Facial Emotion Detection Using Deep Learning with Convolutional Neural Networks

John Russell Castillo
College of Computing and Information Technology
National University
Manila, Philippines
castillojm@students.national-u.edu.ph

***Abstract*— Facial emotion detection plays a crucial role in human-computer interaction and affective computing applications. This paper explores the application of deep learning techniques, specifically Convolutional Neural Networks (CNNs), for automatic facial emotion recognition. The study begins with a comprehensive review of existing methodologies and datasets in the field, highlighting the challenges and advancements. Experimental results demonstrate the efficacy of the proposed method in achieving state-of-the-art performance across various metrics, including accuracy, precision, and recall. Furthermore, the paper discusses practical implications, limitations, and future research directions in the domain of facial emotion recognition using deep learning.**

Keywords*— *Facial Emotion Recognition, Deep Learning, Convolutional Neural Networks (CNNs), Emotion Classification

I. Introduction

Facial expressions are a way to learn more about the physical environment without first-hand experience of the surroundings. One can decide whether to try a new meal, get confirmation to act, or determine whether a threat is nearby by observing another person's face. Developing the ability to correctly interpret others' facial expressions and predict their likely next action would be beneficial [1], [2], [3].

Facial emotion detection is an important task in the fields of artificial intelligence and computer vision, with many real-world applications in entertainment, psychology, and human-computer interaction [4], [5], [6]. The ability to accurately recognize and interpret human emotions from facial expressions can enhance our understanding of human behavior and facilitate more natural and intuitive interactions between humans and machines.

Convolutional Neural Networks (CNNs) have emerged as a powerful tool for facial emotion detection, leveraging their ability to effectively extract and learn hierarchical visual features from image data [4], [6], [7]. CNNs have demonstrated superior performance in various computer vision tasks, including image classification, object detection, and facial recognition, making them a natural choice for facial emotion detection as well.

Several studies have explored the use of CNNs for facial emotion detection, proposing novel architectures and techniques to improve the accuracy and robustness of the models [4], [6], [7], [8]. These approaches often involve preprocessing the input images, designing efficient CNN architectures, and employing advanced training strategies to enhance the model's performance.

The development of large, diverse, and well-annotated facial expression datasets, such as the FER2013 and CK+ datasets, has also been crucial in advancing the field of facial emotion detection using deep learning [5], [6], [8]. These datasets provide the necessary training and evaluation data to develop and assess the performance of CNN-based emotion detection models.

The use of Convolutional Neural Networks for facial emotion detection is a rapidly evolving field, with ongoing research aimed at improving the accuracy, efficiency, and real-world applicability of these systems. The integration of deep learning techniques with facial emotion recognition holds great promise for enhancing human-computer interaction and advancing our understanding of human behavior and emotional expression.

A. Statement of the problem

Human connection depends heavily on emotional intelligence, which affects relationships, communication, and decision-making. But it's still difficult for robots to effectively perceive and react to human emotions. Accuracy and real-time application are two drawbacks of traditional emotion detection techniques like verbal communication analysis and psychological testing.

B. Significance of the Study

The objective of the study is to improve the accuracy and efficiency of face emotion classification systems, which will help the field of artificial intelligence. This can result in better applications in a variety of fields, including improving safety and security, supporting psychological research, and improving user experience in interactive systems

II. Literature Review

This chapter provides a comprehensive review of existing literature related to facial emotion detection. It discusses the overview of emotion recognition, Convolutional Neural Network in emotion recognition, and the current state of the research.

A. Overview of Facial Emotion Detection

Facial emotion recognition (FER) is a field of study that aims to automatically detect and classify emotional states from facial expressions. The general process of FER involves three main steps: face detection, feature extraction, and emotion classification [8], [9].

Face detection is the first step, where the system identifies and locates faces within an image or video frame. Popular techniques for face detection include Haar classifiers and deep learning-based methods [9].

Feature extraction is the process of identifying and representing the unique characteristics of a face that are relevant for emotion recognition. Common feature extraction methods include principal component analysis (PCA), local binary patterns (LBP), active appearance models, and deep learning [9]. PCA generally provides higher recognition rates, while LBP has lower computational complexity [9].

The final step is emotion classification, where the extracted features are used to categorize the facial expression into one or more emotional states. Various machine learning algorithms have been employed for this purpose, such as support vector machines (SVM), AdaBoost, and deep neural networks [9], [10]. The choice of classifier depends on factors like the number of emotions to be recognized, the size and quality of the training data, and the desired trade-off between accuracy and computational efficiency [10].

B. *Convolutional Neural Networks in Emotion Recognition*

Convolutional Neural Networks (CNNs) have emerged as a powerful tool for emotion recognition, revolutionizing the way we understand and process emotional cues from images [7]. CNNs are particularly effective at capturing spatial hierarchies of features in data and have been widely used for various computer vision tasks, including facial emotion recognition [7], [12].

The typical architecture of a CNN for emotion recognition consists of an input layer that receives raw image data, followed by multiple convolutional layers with activation functions to learn hierarchical features [7]. MaxPooling layers are often used to reduce spatial dimensions and computational complexity [7]. The output is then flattened and passed through fully connected layers with a softmax activation function to classify emotions [7].

Emotion recognition using CNNs has numerous practical applications, including human-computer interaction, customer feedback analysis, and mental health monitoring [7]. By accurately identifying and interpreting human emotions from facial expressions, CNNs have the potential to revolutionize various fields that rely on understanding emotional states [7], [11], [12].

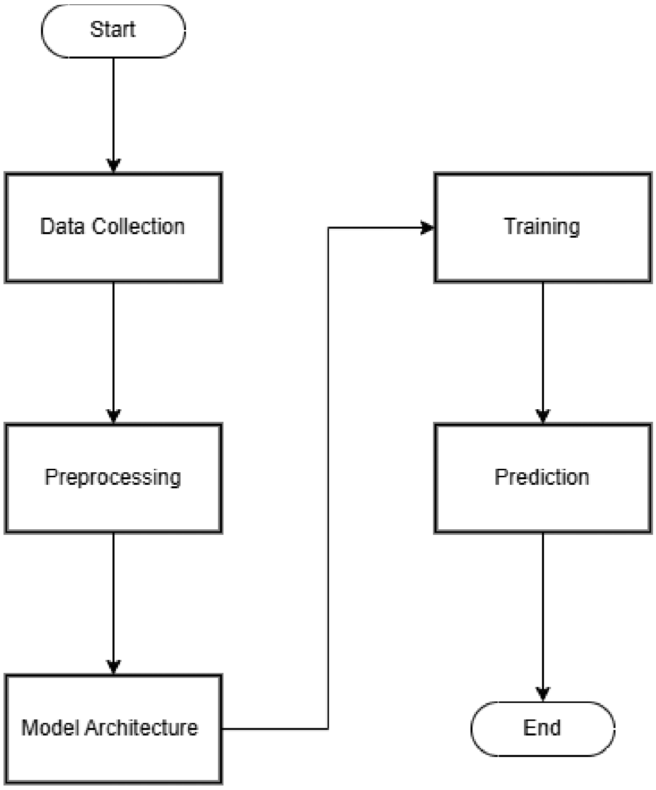
C. *Current State of Research*

The research on facial emotion detection using deep learning with convolutional neural networks (CNNs) has made significant progress, but there are still several gaps and challenges that need to be addressed.

One key challenge is the limited diversity of datasets used in many studies, which often contain images taken under controlled conditions with limited variations in facial expressions, lighting, and other factors. This can lead to models that perform well on these datasets but struggle with more real-world, diverse data

III. METHODOLOGY

In this chapter, the detailed methodology used in the study is described. It includes a step-by-step explanation of data preprocessing techniques specific to facial emotion datasets, the architecture of the CNN model chosen for the study, the training process used to optimize the model, and the evaluation metrics employed to assess the model's performance.



A. *Dataset Collection*

Popular datasets for facial expression recognition tasks are Kaggle's FER (Facial Expression Recognition) 2013 dataset. There are 35,887 grayscale pictures of faces in it that are divided into seven different emotional categories: angry, disgust, fear, happy, neutral, sad, and surprise. Each image is 48x48 pixels. This dataset is frequently used to train and test deep learning models, such as convolutional neural networks (CNNs), for facial emotion recognition.

TABLE I. TRAIN FOLDER

Emotion	Number of Images
Angry	3995
Disgust	436
Fear	4097
Happy	7215
Neutral	4965
Sad	4830
Surprise	3171

TABLE II. TEST FOLDER

Emotion	Number of Images
Angry	958
Disgust	111
Fear	1024
Happy	1774
Neutral	1233
Sad	1247
Surprise	831

B. Preprocessing

The data collected does not include a validation folder, which is an important part of validating the model’s accuracy when training. To solve this, the researcher manually creates a validation folder and gets the half-images from the test folder. The dataset is now split into three folders: train, test, and validation. The researcher also remove the disgust folder because it has a few dataset and will make the accuracy of the model lower

C. Model Architecture

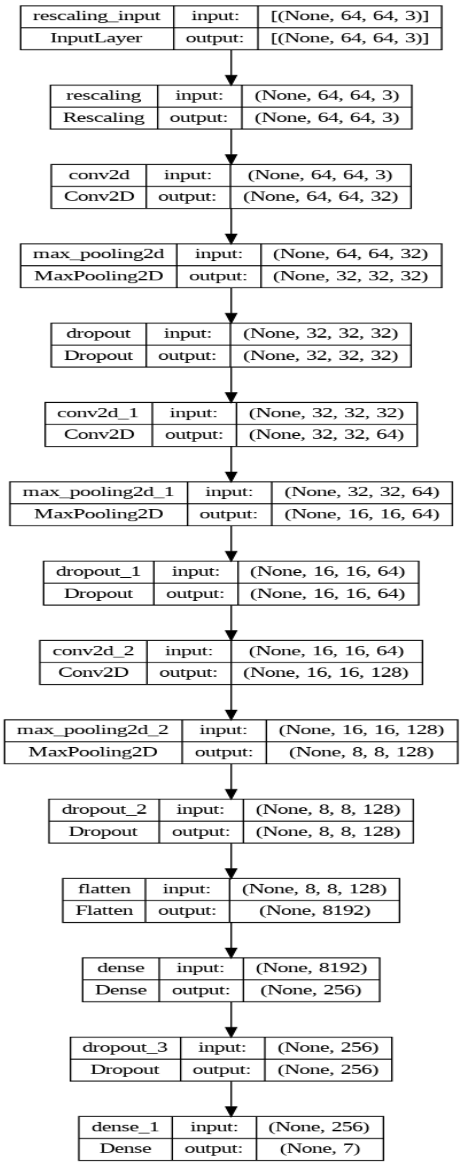


Fig. 1. CNN model visualization

The researcher designed a Convolutional Neural Network (CNN) using TensorFlow and Keras to effectively detect facial emotions from images. The model begins with an input layer that includes a rescaling layer, normalizing pixel values to the range [0, 1] for faster convergence during training. This is followed by three convolutional blocks. Each block consists of a convolutional layer with filters of sizes 32, 64, and 128 respectively, using a 3x3 kernel and ReLU activation function with 'same' padding to preserve the spatial dimensions of the input. These convolutional layers are followed by max pooling layers to reduce spatial dimensions and computational complexity, and dropout layers with a 0.2 dropout rate to prevent overfitting. After the convolutional blocks, a flatten layer converts the 3D output of the last convolutional layer into a 1D vector, which is fed into a fully connected layer with 256 units and ReLU activation. Another dropout layer with a 0.5 rate is included to further prevent overfitting. Finally, the output layer consists of a dense layer with 7 units, corresponding to the seven facial emotion categories, allowing for the prediction of the specific emotion.

D. Training process

The layers stated in the previous section will be compiled using the optimizer adam Using SparseCategoricalCrossentropy will calculate the loss and accuracy as metrics to evaluate the model’s training, a callback function that will early stop the training by monitoring the validation loss of the training in a span of 3 epochs, and lastly, the number of epoch is 30.

E. Evaluation Metrics

In this section, the researcher discusses different methods to evaluate the model for facial emotion recognition: angry, fear, happy, neutral, sad, and surprise. The researcher used plot to see the line graph during the model training and predict the labels. The researcher also used a confusion matrix with precision, recall, f1-score, and an overall accuracy score.

IV. RESULT AND DISCUSION

This chapter presents the findings obtained from applying the methodology. It includes the presentation of results using appropriate tables, figures, and charts to illustrate the performance of the CNN model in detecting facial emotions. The results are analyzed to provide insights into the effectiveness of the proposed approach.

The model's performance during the training was good at first, but in the middle of the training, the validation accuracy was not increasing. At the training stop at epoch 15, the model had 62.26% training accuracy and 97.6% loss. Additionally, the validation accuracy of the model was 54.97%, with a 1.19 validation loss. Figures 2 and Figure 3 will show the line graph of train and validation accuracy and loss.

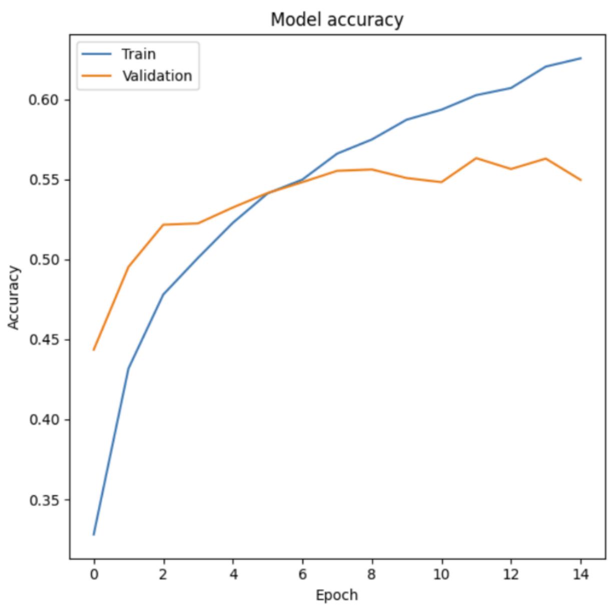


Fig. 2. Model Train and Validation Accuracy Visualization

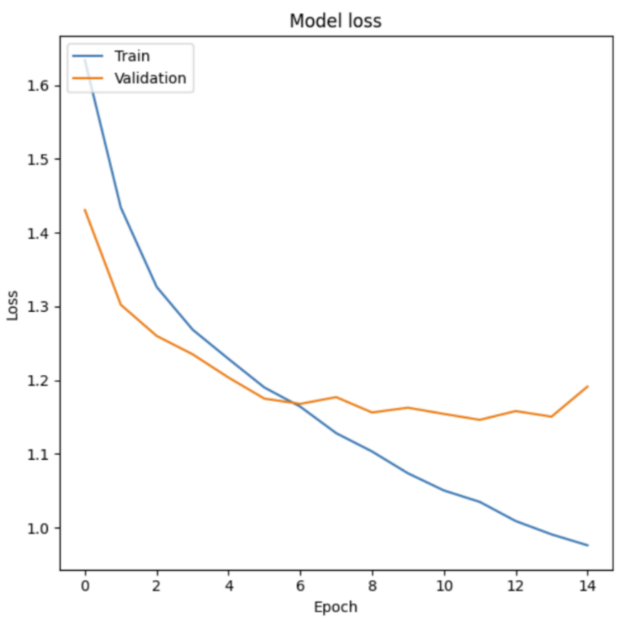


Fig. 3. Model Train and Validation Loss Visualization

Figure 4 shows the precision, recall, and F1 score. The model had 18% overall accuracy when it got new or unseen data. Lastly, Figure 5 will show the model prediction on the test folder

F1 Score:

	precision	recall	f1-score	support
angry	0.13	0.11	0.12	479
fear	0.16	0.10	0.12	512
happy	0.24	0.26	0.25	887
neutral	0.19	0.24	0.21	617
sad	0.19	0.17	0.18	623
surprise	0.12	0.14	0.13	415
accuracy			0.18	3533
macro avg	0.17	0.17	0.17	3533
weighted avg	0.18	0.18	0.18	3533

Fig. 4. F1 Score Visualization

Fig. 5. Model Prediction

V. Conclusion

In this study, the researcher employed a Convolutional Neural Network (CNN) for training the model and achieved 62% training accuracy, 54% validation accuracy, and 56% test accuracy. One of the challenges encountered was that the model often confused the "angry" and "surprise" emotions when the facial expression included an open mouth, which negatively impacted the overall accuracy. The researcher believes that the dataset used is outdated and too small. Additionally, some images in the dataset were blurred, making it difficult for the model to recognize facial features accurately. Increasing the pixel size or enlarging the images would require more training time, which is another challenge given the limitations of the available GPU. To address these challenges, future research should focus on obtaining a more recent and larger dataset with higher-quality images. Enhancing the dataset with more diverse and clear images can help the model better differentiate between similar emotions. Additionally, leveraging more advanced hardware or cloud-based solutions could mitigate the training time issues associated with processing larger and higher-resolution images. Implementing techniques such as data augmentation and transfer learning might also improve the model's accuracy and robustness in recognizing facial emotions.

References

[1] P. Psychology , "The Faces of Emotions," Psychology Today, Dec. 2019. https://www.psychologytoday.com/us/blog/talking-emotion/201909/the-faces-emotions

[2] A. Cuncic, "5 tips to better understand facial expressions," Verywell Mind, Mar. 28, 2023. https://www.verywellmind.com/understanding-emotions-through-facial-expressions-3024851

[3] "Facial Expressions of Emotions (Microexpressions)," Practical Psychology, Dec. 31, 2019. https://practicalpie.com/facial-expressions-of-emotions/

[4] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Applied Sciences, vol. 2, no. 3, Feb. 2020, doi: https://doi.org/10.1007/s42452-020-2234-1.

[5] K. Sarvakar, R. Senkamalavalli, S. Raghavendra, J. Santosh Kumar, R. Manjunath, and S. Jaiswal, "Facial emotion recognition using convolutional neural networks," Materials Today: Proceedings, Aug. 2021, doi: https://doi.org/10.1016/j.matpr.2021.07.297.

[6] D. Dipesh and Patil, "Facial Emotion Recognition using Deep Learning," 2023. Accessed: Jun. 24, 2024. [Online]. Available: https://scholarworks.calstate.edu/downloads/b5645007n

[7] "Emotion Detection Using Convolutional Neural Networks (CNNs)," GeeksforGeeks, Oct. 25, 2023. https://www.geeksforgeeks.org/emotion-detection-using-convolutional-neural-networks-cnns/

[8] P. DursunKarsli, M. Emul, and F. Gencoz, "A Review of the Literature on Emotional Facial Expression and Its Nature," Yeni Symposium, vol. 48, pp. 207–215, Jul. 2010.

[9] D. raval and Mukesh sakle, "A literature review on emotion recognition system using various facial expression," International journal of advance research and innovative ideas in education, vol. 1, no. 2, pp. 326–329, Jan. 2015.

[10] B. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," Sensors, vol. 18, no. 2, p. 401, Jan. 2018, doi: https://doi.org/10.3390/s18020401.

[11]]X. Lu, "Deep Learning Based Emotion Recognition and Visualization of Figural Representation," Frontiers in Psychology, vol. 12, Jan. 2022, doi: https://doi.org/10.3389/fpsyg.2021.818833.

[12] T. Debnath, M. M. Reza, A. Rahman, A. Beheshti, S. S. Band, and H. Alinejad-Rokny, "Four-layer ConvNet to facial emotion recognition with minimal epochs and the significance of data diversity," Scientific Reports, vol. 12, no. 1, p. 6991, Apr. 2022, doi: https://doi.org/10.1038/s41598-022-11173-0.