E-ISSN 2581 – 7957
P-ISSN 2277 – 3916
CVR Journal of Science and Technology, Volume 27, December 2024
DOI: 10.32377/cvrjst2709

Customized Convolutional Neural Network for Detection of Emotions from Facial Expressions

Dr. D. Shyam Prasad[1], Dr. K. Arun Kumar[2]
[1]Sr. Asst. Professor, CVR College of Engineering/EIE Department, Hyderabad, India
Email: pbhanududi@gmail.com
[2]Sr. Asst. Professor, CVR College of Engineering/ECE Department, Hyderabad, India
Email: arun.katkoori@gmail.com

Abstract: **The understanding and accurate interpretation of human emotions are crucial for effective interactions and communication. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized emotion recognition by enabling the analysis of complex facial expressions. This paper focuses on a customized CNN architecture tailored for emotion detection from facial expressions. Utilizing datasets like CK+, FER2013, and JAFFE, preprocessing techniques are employed to enhance model stability and diversity. The proposed CNN architecture comprises convolutional and fully connected layers for feature extraction and classification, respectively. Experimental results, including accuracy, F1 score, and precision, demonstrate the effectiveness of the proposed CNN in accurately identifying various emotions compared to traditional machine learning methods. These findings underscore the potential of CNNs in advancing emotion recognition systems, promising significant applications across diverse fields, from marketing research to virtual reality.**

Index Terms: **Emotion Recognition, Facial Expressions, Convolutional Neural Networks, Customized Architecture, Image Processing.**

I. Introduction

The understanding of human emotions is very important aspect of communication. They play a vital role in our behavior and relationships, and their accurate interpretation and recognition are required in various fields. Due to the emergence of deep learning techniques such as CNNs, which can analyze and understand complex emotions, it has become possible to create systems that can accurately identify and interpret facial expressions. The field of emotion recognition has gained widespread attention due to its applications in various sectors, such as marketing research and virtual reality[1].

Machine learning methods were typically used to identify emotions, but they typically failed to capture the complex patterns that make up human facial expressions. CNNs use a paradigm shift to automatically learn complex hierarchical representations, which helps improve the extraction and classification of emotions. CNNs have demonstrated exceptional capabilities when it comes to accurately identifying and categorizing emotions using large datasets and computational resources [2].

The complexity and variability of human emotions are some of the main challenges that facial expression experts face when trying to identify emotions. These include the interpretation of facial expressions based on cultural and contextual factors. In addition, factors such as facial pose variations and occlusions can also affect the accuracy of this process. These difficulties can be solved by CNNs through the learning of discriminative features in facial images. This allows them to accurately identify various emotions across different environmental and cultural conditions [3].

A CNN architecture that is designed to recognize emotions from facial expressions should be tailored to meet the specific needs of this task. Although off-the-wall CNN platforms have been able to provide promising results in image classification tasks, customized architectures can help improve their efficiency and performance [4].

This paper presents a CNN architecture that is customized to detect facial expressions. We also conduct a comparative analysis of the suggested model's performance against standard frameworks. The proposed model was evaluated and proved to be effective in identifying human emotions using facial expressions. This paves the way for its eventual incorporation into practical applications.

II. Related Works

Machine learning techniques[5][6][7] have been widely used in various fields. It is very important that the machines interact with humans to perform better. In addition, the use of HCI technologies can help identify facial expressions more easily.

Most of the time, researchers focus on developing Active Appearance Model[8] technology to help them identify the facial expressions of people. Machine learning techniques have been used in this field to study the interactions between humans and computers. With the help of HCI technology, it is very easy to identify facial expressions[9].

In the past, researchers[10] have used various types of rules to improve the performance of their machine learning techniques. For instance, they were able to provide results by implementing the fuzzy rules in a support vector machine. Other studies presented by different researchers also proposed methods that involve the use of complex features[11][12].

To improve the performance of their machine learning techniques, researchers have developed different models that can identify the appropriate values for various facial expressions.

Through deep learning, medical researchers have been able to improve the accuracy of their predictions about various

Received on 13-03-2024, Revised on 12-04-2024, Accepted on 12-05-2024.

diseases[13]. According to a study, the use of facial expression was identified as a key component of the deep learning process. The researchers proposed using the ACNN algorithm to recognize facial expressions. This method is commonly referred to as the FACS system, which is a coding system for facial expressions. It is very easy to implement and can be used to improve the accuracy of the prediction [14]. The behavior of people working on technology has changed due to the use of deep learning techniques. It helps visualize the real-time values of the system.

III. Methodology

The block diagram of proposed method is shown in figure 1.

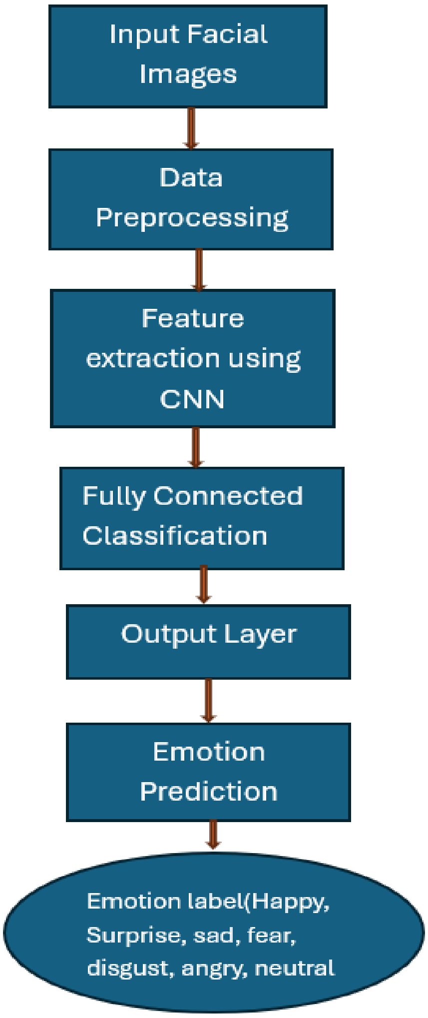

Fig.1. Block diagram of Proposed method

A. Input Data

Various datasets are commonly used in the training and evaluation of CNN models for detecting emotions in humans. One of these is the CK+ dataset, which features 123 facial expressions. It transitions from neutral to emotional sequences and includes expressions such as anger, disgust, happiness, and surprise. The sample images of CK+ dataset is shown in figure 2. The FER2013 dataset contains over 35,000 images of facial expressions extracted from the internet. It features various emotions such as anger, happiness, surprise, fear, and disgust. The sample images of FER2013 are shown in figure 3. On the other hand, the JAFFE dataset features Japanese women's faces with basic expressions. Figure 4 gives the sample images of JAFFE dataset, surprise, Sad, Fear, Disgust, Anger, Joy and Neutral.

Fig. 2. Sample images of CK+ dataset

Fig. 3. Sample images of FER2013 dataset

Fig. 4. Sample images of JAFEE dataset

B. Pre-Processing

The CK+ dataset consists of facial expressions that are posed with consistent faces. To ensure that the model's facial regions are aligned and correctly placed, it should be studied and applied face detection techniques.

The FER2013 dataset contains numerous images that have been extracted from the internet, and these may exhibit varying facial orientation, lighting, and background characteristics. Various techniques can help improve the model's stability and diversity, such as rotation and scaling.

The JAFFE dataset features images of Japanese women with varying emotions. Due to the varying illumination levels, the visibility of certain facial features can be affected. To improve the quality of the data, various techniques, such as gamma correction and histogram equalization, can be used.

C. Feature Extraction

This CNN architecture consists of three convolutional layers followed by max-pooling layers to extract spatial features from input facial images. Figure 5 shows the architecture of customized CNN used in this research.

A spatial feature is a representation of the structures and patterns found within the pixels of a facial image. It can be used to identify the various facial components' spatial layouts. In the context of CNNs, these features play a vital role in identifying facial patterns that are related to certain emotions.

Examples of the spatial features that can be extracted from facial images include:

E-ISSN 2581 – 7957
P-ISSN 2277 – 3916

DOI: 10.32377/cvrjst2709

- ➢ The recognition of contours and edges helps in identifying facial features, such as the eyes, nose, mouth, and eyebrows.
- ➢ A study of texture patterns in facial regions yields information about the surface attributes, like wrinkles and facial expressions.
- ➢ The recognition of local features, such as landmarks and keypoints, helps in identifying facial features that are related to emotional expressions.
- ➢ A study of facial asymmetry and symmetry helps in identifying subtle variations in facial dynamics and expressions linked to different emotions.
- ➢ Through the capture of facial variations, such as eye openness, eyebrow movements, and mouth shape, a network can identify the underlying emotions displayed by an individual.
- ➢ The context of facial components and structures relative to one another provides a basis for assessing emotional expressions and cues.

Max-pooling techniques are used to minimize the spatial dimensions of feature maps while retaining the most important information. This helps in extracting significant features and achieving invariance.

D. Classification

Flattening: After the extraction of the features from the convolutional layers and max pooling, they are transformed into a single vector with a single dimension. This transforms the feature maps into a single array that can be imbibed into the connected layers.

Fully Connected Layers: A feature vector is fed into dense layers, which are also referred to as fully connected ones. These layers are used to classify complex patterns by learning from the extracted ones. The number of neurons in dense layers can vary, which allows the network to capture higher-level representations.

The dropout regularization technique is used to prevent overfitting the layers. When a fraction of the neurons is dropped out during training, the network must learn how to improve generalization and redundant representations.

The network's output layer uses softmax activation to convert the raw output data into probability scores, which represent the likelihood that various classes will be represented in the image.

E. Output

The probability distribution is computed by the SoftMax layer. It shows the likelihood of various emotion classes.

The predicted emotion for the facial image is the result of the highest probability distribution in the output distribution. It is the network's final output.

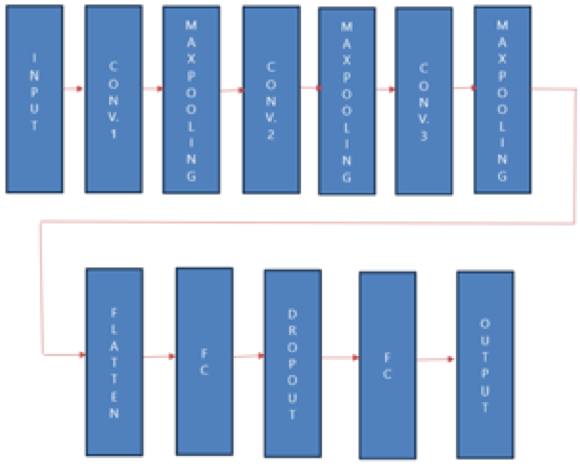

Fig. 5. Customized CNN architecture

IV. EXPERIMENTAL RESULTS

Figure 6 and 7 depict the input from the google images with happy and neutral emotions respectively.

Fig. 6. Emotion-Happy

Fig. 7. Emotion-Neutral

E-ISSN 2581 – 7957
P-ISSN 2277 – 3916

DOI: 10.32377/cvrjst2709

Confusion Matrix

There are three columns for each item in the classification matrix. These are true/actual and predicted, and the confusion matrix records these three values. To maintain consistency, the classification columns are used to represent actual classifications, while the predictions are derived from model predictions.

Corrected-labeled items will be found on the right-hand diagonal from top to bottom-right based on the predicted value and the rater's concurrence. The accuracy of the labels is then evaluated using the confusion matrix.

The table I presents the accuracy rates of Random Forest, Naïve Bayes, Support Vector Machine (SVM), and a proposed Convolutional Neural Network (CNN) in classifying different emotions. Each row corresponds to a specific emotion, while each column represents a model. The values in the table represent the percentage of correct classifications made by each model for the respective emotion. The proposed CNN generally outperforms the other models, achieving the highest accuracy rates for most emotions, such as 91.3% for Happy and 94.2% for Surprise, while SVM follows closely behind. Overall, the results highlight the effectiveness of CNN in emotion classification tasks, showing promising potential for improved accuracy in emotion recognition systems.

TABLE I.
ACCURACY

Emotion	Random forest	Naïve Bayes	SVM	Proposed CNN
Happy	78.2	79.2	81.2	91.3
Surprise	77.6	79.2	79.1	94.2
Sad	81.6	82.6	88	90.3
Fear	81.6	82.6	87.2	91.1
Angry	81.6	80.6	87.6	91.5
Neutral	80.6	82.6	84.6	93.5
Disgust	81.6	85.2	86.3	93.4

Table II presents the F1 scores for Random Forest, Naïve Bayes, Support Vector Machine (SVM), and a proposed Convolutional Neural Network (CNN) in classifying various emotions. Each row represents a specific emotion, and each column represents a model. The F1 score is a measure of a model's accuracy that considers both precision and recall. The table demonstrates the performance of each model in terms of F1 scores for different emotions. The proposed CNN consistently achieves the highest F1 scores across multiple emotions, with notable performances such as 91.1% for Happy and 94.2% for Angry, while SVM also demonstrates competitive performance. These results underscore the effectiveness of CNN in accurately classifying emotions, suggesting its potential for enhancing emotion recognition systems. Figure 8 shows the accuracy of proposed CNN.

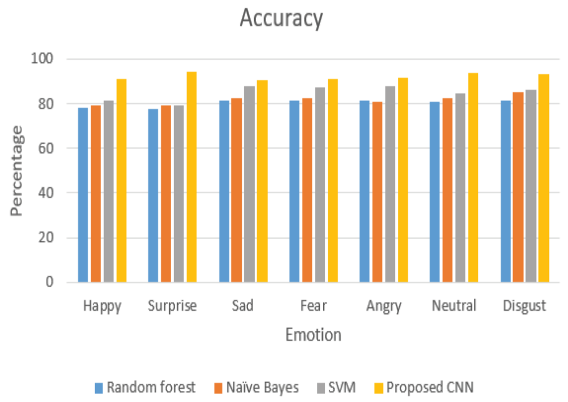

Fig. 8. Accuracy of Proposed CNN

TABLE II.
F1 SCORE

Emotion	Random forest	Naïve Bayes	SVM	Proposed CNN
Happy	81.6	85.3	82.3	91.1
Surprise	73.6	79.6	81.6	91.5
Sad	85.6	88.3	86.1	93.5
Fear	84.6	82.3	87.2	91.3
Angry	71.2	78.9	79	94.2
Neutral	79.9	84.5	89.2	90.3
Disgust	85.6	84.9	88.2	92.3

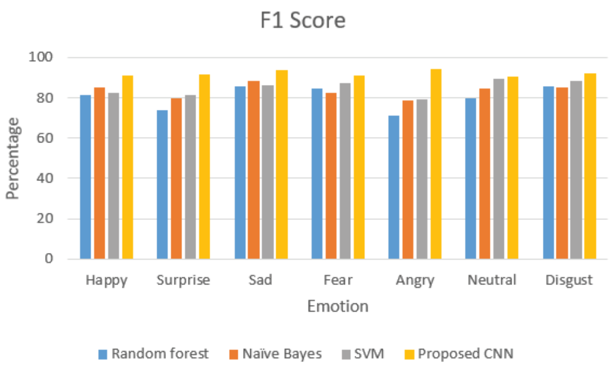

Fig.9. F1-score of Proposed CNN

Table III illustrates the precision scores for Random Forest, Naïve Bayes, Support Vector Machine (SVM), and a proposed Convolutional Neural Network (CNN) in classifying different emotions. Each row corresponds to a specific emotion, and each column represents a model. Precision is a metric that quantifies the accuracy of positive predictions made by a model. The table reveals the precision performance of each model across various emotions. The proposed CNN consistently demonstrates high precision scores, with notable achievements such as 91.2% for Happy and 95.1% for Fear, while Naïve Bayes also displays competitive performance. These findings underscore the effectiveness of CNN in accurately identifying emotions, suggesting its potential for enhancing emotion recognition systems.

TABLE III.
PRECISION

Emotion	Random forest	Naïve Bayes	SVM	Proposed CNN
Happy	82.6	81.6	77.6	91.2
Surprise	82.6	85.2	81.6	91.6
Sad	80.6	81.6	81.6	92.6
Fear	79.2	86.1	81.6	95.1
Angry	78.2	87.2	80.6	94.1
Neutral	81.6	82.6	84.6	91.2
Disgust	81.6	80.6	78.8	90.7

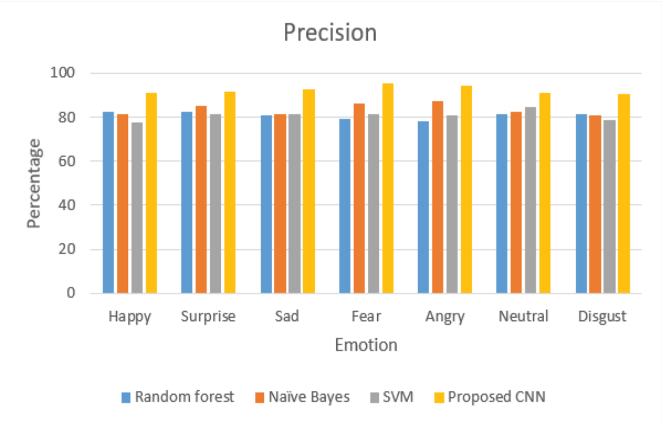

Fig. 10. Precision of proposed CNN

V. CONCLUSIONS

In conclusion, our study presents a customized Convolutional Neural Network (CNN) architecture tailored for emotion detection from facial expressions. Leveraging datasets like CK+, FER2013, and JAFFE, we employed preprocessing techniques to enhance model stability and diversity. The proposed CNN architecture, comprising convolutional and fully connected layers, demonstrated exceptional performance in feature extraction and classification. Experimental results, including accuracy, F1 score, and precision, consistently showcased the superiority of the proposed CNN over traditional machine learning methods. These findings underscore the potential of CNNs in advancing emotion recognition systems, with promising applications across diverse fields such as marketing research and virtual reality. Moving forward, further refinement and validation of our CNN architecture could enhance its efficacy and broaden its utility in real-world applications requiring accurate emotion detection from facial expressions.

REFERENCES

[1] S. Li and W. Deng, "Deep facial expression recognition: a survey," IEEE transactions on affective computing, vol. 13, no. 3, pp. 1195–1215, 2022.

[2] C. Han, L. Zhang, Y. Tang et al., "Understanding and improving channel attention for human activity recognition by temporal-aware and modality-aware embedding," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–12, 2022.

[3] H. Jung, S. Lee, S. Park et al., "Development of deep learning-based facial expression recognition system," in Proceedings of the 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp. 1–4, IEEE, Mokpo, South Korea, January 2015.

[4] Borgalli, Rohan & Surve, Dr. (2022). Deep Learning Framework for Facial Emotion Recognition using CNN Architectures. 10.1109/ICEARS53579.2022.9751735.

[5] Ali, Y., Farooq, A., Alam, T. M., Farooq, M. S., Awan, M. J., & Baig, T. I. (2019). Detection of Schistosomiasis Factors Using Association Rule Mining. IEEE Access, 7, 186108-186114. doi:10.1109/access.2019.2956020

[6] Gupta, M., Jain, R., Arora, S., Gupta, A., Awan, M. J., Chaudhary, G., & Nobanee, H. (2021). AI-enabled COVID-19 outbreak analysis and prediction: Indian states vs. union territories. Computers, Materials and Continua, 67(1)

[7] Alam, T. M., & Awan, M. J. (2018). Domain analysis of information extraction techniques. International Journal of Multidisciplinary Sciences and Engineering, 9, 1- 9.

[8] Samara, A., Galway, L., Bond, R., & Wang, H. (2019). Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), 2175-2184.

[9] Liliana, D. Y., Basaruddin, T., Widyanto, M. R., & Oriza, I. I. D. (2019). Fuzzy emotion: a natural approach to automatic facial expression recognition from psychological perspective using fuzzy system. Cognitive processing, 20(4), 391-403.

[10] Tsai, H.-H., & Chang, Y.-C. (2018). Facial expression recognition using a combination of multiple facial features and support vector machine. Soft Computing, 22(13), 4389-4405.

[11] Salmam, F. Z., Madani, A., & Kissi, M. (2016). Facial expression recognition using decision trees. Paper presented at the 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV).

[12] Nair, P., & Subha, V. (2018). Facial expression analysis for distress detection. Paper presented at the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA).

[13] Awan, M., Rahim, M., Salim, N., Ismail, A., & Shabbir, H. (2019). Acceleration of knee MRI cancellous bone classification on google colaboratory using convolutional neural network. Int. J. Adv. Trends Comput. Sci, 8, 83-88.

[14] Khorrami, P., Paine, T., & Huang, T. (2015). Do deep neural networks learn facial action units when doing expression recognition? Paper presented at the Proceedings of the IEEE international conference on computer vision workshops.