International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 05 | May - 2022 Impact Factor: 7.185 ISSN: 2582-3930

EMOTION CLASSIFICATION – DEEP LEARNING

R. Tamil Roja(tamilrojajpr@gmail.com)

Anjana Venkatasubramani (anjanavenkatasubramani2000@gmail.com)

K. Dhanalakshmi (smileydhana59@gmail.com)

D. Jeffrina Hebzi (jeffrinahebzi21@gmail.com)

Department Of Computer Science and Engineering
Jeppiaar Engineering College, Chennai, India

Abstract - Detecting the emotions using facial features is useful in the various fields. The convolutional neural network is a class of deep neural networks commonly applied to analyze and help to detect images. Here it's been used for facial emotion recognition such as human facial expressions, facial features, and emotions of faces Sequential forward selection algorithms and soft-max activation function are also done in ordering the emotions and finding out in which emotion it belongs. So that this study can easily get a person's feedback without any hindrance or lie from individuals and also this study can easily overcome the language barrier there are four types of datasets used here which are happy, sad, neutral, and angry. As a result of this study we got Ninety eight percent of accuracy in the future this study is expecting to add two more dataset which is guilt and fear so that it will also help us to detect criminals

***Key Words*:** *Deep Learning, TensorFlow, Keras, CNN*

1.INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. The term "data science" has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science. In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic. However, the definition was still in flux. The term "data science" was first coined in 2008 by D.J. Patil, and Jeff Hammerbacher, the pioneer leads of data and analytics efforts at LinkedIn and Facebook. In less than a decade, it has become one of the hottest and most trending professions in the market. Data science is the field of study that combines domain expertise, programming skills, and knowledge of mathematics and statistics to extract meaningful insights from data.

A. Artificial Intelligence

Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.Artificial intelligence (AI) is intelligence demonstrated by machines, as opposed to the natural intelligence displayed by humans or animals. Leading AI textbooks define the field as the study of "intelligent agents" any system that perceives its environment and takes actions that maximize its chance of achieving its goals.

I. **Learning processes.** This aspect of AI programming focuses on acquiring data and creating rules for how to turn the data into actionable information. The rules, which are called algorithms, provide computing devices with step-by-step instructions for how to complete a specific task.

II. **Reasoning processes.** This aspect of AI programming focuses on choosing the right algorithm to reach a desired outcome.

III. **Self-correction processes.** This aspect of AI programming is designed to continually fine-tune algorithms and ensure they provide the most accurate results possible.

B. Natural Language Processing (NLP)

Natural language processing (NLP) allows machines to read and understand human language. A sufficiently powerful natural language processing system would enable natural-language user interfaces and the acquisition of knowledge directly from human-written sources, such as newswire texts. Some straightforward applications of natural language processing include information retrieval, text

mining, question answering and machine translation.

Many current approaches use word co-occurrence frequencies to construct syntactic representations of text. "Keyword spotting" strategies for search are popular and scalable but dumb; a search query for "dog" might only match documents with the literal word "dog" and miss a document with the word "poodle". "Lexical affinity" strategies use the occurrence of words such as "accident" to assess the sentiment of a document. Modern statistical NLP approaches can combine all these strategies as well as others, and often achieve acceptable accuracy at the page or paragraph level. Beyond semantic NLP, the ultimate goal of "narrative" NLP is to embody a full understanding of common sense reasoning. By 2019, transformer-based deep learning architectures could generate coherent text.

C. Machine Learning

- Machine learning is to predict the future from past data. Machine learning (ML) is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed.
- Machine learning focuses on the development of Computer Programs that can change when exposed to new data and the basics of Machine Learning, implementation of a simple machine learning algorithm using python. Process of training and prediction involves use of specialized algorithms

D. *Deep Learning*

Deep learning is a branch of machine learning which is completely based on artificial neural networks, as neural network is going to mimic the human brain so deep learning is also a kind of mimic of human brain. It's on hype nowadays because earlier we did not have that much processing power and a lot of data. A formal definition of deep learning is- neurons Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones. In brain approximately 100 billion neurons all together this is a picture of an individual neuron and each neuron is connected through thousands of their neighbors. The question here is how it recreates these neurons in a computer. So, it creates an artificial structure called an artificial neural net where we have nodes or neurons. It has some neurons for input value and some for output value and in between, there may be lots of neurons interconnected in the hidden layer.

2. Body of Paper

Preparing Dataset

This dataset contains approximately 670 train 182 test image records of features extracted, which were the classified into 4 classes:

- Angry
- Cry
- Happy
- Neutral

PROPOSED METHODOLOGY

CNNs are regularized versions of Multilayer perceptron. Multilayer perceptron usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer.

i.Conv2d: The 2D convolution is a fairly simple operation at heart: you start with a kernel, which is simply a small matrix of weights. This kernel "slides" over the 2D input data, performing an elementwise multiplication with the part of the input it is currently on, and then summing up the results into a single output pixel. The kernel repeats this process for every location it slides over, converting a 2D matrix of features into yet another 2D matrix of features. The output features are essentially, the weighted sums (with the weights being the values of the kernel itself) of the input features located roughly in the same location of the output pixel on the input layer.

Whether or not an input feature falls within this "roughly same location", gets determined directly by whether it's in the area of the kernel that produced the output or not. This means the size of the kernel directly determines how many (or few) input features get combined in the production of a new output feature.

This is all in pretty stark contrast to a fully connected layer. In the above example, we have 5×5=25 input features, and 3×3=9 output features. If this were a standard fully connected layer, you'd have a weight

matrix of 25×9 = 225 parameters, with every output feature being the weighted sum of every single input feature. Convolutions allow us to do this transformation with only 9 parameters, with each output feature, instead of "looking at" every input feature, only getting to "look" at input features coming from roughly the same location. Do take note of this, as it'll be critical to our later discussion.

ii. ***MaxPooling2D layer****:* Downsamples the input along its spatial dimensions (height and width) by taking the maximum value over an input window (of size defined by pool_size) for each channel of the input. The window is shifted by strides along each dimension. The resulting output, when using the "valid" padding option, has a spatial shape (number of rows or columns) of: output_shape = math.floor((input_shape - pool_size) / strides) + 1 (when input_shape >= pool_size)

The resulting output shape when using the "same" padding option is: output_shape = math.floor((input_shape - 1) / strides) + 1

iii. ***Alexnet*** **:** AlexNet is the name of a convolutional neural network which has had a large impact on the field of machine learning, specifically in the application of deep learning to machine vision. AlexNet was the first convolutional network which used GPU to boost performance. AlexNet architecture consists of 5 convolutional layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, and 1 softmax layer. Each convolutional layer consists of convolutional filters and a nonlinear activation function ReLU. The pooling layers are used to perform max pooling.

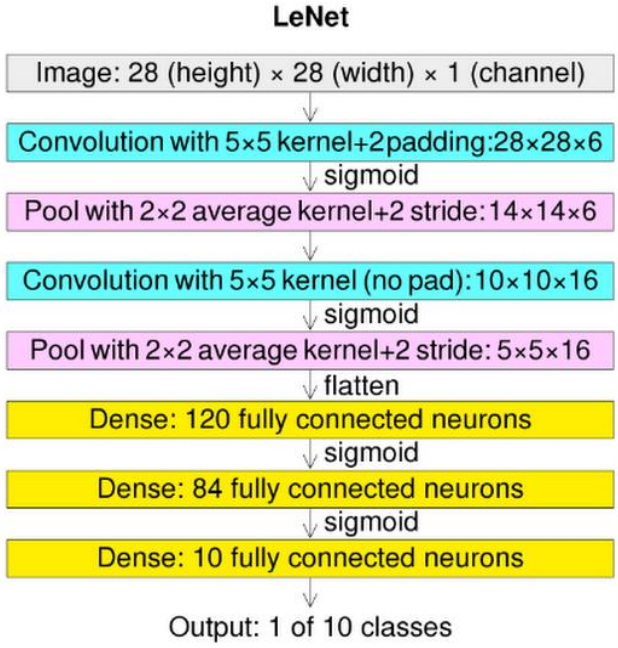
Convolutional layers*:* Convolutional layers are the layers where filters are applied to the original image, or to other feature maps in a deep CNN. This is where most of the user-specified parameters are in the network. The most important parameters are the number of kernels and the size of the kernels.

Pooling layers*:* Pooling layers are similar to convolutional layers, but they perform a specific function such as max pooling, which takes the maximum value in a certain filter region, or average pooling, which takes the average value in a filter region. These are typically used to reduce the dimensionality of the network.

Dense or fully connected layers*:* Fully connected layers are placed before the classification output of a CNN and are used to flatten the results before classification. This is similar to the output layer of an MLP.

iv. ***Lenet****:* LeNet was one among the earliest convolutional neural networks which promoted the event of deep learning. After innumerous years of analysis and plenty of compelling iterations, the end result was named LeNet.

Architecture of LeNet-5: LeNet-5 CNN architecture is made up of 7 layers. The layer composition consists of 3 convolutional layers, 2 subsampling layers and 2 fully connected layers.



Convolutional layers: Convolutional layers are the layers where filters are applied to the original image, or to other feature maps in a deep CNN. This is where most of the user-specified parameters are in the network. The most important parameters are the number of kernels and the size of the kernels.

Pooling layers: Pooling layers are similar to convolutional layers, but they perform a specific function such as max pooling, which takes the maximum value in a certain filter region, or average pooling, which takes the average value in a filter region. These are typically used to reduce the dimensionality of the network.

Dense or Fully connected layers: Fully connected layers are placed before the classification output of a CNN and are used to flatten the results before classification. This is similar to the output layer of an MLP.

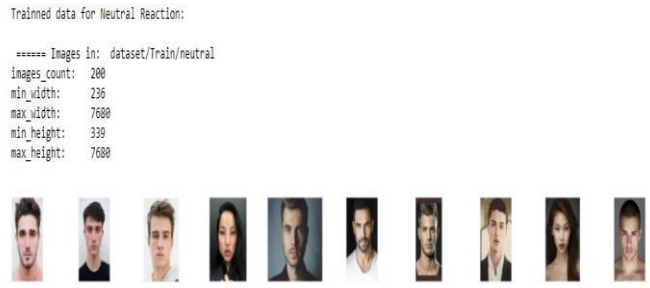
3. Working Process

i. ***Import the given image****:* We have to import our data set using keras preprocessing image data generator function also we create size, rescale, range, zoom range, horizontal flip. Then we import our image dataset from folder through the data generator function. Here we set train, test, and validation also we set target size, batch size and class-mode from this function we have to train using our own created network by adding layers of CNN.

- *Angry*

- *Happy*

- ***Neutral***



- ***Cry***

ii. To train the module by given image datasets: To train our dataset using classifier and fit generator function also we make training steps per epoch's then total number of epochs, validation data and validation steps using this data we can train our dataset.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 75, 75, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 37, 37, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 12, 12, 128)       36992
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 6, 6, 128)         0
_________________________________________________________________
flatten (Flatten)            (None, 4608)              0
_________________________________________________________________
dense (Dense)                (None, 256)               1179904
_________________________________________________________________
dense_1 (Dense)              (None, 4)                 1028
=================================================================
Total params: 1,218,820
Trainable params: 1,218,820
Non-trainable params: 0
```

iii. Working process of layers in CNN model:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.

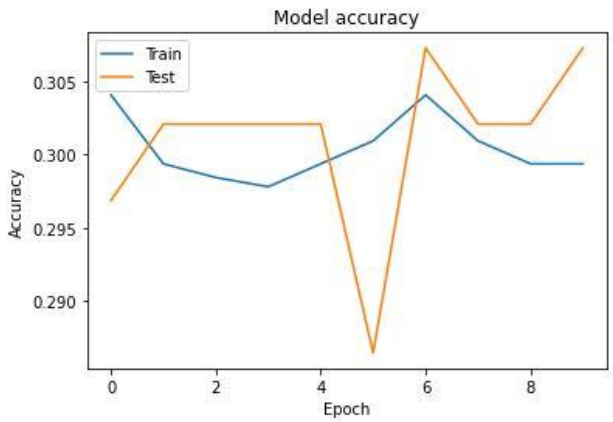
The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. Their network consists of four layers with 1,024 input units, 256 units in the first hidden layer, eight units in the second hidden layer, and two output units.

Input Layer:

Input layer in CNN contain image data. Image data is represented by three dimensional matrixes. It needs to reshape it into a single column. Suppose you have image of dimension 28 x 28 =784, it need to convert it into 784 x 1 before feeding into input.

Convo Layer:

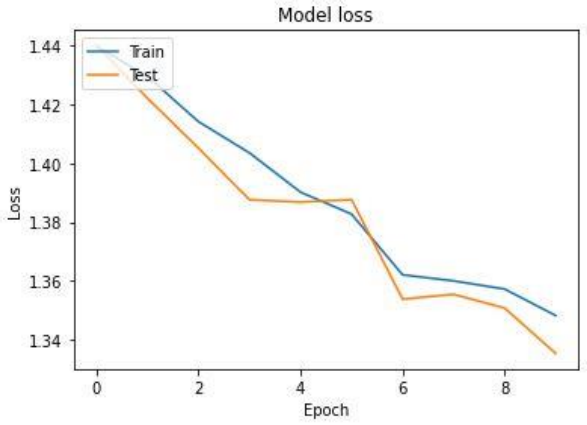
Convo layer is sometimes called feature extractor layer because features of the image are get extracted within this layer. First of all, a part of image is connected to Convo layer to perform convolution operation as we saw earlier and calculating the dot product between receptive field (it is a local region of the input image that has the same size as that of filter) and the filter. Result of the operation is single integer of the output volume. Then the filter over the next receptive field of the same input image by a Stride and do the same operation again. It will repeat the same process again and again until it goes through the whole image. The output will be the input for the next layer.



A. Pooling Layer: Pooling layer is used to reduce the spatial volume of input image after convolution. It is used between two convolution layers. If it applies FC after Convo layer without applying pooling or max pooling, then it will be computationally expensive. So, the max pooling is only way to reduce the spatial volume of input image. It has applied max pooling in single depth slice with Stride of 2. It can observe the 4 x 4 dimension input is reducing to 2 x 2 dimensions**.**

B. Fully Connected Layer (FC): Fully connected layer involves weights, biases, and neurons. It connects neurons in one layer to neurons in another layer. It is used to classify images between different categories by training.

c. Softmax / Logistic Layer: Softmax or Logistic layer is the last layer of CNN. It resides at the end of FC layer. Logistic is used for binary classification and softmax is for multi-classification.

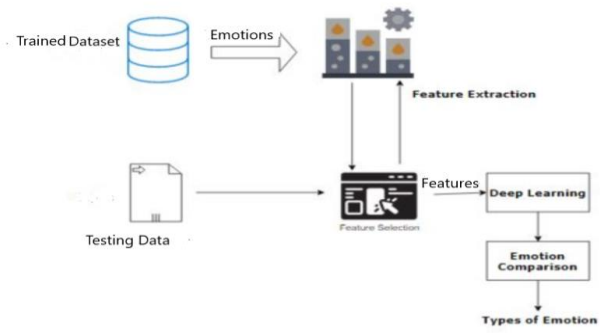


d. ***Output Layer:***

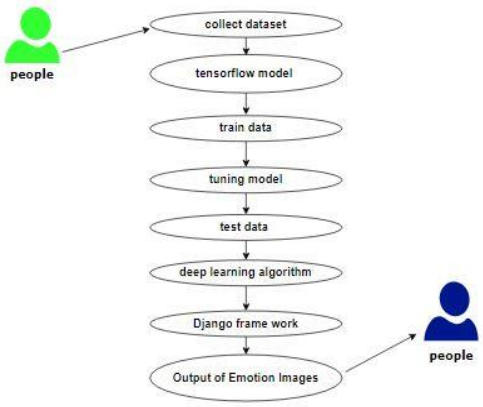
Output layer contains the label which is in the form of one-hot encoded. Now you have a good understanding of CNN.

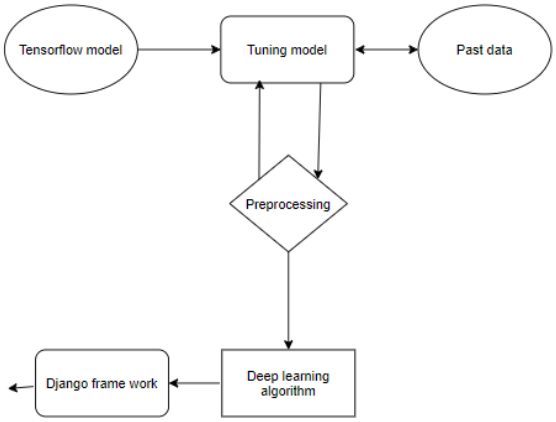
Figures :

Activity Diagram

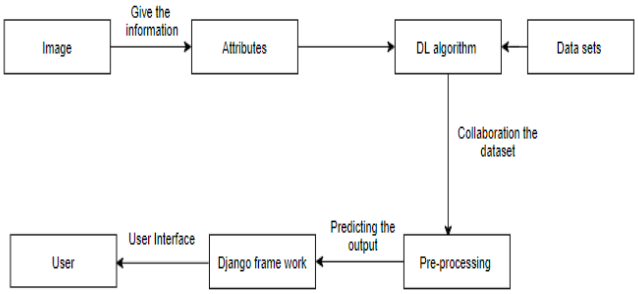


Use case diagram

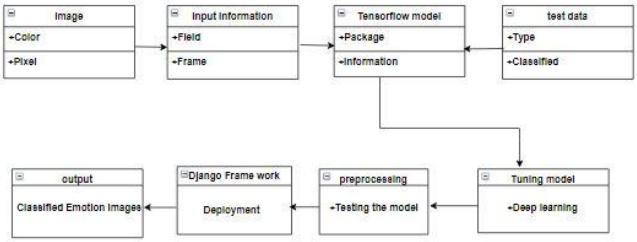




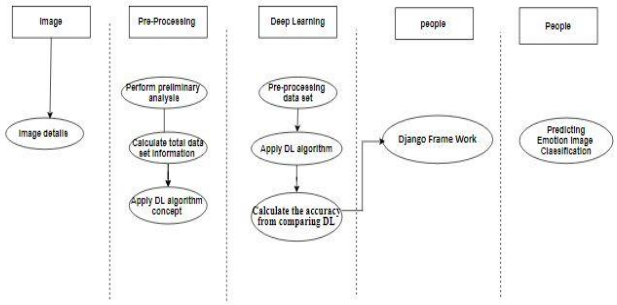
Collaboration Diagram



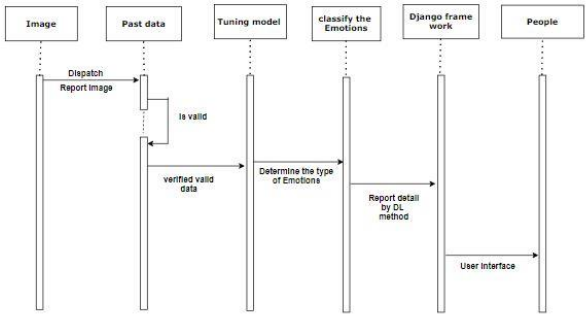
Class Diagram



Activity Diagram



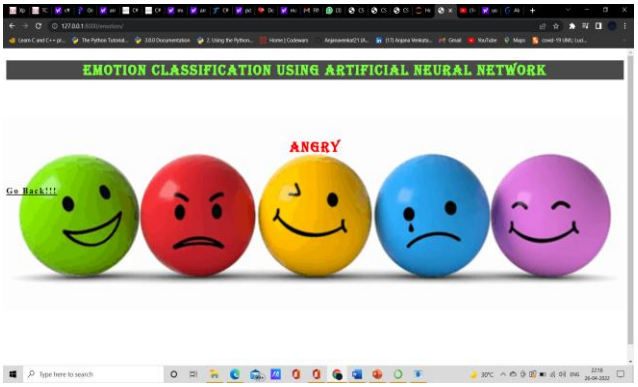
Sequence Diagram

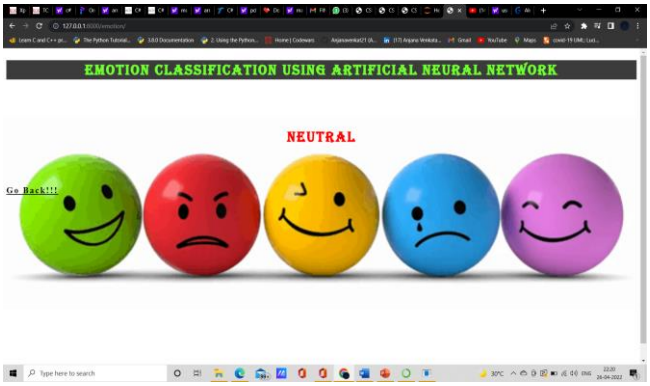


ER-Diagram

Result & Screenshots

- Different emotions has been trained and analysed properly based on the classifier
- Results showed the accuracy around 90% in the emotions of sad, happy, anger and neutral





CONCLUSIONS

In this project, a research to classify facial emotions over static facial images using deep learning techniques was developed. This is a complex problem that has already been approached several times with different techniques. While good results have been achieved using feature engineering, this project focused on feature learning, which is one of DL promises. While feature engineering is not necessary, image pre-processing boosts classification accuracy. Hence, it reduces noise on the input data. Nowadays, facial emotion detection software includes the use of feature engineering. A solution totally based on feature learning does not seem close yet because of a major limitation. Thus, emotion classification could be achieved by means of deep learning techniques.

FUTUREWORK

By training the large number of data sets and by using batch optimization we can improve very good accuracy among more emotions.Training, pre-training on each emotion, and using a larger dataset definitely improves the overall network's performance.Hence, they should be addressed in future research on this topic.

REFERENCES

[1] W. Zhang, P. Song, D. Chen, C. Sheng and W. Zhang, "Cross-corpus Speech Emotion Recognition Based on Joint Transfer Subspace Learning and Regression," in *IEEE Transactions on Cognitive and Developmental Systems*, doi: 10.1109/TCDS.2021.3055524.

[2] Lewenberg, Yoad & Bachrach, Yoram & Volkova, Svitlana. (2015). Using emotions to predict user interest areas in online social networks. 1-10. 10.1109/DSAA.2015.7344887.

[3] Firdaus, Mauajama & Thakur, Nidhi & Ekbal, Asif. (2021). Multi-Aspect Controlled Response Generation in a Multimodal Dialogue System using Hierarchical Transformer Network. 1-8. 10.1109/IJCNN52387.2021.9533886.

[4] Madhavi, Makarand & Gujar, Isha & Jadhao, Viraj & Gulwani, Reshma. (2022). Facial Emotion Classifier using Convolutional Neural Networks for Reaction Review. ITM Web of Conferences. 44. 03055. 10.1051/itmconf/20224403055.

[5] Zhang, Tong, et al. "Spatial–temporal recurrent neural network for emotion recognition."IEEE transactions on cybernetics49.3 (2018): 839-847

[6] S, Manisha & Nafisa, H & Gopal, Nandita & Anand, Roshni. (2021). Bimodal Emotion Recognition using Machine Learning. International Journal of Engineering and Advanced Technology. 10. 189-194. 10.35940/ijeat.D2451.0410421.

[7] Livingstone, S.R. and Russo, F.A, 2018, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English", PloS one, 13(5).

[8] Yadav, S. and Shukla, S, 2016, "Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification", 2016, IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, pp. 78-83.

[9] Yoad Lewenberg, Yoram Bachrach, Svitlana Volkova," Using Emotions to Predict User Interest Areas in Online Social Networks", 2015 IEEE

[10] James A. Russell, "Emotion in Human Consciousness Is Built in Core Affect", Journal of Consciousness Studies,12,No.8-10,pp 26-42, 2005.

[11]. DavidWatson; Auke Tellegan, "Towards a consensual structure of Mood", Psychological Bulletin, Vol. 98, No. 2. 219-235,1985.

[12]. Hugo Lövheim, A new three-dimensional model for emotions and monoamine neurotransmitters, Medical Hypotheses 78 (2012) 341–348. Processing and its Applications.Research in Computing Science 46, 2010, pp. 131- 142.

[13]. Paul Ekman, "Universals and Cultural Differences in Facial Expressions of Emotion", Nebraska Symposium on Motivation, Vol 19,1971.

[14]. Liza Wikarsa, Sherly Novianti Thahir, "A text mining application of Emotion Classifications of Twitters Users using Nave Bayes Method", IEEE 2015.

[15]. Li Yu, Zhifan Yang, Peng Nie, Xue Zhao, Ying Zhang, "Multi-Source Emotion Tagging for Online News", 12th Web Information System and Application Conference 2015.

[16]. Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, Amit P. Sheth," Harnessing Twitter „Big Data" for Automatic Emotion Identification", 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust.

[17]. K Dhanasekaran and R Rajeswari, "Text feature classification approach for effective information extraction via discriminative sequence analysis", International Journal of Applied Engineering Research, Vol. 10 (1), pp.2067- 2079, 2015.

[18]. http://knoesis.org/projects/emotion.

[19]. Sanket Sahu, suraj Kumar Rout, Debasmit Mohanty,"Twitter Sentiment Analysis:A more enhanced way of classification and scoring", 2015 IEEE International Symposium on Nanoelectronic and Information Systems.

[20] Dilbag Singh "Human Emotion Recognition System," in August 2012 MECS(http://www.mecspress.org/) DOI10.5815/ijigsp.2012.08.07).

[21] Zhiwei Deng, Rajitha Navarathna, Peter Carr, Stephan Mandt, Yisong Yue, Iain Matthews, "Factorized Variational Auto encoders for Modelling Audience Reactions to Movies", Greg Mori Simon Fraser University, Disney Research, Caltech.

[22] S. P Khandait, Dr.R.C. Thool & P.D. Khandait, "Automatic Facial Feature Extraction and Expression Recognition based on Neural Network", (IJACSA) International Journal of Advanced Computer Science and Applications. 2, No.1, January 2011.

[23] Octavio Arriaga, Paul G. Ploger, Matias Valdenegro, "Real-time Convolutional Neural Networks for Emotion and Gender Classification".

[24] John Gideon,Soheil Khorram,Zakaria Aldeneh,Dimitrios Dimitriadis2,Emily Mower Provost, "Progressive Neural Networks for Transfer Learning in Emotion Recognition", University of Michigan at Ann Arbor, IBM T. J. Watson Research Centre.

[25] Prathap Nair, Andrea Cavallaro, "3-D Face Detection, Landmark Localization, and Registration Using a Point Distribution Model", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 11, NO. 4, JUNE 2009.

[26] Jayalekshmi J, Tessy Mathew, "Facial Expression Recognition and Emotion Classification System for Sentiment Analysis", 2017 International Conference on Networks & Advances in Computational Technologies (2017).

[27] Kamil Topal, Gultekin Ozsoyoglu, "Movie Review Analysis:Emotion Analysis of IMDb Movie Reviews", 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

[28] Anurag De , Ashim Saha, "A Comparative Study on different approaches of Real Time Human Emotion Recognition based on Facial Expression Detection", 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA),IMS Engineering College, Ghaziabad, India.