International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670 An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

FACIAL EMOTION RECOGNITION FOR DUSKY SKINNED PEOPLE USING CNN

Dr. S. Nandagopal[1], P. Suriya[2], M. Vanmathi[3], V. Vetriselvi[4]

[1] Professor, Department of Information Technology, Nandha College of Technology, Erode-638052, Tamil Nadu, India.

[2,3,4] UG Students - Final Year, Department of Information Technology, Nandha College of Technology, Erode-638052, Tamil Nadu, India.

ABSTRACT - In computer vision, the Convolutional Neural Network is a very popular and useful model for emotion recognition. So, we are using the one of the CNN architectures to analyse the emotions called VGGNet. It is one of the methods or models of Transfer Learning Technology. It has more layers than its predecessors ImageNet and ResNet which have minimal layers when compared to VGGNet. It has 16 or 19 layers in its model to train the dataset. It can train a large dataset at a time. Our main goal is to accurately recognize the dusky skinned people's emotions. And so, for that we have used more sample images of dusky skinned people to be trained with the existing one. By doing so, we can improve the identification of emotions on dusky skinned people. And VGGNet technology also provide a great support to the improvement of the concept.

Keywords: Convolutional Neural Network (CNN), VGGNet architecture, facial emotion recognition, dusky skin.

I. INTRODUCTION

Nowadays, most of the aspects of our daily life has been digitalized and have involvement of technologies. One of such technological software is Facial Emotion Recognition. To analyse the psychological state of patients, intelligent tutoring systems, surveillance the Emotion Recognition can be used. It has been studied by researchers in various fields such as psychology, sociology, healthcare, education, robotics. While using AR and VR technology also it will help. In healthcare, it will allow the doctors to monitor the patients when they are unconscious or in ICU to alert if there is any slight change in the patient's emotion. In robotics, it will help the robots to identify the emotions in the human's face and act upon them.

Emotions are mental state that is expressed through the facial expressions of a person. The face is considered the prime perceptual stimulus. The facial muscles have the ability to form more than 40 expressions. However, according to physiologists, emotions evoke a small number of basic expressions namely: joy, sadness, fear, anger, disgust, surprise, interest, and contempt. These emotions are naturally expressed by humans and are taken as a common reference. These facial expressions are a form of nonverbal communication that can support or even replace verbal communication.

As of now, a research based on the experiment of dusky skinned peoples on three most popular and commercial Facial Emotion Recognition Softwares. Which provided a result of great difference between the accuracy of white skinned people and dusky skinned people.

So, to recognize the emotions, we can use neural network of the deep learning. We are using the Convolutional Neural Network, because it is the best model to such recognitions. In Convolutional neural network we have different types of architectures. One of them is VGGNet, which we are using for the project.

The full name of VGG is the Visual Geometry Group. The input of VGG is set to an RGB image of 224x244 size. The average RGB value is calculated for all images on the training set image, and then the image is input as an input to the VGG convolution network. VGG16 has 13 convolutional layers and 3 fully connected layers and VGG19 has 16 convolutional layers and 3 fully connected layers.

II. RELATED WORK

In the past there have been many research and projects done based on the Facial Emotion Recognition. They had used different types of methods and architecture to analyse and train the images. They also gave a good result for the time being. However, facial expression recognition still faces great challenges.

A deep learning-based model for FER is proposed in [17]. The model is based on CNN. It classifies a person's face image into seven different emotions including sad, fear, happy, anger, neutral, disgust, and surprise. The model was trained and tested over FER2013 dataset [7] with 35,685 grayscale images, where 80% of the proposed dataset was used for training and the remainder was used for testing. Random Search algorithm was used to optimize the hyper-parameters of the CNN. This model achieved an accuracy of 66.7%. The recognized emotions were conveyed in text and audio formats.

Minaee et al. [18] proposed a CNN based FER model. The model focuses attention on specific parts of the face that are believed to have a higher impact on the classification such as the eyes and mouth. Spatial transformer [13] is used to extract the parameters aggregated with features extracted by the CNN layer and passed to the dense layer. The model utilized 28,709, 3500, and 3589 images for training, validating and testing, respectively, using the FER2013 dataset[7].

Another FER system for enhancing online teaching is proposed in [3]. The model aims to identify the facial expressions of students to evaluate their concentration in class.

A lightweight model is proposed in [15] for emotion recognition. The CNN based model detects the face using histograms of gradients

International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670 An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

(HOG) [20]. DenseNet reduces the number of parameters to be trained. The accuracy of the model was 71.73%, tested over FER2013[7] and trained for 250 epochs.

A lightweight emotion recognition (LER) system is proposed in [5]. The model incorporates compression techniques into the connected dense layer to eliminate redundant parameters. Three different models, DenseNet-1, DenseNet2, and DenseNet-3 are proposed. The models were trained and tested over FER2013 [7], and FER+. DenseNet-2 with 218,839 parameters, achieved the highest accuracy of 71.55% and 85.68% on FER2013 and FER+, respectively. A new dataset FERFIN is also created as an enhanced version of FER2013 dataset with less noise and corrected labels. The model achieved an accuracy of 85.89% on FIRFIN dataset.

A FER model for VIP is provided in [12]. The model classifies emotions into three categories; positive, neutral, and negative. It incorporates ResNet model [9] for feature extraction. Extracted features are combined through Gated Recurrent Network (GRU) which is a type of Recurrent Neural Networks (RNNs). Multi-layer Perceptron (MLP) classification system is used to classify the emotion. The model was tested on CK+ dataset and achieved an accuracy of 87%. The developed tool displays the probability of each category with an Emoji that represents the predicted emotion. The authors indicated that they are planning to use three-stage signal with a Braille display [2].

Another FER system is proposed in [1]. The system uses Support Vector Machine (SVM) to classify emotions into three main categories; sad, happy, and surprise. The model was trained on JAFFE dataset combined with some newly added images. The system is incorporated into a desktop application that conveys the classified emotion in audio. However, some important details about the model accuracy, error, and sample size are not provided in the paper [16].

In [10], they have researched about the wireless networks and adhoc and in [19], about the MANET networks to implement the FER along with [11] where they have used CNN and Hybrid Feed Forward Deep Neural Network and its use in using a large amount of dataset for MRI images.

In [6], they have researched about the hybrid emotion recognition using CNN-based features. They tested the accuracy of combining CNN features and the machine learning and the accuracy of using CNN features alone. It provided a good statistical data about the testing and output. The results of our experiments demonstrate that Support Vector Machine (SVM) and Ensemble classifiers outperform the SoftMax classifier on AlexNet architecture. These algorithms were able to achieve improved accuracy of between 7% and 9% on each layer, suggesting that replacing the classifier in each layer of a DCNN with SVM or ensemble classifiers can be an efficient method for enhancing image classification performance.

But all of these papers and experiments doesn’t completely decrease the error accuracy of the dusky skinned peoples. So, we have proposed a solution to overcome the error rate.

Joy Buolamwini[8], a researcher in the MIT Media Lab’s Civic Media group. She applied three commercial facial-analysis systems from major technology companies to her newly constructed data set. Across all three

International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670 An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

commercial facial-analysis systems from major technology companies, the error rates for gender classification were consistently higher for females than they were for males, and for darker-skinned subjects than for lighter-skinned subjects. She had found out that from the Examination of facial-analysis software which shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women. For darker-skinned women — those assigned scores of IV, V, or VI on the Fitzpatrick scale the error rates were 20.8 percent, 34.5 percent, and 34.7. But with two of the systems, the error rates for the darkest-skinned women in the data set were worse still: 46.5 percent and 46.8 percent. Essentially, for those women, the system might as well have been guessing gender at random.

In 2023, Martins E. Irhebhude, Adeola O. Kolawole and Goshit Nenbunmwa Amos [14], done a project to improve the accuracy rate we have discussed above. They used both ResNet algorithm and Rotation Invariant Local Binary Pattern features to train and test the dataset. Which gave a significant increase in the accuracy rate.

III. PROPOSED SYSTEM

3.1 METHODOLOGY

The initial stage of our project involves inputting images and performing feature extraction. The input image can be of any type like a jpeg image, or real time captured images or from a video coverage.

Feature extraction utilizes the pre-trained neural network VGGNet model. The VGGNet model have 16 – 19 convolutional layers to train the dataset which will be more helpful to our project. And then to classify the images we have used a Machine Learning algorithm called Support Vector Machine.

Then, we have used the FER2013 dataset and personally created dataset to train and test the model. By combining the VGGNet and SVM classifier we can get the dataset to be classified and trained more accurately.

To enhance the diversity and robustness of the dataset, data augmentation techniques are applied. These techniques involve applying transformations such as rotation, scaling, flipping, and cropping to the existing images. Data augmentation helps increase the variability of the training data and improves CNN's ability to generalize to unseen images.

The trained CNN model (VGGNet) is utilized to analyse facial expressions in the input images. This model has learned to classify facial expressions into different categories based on the features extracted from the images. The output of the CNN models provides the predicted facial expression for a given input image.

3.2 Datasets

3.2.1 FER-2013

FER-2013 (Facial Expression Recognition 2013) is the most popular facial expression dataset introduced in the Representation learning challenge of ICML (Kaggle facial expression recognition challenge) held in 2013. It is the mostly used dataset for all type facial emotion recognition projects. The training set consists of 28,709 examples and the public test set consists of 3,589 examples. Totally 32,298 images are in the dataset. And these images are divided into seven categories

International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

of emotions (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

3.2.2. Own Dataset

With the existing dataset we have created a dataset of fewer images of dusky skinned people's images with the same variety of emotions (Angry, Disgust, Fear, Happy, Surprise, Neutral).

3.3. VGGNet Architecture

The convolutional neural network model called the VGG model, or VGGNet, that supports 16 layers is also known as VGG16. And the one supports 19 layers is known as VGG19.

The VGGNet 16 Layer Architecture is given in the Fig 3.1. It describes about each layer of the VGGNet 16 architecture.

The VGGNet accepts 224x224-pixel images as input. To maintain a consistent input size for the ImageNet competition, the model's developers chopped out the central 224x224 patches in each image.

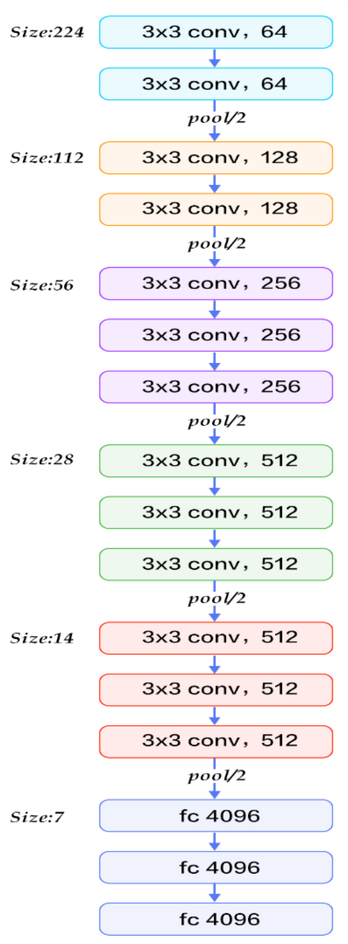


Fig 3.1. VGGNet Architecture

3.4. STEPS IN FER TO TRAIN AND BUILD A MODEL

Step 1:

Dataset Collection

Get the FER2013 dataset to load the dataset. And we have also added the images of dusky skinned people to the existing dataset.

Step 2:

Dataset Preprocessing

First load the datasets in the already setup 'ed environment.

ISSN: 2583-6129
International Scientific Journal of Engineering and Management
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670 An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Then, perform the necessary preprocessing steps on the images, such as normalization, resizing and augmentation and label the images.

Step 3:

Dataset Training

Now convert the labels into suitable format for the model training.

Split the dataset into training, validation and testing sets.

Step 4:

Model Building

Define the VGGNet architecture using your chosen deep learning framework. Configure the layers, neurons and activation functions.

After compiling the model, specify the optimizer, loss function and evaluation metrics.

Step 5:

Model training

Train the model using the training dataset. And utilize the validation set to monitor the model's performance during training.

Now evaluate the trained model on the testing dataset. And analyse the metrics such as accuracy, precision, recall and confusion matrix.

Step 6:

Model Saving

Save the trained model using the save() function and specify the file name to save the model in the Json format. By saving the model we can use it for the future purposes. After saving it, prepare the model for employment ensuring compatibility with the target environment.

Step 7:

To make it real time process

Now we have to connect the project to the real time using the OpenCV algorithm in the Python.

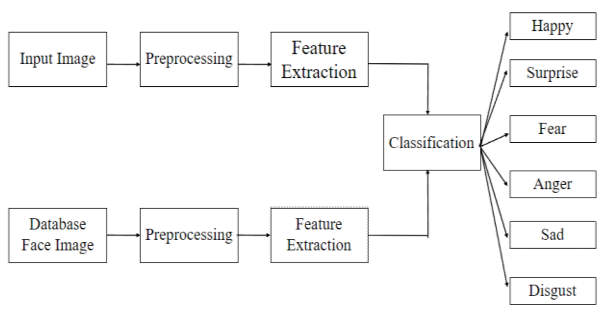
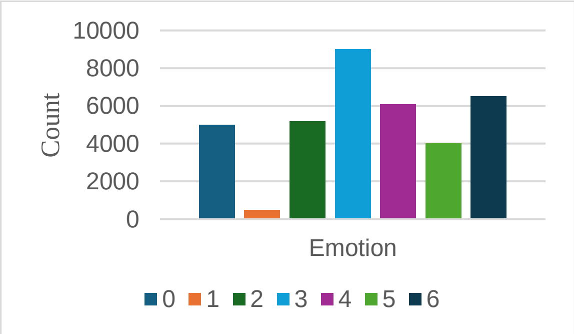


Fig 3.2 Flowchart

Fig 3.2 describes the flow chart of the process in the Facial Emotion Recognition. The flow of the process starts from getting the dataset images and preprocessing them and feature extraction. From the extracted feature, the images are classified. It specifies the emotions such as Happy, Surprise, Fear, Anger, Sad, Disgust. Fig 3.3 describes the level of emotions the system can identify and recognize.



ISSN: 2583-6129
International Scientific Journal of Engineering and Management
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Fig 3.3 Emotion Analysis

IV. CONCLUSION

There are many facial emotion recognition models in usage nowadays. But most of them are not able to completely identify the emotion of the dusky skinned people. Even some models, doesn't recognize the dusky skinned people's face as a human face. So to train the model to identify them we have created a new dataset for dusky skinned people and a different methodology to train the model called VGGNet. And we have also used the existing dataset and combined them.

REFERENCES

[1] A. Ashok and J. John, ''Facial expression recognition system for visually impaired,'' in Proc. Int. Conf. Intell. Data Commun. Technol. Internet of Things. Cham, Switzerland: Springer, 2018, pp. 244–250.

[2] A. Kunz, R. Koutny, and K. Miesenberger, ''Accessibility of co-located meetings,'' in Proc. Int. Conf. Comput. Helping People With Special Needs. Cham, Switzerland: Springer, 2022, pp. 289–294p[‘[

[3] C. Hou, J. Ai, Y. Lin, C. Guan, J. Li, and W. Zhu, ''Evaluation of online teaching quality based on facial expression recognition'' Future Internet, vol. 14, no. 6, p. 177, Jun. 2022.

[4] D. Phutela, ''The importance of non-verbal communication,'' IUP J. Soft Skills, vol. 9, no. 4, p. 43, 2015

[5] G.Zhao, H. Yang, and M. Yu, ''Expression recognition method based on a lightweight convolutional neural network,'' IEEE Access, vol. 8, pp. 38528–38537, 2020.

[6] H.M.Shahzad, Sohail Masood Bhatti, Arfan Jaffar, Sheeraz Akram, Mousa Alhajlah and Awais Mahmood, "Hybrid Facial Emotion Recognition Using CNN-Based Features" Appl. Sci.2023, 13, 5572.

[7] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, and D.-H. Lee, ''Challenges in representation learning: A report on three machine learning contests,'' in Proc. Int. Conf. Neural Inf. Process. Cham, Switzerland: Springer, 2013, pp. 117–124.

[8] Joy Buolamwini "Study finds gender and skin-type bias in commercial artificial-intelligence systems" in MIT, Cambridge, USA.

[9] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[10] Karthick.M, Chandru Vignesh.C, Alfred Daniel.J, Sivaparthipan.C.B, An Efficient Multi-mobile Agent Based Data Aggregation in Wireless Sensor Networks Based on HSSO Route Planning, Ad Hoc & Sensor Wireless Networks, Vol. 57, pp. 187–207, DOI: 10.32908/ahswn.v57.10319.

[11] M. Karthick, Dinesh Jackson Samuel, B. Prakash, P. Sathyaprakash, Nandhini Daruvuri, Mohammed Hasan Ali, R.S. Aiswarya, Real-time MRI lungs images revealing using Hybrid feed forward Deep Neural Network and Convolutional Neural Network, Intelligent Data Analysis 27 (2023) S95–S114, DOI 10.3233/IDA-237436.

International Scientific Journal of Engineering and Management
ISSN: 2583-6129
Volume: 03 Issue: 05 | May – 2024
DOI: 10.55041/ISJEM01670 An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

[12] M. Lutfallah, B. Käch, C. Hirt, and A. Kunz, ''Emotion recognition—A tool to improve meeting experience for visually impaired,'' in Proc. Int. Conf. Comput. Helping People With Special Needs. Cham, Switzerland: Springer, 2022, pp. 305–312.

[13] M. Jaderberg, ''Spatial transformer networks,'' in Proc. Adv. Neural Inf. Process. Syst., vol. 28, 2015, pp. 2017–2025.

[14] Martins E. Irhebhude, Adeola O. Kolawole and Goshit Nenbunmwa Amos "Perspective on Dark-Skinned Emotion Recognition Using Deep-Learned and Handcrafted Feature Techniques" 10.5772/intechopen.109739.

[15] N. Zhou, R. Liang, and W. Shi, ''A lightweight convolutional neural network for real-time facial expression detection,'' IEEE Access, vol. 9, pp. 5573–5584, 2021.

[16] P. Singh, R. Srivastava, K. P. S. Rana, and V. Kumar, ''A multimodal hierarchical approach to speech emotion recognition from audio and text,'' Knowl.-Based Syst., vol. 229, Oct. 2021, Art. no. 107316.

[17] S. K. Lalitha, J. Aishwarya, N. Shivakumar, T. Srilekha, and G. C. R. Kartheek, ''A deep learning model for face expression detection,'' in Proc. Int. Conf. Recent Trends Electron., Inf., Commun. Technol. (RTEICT), Aug. 2021, pp. 647–650.

[18] S. Minaee, M. Minaei, and A. Abdolrashidi, ''Deep-emotion: Facial expression recognition using attentional convolutional network,'' Sensors, vol. 21, no. 9, p. 3046, Apr. 2021.

[19] S.Satheesh Kumar, M.Karthick, An Secured Data Transmission in MANET Networks with Optimizing Link State Routing Protocol Using ACO-CBRP Protocols, IEEE Access, 2018.

[20] W. Zhou, S. Gao, L. Zhang, and X. Lou, ''Histogram of oriented gradients feature extraction from raw Bayer pattern images,'' IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 67, no. 5, pp. 946–950, May 2020.