The 3rd International Conference on Japan–Bangladesh Research and Practice (JBRP2024)
November 29–30, 2024
Online, Coordinated from The University of the Ryukyus, Okinawa, Japan
Organized by the Network of Bangladeshi Researchers in Japan (NBRJ)
Submission Number: 29

Facial Emotion Recognition Using Custom CNN Model: A Comparative Analysis on FER-2013 Dataset

Mst. Jeba Shazida[1], Md. Mahfuzur Rahman[1]*, Bayazid Bustami[1], Angkush Kumar Ghosh[2]
[1] Department of Industrial and Production Engineering,
Jashore University of Science and Technology, Bangladesh
[2] Division of Mechanical and Electrical Engineering,
Kitami Institution of Technology, Japan
* Corresponding Author's Email: mrahman.ipe@just.edu.bd

Track: Natural Sciences, Engineering, and ICT

Keywords: Facial Emotion Recognition, Human-Machine Collaboration, Deep Learning, CNN, Comparison.

Extended Abstract

The successful implementation of Facial Emotion Recognition (FER) has become integral to enhancing Human-Machine or Human-Robot Collaboration (HMC/HRC), a significant feature in smart manufacturing environments. A FER system enhances other manufacturing systems by assessing human emotional states, and thereby improving worker safety, task efficiency, dynamic training and adaptation, as well as adaptive responses during manufacturing [1]. Among many, the FER-2013 dataset includes 35,883 grayscale images, each 48×48 pixels in size, representing seven (7) emotion categories: disgust, fear, anger, happiness, surprise, sadness, and neutral. In this dataset, class imbalance and expression variability, particularly in emotions like "disgust" and "surprise," present challenges to achieving high accuracy compared to other datasets.

Several pre-trained deep learning models (MobileNetV2, ResNet50, DesneNet121, and alike) have been applied to the FER-2013 dataset, as shown in Table 1. MobileNetV2 has 53 layers that basically use ReLU6 activation [2], while ResNet-50 has 50 layers with residual connections [2]. DenseNet121 has 121 layers that connect each layer feed-forwardly [2]. In addition, other custom CNN models like CLCM, FERC, and DCNN (see Table 1), have demonstrated competitive results in recognizing facial emotions in FER-2013 [2, 3].

Table 1: Accuracy of various deep learning models on FER-2013.

Model	Accuracy	Model	Accuracy
MobileNetV2 [2]	68.62%	DCNN [3]	65.68%
FERC [2]	54%	CLCM [2]	63%
ResNet-50 [2]	60%	DeenseNet121 [2]	59%

This study aims to develop a custom CNN model with residual block optimization using hyperparameters (L2 regularization, dropout, and learning rate) for better accuracy in recognizing FER-2013 emotions.

For this, the dataset was extensively augmented after data preprocessing, which involved normalizing pixel values to [0,1]. The custom CNN model consists of 53 layers: one convolution block, three residual blocks, and four fully connected layers. All the layers deploy ReLU activation, except the output layer which uses Softmax. Training was performed with a 32-batch size and 0.0001 learning rate, optimized by cosine annealing with AdamW optimizer. Techniques such as LearningRateScheduler, ModelCheckpoint, ReduceLROnPlateau, and EarlyStopping ensured smooth training via Kaggle's GPU. Additionally, an AdaBoost classifier was used to extract the features from the CNN.

Fig. 1 shows that the model achieved 84.6% training accuracy and 81% testing accuracy. The model-achieved precision is 81.22%, recall at 80.89%, and F1-score of 81.45%. Fig. 2 shows that the model performs well in recognizing emotions called disgust, happiness and surprise. Compared to the models listed in Table 1, the custom CNN with residual blocks shows superior performance in these areas.

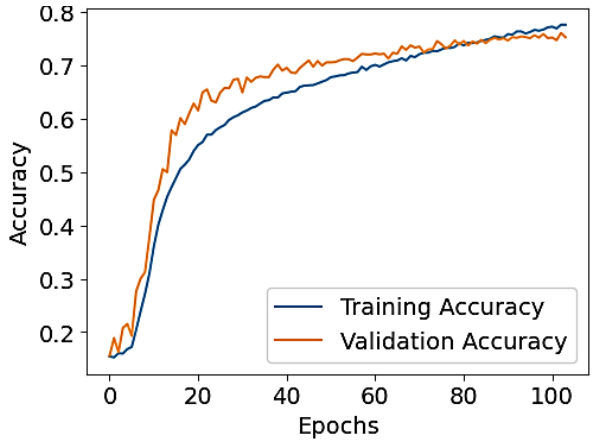

Figure 1: Model Accuracy.

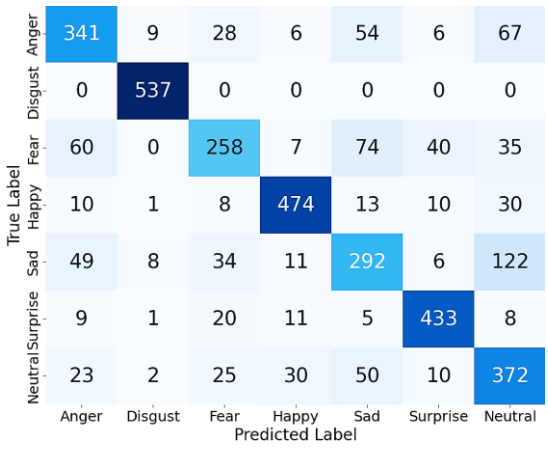

Figure 2: Confusion Matrix.

In conclusion, the custom CNN model developed in this study shows significant performance in FER compared to the traditional models. As such, the model holds promise for effective HMC/HRC in smart manufacturing environments. Future works will involve testing the model with real-time manufacturing shop floor data to evaluate its practical usability. The Monte Carlo Cross-validation (MCCV) will also be employed to verify its consistency and stability. Further enhancement might be considered to improve generalization ability, involving diverse datasets.

References

[1].A. T. Eyam et al., Emotion-Driven analysis and control of Human-Robot interactions in collaborative applications, Sensors, 21, 14, 4626, 2021.
[2].M. Agrawal et al., A Comparative Study on the Effects of Pooling on FER CNN Models, International Journal of Computer Applications, 184, 37, 2022.
[3].D. Bhagat et al., Facial Emotion Recognition (FER) using Convolutional Neural Network (CNN), Procedia Computer Science, 235, 2079–2089, 2024.