Enhancing Facial Emotion Recognition in the Elderly: The Impact of Age-Diverse Training Data on Model Accuracy

Majid Azizi*, Viking Forsman†, Joni Ojala‡, Mattias Tidström§
School of Innovation, Design and, Engineering
Mälardalens University, Västerås, Sweden
Email: *mai20018, †img13001, ‡joa20001, §mtm20001@student.mdu.se

***Abstract*—This paper explores the challenges and methods of Facial Emotion Recognition (FER), with a specific focus on its application within the elderly demographic, a group highly susceptible to dementia. The team is studying the use of advanced models for FER, along with age estimation, to study how well these models perform on both general and elderly-specific datasets. The goal is to answer two key questions: whether the accuracy of FER improves for people aged 60 and above by using transfer learning with data only from this age group, and whether the accuracy improves when the model is trained on data from all age groups. The results show that models trained on a diverse set of datasets, including all age ranges, achieve higher accuracy when tested on elderly subjects. This unexpected outcome emphasizes the importance of having a large and varied dataset rather than one for a specific age group. The study's implications suggest that the key to improving FER accuracy for the elderly lies in the size and variety of the training data, which could lead to more empathetic and effective healthcare solutions for age-related conditions.**

***Index Terms*—Facial Emotion Recognition, Elderly Demographic, Age Estimation, Artificial Intelligence, Healthcare Technology.**

I. Introduction

This academic project, initiated by Börje Bjelke, a geriatrician, along with Martin Hellström and Mikael Ekström, involves applying real-time FER technology to a telepresence robot. The inspiration for this project stemmed from a practice in Norway, where clowns visited hospitals to entertain patients and improve their mood. However, due to the COVID-19 pandemic and subsequent restrictions, the clowns could no longer physically visit hospitals. In response, the Norwegian government implemented a method for the clowns to perform via telepresence robots. Building upon this concept, Börje Bjelke conceived the idea of utilizing these telepresence robots to observe and monitor the mood and well-being of individuals with dementia. A step towards this overarching goal is to be able to recognize the emotions of the elderly, given their higher susceptibility to dementia compared to other age demographics [1].

Research indicates that a significant portion of human communication, approximately two-thirds, occurs through non-verbal means, with facial expressions serving as a pivotal component in this mode of interaction [2]. This phenomenon underscores the importance of FER, a field within artificial intelligence focused on discerning and categorizing human emotions based on facial cues. FER employs machine learning algorithms and computer vision techniques to analyze facial features and expressions, thereby translating them into distinct emotional states [3].

The application of FER holds considerable significance across diverse domains, including human-computer interaction, law enforcement, and mental health assessment [4]. In healthcare settings, FER technology can be instrumental in monitoring the emotional well-being of patients, particularly in contexts such as mental health facilities, to facilitate more accurate diagnosis and treatment of conditions like depression and anxiety. Similarly, in security applications, emotion recognition systems can bolster surveillance efforts by identifying potential threats through the analysis of emotional indicators.

In the realm of emotion classification, Paul Ekman, a renowned figure in psychology, delineated six primary emotions that exhibit universality across cultures and are manifested similarly across diverse societies [5]: happiness, sadness, anger, fear, surprise, and disgust. Later contempt was added as an additional fundamental emotion, thus expanding the list to seven [6]. Achieving a nuanced understanding and accurate description of these emotions necessitates meticulous analytical approaches.

The Facial Action Coding System (FACS) [7] represents one such methodological framework designed to provide objective descriptions of facial expressions. FACS achieves this objective by systematically identifying and categorizing the specific movements of individual facial muscles. Moreover, it acknowledges the presence of a neutral facial state characterized by relaxed muscles, which serves as a baseline for comparative analysis in FER tasks.

This paper will explore whether apparent age is a contributing factor to the difficulties of FER amongst the elderly. The proposed method explores this using State Of The Art **(SOTA)** models for both age estimation and FER. The overarching goal is to answer the following questions:

RQ1. Will the accuracy of FER for individuals aged 60 and above improve if transfer learning is used exclusively with data from this age group?

RQ2. Will the accuracy of FER for individuals aged 60 and above improve if transfer learning is used with data from all age groups?

II. Background

This section will delve into the intersection of Convolutional Neural Networks (**CNN**), Face Detection, Age Estimation, and the datasets specifically tailored to elderly subjects. These components represent crucial pillars in the advancement of both computer vision research and healthcare technology [3].

A. Convolutional Neural Networks

CNNs [8] are inspired by the connectivity pattern of neurons in the human brain and are designed to automatically learn different hierarchical features from data. What makes CNN more efficient than a fully connected neural network, is its local connectivity, shared weights, and hierarchical feature extraction [9]. It also allows CNN to effectively capture spatial and local features from the input, such as edges and textures. All of these features make CNN great for image recognition and classification tasks, such as facial recognition, age detection, and FER.

CNN typically requires a substantial volume of data to achieve optimal performance. However, in scenarios where data availability is limited, the technique of transfer learning can be employed to enable the CNN to adapt to new tasks [10]. This involves taking a pre-trained CNN model and adjusting it to improve its performance on a specific task. Fine-tuning can lead to faster convergence and better performance as the model has already learned useful features from the large pre-training dataset. On the other hand, training a CNN from scratch may require more data and computational resources to achieve comparable performance.

B. Face Detection

According to Kumar et al., "Face detection is a computer technology that determines the location and size of a human face in a digital image" [11]. While this task is effortless for humans, it has historically been one of the most challenging and researched topics within computer vision. However, modern advancements within this field have successfully introduced lightweight models with high accuracy suitable for embedded systems [12]. There are two common approaches to detecting human faces in images: feature-based and image-based [11]. The feature-based approach involves extracting facial features from images and comparing them to known data on facial features. On the other hand, the image-based approach aims to achieve the best possible match between the training and testing images.

C. Age Estimation

Precisely determining the chronological age of a human is a difficult task [13]. Experiments have proved that machine learning has surpassed the human ability to estimate age from images [14]. CNN-based methods have shown improved performance compared to traditional approaches within this field, as they can automatically learn age-related facial features from data, Levi et al. [15] showed in their paper that even with limited data a simple CNN architecture could achieve high performance. They evaluated their method on the Adience benchmark for age and gender estimation and showed that a simple CNN outperformed the age estimation SOTA at that time. CNN-based models have become the predominant approach for age estimation from facial images [16].

D. Datasets with elderly subjects

The lack of publicly available datasets with elderly subjects presents a significant challenge in FER research [3]. While there are numerous datasets available for studying facial expressions in general, many of these datasets predominantly feature young to middle-aged adults. This limitation in data can introduce problems during the training process. The model may not learn to effectively recognize emotions for the underrepresented age groups. In the case of elderly subjects, their facial expressions may alter significantly due to age-related factors such as changes in facial musculature, skin elasticity, and overall appearance [17]. The model's performance may become skewed towards the age groups that are well represented while introducing negative biases toward underrepresented groups. Commonly used datasets such as Affectnet [18], RAF-DB [19], and FER2013 [20] do not include many elderly subjects compared to other age groups, exacerbating this issue. See Figure 1 for a visual representation of the age distribution.

There are some datasets that focus on, or at least have a good representation of, elderly subjects. Examples of such datasets are FACES [21], EISVDB [22], ElderReact [23], and Tsinghua [24], see detailed specification in table I. However, it is important to note that there may be some differences in what the papers define as "elderly," which can further complicate the standardization of data in this domain. Only FACES and Tsinghua in the aforementioned datasets define the ages of the elderly subjects, which are between 58 to 80, see further details in table II.

Table I
FER DATASETS WITH GOOD REPRESENTATION OF ELDERLY SUBJECTS

Dataset	Subjects	Samples	Emotions
FACES	58 young 56 adult 57 elderly	2052 images, 2835x3543 pixels	anger, disgust, fear, happy, neutral, sad
EISVDB	16 elderly	810 videos, 3 to 5 seconds	anger, boredom, disgust, happy, neutral, sad, surprise
ElderReact	46 elderly	1323 videos, 3 to 8 seconds	anger, disgust, fear, happy, sad, surprise
Tsinghua	67 adult 70 elderly	1128 images, 1500x2000 pixels	anger, content, disgust, fear, happy, neutral, sad, surprise

Table II
FER DATASETS DEFINITION OF THE ELDERLY

Dataset	Age Brackets	Mean age	Standard Deviation
FACES	69 – 80 years	73.2 years	±2.8 years
EISVDB	Not defined	Not defined	Not defined
ElderReact	Not defined	Not defined	Not defined
Tsinghua	58 - 72 years	64.4 years	± 3.51 years

III. Related Works

Caroppo et al. [23] presents a deep learning approach based on Stacked Denoising Autoencoders for automatic FER on elderly individuals, validated on benchmark datasets containing expressions by older individuals. The study highlights the complexity of older adults' facial expressions, influenced by factors like fewer facial regions involved but more blended expressions, making classification challenging. The proposed deep learning method outperforms non-deep learning approaches when tested on two datasets with elderly subjects, achieving an accuracy of 88.2% and 93.3%.

Fei et al. [25] presents an approach for the automatic detection of mild cognitive impairment in elderly individuals through FER. They used layers from MobileNet, a lightweight CNN architecture, and an SVM to perform the classifications. An experiment with 61 subjects showed that their approach successfully achieved a detection accuracy of 73.3%.

A paper by Ma et al. [26] introduces the ElderReact dataset, designed for recognizing emotional responses in elderly adults, addressing a research gap in this demographic. They performed experiments comparing traditional machine learning algorithms with deep learning algorithms and found that the latter performed better for FER, with visual features being more indicative than audio. However, Cohen's Kappa metric, which represents the level of agreement between human annotators, reveals a gap in matching human performance. Cross-dataset experiments with the EmoReact dataset for children highlight the need for age-specific models, as models trained on one age group struggle to generalize to another. The study underscores the importance of tailored emotion recognition systems for the elderly.

A study by Jaiyen et al. [27] focuses on utilizing YOLOv7 for detecting facial expressions in Thai elderly individuals. The paper highlights that the facial features of Thai individuals differ from their European counterparts, particularly in the older generations. To facilitate their research, the authors curated a dataset for training and testing purposes, consisting of frames extracted from various sources such as TV programs, movies, and documentaries featuring elderly subjects. The performance of YOLOv7 surpassed that of two other models, namely Faster R-CNN and SSD, as evaluated in the experimental setup.

Kim et al. [28] analyzed the performance of SOTA commercial FER systems on individuals of different age groups. Their result showed that such systems exhibited higher accuracy in recognizing emotions in younger adults compared to middle-aged and older adults. There was a notable age bias in emotion detection, with older adults consistently receiving lower classification accuracy across all systems. Additionally, the study emphasized the importance of sophisticated bias mitigation techniques.

IV. Problem Formulation

The problems faced in this study can be broken down into three points:

(i) Existing FER models have a significantly lower accuracy when applied to elderly individuals [28].
(ii) There is a significant lack of publicly available labeled data for facially expressed emotions of elderly individuals [17].
(iii) Elderly individuals exhibit less obvious facial expressions compared to younger age groups [29].

Training a CNN requires a large amount of labeled training data, this is a challenge to overcome [30]. Due to the lack of elderly subjects in the datasets, age detection will be used to categorize the subjects by their perceived age. The subjects' data will then be split into new datasets based on perceived age and be used in fine-tuning an age-specific model. This way less data is used for training the specified model. The hypothesis is that this approach will reduce the poor representation of elderly facial features in the dataset and overcome the model's age bias.

V. Method

This section outlines the systematic approach employed to investigate the impact of age on FER accuracy among elderly individuals. The study introduces a novel method that integrates advanced models for age estimation, FER, and facial detection. The objective is to address two research questions (**RQ1** and **RQ2**): the effect of model fine-tuning on emotion recognition accuracy for the elderly, and the comparative accuracy of models fine-tuned on datasets of varying age diversity. The methodology is designed to enhance the precision of FER systems when applied to an aging population. The following subsections will go further into detail regarding the choice of models, the data-collection process, and the evaluation method.

A. Data collection

In the initial stage of the project, different datasets were explored for training the model. Access was gained to datasets with a good representation of elderly subjects, namely ElderReact [26], FACES [21], and Tsinghua [24]. The group also explored three larger publicly available datasets, which did not target any specific age groups, namely AffectNet [18], FER2013 [20], and RAF-DB [19]. The decision was made to define 'elderly' subjects in this study as individuals aged 60 years or above, this is equivalent of the threshold used in Tsinghua and slightly younger than the threshold in FACES. The lack of age labeling amongst the different datasets led the group to use the 'perceived' age of the subjects. This approach could potentially be more accurate than using the actual age

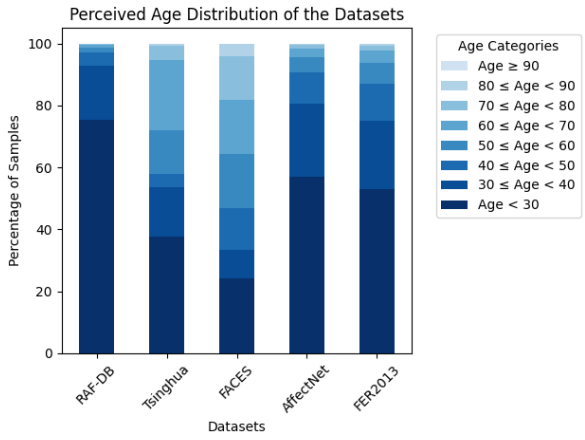

Figure 1. Age distribution of perceived ages in the explored datasets (excluding ElderReact, which uses videos instead of images).

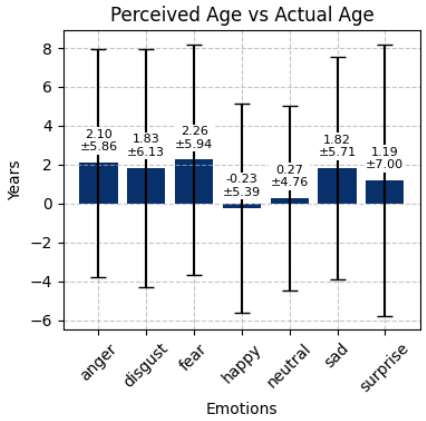

Figure 2. Mean and standard deviation between actual and perceived age (using all ages in FACES and Tsinghua)

since physical aging varies among individuals. However, it is also important to note that the subject's facial expression influenced the age estimation model's predictions. Happiness and neutrality were the most accurate in reflecting the actual age, while other expressions tended to raise the perceived age. See figure 2 for a visual representation on how subjects perceived age deviated based on the emotions.

After exploring the alternatives FACES, Tsinghua, and RAF-DB were chosen to fine-tune the model. Both the FACES and Tsinghua datasets exhibit balanced representation across age groups, emotions, and genders among their subjects [21], [24]. Additionally, they offer high-resolution images captured in controlled environments with optimal lighting conditions, making them highly suitable for training purposes. It should be noted that FACES lacks representation of the emotion "surprise". Tsinghua includes all emotions of interest in the study, but it exclusively features Chinese subjects, which might not generalize well to the other datasets in the study that predominantly have Caucasian subjects. RAF-DB is a dataset collected from various internet sources and the images contain diverse poses, occlusions (e.g., glasses, hats), and lighting conditions to reflect real-world settings [19]. While RAF-DB is large it has a rather minuscule number of elderly subjects.

A decision was made to exclude ElderReact, FER2013, and AffectNet from the fine-tuning process. ElderReact is a video dataset capturing elderly subjects' reactions to visual stimuli [26]. However, one issue is that individual videos in the dataset can be labeled with multiple emotions, either expressed simultaneously or sequentially. The model in this study does not classify combinations of emotions. Additionally, it requires images for the training process, not videos. Both FER2013 and AffectNet are sourced from online images, and are quite large datasets. However, the images in FER2013 have a resolution of 48x48 pixels, which might not be adequate to capture the facial features of elderly subjects. The resolution in AffectNet is higher, but the model that the team decided to use is pre-trained on this dataset, which could introduce bias if used in the validation or testing during the fine-tuning process.

The data collection process yielded two combined datasets: one comprising solely elderly subjects (1113 images) and the other encompassing all age groups (16399 images). These datasets were subsequently partitioned into test (15%), training (70%), and validation (15%) subsets. Additionally, the training subset was augmented using horizontal flips and by obscuring parts of the images.

B. Models

A new dataset was generated by combining and modifying multiple datasets to meet the requirements set by the base model chosen for further training. Following that process, the data samples were sorted and divided based on the subjects' perceived age. Our data consists of a generalized dataset and a separate dataset for elderly subjects since elderly facial features do not generalize as well to other age groups. The overall fine-tuning process is depicted in Figure 3.

MiVOLO [31] is a multi-input transformer model that excels in age and gender estimation by leveraging both facial and body image data. The age estimation process involves several key steps:

- *Dual Input*: MiVOLO integrates age and gender recognition tasks into a unified framework, processing both face and body images to enhance accuracy.
- *Feature Fusion*: The model employs a feature enhancer module for cross-view feature fusion, enriching the features with additional information from both inputs.
- *VOLO Architecture*: Utilizing the two-stage VOLO architecture, MiVOLO processes the fused features through a series of transformers, increasing its accuracy in age and gender predictions.
- *Loss Functions*: For training, MiVOLO uses a combination of WeightedMSE loss for age prediction and BinaryCrossEntropy loss for gender prediction, optimizing for both tasks simultaneously.

This approach allows MiVOLO to deliver SOTA performance in age estimation, even in challenging scenarios where the face is not fully visible.

Poster++ [32] is a FER model built upon Poster but with improvements for both accuracy and computational pressure.

- *Single Stream Design*: The authors of this model decided to remove one previous feature of the model, namely the image-to-landmark branch, and solely focused on the landmark-to-image branch. This decision was made to reduce computational consumption, making the model more lightweight.
- *Cross-fusion*: The model uses a window-based cross-attention mechanism for linear computation. This is done by dividing the image into several non-overlapping windows, extracting features, and then performing a cross-attention for all windows.
- *Multi-scale Feature Extraction*: The model extracts features directly from the facial landmark detector and image backbone. It also includes a vision transformer network for the integration of multi-scale features.

Poster++ is a fast and accurate model, rated highest on the AffectNet database (amongst the models with publicly available code repositories) [33], which uses diverse web-scraped images. This robustness to various subjects and environments makes the AffectNet checkpoint a good base for our fine-tuning.

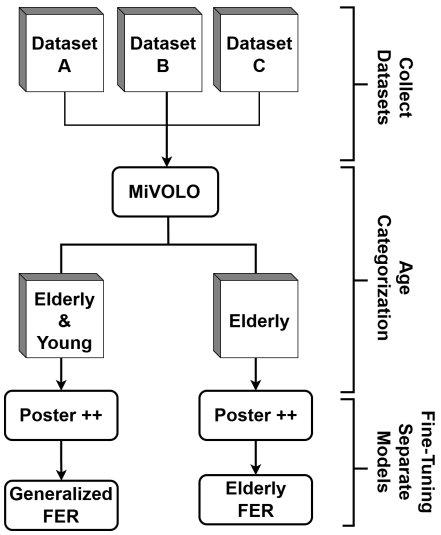

Figure 3. A flowchart depicting the process of fine-tuning emotion recognition models. The original datasets are collected into a dataset, which is then used to fine-tune the model and create a separate model on the left-hand side. Then the data is categorized by age using the MiVOLO model. This step helps to classify data based on different age groups. Finally, the newly classified dataset is used to fine-tune the model that is specific to the elderly.

C. Evaluation

Confusion matrices were used to visualize the evaluation metrics of the FER models, which classify seven distinct emotions. Confusion matrices are a good tool for analyzing the benchmark metrics of FER models [33], particularly in identifying specific emotions that are challenging to recognize and areas where the models may be overfitting or underfitting [34]. It provided insights into the model's performance that can guide further improvements and modifications.

To follow good practice, the labeled data was divided into three distinct subsets: the training set used to train the model, the validation set used to track model parameters and avoid overfitting, and the testing set used to check the model's performance on new data. The different fine-tuned models were evaluated on the same test set for accuracy. The test set is a combination of RAF-DB, Tsinghua, and FACES all classified as having a perceived age above 60, and the data have not been separated before training. This is also why Affectnet was excluded from the test set; the risk of a potential overlap of training and validation data in the way this team divided the datasets compared to the method the Poster++ authors used to divide their datasets [32].

VI. Results

As shown in Figure 4, the model trained with data from all age groups achieved the highest accuracy of 97.63% among the elderly (ages 60+). In contrast, the model trained exclusively on elderly data attained an accuracy of 95.27% when tested on the same dataset. The confusion matrices for the elderly dataset are shown in Figure 5 and Figure 6.

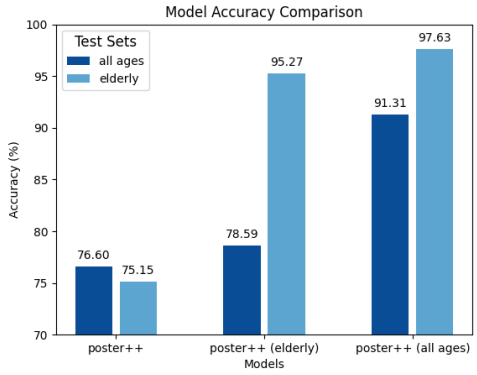

Figure 4. The dataset contains images from RAF-DB, Faces, and Tsinghua datasets. The first model from the left is Poster++, which serves as the base model and is pre-trained on the AffectNet dataset. The second model is built using transfer learning on Poster++ and is fine-tuned with images of people perceived to be aged 60 and above (elderly) from the combined dataset. The third model is also built using transfer learning on Poster++, but it is fine-tuned with images of people of all ages from the same combined datasets. All three models have been benchmarked against the combined datasets, evaluating their performance on both all ages and the elderly age group.

VII. Discussion

As the results show, it seems that a model trained on more data does perform better than an age-specific fine-tuned model. This was also seen when we experimented with another model. When the model was trained using the FACES, Tsinghua, and RAF-DB datasets, which include individuals of all ages, it produced better results when tested exclusively on the elderly

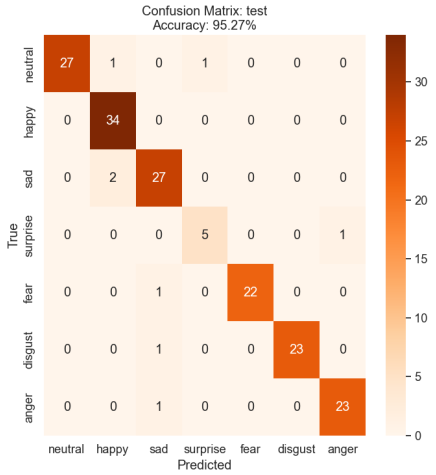

Figure 5. The confusion matrix of Poster++ fine-tuned using the combined datasets that only contain the elderly subjects based on a perceived age of 60 and above. The results show the model's performance when benchmarked on the same dataset.

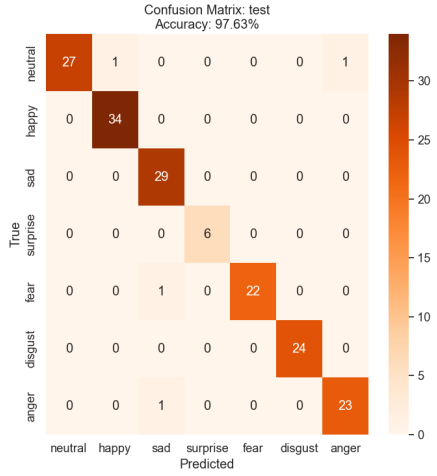

Figure 6. The confusion matrix of Poster++ fine-tuned using the combined datasets, ranging from all age groups. The results show the model's performance when benchmarked on the dataset containing only the elderly subjects.

(ages 60+) from the FACES and Tsinghua test sets. This performance was slightly better compared to when the model was trained solely on the FACES and Tsinghua datasets with individuals of all ages and tested on the same elderly test set. This further strengthens the previously stated assumptions, that the model's performance improved simply by having access to a larger dataset. However, further testing and experimentation are required to strengthen this assumption. Unfortunately we did not have sufficient time to test this hypothesis, and it was not within the scope of our research questions.

Despite the favorable results obtained, they do not necessarily reflect real-world performance. SOTA models are typically evaluated on images that match the quality of the training data, and our results follow this pattern. None of the models have been tested on a universal test set or a comparable standard. In our case using a fixed number of images from each dataset (FACES, Tsinghua, RAF-DB) for the test subsets could perhaps also have been better, since differences in the composition of the two test subsets ("elderly" and "all ages") might have affected the outcome. The all-ages test comprises 12.51% from FACES, 4.66% from Tsinghua, and 82.82% from RAF-DB. The elderly test comprises 65.24% from FACES, 18.45% from Tsinghua, and 16.31% from RAF-DB. Alternatively, future work could involve applying the fine-tuned model in a real-life setting to verify the validity.

The dataset with the most elderly subjects, FACES, lacked images of surprised expressions, leaving the model with less training data for this emotion. Most surprised elderly expressions (66%) came from the Tsinghua dataset, which exclusively features Chinese subjects, creating an imbalance that might cause the model to be less accurate on other demographics. One possible way to address the lack of a dataset for the elderly or to balance out existing datasets is to use generative AI to create synthetic images of elderly subjects, balancing emotional and demographic representation. This approach could improve the model's accuracy across diverse subjects, reducing bias.

It's important to note that while these generated facial features may appear realistic to human observers, their effectiveness in enhancing the model's performance is not guaranteed. What looks visually convincing to humans may not necessarily improve the AI's recognition accuracy. A potentially more reliable solution could be to morph the facial features of pairs of subjects in the dataset to generate images. In an experiment exploring common FER augmentation techniques [35], this method was found to be the most effective, though it was also the most computationally expensive. So this approach is proven to be able to capture accurate details reflective of real-life facial features, which might be transferable in recognizing age-related facial characteristics as well.

VIII. Conclusion

The findings from our study indicate a significant improvement in accuracy, with **RQ1** showing an increase of 20.12% and **RQ2** demonstrating a higher rise of 22.48%. These results lead us to conclude that achieving high accuracy for elderly subjects in facial emotion recognition does not necessarily require an evenly distributed dataset across all age groups. Contrary to initial assumptions, a dataset biased towards middle-aged individuals does not always detract from performance.

Our experiment reveals the necessity of having a sufficient amount of data specifically for the elderly demographic. This conclusion is evident when comparing the results of our models to the baseline performance. Thus, the key factor is not the distribution of age in the dataset but its size, particularly the representation of elderly subjects.

Additionally, the model exhibits good adaptability to various datasets. Even when integrating data of differing quality, the model maintains high performance without becoming overly biased towards the largest subset. This robustness underscores the model's versatility and its potential application across diverse datasets.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the team behind Poster++, who generously made their code publicly available. This contribution significantly propelled our research forward. We also extend our thanks to the authors of Faces, Tsinghua, and Affectnet for providing us with their respective datasets, which were instrumental in accomplishing this research. Special thanks are due to our teachers: **Carl Ahlberg**, a Research Engineer/Technician; **Martin Ekström**, a Senior Lecturer; and **Mikael Ekström**, a Professor in the field of Robotics; for their insights, guidance, and expertise, which greatly enriched our study.

REFERENCES

[1] "Chapter 7. prevalence of dementia at the average age of 81.5 and 87 years." *Acta Psychiatrica Scandinavica*, vol. 79, 1989. [Online]. Available: https://api.semanticscholar.org/CorpusID:221390133

[2] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *sensors*, vol. 18, no. 2, p. 401, 2018.

[3] C. Dalvi, M. Rathod, S. Patil, S. Gite, and K. Kotecha, "A survey of ai-based facial emotion recognition: Features, ml & dl techniques, age-wise datasets and future directions," *IEEE Access*, vol. 9, pp. 165 806–165 840, 2021.

[4] M. Sajjad, F. U. M. Ullah, M. Ullah, G. Christodoulou, F. A. Cheikh, M. Hijji, K. Muhammad, and J. J. Rodrigues, "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandria Engineering Journal*, vol. 68, pp. 817–840, 2023.

[5] P. Ekman, "Universals and cultural differences in facial expressions of emotion." in *Nebraska symposium on motivation*. University of Nebraska Press, 1971.

[6] D. Matsumoto, "More evidence for the universality of a contempt expression," *Motivation and Emotion*, vol. 16, no. 4, pp. 363–368, 1992.

[7] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

[8] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, 2023.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[10] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, "Transfer learning vs. fine-tuning in bilinear cnn for lung nodules classification on ct scans," in *Proceedings of the 2020 3rd International Conference on Artificial Intelligence and Pattern Recognition*, 2020, pp. 99–103.

[11] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: a review," *Artificial Intelligence Review*, vol. 52, pp. 927–948, 2019.

[12] PapersWithCode. Face detection | papers with code. [Online]. Available: https://paperswithcode.com/task/face-detection

[13] A. Clapés, O. Bilici, D. Temirova, E. Avots, G. Anbarjafari, and S. Escalera, "From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation," in *proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 2373–2382.

[14] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 6, pp. 1148–1161, 2014.

[15] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 34–42.

[16] PapersWithCode. Face detection | papers with code. [Online]. Available: https://paperswithcode.com/task/age-estimation

[17] N. Labzour, S. E. Fkihi, S. Benaissa, Y. Zennayi, and O. Bourja, "A survey on facial emotion recognition for the elderly," in *International Conference on Digital Technologies and Applications*. Springer, 2023, pp. 561–575.

[18] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE transactions on affective computing*, vol. 10, no. 1, pp. 18–31, 2019.

[19] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2852–2861.

[20] I. J. Goodfellow, D. Erhan, P. Luc Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, "Challenges in representation learning: A report on three machine learning contests," *Neural networks*, vol. 64, pp. 59–63, 2015.

[21] N. C. Ebner, M. Riediger, and U. Lindenberger, "Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation," *Behavior research methods*, vol. 42, pp. 351–362, 2010.

[22] K. Wang, Z. Zhu, S. Wang, X. Sun, and L. Li, "A database for emotional interactions of the elderly," in *2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS)*, 2016, pp. 1–6.

[23] A. Caroppo, A. Leone, and P. Siciliano, "Facial expression recognition in older adults using deep machine learning." in *AI* AAL@ AI* IA*, 2017, pp. 30–43.

[24] T. Yang, Z. Yang, G. Xu, D. Gao, Z. Zhang, H. Wang, S. Liu, L. Han, Z. Zhu, Y. Tian *et al.*, "Tsinghua facial expression database–a database of facial expressions in chinese young and older women and men: Development and validation," *PloS one*, vol. 15, no. 4, p. e0231304, 2020.

[25] Z. Fei, E. Yang, L. Yu, X. Li, H. Zhou, and W. Zhou, "A novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly," *Neurocomputing (Amsterdam)*, vol. 468, pp. 306–316, 2022.

[26] K. Ma, X. Wang, X. Yang, M. Zhang, J. M. Girard, and L.-P. Morency, "Elderreact: a multimodal dataset for recognizing emotional response in aging adults," in *2019 international conference on multimodal interaction*, 2019, pp. 349–357.

[27] T. Khajontantichaikun, S. Jaiyen, S. Yamsaengsung, P. Mongkolnam, and T. Chirapornchai, "Facial emotion detection for thai elderly people using yolov7," in *2023 15th International Conference on Knowledge and Smart Technology (KST)*. IEEE, 2023, pp. 1–4.

[28] E. Kim, D. Bryant, D. Srikanth, and A. Howard, "Age bias in emotion detection: An analysis of facial emotion recognition performance on young, middle-aged, and older adults," in *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, ser. AIES '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 638–644. [Online]. Available: https://doi.org/10.1145/3461702.3462609

[29] U. Hess, R. B. Adams, A. Simard, M. T. Stevenson, and R. E. Kleck, "Smiling and sad wrinkles: Age-related changes in the face and the perception of emotions and intentions," *Journal of Experimental Social Psychology*, vol. 48, no. 6, pp. 1377–1380, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022103112001126

[30] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[31] M. Kuprashevich and I. Tolstykh, "Mivolo: Multi-input transformer for age and gender estimation," in *International Conference on Analysis of Images, Social Networks and Texts*. Springer, 2023, pp. 212–226.

[32] J. Mao, R. Xu, X. Yin, Y. Chang, B. Nie, and A. Huang, "Poster++: A simpler and stronger facial expression recognition network," *arXiv preprint arXiv:2301.12149*, 2023.

[33] PapersWithCode. Face detection | papers with code. [Online]. Available: https://paperswithcode.com/task/facial-expression-recognition

[34] A. Kulkarni, D. Chong, and F. A. Batarseh, "Foundations of data imbalance and solutions for a data democracy," in *Data democracy*. Elsevier, 2020, pp. 83–106.

[35] S. Porcu, A. Floris, and L. Atzori, "Evaluation of data augmentation techniques for facial expression recognition systems," *Electronics*, vol. 9, no. 11, p. 1892, 2020.