Deep Learning-Based Facial Emotion Recognition with Optimized CNN

Jyostna Devi Bodapati
Dept of ACSE
Vignan's University
Vadlamudi,Guntur
jyostna.bodapati82@gmail.com

Bhavya Sri Maturi
Dept of ACSE
Vignan's University
Vadlamudi,Guntur
bhavyamaturi927@gmail.com

Sai Ramya Tanneru
Dept of ACSE
Vignan's University
Vadlamudi,Guntur
sairamyatanneru@gmail.com

Abstract—Facial Emotion Recognition (FER) is a key aspect of computer vision and artificial intelligence focused on allowing machines to understand human emotions by analyzing facial expressions. This paper presents an innovative approach to facial emotion recognition using deep learning techniques, specifically focusing on Convolutional Neural Networks (CNNs). The proposed CNN model is designed to achieve optimal performance in emotion classification tasks. We used FER-2013, which is a large dataset of 35.9k images, to make our study more thorough and reliable. This research explores the impact of different optimizers and learning rates, highlighting the notable achievement of 65% accuracy.

IndexTerms—Convolutional Neural Networks, Facial Emotion Recognition, Emotion Classification

I. INTRODUCTION

Facial emotion recognition, a cutting-edge technology, employs sophisticated computer vision and image processing techniques to decipher human emotions from facial expressions. The process involves the detection, analysis, and interpretation of facial features to understand the emotional states of individuals. This technology taps into the realm of nonverbal communication, as facial expressions serve as powerful cues for conveying a wide spectrum of human emotions. In daily life, facial emotion recognition finds applications across various domains. One prominent area is Human-Computer Interaction. Interactive systems, such as virtual assistants and augmented reality applications, leverage facial emotion recognition to enhance user experience. By allowing machines to respond appropriately to users' emotions, this technology creates more intuitive and personalized interactions. Moreover, facial emotion recognition has significant implications in healthcare settings. It can aid in diagnosing and monitoring mental health conditions like depression and anxiety. Additionally, it assists healthcare professionals in assessing pain levels in patients who may have difficulty communicating verbally. In market research and advertising, facial emotion recognition plays a crucial role. Businesses utilize this technology to gauge consumer reactions to products, advertisements, and brand experiences. By analyzing emotional responses, companies can refine their marketing strategies and develop products that resonate with their target audience. Similarly, in education and learning, facial emotion recognition offers valuable insights. Educators can use it to gauge student engagement and emotional responses during lessons. This information informs teaching practices, personalized learning experiences, and interventions for students who may require additional support. A study explores the latest advancements in facial emotion recognition from images, aiming to uncover prevalent strategies for interpreting and recognizing facial expressions. Analyzing 51 papers from reputable scientific databases like ACM Digital Library, IEEE Xplore, Science Direct, and Scopus, the researchers identified 94 distinct methods. These methods were then categorized into two main trends: classical and neural network-based approaches. Statistical analysis showed slightly higher recognition precision for classical methods over neural networks, despite reduced generalization capability. Additionally, popular datasets for facial expression recognition were assessed, highlighting their strengths and weaknesses and underlining the necessity for reliable data sources across artificial and natural experimental settings[1]. Despite its potential, recognizing emotions through facial images poses a considerable challenge, demanding the extraction of numerous key features crucial for accurate outcomes. To address this,

researchers have explored diverse facial detection algorithms, discovering that combining Haar Cascade Face Detection with CNN models yields enhanced results [2]. This approach improves accuracy and showcases the efficacy of integrating classical and deep learning techniques in facial emotion recognition systems. Furthermore, advanced methodologies such as the Facial Expression Recognition Network (FERNet) have demonstrated remarkable performance, surpassing human accuracy on benchmark datasets like FER-2013. [3]. The innovation has led to breakthroughs in combining CNNs with Long Short-Term Memory Networks (LSTMs), resulting in a significant performance improvement [4].In a similar study, the researchers introduce a deep learning approach based on an attentional convolutional network, enabling the model to emphasize important facial regions. This novel method demonstrates significant enhancements over previous models across various datasets, including FER-2013, CK+, FERG, and JAFFE. Additionally, they utilize a visualization technique to identify crucial facial regions associated with different emotions detected by the classifier's output[5]. Building on the techniques described, we aimed to experiment with a straightforward Convolutional Neural Network (CNN) model for emotion classification on this dataset. Our CNN model is designed to classify emotions such as Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. We explored variations in optimizers and learning rates to improve the accuracy.

II. RELATED WORK

Facial emotion recognition is a critical component of human-computer interaction, utilizing advanced techniques to interpret emotions from facial expressions. Among these methods, deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated significant potential due to their automatic feature extraction and computational efficiency[6].CNNs have proven to be highly effective in tasks related to computer vision, including image recognition, object detection, and facial recognition. The architecture of CNNs involves convolutional layers that automatically learn hierarchical representations of features from the input data[7]. Facial Emotion Recognition (FER) has evolved from laboratory settings to real-world scenarios, relying heavily on deep learning techniques. Despite its progress, there remain obstacles such as overfitting from limited data and expression variations. This paper delves into the intricacies of deep FER, exploring datasets, algorithms, data selection, and evaluation principles. Moreover, it introduces novel deep neural networks and training methods designed for both static and dynamic images[8].Traditional FER methods using manual feature extraction techniques like SVM with HOG(Histogram of Oriented Gradients)[9] and LBP(Local Binary Patterns)[10] have limitations, especially in uncontrolled environments and complex datasets like FER 2013.Notably, a study by Y. Khaireddin and Z. Chen achieved a remarkable single-network classification accuracy of 73.28% on the FER2013 dataset. This accomplishment involved adopting the VGGNet architecture, fine-tuning hyperparameters, and experimenting with optimization methods [11]. To enhance the accuracy of Facial Emotion Recognition (FER) models, researchers have often turned to utilizing single or combined datasets. However, each dataset presents its own set of limitations and challenges that can impact the performance of the trained model. The model proposed by Ozay Ezerceli and M. Taner Eskil, ¨ leverages a combination of the FER2013 and CK+ datasets for FER2013, achieved an impressive accuracy of 93.7%[12]. Concurrently, a novel CC-CNN model introduced a twolevel approach for feature extraction in facial expressions, incorporating the Cyclopentane Feature Descriptor (CyFD). This unique structure captures significant features, achieving a fusion of local and global features through convolutional layers with cross-connections. Finally, towards the end, the CC-CNN method works by fusing all the features extracted from both levels[13]. In parallel, the development of a Deep Convolutional Neural Network (DCNN) focuses on classifying five distinct human facial emotions. This model undergoes rigorous training, testing, and validation using a meticulously curated image dataset, showcasing its proficiency in accurately classifying a diverse range of human emotions[14]. Another model includes Deep Convolutional Neural networks (DCNN) which integrate convolutional layers and deep residual blocks which significantly improves the performance on benchmark datasets[15]. Another study introduces a novel local gravitational force descriptor for extracting local features, which are then fed into a deep convolutional neural network (DCNN). Experiments were conducted on five different benchmark datasets, resulting in improvements in

terms of accuracy, precision, recall, and F1-score metrics[16].A recent advanced study introduced a novel multimodal methodology, leveraging deep learning techniques, to effectively recognize facial expressions even in masked conditions. The methodology employed two widely used datasets, M-LFW-F and CREMA-D, to capture both facial and vocal emotional expressions. Through fusion techniques, a multimodal neural network was trained, surpassing the performance of traditional unimodal methods in facial expression recognition.The proposed approach performs well in recognizing facial expressions amidst masked conditions[17].A study introduced an attention mechanism tailored for automatic facial expression recognition using traditional feature extraction techniques. The network architecture encompasses four essential components such as the feature extraction module, attention module, reconstruction module, and classification module. Utilizing Local Binary Patterns (LBP) features, texture information is extracted from images, capturing nuanced facial movements to enhance network performance. With the integration of an attention mechanism, the neural network can effectively prioritize pertinent features which resulted in enhanced outcomes[18]. Additionally, a notable study harnessed the Viola-Jones algorithm for robust facial detection, enhancing its capabilities by integrating a pre-trained ConvNet network with attention mechanisms such as SoftMax attention. Attention improves the overall performance resulting in greater accuracy. This process yields feature descriptors for expression recognition. This approach demonstrated impressive results across diverse benchmark datasets, highlighting the effectiveness of attention-driven feature extraction in facial emotion analysis[19]. Another study introduces the effect of parameters such as kernel size and number of filters on classification accuracy. In this study, the researchers propose two novel CNN architectures that achieve a human-like accuracy of 65% on the FER-2013 dataset.[20] Expanding on this, the work of K. Liu, M. Zhang, and Z. Pan reported noteworthy findings, achieving a commendable 62% accuracy with an ensemble CNN[21]. Diverging from the ensemble approach, our research focused on optimizing a standard CNN model, conducting experiments with diverse hyperparameters to surpass this reported accuracy and enhance overall performance.

III. METHODOLOGY

The proposed method includes cleaning the dataset, dividing the dataset into batches, splitting the dataset, and then applying the CNN model. The model is trained to categorize the images into 7 classes: Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Convolutional Neural Network is a type of neural network architecture designed for processing structured grid data, such as images. In this proposed methodology for facial emotion recognition, a Convolutional Neural Network (CNN) architecture is employed within a Sequential model. The model comprises multiple layers, including convolutional layers for feature extraction, Batch Normalization for improved training stability, and max-pooling to capture spatial hierarchies. Dropout layers are strategically introduced to mitigate overfitting. The model takes grayscale images of size 48x48 pixels as input, focusing on a single channel. The final layers consist of fully connected dense layers with a Softmax activation function, facilitating multi-class emotion classification. Training involves categorical cross-entropy loss and the use of different optimizers with different learning rates, with model performance monitored on both training and validation sets. To evaluate the model, metrics such as accuracy, precision, recall, and F1 score are measured on a distinct test dataset.

Fig. 1. Architecture of the Proposed Model

IV. RESULTS

A. DATASET

The FER-2013 dataset comprises 35,900 images, categorized into training and test sets with 28,709 and 3,589 images, respectively. To enhance dataset quality, we eliminated blank images and emojis, resulting in a cleaned version. The emotions represented in the dataset encompass Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral, providing a diverse range for emotion recognition model training and evaluation.

Fig. 2. FER-2013 CLASSES

B. EVALUATION METRICS

In evaluating our proposed method, we assessed various metrics including accuracy, precision, recall, and F1 score. We chose accuracy as the primary metric for the evaluation of our model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

TP (True Positive) is the number of instances correctly predicted as positive. TN (True Negative) is the number of instances correctly predicted as negative. FP (False Positive) is the number of instances incorrectly predicted as positive. FN (False Negative) is the number of instances incorrectly predicted as negative.

C. PERFORMANCE ANALYSIS

Classification Report:

	precision	recall	f1-score	support
happy	0.56	0.55	0.56	957
sad	0.83	0.58	0.68	111
neutral	0.57	0.43	0.49	1024
disgust	0.83	0.84	0.84	1774
angry	0.57	0.60	0.59	1233
surprise	0.49	0.59	0.54	1247
fear	0.81	0.76	0.78	831
accuracy			0.65	7177
macro avg	0.67	0.62	0.64	7177
weighted avg	0.65	0.65	0.65	7177

TABEL 1.
Classification Report

In our research, we found that the performance of our model is significantly influenced by key hyperparameters. Specifically, opting for the 'categorical cross-entropy' loss function and employing the AMSGrad optimizer with a learning rate of 0.01 emerged as crucial factors determining the overall effectiveness of the model. Using these combinations of hyperparameters helped us in achieving an accuracy of 65%

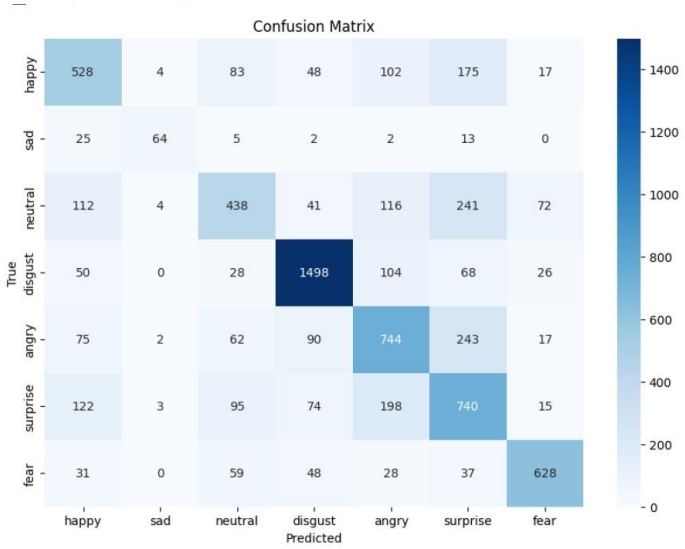

Fig. 3. Confusion Matrix generated by the model on Test Data

V. CONCLUSION

In conclusion, our facial emotion recognition methodology, employing a Convolutional Neural Network (CNN) demonstrated notable success. The process encompassed dataset cleaning, batch division, and model application, achieving a commendable 65% accuracy in categorizing images into seven emotion classes. The CNN architecture, featuring convolutional layers, Batch Normalization, and strategic dropout, proved effective in processing grayscale facial images of size 48x48 pixels. The model's robustness was underscored by performance metrics such as accuracy, precision, recall, and F1 score, assessed on a distinct test dataset. Key hyperparameters, including the choice of the loss function and the use of AMSGrad optimizer with a learning rate of 0.01, significantly influenced model effectiveness

VI. REFERENCES

[1] F. Z. Canal, T. R. Muller, J. C. Matias, G. G. Scotton, A. R. de Sa ¨ Junior, E. Pozzebon, and A. C. Sobieranski, A survey on facial emotion recognition techniques: A state-of-the-art literature review, Information Sciences, 582, 593-617, 2022.

[2] O. C. Oguine, K. J. Oguine, H. I. Bisallah, and D. Ofuani, Hybrid facial expression recognition (FER2013) model for real-time emotion classification and prediction, arXiv preprint arXiv:2206.09509, 2022.

[3] J. D. Bodapati, U. Srilakshmi, and N. Veeranjaneyulu, FERNet: a deep CNN architecture for facial expression

recognition in the wild, Journal of The institution of engineers (India): series B, 103(2), 439-448, 2022.

[4] W. Mellouk and W. Handouzi, Facial emotion recognition using deep learning: review and insights, Procedia Computer Science, 175, 689- 694, 2020.

[5] S. Minaee, M. Minaei, and A. Abdolrashidi, Deep-emotion: Facial expression recognition using attentional convolutional network, Sensors, 21(9), 3046, 2021.

[6] K. O'Shea and R. Nash, An introduction to convolutional neural networks, arXiv preprint arXiv:1511.08458, 2015.

[7] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. AlShamma, et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," Journal of Big Data, vol. 8, pp. 1-74, 2021.

[8] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195- 1215, 2020.

[9] N. Dalal and B. Triggs, Histograms of oriented gradients for human detection, in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005, pp. 886-893.

[10] S. Ke-Chen, Y. A. N. Yun-Hui, C. H. E. N. Wen-Hui, & X. Zhang, Research and perspective on local binary pattern, Acta Automatica Sinica, 39(6), 730-744, 2013.

[11] Y. Khaireddin and Z. Chen, Facial emotion recognition: State of the art performance on FER2013, arXiv preprint arXiv:2105.03588, 2021.

[12] Ozay Ezerceli and M. Taner Eskil, ¨ Convolutional Neural Network (CNN) Algorithm Based Facial Emotion Recognition (FER) System for FER2013 Dataset, in 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), 2022, pp. 1-6. DOI: 10.1109/ICECCME55909.2022.9988371

[13] K. N. Kumar Tataji, M. N. Kartheek, and M. V. Prasad, CC-CNN: A cross connected convolutional neural network using feature level fusion for facial expression recognition, Multimedia Tools and Applications, 1-27, 2023.

[14] E. Pranav, S. Kamal, C. S. Chandran, and M. H. Supriya, Facial emotion recognition using deep convolutional neural network, In 2020 6th International conference on advanced computing and communication Systems (ICACCS), pp. 317-320, IEEE, March 2020.

[15] D. K. Jain, P. Shamsolmoali, and P. Sehdev, Extended deep neural network for facial emotion recognition, Pattern Recognition Letters, 120, 69-74, 2019.

[16] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, Facial expression recognition using local gravitational force descriptor-based deep convolution neural networks, IEEE Transactions on Instrumentation and Measurement, 70, 1-12, 2020.

[17] H. M. Shahzad, S. M. Bhatti, A. Jaffar, M. Rashid, and S. Akram, Multi-Modal CNN Features Fusion for Emotion Recognition: A Modified Xception Model, IEEE Access, 2023.

[18] J. Li, K. Jin, D. Zhou, N. Kubota, & Z. Ju, Attention mechanism-based CNN for facial expression recognition, Neurocomputing, 411, 340-350, 2020.

[19] J. D. Bodapati, D. B. Naik, B. Suvarna, and V. Naralasetti, A deep learning framework with cross pooled soft attention for facial expression recognition, Journal of The Institution of Engineers (India): Series B, 103(5), 1395-1405, 2022.

[20] A. Agrawal and N. Mittal, Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy, The Visual Computer, 36(2), 405-412, 2020.

[21] K. Liu, M. Zhang, and Z. Pan, Facial Expression Recognition with CNN Ensemble, in Proceedings – 2016 International Conference on Cyberworlds (CW), 2016