2025 19th International Conference on Automatic Face and Gesture Recognition (FG)

Synthetic Faces, Real Gains: Improving Age and Gender Classification through Generative Data

Nuno R. Freitas[1] and Andreia M. Costa[1] and João Tremoço[1] and Miguel Lourenço[1]
[1] Youverse, Portugal

2025 IEEE 19th International Conference on Automatic Face and Gesture Recognition (FG) | 979-8-3315-5341-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/FG61629.2025.11099227

Abstract— **Facial image-based age and gender classification is a foundational problem in computer vision, yet performance is often limited by data scarcity and privacy constraints. This paper proposes IDiff-Face-Aged, a novel age transformation framework that utilizes multimodal embeddings to generate realistic and diverse facial images. Two synthetic datasets, Fading (utilizing null text-inversion prompting) and IDiff-Face-Aged[1], were tested across various architectures, including MultiEfficientNet, MultiLightViT, and MultiMobileNet. Three age estimation methods were used to address both continuous and categorical intervals. Experimental results indicate that synthetic data contributes to improved model accuracy, especially by enhancing representation in under-sampled age groups. On the UTKFace dataset, an accuracy gain of over 3% was observed. Moreover, models trained with a mix of synthetic and real data show stronger generalization capabilities, particularly when datasets design and alignment are carefully managed. These findings illustrate the potential of synthetic data to supplement real-world datasets, while also pointing to ongoing challenges related to data realism and artifact avoidance.**

I. INTRODUCTION

Age and gender classification from facial images is a long-standing and widely studied problem in computer vision, with applications in security, social analytics, and personalized user experiences [1]. Estimating age and gender—whether predicting chronological, appearance-based, or perceived age—serves a broad range of real-world applications, including age-invariant face recognition, cross-age face verification, biometric authentication, surveillance, and customer profiling in commercial settings [1], [2]. Recent advancements in deep learning, particularly convolutional neural networks (CNNs) and, more recently, transformer-based architectures, have significantly improved the performance of age and gender classification models [3]. However, achieving robust and generalizable models remains a critical challenge due to inherent biases in existing datasets. Publicly available benchmarks often exhibit imbalanced distributions, limited intra and inter-class diversity, and restricted ethnic representation, all of which contribute to models that underperform on demographically diverse populations [4]. To address these challenges, synthetic data generation has emerged as a promising approach, leveraging generative models such as Generative Adversarial Networks (GANs) [5] and diffusion models (DM) [6] to create realistic face images with controlled variations in age and gender.

This study investigates the role of synthetic aging datasets in the performance of age and gender classification models. Specifically, we assess how two synthetic approaches, Fading [7], which provides progressively aged faces, and IDiff-Face-Aged, adapted from [4], affect classification accuracy when integrated into model training. We provide visual evaluation as well as metrics evaluation for multiple dataset configuration, to determine the extent to which synthetic augmentation enhances or degrades model performance on real-world benchmarks.

Our approach offers a novel framework, IDiff-Face-Aged for age transformation by leveraging multimodal embeddings to guide the generation process in a semantically meaningful manner. Unlike traditional methods that rely solely on age estimation models, we integrate descriptive text embeddings from Florence-2 [8] and CLIP [9] to capture richer facial attribute representations, ensuring that the generated faces align with the target age category. This balance between identity consistency and transformation diversity makes our approach particularly valuable for dataset augmentation in age estimation tasks. Instead of enforcing strict identity preservation, which can lead to artifacts or unrealistic transformations, we encourage controlled variations that enhance the model's ability to generalize across different age groups.

A. Manuscript organization

This paper is structured as follows: Section II reviews related work on age and gender classification and synthetic data generation. Section III details the datasets, models, and experimental setup. Section IV presents the results and discussion, and Section V concludes with key findings and directions for future research.

II. RELATED WORK

A review of existing literature reveals that conventional age and gender estimation models are often trained on limited real-world datasets, which constrains their generalization capabilities. Recent studies have begun to explore the use of synthetic data to augment these datasets, with promising results in terms of improved model accuracy and generalization. However, there remains a need for a comprehensive case study that evaluates the impact of synthetic data on age and gender estimation.

A. Age and Gender estimation

Age and gender estimation from facial images has been extensively studied, with early methods relying on handcrafted

This work was supported by the Portuguese Recovery and Resilience Plan (RPP) under the program 'Agendas para a Inovação Empresarial', reference no. 62 - 'Center for Responsible AI'.

[1] Method under Patent Pending Process

features (e.g., LBP, HOG, Gabor filters) and traditional classifiers like SVMs and k-NNs [1]. These approaches have since been surpassed by deep learning models that automatically learn hierarchical representations.

In 2015, Levi G. and Hassner T. [10] presented a seminal work to explore deep learning CNN for multiclass (age groups) and binary classification on the AdienceFaces dataset. Liu et al. [11] used separate GoogLeNet models for age classification and age regression. Subsequent studies explored lightweight architectures (e.g. MobileNet, EfficientNet and RexNet) [12], [13], improved label distribution learning [14], and multi-branch fusion models [15]. Additionally, advanced loss functions based on optimal transport theory [16] and graph-based learning [17] have been proposed to further enhance performance. This was followed by attention-based methods [18], directing the model to focus on specific age-patch features, and transformer models [19] have been used to further refine feature extraction and aggregation.

For gender recognition, embeddings from FaceNet [20], NN4-based networks with variational loss [21], and Pareto frontier transfer learning approaches [22] have achieved strong performance.

Recent studies include a moving window regression algorithm for ordinal regression [23], and MiVOLO [3], a multi-input transformer model for age and gender estimation, integrating facial and body information. Multimodal large language models [24] have also shown promise, but with high computational costs.

Age estimation remains a challenging task due to its personalized, temporal, and causal nature, compounded by limited and biased datasets. Privacy concerns and the impracticality of collecting massive, diverse real-world datasets have driven interest in synthetic data, which offers scalable and controllable alternatives.

B. Synthetic Data - Face Aging

Synthetic data generation has become a cornerstone in advancing face recognition and aging technologies, enabling the creation of diverse and ethically sourced datasets. Among the leading generative techniques are GANs, DMs and their hybrids.

GANs have significantly advanced the generation of highly realistic facial images like DigiFace-1M [25], SynFace [26], SFace2 [27] and StyleNat [28]. Conditional GANs incorporate age and gender cues [29], [30], [31], [32], [33], while others use RNNs [34] or manipulate the latent space to simulate aging [35], [36]. Techniques such as LATS [37] and AgeSynthGAN [38] adopt style-based architectures to handle appearance variations. However, many GAN-based methods struggle with preserving identity, especially during latent space inversion or extreme pose handling.

DMs, such as DCFace [39] and IDiff-Face [4], offer high-quality synthetic faces through iterative denoising and are promising for subtle changes like wrinkles and skin tone. Early works like SDEdit [40], introduced text-guided editing, while subsequent methods tried mask-conditioned edits [40], [41], [42], to improve fine-grained details. DiffEdit [43] improved region-specific control without the need for a user-provided mask.

Beyond masked editing, other techniques focus on refining text-based control. Prompt-to-Prompt [44] enables edits using slight changes on pairs of text prompts, making modifications purely through textual descriptions. Null-text inversion [45] extends this concept to real images by optimizing the null-text embedding. Similarly, Imagic [46] enables real-image editing by fine-tuning the DM to preserve the input image's appearance and details while applying desired text-driven modifications.

The recent Fading model [7] improves text-guided image editing by leveraging null-text inversion and dynamically refining latent representations for better control and identity preservation. However, it still struggles with precise spatial control, occasionally causing unintended changes in unrelated regions. Meanwhile, DiffAge3D [47] introduced a 3D/aware diffusion aging framework that explicitly accounts for multi-view consistency, demonstrating superior age progression realism while preserving identity across different viewpoints.

Hybrid models like GANDiffFace [48] combine the strengths of both GANs and DMs. These approaches aim to strike a balance between high-quality generation and better control over the output, but fine-tuning such models requires careful balancing of both components to avoid the drawbacks of each method.

Despite these advancements, challenges remain in controlling spatial edits and maintaining identity. Despite these challenges, synthetic data generation remains a powerful tool for training robust and fair face analysis systems.

III. Methodology

In this section, we present the methodology used to improve age and gender estimation through the integration of synthetic data. We first describe datasets and the identity-conditioned diffusion-based synthetic data generation pipeline, followed by the architecture and training of the age and gender classification models.

A. Age and Gender training datasets

This project involves using multiple datasets for age and gender classification. For training and validation, we used a combination of FFHQ [49](#70,000), Adience [10](#26,580), Lagenda [3](#67,159), and IMDB-WIKI [50](#500,000) datasets, CelebAMask-HQ [51](#30,000). To further expand data diversity, SFHQ [52](#425,000), a fully synthetic dataset was used. The UTKFace [53](#23,491) dataset was selected for testing due to its balanced age distribution across various age groups [54]. This ensures a more reliable evaluation of the models' generalization across a broader demographic spectrum. For datasets lacking inherent labels, such as SFHQ, we assign labels using the InsightFace *buffalo_l* [55] age and gender estimation model with a human in loop to ensure accurate annotations.

To enhance model robustness, data pre-processing involved various augmentations, including resizing, cropping, flipping, rotation, brightness/contrast adjustments, color jitter, and noise addition. Compression artifacts were introduced to simulate real-world degradations. All images were normalized using standard ImageNet statistics.

We also leveraged generative-based augmentations. DMs outperform GANs in stability, sample quality, and control. They avoid mode collapse, generalize better, and enable precise conditioning. While GANs are faster, diffusion models excel in generating diverse, high-quality synthetic data, making them ideal for bias-free AI and face synthesis [4]. In this regard, two different strategies were employed: one based on the IDiff-Face [4] with attributes controlling and another using the multimodal Fading [7], model with null text inversion and prompt-to-prompt editing for localized modifications.

IDiff-Face was chosen as the baseline method for dataset augmentation over alternative approaches due to its ability to generate diverse, bias-free, realistic facial variations [4]. While our method, built upon IDiff-Face, does explicitly focus on guaranteeing perfect identity preservation, it conditions the generation process on identity-related features, ensuring that the transformed faces share structural and contextual similarities with the reference. Additionally, by introducing new variations, it enhances the model's ability to learn distinct aging patterns. This identity-conditioned DM leverages an identity encoder with previously extracted identity representations, providing a stronger basis for identity-aware synthesis. To guide age transformations, we conditioned the model on learned aging representations rather than relying solely on a pre-trained age estimation model. Specifically, we used the Florence-2 [8] multimodal model to generate descriptive captions for each image, capturing semantically rich details about facial attributes, including age and gender. These captions were encoded using a pre-trained CLIP text encoder, yielding text embeddings that offer a more context-aware representation of age progression. By incorporating these embeddings, our method intends to align generated faces with the target age category in a semantically meaningful manner.

1) Implementation details for IDiff-Face-aged: In training, we applied paired embeddings corresponding to the target age and gender. During the sampling process, age adjustment was performed by retrieving the embeddings corresponding to an identity of the target age (e.g., 5 years old) and gender (e.g., male), and combining them with a reference image embedding (e.g., a 15-year-old male).We then computed the cosine distance between the reference embeddings and the target age embeddings, selecting those with the smallest distance to ensure the age transformation was realistic. This procedure is explained in III-A.1. While this process encourages identity consistency, it does not enforce strict identity preservation, meaning the generated faces may exhibit variations beyond age transformation. To mitigate excessive aging features drift, we limited the selection process to adjacent age groups, ensuring gradual and plausible aging effects.

Algorithm 1 Sampling for Age Transformation

```
Require: Reference image embeddings: $\mathbf{e}_{ref}$, Target age group: age$_{target}$, Target gender group: gender$_{target}$
Ensure: Transformed image embeddings $\mathbf{e}_{transformed}$
1: Generate descriptive captions for each image using the Florence-2 model.
2: Tokenize and encode captions using the CLIP text-encoder to obtain text embeddings $\mathbf{e}_{text}$.
3: Compute cosine distance between reference embeddings $\mathbf{e}_{ref}$ and identity embeddings of images within the target age and gender group:
       $d(\mathbf{e}_{ref}, \mathbf{e}_{target}) = 1 - \frac{\mathbf{e}_{ref} \cdot \mathbf{e}_{target}}{\|\mathbf{e}_{ref}\| \|\mathbf{e}_{target}\|}$
4: Select identity embeddings with the smallest cosine distance:
       $\mathbf{e}_{selected} = \arg \min_{\mathbf{e}_{target}} d(\mathbf{e}_{ref}, \mathbf{e}_{target})$
5: Retrieve the text embeddings corresponding to the index of $\mathbf{e}_{selected}$.
6: Concatenate the PCA-transformed text embeddings $\mathbf{a}$ with the reference identity embeddings: $\mathbf{e}_{transformed} = \mathbf{Concat}(\mathbf{e}_{ref}, \mathbf{a})$
```

Our data-driven method involved selecting the most similar embeddings from the target age group, allowing us to create age transformations that were both coherent with the original identity and consistent with the desired age progression. This approach enabled us to effectively manipulate age while preserving essential facial features, yielding high-quality identity aging transformations. Fig. 1 shows the proposed pipeline for the new text-conditioned context for age-controlled sampling between reference ID_i and ID_j, where $i \neq j$.

Fading enables localized facial attribute editing by utilizing text prompts and visual features within a multimodal diffusion process. This model requires carefully crafted prompts to guide modifications, making it highly sensitive to the wording and structure of the input. In the prompt-to-prompt editing approach, the model progressively refines the image by iteratively adjusting features according to successive text prompts. This editing occurs within the cross-attention mechanism, where the model attempts to align specific words with the corresponding image features encoded by the vision model. However, this process is highly sensitive to changes in phrasing, leading to different activations in iterations.

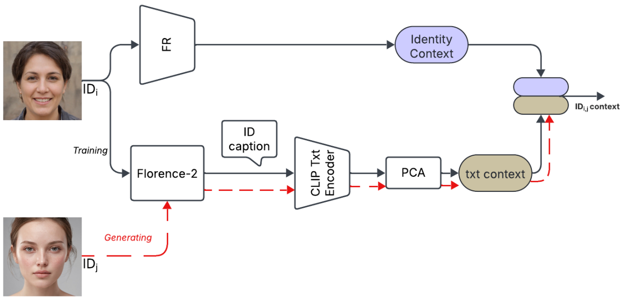

Fig. 1. Flowchart of the training and sampling procedures to generate the new context for the conditional diffusion model IDiff-Face [4].

Additionally, Fading lacks spatial constraints, meaning that certain words or prompts can trigger unintended changes to unrelated facial regions, especially when attributes are not adequately disentangled. The strict dependency on precise prompting can result in inconsistencies, as minor variations between successive prompts may lead to unexpected and sometimes undesirable changes in the generated output [7].

B. Gender and Age classification

To evaluate the impact of synthetic data on lightweight real-time models for gender and age classification, we employed a shared backbone with task-specific classification heads. This approach ensures efficient feature extraction while allowing for independent optimization of each task.

We conducted experiments using three lightweight backbone architectures: EfficientNet-B0, MobileNet, and LightViT-Small. These models were chosen for their computational efficiency and suitability for real-time applications. Each backbone was paired with distinct classification heads, specifically designed for gender classification (binary) and age classification (multi-class, with **8/10/13** categories). Age was treated as a classification problem rather than a regression task due to several key advantages: classification enhances model stability by grouping age into discrete categories, making it more robust to noise and outliers. It also improves interpretability, as age categories provide clearer insights compared to continuous age predictions. Furthermore, classification aligns better with real-world applications, where age is often grouped into ranges (e.g., 18-24, 25-34). This approach also reduces model complexity, resulting in faster training times and better performance, particularly in edge cases. Our approach allows for a direct comparison of how various backbone architectures utilize synthetic data for soft-biometric estimation.

IV. Results and Discussion

In this section we present the results and discussion of synthetic data generation using the two models: IDiff-Face-Aged and Fading, as well as the impact of synthetic data on the age and gender classification models.

A. Synthetic data generation

To enhance data diversity, we employed two distinct datasets for augmentation, leveraging their unique strengths to improve model robustness through complementary feature manipulation strategies.

The Fading approach utilizes CelebA-HQ [51], a high-resolution facial image dataset that emphasizes photorealism and diverse environmental conditions. This dataset is particularly suitable for Fading, as its rich visual details enable controlled modifications—such as changes in lighting, texture, and subtle facial cues—without compromising image realism. These samples were generated for 13 different ages (10, 14, 18, 23, 28, 33, 38, 43, 48, 53, 58, 64, 70 years old). We generated around 70,000 samples for about 5,400 identities. Some examples can be seen in Fig. 2.

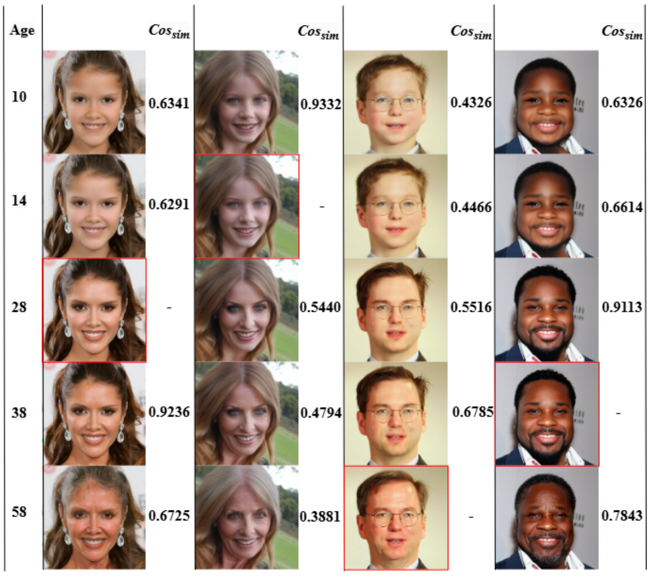

Fig. 2. Samples for Fading generated samples, their reference images (red-squared), and cosine similarities.

For the IDiff-Face-Aged method, we selected SFHQ [52], a dataset characterized by comprehensive identity annotations, a large number of samples, and consistent facial features. This dataset facilitates the employed augmentation strategy by providing a broader search space, thereby introducing greater feature diversity. Specifically, we generated approximately 50,000 synthetic samples of shape 224 × 224 from an original dataset of over 400,000 samples, prioritizing the augmentation of underrepresented age groups (particularly younger ages). Some examples are shown in Fig. 3.

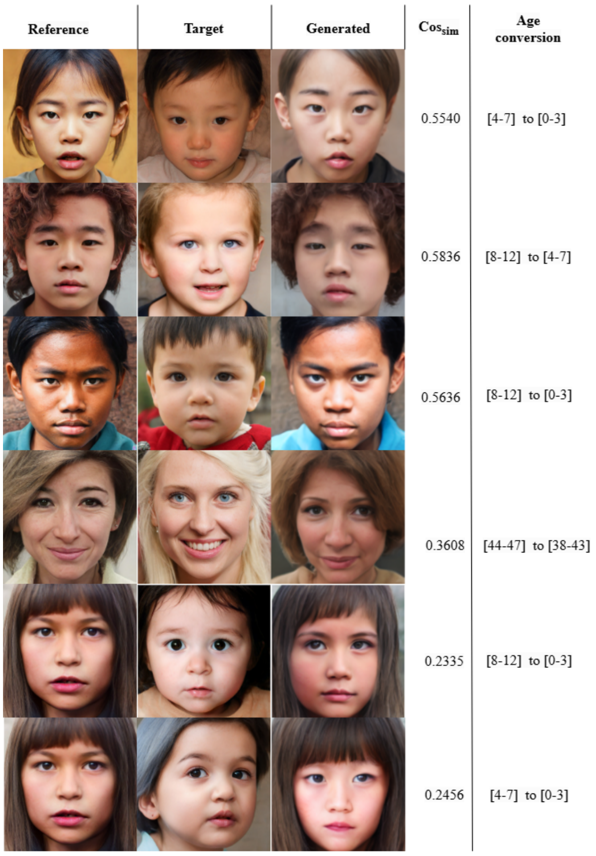

Fig. 3. Samples for IDiff-Face-Aged method, their reference and target images, and cosine similarities.

To evaluate the effect of synthetic data augmentation, we analyze the cosine similarity between each generated sample and its corresponding reference. This was calculated on the identity embeddings extracted with the proprietary face matching algorithm.

The generated samples demonstrate that while some reference features are preserved, there is a noticeable shift in facial structures, resulting in increased intra-class diversity. This is evident in the last two rows of the sample images, where the same reference image is used with different target age groups, demonstrating greater variability in the transformations. The relatively lower cosine similarity scores further support this observation, indicating that the generated faces introduce more substantial modifications. However, this approach still exhibits some stability issues, particularly when attempting to produce an age-modified version that closely resembles the reference image. As seen in the third row, the difficulty in maintaining consistency across groups with significantly different age-related characteristics may be contributing to this instability.

Fig. 4 presents the distribution of cosine similarity scores between generated samples and their corresponding reference images in the Fading (continuous line) and in the IDiff-Face-aged (dashed line). This distribution indicates that the IDiff-Face-Aged may be potentially useful for diverse age transformation while Fading samples may be preferable in applications requiring closer resemblance to the original sample. As depicted, IDiff-Face-Aged produces a unimodal distribution of cosine similarity scores, with a mode around 0.3. This suggests that our method consistently generates augmented samples that exhibit a specific degree of difference from the original images in the chosen feature space. By generating samples with this degree of dissimilarity, IDiff-Face-Aged may be particularly effective in forcing the age estimation model to learn features that are invariant to minor variations while being sensitive to more significant age-related changes. In contrast, the FADING method yields a bimodal distribution of cosine similarity scores. One peak is located at a very high similarity (close to 1.0), suggesting that FADING frequently produces samples that are highly similar to the original. The second is at lower similarity values (around 0.6-0.7). This bimodal nature implies that FADING generates augmentations with two distinct characteristics: many samples retain a very strong resemblance to the original, while others exhibit a more noticeable difference. The high similarity peak suggests that many augmented samples might offer limited benefit for training robustness, as they are very close to the originals. Visual inspection of the generated samples, Fig. 2, indicates that, although this method exhibits improved identity preservation, it occasionally suffers from structural inconsistencies, particularly in specific age groups. For instance, some samples fail to remove age-inconsistent features (e.g., beards in younger generations), suggesting that the model lacks adaptability in certain transformations. Additionally, the observed decrease in cosine similarity for extreme-age transformations implies that the model may rely on modifying a fixed set of features rather than adapting dynamically to different identities. This could result in limited diversity in age-specific edits, as the modifications appear relatively consistent across different individuals.

B. Age and Gender Experiments

Regarding the age and gender classification, different training and validation groups were created, with different amount of synthetic samples and sources, and then tested on

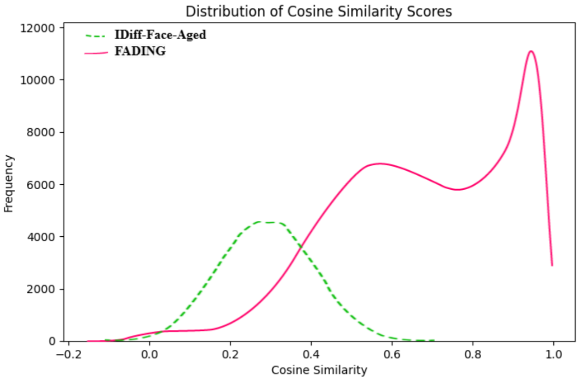

Fig. 4. Distribution of cosine similarity scores for both models' generated samples and their reference images.

UTKFace. For improved readability, the following acronyms are used to refer to the dataset sources: **I**: *IMDB-WIKI* - **L**: *Lagenda* - **A**: *Adience* - **F**: *FFHQ* - **Ia**: *IDiff-Face-aged* - **C**: *CelebA-HQ* - **Fa**: *Fading* - **S**: *SFHQ*.

We defined three distinct classification scenarios to systematically evaluate age and gender estimation performance under varying levels of granularity. The first scenario adheres to the categorical age groups from the Adience dataset (Table I), serving as a benchmark for comparison with existing approaches. This differs from the others because all ages are not presented in the age groups. The second scenario (Table II) utilizes a continuous class labeling scheme, enabling a finer-grained representation of age progression. The third scenario (Table III) merges early-age classes from the continuous labeling approach, aligning with practical applications where broader age group distinctions are more meaningful. For the experiments, all include IMDB-WIKI, Lagenda and Adience, because they are among the largest and most diverse publicly available datasets. To maintain a focused evaluation, we selectively chose dataset combinations, excluding those that performed poorly in initial simpler tests. These tables demonstrate the performance of various models on age and gender classification tasks - accuracy (Acc), precision (P), recall (R), and F1-score (F1), trained with different datasets and augmentation strategies.

Table I, shows that gender classification achieves high accuracy across all models, with Acc ranging from 91.84% to 94.17%. This suggests that gender classification is relatively robust to variations in dataset composition and augmentation strategies. Geometric augmentations (+) slightly enhances performance, but the most significant improvements stem from the model itself, with MultiLightViT achieving the highest accuracy.

In contrast, age classification presents a more complex challenge, with accuracies ranging from 62.99% to 69.38%, and precisions above 71%. Among the models, MultiLightViT stands out with the highest accuracy at 69.38%, highlighting the impact of model architecture in handling the complexities of age classification. These results also underscore the importance of dataset composition, particularly the significant improvements when the FFHQ (*ILAF* subset) was included. This table serves as a general evaluation to benchmark models and establish the best baseline datasets for more intricate and complex age classification scenarios, where synthetic data was not used.

Table II presents results for continuous age classification across 13 classes. Again, MultiLightViT outperforms other models in both gender and age classification tasks. The combination of augmentations (*ILAF* +) achieves the best performance for MultiLightViT, with 95.43% Acc for gender classification and 48.23% for age classification.

The results also highlight the importance of the quality and diversity of the datasets used for training and testing. The combination of datasets (*ILAFIaCFaS* +) yields an Acc of 94.35% for gender classification and 44.83% for age classification for the MultiMobileNet model, but the age performance is lower than that of the combination of only real-world augmentations (*ILAF* +). As shown by the *ILAFIaCFaS* subset, over-representation of synthetic data can cause the model to under-perform on real-world datasets. This suggests that synthetic data, while useful for balancing datasets and capturing diverse features, still struggle to model real-world facial aging characteristics that are crucial for distinguishing adjacent age groups.

In the 10-class age classification scenario (Table III), where ages under 14 are grouped together into a single class due to their practical limited usage and their face features similarity, gender accuracy remains consistently high, between 93.59% and 95.66%. These results show balanced performance across models, with minimal disparity in gender classification accuracy. In contrast, age classification presents a more challenging task, with accuracies ranging from around 48% to 53.79%. MultiLightViT tends to outperform MultiMobileNet in this domain, highlighting its superior capability in handling the nuances of age categorization. The lower F1-scores compared to accuracy metrics suggest potential difficulties in accurately classifying certain age groups or distinguishing between adjacent age ranges. The improved performance of MultiLightViT with more synthetic data, particularly in subsets *ILAFIaS* and *ILAFCFaS*, can be attributed to the additional samples generated for the under-14 age group. This synthetic data helps balance the dataset

TABLE I

PERFORMANCE COMPARISON OF VARIOUS MODELS FOR GENDER (0, 1) AND AGE CLASSIFICATION (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-), INCLUDING ACCURACY, PRECISION, RECALL, AND F1-SCORE METRICS, WITH (+) AND WITHOUT(-) GEOMETRIC AUGMENTATIONS.

Dataset	**Model**	**Gender**				**Age**			
		Acc (%)	**P (%)**	**R (%)**	**F1 (%)**	**Acc (%)**	**P (%)**	**R (%)**	**F1 (%)**
ILA -	MultiMobileNet	91.84	91.84	91.84	91.84	62.99	66.87	62.99	64.19
ILA +	MultiMobileNet	92.02	92.03	92.02	92.02	64.22	67.70	64.22	65.34
ILAF +	MultiMobileNet	92.70	92.70	92.70	92.70	64.69	68.04	64.69	65.92
	MultiEfficientNet-b0	93.80	93.81	93.80	93.80	66.15	71.04	66.15	67.75
	MultiLightViT	**94.17**	**94.18**	**94.17**	**94.17**	**69.38**	**71.60**	**69.38**	**70.26**

TABLE II

CONTINUOUS AGE GROUPS: PERFORMANCE COMPARISON OF VARIOUS MODELS FOR GENDER (0, 1) AND CONTINUOUS AGE CLASSIFICATION (0-3, 4-7, 8-12, 13-14, 15-20, 21-24, 25-32, 33-37, 38-43, 44-47, 48-53, 54-59, 60-100), INCLUDING ACCURACY, PRECISION, RECALL, AND F1-SCORE METRICS WITH (+) GEOMETRIC AUGMENTATIONS.

Dataset	**Model**	**Gender**				**Age**			
		Acc (%)	***P (%)***	***R (%)***	***F1 (%)***	***Acc (%)***	***P (%)***	***R (%)***	***F1 (%)***
ILAF +	MultiMobileNet	93.57	93.57	93.57	93.57	47.75	45.66	47.75	44.41
	MultiEfficientNet-b0	95.28	95.28	95.28	95.28	48.88	47.92	48.88	46.84
	MultiLightViT	**95.43**	**95.44**	**95.43**	**95.43**	**48.23**	**48.60**	**48.23**	**47.57**
ILAFIa +	MultiMobileNet	94.77	94.78	94.77	94.77	46.87	46.69	46.87	45.62
	MultiLightViT	95.40	95.44	95.40	95.40	48.01	47.12	48.01	46.60
ILAFIaCFa +	MultiLightViT	95.36	95.36	95.36	95.36	48.10	48.37	48.10	46.63
ILAFIaS +	MultiMobileNet	95.11	95.12	95.11	95.11	47.03	47.01	47.03	45.97
	MultiLightViT	95.32	95.33	95.33	95.32	48.09	47.09	48.09	46.86
ILAFCFaS +	MultiMobileNet	94.20	94.21	94.20	94.20	45.84	45.56	45.84	44.10
	MultiLightViT	95.30	95.30	95.30	95.30	47.17	47.13	47.17	46.18
ILAFIaCFaS +	MultiMobileNet	94.35	94.35	94.35	94.35	44.83	44.94	44.83	43.38

and aids the model in generalizing better to real-world data.

Overall, our findings underscore the complexity of age and gender classification and the interplay between model architecture, dataset composition, and augmentation strategies. While gender classification remains relatively stable across different models and dataset variations, age classification presents greater challenges due to the fine-grained nature of age progression, the domain gap between synthetic and real-world data. The inclusion of synthetic datasets such as Fading and IDiff-Face-Aged provides valuable age-related variations, helping models capture aging patterns more effectively. However, their impact is highly dependent on how well the synthetic data aligns with real-world distributions. Our results suggest that while synthetic data can enhance model generalization, an excessive reliance on it may introduce biases that reduce real-world performance, and temporal age progression may introduce biases and artifacts on aging features that do not improve performance.

V. CONCLUSIONS AND FUTURE WORKS

This study highlights the substantial impact of dataset composition, augmentation strategies, and the integration of synthetic data on the performance of age and gender classification models. Results show that carefully combining real and synthetic datasets—particularly those like Fading and IDiff-Face-Aged that offer structured aging patterns—can significantly boost classification accuracy and model generalization.

VI. ACKNOWLEDGMENTS

This project is funded by the Portuguese Recovery and Resilience Plan (RPP) under the program 'Agendas para a Inovação Empresarial', reference no. 62 - 'Center for Responsible AI'.

ETHICAL IMPACT STATEMENT

This work does not involve human subject studies or interventions requiring informed consent or IRB approval. All datasets used in this study are either publicly available or synthetically generated, and do not contain any personally identifiable information. As such, ethical board oversight was not required.

Our research investigates the use of synthetic data generated via diffusion-based models to improve age and gender classification tasks. While this can offer performance and fairness benefits by reducing dataset biases, we acknowledge potential risks, particularly in the misuse of synthetic faces in surveillance, deepfakes, or identity manipulation.

TABLE III

Early ages grouping: Performance comparison of various models for gender (0, 1) and continuous age classification (0-14, 15-20, 21-24, 25-32, 33-37, 38-43, 44-47, 48-53, 54-59, 60-100), including accuracy, precision, recall, and F1-score metrics with (+) geometric augmentations.

Dataset	**Model**	**Gender**				**Age**			
		Acc (%)	***P (%)***	***R (%)***	***F1 (%)***	***Acc (%)***	***P (%)***	***R (%)***	***F1 (%)***
ILAF +	MultiMobileNet	94.34	94.38	94.34	94.34	51.08	49.52	51.08	49.46
	MultiEfficientNet-b0	95.06	95.06	95.06	95.06	50.62	50.28	50.62	49.79
	MultiLightViT	95.14	95.17	95.14	95.13	51.08	49.52	51.06	49.44
ILAFIa +	MultiMobileNet	94.47	94.47	94.47	94.47	51.48	50.59	51.48	50.19
	MultiLightViT	95.36	95.37	95.36	95.36	52.64	50.75	52.62	50.63
ILAFFa +	MultiMobileNet	94.06	94.13	94.06	94.06	48.28	47.75	48.28	47.33
ILAFIaCFa +	MultiMobileNet	94.48	94.49	94.48	94.48	50.82	49.23	50.82	49.01
	MultiLightViT	95.38	95.39	95.38	95.38	49.60	50.45	49.60	49.51
ILAFIaS +	MultiMobileNet	94.28	94.33	94.28	94.28	51.67	49.84	51.67	49.58
	MultiLightViT	**95.66**	**95.66**	**95.66**	**95.66**	52.46	51.36	52.46	51.28
ILAFCFaS +	MultiMobileNet	93.59	93.61	93.59	93.59	52.19	52.27	52.19	50.46
	MultiLightViT	95.17	95.18	95.17	95.17	**53.79**	**53.07**	**53.79**	**52.05**

To mitigate these risks, we strictly use synthetic data for model training and never for deceptive content generation. Our data is generated under controlled, research-focused conditions, and we ensure that it cannot be misused to impersonate real individuals. We also emphasize the model's limitations and advocate for transparency in data sources when deploying biometric systems.

We believe the benefits of this work—improving model generalization, fairness, and performance on underrepresented demographics—outweigh the risks, especially when paired with responsible research practices and open discussion of societal implications.

References

[1] R. Angulu, J. R. Tapamo, and A. O. Adewumi, "Age estimation via face images: a survey," *Journal on Image and Video Processing*, vol. 2018, Dec 2018.

[2] G. Panis, A. Lanitis, N. Tsapatsoulis, and T. F. Cootes, "Overview of research on facial ageing using the fg-net ageing database," *IET Biometrics*, vol. 5, no. 2, pp. 37–46, 2016.

[3] M. Kuprashevich and I. Tolstykh, "Mivolo: Multi-input transformer for age and gender estimation," *arXiv preprint*, Jul. 2023.

[4] F. Boutros, J. H. Grebe, A. Kuijper, and N. Damer, "Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion models," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023.

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, (Cambridge, MA, USA), p. 2672–2680, MIT Press, 2014.

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, (Red Hook, NY, USA), Curran Associates Inc., 2020.

[7] X. Chen and S. Lathuilière, "Face aging via diffusion-based editing," in *34th British Machine Vision Conference 2023, BMVC 2023, Aberdeen, UK, November 20-24, 2023*, BMVA, 2023.

[8] B. Xiao, H. Wu, W. Xu, X. Dai, H. Hu, Y. Lu, M. Zeng, C. Liu, and L. Yuan, "Florence-2: Advancing a unified representation for a variety of vision tasks," *arXiv preprint arXiv:2311.06242*, 2023.

[9] G. Ilharco, M. Wortsman, R. Wightman, C. Gordon, N. Carlini, R. Taori, A. Dave, V. Shankar, H. Namkoong, J. Miller, H. Hajishirzi, A. Farhadi, and L. Schmidt, "OpenCLIP," 2024.

[10] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Boston, USA), pp. 34–42, IEEE, 2015.

[11] X. Liu, S. Li, S. Shan, and X. Chen, "Agenet: Deeply learned regressor and classifier for robust apparent age estimation," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2015.

[12] A. V. Savchenko, "Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output cnn," *PeerJ Computer Science*, vol. 4, p. e197, Jul 2018.

[13] A. V. Savchenko, "Facial expression and attributes recognition based on multi-task learning of lightweight neural networks," in *2021 IEEE 19th International Symposium on Intelligent Systems and Informatics (SISY)*, Mar 2021.

[14] H. Zhang, Y. Zhang, and X. Geng, "Practical age estimation using deep label distribution learning," *Frontiers of Computer Science*, vol. 15, Jun 2021.

[15] Y. Deng, S. Teng, L. Fei, W. Zhang, and I. Rida, "A multifeature learning and fusion network for facial age estimation," *Sensors*, vol. 21, Jul 2021.

[16] A. Akbari, M. Awais, S. Fatemifar, S. S. Khalid, and J. Kittler, "A novel ground metric for optimal transport-based chronological age estimation," *IEEE Transactions on Cybernetics*, vol. 52, pp. 9986–9999, Oct 2022.

[17] Y. Shou, X. Cao, H. Liu, and D. Meng, "Masked contrastive graph representation learning for age estimation," *Pattern Recognition*, vol. 158, Feb 2025.

[18] H. Wang, V. Sanchez, and C.-T. Li, "Improving face-based age estimation with attention-based dynamic patch fusion," *IEEE Transactions on Image Processing*, vol. 31, pp. 1–12, Dec 2022.

[19] S. Hiba and Y. Keller, "Hierarchical attention-based age estimation and bias analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 12, pp. 14682–14692, 2023.

[20] A. Swaminathan, M. Chaba, D. K. Sharma, and Y. Chaba, "Gender classification using facial embeddings: A novel approach," in *International Conference on Computational Intelligence and Data Science*, pp. 2634–2642, Elsevier B.V., 2020.

[21] M. Alghaili, Z. Li, and H. A. R. Ali, “Deep feature learning for gender classification with covered/camouflaged faces,” *IET Image Process*, vol. 14, pp. 3957–3964, Dec. 2020.

[22] M. M. Islam, N. Tasnim, and J. H. Baek, “Human gender classification using transfer learning via pareto frontier cnn networks,” *Inventions*, vol. 5, Jun. 2020.

[23] N.-H. Shin, S.-H. Lee, and C.-S. Kim, “Moving window regression: A novel approach to ordinal regression,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Mar. 2022.

[24] M. Kuprashevich, G. Alekseenko, and I. Tolstykh, “Beyond specialization: Assessing the capabilities of mllms in age and gender estimation,” *arXiv preprint*, Mar. 2024.

[25] G. Bae, M. de La Gorce, T. Baltrušaitis, C. Hewitt, D. Chen, J. Valentin, R. Cipolla, and J. Shen, “Digiface-1m: 1 million digital face images for face recognition,” in *2023 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2023.

[26] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao, “Synface: Face recognition with synthetic data,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10880–10890, 2021.

[27] F. Boutros, M. Huber, A. T. Luu, P. Siebke, and N. Damer, “Sface2: Synthetic-based face recognition with w-space identity-driven sampling,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, pp. 1–1, 2024.

[28] S. Walton, A. Hassani, X. Xu, Z. Wang, and H. Shi, “Stylenat: Giving each head a new perspective,” *ArXiv*, vol. abs/2211.05770, 2022.

[29] Z. Zhang, Y. Song, and H. Qi, “Age progression/regression by conditional adversarial autoencoder,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4532–4360, Feb. 2017.

[30] Z. Wang, X. Tang, W. Luo, and S. Gao, “Face aging with identity-preserved conditional generative adversarial networks,” in *2018 Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 7939–7947, IEEE Computer Society, Dec. 2018.

[31] Z. Li, R. Jiang, and P. Aarabi, “Continuous face aging via self-estimated residual age embedding,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15003–15012, Apr. 2021.

[32] G. Antipov, M. Baccouche, and J.-L. Dugelay, “Face aging with conditional generative adversarial networks,” in *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2089–2093, Feb. 2017.

[33] G. S. Hsu, R. C. Xie, and Z. T. Chen, “Wasserstein divergence gan with cross-age identity expert and attribute retainer for facial age transformation,” *IEEE Access*, vol. 9, pp. 39695–39706, 2021.

[34] W. Wang, Y. Yan, Z. Cui, *et al.*, “Recurrent face aging with hierarchical autoregressive memory,” *IEEE Trans Pattern Anal Mach Intell*, vol. 41, no. 3, pp. 654–668, 2019.

[35] F. Makhmudkhujaev, S. Hong, and I. K. Park, “Re-aging gan: Toward personalized face age transformation,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3888–3897, 2021.

[36] Z. Huang, J. Zhang, and H. Shan, “When age-invariant face recognition meets face age synthesis: A multi-task learning framework and a new benchmark,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 7917–7932, 2023.

[37] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher, “Lifespan age transformation synthesis,” in *2020 Computer Vision–European Conference on Computer Vision (ECCV)*, pp. 739–755, 11 2020.

[38] T.-K. Hsieh, T.-J. Liu, and K.-H. Liu, “Agesynthgan: Advanced facial age synthesis with stylegan2,” in *2024 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 1–5, IEEE, 2024.

[39] M. Kim, F. Liu, A. Jain, and X. Liu, “Dcface: Synthetic face generation with dual condition diffusion model,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12715–12725, June 2023.

[40] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, “SDEdit: Guided image synthesis and editing with stochastic differential equations,” in *International Conference on Learning Representations*, 2022.

[41] O. Avrahami, O. Fried, and D. Lischinski, “Blended latent diffusion,” *ACM Trans. Graph.*, vol. 42, jul 2023.

[42] A. Q. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. Mcgrew, I. Sutskever, and M. Chen, “GLIDE: Towards photorealistic image generation and editing with text-guided diffusion models,” in *Proceedings of the 39th International Conference on Machine Learning* (K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, eds.), vol. 162 of *Proceedings of Machine Learning Research*, pp. 16784–16804, PMLR, 17–23 Jul 2022.

[43] G. Couairon, J. Verbeek, H. Schwenk, and M. Cord, “Diffedit: Diffusion-based semantic image editing with mask guidance,” *International Conference in Learning Representations*, 2023.

[44] A. Hertz, R. Mokady, J. Tenenbaum, K. Aberman, Y. Pritch, and D. Cohen-Or, “Prompt-to-prompt image editing with cross attention control,” *arXiv preprint arXiv:2208.01626*, 2022.

[45] R. Mokady, A. Hertz, K. Aberman, Y. Pritch, and D. Cohen-Or, “Null-text inversion for editing real images using guided diffusion models,” *arXiv preprint arXiv:2211.09794*, 2022.

[46] B. Kawar, S. Zada, O. Lang, O. Tov, H. Chang, T. Dekel, I. Mosseri, and M. Irani, “Imagic: Text-based real image editing with diffusion models,” in *Conference on Computer Vision and Pattern Recognition 2023*, 2023.

[47] J. Wahid, F. Zhan, P. Rao, and C. Theobalt, “Diffage3d: Diffusion-based 3d-aware face aging,” *ArXiv*, vol. abs/2408.15922, 2024.

[48] P. Melzi, C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, D. Lawatsch, F. Domin, and M. Schaubert, “Gandiffface: Controllable generation of synthetic datasets for face recognition with realistic variations,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 3086–3095, October 2023.

[49] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4396–4405, 2019.

[50] Y. Lin, J. Shen, Y. Wang, and M. Pantic, “Fp-age: Leveraging face parsing attention for facial age estimation in the wild,” *arXiv*, 2021.

[51] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of gans for improved quality, stability, and variation,” *ArXiv*, vol. abs/1710.10196, 2017.

[52] D. Beniaguev, “Synthetic faces high quality (sfhq) dataset,” 2022.

[53] Z. Zhang, Y. Song, and H. Qi, “Age progression/regression by conditional adversarial autoencoder,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2017.

[54] Melodiz, “Face gender age recognition.” `https://github.com/Melodiz/face-gender-age-recognation/blob/main/demo/age_distribution.jpg`, 2023.

[55] J. Deng *et al.*, “Insightface.” `https://github.com/deepinsight/insightface`, 2019.