Perspectives in Communication, Embedded-Systems and Signal-Processing (PiCES) – An International Journal
ISSN: 2566-932X, Vol. 4, Issue 6, September 2020
Part of the Proceedings of the 1st All India Paper writing Competition on Emerging Research - PaCER 2020

CNN Model for Depression Detection using JAFFE Dataset

Bhavna Singh Parihar
Department of Electronics and Communication,
Usha Mittal Institute of Technology, Mumbai,
India, bhavnap199@gmail.com

Shraddha Sandesh Satam
Department of Electronics and Communication,
Usha Mittal Institute of Technology, Mumbai,
India, shraddhasatam77@gmail.com

Shravani Sandesh Satam
Department of Electronics and Communication,
Usha Mittal Institute of Technology, Mumbai,
India, shravanisatam07@gmail.com

Prof. Kiran Dange
Department of Electronics and Communication,
Usha Mittal Institute of Technology, Mumbai,
India, Kiran.Dange@umit.sndt.ac.in

***Abstract:* Depression is one of the serious mental illnesses and a difficult illness to detect, due to it showing different symptoms in different individuals. It also becomes difficult to treat patients due to them not seeking help because of mental well-being given a backseat in the overall health of an individual, and the stigma present in the society about seeking help from psychologists. It is also seen that people try to downplay the symptoms when talking to a psychologist. Here, we have designed a CNN model which detects whether a person is depressed or not based on the facial features of the person.**

***Keywords:* Convolutional Neural Networks; Depression; Local Binary Pattern; Accuracy; JAFFE dataset**

I. INTRODUCTION

Depression has become a very common mental illness. According to [1] more than 264 million people worldwide are suffering from this disease. Though it seems that as a mental illness it might not affect a person physically, the lack of awareness and social stigma surrounding this topic leads to many disabilities. The effects of depression are long- lasting and can affect a person's functionality and ability to lead a happy life. The major symptoms of depression are lack of appetite/sleep, lack of interests in things that the person previously enjoyed and a general lack of social life which can also sometimes become a contributing factor to a person becoming depressed. According to [2] nearly 15 percent of Indian adults need active intervention for one or more mental health issues and one in about 20 individuals is depressed.

II. RELATED WORK

Each year a challenge is held by ACM called the Audio- Visual Emotion Challenge and Workshop (AVEC) [3] wherein people present their works on emotion recognition by the AI based on the human behavior. The very first challenge posed the problem of detecting discrete emotions from an array of natural behavior data, which was then changed to detection of self-reported severity of depression in the 2nd and 3rd challenges. Then different databases were introduced in the upcoming challenges such as Wizard of Oz [4] wherein the participants interacted with an autonomous machine which was partially controlled by a human. Then they introduced cross-culture affect prediction [5] called 'in-the-wild' wherein the interaction between the human and machine took place in not a controlled environment.

In our CNN model we have used the JAFFE (Japanese Female Facial Expression) database [6] which contains 213 images of 10 Japanese models with 7 facial features.

III. PROPOSED SYSTEM

In the proposed model the database is processed using local binary pattern (LBP) which can capture the micro and macro facial descriptors from the database and when these descriptors are compared with the image of a person, can help in figuring out whether the person is depressed with a good accuracy.

A. Architectural Diagram

We have designed a CNN model which detects whether a person is depressed or not based on the facial features of the person. CNN model takes in an image from the user and compares it with the trained model and returns us the value as depressed or non-depressed.

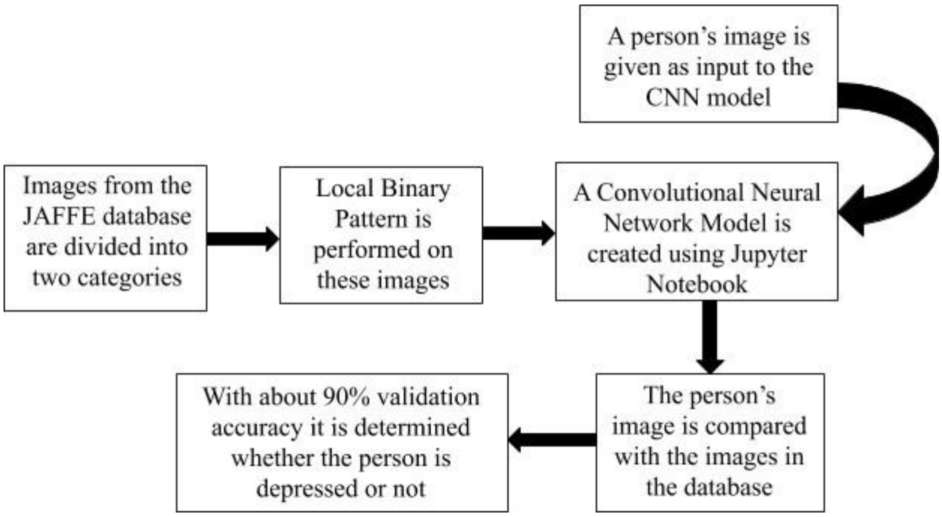

Fig 1. Basic block diagram of the CNN Model

B. Segregating of Training Dataset

[6] is the database that we have used to train our model. The database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. But for the purpose of detecting depression, we have classified images as Non-Depressed (happy, neutral, surprise) and Depressed (angry, disgust, fear, sadness) as depression is mixed emotion. After segregating images into depressed and non-depressed, we apply Local Binary Pattern (LBP) on the set of images. The images are then resized such that they are of 100 X 100 pixel size.

C. Applying Local Binary Pattern on Training Dataset

Local Binary Pattern is used to describe the texture of 2D surfaces. It is used to describe the micro and macro features of the image. [7].

$$LBP(P,R) = \sum_{p=0}^{P-1} (g_p - g_c)2^P \tag{1}$$

$$s(x) = \begin{cases} 1, & if\ x \leq 0 \\ 0, & otherwise \end{cases} \tag{2}$$

where *P* represents the number of samples on the circle of radius *R*; *g(p)* and *g(c)* are the pixel intensities of a pixel on the circle and the central pixel.

In our model we have used *P*=10 points and radius *R* as 5 in uniform mode so the resulting LBP shows patterns where only those combinations are shown where all the black dots are adjacent to each other and the white dots are adjacent to one another.

D. Creation of CNN model

Convolution Neural Network can be formed using Jupyter Notebook. [8] quoted the name "convolutional neural network" indicates that the mathematical operation "convolution" is used in one or more layers of the network instead of the general matrix.

The network majorly contains an input, a output and multiple hidden layers in between. Convolution is central to the efficacy of this algorithm. The algorithm consists of the following steps:

- The image in the training dataset which has already been converted into grayscale and has been processed using the local binary pattern is the input layer.

  Each pixel of the image describes its feature so it can be called as a node and each node has its own weight (also called as kernel). The kernel describes the relationship between the feature and the node, so there must be the same number of nodes in the input layer and all the hidden layers of the network.

- The next layer is the convolutional layer. According to [9] in the convolutional layer a predefined pattern or stamp is run over the entire image starting from the top-left corner. The pixel values covered by the pattern are multiplied with the corresponding stamp value and the products are added. Here we have defined a kernel size of (3 X 3) for the pattern with 64 kernels (at each convolutional layer the network will learn 64 kernels). We have 3 convolutional layers in our CNN model.

- The next layer is the activation layer. The activation layer can be given as an argument of the convolutional layer or can be written as a separate layer in keras. We have used two functions of its functions in our model. According to [10] The ReLU (Rectified Linear Unit) function applies max(x,0) on all the weights of the output of the convolution layer, thus eliminating all the negative values. This function is applied on the first 2 convolutional layers. The other function that we have used in our model is the Sigmoid function applies sigmoid(x)=1/(1-exp(-x)), wherein the values which are less than a threshold value is converted to zero and the values greater than the threshold value are converted to 1. This function is applied to the last convolution layer of our model.

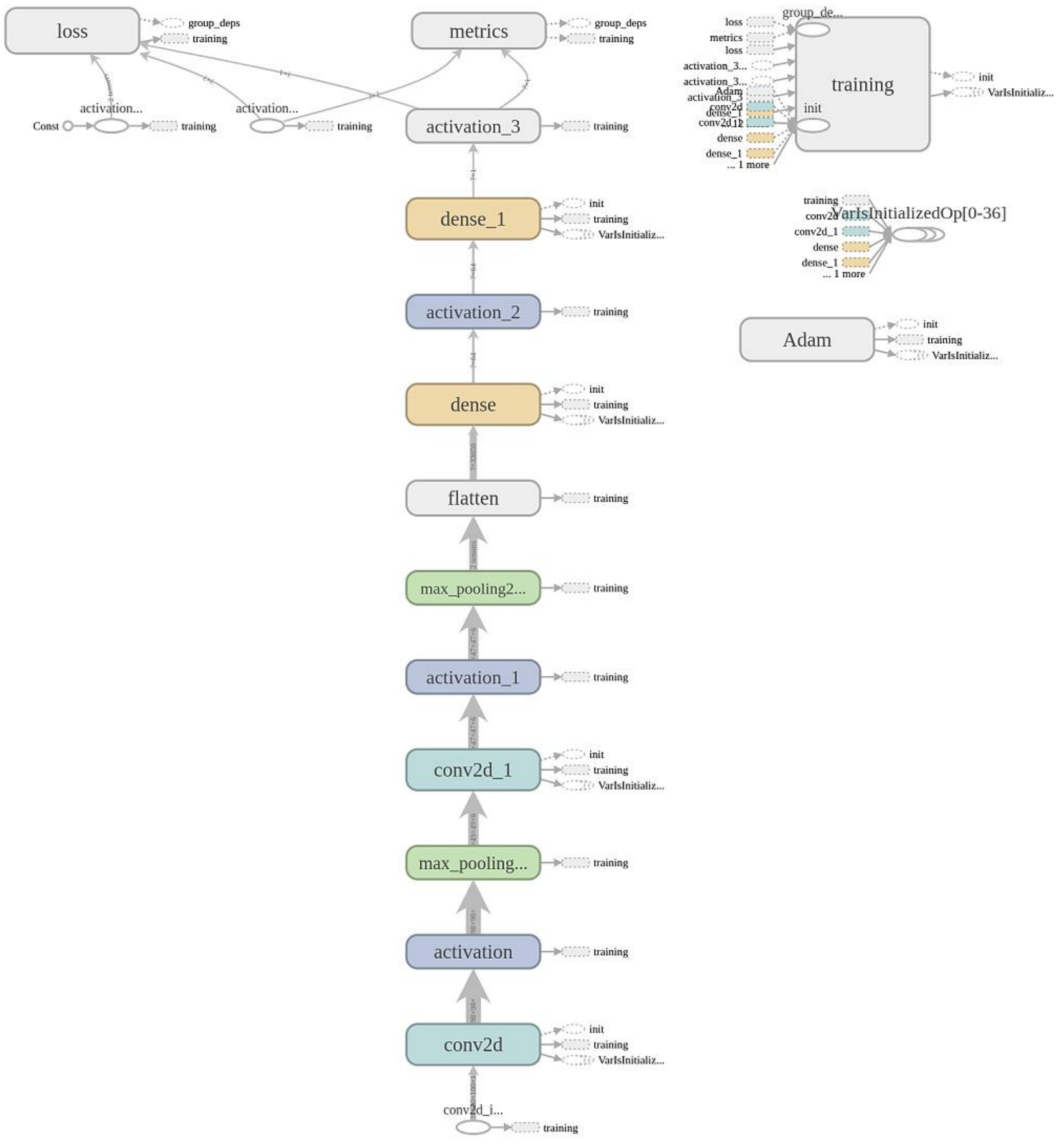

Fig 2. CNN Model

- The next layer is the maxpooling layer. This layer is used to reduce the size of the image so that it becomes easier to compute and requires less time. A kernel size of (2 x 2) is defined in our model. The pattern runs over the output of the last layer and returns the maximum value of that part as output.

- According to [11], the dense layer is a densely connected convolutional layer implements the operation:

$$output = activation(dot(input, kernel)) \quad (3)$$

where activation is the element wise argument passed by the layer and kernel is the matrix weight created by the layer.

- Dropout layer is used to solve the overfitting problem. According to [12], the overfitting problem occurs when the model becomes extremely good at classifying the training dataset but fails to do so when a slightly different data is given to it for classification, so by dropping certain neurons (with their connections) during the training phase itself we are forcing the model to generalize better to the data it hasn't seen before.
- Another way of combatting the overfitting problem of the model is by introducing a validation subset. According to [13], the validation subset is a part of the training dataset which is not trained by the model initially. The model will train on the remaining dataset and then will validate the subset with the features it has just learnt. In our model we have made 0.1 percent of our training dataset as the validation subset.
- The loss function that we have used is called binary cross- entropy. It calculates the loss by computing the following average [14]:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i.\log(p(y_i)) + (1 - y_i).\log(1 - p(y_i)) \quad (4)$$

where N is the total size of the dataset and *y(i)* is the target value. It is necessary to have the last activation layer of the model in sigmoid function to apply this loss function. We have used the 'Adam' optimizer in our model.

E. Comparison of data

The person's image is compared with the image in the dataset. This help us to decide whether a person is depressed or not.

Above model gives us accuracy of 90 percent which is quite good.

IV. RESULTS

A. Local Binary Pattern

Fig 3. Image converted into grayscale

Fig 4. LBP is applied on the Image

Local Binary Pattern is only applied on the training dataset. Note: The images shown here are not from the training dataset, these are just examples.

B. Accuracy and Loss

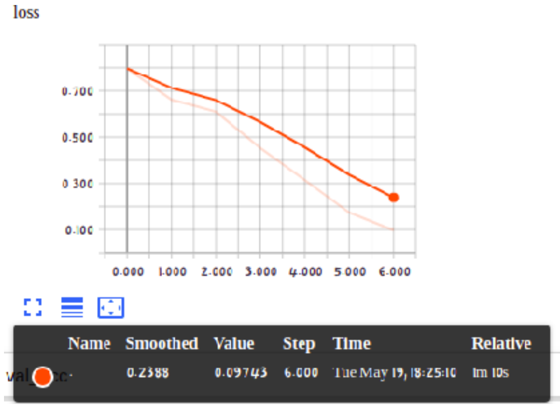

Fig 5. Loss

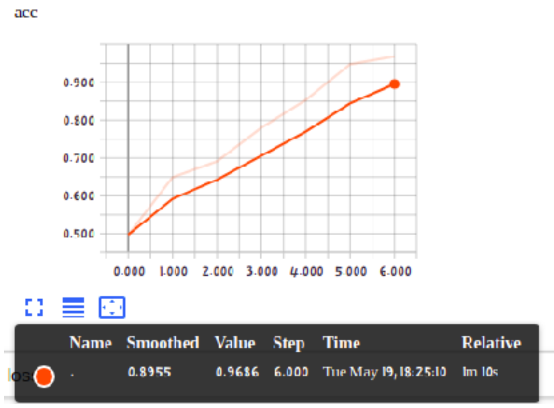

Fig 6. Accuracy

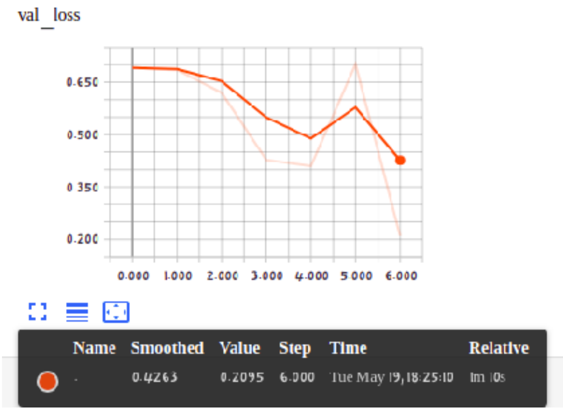

Fig 7. Validation Loss

a) With about 90 percent validation accuracy, it is used to determine whether the person is depressed or not.

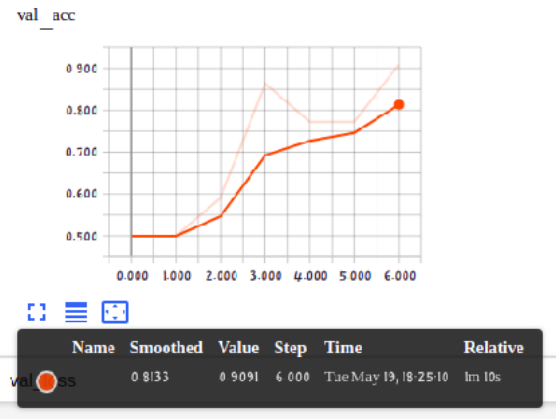

Fig 8. Validation Accuracy

V. CONCLUSION

We have made a basic CNN model that applies grayscale and LBP on training dataset and compare the results with an input image (not from the training dataset) and tells us with about 90 percent accuracy whether the person in the image is depressed or not.

Furthermore, the accuracy of the model can be increased by using different texture detecting algorithms such as OCLBP (Over-Complete Local Binary Pattern), MRELBP (Median Ro- bust Extended Local Binary Pattern), etc. which are extensions of Local Binary Pattern.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Dr. Shikha Nema, HOD (Electronics and Electronics & Communication Department) and our guide, Prof. Kiran Dange for her valuable guidance and continuous support during the project; her patience, motivation, enthusiasm, and immense knowledge. Her direction and mentoring helped us to work successfully on the project topic.

Our sincere gratitude to Dr. Sanjay Pawar, Principal, Usha Mittal Institute of Technology, for his valuable encouragement and insightful comments.

We would also like to thank all the teaching and non-teaching staff for their valuable support. Last but not the least we would like to thank our parents and friends.

REFERENCES

[1] ” Depression”, Who.int, 2020. [Online]. Available: https://www.who.int/health-topics/depressiontab=tab3.

[2] ” Depression”, South-East Asia Regional Office, 2020. [Online]. Available: http://origin.searo.who.int/india/topics/depression/aboutdepression/en/.

[3] ” Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop — ACM Conferences”, Dl.acm.org, 2020. [Online]. Available: https://dl.acm.org/doi/proceedings/10.1145/3347320.

[4] ” Wizard of Oz — Usability Body of Knowledge”, Usabilitybok.org, 2020. [Online]. Available: https://www.usabilitybok.org/wizard-of-oz.

[5] Y. Rogers and P. Marshall,” Research in the Wild Synthesis Lectures on Human-Centered Informatics”, Morganclaypool.com, 2017. [Online]. Available: https://www.morganclaypool.com/doi/abs/10.2200/S00764ED1V01Y201703HCI037

[6] Lyons, Michael, Kamachi, Miyuki, and Gyoba, Jiro, “The Japanese Female Facial Expression (JAFFE) Database”. Zenodo, 14-Apr-1998.

[7] G. Zhao and M. Pietikainen,” Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915-928, 2007. Available: 10.1109/tpami.2007.1110

[8] I. Goodfellow, Y. Bengio and A. Courville. (2016). Deep Learning. MIT Press [online] Available at: http://www.deeplearningbook.org

[9] A. Rosebrock,” Keras Conv2D and Convolutional Layers - PyImageSearch”, PyImageSearch, 2020. [Online]. Available: https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional- layers/.

[10] K. Team,” Keras documentation: Layer activation functions”, Keras.io, 2020. [Online]. Available: https://keras.io/api/layers/activations/.

[11] K. Team,” Keras documentation: Dense layer”, Keras.io, 2020. [Online]. Available: https://keras.io/api/layers/corelayers/dense/.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov,” Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, Journal of Machine Learning Research, vol. 15, pp. 1929-1958, 2014. Available: http://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf.

[13] “ Train, Test, Validation Sets explained”, Deeplizard.com, 2018. [Online]. Available: https://deeplizard.com/learn/video/Zi-0rlM4RDs.

[14] D. Godoy,” Understanding binary cross-entropy / log loss: a visual explanation”, Medium, 2018. [Online]. Available: https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a.