OPEN ACCESS
International Journal of Engineering Innovations in Advanced Technology
ISSN: 2582-1431 (Online), Volume-6 Issue-2, June 2024

A Deep Learning Facial Expression Recongnition Based Scoring System For Restaurants

Venkata Siva Koti Rambabu Vudatha
M.Tech Scholor, Department of CSE, GVR&S College of Engineering & Technology,Guntur, India.
Email: vudatharambabu@gmail.com

Dr K Ramesh Babu
Professor, Department of CSE, GVR&S College of Engineering & Technology,Guntur, India.
Email: dr.rbk.cse@gmail.com

Abstract- **Facial expressions are the primary, immediate, and instinctive means by which humans communicate their ideas and emotions. Automatic facial expression identification is a captivating and challenging field that has a significant impact on the applications of human-computer interaction due to its intricate nature and diverse range of expressions. Academics are intrigued by the topic due to its extensive array of applications, including cognitive science, health care, and video conferencing, among others. In this work, we have devised three distinct methodologies for fer and compared them with existing models documented in the literature. The Deep Layered Representation (DLR) is a novel convolutional neural network architecture designed specifically for facial emotion recognition. Our technique, which utilizes a multilayer deep neural network, has been applied to the FER2013 dataset from Kaggle. The data were analyzed using five generalized activation functions, namely Elu, ReLu, Softplus, Sigmoid, and Selu, in the final dense layer. By using the same dataset, we have conducted a comparison between our top two models based on activation functions and models based on GoogLeNet and VGG16 + SVM. The results have shown that our models have achieved greater accuracy. As an alternative method, we propose using a deep learning framework that employs transfer learning to comprehend facial expressions. This approach adds additional layers to the existing VGG16 model, which has been updated during training. The recommended model is verified using the benchmark datasets JAFFE and CK+. Many academics have used the well recognized facial expression datasets JAFFE (Japanese Female Face Expression) and Extended Cohn-Kanade (CK+) in their studies. The proposed model has been shown to outperform existing techniques, achieving a 94.8% accuracy rate on the CK+ dataset and a 93.7% accuracy rate on the JAFFE dataset. We have implemented the recommended approach on Google Colab-GPU. Currently, automated and unmanned restaurants are more prevalent in industrialized countries because to a shortage of human resources to gather feedback from customers on the food's quality and service. In order to streamline this procedure, the author has introduced the notion of a "Deep Learning Facial Expression Recognition Based Scoring System For Restaurants." In this system, customers evaluate their food and submit a photograph, and the application assesses the customer's degree of contentment by analyzing their facial expressions. We are using the CNN (Convolution Neural Networks) machine learning methodology to extract facial expressions from photographs. This system has the capability to recognize three specific facial emotions from a single image, which are neutral, unsatisfied, and delighted.**

Keywords: **Facial expressions, JAFFE, VGG16, and SVM.**

I. INTRODUCTION

Facial expressions are crucial in determining an individual's emotional state. Because emotions are sensitive and observable, an observer can determine an individual's emotional state. All that the user has entered is the word "state." The majority of individuals agree that the most accurate method of predicting a range of emotions is facial expression ecognition. Facial expression is one of the most distinctive nonverbal communication channels and is essential to interpersonal interactions [1].In the last several years, there has been a lot of effort in the field of facial expression recognition (FER) [2]. The FER system attempts to recognize and classify emotions such as happiness, anger, sorrow, astonishment, disgust, and contempt using face photographs as input [3]. Automated facial expression recognition (FER) is used in many applications nowadays, including human-computer interface systems (HCI), social robots, interactive gaming, and data-driven animation for neuromarketing [4]. These days, face expression recognition using machine learning algorithms is a challenging problem. For reliable face recognition, having good facial features is essential. Reliable face recognition depends on improved facial images obtained via effective pre-processing image enhancement techniques. A crucial first step in face Expression Recognition (FER) is the extraction of meaningful characteristics from the face picture [5]. Appearance-based techniques and geometry-based approaches are the two primary categories of feature extraction algorithms [6]. Common names for these two techniques are form and attribute qualities related to a material's feel. When features based on appearance are to be retrieved, the whole face or specific regions of the face are taken into consideration [7]. On the other hand, landmark points—pre-defined geometric points—are used in features that rely on geometry.

The forms and positions of immovable features such as the mouth, nose, eyebrows, and eyes are represented by the points [8]. Photographic feature extraction seems to be enhanced by content-based feature extraction using Scale Invariant Feature Transform (SIFT), Oriented FAST, and Rotated BRIEF (ORB) [9], [10]. Furthermore, the facial detection method is essential for recognizing facial expressions. By using complex face recognition algorithms, the Facial Expression Recognition (FER) system's efficiency may be increased [11]. A precise drawing of the face is necessary for effective categorization. It is challenging to identify which face characteristics best convey a certain emotional quality [12]. Some academics employ shape models that integrate distinctive face characteristics to characterize facial emotions.

It is difficult to predict face landmarks in real-world scenarios, because the previously discussed methods depend on precise and trustworthy forecasting [13]. The term *Facial Emotion Recognition* is shortened. Deep learning is shown great promise in feature learning for FER. Unfortunately, owing to a few special aspects of deep learning, existing datasets are insufficient for deep learning to achieve human levels of training and accuracy [14]. An image acquisition stage that recognizes the facial region in the incoming video or picture is one part of the FER system. The specified area of the image is cropped to remove the other portions [15]. Extracting several properties from the area of interest is the next stage. To increase the performance of the framework, it is crucial to extract its core characteristics. The framework can only classify emotions based on the collected characteristics. The requirements for the next phase should be simple to comprehend and apply [16].

To satisfy the requirements, the features must be able to accurately represent changes in intensity in a lower-dimensional space and be succinctly explained. Frequently used methods include the active appearance model (AAM) [18], dynamic cascaded classifier [17], K-Nearest Neighbors (KNN) [19], and Convolutional Neural Networks (NN) [20]. deep learning [23][24], VGG16 with a transfer learning model [25], YOLOv3 [27], decision trees [28], Support Vector Machines (SVM) [21][22], deep layered convolutional neural networks (CNN) [26], and many more. Not to mention, the FER system selects a classifier using the previously acquired characteristics. Following their retrieval, the features are sent into the chosen classifier, which classifies data based on feature similarities. Support Vector Machines (SVMs) outperform other classifiers in FER (Facial Expression Recognition) [29].

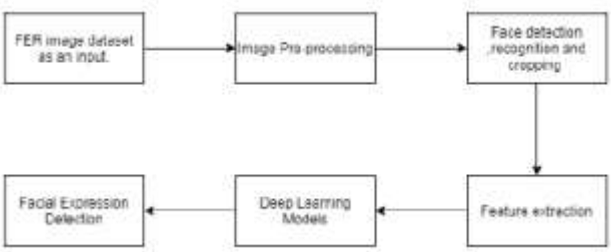

Figure 1.1: Basic model for Facial Expression Recognition

II LITERARURE SURVEY

Utilizing a collection of 14,051 grayscale photos extracted from the FERET face database, we have conducted a comparative study between the Local Binary Pattern (LBP) methodology and other contemporary face recognition methods.

This database contains data on lighting, position angles, face expressions, and other topics.

1.This experiment just looks at the front faces.X: shows 1,195 individuals in a front-on photo.

2. Y: Please give us an alternative sentence. Test1 and subTest2 are included in the snapshot, which was captured at different times. The pictures from the previous year are included in the subgroup of Test 1.

Table 2.1 presents the comparative performance of many methods, including MAP, EBGM, weighted and nonweighted Local Binary Pattern (LBP), and Bayesian Maximum A Posteriori (MAP). These algorithms' effectiveness is quantified. Compared to competing techniques, the mean accuracy is greater when using LBP weighted.

The results of comparing ranks and cumulative scores are shown in Figure 2.1.Comparing data sets from Test1 and Test2 in terms of categories or clusters is the process of ranking. The LBP algorithm performs better than the competition in face recognition, as seen by the rank curve [47].

Table 2.1: LBP comparison with other Face recognition algorithm

Method	X	Y	Test1	Test2	Mean Accuracy
LBP Weighted	0.96	0.78	0.67	0.65	0.81
LBP No weighted	0.92	0.52	0.62	0.51	0.75
Bayesian MAP	0.82	0.37	0.52	0.32	0.72
EBGM Optimal	0.89	0.41	0.45	0.23	0.65

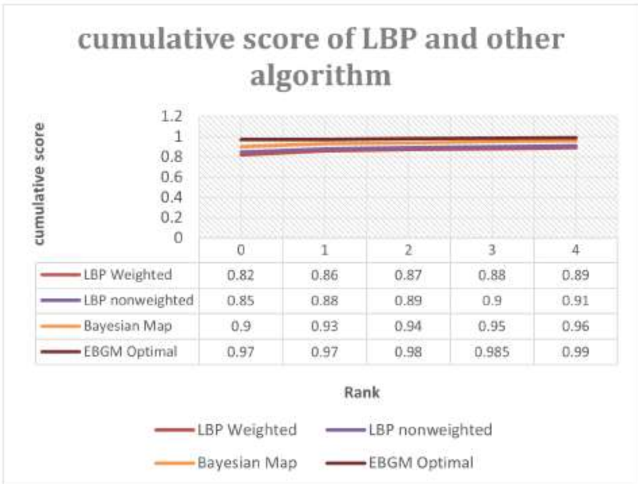

Figure 2.1: Feature extraction methods used for FER

For emotion recognition, the conventional two-stage machine learning approach is used. Taking a set of distinct features or attributes out of the pictures is the first stage. Using a classifier is the second step.

To interpret human emotions, we use emotion detection techniques such as Support Vector Machines (SVM) [76], neural networks [77], and random forests [78]. In recent years, a great deal of time and effort has been dedicated to the research of human emotion recognition. This study has improved our knowledge of psychological behavior in humans. The first step in facial expression recognition is feature extraction, which is followed by model training using these characteristics. The CK+ and JAFFE datasets have piqued the curiosity of several eminent scholars. A mixed deep neural network was utilized by N. Jain et al. [79] to identify facial expressions. Zhang et al. [80] suggested a spatial-temporal Recurrent Neural Network (RNN) for expression detection on the CK+ dataset.

Busso et al.'s evaluation of the efficacy of models based on facial expressions [81]. According to Swapna Agarwal et al. [82], FER—an enhanced platform—may result from fusing pure and mixed emotions.

An automatic facial recognition system was described by Arora and Kumar [83] using a mix of PCA and particle swarm optimization (PSO). The approach combined feature extraction and optimization to achieve accurate facial emotion recognition.

The efficacy of the created technique is evaluated using the JAFFE dataset, which monitors the facial expressions of Japanese women. An automated method for Facial Expression Recognition (FER) was created by Li et al. [84] employing a novel end-to-end network trained using an attention-based methodology.

The network is composed of modules for feature extraction, attention, reconstruction, and classification. The LBP features were extracted using texture data from the face picture. The following content:

We captured even the smallest of facial expressions in order to optimize the network's performance. The attention technique uses a neural network to extract valuable characteristics.
Better outcomes were obtained by using LBP features to improve the attention method. Zhang et al. [84] introduced an unbiased deep learning model for position-independent face expression recognition (FER) and facial image synthesis. The generative adversarial network, or GAN, is the foundation of the recommended technique.

III COMPARATIVE STUDY OF CONVOLUTION NEURAL NETWORK'S ACTIVATION FUNCTIONS

Using the Rectified Linear Unit (ReLU) and Leaky-ReLU activation functions for the inner layer and the softmax activation function for the output layer, we trained a Convolutional Neural Network (CNN) model in this chapter. We examined the impact of various activation functions on image processing.

obtaining information. In addition to the previously mentioned activation functions, this section also describes the Elu, Selu, Softplus, and Sigmoid activation functions.

A. *Role of Relu and Leky-Relu Activation function for image dataset*

Activation functions considerably facilitate analysis. Different outcomes are obtained from linear and nonlinear models, depending on the complexity and number of features in the dataset. Complex datasets are not well-suited for linear models, but datasets with plenty of characteristics are often employed with nonlinear models. These features may be used in order to

Use modeling to represent intricate non-linear relationships. It is capable of acquiring knowledge from both continuous and categorical data. Weights and biases in an artificial neural network can only operate nonlinearly with the aid of functions. Backpropagation is made possible by the activation function, which allows gradients to alter their weights and biases in response to erroneous values. Since these functions are monotonic, the error surface associated with the model is unquestionably convex. Regretfully, not all activation functions are capable of managing all potential scenarios. For instance, the sigmoid activation function has a sluggish convergence rate and is well-known for generating gradient vanishing. Nonetheless, the issue with the tanh activation function is the cause of the fading gradients. Certain gradients may become unstable and susceptible during training, which might result in their termination (Relu[97]-[99]). It might, in short, lead to neuronal death. The binary step function is not continuously differentiable at zero because it is inappropriate for gradient-based optimization. Permeable Relu: During forward propagation, the increasing gradient issue arises when the learning rate is set too high.

The Rectified Linear Unit (ReLU) and the Leaky ReLU, two widely used activation functions today, differ in a few important ways. Figure 1 illustrates it.The monotonic behavior of both functions is further supported by Figure 3.1.

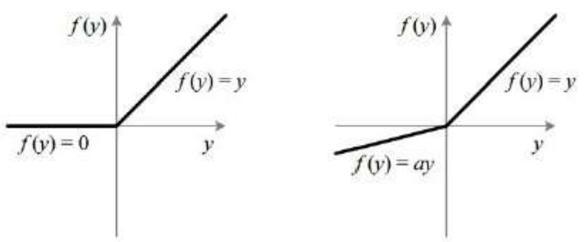

Figure 3.1: Mathematical expression of Relu and leaky-Relu

IV AUTOMATIC FACIAL RECOGNITION USING VGG16 BASED TRANSFER LEARNING MODEL

This section proposes a unique framework for face expression identification using deep learning and transfer learning. Using this approach, the VGG16 model that currently exists is

International Journal of Engineering Innovations in Advanced Technology
ISSN: 2582-1431 (Online), Volume-6 Issue-2, June 2024

layered with a modified trained model. Then, by including more layers, the underlying model is improved.

Complete the task. The proposed model is verified on two benchmark datasets: JAFFE and CK+. Scholars that study facial expressions often use the Cohn-Kanade (CK+) and JAFFE databases. The suggested model outperformed previous techniques, achieving an accuracy of 94.8% on the CK+ dataset and 93.7% on the JAFFE dataset. The given technique has been successfully implemented on Google Colab-GPU.

A. Preprocessed Model

Figure 4.1 presents a model recommendation using VGG16 for transfer learning. Transfer learning uses the weights and layers from an existing model to expedite the training of a new, untrained model.

the acquisition of a cutting-edge model. Our automated facial expression detection system was constructed by transfer learning with the pre-existing VGG16 model. Next, we deployed a newly constructed, untrained model on the CK+ and JAFFE datasets. The VGG16 network model is built on the convolutional layer. Recent victories in the ImageNet 2014 competition by several computer vision techniques indicate to their growing acceptance. After removing the top layers, we added a Flatten, Dense, Drop, and dense-SoftMax layer to aid with categorization. The way our model is designed, this is quite clear. The dropout layer eliminates some data at random to prevent overfitting. The SoftMax layer is used for multiclassification of emotions. All of the layers have used the ReLu activation function, with the exception of the final layer.

The JAFFE and CK+ datasets were used in both the training and testing phases.

A CK+ dataset is an improved version of the original CK dataset. There are 22% more video clips in it. The whole dataset has been divided in half, with 20% designated for testing and 80% for training.

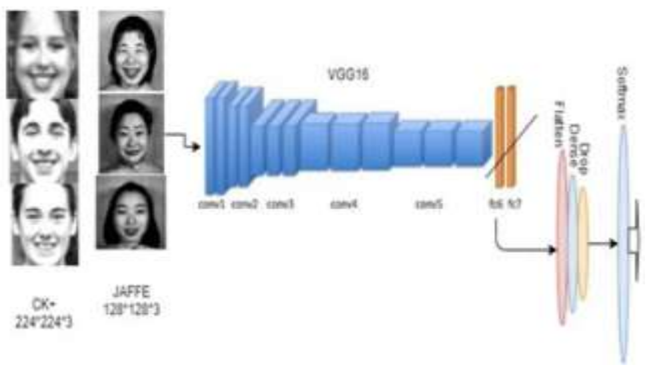

Figure 4.1: A proposed model based on transfer learning using VGG16

Figure 4.2: Sample images of CK+ and JAFFE

We chose 4,000 photographs from the CK+ collection that represented a range of emotions, including neutrality, surprise, pleasure, sorrow, rage, hate, and terror. The JAFFE collection consists of 213 photos that each depict one of seven distinct moods. Six fundamental emotions and one neutral emotion are recognized. Giving an illustration

Pictures from the two databases are shown in Figure 4.2. To make training and testing easier, we resized the images of JAFFE to 128*128*3 and CK+ to 224*224*3. Utilizing both datasets, the proposed approach improves transfer learning outcomes using VGG16.

V FACIAL EXPRESSION RECOGNITION BASED SCORING SYSTEM FOR RESTAURANTS

In order to properly solve the problem that was described before, it is essential to urge each and every customer to offer a rating. A technique for a restaurant rating system is presented in this article. The approach asks each customer for a review after they have visited the establishment. The goal of this method is to increase the total number of ratings. This technology, which employs pre-trained convolutional neural network (CNN) models for the purpose of recognizing facial expressions of emotion, has the potential to be used in open-air dining establishments. It is possible for the customer to evaluate the food by taking a picture of his face that conveys the proper feelings. When compared to a rating system that is based on language, there is a much less amount of data and no collecting of individual experience reports. However, this straightforward, fast, and enjoyable technique of grading ought to give a larger diversity of perspectives about the experiences that customers have had with the concept of the restaurant.

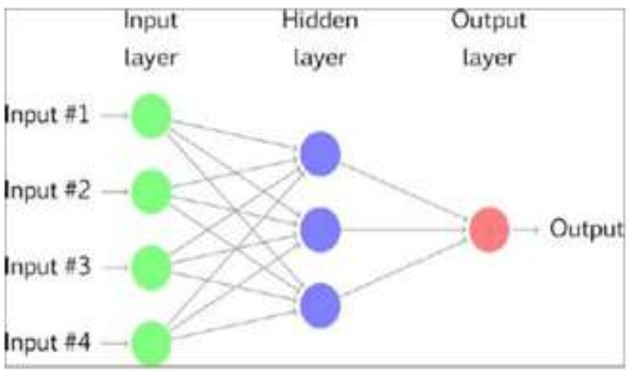

Figure 5.1 System Architecture

Figure 5.2: In above screen click on 'User' link to get below screen where user can upload photo and give ratings

CONCLUSION

Data bias and incorrect annotations are widespread in facial expression datasets due to differences in collecting circumstances and the subjective nature of annotation. Academics often evaluate their algorithms by using a specific dataset and may get favorable outcomes. Preliminary cross-database tests have shown that variations in construction indicators and collecting settings across databases result in disparities [12]. Consequently, algorithms evaluated using protocols inside a single database cannot be applied to new test data, and their performance is notably degraded when applied to other datasets. Other options to reduce this bias include knowledge distillation and deep domain adaptation. In addition, while the training data is expanded by combining many datasets, the performance of Facial Expression Recognition (FER) cannot increase further owing to inconsistent labeling of facial expressions. Class imbalance is a common problem in facial expression analysis that occurs due to the practical difficulties in collecting data. It is easier to capture and label a smile compared to other facial expressions.

A deep learning architecture based on transfer learning is used in the facial expression technique. In this framework, the pre-existing VGG16 model is based upon a modified trained model. Adding extra layers improves this model much better. The current model layers have been frozen, and fresh ones have been trained using each dataset independently. The model showed an increase in performance on the JAFFE and CK+ datasets. The CK+ accuracy of 94.84% and the JAFFE accuracy of 93.75% are attained by our suggested model. In the future, researchers may get better results if they could train a model on a specialized dataset and freeze intermediate-level layers at the same time.

REFERENCES

[1] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016,pp. 1–10. doi: 10.1109/WACV.2016.7477450.

[2] C. Zhang, P. Wang, K. Chen, and J.-K. Kamarainen, "Identity-aware convolutional neural networks for facial expression recognition,"*Journal of Systems Engineering and Electronics*, vol. 28, no. 4,pp. 784–792, 2017. doi: 10.21629/JSEE.2017.04.18.

[3] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie,"Facial expression recognition via learning deep sparse autoencoders,"*Neurocomputing*, vol. 273, pp. 643–649, 2018, issn: 0925-2312. doi:https : / / doi . org / 10 . 1016 / j . neucom . 2017 . 08 . 043. [Online].Available: https : / / www . sciencedirect . com / science / article /pii/S0925231217314649.

[4] A. T. Lopes, E. de Aguiar, A. F. D. Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognition*,2016, issn: 0031-3203. doi: http://dx.doi.org/10.1016/j.patcog.2016.07.026.

[5] A. Ucar, Y. Demir, and C. Guzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," 2016.

[6] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition in video with multiple feature fusion," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 38–50, 2018. doi: 10.1109/TAFFC.2016.2593719.

[7] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "From facial expression recognition to interpersonal relation prediction," *Int. J. Comput. Vision*, vol. 126, no. 5, 550–569, 2018, issn: 0920-5691. doi: 10.1007/s11263-017-1055-1. [Online]. Available: https://doi.org/10.1007/s11263-017-1055-1.

[8] S. K. A. Kamarol, "Spatiotemporal feature extraction for facial expression recognition," *IET Image Processing*, vol. 10, 534–541(7), 72016, issn: 1751-9659.

[9] A. Kumar, A. Kaur, and M. Kumar, "Face detection techniques: Areview," *Artificial Intelligence Review*, pp. 1–22, 2018.

[10] P. Chhabra, N. K. Garg, and M. Kumar, "Content-based image retrieval system using ORB and SIFT features," *Neural Computing and Applications*, vol. 32, no. 7, pp. 2725–2733, Aug. 2018. doi: 10.1007/s00521-018-3677-9. [Online]. Available: https://doi.org/10.1007/s00521-018-3677-9.

[11] M. Kumar, P. Chhabra, and N. K. Garg, "An efficient content based image retrieval system using bayesnet and k-nn," *Multimedia Tools Appl.*, vol. 77, no. 16, 21557–21570, 2018, issn: 1380-7501. doi: 10 .1007/s11042-017-5587-8.

[12] K. Mistry, L. Zhang, S. C. Neoh, C. P. Lim, and B. Fielding, "A micro-ga embedded pso feature selection approach to intelligent facial emotion recognition," *IEEE Transactions on Cybernetics*, vol. 47, no. 6,pp. 1496–1509, 2017. doi: 10.1109/TCYB.2016.2549639.