Stress Detection Based on Emotion Recognition Using Deep Learning

Anjali R[1], J. Babitha[2], Rithika W[3], Ms.Reeja S.L[4]
[1,2,3]*Associate Software Engineer, Ospyn Technologies, Thiruvananthapuram, India*
[4]*Research Scholar, APJ Abdul Kalam KTU*

***Abstract* - Automatic emotion recognition based on facial expression is an interesting research field, which has presented and applied in several areas. Also, Machine learning and deep learning algorithms have gained great success in different applications such as classification systems, recommendation systems, pattern recognition etc. Human face conveys many information including age, gender, identity, personality, emotions, etc. Emotion recognition refers to identifying human emotions typically from facial expressions. This project aims to develop a facial emotion recognition system to identify whether the person is stressed or not. To classify the emotion on a person's face, use a deep convolutional neural network. Dataset having 7 facial expressions is used to train the CNN network. This work is a real time application, in which facial emotion can be detected in live video stream. To detect faces in each frames in the webcam, Haarcascade technique is used.**

***Index Terms* – Facial Emotion Recognition, Convolutional Neural Network , Deep Convolutional Neural Network.**

I.INTRODUCTION

The development and usage of computer systems, software and networks are growing tremendously. These systems have an important role in our everyday life and they make human life much easier. Facial emotion recognition system assumes a lot of importance in this era since it can capture the human behaviour, feelings, intentions etc. The conventional methods have limited speed and have less accuracy while facial emotion recognition system using deep learning has proved to be the better one. This system aims to build a deep convolutional neural network model that recognizes 7 different human facial emotions and this can be used for applications such as customer feedback analysis, face unlocking etc. The rapid growth of artificial intelligence has contributed a lot to the technology world. As the traditional algorithms failed to meet the human needs in real time, Machine learning and deep learning algorithms have gained great success in different applications such as classification systems, recommendation systems, pattern recognition etc. Emotion plays a vital role in determining the thoughts, behaviour and feeling of a human. An emotion recognition system can be built by utilizing the benefits of deep learning and different applications such as feedback analysis, face unlocking etc. can be implemented with good accuracy. The main focus of this work is to create a Deep Convolutional Neural Network (DCNN) model that classifies 5 different human facial emotions. The model is trained, tested and validated using the manually collected image dataset. In the field of computer science, machine learning is one of the emerging technologies that is considered to have an impact of 90% in the next 4 years. Deep learning, a subset of machine learning uses artificial neural network, which is an algorithm inspired from the human brain. Convolutional Neural Network (CNN) is a class of deep neural network that uses convolution as the mathematical operation. As the dataset consists of images, the system uses a 2D CNN for the recognition task. The proposed deep convolutional neural network is trained not only to classify 5 different human facial emotions, but also to yield a good accuracy.

The main motive of our project is to detect stress using vivid deep convolutional neural network and image processing techniques. Our system is an upgraded version of the old stress detection systems which excluded the live detection.

II. LITERATURE REVIEW

Many researches are carried out related to the Stress detection System based on Deep learning. In 2019 G. Cao, Y. Ma, X. Meng, Y. Gao, and M. Meng

[1]developed a system in which deep convolutional neural network is not only capable of recognising five distinct human facial emotions, but it also does so with a high degree of accuracy. The model is trained using the dataset, which was collected manually using a cell phone camera.

In 2020 AnkitaPatil1,Rucha Mangalekar2, NikitaKupa
wdekar3, Viraj Chavan4, Sanket Patil5, Ajinkya Yadav6 [6] developed a stress detection system which is an upgraded version of the old stress detection systems which excluded the live detection and the personal counseling but this system comprises of live detection and periodic analysis of employees and detecting physical as wellas mental stress levels in his/her by providing them with proper remedies for managing stress by providing survey from periodically.

In 2017 C. J. L. Flores, A. E. G. Cutipa and R. L. Enciso, [9]developed a system to recognize the static hand gestures taken under in variations features as scale, rotation, translation, illumination, noise and background. This system use the alphabet of sign language of Peru (LSP). For this purpose, digital image processing techniques are used to eliminate or reduce noise, to improve the contrast under a variant illumination, to separate the hand from the background of the image and finally detect and cut the region containing the handgesture. Convolutional neural networks (CNN) to classify the 24 hand gestures.

From the above reviews, it can be concluded that there is a need for a full fledged development that gives a real time result needs. Our system proposes a cost efficient technique for analyzing the emotion and detecting stress to happy ratio

A. Existing System

In the existing system, image processing techniques is used to detect emotions, which is not accurate. Some of the existing work includes machine learning algorithms, in which facial features is extracted from the input image by using image preprocessing techniques. Thus, it leads to lower the accuracy. Also, it classifies emotions to a limited number of emotions i.e., 2 or 3 emotions. Many of the existing methods use of still images and emotion were perceived by measuring the dimensions of lips and eyes. Currently, the system failed to identify whether the person is stressed or not based on the predicted emotion as well as the amount of stress in the input image, i.e., stress ratio or happiness ratio. Existing work make predictions on the uploaded input facial image, which is not real- time stress detection.

B. Proposed System

The main contribution of the proposed is to model emotion expression on static images for recognizing seven (happy, surprise, angry, disgust, fear, sad and neutral) facial emotional states. To this end, first images have adjusted by pre-processing algorithms to have exact part of face, and the train a deep CNN for classify the emotions.

Deep learning is used for emotion recognition. The image dataset contains number of images. These layers together extract the features from the input image. The first step is to identify the facial region of the input image. Then, it is given to the CNN network for training. After training completed, this model can be used for stress analysis using emotion recognition from live video stream.

III. WORKING

The most important part of developing a deep learning model is analyzing the data. Data analysis consists of two important parts:

i. Data Collection and Exploration:

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data we collect to develop practical artificial intelligence (AI) and machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand. Collecting data allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future changes.

Predictive models are only as good as the data from which they are built, so good data collection practices are crucial to developing high-performing models. The data need to be error-free (garbage in, garbage out) and contain relevant information for the task at hand. For example, a loan default model would not benefit from tiger population sizes but could benefit from gas prices over time. Data collection is the systematic approach to gathering and measuring information from a variety of sources to get a complete and accurate picture of an area of interest. Data collection enables a person or organization to answer relevant questions, evaluate outcomes and make predictions about future probabilities and trends.

ii. Data Preprocessing :
In the preprocessing stage, prepare the data to be fed to the Keras model. The first step is clearing the dataset of null values. Python has become the go-to language for Machine Learning and many of the most popular and powerful deep learning libraries and frameworks. Python provide efficient libraries for data preprocessing.
As the images can be of varying size and complexity without analyzing and processing, the data would not be in a proper format to be used for model development.

IV. METHODOLOGIES

Data collection
The process starts with the collection of data. Data collection is the process of gathering and measuring information from different sources. However, in our case we are using the 'FER-2013 dataset' which is available in the public domain. The dataset contains face images of different persons with different facial expressions. It includes 7 categories of facial expressions such as, angry, disgusted, fearful, happy, neutral, sad, and surprised. Dataset contains total of 35,887 images with .png extension in which images are classified in 7 categories. Angry- 4953 images, disgusted- 547 images, fearful- 5121 images, happy- 8989 images, neutral- 6198 images, sad- 6077 images, surprised- 4002 images.

Data preprocessing
Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. The aim of pre-processing is an improvement of the data for further processing. Pre-processing of data is carried out before model is built and training process is executed. Since the data are images, thus image preprocessing steps are used. Image rescaling is done to convert all images to a common scale. Image scaling refers to the resizing of a digital image. RGB to Grayscale conversion is also done. Haarcascade technique is used to detect faces in each image or each frame in the video. It is an Object Detection Algorithm used to identify faces in an image or a real time video. The algorithm uses edge or line detection features proposed by Viola and Jones.

Data split
The overall dataset provided by fer-2013 contains 35,887 images with .png extension that fall under the 7 classes. The dataset was originally split up into training and testing data. 80% of the images in the dataset for training set and remaining 20% images for testing set The main difference between the two is that the training data contained the ground truth labels of each image, while the testing images did not. This allows us to evaluate the model and also test the model against unseen data and apply our own testing measures.

Feature extraction
Feature extraction is a type of dimensionality reduction where a large number of pixels of the image are efficiently represented in such a way that interesting parts of the image are captured effectively. In deep neural networks(CNN), feature extraction is not manual. The network itself learns to extract features while training. Inside the CNN network, convolution layers extracts features from the input image.

Model development
A CNN architecture is formed by a stack of distinct layers that transform the input volume into an output volume (e.g. holding the class scores) through a differentiable function. A few distinct types of layers are commonly used. We discuss them further below:

Convolutional layer (CNV): The convolutional layer is the core building block of a CNN. The layer's parameters consist of a set of learnable filters (or kernels), which have a small receptive field, but extend through the full depth of the input volume. During the forward pass, each filter is convolved across the width and height of the input volume, computing the dot 16 product between the entries of the filter and the input and producing a 2- dimensional activation map of that filter. As a result, the network learns filters that activate when it detects some specific type of feature at some spatial position in the input. Stacking the activation maps for all filters along the depth dimension forms the full output volume of the convolution layer. Every entry in the output volume can thus also be interpreted as an output of a neuron that looks at a small region in the input and shares parameters with neurons in the same activation map.

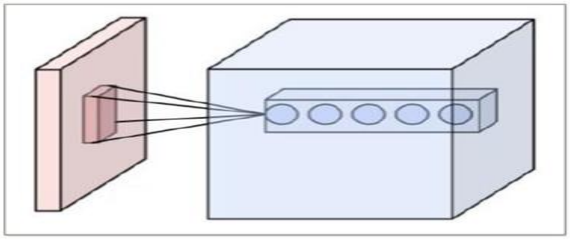

Fig 1: Convolutional Layer

Pooling layer (PL): Another important concept of CNNs is pooling, which is a form of nonlinear down-sampling. There are several non-linear functions to implement pooling among which max pooling is the most common. It partitions the input image into a set of non overlapping rectangles and, for each such sub-region, outputs the maximum. The intuition is that the exact location of a feature is less important than its rough location relative to other features. The pooling layer serves to progressively reduce the spatial size of the representation, to reduce the number of parameters and amount of computation in the network, and hence to also control overfitting. It is common to periodically insert a pooling layer between successive convolutional layers in a CNN architecture. The pooling operation provides another form of translation invariance. The pooling layer operates independently on every depth slice of the input and resizes it spatially. The most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples at every depth slice in the input by 2 along both width and height, discarding 75% of the activations. In this case, every max operation is over 4 numbers. The depth dimension remains unchanged. This is shown in Fig. In addition to max pooling, the pooling units can use other functions, such as average pooling or L2-norm pooling.

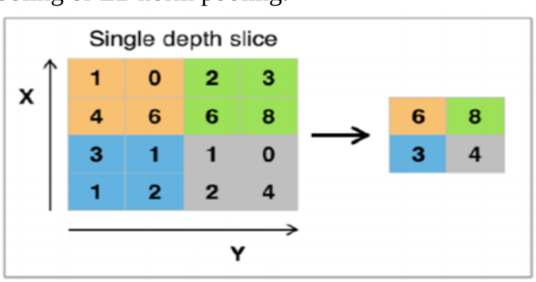

Fig 2 : Pooling Layer

Fully connected layer (FC): Finally, after several convolutional and max pooling layers, the high-level reasoning in the neural network is done via fully connected layers. Neurons in a fully connected layer have connections to all activations in the previous layer, as seen in regular neural networks. Their activations can hence be computed with a matrix multiplication followed by a bias offset. Classification Layer (CL): The classification layer specifies how training penalizes the deviation between the predicted and true labels and is normally the final layer. Various loss functions appropriate for different tasks may be used there. SoftMax loss is used for predicting a single class of K mutually exclusive classes. Sigmoid cross-entropy loss is used for predicting K independent probability values in [0,1] A typical CNN architecture is shown below A simple CNN (Convolutional Neural Net) is created as an initial step. This is then trained with training data.

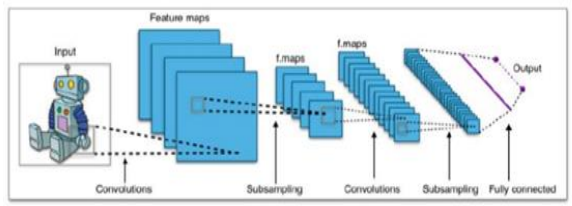

Fig 3: CNN Architecture

Following are the steps carried out for training a CNN:

1. The entire dataset i.e. train; validation and test dataset are pre-processed as discussed in Data pre-processing.
2. A simple CNN is created to classify images. The CNN consists of 4 convolutional layers with 3 max pooling layers in between.
3. Dropout Layer is added.
4. Flattening layer along with two fully connected layers were added at the end of the CNN.
5. Number of nodes in the last fully connected layer were setup as 7 which is number of categories in the dataset, along with using SoftMax activation function. This layer is being used as classification layer.
6.' Relu' activation function was used for all other layers.
7. The model is compiled with 'Adam' (Adaptive Moment Estimation) as the optimizer and 'categorical_crossentropy' as the loss function.
8. After building the model it is trained. During training process the model parameters were saved

when there is an improvement of loss for validation dataset.

9. Save the model for future predictions.

V. RESULT

The authentication is tested effectively and only valid Users can login to this portal .

ID	Scenario	Steps	Data	Expected Results	Received	Pass/Fail
1	Admin Login with correct data	1. Open URL 2. Enter User-ID and Password 3. Click the Login Button	admin for both username and password	Successful Login	Login Done	Passed
2	Fill the details to Add A staff	1. Login as admin 2. Fill all mandatory data in correct format. 3. Click to register	Give all valid data in all fields	Registration successful	Registration done	Passed
3	Login with incorrect data	1. Open URL 2. Enter invalid details 3. Click to login	Give incorrect details	Cannot login	login failed	Passed
4	Analyze Details	1. Start analyses 2. Click report	1. data to DB	Store successfully	Give report	Passed

The application is working properly without any error and the concerned users can use the application effectively.

VI. CONCLUSION

This paper proposes a convolution network model for facial emotion recognition. The model classifies different facial emotions from the image dataset. CNN performs well in emotion recognition. The output from the CNN model is used to detect whether the person is stressed or not. The proposed work detect emotions of person in live video as well as predicts the stress ratio. Stress Detection System is designed to predict stress by monitoring captured images of authenticated users which makes the system secure. Then the system will analyze the stress levels by using Machine Learning algorithms which generates the results that are more efficient.

REFERENCE

[1] G. Cao, Y. Ma, X. Meng, Y. Gao and M. Meng, "Emotion Recognition Based On CNN," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8627-8630

[2] G. Yang, J. S Saumell, J Sannie, " Emotion Recognition using Deep Neural Network with Vectorized Facial Feature" 2018 IEEE International Conference on Electro/Information Technology (EIT).

[3] Lu Lingling liu, "Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network", 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA).

[4] I. Rocco, R. Arandjelovic and J. Sivic, "Convolutional Neural Network Architecture for Geometric Matching," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017,pp. 39-48.

[5] J. Shijie, W. Ping, J. Peiyi and H. Siping, "Research on data augmentation for image classification based on convolution neural networks," 2017 Chinese Automation Congress (CAC), Jinan, 2017, pp. 4165-4170.

[6] Ankita Patil1, Rucha Mangalekar2, Nikita Kupawdekar3, Viraj Chavan4"Stress detection in IT proffessionals by Image processing and Machine learning" vol:3,January-2020

[7] Diederik P. Kingma, Jimmy Ba, "Adam: A Method for Stochastic Optimization," 33rd International Conference for Learning Representations, San Diego, 2015

[8] A. Krizhevsky, I. Sutskever, and G. E. A Convolutional Neural Network Hand Tracker", in proceedings: Neural Information Processing Systems Foundation. 1995Hinton, "Imagenet classification with deep convolutional neural networks," in Proc. Annual Conference on Neural Information Processing Systems (NIPS), 2012, pp. 1106–1114.

[9] C. J. L. Flores, A. E. G. Cutipa and R. L. Enciso, "Application of convolutional neural networks for static hand gestures recognition under different invariant features," 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Cusco, 2017, pp. 1-4.

[10] K. Bouaziz, T Ramakrishnan, S. Raghavan, K. Grove, A.A.Omari, C Lakshminarayan, " Character Recognition by Deep Learning: An Enterprise solution.", 2018 IEEE Conference on Big Data

[11] S. Suresh, H. T. P Mithun and M. H. Supriya, "Sign Language Recognition System Using Deep Neural Network," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 2019, pp. 614-618. doi: 10.1109/ICACCS.2019.8728411