ISSN: 2633-4828 Vol. 4 No.3, December, 2022

International Journal of Applied Engineering & Technology

FACIAL EMOTION DETECTION USING RESIDUAL NEURAL NETWORK

Naveen Kumari
Assistant Professor (Computer Science), Department of Statistics, Punjabi University, Patiala

ABSTRACT
Facial emotion detection (FER) is now a fundamental component of human-computer interaction, allowing systems to successfully understand and detect to human emotions. However, developing accurate FER systems often requires the use of significant annotated datasets and significant computational resources. This research proposes a model for facial emotion detection based on transfer learning. In this paper, ResNet50, residual neural network architecture with 50 layers *is used for recognition of human emotion. The suggested model is tested on CK+ and FER+ datasets. The suggested model outperformed other baseline techniques with classification accuracy of 62.8% on FER+ and 97.6% on CK+.*

Keywords: *Emotions, Facial Emotion Detection, Transfer Learning, ResNet 50, CK+, FER+*

1. INTRODUCTION
One of the most important aspects of nonverbal communication is the understanding of human emotions from facial expressions. As technology becomes more and more integrated into everyday life, a key area of research in artificial intelligence is how to make machines understand and react to human emotions. Facial Emotion Recognition (FER) is used in a wide range of applications, including evaluations of mental wellness and personalized learning environments, as well as driver monitoring systems and interactive forms of entertainment [1].

Despite increased interest, FER remains a challenging endeavor due to the broad range of facial emotions across humans, traditions, environments, and obstacles. Conventional machine learning techniques for facial emotion detection rely on handcrafted features, emotion-specific datasets, which could restrict ability to be generalized and scalability [2]. The latest developments in deep learning, specifically Convolutional Neural Networks (CNNs), have considerably increased facial emotion classification performance. But developing deep neural networks from beginning is extremely expensive and often needs large volumes of information with labels, which are frequently insufficient in emotion recognition applications. To solve these restrictions, this study investigates the use of transfer learning that have been trained for huge-scale image classification tasks and applies them to the FER. This study aims to develop a transfer learning-based FER model on publically available datasets.Several already trained networks are currently accessible, such as DeepNet, ResNet, GoogleNet, and AlexNet[3-6] .

This research is divided into five phases. In the first, the problem's statement and inventive ideas are presented. The associated material is illustrated in the second part. In the third section, the suggested model is discussed. The outcomes of the study are shown in the fourth section. The summary of the work is presented in the last part.

2. RELATED WORK
Traditional two-step method of machine learning has been applied to detecting emotion. In the initial phase, a classifier like the Support Vector Machine , a neural network, a random forest , etc. is applied to identify the emotions, and in the second step, a variety of significant characteristics as well as features are obtained from the images[9-10].

Q. Wei [11] presented a salient feature extraction based deep convolutional network. Khan, et al. [12] proposed a novel saliency map methods for recognition of human facial emotion.

Zhao et al. [13] proposed a saliency maps based method to compare different techniques qualitatively. Authors [14] implemented deep-learning-based based using "LayerWise Relevance Propagation" (LRP) and "Spectral Relevance Analysis" (SpRAy). Khattak et al. [15] developed an effective deep learning approach to classifying age as well as gender from expressive facial features and for recognizing emotional responses. Using a modified joint trilateral filter and a deep convolutional neural network, Kumari et al. [16] designed an efficient emotion-based facial recognition model. Kong et al. [17] used the voting technique to identify a cromegaly from human facial photos.

Using a dataset of 26,000 patient cases, Gurovich et al. [18] developed a deep convolutional neural network and achieved accuracy of 60%.

A transfer learning-based approach was presented by Jin et al. [19] to identify several disorders from facial photos, including leprosy, Down syndrome, hyperthyroidism, and beta-thalassemia.

Several models were developed by the authors using pre-trained models such as VGG 16, ResNet50, and AlexNet as feature extractors and used disease-specific face (DSF) dataset [20].

Razavian et al. [21] showed how well convolutional neural networks trained on massive datasets, like ImageNet, operate as feature extractors in the context of applying transfer learning to medical diagnosis. A system for identifying six developmental diseases from facial photos was presented by Shukla et al. [22] using deep convolutional neural networks.

In conclusion, it is clear that CNN architecture produces good results for classification tasks and can be applied to FER applications. Transfer learning can be used to reduce computational complexity and enhance accuracy by transferring knowledge from a pre-trained network architecture trained with high resolution images from a large dataset to learn on a new, smaller dataset without losing classification performance.

In this research paper, developed an efficient model based on ResNet50 model, which was previously trained on ImageNet dataset. This method concatenates additional layers on top of the previous ResNet50 model to create a modified trained model. The model is then tested on the benchmark datasets.

3. PROPOSED MODEL

Figure 1 presents the suggested transfer learning model with ResNet50 network. In this proposed work, important image features are extracted using the convolution process and their dimensions are reduced via pooling. To improve the model's performance, batch normalization and activation function ReLU are applied.

Architecture of Resnet50 network

The architecture of ResNet50 is separated into five different layers, as shown in the figure 1. The size of the input image is 128×128×3. First layer is convolutional layer, second layer is batch normalization, third layer is ReLu and fourth layer is maxpool. In stage 2, 3, 4, first layer is convolutional layer second layer is Id block. For classification, Average pool, flattening and fully connected layers is used.

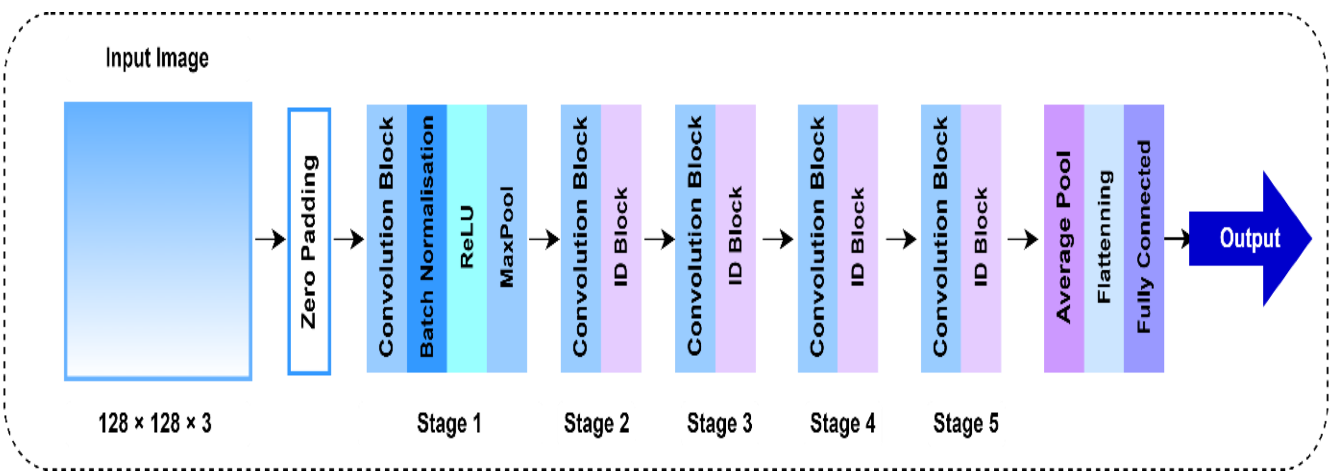

Figure 1: Architecture of ResNet50 for facial emotion recognition.

4. EXPERIMENT AND RESULTS

The proposed model is trained on a 2GB NVIDIA GeForce MX150 GPU with 8th Gen Intel Core i5-8250U processor, 1.6 GHz base processor speed (6MB Cache, up to 3.4 GHz), All the parameters have been fixed during implementation. The Learning rate (η) was set to 0.01, batch size to 16. The proposed model is tested on CK+[7] and FER+[8] benchmark dataset.

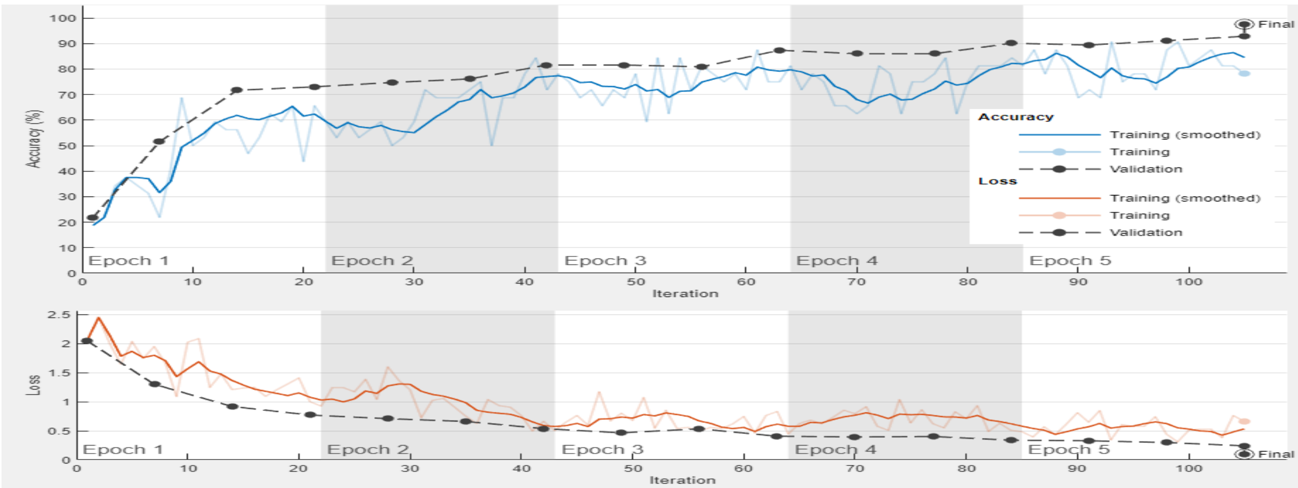

Figure 2: Training accuracy vs loss of ResNet50 on CK+ dataset.

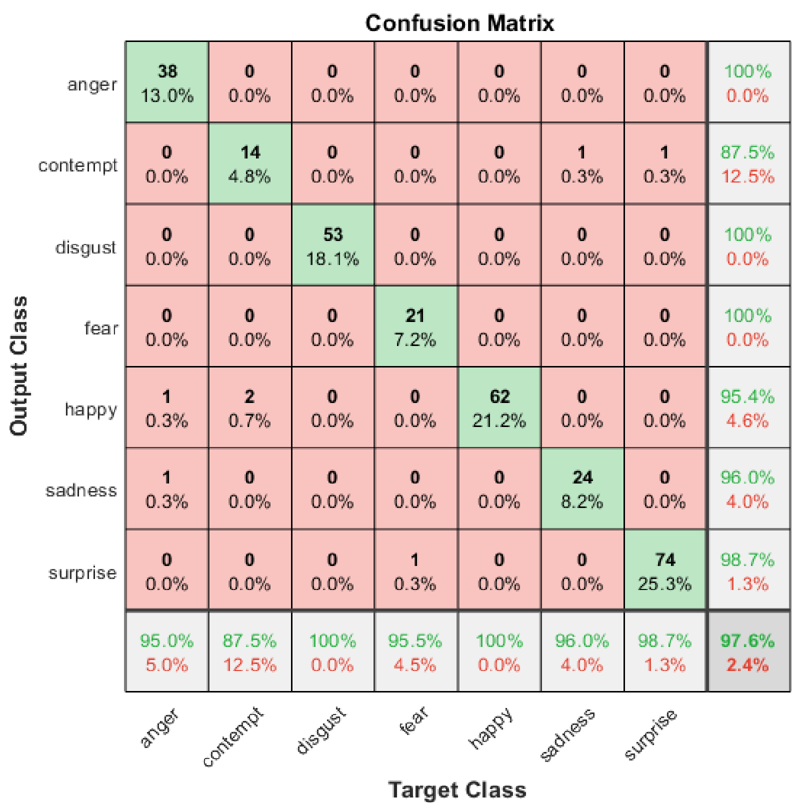

Figure 3: Confusion matrix of CK+ Dataset

Figure 2 illustrates the accuracy and loss curves for the CK+ dataset over the training epochs. The training loss at the final epoch is approximately 0.0243, while the testing accuracy reaches 97.6%. Figure 3 further highlights the performance of the proposed model on the CK+ dataset, reporting a testing accuracy of 97.64%.

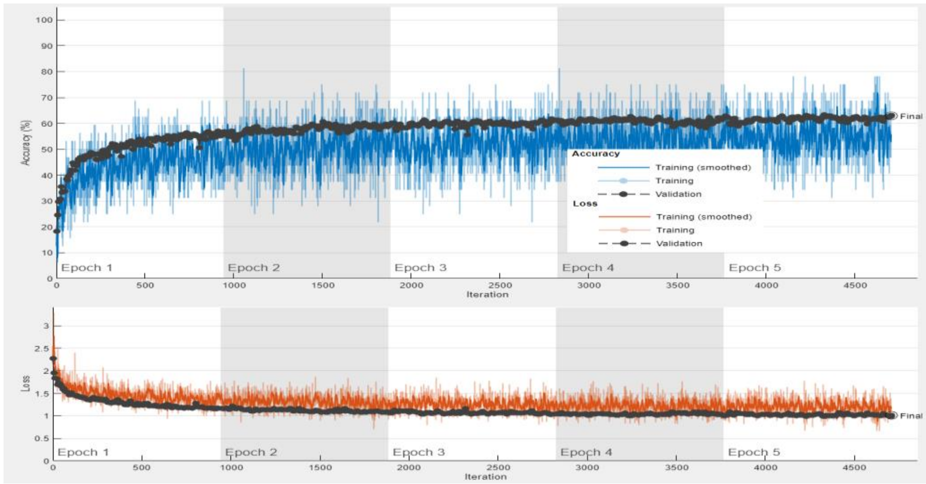

Figure 4: Training accuracy vs loss of ResNet50 on FER+ dataset.

Figure 4 presents the accuracy and loss curves for the FER+ dataset over the training epochs. It shows that the training loss at the final epoch is approximately 1, while the accuracy reaches 97.6%. Figure 5 illustrates the performance of the proposed model on the FER dataset, where it achieves a testing accuracy of 62.8%.

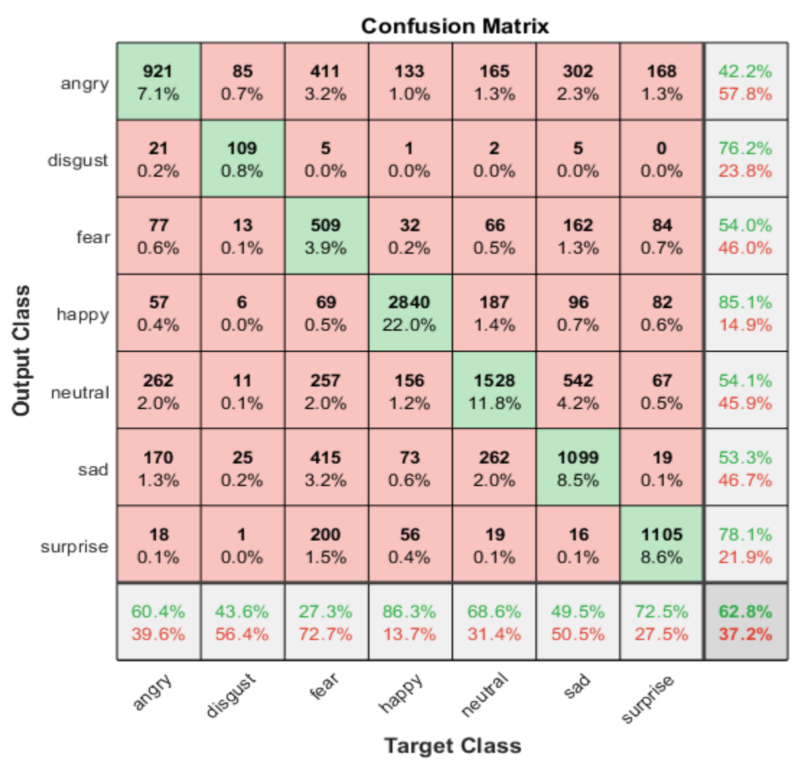

Figure 5: Accuracy of Resnet50 for each class on FER+ dataset.

CONCLUSION

This research work presents a deep residual network, ResNet-50, for facial emotion detection. The proposed framework employs convolutional layers for feature extraction and pooling layers to reduce feature dimensionality. To improve performance, batch normalization and the ReLU activation function are incorporated. The proposed research model is evaluated on two benchmark datasets CK+ and FER+. On the CK+ dataset, it achieves the highest accuracy of 97.6%, whereas on the FER+ dataset, it attains 62.8%.

Funding: This research received no specific grant from any funding agency.

Data Availability: Datasets is freely available on Kaggle website https://www.kaggle.com/shawon10/ckplus.

Conflict of interest: The author declare no competitive interest regarding the publication of this paper

REFERENCES

1. Bin Li, Dimas Lima, Facial expression recognition via ResNet-50, International Journal of Cognitive Computing in Engineering, Volume 2, 2021,Pages 57-64, ISSN 2666-3074, https://doi.org/10.1016/j.ijcce.2021.02.002.(https://www.sciencedirect.com/science/article/pii/S2666307421000073)
2. Shaees, Shamoil & Naeem, Muhammad Rashid & Naeem, Hamad & Syed, Hamza & Arslan, Muhammad & Aldabbas, Hamza. (2020). Facial Emotion Recognition Using Transfer Learning. 10.1109/ICCIT-144147971.2020.9213757.
3. Kumari, N., Bhatia, R. Systematic Review of Various Feature Extraction Techniques for Facial Emotion Recognition System, International Journal of Intelligent Engineering Informatics 2021 Vol.9 No.1,DOI: 10.1504/IJIEI.2021.116088.
4. A. Krizhevsky, I. Sutskever, and G.E. Hinton. “Imagenet classification with deep convolutional neural networks”, Comm. ACM., Vol.60, No.6, pp.84-90, 2017.
5. C. Szegedy, W. Liu, Y. Jia, and P. Sermanet, “Going deeper with convolutions”, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1-9, 2015.
6. 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, pp.770-778, 2016.
7. 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," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.
8. Emad Barsoum and Cha Zhang and Cristian Canton Ferrer and Zhengyou Zhang,” Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution (2016), https://arxiv.org/abs/1608.01041.
9. Shervin Minaee and Amirali Abdolrashidi. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network (2019), arXiv: 1902.01019v1.
10. P. Ekman and Wallace Friesen. Facial action coding system: a technique for the measurement of facial movement. Palo Alto: Consulting Psychologists (1978).
11. Q. Wei, "Saliency Maps-Based Convolutional Neural Networks for Facial Expression Recognition," in *IEEE Access*, vol. 9, pp. 76224-76234, 2021, doi: 10.1109/ACCESS.2021.3082694.
12. Khan, R.A., Meyer, A., Konik, H. *et al*. Saliency-based framework for facial expression recognition. *Front. Comput. Sci.* **13**, 183–198 (2019). https://doi.org/10.1007/s11704-017-6114-9.

13. K. Zhao, W.-S. Chu, H. Zhang, Deep region and multi-label learning for facial action unit detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3391–3399.
14. Finzel, Bettina ; Kollmann, Rene; Rieger, Ines ; Pahl, Jaspar ; Schmid, Ute .Deriving Temporal Prototypes from Saliency Map Clusters for the Analysis of Deep-Learning-based Facial Action Unit Classification. Proceedings of the LWDA Workshops: FGWM, KDML, FGWI-BIA, and FGIR, 2993 (2021), S. 86-97.
15. Khattak, A., Asghar, M.Z., Ali, M. *et al.* An efficient deep learning technique for facial emotion recognition. *Multimed Tools Appl* **81**, 1649–1683 (2022). https://doi.org/10.1007/s11042-021-11298-w.
16. Kumari, N., Bhatia, R. Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter. *Soft Computing* **26**, 7817–7830 (2022). https://doi.org/10.1007/s00500-022-06804-7.
17. Kong, X., Gong, S., Su, L., Howard, N., & Kong, Y. (2018). Automatic detection of acromegaly from facial photographs using machine learning methods. EBioMedicine, 27, 94–102.
18. Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., Basel-Salmon, L., Krawitz, P. M., Kamphausen, S. B., Zenker, M., Bird, L. M., & Gripp, K. W. (2019). Identifying facial phenotypes of genetic disorders using deep learning. Nature Medicine, 25(1), 60–64.
19. Jin, B., Cruz, L., & Gonçalves, N. (2020). Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis. IEEE Access, 8, 123649–123661.
20. Jin, B. (2020). Disease-specific faces. IEEE Dataport. http://dx.doi.org/10.21227/rk2v-ka85[Last visited 10thJuly 2021].
21. Razavian, A. S., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 806–813).
22. Shukla, P., Gupta, T., Saini, A., Singh, P., & Balasubramanian, R. (2017, March). A deep learning framework for recognizing developmental disorders. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 705-714).