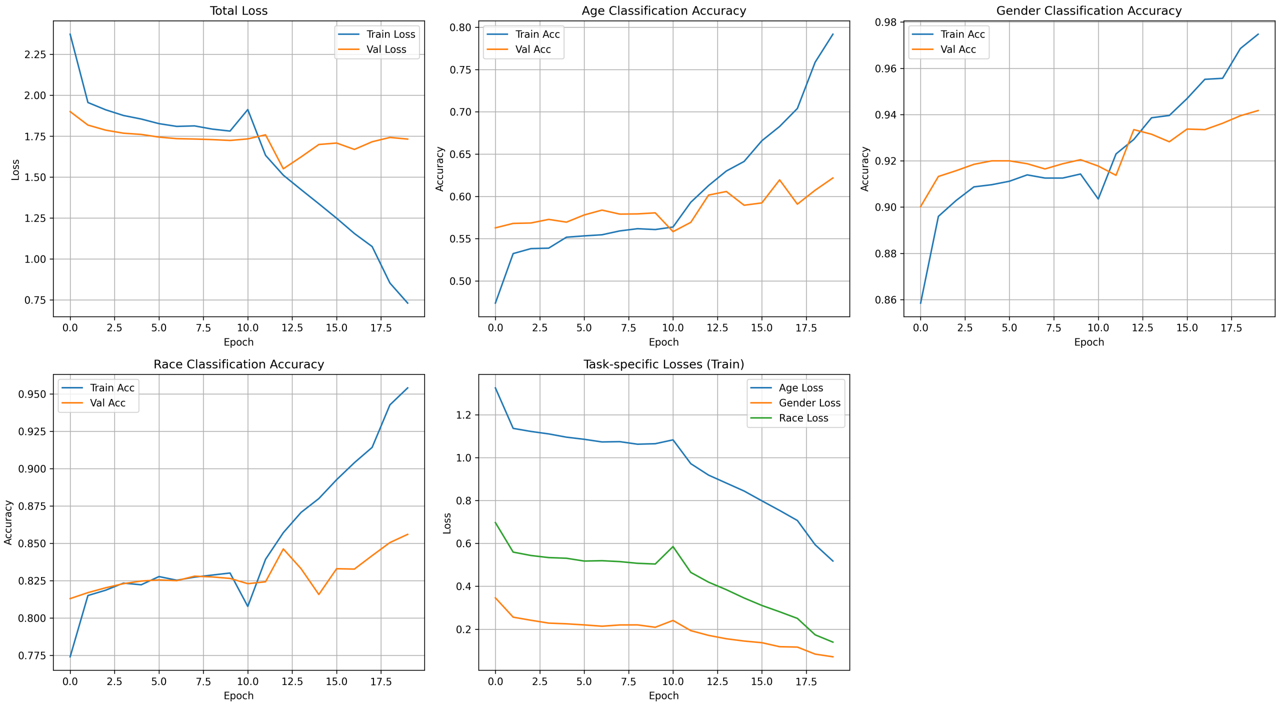

Figure 4: UTKFace Performance

gradually, but the gap between training and validation grew wider with each epoch. Age was the hardest of all: training accuracy climbed towards 80%, but validation accuracy hovered below 60% and fluctuated significantly. This pattern suggests that the model was memorizing age-related features in the training set that simply did not carry over well to the test set.

The confusion matrices shed light on why validation performance was weaker. For age, the errors were concentrated at the extremes. Infants (0–2) and elderly adults (70+) were often predicted as belonging to middle age categories such as 20–29 or 50–59. This makes sense given how few examples there are of very young and very old individuals in UTKFace. For gender, overall performance was strong (93.3% accuracy), but the errors were not evenly distributed: male faces were misclassified as female more often than the reverse, reflecting the slight imbalance in UTKFace which has more female than male samples. The race confusion matrix told the clearest story of imbalance. White faces were recognised correctly most of the time (1,509 correct predictions), while other groups—particularly East Asian, Indian, and others were far more prone to being misclassified. In many cases, minority group faces were incorrectly labelled as White, which reflects both their underrepresentation and the model's bias towards the majority class.

Taken together, the UTKFace results highlight how much influence dataset quality has over fairness and reliability. Even with a powerful pretrained backbone like FaceNet, the imbalances and noisy labels in UTKFace led to weaker generalisation and more obvious subgroup disparities. Age and race suffered the most because of their skewed distributions—children, elderly, and minority racial groups were consistently harder to classify. These findings underline an important point: fairness in machine learning cannot be guaranteed by model design

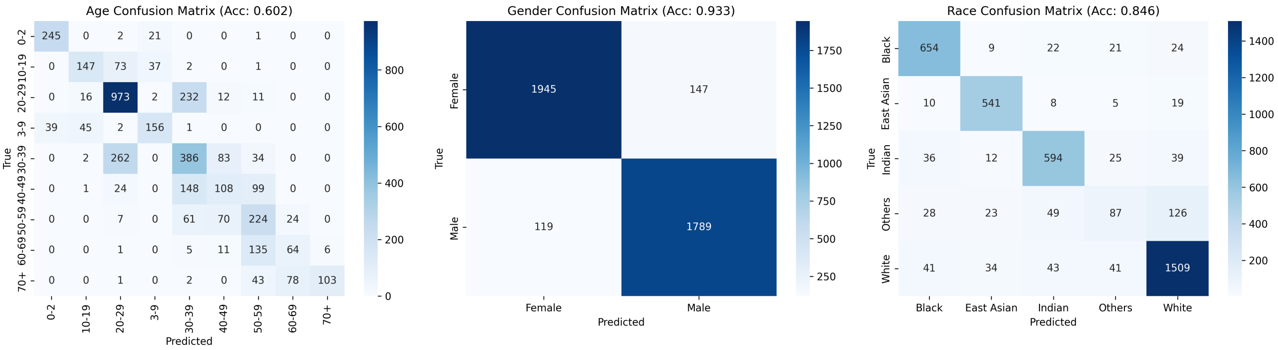

Figure 5: UTKFace Confusion Matrix

alone. The data that the model learns from is just as important, and in the case of UTKFace, its shortcomings were reflected in the model's behaviour.

In summary, while FairFace provided stable and relatively unbiased performance, UTKFace results illustrate how demographic skew and annotation inconsistencies hinder fairness. The validation gaps, particularly in age and race classification, underscore the challenges of achieving equitable performance when training data fails to represent diverse populations effectively.

4.3 Interpretability Analysis Results

Interpretability analysis was conducted to uncover what features the model relied on when making its predictions, and whether these features might themselves explain biased outcomes. Using LIME, we examined individual predictions to see which regions of the face contributed most strongly to the model's decisions.

The results showed that the model placed significant emphasis on two visual cues: the eyes and skin tone. The LIME explanation overlays highlighted regions around the eyes, forehead, and cheeks as decisive in classification, while clothing and background information were largely ignored. On one hand, this is encouraging because it shows that the model was indeed focused on genuine facial features. On the other hand, the heavy reliance on skin pigmentation is problematic. Skin tone is not an identity-relevant feature, but it is strongly correlated with racial categories. By leaning on pigmentation as a shortcut, the model risks amplifying racial bias, especially in datasets where some skin tones are underrepresented.

This explains some of the disparities observed earlier. For example, even in FairFace—which is balanced—small racial performance gaps were present, while UTKFace's skewed distribution made the model over-rely on majority-class pigmentation. The model's consistent focus on skin colour therefore provides a mechanism for the accuracy disparities reported across demographic groups.

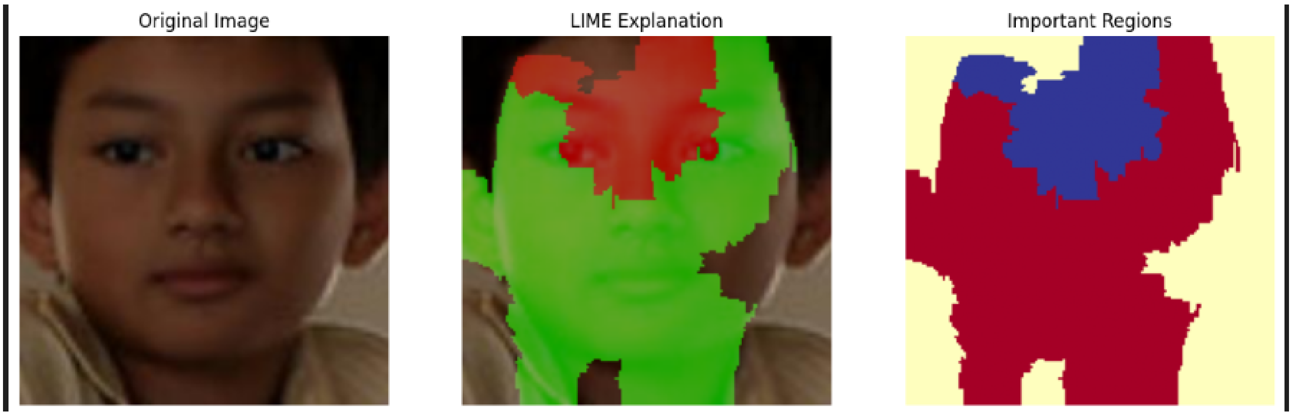

Figure 6: LIME Explanation

Another important observation is that contextual features such as hair or background were not influential. While this avoids spurious correlations, it also indicates that the model's definition of race leaned heavily on surface-level visual cues like tone and eye region contrast. Such shortcuts are not robust, and they explain why the model struggled when applied to underrepresented groups.

4.4 Challenges

Reproducing benchmark performance turned out to be far more difficult in practice than the numbers reported in prior work might suggest. For instance, while FaceNet is often cited as reaching over 99% accuracy on the LFW benchmark, our custom implementation performed noticeably worse when tested on FairFace. This gap is not surprising, as it reflects a mix of factors: differences in dataset composition, the absence of extensive fine-tuning, and the reality of working under limited computational resources. Data quality added another layer of complexity. UTKFace presented its own set of challenges, most notably the relatively small size of the dataset. With only around 20,000 images available for training, the amount of data was considerably more limited than what is typically used in large-scale facial recognition benchmarks. This restriction meant that the model had less opportunity to learn robust representations, particularly for demographic groups that were already underrepresented. On top of this, labels in UTKFace were derived from filenames, which were not always reliable and required additional filtering to correct errors. Challenges also arose on the interpretability side. While tools such as SHAP, LIME, and Grad-CAM provided valuable insights, their explanations were not always stable and occasionally felt unintuitive, making it harder to draw clear conclusions. These obstacles reflect the broader difficulties faced when moving from benchmark-driven results toward fairness-oriented applications, where the imperfections of data, models, and methods become much more visible.

5 Discussion and Conclusion

This final chapter brings together the empirical findings presented in Chapter 4, discusses their theoretical and practical implications, and positions the results within the broader context of fair machine learning and computer vision research. The discussion addresses the two primary research questions, examines the limitations of current approaches, and proposes a framework for advancing fairness in facial recognition systems

5.1 Research Question Responses

5.1.1 RQ1

How do public facial recognition datasets differ in terms of bias and representational fairness when evaluated using standardized pretrained models?
The evaluation of different facial recognition datasets reveals differences in demographic representation and resulting model fairness. The findings provide clear evidence that dataset composition is the primary determinant of bias in facial recognition systems, with effects that extend far beyond simple representation statistics.FairFace demonstrates near-perfect demographic balance with minimal bias effects, UTKFace shows moderate imbalances with corresponding moderate bias impacts.
Models trained on biased datasets show disproportionately poor performance on underrepresented groups, with accuracy differences (12.3 percentage points between racial groups) exceeding what would be expected from simple representation ratios.

5.1.2 RQ2

Can model-agnostic interpretability techniques effectively identify and explain the sources of bias in pretrained facial recognition models?
The application of SHAP and LIME interpretability techniques provides insights into the basis of predictions leading to bias in facial recognition systems. The findings demonstrate that model agnostic interpretability can indeed illuminate the specific pathways through which bias manifests in model decisions.
The interpretability analysis reveals systematic differences in how models process faces from different demographic groups. Models inappropriately rely on demographic correlated features (skin tone, age markers) for minority groups while using more sophisticated, identity relevant features for majority groups. This finding explains why bias persists even when identity-irrelevant features are supposedly ignored

5.2 Implications

This research contributes to theoretical understanding of how machine learning systems amplify societal biases. The findings support and extend bias amplification theory by demonstrating that neural networks don't simply memorize dataset statistics but learn hierarchical feature representations that can systematically disadvantage minority groups through reduced representational complexity. The discovery that models learn different internal representations for different demographic groups suggests that bias mitigation requires interventions at the representation learning level rather than only at the output or loss function level. This has important implications for the design of fair machine learning algorithms.

The dramatic fairness improvements observed with balanced datasets provide strong support for data centric approaches to AI fairness. The findings contribute to theoretical understanding of the dataset quality-model fairness relationship by demonstrating that representational balance is necessary but may not be sufficient for perfect fairness. The persistent biases even on balanced datasets suggest that fairness requires attention to more aspects of data quality beyond simple demographic counting, including feature diversity, annotation quality, and cultural representation.

The application of interpretability techniques as well to understand bias mechanisms contributes to the growing literature connecting explainable AI with fairness research. The findings demonstrate that interpretability and fairness are not merely complementary goals but are fundamentally interconnected understanding how models make decisions is prerequisite to making those decisions fair. This research establishes interpretability techniques as essential tools for bias auditing, moving beyond simple fairness metrics to provide mechanistic understanding that enables targeted interventions.

5.3 Contribution

This research addresses the limitations of existing approaches that treat bias detection and explanation as separate problems, arguing instead that fairness evaluation must be tightly integrated with interpretability analysis. The study demonstrates that relying solely on fairness metrics can obscure underlying mechanisms of bias, whereas the combined use of SHAP, LIME and Grad-CAM provides a more comprehensive view of how demographic disparities emerge within facial recognition systems. Through systematic experiments, the findings offer evidence that balancing datasets substantially reduces demographic bias, but does not fully eliminate it.

Importantly, the results also reveal that improvements in fairness are highly context-specific, with gains observed in one dataset failing to generalize across different populations. By integrating interpretability methods with fairness assessment, the study highlights how explanations derived from model behavior can uncover sources of bias that remain hidden when analysis is limited to statistical measures alone. Building on these insights, the research proposes a reproducible framework for fairness auditing that unifies dataset profiling, model-centric evaluation, and interpretability analysis into a single pipeline, providing both methodological standards and practical tools for bias diagnosis in machine learning systems.

5.4 Limitations and Future Work

While this research contributes valuable insights into the evaluation of bias in pretrained facial recognition systems, it is important to acknowledge its limitations. First, computational resources imposed a significant constraint on the extent of experimentation.The original objective was to build on top of a pretrained FaceNet model in order to assess how dataset balance influences downstream classification. In practice, this approach allowed the study to surface meaningful dataset-induced biases, but it also revealed important limitations. Relying only on frozen pretrained representations may not be sufficient to capture the subtler ways in which bias manifests within deeper layers. To gain a more nuanced understanding of how models adapt to balanced versus imbalanced data, some degree of fine-tuning may be necessary, as it enables the representations themselves to adjust rather than constraining learning to the final classification layers. Future research with access to larger-scale computational infrastructure could address this gap.

A second limitation concerns the scope and quality of datasets employed. The study was restricted to FairFace and UTKFace, which, while widely used and representative of two distinct dataset philosophies (balanced versus naturally imbalanced), do not fully capture the demographic diversity of the global population. FairFace provided deliberate balance across race, gender, and age groups, but residual biases still emerged, suggesting that balance alone is not a guarantee of fairness. UTKFace, by contrast, exposed clear overfitting tendencies and subgroup disparities due to its relatively small size and noisier annotations, particularly for underrepresented categories such as infants, elderly individuals, and minority racial groups. Including additional datasets such as BUPT-BalancedFace would likely improve the generalisability of the findings.

Third, although fairness was evaluated through subgroup accuracies, confusion matrices, and loss comparisons, the study did not implement fairness-specific metrics such as demographic parity, equalized odds, or disparate impact. These metrics would have offered a more comprehensive perspective on equity across demographic subgroups. Their omission reflects both practical time constraints and the ongoing absence of universal standards in fairness evaluation across the research community.

A further limitation lies in the interpretability analysis. While the methodology anticipated the use of SHAP, LIME, and Grad-CAM, only LIME was applied systematically. SHAP and Grad-CAM were explored but produced unstable or computationally prohibitive results, limiting their inclusion in the final analysis. Consequently, the interpretability insights reported here are less broad than originally envisaged. This limitation underscores the need for continued development of scalable, image-specific interpretability tools that can reliably diagnose sources of bias in deep learning models.

These limitations highlight the challenges of fairness research in facial recognition. They also point to clear directions for future work: the need for broader and more culturally diverse datasets, the integration of fairness-specific evaluation metrics, and the application of more powerful computational resources to enable full fine-tuning and advanced interpretability analyses. Addressing these gaps will be crucial for developing robust, generalisable, and ethically responsible frameworks for fair and interpretable artificial intelligence.

5.5 Reflections

This research demonstrates both the severity of current bias problems in facial recognition systems and the feasibility of developing more fair alternatives. The dramatic bias reductions achieved through balanced datasets provide hope that technical interventions can meaningfully advance algorithmic justice, while the complexity of bias mechanisms revealed through interpretability analysis underscores the continued need for comprehensive, multi-faceted approaches to fairness.

The path toward fair facial recognition systems requires coordinated efforts across multiple stakeholders: researchers must develop better evaluation methods and mitigation techniques, practitioners must prioritize fairness in system design and deployment, regulators must establish appropriate oversight frameworks, and affected communities must be centered in discussions about fairness definitions and priorities.

Ultimately, this research contributes to a growing body of work demonstrating that fair AI is not only possible but essential for the ethical development and deployment of intelligent systems. As facial recognition technology becomes increasingly pervasive in society, ensuring its fairness across all demographic groups is not merely a technical challenge but a moral imperative that requires sustained commitment from the entire AI community.

The tools, insights, and recommendations provided in this thesis offer a foundation for continued progress toward this goal, but achieving truly fair facial recognition systems will require ongoing innovation and collaboration across disciplinary and community boundaries. The stakes are high, but the potential for positive impact through fair AI systems makes this one of the most important challenges facing the computer vision and machine learning communities today.

References

Albiero, V., S, K. K., Vangara, K., Zhang, K., King, M. C. and Bowyer, K. W. (2020), 'Analysis of gender inequality in face recognition accuracy', *CoRR* **abs/2002.00065**.
URL: *https://arxiv.org/abs/2002.00065*

Alvi, M., Zisserman, A. and Nellaker, C. (2018), Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings, *in* 'Workshop on Bias Estimation in Face Analytics, ECCV'.

Buolamwini, J. and Gebru, T. (2018), Gender shades: Intersectional accuracy disparities in commercial gender classification, *in* S. A. Friedler and C. Wilson, eds, 'Proceedings of the 1st Conference on Fairness, Accountability and Transparency', Vol. 81 of *Proceedings of Machine Learning Research*, PMLR, pp. 77–91.
URL: *https://proceedings.mlr.press/v81/buolamwini18a.html*

Chuanjie, Z. and Changming, Z. (2020), Facial expression recognition integrating multiple cnn models, *in* '2020 IEEE 6th International Conference on Computer and Communications (ICCC)', pp. 1410–1414.

Colares, W. G., Costa, M. G. F. and Costa Filho, C. F. F. (2024), Enhancing emotion recognition: A dual-input model for facial expression recognition using images and facial landmarks, *in* '2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)', pp. 1–5.

Devireddy, K. (2025), 'A comparative study of explainable ai methods: Model-agnostic vs. model-specific approaches'.
URL: *https://arxiv.org/abs/2504.04276*

Grother, P., Ngan, M. and Hanaoka, K. (2019), Face Recognition Vendor Test Part 3: Demographic Effects, Technical Report 8280, National Institute of Standards and Technology.

Haliburton, L., Leusmann, J., Welsch, R., Ghebremedhin, S., Isaakidis, P., Schmidt, A. and Mayer, S. (2024), 'Uncovering labeler bias in machine learning annotation tasks', *AI and Ethics* **5**, 2515–2528.

He, K., Zhang, X., Ren, S. and Sun, J. (2015), 'Deep residual learning for image recognition', *CoRR* **abs/1512.03385**.
URL: *http://arxiv.org/abs/1512.03385*

He, Z., Yuan, X. and Qingge, L. (2025), A study of bias in gender and racial classification from face images using facenet, *in* '2025 IEEE Symposium for Multidisciplinary Computational Intelligence Incubators (MCII Companion)', IEEE, pp. 1–5.

Kamatala, S., Naayini, P. and Myakala, P. (2025*a*), 'Mitigating bias in ai: A framework for ethical and fair machine learning models', *INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS* **12**, 848–853.

Kamatala, S., Naayini, P. and Myakala, P. K. (2025*b*), 'Mitigating bias in ai: A framework for ethical and fair machine learning models', *Available at SSRN 5138366* .

Kärkkäinen, K. and Joo, J. (2021), Fairface: Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation, *in* '2021 IEEE Winter Conference on Applications of Computer Vision (WACV)', pp. 1547–1557.

Liu, B., Ran, L., Chen, Z., Guo, W. and Li, Q. (2024), A lightweight facial recognition model for power sites, *in* '2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS)', pp. 339–343.

Liu, Y. (2023), Facial expression recognition model based on improved vggnet, *in* '2023 4th International Conference on Electronic Communication and Artificial Intelligence (ICE-CAI)', pp. 404–408.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A. (2021), 'A survey on bias and fairness in machine learning', *ACM Comput. Surv.* **54**(6).
URL: *https://doi.org/10.1145/3457607*

Merler, M., Ratha, N., Feris, R. S. and Smith, J. R. (2019), 'Diversity in faces'.
URL: *https://arxiv.org/abs/1901.10436*

Messalas, A., Aridas, C. and Kanellopoulos, Y. (2020), Evaluating mashap as a faster alternative to lime for model-agnostic machine learning interpretability, *in* '2020 IEEE International Conference on Big Data (Big Data)', pp. 5777–5779.

Ribeiro, M. T., Singh, S. and Guestrin, C. (2016), '"why should I trust you?": Explaining the predictions of any classifier', *CoRR* **abs/1602.04938**.
URL: *http://arxiv.org/abs/1602.04938*

Schroff, F., Kalenichenko, D. and Philbin, J. (2015), Facenet: A unified embedding for face recognition and clustering, *in* '2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', IEEE, p. 815–823.
URL: *http://dx.doi.org/10.1109/CVPR.2015.7298682*

Slack, D., Hilgard, S., Jia, E., Singh, S. and Lakkaraju, H. (2019), 'How can we fool LIME and shap? adversarial attacks on post hoc explanation methods', *CoRR* **abs/1911.02508**.
URL: *http://arxiv.org/abs/1911.02508*

Srinivasan, R. and Chander, A. (2021), 'Biases in ai systems', *Communications of the ACM* **64**, 44–49.

Wang, M., Deng, W., Hu, J., Tao, X. and Huang, Y. (2019), Racial faces in the wild: Reducing racial bias by information maximization adaptation network, *in* 'The IEEE International Conference on Computer Vision (ICCV)'.

Yeung, M., Teramoto, T., Wu, S., Fujiwara, T., Suzuki, K. and Kojima, T. (2025), 'Variface: Fair and diverse synthetic dataset generation for face recognition'.
URL: *https://arxiv.org/abs/2412.06235*

Zhang, Z., Song, Y. and Qi, H. (2017), 'Age progression/regression by conditional adversarial autoencoder'.
URL: *https://arxiv.org/abs/1702.08423*

A FairFace Source Code

```python
if os.path.exists(DATASET_PATH):
print("Path exists.")
else:
print("Path does not exist.")
expected_files = [
"train_labels.csv",
"val_labels.csv",
"train/",
"val/"
]
print(f"Loading dataset from: {DATASET_PATH}")
missing_files = []
for item in expected_files:
if not os.path.exists(os.path.join(DATASET_PATH, item)):
missing_files.append(item)
if missing_files:
raise FileNotFoundError(
f"Dataset incomplete. Missing: {missing_files}\n"
f"Expected structure:\n"
f"{DATASET_PATH}/\n"
f"          train_labels.csv\n"
f"          val_labels.csv\n"
f"          train/ [contains images]\n"
f"          val/   [contains images]"
)
TRAIN_CSV_PATH = os.path.join(DATASET_PATH, "train_labels.csv")
VAL_CSV_PATH = os.path.join(DATASET_PATH, "val_labels.csv")
TRAIN_IMG_DIR = os.path.join(DATASET_PATH, "train")
VAL_IMG_DIR = os.path.join(DATASET_PATH, "val")
print("\nDataset structure verified:")
print(f"Training CSV:   {TRAIN_CSV_PATH}")
print(f"Validation CSV: {VAL_CSV_PATH}")
print(f"Training images: {TRAIN_IMG_DIR} ({len(os.listdir(TRAIN_IMG_DIR))} files)")
print(f"Validation images: {VAL_IMG_DIR} ({len(os.listdir(VAL_IMG_DIR))} files)")

class FaceNetMultiTask(tf.keras.Model):
def __init__(self, num_age_classes, num_gender_classes, num_race_classes,
freeze_backbone=False):
super(FaceNetMultiTask, self).__init__()
self.backbone = FaceNet().model
if freeze_backbone:
self.backbone.trainable = False
self.shared_dense = layers.Dense(512, activation='relu', name='shared_dense')
self.shared_dropout = layers.Dropout(0.7, name='shared_dropout')
self.age_classifier = tf.keras.Sequential([
layers.Dense(128, activation='relu', name='age_dense'),
layers.Dense(num_age_classes, activation='softmax', name='age_output')
], name='age_head')
self.gender_classifier = tf.keras.Sequential([
layers.Dense(1, activation='sigmoid', name='gender_output')
], name='gender_head')
self.race_classifier = tf.keras.Sequential([
layers.Dense(num_race_classes, activation='softmax', name='race_output')
], name='race_head')
def call(self, inputs, training=None):
embeddings = self.backbone(inputs, training=training)
shared_features = self.shared_dense(embeddings, training=training)
shared_features = self.shared_dropout(shared_features, training=training)
age_pred = self.age_classifier(shared_features, training=training)
gender_pred = self.gender_classifier(shared_features, training=training)
race_pred = self.race_classifier(shared_features, training=training)
return {
'age_output': age_pred,
'gender_output': gender_pred,
'race_output': race_pred
}
def create_and_compile_model(num_age_classes, num_gender_classes, num_race_classes, freeze_backbone=False):
"""Create and compile the multi-task model """
model = FaceNetMultiTask(
num_age_classes=num_age_classes,
num_gender_classes=num_gender_classes,
num_race_classes=num_race_classes,
freeze_backbone=freeze_backbone
)
```

```
model.build((None, 160, 160, 3))
initial_lr = 0.001
model.compile(
optimizer=tf.keras.optimizers.Adam(learning_rate=initial_lr),
loss={
'age_output': losses.SparseCategoricalCrossentropy(),
'gender_output': tf.keras.losses.BinaryCrossentropy(),
'race_output': losses.SparseCategoricalCrossentropy()
},
loss_weights={
'age_output': 1.0,
'gender_output': 1.0,
'race_output': 1.0
},
metrics={
'age_output': ['sparse_categorical_accuracy'],
'gender_output': ['binary_accuracy'],
'race_output': ['sparse_categorical_accuracy']
}
)
print(f"\n Model compiled successfully!")
print(f"   - Backbone frozen: {freeze_backbone}")
print(f"   - Learning rate: {initial_lr}")
return model
```

Listing 1: FairFace implementation, lines 1–99

B UTKFace Source Code

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers, models, optimizers, losses, metrics, Model
from keras_facenet import FaceNet
import pandas as pd
import numpy as np
import os
from pathlib import Path
import zipfile
import gdown
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
```

```python
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

if os.path.exists(DATASET_PATH):
    print("Path exists.")
else:
    print("Path does not exist.")
expected_files = [
    "train_labels.csv",
    "val_labels.csv",
    "train/",
    "val/"
]
print(f"Loading dataset from: {DATASET_PATH}")

class FaceNetMultiTask(tf.keras.Model):
    def __init__(self, num_age_classes, num_gender_classes, num_race_classes,
                 freeze_backbone=False):
        super(FaceNetMultiTask, self).__init__()
        self.backbone = FaceNet().model
        if freeze_backbone:
            self.backbone.trainable = False
        self.shared_dense = layers.Dense(512, activation='relu', name='shared_dense')
        self.shared_dropout = layers.Dropout(0.7, name='shared_dropout')
        self.age_classifier = tf.keras.Sequential([
            layers.Dense(128, activation='relu', name='age_dense'),
            layers.Dense(num_age_classes, activation='softmax', name='age_output')
        ], name='age_head')
        self.gender_classifier = tf.keras.Sequential([
            layers.Dense(num_gender_classes, activation='softmax', name='gender_output')
        ], name='gender_head')
        self.race_classifier = tf.keras.Sequential([
            layers.Dense(num_race_classes, activation='softmax', name='race_output')
        ], name='race_head')
    def call(self, inputs, training=None):
        embeddings = self.backbone(inputs, training=training)
        shared_features = self.shared_dense(embeddings, training=training)
        shared_features = self.shared_dropout(shared_features, training=training)
        age_pred = self.age_classifier(shared_features, training=training)
        gender_pred = self.gender_classifier(shared_features, training=training)
        race_pred = self.race_classifier(shared_features, training=training)
        return {
            'age_output': age_pred,
            'gender_output': gender_pred,
            'race_output': race_pred
        }
def create_and_compile_model(num_age_classes, num_gender_classes, num_race_classes, freeze_backbone=False):
    """Create and compile the multi-task model"""
    model = FaceNetMultiTask(
        num_age_classes=num_age_classes,
        num_gender_classes=num_gender_classes,
        num_race_classes=num_race_classes,
        freeze_backbone=freeze_backbone
    )
    model.build((None, 160, 160, 3))
    initial_lr = 0.001
    model.compile(
        optimizer=tf.keras.optimizers.Adam(learning_rate=initial_lr),
        loss={
            'age_output': losses.SparseCategoricalCrossentropy(),
            'gender_output': losses.SparseCategoricalCrossentropy(),
            'race_output': losses.SparseCategoricalCrossentropy()
        },
        loss_weights={
            'age_output': 1.0,
            'gender_output': 1.0,
            'race_output': 1.0
        },
        metrics={
            'age_output': ['sparse_categorical_accuracy'],
            'gender_output': ['sparse_categorical_accuracy'],
            'race_output': ['sparse_categorical_accuracy']
        }
    )
    print(f"\nModel compiled successfully!")
    print(f"   - Backbone frozen: {freeze_backbone}")
    print(f"   - Learning rate: {initial_lr}")
    return model
```

Listing 2: UTKFace implementation, lines 1–92

C Training Logs

Table 1: Fairface Training Log

epoch	age$_{o}utput_{l}oss$	age$_{o}utput_{s}parse_{c}ategorical_{a}ccuracy$	gender$_{o}utput_{b}inary_{a}ccuracy$	gender$_{o}utput_{l}oss$	learning$_{r}ate$	loss	race$_{o}utput_{l}oss$	race$_{o}utput_{s}parse_{c}ategorical_{a}ccuracy$	val$_{a}ge_{o}utput_{l}oss$	val$_{a}ge_{o}utput_{s}parse_{c}ategorical_{a}ccuracy$	val$_{g}ender_{o}utput_{b}inary_{a}ccuracy$	val$_{g}ender_{o}utput_{l}oss$	val$_{l}oss$	val$_{r}ace_{o}utput_{l}oss$	val$_{r}ace_{o}utput_{s}parse_{c}ategorical_{a}ccuracy$
0	1.2897334098815918	0.4657441973686218	0.8558720946311951	0.3333447277545929	0.001	2.7779247760772705	1.1537314653396606	0.5668139457702637	1.167102336883545	0.513511062	0.8725579977035522	0.28564509749412537	2.4832065105438232	1.0224376916885376	0.6091838479042053
1	1.225222110748291	0.49154651165008545	0.8679883480072021	0.29875943064689636	0.001	2.6328964233398438	1.107894778251648	0.5812790989875793	1.1446130275726318	0.5259265899658203	0.8799525499343872	0.2741459608078003	2.4398210048675537	1.013224720954895	0.6132006645202637
2	1.2126197814941406	0.4932093024253845	0.8701744079589844	0.2940874397754669	0.001	2.6095428466796875	1.101835012435913	0.5832442045211792	1.145728588104248	0.5265656113624573	0.8795873522758484	0.2692716717720032	2.4305968284606934	1.007926344871521	0.6120138764381409
3	1.2067450284957886	0.4973255693912506	0.8713604807853699	0.2925928235054016	0.001	2.596278667449951	1.096028208732605	0.5843372344970703	1.1330928802490234	0.5277523994445801	0.8795873522758484	0.2678548991680145	2.413529396057129	1.0047814846038818	0.6168522834777832
4	1.2023062705993652	0.4984069764614105	0.8726046681404114	0.28878673911094666	0.001	2.58691144	1.0948097705841064	0.5865116119384766	1.1366534233093262	0.5235530138015747	0.8809567093849182	0.2662157416343689	2.4152915477752686	1.0048788785934448	0.6157568097114563
5	1.200564980506897	0.5007790923118591	0.872383714	0.2888028025627136	0.001	2.5861005783081055	1.095852255821228	0.5863953232765198	1.136475920677185	0.5260179042816162	0.8820522427558899	0.26642224192619324	2.4143104553222656	1.0038377046585083	0.6163958311080933
6	1.1959344148635864	0.5013488531112671	0.8737325668334961	0.2866191565990448	0.001	2.575730085372925	1.0921285152435303	0.5870581269264221	1.138403058052063	0.527204692	0.8814131617546082	0.2649272680282593	2.4095025062561035	0.9983705282211304	0.618221641
7	1.195930004119873	0.49919766187667847	0.8738953471183777	0.28493213653564453	0.001	2.571227550506592	1.0894310474395752	0.588209331	1.1340941190719604	0.5292130708694458	0.8835128545761108	0.26405540108680725	2.4064972400665283	1.0006606578826904	0.6167610287666321
8	1.1897355318069458	0.5025581121444702	0.8738721013069153	0.28478220105171204	0.001	2.5641987323760986	1.0886777639389038	0.5881162881851196	1.131489634513855	0.5237356424331665	0.8828738331794739	0.26529791951179504	2.402033805847168	0.9977157115936279	0.618495524
9	1.188187599182129	0.5057441592216492	0.8733255863189697	0.28671929240226746	0.001	2.5641560554504395	1.0883272886276245	0.5869651436805725	1.134525179862976	0.5261091589927673	0.8832389712333679	0.26371777057647705	2.403578996658325	0.9976597428321838	0.61740005
10	1.1883904933929443	0.5042325854301453	0.8754186034202576	0.2836840748786926	0.001	2.5618834495544434	1.0888891220092773	0.5885581374168396	1.129785418510437	0.5283001661300659	0.8829651474952698	0.2635944187641144	2.399000883102417	0.9980509281158447	0.6189519762992859
11	1.1875596046447754	0.505732536	0.8740581274032593	0.2843858003616333	0.001	2.5569040775299072	1.0839869976043701	0.5872442126274109	1.1332379579544067	0.5277523994445801	0.8825086951255798	0.26269519329071045	2.4042134284973145	1.0006204843521118	0.6159393787384033
12	1.183933138847351	0.5055232644081116	0.8760814070701599	0.28241533041000366	0.001	2.5523581504821777	1.0850833654403687	0.5908371806144714	1.1306298971176147	0.5293043851852417	0.881778359	0.26292312145233154	2.3972184658050537	0.9962541460990906	0.617856503
13	1.1843260526657104	0.5035930275917053	0.8756744265556335	0.2818249464035034	0.001	2.5537984371185303	1.0867671966552734	0.5887325406074524	1.1293151378631592	0.5232791900634766	0.8845170736312866	0.26296621561050415	2.3985137939453125	0.9987220764160156	0.618221641
14	1.1816365718841553	0.5063837170600891	0.8753721117973328	0.28118813037872314	0.001	2.5485198497772217	1.084775686264038	0.5874767303466797	1.1252387762069702	0.5271133780479431	0.8828738331794739	0.262847364	2.393333911895752	0.9975952506065369	0.618039072
15	1.1813033819198608	0.5052441954612732	0.8761511445045471	0.2812676727771759	0.001	2.5474278926849365	1.0840476751327515	0.5893488526344299	1.1290839910507202	0.5267482399940491	0.8832389712333679	0.26162341237068176	2.391813039779663	0.9934831857681274	0.6189519762992859
16	1.1794272661209106	0.5070697665214539	0.8746628165245056	0.28236278891563416	0.001	2.5471463203430176	1.0844528675079346	0.5891278982162476	1.1260383129119873	0.5314040780067444	0.8824173808097839	0.26353389024734497	2.3906009197235107	0.9932906627655029	0.6183129549026489
17	1.1794235706329346	0.507906973	0.8752209544181824	0.2827748656272888	0.001	2.548365592956543	1.085166096687317	0.5892093181610107	1.1273308992385864	0.5269308090209961	0.8833302855491638	0.26271724700927734	2.3923027515411377	0.9946978092193604	0.618860662
18	1.1798968315124512	0.5065000057220459	0.8763837218284607	0.28175392746925354	0.001	2.5482680797576904	1.0857316255569458	0.5895116329193115	1.132564663887024	0.5259265899658203	0.8853386640548706	0.2615387439727783	2.395092248916626	0.9935449361801147	0.6198648810386658
19	1.1780893802642822	0.5066744089126587	0.8767558336257935	0.28006109595298767	0.001	2.543213129043579	1.0840486288070679	0.5903023481369019	1.1253223419189453	0.5331385731697083	0.8837867379188538	0.2622824013233185	2.389691114425659	0.9946097731590271	0.6167610287666321
20	1.1768099069595337	0.5050930380821228	0.8763023018836975	0.28054386377334595	0.001	2.539485216140747	1.0811291933059692	0.5896976590156555	1.1278961896896362	0.5256527066230774	0.8843345046043396	0.2613166272640228	2.3897666931152344	0.9932669997215271	0.6216906905174255
21	1.1783589124679565	0.5063139796257019	0.8746744394302368	0.2822168171405792	0.001	2.5426223278045654	1.0811032056808472	0.5905232429504395	1.129272222518921	0.527387261	0.882326066	0.2601879835128784	2.3880224227905273	0.991006613	0.6209603548049927
22	1.1774734258651733	0.5075348615646362	0.8763488531112671	0.28048714995384216	0.001	2.542309522628784	1.083373785018921	0.5902674198150635	1.1267471313476562	0.5326821208000183	0.8812305927276611	0.26337188482284546	2.391234874725342	0.9936549067497253	0.6215081214904785
23	1.1770914793014526	0.5084883570671082	0.8761860728263855	0.2808637320995331	0.001	2.5414130687713623	1.0824609994888306	0.5896627902984619	1.1279411315917969	0.5292130708694458	0.8833302855491638	0.2630385160446167	2.3918609619140625	0.9932060837745667	0.6223297715187073
24	1.1768755912780762	0.5091279149055481	0.8769534826278687	0.28065431118011475	0.001	2.540321111679077	1.0819027423858643	0.5895581245422363	1.1228524446487427	0.5326821208000183	0.8840606212615967	0.2614687383174896	2.3840389251708984	0.9921976327896118	0.6226035952568054
25	1.1744056940078735	0.5084999799728394	0.8758720755577087	0.28113654255867004	0.001	2.5402164459228516	1.0837228298187256	0.5895000100135803	1.1230251789093018	0.5314040780067444	0.8828738331794739	0.2631818950176239	2.387556314468384	0.9937855005264282	0.6204126477241516
26	1.1759281158447266	0.5080116391181946	0.8774535059928894	0.27910974621772766	0.001	2.5384907722473145	1.08242929	0.5897907018661499	1.1274871826171875	0.5313127636909485	0.8834215998649597	0.26487722992897034	2.3922109603881836	0.9922240376472473	0.6215994358062744
27	1.1741044521331787	0.5090232491493225	0.8758023381233215	0.28011462092399597	0.001	2.538632392883301	1.083382248878479	0.5900116562843323	1.1256346702575684	0.5313127636909485	0.8854299783706665	0.2617639899253845	2.389010190963745	0.9940107464790344	0.6192258596420288
28	1.1739909648895264	0.5104069709777832	0.8756744265556335	0.27988919615745544	0.001	2.5368549823760986	1.0819363594055176	0.5898371934890747	1.1242454051971436	0.5270220637321472	0.88259995	0.2623029351234436	2.3878846168518066	0.9938288927078247	0.6183129549026489
29	1.174655556678772	0.5080465078353882	0.8765465021133423	0.2804047465324402	0.001	2.5360286235809326	1.0799823999404907	0.5915232300758362	1.1249706745147705	0.5300346612930298	0.8848822116851807	0.26116272807121277	2.390082359313965	0.9965620040893555	0.6194084286689758
30	1.1647976636886597	0.5126860737800598	0.877488375	0.27750450372695923	0.0005	2.5122101306915283	1.0688878297805786	0.5951046347618103	1.1181755065917969	0.5315866470336914	0.8835128545761108	0.2609875202178955	2.378133773803711	0.9914880990982056	0.6213255524635315
31	1.1608504056930542	0.5141860246658325	0.8781395554542542	0.2767392098903656	0.0005	2.5037569999694824	1.0651731491088867	0.5960000157356262	1.1182183027267456	0.5335037708282471	0.8836954832077026	0.2608073949813843	2.374861240386963	0.9882318377494812	0.6231513619422913
32	1.1578681468963623	0.5148720741271973	0.8771278858184814	0.27636969089508057	0.0005	2.500349521636963	1.0651333332061768	0.5979418754577637	1.1243665218353271	0.5303085446357727	0.8842431902885437	0.2618512213230133	2.382429838180542	0.9887435436248779	0.6227862238883972
33	1.1574169397354126	0.5145813822746277	0.876976728	0.27602019906044006	0.0005	2.499014377593994	1.0646452903747559	0.5986279249191284	1.1198337078094482	0.5308563113212585	0.8838780522346497	0.2596891224384308	2.375485897064209	0.9884622097015381	0.6198648810386658
34	1.1561317443847656	0.5165930390357971	0.8788022994995117	0.27623099088668823	0.0005	2.493950605392456	1.0606132745742798	0.5991976857185364	1.1194339990615845	0.5305824279785156	0.8833302855491638	0.2601400315761566	2.376662254333496	0.989556074	0.6191345453262329
35	1.159149408340454	0.5130348801612854	0.8789418339729309	0.27518078684806824	0.0005	2.4979960918426514	1.0627601146697998	0.5975232720375061	1.1192306280136108	0.5298520922660828	0.8845170736312866	0.259977788	2.3744707107543945	0.9878019094467163	0.6208691000938416
36	1.1555719375610352	0.5167558193206787	0.8788372278213501	0.27429822087287903	0.0005	2.4956605434417725	1.064974069595337	0.5981046557426453	1.120345950126648	0.5322256684303284	0.882782519	0.2602786421775818	2.3765649795532227	0.988368809	0.6208691000938416
37	1.1540567874908447	0.5168721079826355	0.8792790770530701	0.273764431	0.0005	2.491205930709839	1.0627176761627197	0.5998837351799011	1.1185859441757202	0.5346904993057251	0.8835128545761108	0.26202431321144104	2.376995325088501	0.9889753460884094	0.6223297715187073
38	1.1547571420669556	0.517918587	0.8790697455406189	0.27596914768218994	0.0005	2.4957549571990967	1.0642532110214233	0.5955930352210999	1.1159294843673706	0.5330473184585571	0.8854299783706665	0.259924054	2.3705220222473145	0.9872325658798218	0.6238816976547241
39	1.1545095443725586	0.518558145	0.879790723	0.27333980798721313	0.0005	2.4920856952667236	1.0633716583251953	0.599046528	1.1151988506317139	0.5336863398551941	0.8841519355773926	0.2591712772846222	2.3684091567993164	0.986534834	0.6235164999961853
40	1.1543771028518677	0.515686035	0.8795232772827148	0.2731168568134308	0.0005	2.487708568572998	1.0591782331466675	0.5998953580856323	1.1167597770690918	0.5371553897857666	0.8841519355773926	0.2583792805671692	2.369513988494873	0.9869215488433838	0.6217820048332214
41	1.152290940284729	0.517604649	0.8799883723258972	0.27377834916114807	0.0005	2.487985372543335	1.0610698461532593	0.5968255996704102	1.1155319213867188	0.534416676	0.8845170736312866	0.25983425974845886	2.3700361251831055	0.9872928261756897	0.6224210262298584
42	1.1545356512069702	0.5179069638252258	0.8802441954612732	0.27387499809265137	0.0005	2.4897475242614746	1.0603528022766113	0.5988720655441284	1.1174849271774292	0.534599245	0.8837867379188538	0.25978556275367737	2.3718690872192383	0.9871259331703186	0.6220558881759644
43	1.1499851942062378	0.5180813670158386	0.8797674179077148	0.27300482988357544	0.0005	2.484539747238159	1.0605367422103882	0.5999302268028259	1.1140321493148804	0.5368815064430237	0.8853386640548706	0.2599751353263855	2.3675498962402344	0.9860589504241943	0.6233339309692383
44	1.1533863544464111	0.5167325735092163	0.8797441720962524	0.27492544054985046	0.0005	2.4895505905151367	1.0603300333023071	0.5998139381408691	1.1182407140731812	0.534599245	0.8859777450561523	0.2602005898952484	2.3726346492767334	0.9866812229156494	0.6242468357086182
45	1.1521084308624268	0.5182906985282898	0.8787790536880493	0.2734508812427521	0.0005	2.488023519515991	1.0616780519485474	0.5983256101608276	1.1172271966934204	0.5366989374160767	0.8845170736312866	0.259522945	2.3696377277374268	0.9854355454444885	0.6237903833389282
46	1.1508326530456543	0.5198604464530945	0.880616307	0.2738422155380249	0.0005	2.485933303833008	1.060418725013733	0.5977557897567749	1.116279125213623	0.5335950255393982	0.8843345046043396	0.25983762741088867	2.3706536293029785	0.9869569540023804	0.6263465285301208
47	1.1536304950714111	0.5170232653617859	0.8797325491905212	0.2734162211418152	0.0005	2.4904897212982178	1.062551498413086	0.5988023281097412	1.1142839193344116	0.5377944111824036	0.8844257593154907	0.2602122128009796	2.3675191402435303	0.9854936599731445	0.6239729523658752
48	1.151414394378662	0.5191743969917297	0.8787325620651245	0.2755967974662781	0.0005	2.4870965480804443	1.059211254119873	0.6008023023605347	1.1147247552871704	0.5334124565124512	0.8853386640548706	0.25977200269699097	2.3671460151672363	0.9851544499397278	0.6216906905174255
49	1.1525671482086182	0.5171046257019043	0.8798023462295532	0.27443593740463257	0.0005	2.491182804107666	1.0633726119995117	0.5980348587036133	1.1161837577819824	0.5357860326766968	0.8856125473976135	0.2608121335506439	2.3706552982330322	0.9861860275268555	0.6212342381477356

Table 2: UTKFace Training Log

epoch	age$_output_loss$	age$_output_sparse_categorical_accuracy$	gender$_output_binary_accuracy$	gender$_output_loss$	learning$_rate$	loss	race$_output_loss$	race$_output_sparse_categorical_accuracy$	val$_age_output_loss$	val$_age_output_sparse_categorical_accuracy$	val$_gender_output_binary_accuracy$	val$_gender_output_loss$	val$_loss$	val$_race_output_loss$	val$_race_output_sparse_categorical_accuracy$
0	1.2897334098815918	0.4657441973686218	0.8558720946311951	0.3333447277545929	0.001	2.7779247760772705	1.1537314653396606	0.5668139457702637	1.167102336883545	0.513511062	0.8725579977035522	0.28564509749412537	2.4832065105438232	1.0224376916885376	0.6091838479042053
1	1.225222110748291	0.49154651165008545	0.8679883480072021	0.29875943064689636	0.001	2.6328964233398438	1.107894778251648	0.5812790989875793	1.1446130275726318	0.5259265899658203	0.8799525499343872	0.2741459608078003	2.4398210048675537	1.013224720954895	0.6132006645202637
2	1.2126197814941406	0.4932093024253845	0.8701744079589844	0.2940874397754669	0.001	2.6095428466796875	1.101835012435913	0.5832442045211792	1.145728588104248	0.5265656113624573	0.8795873522758484	0.2692716717720032	2.4305968284606934	1.007926344871521	0.6120138764381409
3	1.2067450284957886	0.4973255693912506	0.8713604807853699	0.2925928235054016	0.001	2.596278667449951	1.096028208732605	0.5843372344970703	1.1330928802490234	0.5277523994445801	0.8795873522758484	0.2678548991680145	2.413529396057129	1.0047814846038818	0.6168522834777832
4	1.2023062705993652	0.4984069764614105	0.8726046681404114	0.28878673911094666	0.001	2.58691144	1.0948097705841064	0.5865116119384766	1.1366534233093262	0.5235530138015747	0.8809567093849182	0.2662157416343689	2.4152915477752686	1.0048788785934448	0.6157568097114563
5	1.200564980506897	0.5007790923118591	0.872383714	0.2888028025627136	0.001	2.5861005783081055	1.095852255821228	0.5863953232765198	1.136475920677185	0.5260179042816162	0.8820522427558899	0.26642224192619324	2.4143104553222656	1.0038377046585083	0.6163958311080933
6	1.1959344148635864	0.5013488531112671	0.8737325668334961	0.2866191565990448	0.001	2.575730085372925	1.0921285152435303	0.5870581269264221	1.138403058052063	0.527204692	0.8814131617546082	0.2649272680282593	2.4095025062561035	0.9983705282211304	0.618221641
7	1.195930004119873	0.49919766187667847	0.8738953471183777	0.28493213653564453	0.001	2.571227550506592	1.0894310474395752	0.588209331	1.1340941190719604	0.5292130708694458	0.8835128545761108	0.26405540108680725	2.4064972400665283	1.0006606578826904	0.6167610287666321
8	1.1897355318069458	0.5025581121444702	0.8738721013069153	0.28478220105171204	0.001	2.5641987323760986	1.0886777639389038	0.5881162881851196	1.131489634513855	0.5237356424331665	0.8828738331794739	0.26529791951179504	2.402033805847168	0.9977157115936279	0.618495524
9	1.188187599182129	0.5057441592216492	0.8733255863189697	0.28671929240226746	0.001	2.5641560554504395	1.0883272886276245	0.5869651436805725	1.134525179862976	0.5261091589927673	0.8832389712333679	0.26371777057647705	2.403578996658325	0.9976597428321838	0.61740005
10	1.1883904933929443	0.5042325854301453	0.8754186034202576	0.2836840748786926	0.001	2.5618834495544434	1.0888891220092773	0.5885581374168396	1.129785418510437	0.5283001661300659	0.8829651474952698	0.2635944187641144	2.399000883102417	0.9980509281158447	0.6189519762992859
11	1.1875596046447754	0.505732536	0.8740581274032593	0.2843858003616333	0.001	2.5569040775299072	1.0839869976043701	0.5872442126274109	1.1332379579544067	0.5277523994445801	0.8825086951255798	0.26269519329071045	2.4042134284973145	1.0006204843521118	0.6159393787384033
12	1.183933138847351	0.5055232644081116	0.8760814070701599	0.28241533041000366	0.001	2.5523581504821777	1.0850833654403687	0.5908371806144714	1.1306298971176147	0.5293043851852417	0.881778359	0.26292312145233154	2.3972184658050537	0.9962541460990906	0.617856503
13	1.1843260526657104	0.5035930275917053	0.8756744265556335	0.2818249464035034	0.001	2.5537984371185303	1.0867671966552734	0.5887325406074524	1.1293151378631592	0.5232791900634766	0.8845170736312866	0.26296621561050415	2.3985137939453125	0.9987220764160156	0.618221641
14	1.1816365718841553	0.5063837170600891	0.8753721117973328	0.28118813037872314	0.001	2.5485198497772217	1.084775686264038	0.5874767303466797	1.1252387762069702	0.5271133780479431	0.8828738331794739	0.262847364	2.393333911895752	0.9975952506065369	0.618039072
15	1.1813033819198608	0.5052441954612732	0.8761511445045471	0.2812676727771759	0.001	2.5474278926849365	1.0840476751327515	0.5893488526344299	1.1290839910507202	0.5267482399940491	0.8832389712333679	0.26162341237068176	2.391813039779663	0.9934831857681274	0.6189519762992859
16	1.1794272661209106	0.5070697665214539	0.8746628165245056	0.28236278891563416	0.001	2.5471463203430176	1.0844528675079346	0.5891278982162476	1.1260383129119873	0.5314040780067444	0.8824173808097839	0.26353389024734497	2.3906009197235107	0.9932906627655029	0.6183129549026489
17	1.1794235706329346	0.507906973	0.8752209544181824	0.2827748656272888	0.001	2.548365592956543	1.085166096687317	0.5892093181610107	1.1273308992385864	0.5269308090209961	0.8833302855491638	0.26271724700927734	2.3923027515411377	0.9946978092193604	0.618860662
18	1.1798968315124512	0.5065000057220459	0.8763837218284607	0.28175392746925354	0.001	2.5482680797576904	1.0857316255569458	0.5895116329193115	1.132564663887024	0.5259265899658203	0.8853386640548706	0.2615387439727783	2.395092248916626	0.9935449361801147	0.6198648810386658
19	1.1780893802642822	0.5066744089126587	0.8767558336257935	0.28006109595298767	0.001	2.543213129043579	1.0840486288070679	0.5903023481369019	1.1253223419189453	0.5331385731697083	0.8837867379188538	0.2622824013233185	2.389691114425659	0.9946097731590271	0.6167610287666321
20	1.1768099069595337	0.5050930380821228	0.8763023018836975	0.28054386377334595	0.001	2.539485216140747	1.0811291933059692	0.5896976590156555	1.1278961896896362	0.5256527066230774	0.8843345046043396	0.2613166272640228	2.3897666931152344	0.9932669997215271	0.6216906905174255
21	1.1783589124679565	0.5063139796257019	0.8746744394302368	0.2822168171405792	0.001	2.5426223278045654	1.0811032056808472	0.5905232429504395	1.129272222518921	0.527387261	0.882326066	0.2601879835128784	2.3880224227905273	0.991006613	0.6209603548049927
22	1.1774734258651733	0.5075348615646362	0.8763488531112671	0.28048714995384216	0.001	2.542309522628784	1.083373785018921	0.5902674198150635	1.1267471313476562	0.5326821208000183	0.8812305927276611	0.26337188482284546	2.391234874725342	0.9936549067497253	0.6215081214904785
23	1.1770914793014526	0.5084883570671082	0.8761860728263855	0.2808637320995331	0.001	2.5414130687713623	1.0824609994888306	0.5896627902984619	1.1279411315917969	0.5292130708694458	0.8833302855491638	0.2630385160446167	2.3918609619140625	0.9932060837745667	0.6223297715187073
24	1.1768755912780762	0.5091279149055481	0.8769534826278687	0.28065431118011475	0.001	2.540321111679077	1.0819027423858643	0.5895581245422363	1.1228524446487427	0.5326821208000183	0.8840606212615967	0.2614687383174896	2.3840389251708984	0.9921976327896118	0.6226035952568054
25	1.1744056940078735	0.5084999799728394	0.8758720755577087	0.28113654255867004	0.001	2.5402164459228516	1.0837228298187256	0.5895000100135803	1.1230251789093018	0.5314040780067444	0.8828738331794739	0.2631818950176239	2.387556314468384	0.9937855005264282	0.6204126477241516
26	1.1759281158447266	0.5080116391181946	0.8774535059928894	0.27910974621772766	0.001	2.5384907722473145	1.08242929	0.5897907018661499	1.1274871826171875	0.5313127636909485	0.8834215998649597	0.26487722992897034	2.3922109603881836	0.9922240376472473	0.6215994358062744
27	1.1741044521331787	0.5090232491493225	0.8758023381233215	0.28011462092399597	0.001	2.538632392883301	1.083382248878479	0.5900116562843323	1.1256346702575684	0.5313127636909485	0.8854299783706665	0.2617639899253845	2.389010190963745	0.9940107464790344	0.6192258596420288
28	1.1739909648895264	0.5104069709777832	0.8756744265556335	0.27988919615745544	0.001	2.5368549823760986	1.0819363594055176	0.5898371934890747	1.1242454051971436	0.5270220637321472	0.88259995	0.2623029351234436	2.3878846168518066	0.9938288927078247	0.6183129549026489
29	1.174655556678772	0.5080465078353882	0.8765465021133423	0.2804047465324402	0.001	2.5360286235809326	1.0799823999404907	0.5915232300758362	1.1249706745147705	0.5300346612930298	0.8848822116851807	0.26116272807121277	2.390082359313965	0.9965620040893555	0.6194084286689758
30	1.1647976636886597	0.5126860737800598	0.877488375	0.27750450372695923	0.0005	2.5122101306915283	1.0688878297805786	0.5951046347618103	1.1181755065917969	0.5315866470336914	0.8835128545761108	0.2609875202178955	2.378133773803711	0.9914880990982056	0.6213255524635315
31	1.1608504056930542	0.5141860246658325	0.8781395554542542	0.2767392098903656	0.0005	2.5037569999694824	1.0651731491088867	0.5960000157356262	1.1182183027267456	0.5335037708282471	0.8836954832077026	0.2608073949813843	2.374861240386963	0.9882318377494812	0.6231513619422913
32	1.1578681468963623	0.5148720741271973	0.8771278858184814	0.27636969089508057	0.0005	2.500349521636963	1.0651333332061768	0.5979418754577637	1.1243665218353271	0.5303085446357727	0.8842431902885437	0.2618512213230133	2.382429838180542	0.9887435436248779	0.6227862238883972
33	1.1574169397354126	0.5145813822746277	0.876976728	0.27602019906044006	0.0005	2.499014377593994	1.0646452903747559	0.5986279249191284	1.1198337078094482	0.5308563113212585	0.8838780522346497	0.2596891224384308	2.375485897064209	0.9884622097015381	0.6198648810386658
34	1.1561317443847656	0.5165930390357971	0.8788022994995117	0.27623099088668823	0.0005	2.493950605392456	1.0606132745742798	0.5991976857185364	1.1194339990615845	0.5305824279785156	0.8833302855491638	0.2601400315761566	2.376662254333496	0.989556074	0.6191345453262329
35	1.159149408340454	0.5130348801612854	0.8789418339729309	0.27518078684806824	0.0005	2.4979960918426514	1.0627601146697998	0.5975232720375061	1.1192306280136108	0.5298520922660828	0.8845170736312866	0.259977788	2.3744707107543945	0.9878019094467163	0.6208691000938416
36	1.1555719375610352	0.5167558193206787	0.8788372278213501	0.27429822087287903	0.0005	2.4956605434417725	1.064974069595337	0.5981046557426453	1.120345950126648	0.5322256684303284	0.882782519	0.2602786421775818	2.3765649795532227	0.988368809	0.6208691000938416
37	1.1540567874908447	0.5168721079826355	0.8792790770530701	0.273764431	0.0005	2.491205930709839	1.0627176761627197	0.5998837351799011	1.1185859441757202	0.5346904993057251	0.8835128545761108	0.26202431321144104	2.376995325088501	0.9889753460884094	0.6223297715187073
38	1.1547571420669556	0.517918587	0.8790697455406189	0.27596914768218994	0.0005	2.4957549571990967	1.0642532110214233	0.5955930352210999	1.1159294843673706	0.5330473184585571	0.8854299783706665	0.259924054	2.3705220222473145	0.9872325658798218	0.6238816976547241
39	1.1545095443725586	0.518558145	0.879790723	0.27333980798721313	0.0005	2.4920856952667236	1.0633716583251953	0.599046528	1.1151988506317139	0.5336863398551941	0.8841519355773926	0.2591712772846222	2.3684091567993164	0.986534834	0.6235164999961853
40	1.1543771028518677	0.515686035	0.8795232772827148	0.2731168568134308	0.0005	2.487708568572998	1.0591782331466675	0.5998953580856323	1.1167597770690918	0.5371553897857666	0.8841519355773926	0.2583792805671692	2.369513988494873	0.9869215488433838	0.6217820048332214
41	1.152290940284729	0.517604649	0.8799883723258972	0.27377834916114807	0.0005	2.487985372543335	1.0610698461532593	0.5968255996704102	1.1155319213867188	0.534416676	0.8845170736312866	0.25983425974845886	2.3700361251831055	0.9872928261756897	0.6224210262298584
42	1.1545356512069702	0.5179069638252258	0.8802441954612732	0.27387499809265137	0.0005	2.4897475242614746	1.0603528022766113	0.5988720655441284	1.1174849271774292	0.534599245	0.8837867379188538	0.25978556275367737	2.3718690872192383	0.9871259331703186	0.6220558881759644
43	1.1499851942062378	0.5180813670158386	0.8797674179077148	0.27300482988357544	0.0005	2.484539747238159	1.0605367422103882	0.5999302268028259	1.1140321493148804	0.5368815064430237	0.8853386640548706	0.2599751353263855	2.3675498962402344	0.9860589504241943	0.6233339309692383
44	1.1533863544464111	0.5167325735092163	0.8797441720962524	0.27492544054985046	0.0005	2.4895505905151367	1.0603300333023071	0.5998139381408691	1.1182407140731812	0.534599245	0.8859777450561523	0.2602005898952484	2.3726346492767334	0.9866812229156494	0.6242468357086182
45	1.1521084308624268	0.5182906985282898	0.8787790536880493	0.2734508812427521	0.0005	2.488023519515991	1.0616780519485474	0.5983256101608276	1.1172271966934204	0.5366989374160767	0.8845170736312866	0.259522945	2.3696377277374268	0.9854355454444885	0.6237903833389282
46	1.1508326530456543	0.5198604464530945	0.880616307	0.2738422155380249	0.0005	2.485933303833008	1.060418725013733	0.5977557897567749	1.116279125213623	0.5335950255393982	0.8843345046043396	0.25983762741088867	2.3706536293029785	0.9869569540023804	0.6263465285301208
47	1.1536304950714111	0.5170232653617859	0.8797325491905212	0.2734162211418152	0.0005	2.4904897212982178	1.062551498413086	0.5988023281097412	1.1142839193344116	0.5377944111824036	0.8844257593154907	0.2602122128009796	2.3675191402435303	0.9854936599731445	0.6239729523658752
48	1.151414394378662	0.5191743969917297	0.8787325620651245	0.2755967974662781	0.0005	2.4870965480804443	1.059211254119873	0.6008023023605347	1.1147247552871704	0.5334124565124512	0.8853386640548706	0.25977200269699097	2.3671460151672363	0.9851544499397278	0.6216906905174255
49	1.1525671482086182	0.5171046257019043	0.8798023462295532	0.27443593740463257	0.0005	2.491182804107666	1.0633726119995117	0.5980348587036133	1.1161837577819824	0.5357860326766968	0.8856125473976135	0.2608121335506439	2.3706552982330322	0.9861860275268555	0.6212342381477356