EMOTION RECOGNITION USING CNN

[1]Chaithra K, [2]Banda Preeta, [3]Aanchal Jain, [4]Haasini T S, [5]Amogh G Padukone, [6]Nazmin Begum
[1,2,3,4,5,6]Department of computer science, Dayananda Sagar University, Banglore, India
[1]chaithra60120@gmail.com, [2]preetabgoud@gmail.com, [3]aanchaljain1611@gmail.com,
[4]haasini00@gmail.com, [5]amoghgopal@gmail.com, [6]nazmin-cse@dsu.edu.in

***Abstract*— Recognition of human emotions plays a vital role in understanding of the human brain when exposed to different situations. Although humans can recognize a person's facial expressions without problem or delay, machine expression identification remains a challenge. Several studies on feature extraction, face identification, and expression classification approaches have been published in recent years, but developing an automated system that does this task is difficult. We have implemented a machine learning model using fer2013 dataset(implemented on google colab/jupyter). This model has been developed using deep learning techniques such as convolutional neural networks(CNN) and python libraries(openCV). In comparison to other image classification techniques, CNNs require very little preprocessing. The model has been trained with input as images which produces output in the form of text. One of the most important findings of our study was that most of the models used CNN approach with an accuracy of 75% to 90%**

***Keywords*— Convolutional Neural Networks, Deep Learning, Emotion Recognition, Fer2013, Machine Learning**

I. INTRODUCTION

Human facial expressions are considered as critical components in understanding one's emotions. Human emotions and intentions are expressed through facial expressions and these facial expressions convey non-verbal cues, which play an important role in interpersonal relations. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioural science and in clinical practice.

Convolutional neural networks (CNN) is one of the fundamental deep learning techniques which have provided support and platform for analysing visual imagery. CNN is a well-known deep learning algorithm that has achieved outstanding performance in image processing by applying a filter to an input and resulting in an actuation. Convolution neural network model takes an image dataset as input. It is trained by using multiple training datasets.

Convolutional layers are the layers in a deep CNN that apply filters to the original picture or other feature maps. The majority of the network's user-specified parameters are located here. The image is broken into three colour channels: Red, Green, and Blue. Each of these colour channels corresponds to a pixel in the image. The computer then determines the size of the image by recognising the value associated with each pixel.

The real-time emotion recognition can be achieved using visual-based techniques or sound-based techniques. This paper will mostly focus on image-based emotion recognition using deep learning. The fundamental piece of the message is the facial expression, which establishes 55% of the general impression.

II. LITERATURE SURVEY

1. Authors [1] here have proposed the Venturi architecture, in terms of training accuracy, training loss, testing accuracy, and testing loss it is being compared to the performance of two current deep neural network architectures. The Karolinska Directed Emotional Faces dataset, which contains 4900 images of human facial emotions, is used in this study. The features from the photos were convoluted using two layers of feature maps, and then passed on to a deep neural network with up to six hidden layers. The proposed Venturi architecture enhances accuracy greatly when compared to modified triangular and rectangular layouts. The show Modified Triangle Architecture features modified triangular architecture. When compared to the modified triangular and rectangular architectures, the proposed Venturi architecture improves accuracy significantly. With a training accuracy of 98.29%, the Modified Triangular Architecture had the lowest results. In comparison to the other two architectures, the Venturi Architecture has the best testing or validity accuracy (86.78%), while the Rectangular Architecture has the weakest validity accuracy (79.61%). The proposed venturi architecture outperforms the modified triangular architecture by 4.08 % and the rectangular architecture by 7.17 % of accuracy.
2. Authors [2] have utilised convolutional neural networks (CNNs) model in this study to train on grayscale photos

 DOI : 10.35291/2454-9150.2022.0020

from the fer2013 dataset. To acquire the best accuracy, they experimented with different depths and max pooling layers, eventually achieving 89.98 percent accuracy. They have employed strategies like dropout to fight overfitting, as well as examining the performance of different network topologies like shallow networks and contemporary deep networks in understanding human emotion. They also demonstrate a web-camera implementation of real-time emotion identification that yields correct results for numerous faces at the same time. As a result of comparing the accuracy at different depths it has achieved 89.98% as maximum accuracy using swish function.

3. In this research the authors [3] investigates the use of visual and textual signals in speech emotion identification using a hybrid fusion method known as multimodal attention network (MMAN). They propose cLSTM-MMA, an unique multimodal attention mechanism that allows attention across three modalities while selectively fusing data. In late fusion, cLSTM-MMA is fused with other uni-modal subnetworks. The results reveal that visual and textual signals help speech emotion detection greatly, and the proposed cLSTM-MMA alone is as accurate as existing fusion approaches, but with a much more compact network topology. On the IEMOCAP database for emotion recognition, the proposed hybrid network MMAN delivers state-of-the-art performance. On the dataset IEMOCAP, they claimed MMAN obtains a state-of-the-art performance of 73.94%.

4. Happy Emotion Recognition is the name of the model proposed by the authors[4] , which employs the 3D hybrid deep and distance features (HappyER-DDF) method to improve accuracy by leveraging and extracting two separate types of deep visual features. To extract dynamic spatial-temporal information across sequential frames, they used a hybrid 3D Inception-ResNet neural network with long-short term memory (LSTM). Second, they identified the features of facial landmarks and calculated the distance between each landmark and a reference point on the face (e.g., the nose peak) to capture changes as a person smiles (or laugh).The trials were then carried out on three unconstrained video datasets using both feature-level and decision-level fusion approaches. The suggested HappyER-DDF technique detects happy emotions with an accuracy of 95.97 % for the AM-FED+ dataset, 94.89 % for the AFEW dataset, and 91.14 % for the MELD dataset, according to testing using three unconstrained video datasets.

5. On numerous datasets, including FER-2013, CK+, FERG, and JAFFE, authors [5] have employed a convolutional network that was able to focus on essential areas of the face and produced considerable improvements over earlier models. They also employed a visualisation technique that allowed them to discover relevant face regions based on the classifier's output to detect distinct moods. For recognising facial emotions, careful attention to specific regions is critical, which has allowed neural networks with fewer than ten layers to compete with (and even exceed) much deeper networks in emotion identification. Through experimental results, they showed that different emotions are sensitive to different parts of the face.

6. Facial expressions are most significant parameters for non-verbal communication. Here the authors [6] of this paper explores the performance of CNN with Radial Basis Function for expression recognition. The psychological conditions of the patients is additionally easily analysed and also the remedial actions could also be dispensed at a faster rate. Experimental results helps us to know that combination of methods will lead to better accuracy. The face image dataset employed within the experiment is FER 2013 which consists of 37,000 samples(28,000 labelled samples within the training set, and 3,500 labelled samples in development set, and 3,500 images in test set) such as human faces with seven countenance like surprise, fear, disgust, neutral, anger, happiness, and sadness. The mixture of CNN with RBF provides better accuracy of 95.4%. The output results of enhanced CNN had an accuracy rate of 98%.

7. In this paper the authors [7] have considered the FER2013 dataset and performed it with different machine learning models like AdaBoost, CNN ,DNN, Logistic Regression. The CNN model was able to distinguish the test samples more easily compared to other models despite not having many training samples to train the model. Here the prediction was based on a confusion matrix which explains the weaker models tended to accidently predict wrong emotions. CNN algorithm was shown superior on the FER2013 dataset accomplishment with the task. It was not only able to remain robust and generalizable in the face of unbalanced training, but it was also able to recognise emotions as well enough without any mistake like other models. CNN model achieved an accuracy of 64% which is as very much good as making human intelligence for performing a limited task.

8. The proposed method by the authors [8] proves that employing a well-trained CNN followed by RNN is equally effective for Video face expression Recognition as for other similar tasks e.g. Action Recognition. the results of attention-based networks and implications in continuous video analysis are studied in future. Videos provide an excellent insight during a lot of domains and supply endless possibilities to develop smart technologies supported beholding. The work

 DOI : 10.35291/2454-9150.2022.0020

provides the bottom ground to increase its application in various important tasks like suspicious behaviour detection, video summarization and emotionally aware interactive e-learning solutions. The trained model was ready to classify test videos with 61% accuracy. This has demonstrated video classification performance on this dataset using only video frames, this may be considered because the first benchmark on RAVDESS video dataset using only visual features.

9. The facial emotion recognition system based on CNN designed by the authors [9] in this particular paper achieves the RMSE index pf 0.0857+0.0064.The cascade classifier is used to capture the face region and then input to the CNN trained by CK+ and Fer2013 expression datasets to obtained the predicted results. The emotion recognition method which is very much stable and will not be affected by the facial physiological variations between the subjects .By considering that it is difficult to classify the valence dimensions of nine grades, the difference in the expression of the adjacent valence dimensions is not obvious, here the method will not directly select the valence grade of maximum probability as the output, but selects the weighted fusion of each effect value and the prediction probability. Some dropout of regularisation are applied since there are some overfitting in the training and testing phase due to insufficient number of facial expression images with the valued dimensions specified. Finally this method can correctly able to identify the different emotional categories .The generalisation ability of the CNN model is improved by the combination of L2 regularisation and the dropout. The CNN network model is optimised by introducing Adam algorithm.

10. Two datasets were employed in this paper were ADFES-BIV and WSEFEP. An algorithm for video-based emotion recognition with no manual design of features employing a DCNN. The authors [10] have proposed a model considered the sight only and achieved a wonderful recognition rate for the ten used emotions. This automatic system utilises the sparse-representation-based classifier and obtains the highest accuracy of 80% by considering the information intrinsically present within the videos. Then WSEFEP dataset was used in this method for testing by merging it with the testing frames taken from the primary dataset. The popularity rate accuracy is 95.12%. DCNN was used as it requires a large amount of training data since the inference accuracy improves by considering more data, and because what really mattered to us here was to recognize facial emotions.

III. ANALYSIS

SL.No	Author and Paper Title	Algorithms	Data Set	Accuracy
1.	Verma, A., Singh, P., & Alex, J. S. R. (2019, June). Modified convolutional neural network architecture analysis for facial emotion recognition.	CNN, DNN	Karolinska Directed Emotional Faces dataset	82.70%
2.	Pathar, R., Adivarekar, A., Mishra, A., & Deshmukh, A. (2019, April). Human emotion recognition using convolutional neural networks in real time	CNN	FER2013	90.22%
3.	Pan, Z., Luo, Z., Yang, J., & Li, H. (2020). Multi-modal attention for speech emotion recognition.	LTSM (MNAN,Fusion method)	IEMOCAP	73.8%
4.	Samadiani, N., Huang, G., Hu, Y., & Li, X. (2021). Happy Emotion Recognition From Unconstrained Videos Using 3D Hybrid Deep Feature	3D INCEPTION-ResNet (3DIR) NEURAL NETWORK	AM-FED+,AFEW and multi-party conversational (MELD),	95.12%
5.	Minaee, S., Minaei, M., & Abdolrashidi, A.(2021). Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors, 21(9), 3046.	CNN	FER2013, CK+, JAFFE, FERG	92.8%
6.	Meryl, C. J., Dharshini, K., Juliet, D. S., Rosy, J. A., & Jacob, S. S. (2020, July). Deep Learning based Facial	CNN (Radial Basis Function)	FER-2013	95.4 %

 DOI : 10.35291/2454-9150.2022.0020

	Expression Recognition for Psychological Health Analysis.			
7.	Gory, S., Al-Khassaweneh, M., & Szczurek, P. (2020, July). Machine Learning Approach for Facial Expression Recognition.	AdaBoost, CNN ,DNN, Logistic Regression.	FER-2013	60% (CNN model)
8.	Abdullah, M., Ahmad, M., & Han, D. (2020, January). Facial expression recognition in videos: An CNN-LSTM based model for video classification	CNN,RNN	RAVDESS video dataset	61%
9.	Liu, S., Li, D., Gao, Q., & Song, Y. (2020, November). Facial Emotion Recognition Based on CNN.	CNN	CK+ and FER-2013	----
10.	Abdulsalam, W. H., Alhamdani, R. S., & Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks	Deep CNN	Amsterdam Dynamic Facial Expression Set Bath Intensity Variations (ADFES-BIV)	95.12%

A CNN is a type of artificial neural network that is used in image recognition and processing the image is specifically designed to process pixel data. CNN is one of the most powerful algorithms for Image processing . It gives a more accurate result. Convolutional neural network is then a feed-forward network because CNN has features parameters that share and dimensionality reduce, because of parameter sharing in CNN, the number of parameters is reduced thus the computations also decreased. The advantage of using a CNN compared to its predecessors is that it can automatically detect the important features without any human supervision. CNN is little dependent on pre-processing, decreasing the needs of human effort developing its functionalities. It is easy to understand the algorithm and fast to implement. While using CNN, the training time is significantly smaller compared to any other model. state the units for each quantity that you use in an equation.

IV. CONSLUION

The expression analysis provided for best results when combined with the CNN model as it provides a method for image processing. This enables us to get the best results out of all the techniques. Out of all the combinations the CNN radial basis function provided for the best result with an accuracy of 95.4%. Overall we can observe that with different combinations of the CNN model we have an approximate accuracy of about 78% to 80%.

REFERENCES

[1] Verma, A., Singh, P., & Alex, J. S. R. (2019, June). Modified convolutional neural network architecture analysis for facial emotion recognition. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)* (pp. 169-173). IEEE.

[2] Pathar, R., Adivarekar, A., Mishra, A., & Deshmukh, A. (2019, April). Human emotion recognition using convolutional neural networks in real time. In *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)* (pp. 1-7). IEEE.

[3] Pan, Z., Luo, Z., Yang, J., & Li, H. (2020). Multi-modal attention for speech emotion recognition. *arXiv preprint arXiv:2009.04107*.

[4] Samadiani, N., Huang, G., Hu, Y., & Li, X. (2021). Happy Emotion Recognition From Unconstrained Videos Using 3D Hybrid Deep Features. *IEEE access*, *9*, 35524-35538.

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

[6] Meryl, C. J., Dharshini, K., Juliet, D. S., Rosy, J. A., & Jacob, S. S. (2020, July). Deep Learning based Facial Expression Recognition for Psychological Health Analysis. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 1155-1158). IEEE.

[7] Gory, S., Al-Khassaweneh, M., & Szczurek, P. (2020, July). Machine Learning Approach for Facial Expression Recognition. In *2020 IEEE International Conference on Electro Information Technology (EIT)* (pp. 032-039). IEEE.

[8] Abdullah, M., Ahmad, M., & Han, D. (2020, January). Facial expression recognition in videos: An CNN-LSTM based model for video classification. In *2020 International Conference on Electronics, Information, and Communication (ICEIC)* (pp. 1-3). IEEE.

[9] Liu, S., Li, D., Gao, Q., & Song, Y. (2020, November). Facial Emotion Recognition Based on CNN. In *2020 Chinese Automation Congress (CAC)* (pp. 398-403). IEEE.

[10] Abdulsalam, W. H., Alhamdani, R. S., & Abdullah, M. N. (2019). Facial emotion recognition from videos using deep convolutional neural networks. *International Journal of Machine Learning and Computing*, *9*(1), 14-19.

 DOI : 10.35291/2454-9150.2022.0020