Deep Learning for Automatic Stress Recognition in Real-Life Situations

Mahesh Babu U[1], Rohit Satya Sai P[2], Sai Thilak K[3], Sai Shankar CH[4], Swarna Kuchibhotla [5]

[1,2,3,4,5] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, AndhraPradesh, India

[1] maheshbabu.uppada@gmail.com

ABSTRACT

Stress recognition is important for facial acknowledgment which is acquiring prevalence and the requirement for it increments massively. Even though there are strategies to perceive feeling utilizing AI procedures. The Deep Learning Methods to perceive pressure to group the feeling as indicated by outward appearances. Stress Recognition manages the investigation of deriving feelings, strategies utilized for gathering. Stress Emotion can be perceived from outward appearances, discourse signals, and so on Applying the Convolution Neural Network [1] calculation and Hidden Markov Model calculation, the model consequently perceive the individual if the person in question in pressure. This Model first catches the outward appearance, then sends picture to the Convolution neural network first, perceives the facial highlights, and afterward the model acknowledgment whether the individual in pressure or neural dependent on facial highlights.

Keywords

Face features, Classifier, Selection, Suitable Crop.

Article Received: 10 August 2020, Revised: 25 October 2020, Accepted: 18 November 2020

I. Introduction

Artificial intelligence has had to expand contribution to any extent of human existence. The advances are adjusted to the requirements of the person and man-made consciousness is the thing that makes this variation among innovation and people conceivable.

These procedures are utilized in calculations for the acknowledgment of human feelings. At the point when people attempt to speak with others, a high rate is addressed by non-verbal correspondence. Numerous examinations show that outward appearances have an association with human feelings. The capacity of people to distinguish and recognize these feelings makes it workable for us to see one another. The fundamental goal of this piece of man-made brainpower is to utilize learning strategies to get the machine fit for distinguishing these feelings.

Robots don't feel any feeling, they don't have sympathy with people, they can't perceive the feelings that an individual is feeling contrasted with their own, as people do. The data accessible is enormous grids that address the pictures caught and other extra data they may have like sensors or amplifiers. It is starting there where need to start working. Hence, the acknowledgment of feelings by PC is so convoluted. To accomplish this, fundamentally find what feelings the individual truly feels, what innovation is the most suitable to catch them, and which models and calculations are the best. It is for every one of these factors that this issue has not yet been adequately addressed. In this theory, gauge some AI and profound learning models including Support Vector Machine (SVM) [2], Multilayer Perceptron (MLP), and Convolutional Neural Networks (CNNs) [1].

The evaluation focuses on the task of emotion recognition through facial expression has been used by the algorithm Convolutional Neural Networks (CNNs) and Speech recognition using Hidden Markov Model (HMM).

II. Literature Survey:

As humans, with age, individually learn to react to situations with empathy, personally learn new emotions and behavior skills but machines are not capable of doing things on their own so, let us to train the machines to capable of recognizing by using machine learning or deep learning. In present days, machines are capable of accomplishing many tasks that humans can do. In fact, machines perform more complex tasks than humans but recognizing mien and facial expressions is challenging. The first paper is the detecting emotional stress from facial expressions for driving safety in this paper they developed a real-time non-intrusive monitoring system, which

detects the emotional conditions of the driver by dissecting outward appearances. In this project, the index terms are emotion, stress, detection, facial expressions technique used Support Vector Machine, Supervised descent method (SDM) and future scope is important to explore further the definition and subject ward qualities of passionate pressure. The example of worldly elements of outward appearances and activities can likewise be coordinated into the model preparing. Different prompts, for example, head movement and acoustic signs could likewise be incorporated to accomplish better execution [3]. The other paper is about Facial Expression Recognition Method Using Convolutional Neural Networks Based on Edge Computing in this paper they found the imbalanced number and the high likeness of tests in the appearance information base can prompt overfitting in facial acknowledgment neural organizations. To defeat the deficiency that the roundabout agreement ill-disposed organization model must be planned coordinated, then build a compelled roundabout agreement generative antagonistic organization by adding a class limitation data. Technique used edge computing, deep convolutional neural networks, Generative adversarial networks (GAN) and future scope is to upgrade the information without restricting the articulation condition of the info picture is likewise a bearing that can be improved [4]. Picture Based Stress Recognition Using a Model-Based Dynamic Face Tracking System this paper presents a technique to identify pressure from dynamic facial picture successions. The picture groupings comprise of individuals exposed to different mental tests that instigate high and low-stress Hidden Markov Models (HMMs) are a compelling instrument to display the fleeting reliance of the facial developments technique used Hidden Markov Models (HMMs) and future scope is they have introduced a DDDAS for the acknowledgment of stress from outward appearances. Our strategy depends on the utilization of deformable models and HMMs that can manage the powerfully changing information and fluctuations in the outflow of stress among individuals [5]. Next paper is based on FER is one of the applications in the human-PC communication. In this model, they train the facial expression recognition (FER) by using the deep network as a CNN algorithm. Technique used CNN, Facial Expression Recognition (FER) and future scope is required to apply the proposed network alongside the preparation procedures to different areas with little datasets [6].

III. Methodology:

In this paper the proposed model was to initially give images as input to image preprocessing. In that haar feature based cascade.xml is to extract facial feature and those facial feature [7] was to send model.h5 in that model the facial features was predicted as neutral or stress and that predicted value was send as output shown in figure -1.

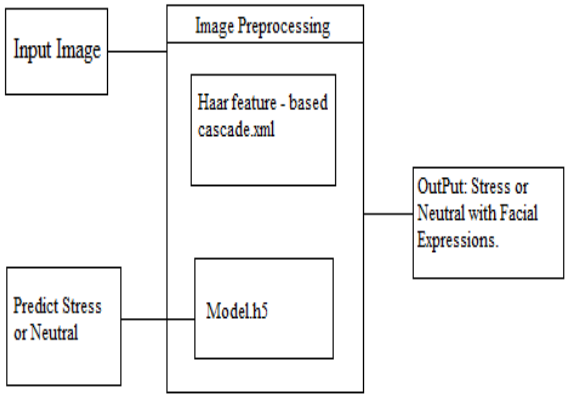

Figure – 1: Block Diagram of Proposed Model

3.1 Data base:

In this paper used database is FER2013 [8] consists of 48*48 pixels sized face images. In this database, it contains 7 types of images labelled as: 4593 Angry images, 547 Stress images, 5121 Fear images, 8989 Happy images, 6077 Sad images, 4002 Surprise images, 6198 Neutral images are there in this Model, in that only two types of emotions are used in this paper those are stress and Neutral emotions. The trained dataset consists of two columns emotion and pixels. The emotion column consists of 0 and 1, 0 is represented as stress and 1 is represented as Neutral.

3.2 Preprocessing:

Image pre-processing is a crucial process in any image analyzing system [9]. Without proper pre-processing, the recognition will be ineffective or may give improper results in later stages. Here to improve the quality will be processed further. Various processes that in this model going to apply converting an RGB image to Grayscale image, etc.

The convolution activity is utilized with certain channels for identifying edges. Assume you have a grayscale picture with measurements 6 x 6 and a channel of measurements 3 x 3(say). At the point when 6 x 6 greyscale pictures convolve with 3 x 3 channel, then result is 4 x 4 pictures. Most importantly 3x3 channel network gets duplicated with the initial 3 x 3 size of our greyscale picture, at that point move one segment straight up to the end, after that, move one column, etc.

3.3 CNN classifier:

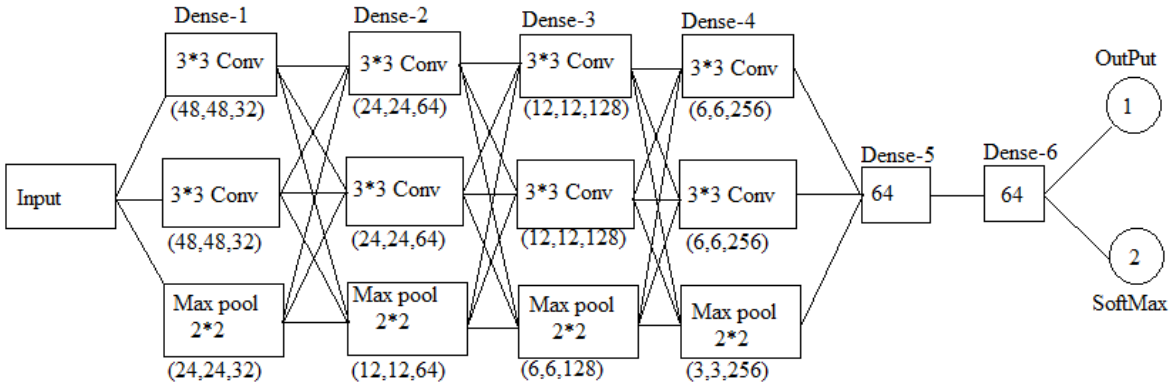

Figure – 2: Flow Chat of the Model

In this model, create an application that takes the face of the person of the image as an input and for the format module, as input, then detect whether he or she in stress or neutral. This paper for recognition of the input images the algorithm used is CNN.

CNN is a neural network algorithm. The Convolution neural network extricates the highlights of the picture and diminishes the measurement without losing its highlights. Convolutional Layer is the principal layer in a CNN. The algorithm takes the input image as a matrix form of height*width*depth. Kernel mean matrix which has smaller the size of an Image it is named as convolution matrix. The kernel matrix also in a format of height*width*depth for each kernel have respect bias which is of a scalar quantity.

The dot product of the convolution matrix (kernel) and the input image give the convolved feature. From the above image fig-1 the convolution activity is utilized with certain channels for identifying edges. Assume you have a grayscale picture with measurements 6 x 6 and a channel of measurements 3 x 3(say). At the point when 6 x 6 greyscale pictures convolve with 3 x 3 channel, then result is 4 x 4 pictures. Most importantly 3x3 channel network gets duplicated with the initial 3 x 3 size of our greyscale picture, at that point move one segment straight up to the end, after that, move one column, etc.

MaxPool layer is utilized to diminish the size of the convolved pictures. When apply a fully connected layer without applying the Max pool layer between two convolutional layers then the values are computationally expensive so apply fully connected after applying max-pooling of two convolution layers.

ReLU is an actuation function that is in a non-straight structure that has picked up prominence in the profound learning area. The principle favorable position of the ReLU actuation work is it doesn't enact all the neurons simultaneously. This sort of actuation work is direct in the positive measurement, however zero in the negative measurement. Subsequently, this initiation work returns 0 at whatever point it gets negative info, yet in the event that any sure information x it restores that esteem. So numerical spoke to as f(x) =max (0, x).

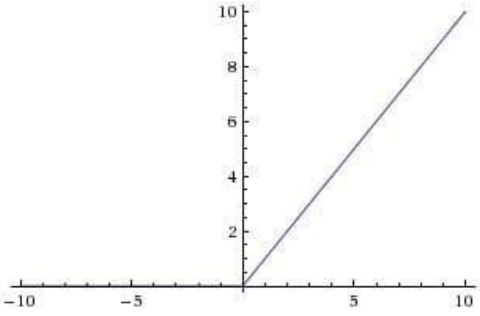

Figure – 3: RELU activation function graph

The completely associated layer is the feed-forward layer in the Convolution neural network these layers are in the last couple of layers in the network the yield of the maximum pool layer is the contribution of the completely associated layer. The utilization of a completely associated layer is to order the picture into a mark.

The softmax is the enactment work which is as calculated relapse that standardizes an information esteem into a vector of significant worth that follows a likelihood circulation whose absolute summarizes to 1. In this paper softmax function is to yield the layer for the CNN calculation in light of the fact that for the most part, softmax created values are between the scope of 0,1 can be away from double arrangement and added the same number of classes in our neural organization model. The softmax work is utilized to process misfortune that can be normal when preparing a dataset. In the event that one of the information sources is little, the softmax transforms it into a little likelihood, and in the event that the information is huge, at that point it transforms it into an enormous likelihood, yet it will consistently stay somewhere in the range of 0 and 1. That's why it is also known as multinomial logistic regression. The mathematical represented as:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Where

σ = SoftMax

Z = Input Vector

K = No. of multi class classifier

e^{zi} = Exponential function of input vector

e^{zj}= Exponential function of output vector

IV. Experimental result:

4.1 Recognition Model:

Model is saved as Model.h5 so this Model.h5 is used as a classifier. The input image is trained with this model and predict the output. At the recognition stage, in this paper use Haar feature-based cascade classifier XML by using this XML then extract the facial features in the image. Haar feature-based cascade classifier XML depends on AI where course work is prepared from a lot of pictures where Model.h5 is a deep learning model because in this model uses CNN algorithm as classification algorithm.

4.2 Analyze Model:

In this model the input image is train with the model.h5 it predict the output the output is show by the xml file the facial features. Finally, conclude with the result by generating the Stress or Neutral and display the facial features which extracted.

Precision means exactness is the proportion between the True Positives and all the Positives. Accuracy additionally gives us a proportion of the significant information point shown in table-1.

Recall means the review is the proportion of our model accurately recognizing True Positives. Review likewise gives a proportion of how precisely our model can recognize the applicable information shown in table-1.

F1-score is the Symphonious mean of the Exactness and Review. This is simpler to work with since now, rather than adjusting exactness and review, simply focus on a decent F1-score and that would be demonstrative of a decent Accuracy and a decent Review incentive too shown in table-1.

Support means scores relating to each class will disclose to you the precision of the classifier in grouping the information focuses on that specific class contrasted with any remaining classes. The help is the number of tests of the genuine reaction that lie in that class shown in table-1.

	precision	recall	f1-score	support
Stress	0.72	0.65	0.68	1018
neutral	0.73	0.79	0.76	1216
micro avg	0.73	0.73	0.73	2234
macro avg	0.73	0.72	0.72	2234
weighted avg	0.73	0.73	0.72	2234
samples avg	0.73	0.73	0.73	2234

Table - 1) Accuracy of output quality

The training information well, however it can't sum up and make precise forecasts for information it hasn't seen previously. while the validation set is simply used to assess the model's presentation. The validation set that let you get a proportion of the nature of the model show in figure- 4.

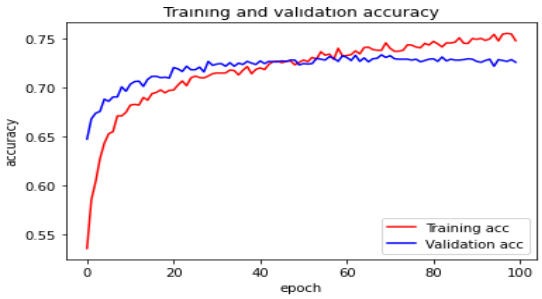

Figure-4) Difference between the Training and Validation accuracy.

Training misfortune is the blunder on the training set of information. Validation misfortune is the mistake in the wake of running the validation set of information through the prepared organization. Train/substantial is the proportion between the two. Startlingly, as the ages increment, both validation and training mistakes drop. Validation misfortune is similar measurement as training misfortune, yet it isn't utilized to refresh the loads show in figure-5.

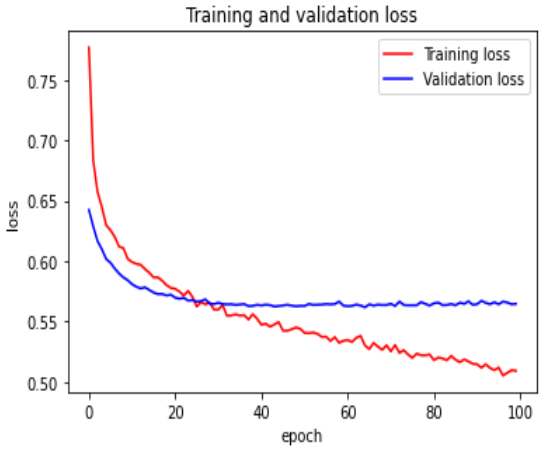

Figure-5) Difference between the Training and Validation Loss.

A Confusion matrix is a N x N matrix utilized for assessing the exhibition of a characterization model, where N is the number of target classes. The matrix contrasts the real objective qualities and those anticipated by the machine learning model. The columns speak to the anticipated estimations of the objective variable shown in Figure-6.

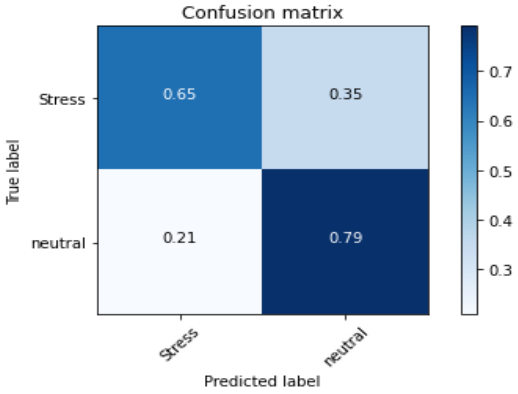

Figure-6) Confusion Matrix

```
Epoch 100/100
182/182 [==============================]
45/45 [==============================] -
[INFO] accuracy: 72.61%
[INFO] Loss: 0.5647209286689758
Time:  0:01:34.466356
```

Figure-7) Accuracy and Loss

Figure-8) Predicted Outputs

V. Conclusion and Feature scope:

Stress recognition is actualized utilizing various calculations and procedures. Each technique has its preferences and drawbacks. Our proposed model at first does the pre-processing in that step it converts every image into a grayscale image. After the extraction of facial expression by using

Haar feature-based cascade classifier. XML in this paper used CNN algorithm model. Training our system with the help model automatically recognition whether the face is stressful or neural efficient to recognize the class correctly.

New findings are up to now FER2013 database was not used for stress recognition so in this paper FER2013 used for stress. The accuracy given by this database was 72.61 percentage. Although let us see that so many algorithms have been implemented in various previous projects, in order to make a robust system for automatic Stress recognition, there are still many loop holes left in the model. Allows us to increase the accuracy by adding more epochs and by adding more input images in the database.

VI. References:

[1] [1] "Learning salient features for speech emotion recognition using convolutional neural networks 16(8), Mao Q, Dong M, Huang Z, & Zhan Y December-(2014)."

[2] [2] "A Comprehensive Survey on Techniques for Facial Emotion Recognition. Deshmukh, Renuka, and M. E. Scholar. International Journal of Computer Science and Information Security (IJCSIS) 15, no. 3 (2017)."

[3] [3] "Detecting emotional stress from facial expressions for driving safety. Gao, Hua, Anil Yüce, and Jean-Philippe Thiran. 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014."

[4] [4] "A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing. Chen A, Xing H, & Wang F. (2020) IEEE Access, 8, 49741-49751."

[5] [5] "Image-based stress recognition using a model-based dynamic face tracking system. Metaxas D, Venkataraman S, & Vogler C. (2004, June). In International Conference on Computational Science (pp. 813-821). Springer, Berlin, Heidelberg."

[6] [6] "Physiological inspired deep neural networks for emotion recognition. Ferreira, P. M., Marques, F., Cardoso, J. S., & Rebelo, A. (2018) IEEE Access, 6, 53930-53943."

[7] [7] "A survey of face recognition techniques. journal of information processing systems 5, no. 2 (2009): 41-68. Jafri, Rabia, and Hamid R. Arabnia."

[8] [8] "A face emotion recognition method using convolutional neural network and image edge computing. Zhang, Hongli, Alireza Jolfaei, and Mamoun Alazab. IEEE Access 7 (2019): 159081-159089."

[9] [9] "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network. Abuzneid, Mohannad A., and Ausif Mahmood IEEE access 6 (2018): 20641-20651."