

Sentiment Analysis of Images using Machine Learning Techniques

Yash Gherkar¹, Parth Gujar¹, Amaan Gaziyani¹, and Siddhi Kadu²

¹Student, Department of Computer Engineering, Ramrao Adik Institute of Technology, DY Patil Deemed to be University, Nerul, Navi Mumbai

²Assistant Professor, Department of Computer Engineering, Ramrao Adik Institute of Technology, DY Patil Deemed to be University, Nerul, Navi Mumbai

Abstract. Sentiment analysis is the process of identifying the idea of a text. People share the comments on social media stating their knowledge of the event and would like to know if most other people had a good or bad experience at the same event. This distinction can be made through Emotional-Analysis. Sentiment analysis captures informal text comments, posts and images from all comments shared by different users and classifies comments into different categories as neutral, negative or positive. This is also called as polarity separation. Various different types of ML and in-depth learning methods may be utilised in Sentiment Analysis like Support Vector Machines, NB, Haar Cascade, LBPH, CNN, etc. Emerging rise in popularity in Social Media has established a trend of posting images in restaurants to express their opinion on the food, ambience, etc which can be a useful resource to obtain opinion and feedback from the Customers. In this paper, the implementation of Sentiment Analysis on images containing users along with their faces from the restaurants review revealing it more efficacious in classifying and identifying sentiments of review-images.

Keyword :

Sentiment Analysis, Product Reviews,
Image Sentiment Analysis, LBPH, Haar Cascade

1 Introduction

Online updates become one of the main ways to explore a wide range of options. The user might check out Yelp reviews when choosing a restaurant. Review's usefulness is found in their role in accurately capturing prior customer information. In reality, one important piece of information that needs to be obtained get from the reviews is the feeling that the consumer is expressing, in the end whether you have had a good or a bad experience. Understanding emotions is an important and important function of analysing reviews, as emotions may express users' preferences, furthermore weaknesses and strengths. Information similar to these are crucial for designing the product, their recommendation, advertising, etc.

There has been extensive work done on detecting sentiments from Text reviews on any product but not many focus on the review images people post online. Most of the research paper and projects have implemented either Text based Sentiment Analysis on reviews such Amazon products review, etc or have implemented facial sentiment analysis without any real word application. With such massive growth in social media, it has provided a huge platform for influencers, critiques, bloggers, who post each and every experience they have with a business or a product. Even when a casual social media user visits a restaurant, it is observed that they take multiple Image to share in online

with their peers, followers, etc. Yet even with such a massive food trend online, there has been a very few projects that have researched in extracting opinions from these images. The images that were found on Yelp indicate that a good majority of people are willing to upload their own pictures while dining. These images which contain facial expression can be a great resource for extracting opinions, it can also be used by business to understand their user's feedback better.

The main contribution of this study is that it provides the users a platform to evaluate business with the aid of Image reviews posted by others users. It also enables the businesses to acquire a more in-depth feedback which involves sentiments expressed by the users utilizing facial expressions. By the information gathered from such different platforms, a model was developed that would help users find the restaurant that serves the type of food they are looking for. Also, it would help the restaurant owners in improving the quality of food based on the reviews submitted by their customers.

The paper starts with the Literature Survey, that includes various papers that were critically analysed and were used for the research purpose. The next part contains the Proposed Methodology, consisting of all the techniques used in the project and it also includes the System Design of the project. Result Analysis, the next part, contains the analysis of the project, its accuracy calculated through Confusion matrix and the various results, the project produced in the testing stage. Then, the next part was the Conclusion of the project and its future scopes. The last part,

References, includes the work that were used to produce the paper.

2 Literature Survey

There are various scholarly publications such as journal articles, research work etc. that are done in the field related to Sentiment Analysis. Many such articles and research reports were referred before making the project report. The reports/articles that have influenced the project, the most are :

Jin Ye, et al.[1] proposed a project focusing on product reviews. They have used a multi-modal approach to detect sentiments in an e-commerce reviews. Pytorch method is used to implement this method. A pre-trained model called as ImageNet is used to fine tune the dataset and for textual models, words are inserted using 'nn.embedding' in Pytorch. Both of these are then used in an CNN or GRU layer. A pre-trained model called as ImageNet is used to fine tune the dataset and for textual modes, words are embedded using "nn-embedding" in Pytorch. Both of these are then used in an CNN or GRU layer. The accuracy for only Image based analysis was low and text was used to reinforce the accuracy. Anthony Hu, et al.[2] proposed the visual models, a pre-trained model called "Inception" model. This model is trained to recognize image features with a deep architecture of 22 layers. Different features such as colors and shapes of an image are extracted. As for the textual model, Natural Language Processing (NLP) is used which implements methods such as Word embedding, Sequence input, etc to extract sentiments from the text. The input image, Inception network The test start network appears in a high-resolution area that exceeds the LSTM layer. The two methods are then combined and fed by a dense layer. Also no consumer driven application was developed. The proposed method also provide an important tool for the increasing research of images both-memes and photographs-in the social-network.

Quoc-Tuan Truong, et al.[3] proposes a method that is particularly focused on the sentiment that are a review image is displayed. It's not a multi-modal approach and only focus on Image Sentiment Analysis. They realized that the feeling expressed in the review-image, probably is governed by 3 factors: the image aspect (the emotions coded in image); user feature (emotions that were indicated by the reviewer in the images); & the factor of the user (feelings associated with image due to something). The paper utilizes two different model i.e., User-oriented Model: uVS-CNN and then item-oriented CNN. User Oriented model has a slightly lower accuracy. It focuses on Review Images. The experiments proved much successful in recognizing review image's sentiments and it was followed on the images that were gathered from the reviews from restaurant. Pawel Tarnowski, et al.[4] proposed a model that was capable of recognising 7 states of emotions i.e. fear, anger, disgust, sadness, joy, neutral and surprise based on expression on their face. The classification was done with the help of six subject were performed using a kinect device, and the classifications was performed using MLP neural network and k-NN classifier. The emotions were calculated

using 2 methods-> subject-independent(all users together) and subject-dependent(separate for each user). They found out after analysing that the user-independent method was more accurate than user-dependent method in recognising the emotions from facial expressions.

Santosh Kumar, et al.[5] proposed a Image sentiment analysis model which was used on to detect sentiments from Facial Features using Eigen faces, Fisherfaces, LBP, Speeded Up Robust Features(SURF) and FLANN. It focuses on extracting sentiments from Facial Expressions only. Image Sentiment Analysis is used on to detect sentiments from facial features using Eigen faces, Fisherfaces, LBP, Speeded Up Robust Features(SURF). It focuses on extracting sentiments from facial sentiments from Facial Expressions only therefore No real-world application. Paul Viola, et al.[6] proposed a 38 layered classifier which was trained to detect frontal upright faces and set of faces and non-faces images were utilized in classifier's training. The classifier they presented had high detection rate with minimum computation time required. The classifier was also trained to classify and detect faces in certain difficult circumstances.

Xing Fang, et al.[7] proposed a project in which he collected the data in form of product reviews from amazon.com in which the product mainly belongs to 4 major category namely beauty, electronic, book, home. Then the reviews are categorised into negative, positive or neutral by max-entropy POS tagger. After that, multiple ML-Algorithms like the SVM, Random Forest, NB are used compared with each other to implement Text Based Sentiment Analysis. DataSet containing Sentiment Tokens/Values is used. Each word to be evaluated is given a polarity according to the words from the data set. The aggregate Sentiment score of these words are considered to determine the sentiment of provided text.

Akriti Jaiswal, et al.[8] suggested a model that is based on Deep-learning architecture and worked on Convolution Neural Network(CNN) to find the emotions of the images. The working performance of this model was calculated with the help of two datasets namely JAFFE and FER-2013. The accuracy of the model produced was 98.65 for JAFFE dataset and it was 70.14 for the FER-2013 dataset. Stuti Jindal, et al.[9] suggested a model where the prediction of the sentiment was done with the help of Convolution Neural Networks or CNN. Some exclusive experiments were conduction on the Flickr image dataset that were already labeled. They reached on the conclusion that their proposed CNN training achieved better performance than the other competitive image sentiment analysis models.

3 Proposed Methodology

3.1 Dataset

The Dataset used was majorly gathered from the GitHub repository that contained various different facial expressions along with the respective files that were sorted according to their respective sentiments expressed. A CSV

file containing the names of the image files with their Sentiment score is used to sort the images. Furthermore, the dataset was also gathered by the process of Web-Scraping using a python library 'BeautifulSoup' from the popular restaurant review website- Yelp.com and then the data that contained the reviews with the facial data was sorted and used. The rest of images for the dataset was gathered by downloading the images from the dataset used in [3].

3.2 System Design

Fig.1 shows the entire system flow. The data set is fetched from a GitHub repository that has Facial data set. Using a list of predetermined images classify the images into different folders. Train the classifier using the classified images. Using Haar-Cascade and LBPH the model is trained. The Images are then scraped off the URL of the restaurant page on yelp provided by the user, then the review images from that web page are fetched and then processed to detect the sentiments in those Images. The overall Sentiment along with processed Images are then presented to the user.

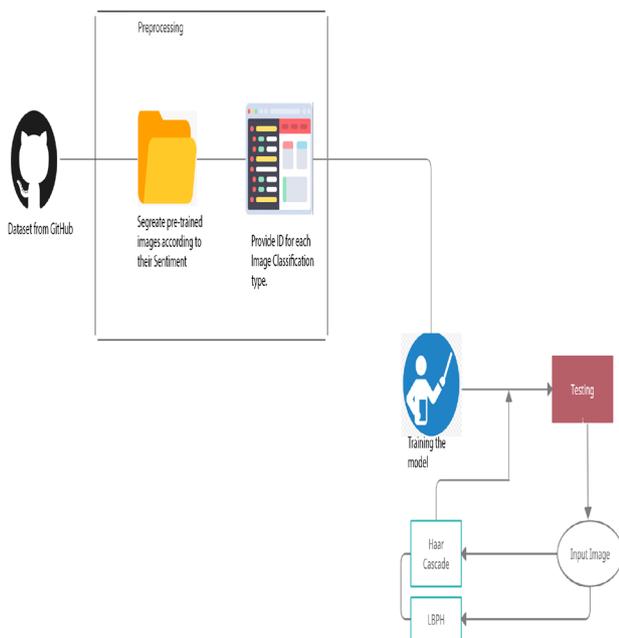


Fig.1. System design

3.3 Implementation Details

- 1.For this project, a data set which includes about 12,000 text and images from GitHub repository along with the images scraped of Yelp reviews that contained facial data. In that repository there is a separate CSV file which gives us the information of the images sentiments whether it was positive, neutral or negative.
- 2.The repository has separate folders based on the sentiments of the images. In that folder all the images with sentiment matching to that of folder's name has been included.
- 3.Then, the pre-processing of the data is carried out by removing the background of the images which have nothing

to do with the sentiment analysis.

4.Then, a training file is created, which is made by going through each of the images in our dataset. This training file will contain the information of all the expression their emotions from their faces in images.

5.After that, the model is provided with images which are obtained by scraping off the reviews in order to carry out testing.

3.4 Techniques Used

3.4.1 Haar-Cascade

Haar Cascade classifiers are a very effectual and productive and classifier way for object detection. Haar-Cascade is a ML built method where lots of negative positive images are being used in training the classifier. The main goal is to calculate the summation of all the pixels that are on dark-side of Haar-features & the summation of every pixels on the light-side of the Haar-features. Haar-Cascade separators is an effective way to find an object. This method was suggested by M. Jones P. Viola in [6] and the working of the same is shown in Fig.2.

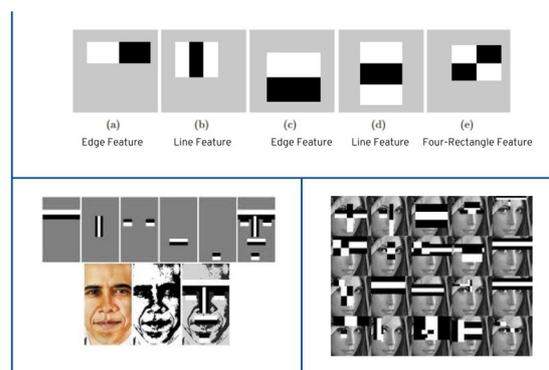


Fig.2. Haar Cascade Working[8]

Making a Haar Cascade Classifier

This algorithm is explained in 4 stages:

- Calculation of the Haar-Features
- Creation of Integral Images
- Use of Ada-boost
- Implementation of the Cascading Classifiers

It is of utmost importance remembering this algorithm needs lots of good facial images and negative non-facial images in training the classifier, similar to other ML models.

3.4.1.1 Calculation of the Haar Features

The primary goal is the collection of Haar-Features. A Haar feature is mainly the computation which is done performed in nearby rectangular areas somewhere in the detection-window. The computation includes adding up the intensities of the pixel in particular area or region and calculation of the difference between both the summation. Some of the examples of Haar Features are shown in Fig.3.

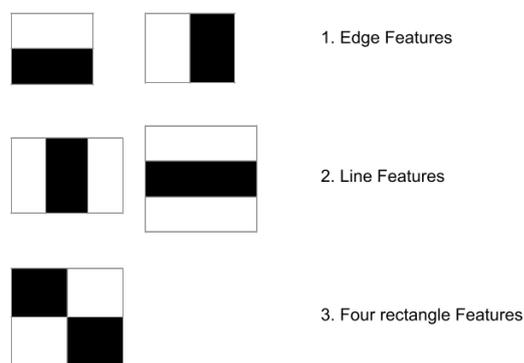


Fig.3. Haar Features[8]

These factors may be a little tricky in finding out for a big and huge picture. That's where the integral-pictures start to play as the no. of tasks is gets reduced by the use integral picture.

There are several advantages of Haar-Cascade that are it has low level of complexity, it has high level of computing. It captures the face more accurately and it has faster rate of recognition.

Some of the disadvantages of Haar-Cascade is that it is generally complex and hence slow, it's not suitable for black faces and turns out to be less accurate in this case. Haar-Cascade also requires long training time and if the lightening conditions are not favourable it have lots of false positive and false negatives. It is also less robust in case of occlusion.

3.4.2 LBPH

LBPH - Local Binary Pattern (LBPH) is an uncomplicated yet excessively effective texture-operator that highlights image's pixels with the help of thresh-holding the area of each of the pixel & then considering the output to be a binary number. Because of its computer simplicity & discriminatory power, the texture-operator based on LBPH have became one of the most popular feature in various programs. It can be viewed as way of integrating different mathematical models with a structure and analysis of texture. Probably, the majorly necessary attribute of LBPH-Operator in actual life programs is that it is robust in terms of changes in the monotonic gray-scale caused by, for ex-light changes. Another utmost-important asset is that it is simple, computation wise, which makes possible for analyzing the pictures in real time challenging situations.

The LBPH algorithm utilizes four parameter:

1. Neighbors: No. of sample points that are used in building the circular local binary pattern.
2. Radius: radius that is used to create a round binary area pattern design and constitutes the radius throughout the pixel in the center. It is normally adjusted as 1.
3. Grid X: Grid X represents summation of all the cells that are located in the horizontal-direction. The greater, the number of cells, the finer or superior the grid will be, the

higher the dimensional of the emerging feature-vector. It's normally adjusted to 8.

4. Grid Y: Grid Y represents the summation of all the cells that are present in the vertical-direction. The greater the no. of cells, the finer or superior the grid will be, the higher would be the dimensional of results in feature-vector. It's generally adjusted to 8.

3.4.2.1 Training the Algorithm

The foremost step forward is the algorithm's training. To do that, a database that contains the faces of people is used. There is also a need to set the ID for every images, so that the algorithm will use the information in identifying the inserted image will give the output. Photos of same individual must contain the unique identity. The LBPH calculating steps are.

Applying LBPH-Operation: The foremost part is to generate an intermediary image that grant the information of the actual image in a much superior way, by underlining the facial features. In order to do this, the algorithm utilizes the theory of a sliding-window, which is formulated on the radius, neighbors & parameters.

The figure(Fig.4.) shown beneath explains that method:

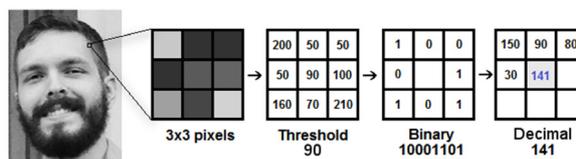


Fig.4. LBPH Working[9]

Based on the picture that was used in the above space, following steps are taken in the algorithm:

- Suppose, the face-image is present in gray-scale.
- Then the part of the picture is obtained in a 3x3 window.
- It maybe shown like a 3x3-matrix that contains intensity or magnitude of every pixel(0 255).
- After that, the value in the middle of the matrix is taken as thresh-hold.
- This thresh-hold value would be utilized in setting the values from 8 neighbours.
- For every adjacent of central value that is the thresh-hold, a fresh binary-value will be set. 1 will be set for values that are equal than or greater than the thresh-hold and 0 will be set for values that are lower than the thresh-hold.
- The matrix must have only binary value, then combination of each and every binary value from the matrix is taken and adjusted line-by-line accordingly to make a new binary-value. For example 10001101.
- Transform this obtained binary-value into the corresponding decimal value and set this decimal value at the center of the matrix, which actually is one of the pixel from the original picture.

-Finally, in the end of the LBP-Procedure, a latest image which have will have superior properties than the actual picture will be obtained as shown in Fig.5.

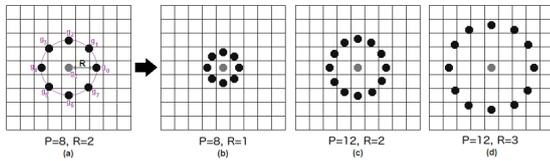


Fig.5. LBP-Bilinear Interpolation[7]

4 Result Analysis

```

id = 0
names = ['Positive', 'Neutral', 'Negative', 'None', 'None', 'None']
cam = cv2.VideoCapture(0)
cam.set(3, 640)
cam.set(4, 480)
minW = 0.1*cam.get(3)
minH = 0.1*cam.get(4)
img = cv2.imread("img1.jpg")
gray = cv2.cvtColor(img,cv2.COLOR_BGR2GRAY)
faces = faceCascade.detectMultiScale(
    gray,
    scaleFactor = 1.2,
    minNeighbors = 5,
    minSize = (int(minW), int(minH)),
)
for(x,y,w,h) in faces:
    cv2.rectangle(img, (x,y), (x+w,y+h), (0,255,0), 2)
    id, confidence = recognizer.predict(gray[y:y+h,x:x+w])
    # Check if confidence is less than 100 ==> "0" is perfect match
    if (confidence < 100):
        id = names[id]
        confidence = " {0}%".format(round(100 - confidence))
    else:
        id = "unknown"
        confidence = " {0}%".format(round(100 - confidence))
    cv2.putText(img, str(id), (x+5,y-5), font, 1, (255,255,255), 2)
    # cv2.putText(img, str(confidence), (x+5,y+h-5), font, 1, (255,255,0), 1)
cv2.imwrite("imgFinal.jpg",img)
print("\n [INFO] Done detecting and Image is saved")
    
```

Fig.6. Code to categorize images as positive,negative,neutral

The figure above(Fig.6.) represents the code to categorize sentiment of images (categories are positive, negative, neutral).Fig.7. shows the classification report based on our testing data. It consists of factors like false negative rate, accuracy, false discovery rate, etc. Fig.8. represents the confusion matrix for our testing data.Our model gives out an accuracy of 80.46 .

Measure	Value
Sensitivity	0.7963
Specificity	0.8182
Precision	0.8776
Negative Predictive Value	0.7105
False Positive Rate	0.1818
False Discovery Rate	0.1224
False Negative Rate	0.2037
Accuracy	0.8046
F1 Score	0.8350
Matthews Correlation Coefficient	0.6011

Fig.7. Classification Report

Fig.7. represents the various different parameters including Sensitivity, Specificity, Precision, F1 score, Accuracy etc that were calculated with the help of the number of images tested with the model. Similarly, the Confusion Matrix shown in Fig.8. is obtained with the help of number of true-positive,false-positive,true-negative and false-negative results obtained when the model was already trained with the Dataset and is now being tested. The accuracy of our model came out to be 80.46% that was calculated by the summation of True Negatives and True Positives.

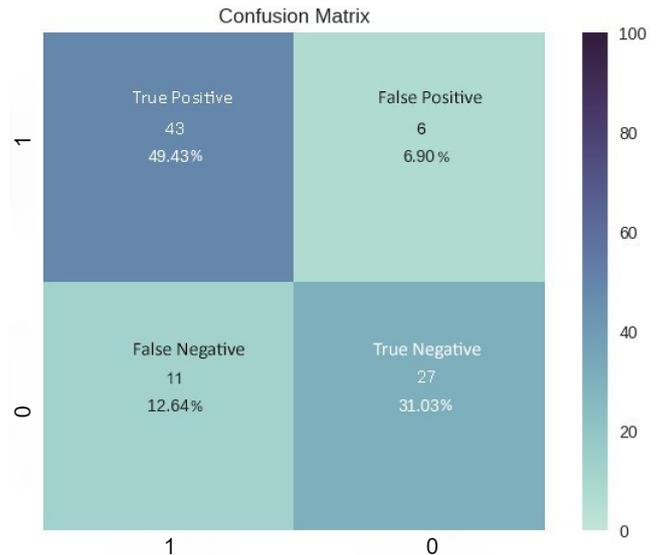


Fig.8. Confusion Matrix

The figures below[Fig.9., Fig.10., Fig.11.] represents images classified as either positive or neutral or negative based on the sentiments derived by the model from these images.



Fig.9. Image identified as a positive image

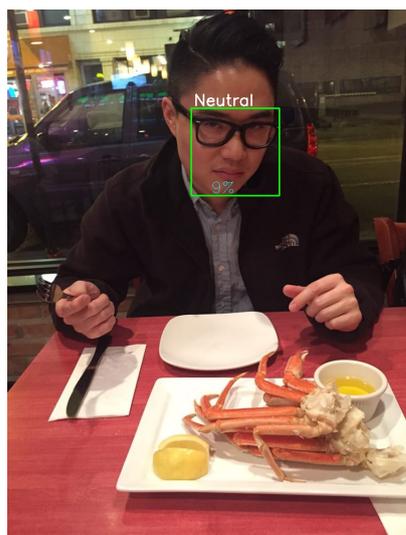


Fig.10. Image identified as a neutral image

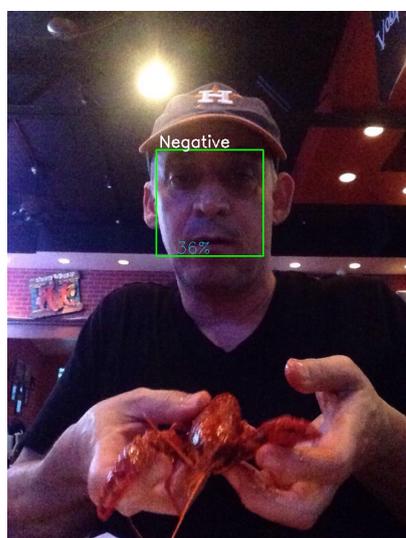


Fig.11. Image identified as a negative image

5 Conclusion

In the proposed system, the primary aim was to construct an accurate and dynamic sentiment analysis model that would help the restaurant owners to get accurate feedback from their customers. Our system uses Haar-Cascade to calculate the summation of every pixels that lies under the dark-area of the Haar-feature and summation of every pixels that lies under the light-area of the Haar-feature and LBP (Local Binary Pattern) that labels photo pixels by thresh-holding neighbor-hood of every pixel. Objectives like face recognition and emotion recognition were the primary objectives as they were the main point of concentration in the entire project. Other objectives like data collection, treating false positives were also taken care of to the best.

In future, detecting sentiments from food review images that don't contain any user using SIFT feature extraction method will be the primary goal along with using text shared by the reviewer in order to make the model more efficient and accurate. Also, the restaurant owner will be given a detailed report of the food items

reviewed by their customers that would help them make changes accordingly and serve their customers with the type of food they crave for.

References

- [1] J. Yin, X. Peng, Y. Qiao, H. Xing, J. Li, R. Ji, *Visual-textual sentiment analysis in product reviews*, *IEEE ICIP*, pp. 869-873, doi: 10.1109/ICIP.2019.8802992 (2019)
- [2] A. Hu, S. Flaxman, *Multimodal Sentiment Analysis to explore the structure of emotions*, *ACM SIGKDD, International conference on Knowledge Discovery and Data Mining*, pp. 350-358, doi: 10.1145/3219819.3219853 (2018)
- [3] Q.Tuan Truong, HW. Lauw, *Visual Sentiment Analysis for review images with item-oriented and user-oriented CNN*, *ACM International Conference on Multimedia*, pp. 1274-1282, doi: 10.1145/3123266.3123374 (2017)
- [4] P. Tarnowski, M. Kołodziej, A. Majkowski, RJ. Rak, *Emotion recognition using facial expressions*, *ICCS*, pp.1175-1184, doi: 10.1016/j.procs.2017.05.025 (2017)
- [5] S. Kumar, S. Jaiswal, R. Kumar, and SK. Singh, *Emotion Recognition Using Facial Expression*, *IGI Publications, USA*, pp. 1-19, doi: 10.4018/978-1-4666-8723-3.ch013 (2016)
- [6] P. Viola, M. Jones. *Rapid object detection using adaboosted cascade of simple features*, *IEEE Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol.* pp. 1-1, doi: 10.1109/CVPR.2001.990517 (2001)
- [7] X. Fang, J. Zhan, *Sentiment analysis using product review data*, *Journal of Big Data 2.1*, article no. 5, doi: 10.1186/s40537-015-0015-2 (2015)
- [8] A. Jaiswal, AK Raju, S. Deb, *Facial emotion detection using deep learning*, *IEEE INCET*, pp. 1-5, doi: 10.1109/INCET49848.2020.9154121 (2020)
- [9] S. Jindal, Sanjay Singh, *Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning*, *International Conference on Information Processing IEEE*, pp. 447-451, doi: 10.1109/INFOP.2015.7489424 (2015)
- [10] S. Ramalingam, F. Garzia, *Facial expression recognition using transfer learning*, *IEEE ICCST, International Carnahan Conference on Security Technology*, pp. 1-5, doi: 10.1109/CCST.2018.8585504 (2018)
- [11] AM. Jagtap, V. Kangale, K. Unune and P. Gosavi, *A Study of LBPH, Eigenface, Fisherface and Haar-like features for Face recognition using OpenCV*, *IEEE ICISS*, International Conference on Intelligent Sustainable System, pp. 219-224, doi: 10.1109/ISS1.2019.8907965 (2019)
- [12] N. Mehendale, *Facial Emotion Recognition using Convolutional Neural Networks (FERC)*. *SN Applied Sciences 2*, 446, pp.1-8, doi: 10.1007/s42452-020-2234-1 (2020)