

Real Time Emotion Detection in Psychiatry Patients using Deep Learning

Preethu S¹, Puneeth Kumar V H², Rakshitha M³, Rohith Gowda N S⁴, Prof. Ladly Patel⁵.

⁵ Faculty CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India

^{1,2,3,4}CSE Department, Sri Krishna Institute of Technology, B'lore-560090, India

Abstract:-This is a real-time project on human emotion recognition system that tracks a person's mood particularly meant for psychiatric patients. Humans use their expressions to express their mood and, on occasion, their needs. It could be a happy face or a frowning face. Words aren't always as powerful as our expressions. The models in this project were created using a variety of machine and deep learning algorithms. It also makes use of some of Python's most powerful packages to create a programme that recognizes human expression in real time. Tensor Flow, Keras, OpenCV, and Matplotlib are examples of libraries. This implementation is adaptable to different environments and platforms. The first example is customer feedback on services and food at restaurants and hotels. In the military, it can have a significant impact. Its very use can aid in recognizing people's behavior in border areas and identifying suspects among them. This project is merely an environment implementation, not actual software that can be used in a real-time setting. This project is divided into two sections processing and building the model for the application using various techniques, and using the model to recognize human facial expressions using OpenCV.

Keywords: *Tensor Flow, Keras, OpenCV, Matplotlib, CNN, MTCNN*

I. INTRODUCTION

This project makes use of a Kaggle dataset containing 48x48-pixel grayscale images of faces. This project is more concerned with improving the accuracy of previous models. Few emotions can be included in grayscale photographs of the face pixels of the forehead. The model provided in this project is more accurate and faster than earlier models. Tensor Flow is a powerful deep learning module in Python that allows you to run deep learning neural Convolutional networks for digits (handwritten) classification, image pre-processing and recognition, sequential models for translation, natural language processing (NLP), and partial differential equation (PDE) based tasks. This model employs the OpenCV package for real-time image processing. For loading, displaying, and analyzing the data, various libraries such as NumPy, Pandas, and Matplotlib were employed. Keras, one of the libraries used to code deep learning models, is also employed in this model. Tensor Flow is used in the backend. This project contains many of the characteristics required for accurate facial expression recognition. Though there are seven of them, two of them have had their tuples eliminated for better outcomes. This project draws inspiration from everyday life, which many people dismiss as unimportant. Consider a restaurant with a large number of repeat clients. A single guest leaving to go to another restaurant would be a disastrous day for the restaurant owner. Many guests are hesitant or unwilling to provide feedback about their restaurant or hotel experience.

And the individual who provides service to them frequently misses or misunderstands what the consumers are feeling. What if you could install a system that tracked people's emotions from their faces at the cash register? It can be utilized to improve their service based on the feedback they receive. Consider the security threat that a crowd poses during a protest. It may appear tranquil, but no one knows what is going on within their heads. What if each security personnel's vest had an effective camera that not only recognized faces but also emotions? If a person is not professional, it is quite difficult to conceal their motives. If security personnel have enough tools, including that facial emotion recognition camera, a violent crowd can be swiftly brought back to normalcy.

Computer Vision:

Leading computer vision research entails the capacity to give any machine with the same level of understanding of images and movies as humans. In the realm of Artificial Intelligence, Computer Vision is the next step. It allows the machine to obtain much more information than a human can. A good example is the medical imaging of bones and muscles in various sections of the body, which is so complex that even a doctor sometimes has trouble understanding it. This section focuses on picture analysis, processing, and interpretation in the form of machine code. The difficulty of artificially extracting information from digital photos is the subject of this research or discovery. There are no neurons in a computer that act like neurons in the brain to boost the capacity to understand images. These studies are usually supported by businesses and organizations in order to identify solutions that a human intellect cannot. Researchers have classified the computer domains in scene reconstruction, video tracking, object recognition and detection, motion estimation, and image restoration. Everyone Dance Now is one of the most remarkable and driven works in this sector to date. Motion translation is included. The transition between two frames, as well as between two people. One can be shown dancing properly in a video by changing the frame motion with another frame in this highly motivated undertaking. There is still a lot of research to be done in this area. This has also been made available to the public. In the fields of neuroscience, signal processing, information engineering, and artificial intelligence, another significant milestone has been reached. All of them are assisting individuals in increasing their efficiency and achieving greater achievements on a daily basis. It is constantly used to assist humans in recognizing activities at the most basic level.

II. BACKGROUND STUDY

[1] Zhou, Ning, Renyu Liang, and Wenqian Shi it employs (MTCNN) Multi-tasking cascaded convolutional networks to

detect face. The detected face coordinates will be transmitted to the final system's. the network model reduces the parameters in convolutional layer by eliminating the fully connected layer.

[2] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab to improve accuracy by extracting 2 different types of deep visual features using 3d hybrid deep and distance features(happyER-DDF). This method adopts a hybrid deep neural network to recognize from unconstrained videos.

[4] Miao, Si, et al it defines about Major Depressive Disorder (MDD). It provides improved generalizable approach to MDD automatic assessment from videos. the maximum probability, average probability of happiness and maximum of AU4(brow lowered) are selected as top three features in each video

[5] Jiang, Zifan, et al in this, system becomes better when multi class detection is employed up to seven categories facilitates clinical analysis of infants and enables to combine these states with disease analysis like GERD. this experiment results a mean-average precision of 81.9% and 84.8% for 4 infants.

A. Abbreviations and Acronyms

MTCNN: Multi-Tasking Cascading Convolutional Neural Network

CNN: Convolutional Neural Network

RIBE: Reaction Ion Beam Etching

DDF: Drug Disease Free

SCEP: Spatial and Channel-wise attention-based emotion prediction

PPG: Photoplethysmography

ECG: Electrocardiogram

EEG: Electroencephalography

III. METHODOLOGY

Automation has long been a hot topic in the twenty-first century. While technology eliminates jobs, it also gives a better and more standard way of life in society. Traditional firms are on the point of practically disappearing in the current day, as new market leaders come up with intelligent company ideas. A company made up of extremely advanced systems for supplying and making decisions for enterprises. What if we could create a real-world application that could collect feedback from customers directly through their facial expressions? What if we could create a system that could detect any antisocial behavior before it occurs based on a mob's mood? What if we could analyze the mental state of psychiatry patient with manual monitoring? Human face expression recognition has a wide range of applications. From customer comments to criminal confessions, and even seeing anti-social elements in the crowd, there's a lot to consider. The major emphasis of this project is grayscale picture conversion. The data is primarily in grayscale, which is a matrix. It also makes use of some of the most widely used open-source libraries, which come pre-loaded with a variety of algorithms. It includes code and implementations. The utilization of these libraries aids the correctness of the model that is later developed. The suggested model is divided into two parts: the first is concerned with the elimination of tuples with the least

important attributes, and the second is concerned with the addition of additional technique to improve the model. Both for testing and training, the characteristics of disgust and fear have the fewest tuples. The model's performance increased when these features were removed.

Initially in the training phase MobileNet model is pre-trained on ImageNet dataset and pre-trained model model is obtained. MobileNets are low-latency, low-power models that have been parameterized to match the resource restrictions of various use cases. They can be used for classification, detection, embeddings, and segmentation in the same way that other prominent large-scale models like Inception are. MobileNets strike a balance between latency, size, and accuracy, and they outperform popular models in the literature. The ImageNet dataset has thousands of classes and hence the pre-trained model has to be adapted to emotion recognition by adding new dense layers and removing the layers that are not required. The model is now pre-trained with new dense layer.

Now the facial expression dataset is neatly cropped for the faces and clean dataset is fine-tuned with pre-trained model with new dense layer to get Emotion Recognition-model. This ends the training phase.

Now in testing phase, the test image is neatly cropped for the face detection and the cropped face is given as input to Emotion Recognition Model. The model then classifies the image into various classes as shown in figure1

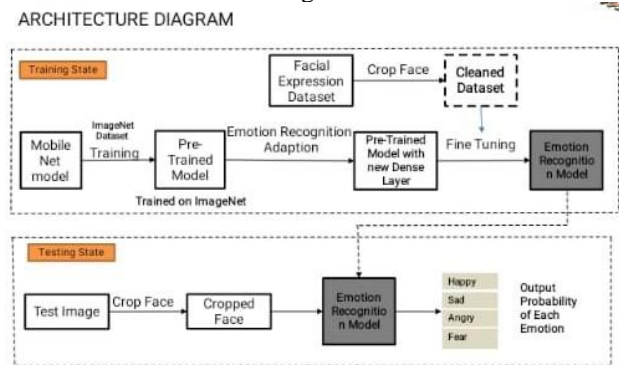


Figure 1: Architecture Diagram

Convolutional Neural Network:

CNN (Convolutional Neural Network) or Conv-Net is a Deep Learning algorithm. An input image is fed into the algorithm, which assigns learnable weights and biases to various qualities in the image and tries to find value to them. These networks help to distinguish one characteristic from the others. The essential aspect of CNN is that it requires far less pre-processing than other methods (classification). The layout of network neurons in a Convolutional Neural Network is comparable to the patterns that human brain cells have when connecting to each other. The visual field of the narrow region where single neurons respond to stimuli is known as the receptive field. A set of such fields covers the entire visual area and overlaps. The example below shows how an image of a handwritten digit is fed into the Conv-Nets and then via the pooling layers.

IV. IMPLEMENTATION:

The proposed model's code is developed in Python3. Later, the model was used to generate outputs for input photographs in

order to determine the person's emotion. 75 percent of the tuples were used for training and 25% for testing in this model. Furthermore, the majority of the work on this project was completed using sublime text, a powerful Python 3 interpreter editor. Library dependencies are also affected by the versions used. This project uses Python3.5, the most recent stable version of Python3. The libraries were acquired from PYPI, or Python Package Index, which is a software repository with a large number of works. It currently has 113,000 libraries, with over 10,000 dedicated to data science. Even though the CPU has been used for the Convolutional Neural Network's epochs, this project does not monitor the system requirements because it cannot function effectively on anything less than a GPU.

The first module comprises pre-processing processes as well as numerous CNN filters and Max Pooling. The data is pre-processed here, and the model that is used in the following module is generated after the data is filtered and pooled over multiple iterations.

The images in the second module are taken in real time from the video capture in OpenCV. After that, the photos are transformed to grayscale images of 48x48 pixels. The grayscale image is then compared to the model created in the first module. To detect the user's faces, this module uses the HaarCascade classifier. Before the image can be converted to grayscale, faces must be recognized. The following are the major steps in implementation-

1. Input Data Set

The input is entered by the user.

2. Data pre-processing

Data pre-processing refers to the control or removal of data before it is used to ensure or improve execution, and it is an important stage in the data mining process.

The pre-processing of data is an important step in the creation of an AI model. Information may not be perfect or in the required configuration for the model at first, which might lead to misleading findings. We change information into the necessary arrangement during pre-handling. It is used to manage the dataset's commotions, copies, and missing upsides. Bringing in datasets, splitting datasets, quality scaling, and other exercises are part of the information pre-handling process. Pre-processing of data is expected when working on the model's precision.

3. Split dataset into train phase and test phase

The train-test split is used to evaluate the display of AI calculations suitable for Algorithms/Applications. This strategy is quick and easy to implement to the point where we can compare the outcomes of our own AI models to machine results.

4. Train the model

Around 75% of FER-2013 dataset is used for training the model by fine tuning the dataset with pre-trained model with new dense layer to obtain Emotion Recognition Model.

5. Validate

The term "model approval" refers to the process of verifying that the model achieves its intended goal. In general, this will entail confirmation that the model is accurate in the conditions of its intended application.

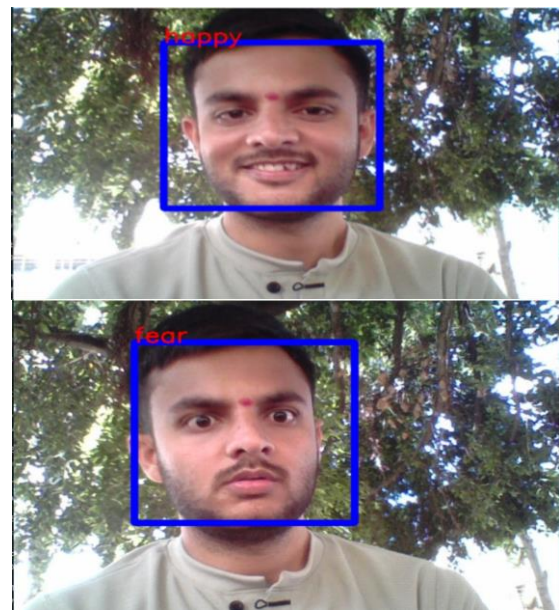
6. Output or result

Emotion Recognition Model is fed a finely cropped image of a face, and the model classifies the image into its appropriate class and outputs the class label.

V. Results and Discussions:

The following information demonstrates how well the model performed:

After half of the epochs had gone through pooling and filtration via numerous layers, the accuracy was 53.3 percent. The final percentage was roughly 69%. The train accuracy was 94.93 percent, and the test accuracy was 58.82 percent, according to the results of the model.



CONCLUSION:

This study can be analyzed and investigated further in order to develop more accurate models using various algorithms and image processing approaches. With more people participating in this field of study, there is a probability that a completely automated facial expression detection system with 100% accuracy can be introduced to the market. These models will aid researchers in developing effective Artificial Intelligence. A humanoid cannot exist without the ability to understand how another person feels in order to assist or serve them. The ability to automatically input photographs into datasets after converting them to grayscale will boost the likelihood of creating a new dataset and generating models. It may also be fed into any microcontroller to turn it into a live or IoT project. The Raspberry Pi is the ideal microcontroller since it functions as an operating system, reducing the effort required to write microcontroller code; all that is required is to dump these programs into it.

This project's code is well-organized and up-to-date, with each dependent capable of meeting the environment variables. As a result, even if the versions are upgraded, there will be no issues with code usability and maintenance. One can never fathom a world without automation in this age of cutting-edge technology. The first thought that comes to mind is how a computer visualizes things around them, aside from logical

reasoning. This is the answer to that imagination. This project's multiple uses include emotion detection of psychiatric patient, security dangers posed by the general public, as well as gathering input from customers at hotels, restaurants, and other profitable companies. It can also be used to make algorithms better. Because this code is dataset independent, the researchers will be able to investigate various choices for constructing models. It just requires photos that will be converted to grayscale as input.

The future work is to apply SCEP's visual attention mechanism to video sentiment analysis and use emotional guidance to construct video summaries and also since the facial expression can be fake, bio-signals or physiological signals such as PPG, ECG and EEG can be used in real world to measure the intensity of emotion and its genuineness.

VI. ACKNOWLEDGEMENT

We would like to thank Prof Ladly Patel for her valuable suggestion, expert advice and moral support in the process of preparing this paper.

REFERENCES

- [1] Zhou, Ning, Renyu Liang, and Wenqian Shi. "A Lightweight Convolutional Neural Network for RealTime Facial Expression Detection." *IEEE Access* 9 (2020): 5573- 5584.
- [2] Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. "A face emotion recognition method using convolutional neural network and image edge computing." *IEEE Access* 7 (2019): 159081-159089.
- [3] Li, Cheng, et al. "Infant facial expression analysis: towards a real-time video monitoring system using r-cnn and hmm." *IEEE Journal of Biomedical and Health Informatics* 25.5 (2020): 1429-1440.
- [4] Miao, Si, et al. "Recognizing facial expressions using a shallow convolutional neural network." *IEEE Access* 7 (2019): 78000-78011.
- [5] Jiang, Zifan, et al. "Classifying Major Depressive Disorder and Response to Deep Brain Stimulation Over Time by Analyzing Facial Expressions." *IEEE Transactions on Biomedical Engineering* 68.2 (2020): 664-672.
- [6] Zheng, Kun, et al. "Recognition of Teacher's Facial Expression Intensity Based on Convolutional Neural Network and Attention Mechanism." *IEEE Access* 8 (2020): 226437-226444.
- [7] Samadiani, Najmeh, et al. "Happy Emotion Recognition From Unconstrained Videos Using 3D Hybrid Deep Features." *IEEE access* 9 (2021): 35524- 35538.
- [8] Yang, Jiannan, et al. "Real-Time Facial Expression Recognition Based on Edge Computing." *IEEE Access* 9 (2021): 76178-76190.
- [9] Li, Bo, et al. "SCEP—A New Image Dimensional Emotion Recognition Model Based on Spatial and Channel-Wise Attention Mechanisms." *IEEE Access* 9 (2021): 25278- 25290.
- [10] Patel, Ladly. "Music Therapy-Based Emotion Regulation Using Convolutional Neural Network." *Applications of Machine Learning and Artificial Intelligence in Education*. IGI Global, 2022. 73-96.
- [11] Wang, Su-Jing, et al. "MESNet: A convolutional neural network for spotting multi-scale micro-expression intervals in long videos." *IEEE Transactions on Image Processing* 30 (2021): 3956-3969.