

Article

Facial Emotion Recognition Analysis Based on Age-Biased Data

Hyungjoo Park, Youngha Shin, Kyu Song, Channyeong Yun and Dongyoung Jang *

Korea Electronics-Machinery Convergence Technology Institute, Seoul 01811, Korea

* Correspondence: dyjang@kemcti.re.kr

Abstract: This paper aims to analyze the importance of age-biased data in recognizing six emotions using facial expressions. For this purpose, a custom dataset (adults, kids, mixed) was constructed using images that separated the existing datasets (FER2013 and MMA FACILE EXPRESSION) into adults (≥ 14) and kids (≤ 13). The convolutional Neural Networks (CNN) algorithm was used to calculate emotion recognition accuracy. Additionally, this study investigated the effect of the characteristics of CNN architecture on emotion recognition accuracy. Based on the variables of Accuracy and FLOP, three types of CNN architectures (MobileNet-V2, SE-ResNeXt50 (32×4 d), and ResNeXt-101 (64×4 d)) were adopted. As for the experimental result, SE-ResNeXt50 (32×4 d) showed the highest accuracy at 79.42%, and the model that learned by age obtained 22.24% higher accuracy than the model that did not learn by age. In the results, the difference in expression between adults and kids was greatest for fear and neutral emotions. This study presented valuable results on age-biased learning data and algorithm type effect on emotion recognition accuracy.

Keywords: facial emotion recognition; deep learning; convolution neural networks; age bias

Citation: Park, H.; Shin, Y.; Song, K.; Yun, C.; Jang, D. Facial Emotion Recognition Analysis Based on Age-Biased Data. *Appl. Sci.* **2022**, *12*, 7992. <https://doi.org/10.3390/app12167992>

Academic Editor: Christian W. Dawson

Received: 5 July 2022

Accepted: 8 August 2022

Published: 10 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Deep learning is used in many applications such as computer vision, pattern recognition, robotics, and autonomous driving [1]. In particular, in the field of computer vision, deep learning-based emotion classification technology is attracting attention not only from researchers but also from industries and the public. It is used in a variety of fields, from special environments such as interviews and counseling centers to everyday situations such as schools and supermarkets [2]. With the recent development of deep learning technology, many developments have been made in object classification technology, and performance has been improved. However, emotion classification is still a challenging task. Convolutional neural networks (CNNs) perform face normalization, facial expressions, and emotional classification using real images as their main functions and are frequently adopted and used in computer vision applications [3–6]. The accuracy of the CNN-based emotion classification system has been improved through pre-or post-processing [2,5,7] and the development of new algorithms in the architecture. Emotional recognition accuracy is fundamentally determined by the type and number of training data. However, existing studies [1–3] on facial expression emotion detection using deep learning techniques focus only on algorithm development. Therefore, in this study, we considered data pre-processing, particularly data modification. A few studies for improving the accuracy of emotion classification by modifying and changing existing learning data are described below.

- LoBue et al. conducted a study of emotional classification consisting only of kids under the age of 13. They organized kids into groups of various races and studied whether the kids's race and ethnicity affected the accuracy in identifying emotional expressions. The accuracy of identifying kids's emotions was, on average, 66%. The average of accurate responses for each of the seven categories of facial expressions did not show significant differences or interactions due to race and ethnicity [8].

- Gonçalves et al. conducted an emotion classification study based on a dataset that separated adults and the elderly from the age of 55. The results of the study showed that compared to the elderly, adults had higher accuracy in identifying emotions such as anger, sadness, fear, surprise, and happiness. In this result, it is said that aging and the classification of emotions are related and can affect emotional identification [9].
- Kim et al. conducted a study on emotion classification by constructing datasets of adults (19–31), middle-aged (39–55), and elderly (69–80) participants. The accuracy, using Amazon Recognition, was 89% for adults, 85% for middle-aged, and 68% for the elderly. The emotion classification accuracy between adults and middle-aged people showed a difference of less than 4%, but it was confirmed that the accuracy of emotion classification for the elderly was about 17% lower than that of adults and middle-aged people [10].
- Sullivan et al. studied the classification of emotions of the elderly, and they determined that it was difficult to classify emotions of the elderly when they were consistently expressing anger, fear, and sadness. Therefore, differences in emotional expression methods were studied in the elderly and adults [11].
- A study by Thomas et al. found that adults expressed fear and anger more accurately than kids. Adults confirmed higher accuracy in the classification of fear and anger emotions than kids [12].
- Lida et al. studied real-time emotion recognition using CNN MobileNetv2. In this study, the background of the image was learned so that it did not affect the emotional classification. CK+, JAFFE, and FER2013 were used as learning data. The accuracy of emotional classification was CK+ 86.86%, JAFFE 87.8%, and FER2013 66.28% [13].
- A study by Ishika et al. used pre-processing steps and several CNN topologies to improve accuracy and training time. MMA FACILE EXPRESSION was used as the learning data, and 35,000 images were used. The accuracy of emotion classification was 69.2% [14].
- Yahia et al. used a face-sensitive convolutional neural network (FS-CNN) for human emotion recognition. The proposed FS-CNN is used to detect faces in high-scale images and analyze landmarks on faces to predict facial expressions for emotion recognition. The proposed FS-CNN was trained and evaluated on the UMD Faces dataset. The experiment resulted in an emotion classification accuracy of 94% [15].

In related studies, although differences according to race and gender bias are receiving much attention, age-related biases are not considered. However, emotional expressions are also developed when kids become adults [16], and there are previous studies showing differences in the expression of emotions between kids and adults. In order to perform emotion recognition, it is necessary to construct an emotion recognition system divided by age that can be used not only in a group consisting of only adults but also in places such as daycare centers and elderly homes. Additionally, in previous studies [13,14], FER2013 obtained 66.28% and MMA FACILE EXPRESSION 69.2%, and the accuracy was improved by increasing the resolution of the input image [15]. However, in this study, an accuracy of 81.57% was obtained by training the same dataset through only the data separation task. In addition, data processing speed was fast because high-resolution data was not required. The CNN algorithm was used for emotion recognition using images. An emotion recognition system for each age was constructed by classifying emotion recognition learning data by age. Accuracy, recall, precision, and *F1* score were used as outcome metrics. The remainder of this article is organized as follows. Section 2 presents the experimental method adopted to carry out the study. The results obtained and their discussion is presented in Section 3. The conclusions end the paper in Section 4.

2. Experimental Method

Table 1 shows the experimental parameters. All experiments were conducted using a workstation equipped with an Intel i7 CPU with 16 GB of RAM (Intel, Santa Clara, CA, USA) and an Nvidia GTX3050 GPU with 8 GB of memory (Nvidia, Santa Clara, CA, USA).

The CNN used in this study was developed using the PyTorch (FAIR, Menlo Park, CA, USA) deep learning framework, including the CUDA acceleration library and cuDNN library support. The images used to build the custom dataset were sorted from the OpenCV library. A custom dataset was created by classifying adults and kids from the existing FER2013 [17] and MMA FACILE EXPRESSION [18] datasets. As shown in Figure 1, FER2013 consists of a grayscale image of 48×48 pixels. MMA FACILE EXPRESSION consists of color images with the same number of pixels. Both datasets included six identical emotion classes (anger, fear, happy, neutral, sad, and surprised). The custom datasets were divided into 3 categories, consisting of 50,000 images each for adults only, kids only, and mixed. The training process was performed using 45,000 images out of 50,000 images in each dataset, and the remaining 5000 images were used for testing.

Table 1. Experimental variables.

	Dataset	Adult Type, Kid Type, Adult + Kid Type
Ages		Adult type (≥ 14), Kid type (≤ 13)
Number of images in the dataset		50,000 (Train: 45,000, Test: 5000)
Type of emotion		Angry, Fear, Happy, Neutral, Sad, Surprise

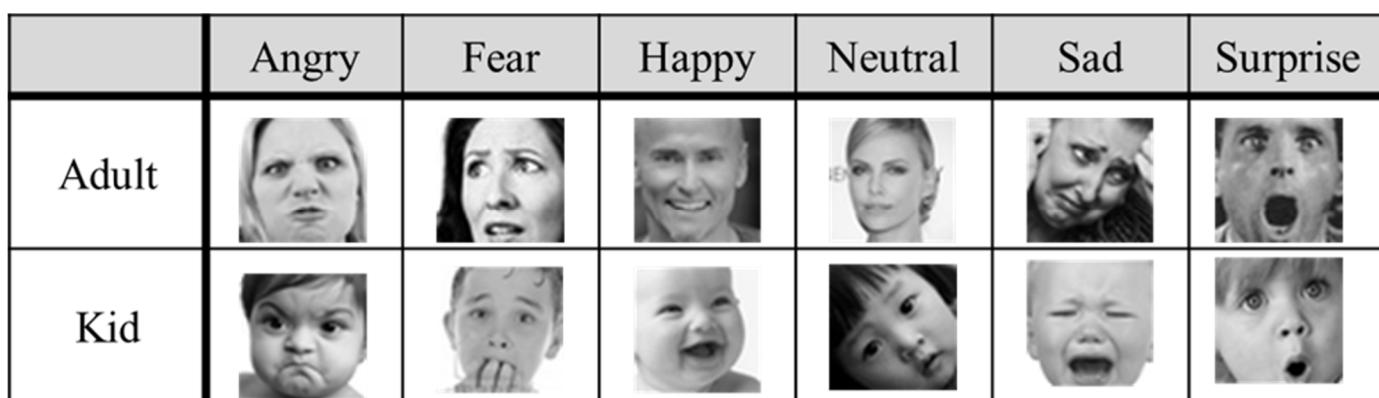


Figure 1. Example of FER2013.

The experimental variables affect the analysis sequence for the accuracy of emotion recognition, as shown in Figure 2. In this study, experiments were conducted using three types of training data (adults, kids, and mixed) and two types of test data (adults and kids). The three types of architecture were trained to produce six age-specific outcomes (A/A, K/K), opposite-age outcomes (A/K, K/A), and mixed-age outcomes (A + K/A, A + K/K). The architecture that was selected was MobileNet-V2 [19], SE-ResNeXt50 (32×4 d) [20,21], and ResNeXt101 (64×4 d) [21] to compare the accuracy of emotions. This architecture was adjusted based on the accuracy and FLOP variables of the object recognition benchmark test [22] in Figure 3.

In this study, pre-processing was performed to apply the custom dataset to the parameters required according to the architecture type in Table 2. The data's pre-processing parameters are as follows: all image sizes of the custom dataset were converted from 48×48 to 224×224 to allow them to be input into the architecture [23–25]. Since the FER2013 data is grayscale, all images were converted to grayscale. Afterward, the image data's tensor values were normalized.

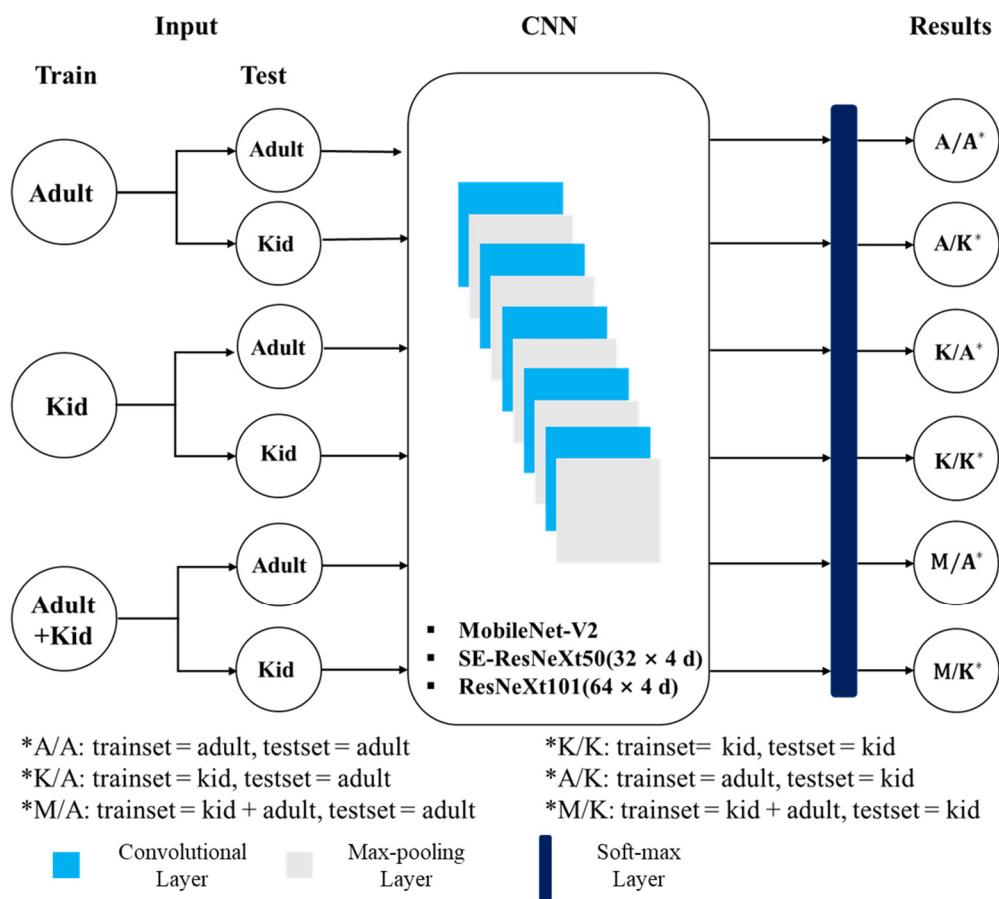


Figure 2. Schematic diagram of the experimental sequence.

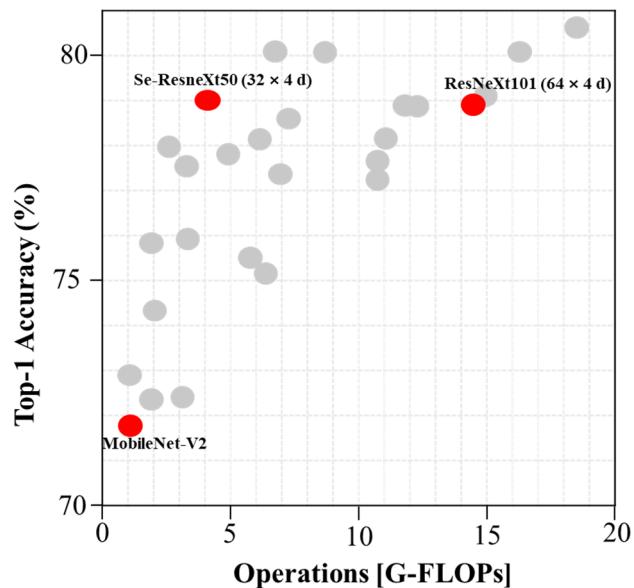


Figure 3. Results of the architecture used in the benchmarks.

Table 2. Experimental parameters of the architectures.

	Parameter	MobileNet-V2	SE-ResNeXt50 (32 × 4 d)	ResNeXt101 (64 × 4 d)
Pre-Processing	Input dimension (pixel) Batch size	3 × 224 × 224 32	3 × 224 × 224 32	3 × 224 × 224 12
Learning	Optimization algorithm	RAdam	RAdam	RAdam
	Learning rate (s)	1×10^3	1×10^3	1×10^3
	Learning scheduler	Custom	Custom	Custom
	Epochs	50	50	50
	Shuffle	True	True	True
	Elapsed time (min)	235	955	1563

The learning parameters of each architecture were set as follows. Considering the memory performance of GPU, the batch size was set to 32 for MobileNet-V2, 32 for SE-ResNeXt50 (32 × 4 d), and 12 for ResNeXt101 (64 × 4 d). The loss function used cross-entropy loss [23] and set weights to solve the problem of unbalanced data. RAdam (Microsoft Research, Redmond, WA, USA) [24] was used as the optimizer, and the initial learning rate was set to 1×10^3 s. The learning scheduler created a custom scheduler that was slightly modified from the existing CosineAnnealing WarmRestarts (University of Freiburg, Freiburg, Germany) scheduler [25]. Each architecture had different elapsed times due to having different sizes and types of parameters such as instance activation functions, pooling layers, fully connected layers, regularization layers, and a number of input and output hidden layers.

For each architecture, both adults and kids obtained accuracy, recall, precision, and $F1$ score by learning the model and accuracy in class. Accuracy is usually a metric that describes how a model performs in any class. This was a reasonable measure because the weights were adjusted so that the importance of each grade was the same [26].

$$\text{Accuracy} (\%) = \frac{TP + TN}{TP + FP + FN + TN} * 100 \quad (1)$$

$$\text{Recall} (\%) = \frac{TP}{TP + FN} * 100 \quad (2)$$

$$\text{Precision} (\%) = \frac{TP}{TP + FP} * 100 \quad (3)$$

$$F1 (\%) = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} * 100 \quad (4)$$

TP = true positive, TN = true negative, FP = false positive, FN = false negative.

3. Results and Discussion

3.1. Comparison of the Average and Emotional Evaluation Metrics of the Architecture by Using the Learning Model

Figure 4 shows that there was a difference in the accuracy between the learning model by age and other models in the experimental results. In the results of the age-specific, opposite-age, and mixed-age models of all architectures, the age-specific learning model showed the highest accuracy. The average accuracy of the learning model by age was 71.06%, 79.42%, and 72.27% for MobileNet-V2, SE-ResNeXt50 (32 × 4 d), and ResNeXt101 (64 × 4 d), respectively. Table 3 shows other metrics: recall, precision, and $F1$ score. Those metrics showed the same trend as accuracy. This showed the same trend as the benchmark results in Figure 3. The accuracy of the opposite-age model was the lowest, and the mixed-age model was also lower than that of the age-specific learning model. From this result, it can be seen that it is necessary to separate the learning data in order to identify the emotions of a specific age.

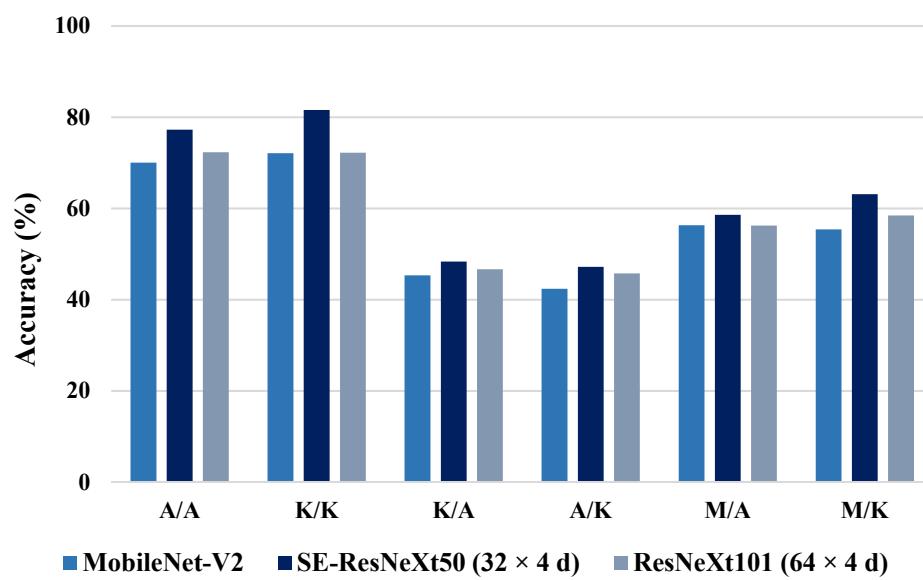


Figure 4. Emotion recognition accuracy by the architecture type.

Table 3. Deviation between the three age-specific, opposite-age, and mixed-age architectures.

Dataset	MobileNet-V2	SE-ResNeXt50 (32x4d)	ResNeXt101 (64x4d)
Accuracy (%)	A/A	70.03	77.26
	A/K	42.36	47.17
	K/A	45.31	48.33
	K/K	72.09	81.57
	M/A	56.30	58.96
	M/K	55.41	63.13
Recall (%)	A/A	72.20	75.25
	A/K	38.43	45.09
	K/A	41.78	47.07
	K/K	69.81	78.61
	M/A	56.37	62.80
	M/K	51.47	55.36
Precision (%)	A/A	72.62	75.48
	A/K	38.50	46.96
	K/A	41.78	47.63
	K/K	69.81	79.97
	M/A	56.37	64.81
	M/K	51.47	58.30
F1 score (%)	A/A	72.33	74.78
	A/K	37.97	44.33
	K/A	42.19	51.23
	K/K	68.04	79.07
	M/A	56.23	63.35
	M/K	49.97	55.93

In Figures 5–7, the accuracy results subjected to the types of emotion and architecture were presented. The accuracy of the age-specific learning model was higher than both the opposite-age models and the mixed-age models in all classes. The detailed values are specified in Table 4. This shows that it is a reasonable measure for the accuracy derivation performance because the recognition of all emotion accuracy for all emotions had the same trend. Since the accuracy of emotion recognition is different for each architecture, it means that it is necessary to develop an architecture dedicated to emotion recognition.

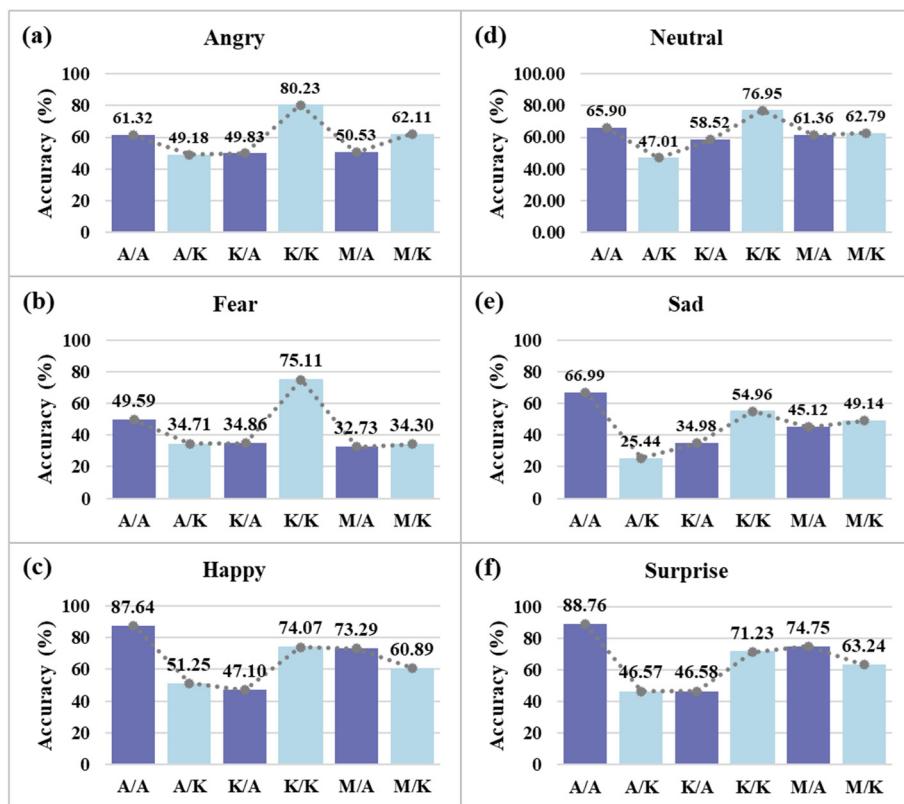


Figure 5. MobileNet-V2's accuracy according to different emotions: (a) angry, (b) fear, (c) happy, (d) neutral, (e) sad, (f) surprise.

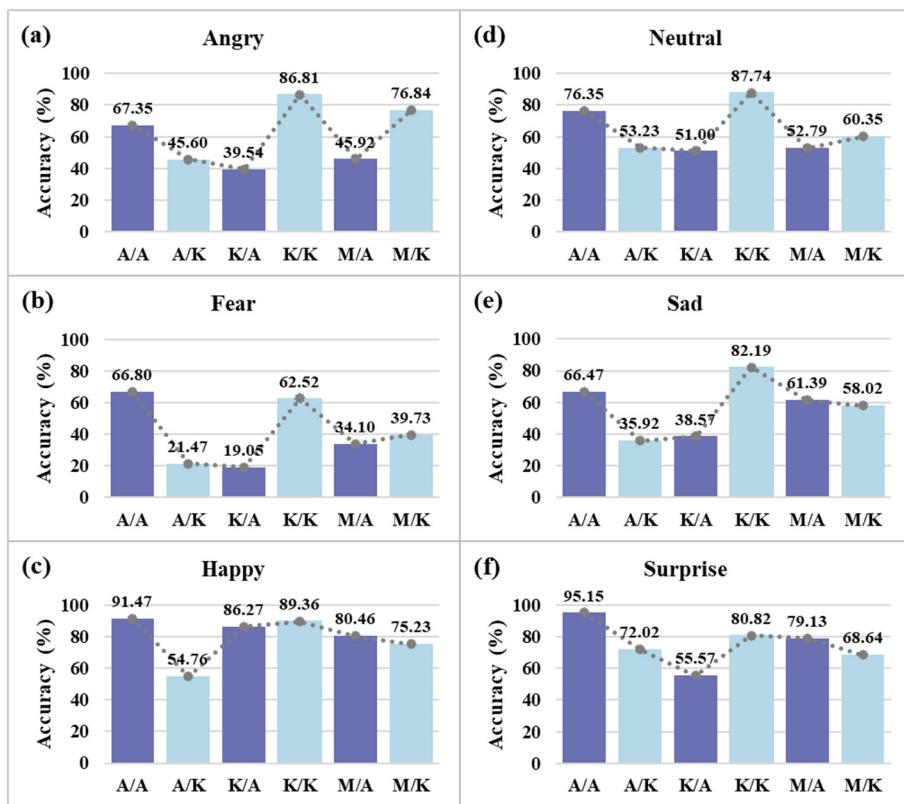


Figure 6. SE-ResNeXt50 (32 × 4 d)'s accuracy according to different emotions: (a) angry, (b) fear, (c) happy, (d) neutral, (e) sad, (f) surprise.

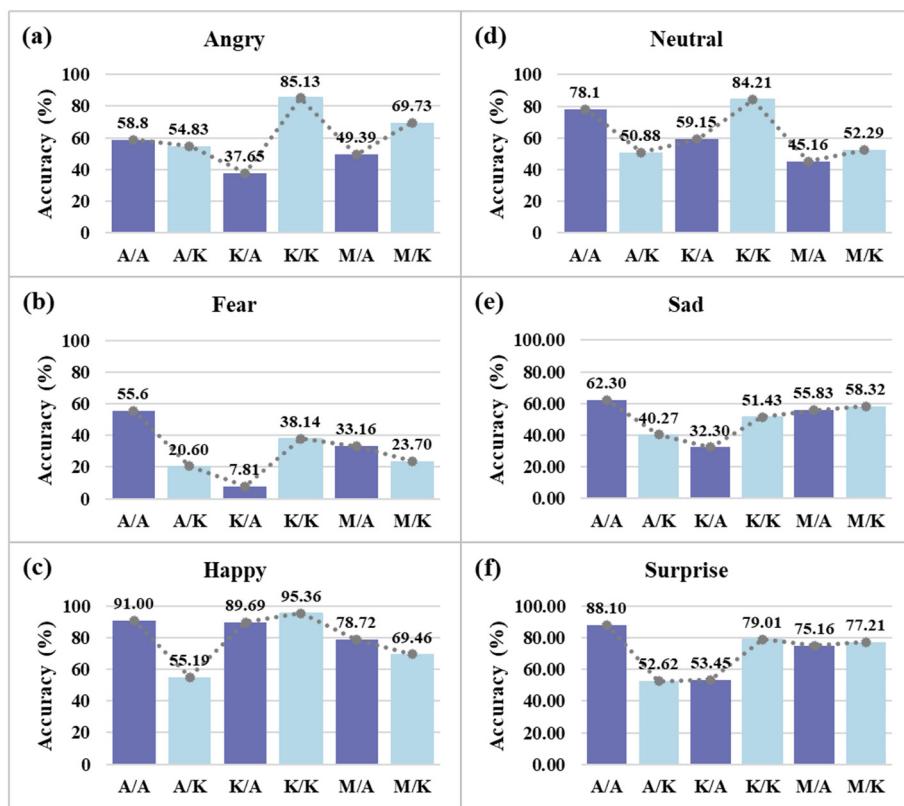


Figure 7. ResNeXt101 (64 × 4 d)'s accuracy according to different emotions: (a) angry, (b) fear, (c) happy, (d) neutral, (e) sad, (f) surprise.

Table 4. Recognition accuracy of each emotion for all architecture.

Architecture	Dataset	Accuracy (%)						
		Angry	Fear	Happy	Neutral	Sad	Surprise	AVG
MobileNet-V2	A/A	81.17	42.71	77.83	72.30	55.43	90.74	70.03
	A/K	49.18	34.71	51.25	47.01	25.44	46.57	42.36
	K/A	49.83	34.86	47.10	58.52	34.98	46.58	45.31
	K/K	80.23	75.11	74.07	76.95	54.96	71.23	72.09
	M/A	50.53	32.73	73.29	61.36	45.12	74.75	56.30
	M/K	62.11	34.30	60.89	62.79	49.14	63.24	55.41
SE-ResNeXt50 (32 × 4 d)	A/A	67.35	66.80	91.47	76.35	66.47	95.15	77.26
	A/K	45.60	21.47	54.76	53.23	35.92	72.02	47.17
	K/A	44.18	10.44	81.63	42.43	45.61	65.70	48.33
	K/K	86.81	62.52	89.36	87.74	82.19	80.82	81.57
	M/A	72.02	52.18	59.73	54.11	47.58	68.15	58.96
	M/K	76.84	39.73	75.23	60.35	58.02	68.64	63.13
ResNeXt101 (64 × 4 d)	A/A	58.80	55.60	91.00	78.10	62.30	88.10	72.32
	A/K	54.83	20.60	55.19	50.88	40.27	52.62	45.73
	K/A	37.65	7.81	89.69	59.15	32.30	53.45	46.67
	K/K	85.13	38.14	95.36	84.21	51.43	79.01	72.21
	M/A	49.39	33.16	78.72	45.16	55.83	75.16	56.24
	M/K	69.73	23.70	69.46	52.29	58.32	77.21	58.45

3.2. Comparison of the Emotional Accuracy in the Age-Specific Learning Model

Figure 8 represents the recognition accuracy for each emotion using SE-ResNeXt50 (32 × 4 d). The average emotion recognition accuracy of SE-ResNeXt50 (32 × 4 d) was 77.26 for adults and 81.57% for kids. In the results for the emotion recognition accuracy for adults, angry (67.35%), fear (66.80%), neutral (76.35%), and sad (66.47%) had lower than

average accuracy. The above results are similar to the results of a previous study by Sullivan et al. [11]. On the other hand, in the case of kids, fear (62.62%) and surprise (80.82%) were lower than the average. This shows that adults had difficulties in recognizing anger, fear, and sadness when these emotions were expressed consistently. On the other hand, kids, in contrast to adults, did not have low accuracy when recognizing angry emotions but obtained low accuracy with fear and surprise emotions. In order to solve this problem, this study showed the need to not only separate data by age but also to segment specific emotions in a specific age group or to contribute more to data processing.

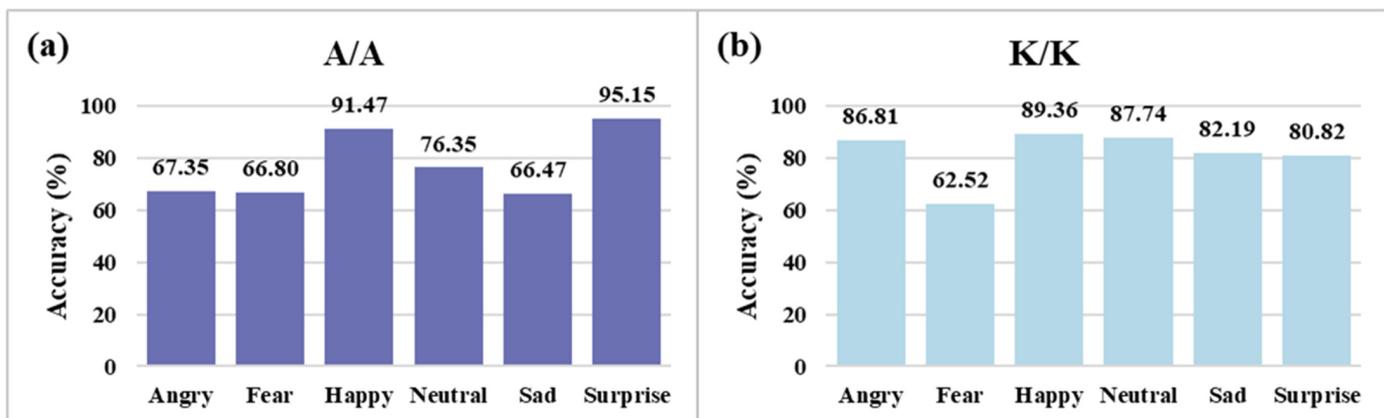


Figure 8. Age-specific accuracy of SE-ResNeXt50 (32×4 d) according to different emotions: (a) A/A model and (b) K/K model.

Table 5 summarizes the recognition accuracy of existing reports of the FER2013. The previous best model for single-network accuracy is 72.7% [26]. In this work, our model achieved an accuracy of 79.42%. This result shows that the accuracy of the separate dataset of adults and kids is higher than that of the existing dataset. The reason is that there is a difference in facial expressions in the emotional expression of adults and kids, so the existing dataset learns by classifying images of different facial expressions as the same emotion.

Table 5. Average accuracy of the Fer2013 public benchmark.

Method	Accuracy (%)
GoogleNet [27]	65.20
MobileNet-V2 [13]	66.28
ARM (ResNet-18) [28]	71.38
Inception [29]	71.60
ResNet [29]	72.40
VGG [29]	72.70
Our model (SE-ResNeXt50 (32×4 d))	79.42

3.3. Comparison of Emotional Accuracy Deviation between the Learning Model

Figure 9 shows the accuracy deviation between the learning models for each age and non-age models of SE-ResNeXt50 (32×4 d) for adults and kids. There were accuracy deviations in all classes, and the age-specific model showed higher accuracy in all classes than the rest of the models. The difference between the age-specific and opposite-age models was greatest in A/A-K/A with fear (47.75%) and sad (46.27%) with K/K-A/K. The difference between the age-specific model and the mixed model was the greatest, with fear (32.70%) in A/A-M/A and neutral (27.39%) in K/K-M/K. In Section 3.2, it was found that adults have difficulties in recognizing fear emotions. In the deviation between adults and kids, there was also a large deviation in the expression of fear emotions. This result occurs because the facial expressions of adults and kids are different emotional expressions [12]. Based on the experimental results, it was found that learning to separate data by age can

solve the problem of the expression differences in expression between adults and kids. As a result, when data processing was performed based on adults, kids, or other anthropological information in the emotional recognition deep learning mechanism, results that helped improve performance were obtained.

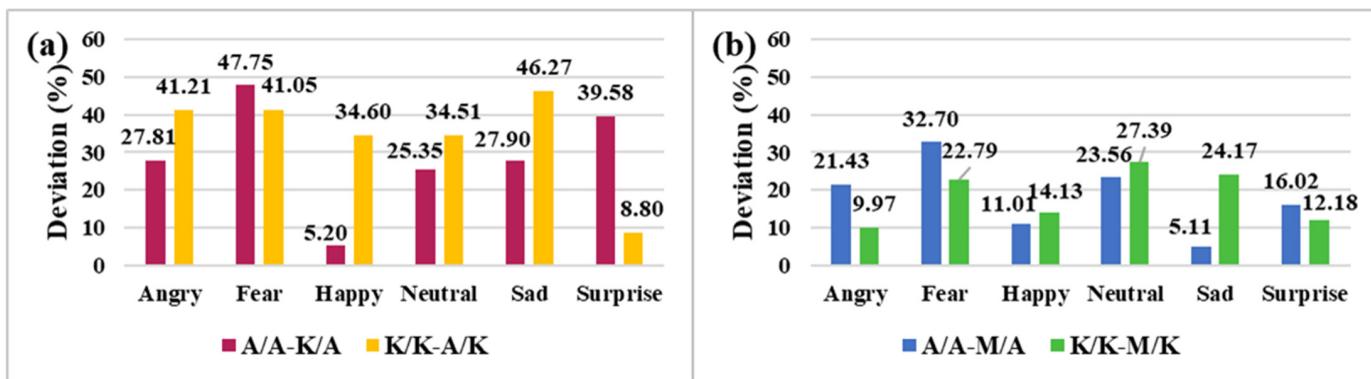


Figure 9. Accuracy deviation between age-specific and nonage-specific of SE-ResNeXt50 (32×4 d): (a) deviation between the age-specific and opposite-age groups, (b) deviation between the age-specific and mixed-age groups.

Table 6 shows that using data separation can lead to an improvement in accuracy (2.981% for MobileNet-V2, 2.020% for SE-ResNeXt50 (32×4 d), and 1.384% for ResNeXt101 (64×4 d)). After data separation, accuracy has achieved 11.36% for the MobileNet-V2, 17.13% for SE-ResNeXt50 (32×4 d), and 13.94% for ResNeXt101 (64×4 d) in FER2013 datasets can be obtained.

Table 6. Comparison of our dataset and the existing dataset of each architecture.

Architecture	MobileNet-V2				SE-ResNeXt50 (32×4 d)				ResNeXt101 (64×4 d)	
	Dataset	Age-Specific	FER2013	MMA	Age-Specific	FER2013	MMA	Age-Specific	FER2013	MMA
Accuracy (%)	71.06	59.70	54.65	79.42	62.29	62.87	72.27	58.33	60.76	
Recall (%)	71.00	58.19	54.40	76.93	59.86	52.66	66.31	56.54	49.59	
Precision (%)	70.84	56.88	47.92	77.73	61.02	57.41	67.69	55.12	56.14	
F1 score (%)	70.19	57.25	48.86	76.93	59.43	53.98	66.03	55.38	51.08	

4. Conclusions

In this study, when recognizing facial expressions using the CNN algorithm, an accuracy improvement system based on solving the age bias problem of data was proposed. We provided a performance comparison of three CNN architectures in emotional recognition. Using the adopted architecture, we compared the performance of the age-specific learning model and the non-age-specific learning model. In the age-specific learning model, the recognition performance of each emotion of adults and kids was compared. In addition, the emotional deviations of adults and kids were compared. The framework proposed in the results of this study was able to obtain higher accuracy than the existing classifier in FER2013 and MMA FACIAL EXPRESSION. Although some of the previous studies outperformed our model in accuracy, the method had a much higher elapsed time. Based on the results of this study, the following conclusions can be drawn.

1. Based on the CNN algorithm, emotion recognition accuracy was derived using three types of architecture, and the accuracy of the architecture was verified through comparison with the object recognition benchmark chart.
2. When the data were trained by age, the average accuracy was 22% higher than that of the non-age-specific learning model.

3. In the case of the adult learning model, it was difficult to recognize the emotions of angry, fear, and sad. In the case of the kid learning model, only the fear emotion was derived with low accuracy.
4. The age-specific learning model and the mixed learning model also achieved up to 32.70% improvement in the recognition accuracy of the fear emotion.

The above results prove the importance of the age bias problem in emotional recognition. They also show the need for pre-processing the learning model according to age and emotion. This study provides comparative data for emotion and age to researchers in the field of emotion recognition and proposes insight into the age bias problem experienced in emotion recognition learning models. In our future work, data processing or emotion recognition architecture will be developed to improve the accuracy of specific emotions for which there is low accuracy by age.

Author Contributions: Conceptualization, H.P. and Y.S.; methodology, H.P., Y.S. and K.S.; software, H.P.; validation, D.J., K.S. and C.Y.; formal analysis, H.P.; investigation, D.J.; resources, D.J.; data curation, H.P.; writing—original draft preparation, H.P. and Y.S.; writing—review and editing, D.J., K.S. and C.Y.; visualization, H.P. and Y.S.; supervision, D.J.; project administration, D.J.; funding acquisition, D.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study is a research project conducted with the support of the Ministry of SMEs and Startups with funds from the Ministry of Trade, Industry, and Energy (No. S2640869).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Authors appreciate the kind support from the Korea Electronics-Machinery Convergence Technology Institute (KEMCTI) and Ministry of SMEs and Startups with funds from the Ministry of Trade, Industry, and Energy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jayawickrama, N.; Ojala, R.; Pirhonen, J.; Kivekäs, K.; Tammi, K. Classification of Trash and Valuables with Machine Vision in Shared Cars. *Appl. Sci.* **2022**, *12*, 5695. [[CrossRef](#)]
2. Kim, J.C.; Kim, M.H.; Suh, H.E.; Naseem, M.T.; Lee, C.S. Hybrid Approach for Facial Expression Recognition Using Convolutional Neural Networks and SVM. *Appl. Sci.* **2022**, *12*, 5493. [[CrossRef](#)]
3. Jahangir, R.; Teh, Y.W.; Hanif, F.; Mujtaba, G. Deep learning approaches for speech emotion recognition: State of the art and research challenges. *Multimed. Tools Appl.* **2021**, *80*, 23745–23812. [[CrossRef](#)]
4. Le, D.S.; Phan, H.H.; Hung, H.H.; Tran, V.A.; Nguyen, T.H.; Nguyen, D.Q. KFSENet: A Key Frame-Based Skeleton Feature Estimation and Action Recognition Network for Improved Robot Vision with Face and Emotion Recognition. *Appl. Sci.* **2022**, *12*, 5455. [[CrossRef](#)]
5. El-Hasnony, I.M.; Elzeki, O.M.; Alshehri, A.; Salem, H. Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction. *Sensors* **2022**, *22*, 1184. [[CrossRef](#)]
6. ElAraby, M.E.; Elzeki, O.M.; Shams, M.Y.; Mahmoud, A.; Salem, H. A novel Gray-Scale spatial exploitation learning Net for COVID-19 by crawling Internet resources. *Biomed. Signal Process. Control* **2022**, *73*, 103441. [[CrossRef](#)]
7. Wang, Z.; Ho, S.B.; Cambria, E. A review of emotion sensing: Categorization models and algorithms. *Multimed. Tools Appl.* **2020**, *79*, 35553–35582. [[CrossRef](#)]
8. LoBue, V.; Thrasher, C. The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Front. Psychol.* **2015**, *5*, 1532. [[CrossRef](#)]
9. Gonçalves, A.R.; Fernandes, C.; Pasion, R.; Ferreira-Santos, F.; Barbosa, F.; Marques-Teixeira, J. Effects of age on the identification of emotions in facial expressions: A meta-analysis. *PeerJ* **2018**, *6*, e5278. [[CrossRef](#)]
10. Kim, E.; Bryant, D.A.; Srikanth, D.; Howard, A. 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, Virtual, 19–21 May 2021; pp. 638–644.
11. Sullivan, S.; Ruffman, T.; Hutton, S.B. Age differences in emotion recognition skills and the visual scanning of emotion faces. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* **2007**, *62*, P53–P60. [[CrossRef](#)]

12. Thomas, L.A.; De Bellis, M.D.; Graham, R.; LaBar, K.S. Development of emotional facial recognition in late childhood and adolescence. *Dev. Sci.* **2007**, *10*, 547–558. [[CrossRef](#)] [[PubMed](#)]
13. Hu, L.; Ge, Q. Automatic facial expression recognition based on MobileNetV2 in Real-time. *J. Phys. Conf. Ser.* **2020**, *1549*, 2. [[CrossRef](#)]
14. Agrawal, I.; Kumar, A.; Swathi, D.; Yashwanthi, V.; Hegde, R. Emotion Recognition from Facial Expression using CNN. In Proceedings of the 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC), Bangalore, India, 30 September–2 October 2021; pp. 1–6.
15. Said, Y.; Barr, M. Human emotion recognition based on facial expressions via deep learning on high-resolution images. *Multimed Tools Appl.* **2021**, *80*, 25241–25253. [[CrossRef](#)]
16. Nook, E.C.; Somerville, L.H. Emotion concept development from childhood to adulthood. In *Emotion in the Mind and Body*; Neta, M., Haas, I., Eds.; Nebraska Symposium on Motivation; Springer: Cham, Switzerland, 2019; Volume 66, pp. 11–41.
17. Goodfellow, I.J.; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 117–124.
18. MMA FACILE EXPRESSION | Kaggle. Available online: <https://www.kaggle.com/mahmoudima/mma-facial-expression> (accessed on 22 February 2022).
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–21 June 2018; pp. 4510–4520.
20. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–21 June 2018; pp. 7132–7141.
21. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
22. Bianco, S.; Cadene, R.; Celona, L.; Napoletano, P. Benchmark analysis of representative deep neural network architectures. *IEEE Access* **2018**, *6*, 64270–64277. [[CrossRef](#)]
23. Zhang, Z.; Sabuncu, M. Generalized cross-entropy loss for training deep neural networks with noisy labels. *Adv. Neural Inf. Process. Syst.* **2018**, *31*.
24. Liu, L.; Jiang, H.; He, P.; Chen, W.; Liu, X.; Gao, J.; Han, J. On the variance of the adaptive learning rate and beyond. *arXiv* **2019**, arXiv:1908.03265.
25. Loshchilov, I.; Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
26. Daskalaki, S.; Kopanas, I.; Avouris, N. Evaluation of classifiers for an uneven class distribution problem. *Appl. Artif. Intell.* **2006**, *20*, 381–417. [[CrossRef](#)]
27. Giannopoulos, P.; Perikos, I.; Hatzilygeroudis, I. Deep learning approaches for facial emotion recognition: A case study on FER-2013. *Smart Innov. Syst. Tech.* **2018**, *85*, 1–16.
28. Shi, J.; Zhu, S.; Liang, Z. Learning to Amend Facial Expression Representation via De-albino and Affinity. *arXiv* **2021**, arXiv:2103.1018.
29. Pramerdorfer, C.; Kampel, M. Facial expression recognition using convolutional neural networks: State of the art. *arXiv* **2016**, arXiv:1612.02903.