International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

Detecting Several Facial Emotions Using CNN Algorithm

Dr. B. Rajesh Kumar[1]
Department of Software Systems
Sri Krishna Arts and Science College

Kavya S[2]
Department of Software Systems
Sri Krishna Arts and Science College

***Abstract*— Facial expression is a vital component of nonverbal indicators in interpersonal interactions. The Cognitive Emotions AI system determines an individual's emotional state in this way. Our main objective is to develop a reliable system that can identify and track human emotions in real-time programs. There are many universal emotions, such as anger, sadness, love, surprise, fear, scorn, and indifference. The Viola Jones's algorithm is used to extract a face's Haar-cascade features, which completes the facial identification process. Artificial intelligence algorithms are then used to confirm and recognize the emotion. As soon as the system receives a snapshot or frames as input, the model preprocesses it, selects its attributes, and then predicts the image's emotional state.**

Keywords:* *Facial Expressions, Deep Learning, Image Pre-processing, Emotional Intelligence

I. INTRODUCTION

People commonly use their facial expressions to convey their feelings. One of the easiest, most effective, and natural ways for people to express their intentions and feelings is through their looks. People frequently have to conceal their emotions, such as while hospitalized or when they are ill, therefore having a greater awareness of others' moods would lead to more fruitful conversations. artificial intelligence science. Artificial Intelligence (AI) is the ability of a machine to "decide or act humanely or rationally". Machines can now process the massive amount of data in real time and respond appropriately. However, these machines' emotional intelligence (EI) was always lacking, even with high IQs (intellectual quotients). As technology advances and the world becomes increasingly virtual, there are worries that we would lose our ability to connect and communicate with one another. However, what if our gadgets could replace such interactions? Its ability to develop a system that understands the feelings of people is the main challenge of the day. Studies in emotional analysis and research are not new. According to Charles Darwin's book, "Facial emotions of emotion are ubiquitous, yet learned individually in each culture". Paul Ekman, a psychologist, carried out the fundamental research on the subject. He was the very first to classify emotions.

1. Enforcing the eyelids and raising the outer edges of one's mouth into a noticeable smile are signs of joy or happiness.

2. The expressions of surprise include a little drop in jaw, arched brows, and wide, whiter-looking eyes.

3. A sad expression would have drooping eyelids, lowered brows to inner corners, and lowered lip corners.

4. The expressions of anger are lowered eyebrows, pursed lips, and enlarged eyes.

5. Rising the upper lip, crowing through the nasal bridge, and puffing up the cheeks are signs of disgust.

6. Fear is represented by the lips spreading horizontally, the eyes widening, and the top eyelids raising.

7. Disdain is shown by lifting half of the upper lip and frequently leaning the head slightly back.

It is believed that these emotions are the most prevalent. Researchers and programmers have created artificial intelligence to create systems that can recognize and react to emotions in addition to thinking and acting like people. Despite substantial variations in emotional recognition, humans are universally equipped to identify emotions across cultural boundaries. The only way we can better our relationships with other humans and the robots around us is if we can educate them to identify our emotions. Providing individualized customer experiences that can improve lives is the aim of the investigations for this project. We decided to test a handful of these because several multinational companies have already started investigating emotion recognition. The goal is to understand the variety of features they offer, the level of identification accuracy, and the breadth of use cases that they may be applied to. Major services were evaluated after being put to the test on a variety of equipment, operating systems, etc. Motivated by the provided answers, we decided to investigate emotion recognition through neural network experiments. Thanks to recent advancements in machine learning and open resource tools like Google's Tensorflow, creating and training models is now lot simpler. One of the largest obstacles is getting the dataset required to train the model. The "Challenges in Representing Education: facial expression Prediction Challenge" on Kaggle was then found to be an emotion identification challenge. 56 teams competed in this competition, which used a variety of techniques like as neural networks, SIFT, Haar, and Hog. The FER2013 dataset that was made available for this challenge can be downloaded for free. 35887 images total, comprising 5124 images of "Anger," 638 images of "Disgust," 3442 images of "Fear," 6752 images of "Happiness," 4125 images of "Sadness," 4002 images of "Surprise," and 7254 images of "Neutral." A model was developed using Tensorflow and Keras, accounting for the latest advancements in neural networks. It is a lightweight, real-time model that can be used in systems with less hardware. A comparable degree of accuracy is obtained during testing using live video streams.

II. RELATED WORKS

The technique by which a person uses their facial muscles to convey their emotions is known as facial expression. It gives information on how the person is feeling. Emotion is a mental state that an individual goes through. He is reacting inside to what is happening to him externally. Many times, a person's facial features can provide insight into their mental health. Facial analysis has application in many industries, such as robotics, painting, and lie detectors. Improved facial expression detection skills are required in order for an autonomous system to communicate with humans and robots in machine-human cooperative and robot-robot interaction [1]. In [2].

Considering that the subject subject matter has been ongoing for years, the advancements made through study on this topic are remarkable [3] [4]. Fluctuation in rates of identification among classes is a concern for most research because students have a lower rate of understanding for emotions such as fear and revulsion [5] [6].

The goal of this research is to develop a face expression identification system that can identify seven distinct emotional groups in an image. It will also cover how to maintain a respectable and almost equal recognition rate for each class while improving validation accuracy relative to other current systems. A small number of research have employed CNN for facial expression recognition; however, the bulk of these studies suffer from uneven recognition rates across classes, as evidenced by the lower classification rates for annoyance and dread [7]. "Real-Time Human emotion identification using Deep A convolution Neural Network" was published in 2021 by Adil Mohammed Adnan and associates. Six distinct facial emotions were identified in this study using a CNN: surprise, fear, fury, sadness, and happiness. Using a test batch of 1,000 images, the CNN achieved a precision of 80%. It was trained using a data set of 4,000 face shots. "Deep Intelligence for real-time Emotions Recognition of Achievement: A Survey" was released by Zhang et al. (2018). This survey research provides an overview of the use of deep learning techniques for real-time emotion recognition. It covers the challenges and limitations of this approach in addition to the many deep neural network variants that have been successfully used for this goal. Chen et al. published "Convolutional neural network Neural Networks for immediate Emotion Recognition of Excellence A Review" in 2019. This review research discusses the application of CNN algorithms. for real-time emotion recognition. It examines how several CNN architectures have been used to solve this issue and how effectively these models function across a range of datasets.

"Actual Time Recognition of Emotions Using Convolutional Neural Networks with Deep Roots and Transfer Learning" was published by Zhang et al. in 2020. In this work, transfer learning was used to improve a CNN's real-time emotional recognition performance. The CNN was trained using a sizable dataset of face image data, and its weights were adjusted using a smaller sample of emotion-labeled picture data. This approach achieved 90% accuracy on a practice set of 2,000 photographs. A CNN architecture that is not dependent on manually created features was proposed by Burkert et al. The first part of this architecture consists of four sections where a convolutional layer is used to automatically preprocess the pictures. The pictures in the second section are sampled by the pooling layer down. The fundamental element of this design inspired by GoogleNet is the FeatEx, which is the next item. To do the classification, the features that were acquired by concatenating two FeatEx blocks are subsequently sent into an entirely linked layer. Its validity is demonstrated by visualizing the complex characteristics of different levels. The MMI & CK+ benchmark data sets are used for evaluation. They analyzed seven different emotions (anger, disgust, fear, gladness, sorrow, and surprise) using the CK+ dataset, and they achieved a 99.6% recognition rate.

III. PROPOSED SYSTEM

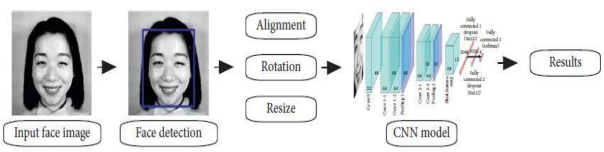

Figure 3.1 Proposed Overall Architecture

In the picture 3.1 is representing the way of how images are focused from normal images to predictive image using CNN algorithm. The CNNs are perfect for instantaneous fashion emotion recognition since they can be utilized on hardware systems with sufficient processing power to analyze images rapidly. Additionally, CNNs can withstand positional, occlusion, and illumination variations with some degree of resilience; these factors can all affect how accurately CNNs recognize emotions. The proposed minimal CNN model consists of four convolutions, one completely interconnected layer, and the output layer. CNNs are among the deep learning algorithms designed mostly for image processing applications. Their ability to discern features from images that are relevant to the task at hand allows them to operate. In the realm of emotion recognition, CNN simulations can be trained to distinguish characteristics from pictures of facial features that are linked to different emotions.

The following procedures would be included in the suggested system for a deep CNN-based real-time emotion recognition system:

DATA COLLECTION:

To compile a dataset of facial images with corresponding emotion classifications, multiple approaches can be considered. One method involves utilizing existing publicly available datasets while ensuring diversity in terms of demographics, expressions, and lighting conditions. Additionally, leveraging crowdsourcing platforms to collect labeled facial images can enhance dataset diversity and size. It is crucial to prioritize ethical considerations such as obtaining consent and ensuring data privacy throughout the collection process.

IMAGE PREPROCESSING:

Preprocessing of facial images involves various steps aimed at enhancing the quality and uniformity of the dataset. These steps may include resizing images to a standard resolution, adjusting lighting conditions, removing background noise, and aligning facial landmarks. By standardizing these aspects, the subsequent stages of feature extraction and model training can be more effective.

FEATURE EXTRACTION:

Feature extraction is a critical stage where relevant information is extracted from facial images to facilitate emotion recognition. Convolutional Neural Networks (CNNs) are commonly employed for this task due to their ability to automatically learn hierarchical features from data. By training a CNN on a large and diverse dataset of labeled facial images, the network can capture discriminative features associated with different emotions.

MODEL TRAINING:

Once features are extracted, they are used to train a CNN for emotion recognition through supervised learning. This involves optimizing the network's parameters using gradient-based optimization algorithms such as stochastic gradient descent. The training data comprises pairs of facial images and corresponding emotion labels, allowing the CNN to learn to associate specific features with different emotions.

MODEL EVALUATION:

After training the CNN, it is essential to evaluate its performance on a held-out validation set to assess generalization ability and prevent overfitting. Various metrics such as accuracy, precision, recall, and F1 score can be used to evaluate the model's performance across different emotion categories. Additionally, techniques like cross-validation can provide insights into the model's robustness and reliability.

REAL-TIME IMPLEMENTATION:

To deploy the CNN for real-time emotion recognition, it needs to be implemented on hardware platforms capable of supporting inference tasks efficiently. This could involve deploying the model on embedded devices, laptops, or smartphones, depending on the specific application requirements. Optimizing the model for inference speed and memory efficiency is crucial for achieving real-time performance in practical settings. Additionally, considerations such as power consumption and latency should be taken into account when deploying the system in real-world scenarios.

CNN *ALGORITHM*

- Data Collection.
- Install needed packages.
- Import needed packages.
- Initialize image data
- creator with rescaling.
- Preprocess all train images.
- Preprocess all Test Image.
- produce CNN Model Structure.

Action for Accuracy and loss Graph A convolutional neural network-Bi-directional Long Short- Term Memory (CNN- BiLSTM) algorithm is used to dissect the emotion through speech and images of learners in the intelligent literacy terrain. The machine learning algorithm grounded on bracket improvement is used for speech emotion recognition CNN is veritably effective for emotion recognition tasks (18). They can prize features from input images, and also use these features to train a classifier. Once the classifier is trained, it can be used to fete images. The advantage of using a CNN is that it's suitable to learn complex patterns in input images.

Step 1: Input layer: obtains the image data, which is commonly displayed as a three-dimensional in nature tensor comprising channels (color or intensity), width, and height.

Here's a breakdown of the components of this tensor:

1.Channels: The first dimension of the tensor represents the channels. Channels refer to different aspects of the image data. In the case of color images, the channels typically represent the color information, such as red, green, and blue (RGB). Each channel contains intensity values corresponding to the presence of that color in each pixel of the image. For grayscale images, there is only one channel representing the intensity of each pixel.

2. Width: The second dimension of the tensor represents the width of the image. It indicates the number of pixels horizontally across the image. Each row of pixels in the image is represented in this dimension.

3. Height: The third dimension of the tensor represents the height of the image. It indicates the number of pixels vertically across the image. Each column of pixels in the image is represented in this dimension.

Step 2: Convolutional layers: At the core of convolutional layers is the convolution operation. This operation involves moving a small matrix called a filter or kernel across the input image and computing the dot product between the filter and the local region of the image it is currently covering. This process is performed iteratively across the entire image. The filter serves as a feature detector, capturing patterns or features from the input data.

Step 3: Pooling layers: Convolutional layers are often followed by pooling layers, which down sample the feature maps produced by the convolutional layers. Pooling helps in reducing the spatial dimensions of the feature maps while retaining the most important information. Common pooling operations include max pooling, which retains the maximum value within each pooling region, and average pooling, which computes the average value within each region. By down-sampling the map of features from the convolutional layers, the size of the data can be decreased. This increases computing efficiency and facilitates control over fitting.

Step 4: Activation functions: Non-linear Activation: After the convolution operation, a non-linear activation function such as ReLU (Rectified Linear Unit) is typically applied element-wise to introduce non-linearity into the network. This helps

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

the network learn complex relationships between features and enables it to approximate non-linear functions. Give the network non-linearity so that it can learn intricate correlations between features.

Step 5: Fully connected layers: These layers function similarly to conventional neural networks in that they incorporate features from earlier layers to produce the desired final output, which may be an item detection box, classification, or another outcome.

IV. RESULT AND DISCUSSION

Using movements of the face and deep learning techniques, it has been demonstrated that convolutional neural networks with deep learning (CNNs) are highly successful at real-time emotion recognition. Six different facial emotions were trained to identify a CNN in an experiment conducted through Adil Mohammed Adnan et al. (2021): happiness, sadness, rage, fear, surprise, and neutral. Using a test set of 1,000 photos, the CNN attained an accuracy of 80%. It was trained using an array of 4,000 facial photographs. The study also discovered that the CNN exhibited a reasonable degree of robustness against changes in stance, illumination, and occlusion. This is crucial for actual time emotion recognition since, in certain situations, the face can look partially veiled or under low illumination. According to the study's findings, CNNs are a viable tool for recognizing emotions in real time utilizing deep learning techniques and facial expressions. Nevertheless, before CNNs are extensively used for this purpose, a few issues still need to be resolved. The fact that CNNs need a lot of data for learning presents one difficulty. The collection of this data can be costly and time-consuming, particularly if emotion annotations need to be added to the data. Furthermore, the computational cost of training and deploying CNNs can be a deterrent for some applications.

Figure 4.1 Angry Face Prediction Result Screenshot for Video with Accuracy

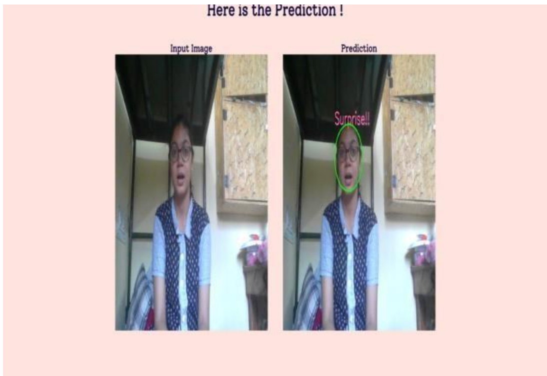

Figure 4.2 Screenshot of the surprise face prediction result from the video, with accuracy level

International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

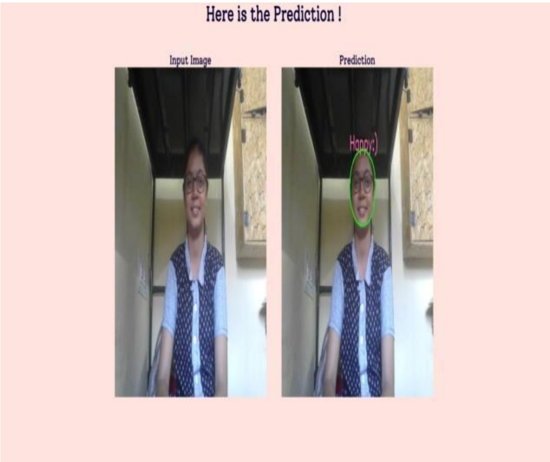

Figure 4.3 Happy Face Prediction Result Screenshot with Accuracy Level for Video

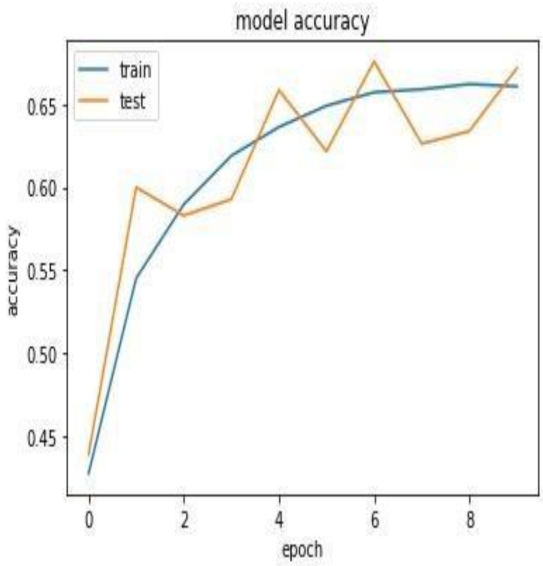

Figure 4.4 Plot for Predictive Model Accuracy

V. CONCLUSION

Deep convolution neural network models (CNNs) are a viable option for immediate form emotion recognition using facial movements and deep learning techniques. It has been shown that CNNs can reliably detect emotions from face images and are not too sensitive to changes in illumination, location, or occlusion. There are still certain problems to be resolved before CNNs are extensively employed for real-time emotion recognition. One challenge is that CNNs require large amounts of data in order to train. It can be expensive and time-consuming to gather this data, especially if emotional remarks are required. Furthermore, CNN deployment and training can be computationally expensive, which may be a barrier for some applications. Another challenge is that CNNs could be vulnerable to the choice of hyperparameters. This suggests that the model's training parameter values may have an impact on the CNN's performance. As such, determining the optimal set of hyperparameters for a specific application can be difficult.

REFERENCES

[1] A. Dhall, O. V. R Murthy, R. G.J., and T.Gedeon. Video and image-based emotion recognition challenges in the wild: Emotiw 2021. In the 2015 ACM on international Conference on Multimodal interaction (2021), ACM, pp. 423–426.

[2] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., et al. Tensor- flow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2021).

[3] Abaya, W. F., Basa, J., Sy, M., Abad, A. C., and Dadios, E. P. Low-cost smart security camera with night vision capability using raspberry pi and opencv. In Humanoid, Nanotechnology, Information Technology, Communication and Control,

Environment and Management (HNICEM), 2021 International Conference on (2021), IEEE, pp. 1–6.

[4] Black, M. J., and Yacoob, Y. Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion. In Computer Vision, 1995. Proceedings., Fifth International Conference on (1995), IEEE, pp. 374–381.

[5] Bradski, G., et al. The opencv library. Doctor Dobbs Journal 25, 11 (2021),120–126.

[6] Bradski, G., and Kaehler, A. Learning OpenCV: Computer vision with the OpenCV library." O'Reilly Media, Inc.", 2021.

[7] Burkert, P., Trier, F., Afzal, M. Z., Dengel, A., and Liwicki, M. Dexpression:

[8] M. Liu, S. Li, S. Shan, R. Wang, "Deeply learning deformable facial action parts model for dynamic expression analysis", in Proceedings of the IEEE International Conference on Computer Vision.

[9] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D. Lee, "Challenges in representation learning: A report on three machine learning contests", Neural Networks, vol. 64, pp. 59-63, 2020.

[10] Y. Sun, X. Wang, X. Tang, "Deep learning face representation by joint identification-verification", in Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2021.