İSTANBUL UNIVERSITY CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2024; 24(3): 818-825

Facial Emotion Recognition Using Residual Neural Networks

Serap Kırbız

Department of Electrical and Electronics Engineering, MEF University, Faculty of Engineering, İstanbul, Türkiye

Cite this article as: S. Kırbız, "Facial emotion recognition using residual neural networks," *Electrica*, 24(3), 818-825, 2024.

ABSTRACT

Facial emotion recognition (FER) has been an emerging research topic in recent years. Recent automatic FER systems generally apply deep learning methods and focus on two important issues: lack of sufficient labeled training data and variations in images such as illumination, pose, or expression-related variations among different cultures. Although Convolutional Neural Networks (CNNs) are widely used in automatic FER, they cannot be used when the number of layers is large. Therefore, a residual technique is applied to CNNs and this architecture is named residual neural network. In this paper, an automatic facial emotion recognition method using residual networks with random data augmentation is proposed on a merged FER dataset consisting of 41,598 facial images of size 48 × 48 pixels from seven basic emotion classes. Experimental results show that ResNet34 with data augmentation performs better than CNN with a classification accuracy of 81%.

***Index Terms*—**Deep learning, facial emotion recognition, residual neural networks, ResNet34

I. INTRODUCTION

Facial expressions are the signs that most clearly reveal people's emotional states and intentions when communicating [1]. Humans have the ability to convey emotions during communication through facial expressions, mimics, body language, and sound. Several works have been proposed on automatic facial emotion recognition in areas such as human–computer interaction systems, marketing, security, and healthcare. Ekman and Friesen [2] defined six basic emotions based on cross-cultural work. These emotions can be categorized as anger, disgust, fear, happiness, sadness, and surprise. Contempt was then added as the seventh emotion [3].

Automatic facial emotion recognition (FER) methods can be categorized into two main categories: traditional methods and deep learning based methods. Most of the traditional methods obtain feature vectors representing the facial expression depending on the positions of the turning points (eye, mouth, nose, and eyebrow) [4] or shallow learning, such as local binary patterns (LBPs) [5], non-negative matrix factorization (NMF) [6], scale-invariant feature transform (SIFT) [7].

The widespread use of deep learning has increased the accuracy of FER tasks. Convolutional Neural Network (CNN) models are one of the most commonly used models in FER. A CNN architecture has many layers consisting of convolution, pooling, and fully connected (FC) layers. Recently, deep CNNs have been used for feature extraction [8-10], which has significantly improved the performance of automatic facial emotion recognition. It is straightforward to increase the number of layers in deep learning for better learning. However, increasing the number of layers in CNNs results in degraded accuracy due to vanishing gradients. The vanishing gradient problem has been solved by adding residual connections in CNNs. This leads to more efficient learning schemes called Deep Residual Network (Deep ResNet) [11]. ResNets have been successfully applied in image recognition and image classification tasks due to their efficiency. In this paper, ResNet34, a residual neural network architecture with 34 layers, is used for facial emotion recognition with data augmentation on a merged dataset.

Although CNN-based automatic FER systems have outperformed the traditional methods, the lack of sufficient labeled data still remains an issue. In this research, in order to obtain a larger and more balance dataset, Facial Emotion Recognition (FER+) [12], the Extended Cohn-Kanade (CK+) [13] and Karolinska Directed Emotional Faces (KDEF) [14] are merged. Even though the

Corresponding author:
Serap Kırbız

E-mail:
kirbizs@mef.edu.tr

Received: July 26, 2024

Revision Requested: September 26, 2024

Last Revision Received: October 6, 2024

Accepted: October 19, 2024

Publication Date: November 8, 2024

DOI: 10.5152/electrica.2024.24098

dataset size has increased, the dataset is imbalanced and there is still not enough labeled data from each class. Thus, data augmentation is applied to prevent overfitting.

The rest of the paper is organized as follows. Section II summarizes some studies related to automatic FER systems based on deep learning. Section III explains the proposed method in detail. Section IV reports the classification performance of the proposed method, and Section V summarizes the paper.

II. DEEP FACIAL EMOTION RECOGNITION

In this section, we briefly explain the three main steps of automatic FER approaches based on deep learning: preprocessing, feature learning, and classification.

A. Preprocessing

Preprocessing is applied to facial images before training the deep neural network in order to minimize the variations among the images, such as background, illumination, head poses, etc.

Face detection is the first step in data preprocessing in order to remove the background. The Viola–Jones face detector [15] has been commonly used in the literature due to its robustness and low computational load.

Deep Neural Networks (DNNs) require a large amount of labeled data for training. However, most of the publicly available FER datasets do not have sufficient data for training. Therefore, data augmentation techniques are applied to avoid overfitting. The most commonly used augmentation methods are rotation, shift, scaling, adding noise, etc. [1]. Applying combinations of several augmentation methods can generate more diverse images and make the network more robust to variations among head poses or illumination. Besides these image processing techniques, deep learning can be used to generate synthetic images by three-dimensional CNN (3D CNN) [16] or Generative Adversarial Network (GAN) [17].

B. Feature Learning

In the last decades, deep learning has achieved state-of-the-art performance for several applications. In deep learning, the aim is to obtain high-level abstractions from hierarchical architectures consisting of many layers of nonlinear transformations and representations. In this section, we briefly explain two models that have been commonly applied in deep learning-based techniques in FER.

1) Convolutional Neural Networks

Convolutional Neural Network has been widely used in several applications of computer vision, including automatic FER. In the last decades, several studies in the FER literature [18] have shown that CNN is robust to changes in face location, variations in scale, and performs better than the multilayer perceptron (MLP) in the case of previously unseen face pose variations.

A CNN has three types of layers: convolutional layers, pooling layers, and fully connected (FC) layers. One of the most important characteristics of CNNs, unlike other deep learning models, is their special convolutional structure, which inherently imposes sparsity and significantly reduces the number of parameters. The convolutional layer has a set of learnable kernels which convolve with the input image and produce various types of activations. There are three primary benefits of the convolution process: local connectivity, which captures correlations among neighboring pixels; weight sharing, which significantly reduces the number of learnable parameters; and shift-invariance with respect to the location of the object [1]. The convolution layer is followed by the pooling layer in order to reduce the size of the features and the computational load of the neural network. The most commonly used nonlinear downsampling methods in neural networks are average pooling and max pooling, which also provide translation invariance. The last layer is generally the FC layer, which ensures that all the neurons in the current layer are connected to all activations in the previous layer. The FC layer also enables the conversion of the two-dimensional feature maps into one-dimensional feature maps for classification.

Adding more layers to CNN helps to detect different features. However, increasing the number of layers in CNN architectures generally causes performance degradation due to the vanishing gradient problem and overfitting.

To alleviate the overfitting problem, many studies have proposed finetuning their networks on known pretrained models like AlexNet, VGGNet, and ResNet [1]. Knyazev et al. [19] have shown that pretraining the model on a larger face recognition dataset and finetuning with FER datasets improves emotion recognition performance. Although the network size is very large for Visual Geometry Group (VGG) models and it requires more time to train, they have promising FER performance [20]. EfficientNet [21] has been proposed as a finetuned CNN model with high performance in FER. EfficientNet [21] uses a compound coefficient technique to scale up depth, width, and resolution of models effectively.

In our recent work, CNN was used for automatic FER on merged datasets using several augmentation techniques [8].

2) Residual Neural Networks

Although CNN models have been used for automatic FER and achieved state-of-the-art results, it is observed that as the number of layers increases beyond five, the classification accuracy decreases [11]. This problem can be eliminated by adding residual connections between every other layer and propagating the value of features.

In this paper, the Residual Neural Network (ResNet) [11, 22] is preferred for automatic FER because of its residual connections, which eliminated the vanishing gradient problem of CNNs with many layers. The ResNet proposes two types of mapping: identity and residual. As can be seen from Fig. 1, for input x, the output can be obtained as $y = F(x) + x$, in which x corresponds to the identity

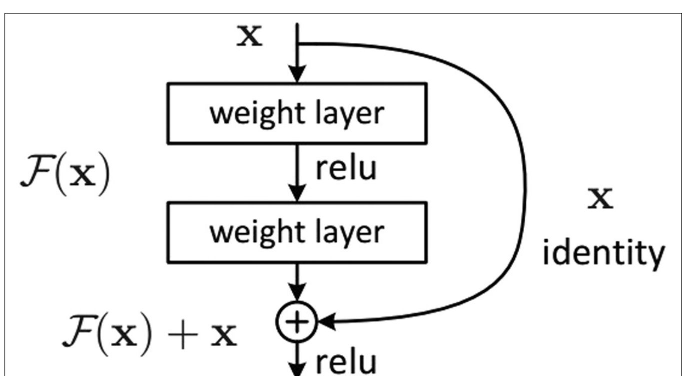

Fig. 1. The residual block.

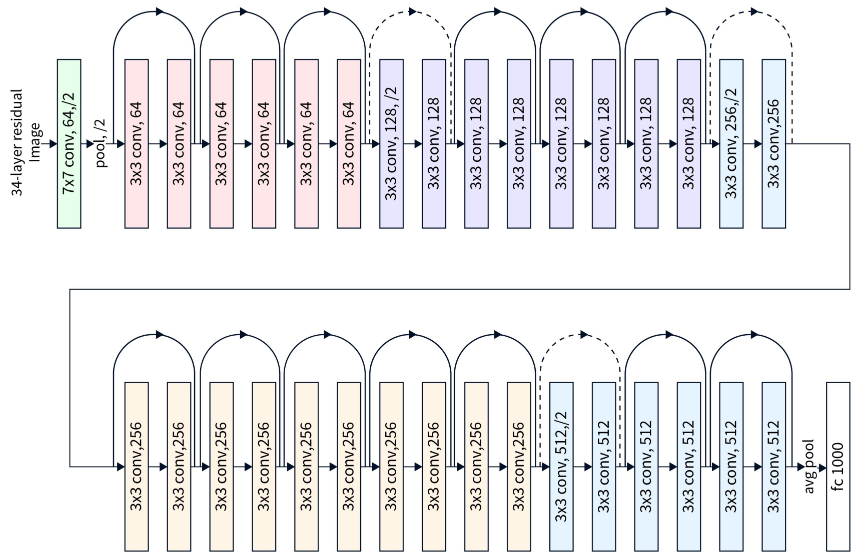

Fig. 2. The structure of ResNet34 [23] .

mapping, and the residual mapping corresponds to the difference $y - x = F(x)$. The ResNet model reuses the activations from the previous layers for faster learning. There are several ResNet architectures with different numbers of layers, e.g., ResNet14, ResNet34, ResNet50, etc. The number at the end of ResNet indicates the number of layers in the network.

In this paper, the ResNet model used with depth 34 is preferred due to its classification performance in image classification. Compared to CNNs, the deep Residual Networks 1) are easier to optimize; 2) represent better; and 3) have higher classification performance using a deeper architecture.

A. Classification

At the final step, the learned deep features are used to classify the facial emotion into one of the seven basic emotion classes: anger, fear, disgust, happiness, sadness, neutral, and surprise. In traditional methods, the prediction probabilities of the classes are obtained in an end-to-end manner at the output of a loss layer that is added at the end of the network.

In CNNs, generally softmax loss is used to minimize the cross entropy between the ground-truth distribution and the estimated class probabilities.

III. PROPOSED METHOD

In this paper, we propose an automatic FER method based on residual neural networks consisting of 34 convolutional layers (ResNet34). The ResNet34 architecture used in this paper can be seen in Fig. 2.

As can be seen from Fig. 2, ResNet34 consists of one convolution and pooling step followed by four convolutional layers of similar structure. Each of the four layers performs 3 × 3 convolutions with fixed feature map dimensions of 64, 128, 256, and 512, respectively. Each of the four layers bypasses the input every two convolutions. Moreover, the dimensions do not change within each layer, and the dimension of the input changes for the dotted lines due to convolution. Each of the convolution blocks consists of three layers: a convolution layer, a rectified linear unit (ReLU) activation layer, and a batch normalization (BN) layer. The convolutional blocks are used for feature extraction, and the last FC layer is used for classification.

The main problem in facial emotion recognition is the lack of sufficient labeled training data. In order to minimize this problem, we merged three commonly used datasets: FER+ [12], the extended CK+ [13] and KDEF [14]. Fig. 3 displays the distribution of facial images for each emotion class in the merged dataset. As can be seen in Fig. 3, even the training data has sufficient labeled data for most of the classes, there is still not enough data, especially for the disgust and fear classes. This imbalanced dataset problem might reduce the accuracy of the model and cause the algorithm to overfit the majority classes. In order to solve the problems caused by the imbalanced dataset, data augmentation is applied to the training dataset using RandAugment [21] which includes auto-contrast, equalize, invert, rotate, posterize, solarize, color, and contrast.

Each emotion dataset has different characteristics such as illumination, resolution, pose, etc. In order to decrease the variations in the

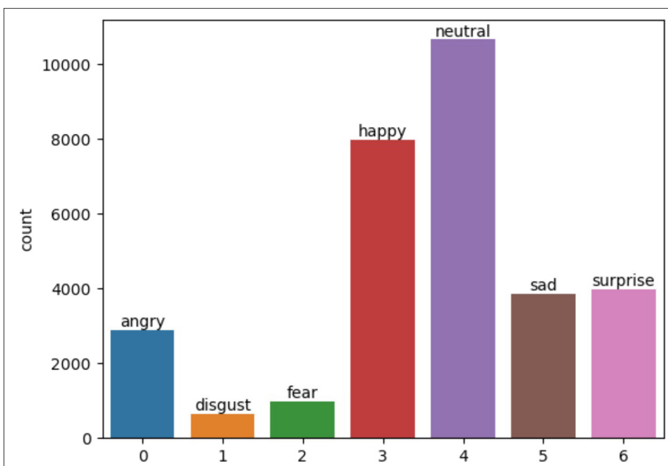

Fig. 3. The distribution of labeled data in the merged database.

images caused by merging different datasets, all the images are converted to grayscale. The Viola–Jones algorithm [15] is applied to the images in order to remove the background. Then, all the images are downsampled to 48×48 pixels.

IV. EXPERIMENTS

In order to evaluate the proposed method, a comparison between three deep learning methods is conducted: CNNs [8], EfficientNet [21] and the proposed Resnet34 model. First, the dataset for this experiment is introduced. Then, the evaluation criteria and experimental setup are explained. Finally, the evaluation results are reported.

A. Dataset

We merged three distinct datasets, FER+ [12], CK+ [13], and KDEF [14], for evaluation of the ResNets.

The first dataset, FER+ [12] consist of 35,771 gray-scale labeled facial images of size 48×48 pixels: 28,559 for training, 3,579 for evaluation and 3573 for test.

The second database is the CK+ [13]. CK+ contains 927 images of size 48×48 pixels from six emotions. The CK+ database does not include any images from the neutral class.

The third database, KDEF [14] is a dataset of a total of 4,900 facial images of size 562×762 from seven basic emotions.

The merged dataset includes 41,598 facial images from seven basic emotions. Fig. 4 illustrates example facial images from the FER+, CK+, and KDEF databases of (a) angry, (b) happy, (c) neutral, and (d) surprised classes. In the merged dataset, we adopted 80%, 10%, and 10% of available annotated data for the learning, validation, and testing of deep learning models, respectively. The input data of all deep learning models are 48×48.

B. Evaluation Criteria

In order to measure the performance of the proposed method, we use the confusion matrix, precision, recall, F1 score, and accuracy rate.

Confusion matrix is used to demonstrate the number of correctly and incorrectly predicted samples by a classifier. The confusion matrix is calculated based on four items: True Negative (TN), True Positive (TP), False Negative (FN), and False Positive (FP).

Precision measures the ratio of the true positive predictions (TP) to all positive predictions (TP + FP):

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall measures the ratio of the true positive predictions (TP) to the sum of true positive predictions (TP) and false negative predictions (FN) as follows:

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

F1 score is commonly used to evaluate classification tasks, and it is calculated as the harmonic mean of recall and precision:

$$F1Score = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \tag{3}$$

The accuracy rate is defined as the ratio of correct predictions for the given test set.

C. Experimental Setup

We run all experiments for 100 epochs, optimizing the cross-entropy loss. The image dataset is shuffled, and samples are selected

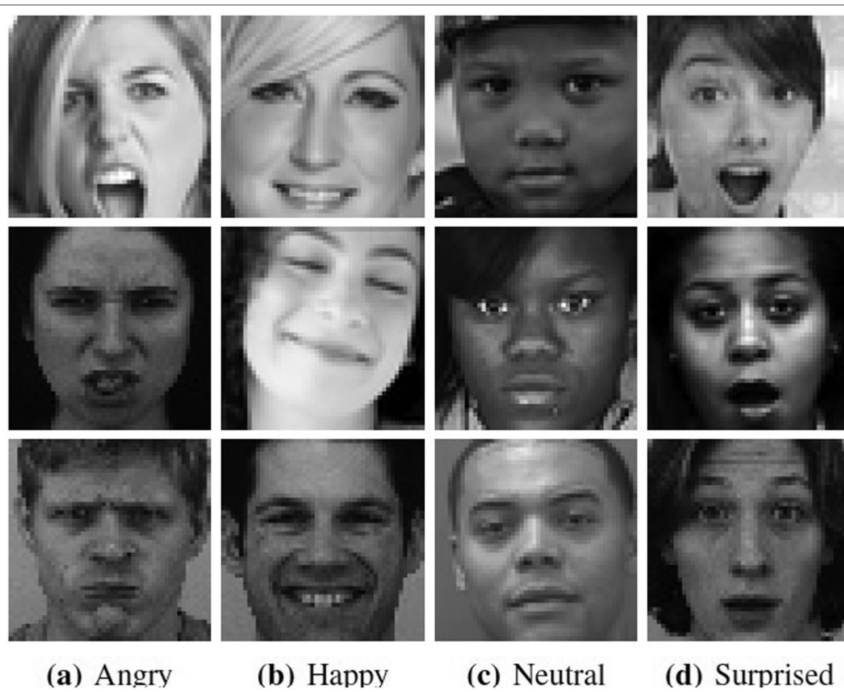

Fig. 4. Sample images from (a) angry, (b) happy, (c) neutral, and (d) surprise classes of FER+ (top row), CK+ (middle row), and KDEF (bottom row) databases.

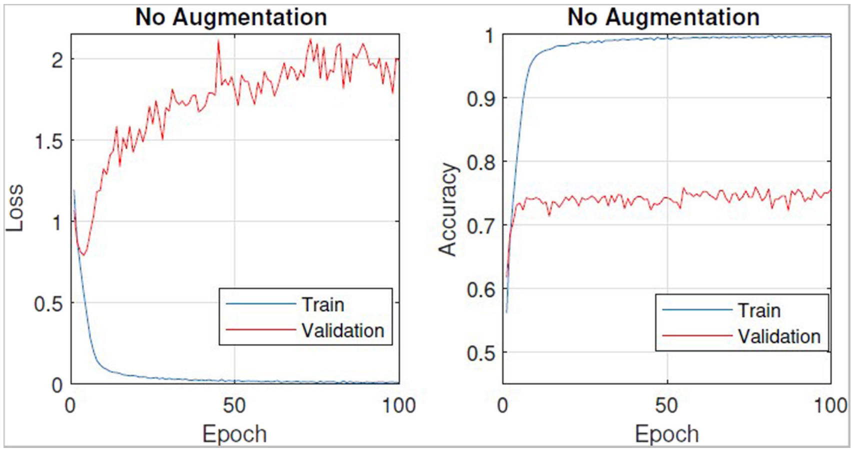

Fig. 5. The accuracy and the loss of training and validation data of the merged dataset on the ResNet-based model over 100 epochs.

randomly and entered into the training, validation, and test sets, in order for the experiment to be performed again.

Adam optimizer is used with a learning rate of 0.001. The model parameters are saved for the best validation accuracy to avoid overfitting.

All the methods are trained on an NVIDIA Tesla K80 GPU and implemented with PyTorch.

V. RESULTS

We report the classification performance of the proposed ResNet34 model with and without data augmentation in order to see the effect of data augmentation. Data augmentation is performed using seven different methods:

- Flip: flips the given image horizontally and randomly with a given probability. The default probability of 0.5 is used.
- Rotate: rotates the image by an angle. The range of the rotation degrees is given as (−20, 20).
- Flip + Rotate: Both horizontal flip and rotation are applied.
- AugMix [24]: applies several augmentations separately using photometric and geometric transformations and combines the augmented images into a new image using element-wise pixel combinations.
- AutoAugment [25]: searches for the optimal augmentation from various augmentation techniques by using reinforcement learning.
- TrivialAugment [26]: randomly chooses an augmentation and applies the chosen augmentation with a selected strength. TrivialAugment applies only one augmentation to each image.
- RandAugment [27]: randomly chooses augmentation. Different than TrivialAugment, RandAugment may apply more than one augmentation to each image.

First, the change in loss and accuracy versus training error is displayed in Fig. 5 for the proposed ResNet34 model. We can see that although the training loss decreases with increasing epochs, the validation loss starts to increase, which indicates overfitting. We can observe the same pattern for the classification accuracy. To optimize the learning parameters, we save the model parameters for the best validation accuracy. Fig. 6 displays the change in loss and accuracy versus training epoch when the augmented training dataset is used for training and RandAugment is used for data augmentation. If we compare Fig. 5 and Fig. 6, we can see that we get a better training loss and training accuracy when we train our model with data

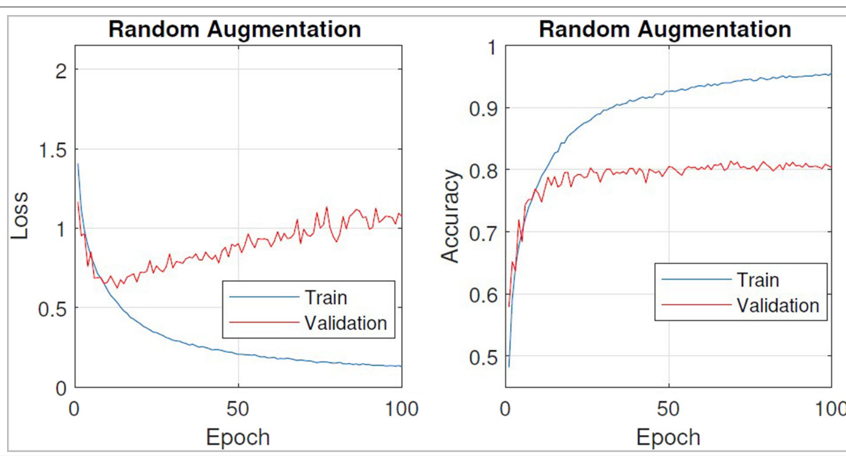

Fig. 6. The accuracy and the loss of training and validation data of the merged dataset on the ResNet-based model over 100 epochs when random augmentation is applied.

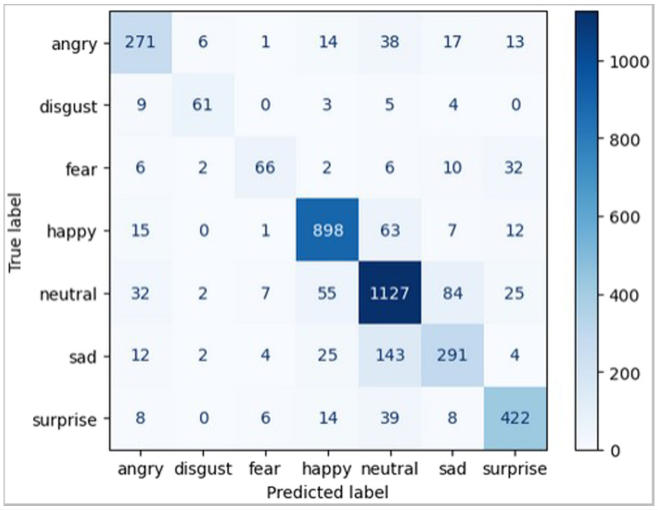

Fig. 7. The confusion matrix obtained by the ResNet34-based model when random augmentation is applied.

augmentation. Although the training loss increases when we train with data augmentation, the validation loss decreases significantly and the validation accuracy increases by more than 5%, proving that data augmentation prevents overfitting.

Fig. 7 displays the proposed model's confusion matrix on the merged test set. The model shows the best classification on the "happy," "neutral," and "surprise" emotions. On the other hand, it has the lowest classification accuracy for the "fear" and "disgust" classes due to the low number of samples in the training dataset.

Table I compares the classification accuracy obtained on the test set using CNN [8], EfficientNet [21] and the proposed ResNet34 model with and without data augmentation. It can be seen from Table I that CNN and ResNet34 outperform EfficientNet. Moreover, the applied data augmentation methods improve the classification accuracy. Among the augmentation methods, RandAugment [27] applied during ResNet34 training, performs the best classification with an average accuracy rate of 81%.

TABLE I. ACCURACY RATE USING CNN, EFFICIENTNET, AND RESNET34 ON TEST SET

Model	Accuracy Rate (%)
CNN [8]	76.5%
EfficientNet [21]	71.6%
ResNet34	74.4%
ResNet34 (flip)	76.4%
ResNet34 (rotate)	76.2%
ResNet34 (flip + rotate)	77.9%
ResNet34 (AugMix)	74.5%
ResNet34 (AutoAugment)	76.3%
ResNet34 (TrivialAugment)	78.1%
ResNet34 (RandAugment)	81%

TABLE II. CLASSIFICATION PERFORMANCE IN TERMS OF PRECISION, RECALL, AND F1 SCORE

Emotion	Recall	Precision	F1-Score
Angry	0.753	0.768	0.760
Disgust	0.744	0.836	0.787
Fear	0.532	0.776	0.632
Happy	0.902	0.888	0.895
Neutral	0.846	0.793	0.819
Sad	0.605	0.691	0.645
Surprise	0.849	0.831	0.840

Table II reports the classification performance of the proposed ResNet34 model with random augmentation in terms of precision, recall, and F1 score. The proposed method estimates the emotions with high precision, recall, and F1 score for most emotions. The performance can be improved by adding more images from larger and diverse datasets.

VI. CONCLUSIONS

In this paper, we have used a deep ResNet34 model with various data augmentation techniques for facial emotion recognition using a merged facial emotion dataset. According to the experimental results, we can see that the ResNet34 model trained with random data augmentation achieved very competitive classification results and increased the classification accuracy significantly. In the future, we will increase the size of the dataset, involve more hyperparameters, use bigger architecture, and involve transfer learning.

Availability of Data and Materials: The data that support the findings of this study are available on request from the corresponding author.

Peer-review: Externally peer-reviewed.

Declaration of Interests: The author has no conflict of interest to declare.

Funding: The author declared that this study has received no financial support.

REFERENCES

1. S. Li, and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affect. Comput.*, vol. 13, no. 3, pp. 1195–1215, 2020. [CrossRef]
2. P. Ekman, and W. V. Friesen, "Constants across cultures in the face and emotion," *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, 1971. [CrossRef]
3. D. Matsumoto, "More evidence for the universality of a contempt expression," *Motiv. Emot.*, vol. 16, no. 4, pp. 363–368, 1992. [CrossRef]
4. M. Sari, A. Moussaouı, and A. Hadid, "Automated facial expression recognition using deep learning techniques: An overview," *Int. J. Inform. Appl. Math.*, vol. 3, no. 1, pp. 39–53, 2020.
5. C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, vol. 27, no. 6, pp. 803–816, 2009. [CrossRef]
6. J. Zhou, S. Zhang, H. Mei, and D. Wang, "A method of facial expression recognition based on gabor and nmf," *Pattern Recognit. Image Anal.*, vol. 26, no. 1, pp. 119–124, 2016. [CrossRef]
7. Y. Shi, Z. Lv, N. Bi, and C. Zhang, "An improved sift algorithm for robust emotion recognition under various face poses and illuminations," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9267–9281, 2020. [CrossRef]
8. B. Kayaoğlu, T. Toktaş, and S. Kırbız. "CNN-Based emotion recognition using data augmentation and preprocessing methods." 2023

Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, 2023.
9. M. R. A. Borgalli, and S. Surve, "Deep learning for facial emotion recognition using custom cnn architecture," *J. Phys. Conf. Ser.*, vol. 2236, no. 1, p. 012004, 2022.
10. O. Ezerceli, and M. T. Eskil, "Convolutional neural network (cnn) algorithm based facial emotion recognition (fer) system for fer-2013 dataset," in 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), IEEE PUBLICATIONS, 2022, pp. 1–6. [CrossRef]
11. Z. Zhong, J. Li, L. Ma, H. Jiang, and H. Zhao, "Deep residual networks for hyperspectral image classification," in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE PUBLICATIONS, 2017, pp. 1824–1827. [CrossRef]
12. E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution," in Proceedings of the 18th ACM international conference on multimodal interaction, 2016, pp. 279–283. [CrossRef]
13. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in 2010 IEEE computer society conference on computer vision and pattern recognition-workshops, IEEE PUBLICATIONS, 2010, pp. 94–101. [CrossRef]
14. M. G. Calvo, and D. Lundqvist, "Facial expressions of emotion (kdef): Identification under different display-duration conditions," *Behav. Res. Methods*, vol. 40, no. 1, pp. 109–115, 2008. [CrossRef]
15. P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Vol. 1. Ieee Publications, 2001, pp. I–I. [CrossRef]
16. I. Abbasnejad, S. Sridharan, D. Nguyen, S. Denman, C. Fookes, and S. Lucey, "Using synthetic data to improve facial expression analysis with 3d convolutional networks," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1609–1618. [CrossRef]
17. I. Goodfellow et al., "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, 2014.
18. B. Fasel, "Head-pose invariant facial expression recognition using convolutional neural networks," in. Fourth IEEE international conference on multimodal interfaces. IEEE PUBLICATIONS, 2002, pp. 529–534. [CrossRef]
19. B. Knyazev, R. Shvetsov, N. Efremova, and A. Kuharenko, "Leveraging large face recognition data for emotion classification." 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). IEEE PUBLICATIONS, 2018. [CrossRef]
20. M. Bentoumi, M. Daoud, M. Benaouali, and A. Taleb Ahmed, "Improvement of emotion recognition from facial images using deep learning and early stopping cross validation," *Multimedia Tool. Appl.*, vol. 81, no. 21, pp. 29887–29917, 2022. [CrossRef]
21. M. Tan, and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proceedings of the 36th International Conference on Machine Learning. Long Beach, CA, USA: PMLR, 10–15 June 2019.
22. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778. [CrossRef]
23. P. C. Shih, C.-C. Hsu, and F.-C. Tien, "Automatic reclaimed wafer classification using deep learning neural networks," *Symmetry*, vol. 12, no. 5, 2020. [Online]. Available: https://www.mdpi.com/2073-8994/12/5/705.
24. D. Hendrycks et al., "Augmix: A simple data processing method to improve robustness and uncertainty". International Conference on Learning Representations.
25. E. D. Cubuk et al., "Autoaugment: Learning augmentation policies from data." *arXiv Preprint ArXiv:1805.09501*, 2018.
26. S. G. Müller, and F. Hutter, "Trivialaugment: Tuning-free yet state-of-the-art data augmentation," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, 754–762. [CrossRef]
27. E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 3008–3017. [CrossRef]

Serap Kırbız received her Ph.D. from İstanbul Technical University, İstanbul, Türkiye. During her Ph.D., she worked at Philips Research Center, Eindhoven, Netherlands, and as a visiting researcher with the Department of Cognitive and Neural Systems, Boston University, Boston, USA. She worked as a Postdoctoral Researcher at Gipsa-Lab, Grenoble, France. She has been working as an assistant professor in the Electrical and Electronics Engineering Department, MEF University, İstanbul, Türkiye. Her research interests include digital signal processing, pattern recognition, audio watermarking, audio source separation, matrix and tensor decomposition, facial emotion recognition, and machine learning.