Available online at http://scik.org

J. Math. Comput. Sci. 11 (2021), No. 5, 5474-5486

https://doi.org/10.28919/jmcs/5835

ISSN: 1927-5307

ANALYSIS OF SENTIMENTAL IMAGES USING DEEP LEARNING APPROACH

G. HEREN CHELLAM[1], V. ROSELINE[2,†,*]

[1]Department of Computer Science, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012, Tamil Nadu, India

[2]Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli – 627 012, Tamil Nadu, India

Abstract: Deep learning also known as universal learning approach is a kind of machine learning used to carry out classification tasks straightforwardly from Medias like images, text or sound. The paper centers on execution examination of three pre-trained deep learning network with an end goal of classification of images which are related to sentimental analysis. The pre-defined convolutional neural networks (CNN) handled are AlexNet, ResNet50 and VGG16 with different Epoch. These networks are pre-trained on Twitter dataset. We focus on the structure of feelings deduced by our model and contrast it with what has been proposed in the psychology literature, and confirm our model on a bunch of pictures that have been utilized in psychology studies. At long last, our work likewise gives a helpful instrument to the developing scholarly investigation of pictures consists of both photographs and memes on social networks. The network architectures are analyzed dependent on different means including, accuracy, precision, recall and F1-score. As per the experiment, out of three networks AlexNet gives better outcome as far as precision when compared to other networks.

Keywords: CNN; deep learning; pre-trained networks; sentimental analysis.

2010 AMS Subject Classification: 91D30.

---

*Corresponding author

E-mail address: rose_vasee@yahoo.com

†Research Scholar, Register Number: 18221172162009

Received April 9, 2021

1. INTRODUCTION

The online social network has become an indispensable piece of our ordinary life. Users are sharing a ton of a lot of scholarly and visual substance to convey their emotions and sentiments. These contents exhibit the emotions and practices of billions of individuals all through the world. Social networks are offering various types of assistance for their users to communicate and exchange information. Users use these services to share various occasions of their life, to communicate conclusions on various issue and to show care and backing towards companions and society. Examining these user created contents can help comprehend and foresee user behavior. Information gained from such frameworks can profit a few applications like such as predictive modeling, product, and service recommender system, online marketing etc. Researchers have analyzed this pattern and tons of investigations have been performed to analyze sentiment and opinion mining through textual contents of social networks.

In recent days, visual contents acquired generally more fame than textual contents among the users of different social networks such as Facebook, Instagram, SnapChat, Flickr, Twitter, etc. Status or posts with visual substance regularly contain a short literary depiction or no content by any means. Along these, the visual highlights express a huge part of the people feeling or assessment in these kinds of contents. In addition, pictures can defeat language limit and are more obvious. Fig. 1 shows some picture tweets gathered from Twitter where various sorts of feelings are communicated [13]. While there are huge measure of work for examining the feeling of literary substance, research on visual supposition examination is as yet in its rudimentary stage. Since dissecting supposition from the picture is challenging because of a few reasons. While object acknowledgment is commonly very much characterized, picture assessment investigation is more theoretical in nature. Visual sentiment analysis includes the capacity to perceive object, scene, action and their emotional context producing hand-created highlights from pictures for foreseeing sentiment requires a significant requires a lot of human exertion and time. On the other side, supervised algorithms need an immense volume of regulated preparing

information which is hard to gather for pictures of various spaces. As a result, passionate parts of pictures are decently ignored contrasted with other computer vision activities such as object recognition, detection, and tracking.

Fig. 1 Sample Tweet images from Twitter

Deep learning is sometimes called universal learning since it tends to be applied to practically any application space deep learning don't need the plan of highlights early. Highlights are consequently discovered that are ideal for the main job. As a result, the regular varieties in the information are consequently educated. The same deep learning approach can be utilized in various applications or with various data types. This approach is often called transfer learning. Likewise, this methodology is useful where the issue doesn't have adequate accessible information. The profound learning approach is exceptionally adaptable. There is a major initiative at Lawrence Livermore National Laboratory (LLNL) in creating systems for networks this way, which can execute.

2. RELATED WORKS

Sentiment analysis on text is a well-developed research area in both computer science and psychology, and sentiment analysis has been used to answer psychological questions. However, researchers have cautioned that sentiment analysis focuses on the positive or negative sentiment

expressed by a piece of text, rather than on the underlying emotional state of the person who wrote the text [2] and thus is not definitely a reliable measure of latent emotion. As my contribution I have done novelty in the text mining in regards of sentimental analysis with PS-POS for text extraction and sentimental analysis was done using the CNN technique Bi-LSTM giving out a drastic output of 93.05% accuracy [1].

Recently there is an improved interest from various research communities in understanding the emotional response of the viewer during interaction with social media. A psychological study on the effect of colors on emotions based on Pleasure, Arousal and Dominance model shows that more brighter tones are more lovely, less stirring, and prompt less predominance than the more obscure colors [3]. In [4], researchers used factor analysis method and investigated how eleven emotion scales are related with three color emotion factors (i.e., color activity, color weight and color heat) of single colors, which shows that there is stability in the way people perceive colors. To computationally tackle this issue, scientists have done a ton of deals with this. In [5], they used Supporting Vector Machines to estimate the local image statistics. Sartori et al. [6] proposed to use both visual and text information in a combined learning model for abstract painting emotion recognition. K. He et al. [2] used a multi-task learning approach for painting style analysis. These models are all traditional statistic models and don’t apply deep neural networks. Subsequently, higher-level visual semantics such as image aesthetic analysis [7] and visual sentiment analysis [8] are getting increasingly manageable. You et al. [9] utilized CNN to learn highlights which are valuable for visual examination.

3. METHODOLOGY

The schematic outline of the methodology is shown in Fig. 2. At first all the images are preprocessed. Next feature extraction and classification are performed, which are carried out by using pre-trained CNN architecture which includes AlexNet, VGG16 and ResNet50. Performance analysis of all networks is detailed in the later sections.

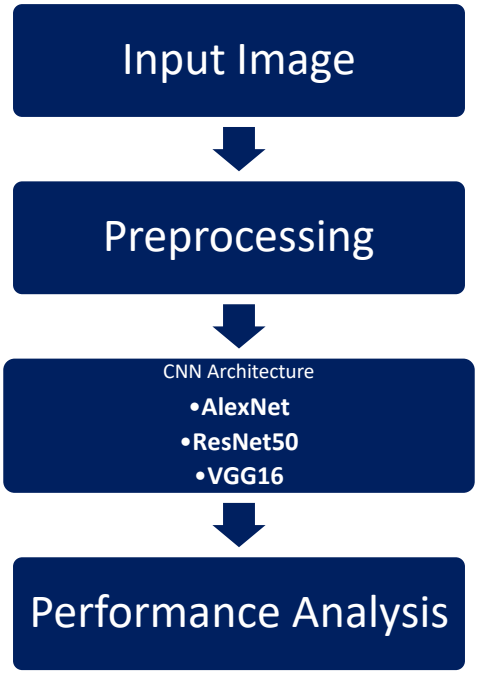

Fig. 2 Block diagram of the proposed method

3.1 Preprocessing

The purpose of preprocessing is to upgrade the picture to required level. Data augmentation is the process by which the total number of images can be increased to manifold. It is basically utilized for avoiding the issue of over fitting if there is less number of images is accessible in the dataset. The processes included for data augmentation are resizing, rotation, translation and reflection. The images are resized as required by the chosen network. The other one rotation incorporates turning the pictures in certain measure of degree. In translation, the images are vertically and horizontally aligned by a defined numeric vector. In reflection, images are flipped from left to right.

3.2 CNN Architecture

The CNN, one of the deep learning architecture belongs to the class of feed forward neural network [10]. The two significant steps involved are learning features and classifying data performed by input layer, hidden layer and output layer. The images with pre-defined size are stacked to the input layer. The size of image is in the format height, width and depth. For RGB, depth is 3 and 1 for gray scale image. The initial segment constitutes the feature learning layer, where the main two tasks carried out are convolution and pooling. The other layers which are under consideration are normalization layer and activation layer. The convolution layer is indicated by the filter size, number of filters, stride and padding. The aftereffect of convolution layer is a feature map and the size of which is given by Eqn 1.

$$O = \frac{W-K+2P}{S} + 1 \quad (1)$$

where, W denotes the input height/length, K is the filter size, P means number of zero padding and S is the stride. When there is an increase the number of convolution layer, more complex features can be learned. The goal of batch normalization followed by convolution layer is for regularization. The activation function used is ReLU (rectified linear unit). Eqn 2 gives relation for ReLU.

$$f(x) = \max(0; x) \quad (2)$$

The max pooling operation performed here is for downsampling, helps in minimizing the size of feature map. The output layer is one where the classification is performed. Which consist of a fully connected layer, softmax layer and classification layer. The input to fully connected layer is given by the relation

$$y = WT\ x + b \quad (3)$$

Here weight matrix W is multiplied with the input obtained in hidden layer and is added with a bias where, y is the output class obtained, x is the feature vector. The softmax layer, it converts the raw values of the output classes into normalized score. The result of prediction is a probability value of class occurrence. Finally classification layer provides class label according to the probability. The following section explains the three pre-trained CNN architectures [15].

i) ResNet50

The basic block diagram of the ResNet50 architecture is depicted in Fig. 3. ResNet50 is a traditional feed forward network with a residual connection. The output of a residual layer can be defined based on the outputs of $(l-1)h$ which comes from the past layer defined as x_{l-1}. $F(x_{l-1})$ is the yeild after performing various operations (e.g. convolution with various size of filters, Batch Normalization (BN) trailed by an activation function like the ReLU on x_{l-1}). The final yeild of residual unit is xl which can be defined with the following equation:

$$x_l = F(x_{i-1}) + x_{l-1} \quad (4)$$

The residual network encompass of few fundamental residual blocks. However, the works in the residual block can be altered based on the distinct architecture of residual networks [9].

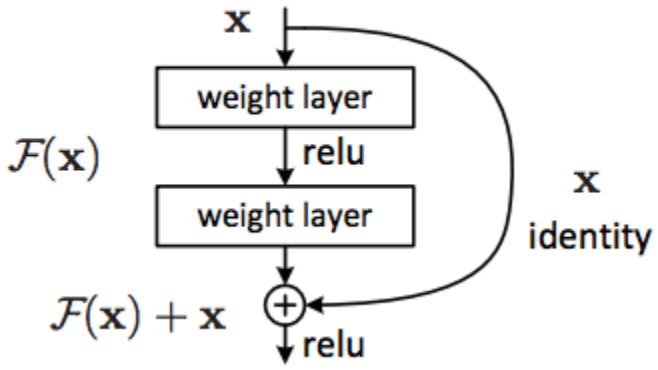

Fig. 3 ResNet50 Architecture

ii) VGG16

The flow chart of the VGG-16 network shown in Fig. 4 as follows:

- The first and second convolutional layers comprise of 64 feature kernel filters and size of the filter is 3×3. As input image (RGB image with depth 3) moves into first and second convolutional layer, dimensions changes to 224x224x64. Then the subsequent yield is passed to max pooling layer with a stride of 2.

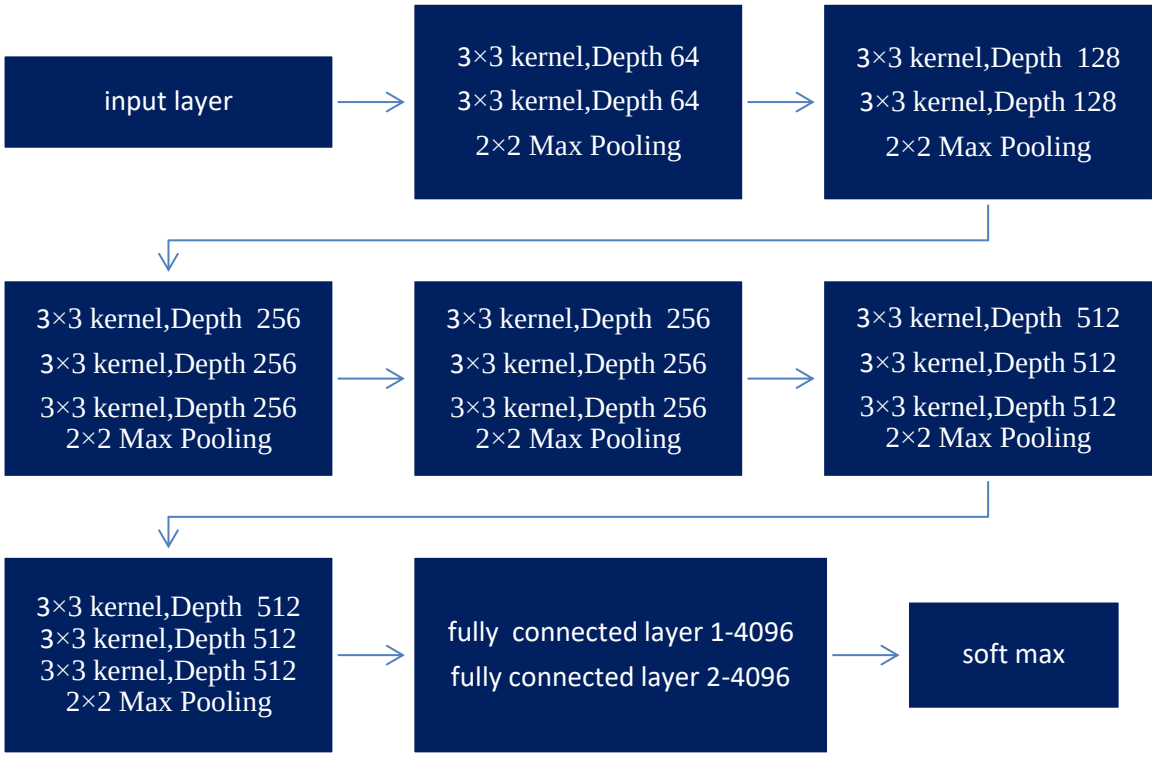

Fig. 4 VGG16 Architecture

- The third and fourth convolutional layers are of 124 feature kernel filters and size of filter is 3×3. These two layers are passed on to a max pooling layer with stride 2 and the subsequent yield will be reduced to 56x56x128.
- The fifth, sixth and seventh layers are convolutional layers with kernel size 3×3. All three use 256 feature maps. These layers are passed on to a max pooling layer with stride 2.
- Eighth to thirteen are two sets of convolutional layers with kernel size 3×3. The full sets of convolutional layers have 512 kernel filters. These layers are passed on to a max pooling layer with stride of 1.
- Fourteen and fifteen layers are fully connected hidden layers of 4096 units succeeded by a softmax output layer (Sixteenth layer) of 1000 units [16].

iii) AlexNet

The precise structure of AlexNet is shown in Fig. 5. The first convolutional layer carries out convolution and max pooling (MXP) with Local Response Normalization (LRN) where 96 distinct receptive filters are used which is in size 11×11. The max pooling operations are performed with 3×3 filters with a stride size of 2. The same operations are performed in the second layer with 5×5 filters. 3×3 filters are utilized in the third, fourth, and fifth convolutional layers with 384, 384, and 296 feature maps respectively. Two fully connected (FC) layers are used with dropout succeeded by a Softmax layer at the end. Two networks with similar structure and the similar number of feature maps are trained in parallel for this model. Two new ideas, Local Response Normalization (LRN) and dropout are presented in this network. LRN can be applied in two distinct manners: first applying on single channel or feature maps, where an N×N patch is selected from same feature map and normalized based on the neighborhood values. Second, LRN can be applied across the channels or feature maps (neighborhood along the third dimension but a single pixel or location) [13].

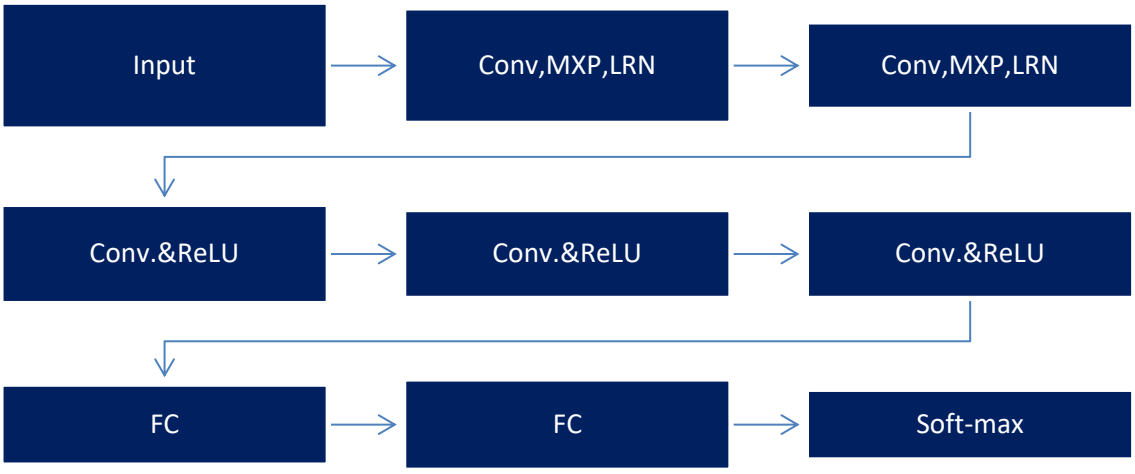

Fig. 5 AlexNet Architecture

4. PERFORMANCE ANALYSIS

4.1 Dataset

In this paper, we examine the performance of Resnet50, AlexNet and VGG16 by using Twitter dataset with 8288 images taken from different tweets. The examination is done in different epoch to identify the better approach in above three algorithms. We used Keras and TensorFlow as backend.

4.2 Metrics

The performance of all networks are compared using various metrics mentioned below, which are determined from a matrix called confusion matrix. The performance metrics used here are accuracy, precision, recall and F1-score [12].

$$\text{Classification accuracy} = \frac{\text{correct prediction}}{\text{All Predicion}}$$

$$\text{Precision} = \frac{\text{positive predictedcorrectly}}{\text{all positive prediction}}$$

$$\text{Recall (Sensitivity)} = \frac{predicted\ to\ be\ positive}{all\ positive\ observation}$$

$$\text{F1Score} = 2\left(\frac{precision \times recall}{precision + recall}\right)$$

Table I. Performance comparison of different CNN architecture

CNN Architecture	Performance Analysis **%**	10 Epoch	20 Epoch	30 Epoch
ResNet50	Accuracy%	56	60	56
	Precision%	17	20	18
	Recall%	32	33	33
	F1 Score%	24	25	24
AlexNet	Accuracy%	54	59	58
	Precision%	44	55	50
	Recall%	36	46	46
	F1 Score%	34	47	48
VGG16	Accuracy%	61	63	59
	Precision%	20	21	21
	Recall%	33	33	33
	F1 Score%	25	26	26

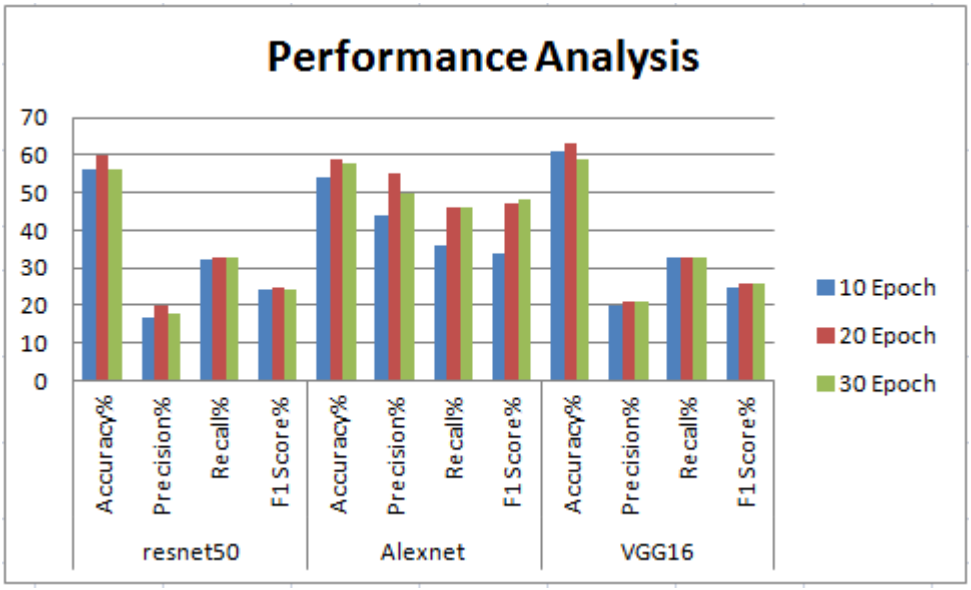

Fig 6: Chart showing the performance of the three network

On comparing the performance of above three networks for the proposed method, after performing 10, 20, 30 Epochs of training AlexNet gives better results in terms of precision, recall and F1Score compared to other architecture. Even in accuracy it is almost equal to VGG16. So on comparison the chart given in Fig 6 depicts that AlexNet is the better architecture out of the three. Table I above shows the comparison of better performing architecture with recent similar methodologies.

5. CONCLUSION

Our goal was to explore a core area of psychology, the study of emotion, using a huge and novel social media dataset. The aim of this work concentrated on comparing the performance of AlexNet, ResNet50 and VGG16 the undertaking of breaking down the sentiment in Twitter images with different performance metrics. The correlation of these architectures presents the benefits, in which they do not need any tedious pre-processing, and they are faster and a profitable training performance. The goal was to find the more suitable model. The AlexNet model with all three layers has given overall better results, which is highly statistically

significant and demonstrates the effectiveness of analyzing images with the combination of CNN and fine-tuning adjustment. In the future, we will assess our model on other picture and text upgrades datasets that have been developed for psychological studies and investigate whether human judges are pretty much precise than our model. Finally, we will explore other psychological components of the structure of emotion, for example day to day and day of week trends in emotion.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] V. Roseline, G.H. Chellam, PS-POS embedding target extraction using CRF and BiLSTM, Int. J. Adv. Sci. Technol. 29(3) (2020), 10984-10995.

[2] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[3] P. Valdez, A. Mehrabian, Effects of color on emotions., J. Experiment. Psychol.: Gen. 123 (1994), 394–409.

[4] L.-C. Ou, M.R. Luo, A. Woodcock, A. Wright, A study of colour emotion and colour preference. Part I: Colour emotions for single colours, Color Res. Appl. 29 (2004), 232–240.

[5] V. Yanulevskaya, J.C. van Gemert, K. Roth, A.K. Herbold, N. Sebe, J.M. Geusebroek, Emotional valence categorization using holistic image features, in: 2008 15th IEEE International Conference on Image Processing, IEEE, San Diego, CA, USA, 2008: pp. 101–104.

[6] A. Sartori, Y. Yan, G. Ozbal, et al. Looking at Mondrian's victory Boogie-Woogie: What do I feel? in: Proceedings of the 24th International Conference on Artificial Intelligence, 2015, pp. 2503–2509.

[7] X. Lu, Z. Lin, H. Jin, J. Yang, J.Z. Wang, RAPID: Rating Pictorial Aesthetics using Deep Learning, in: Proceedings of the 22nd ACM International Conference on Multimedia, ACM, Orlando Florida USA, 2014: pp. 457–466.

[8] D. Borth, R. Ji, T. Chen, T. Breuel, S.-F. Chang, Large-scale visual sentiment ontology and detectors using adjective noun pairs, in: Proceedings of the 21st ACM International Conference on Multimedia - MM '13, ACM Press, Barcelona, Spain, 2013: pp. 223–232.

[9] Q. You, J. Luo, H. Jin, J. Yang, Robust image sentiment analysis using progressively trained and domain transferred deep networks. In: Proceedings of the twenty-ninth AAAI conference on artificial intelligence, (2015), pp. 381–388.

[10] S.R. Flaxman, Y.-X. Wang, A.J. Smola, Who Supported Obama in 2012?: Ecological Inference through Distribution Regression, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Sydney NSW Australia, 2015: pp. 289–298.

[11] A.C. Wojnicki, D. Godes, Word-of-Mouth as Self-Enhancement (April 25, 2008). HBS Marketing Research Paper No. 06-01, http://dx.doi.org/10.2139/ssrn.908999

[12] M.Z. Alom, T.M. Taha, C. Yakopcic, et al. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches, ArXiv:1803.01164 [Cs]. (2018).

[13] A. Hu, S. Flaxman, Multimodal Sentiment Analysis To Explore the Structure of Emotions, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. (2018) 350–358.

[14] J. Islam, Y. Zhang, Visual Sentiment Analysis for Social Images Using Transfer Learning Approach, in: 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), IEEE, Atlanta, GA, USA, 2016: pp. 124–130.

[15] A.P. Rahmathunneesa, K.V. Ahammed Muneer, Performance Analysis of Pre-trained Deep Learning Networks for Brain Tumor Categorization, in: 2019 9th International Conference on Advances in Computing and Communication (ICACC), IEEE, Kochi, India, 2019: pp. 253–257.

[16] S. Tammina, Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images, Int. J. Sci. Res. Publ. 9 (2019), 143-150.