Improved Model for Facial Expression Classification for Fear and Sadness Using Local Binary Pattern Histogram

Adebola K. Ojo[1*] and Temitope Ololade Idowu[1]

[1]Computer Science Department, University of Ibadan, Nigeria.

Authors' contributions

Original Research Article

Abstract

In this study, a Local Binary Pattern Histogram model was proposed for Facial expression classification for fear and sadness. There have been a number of supervised machine models developed and used for facial recognition in past researches. The classifier requires human effort to perform feature extraction which has led to unknown changes in the expression of human face and incomplete feature extraction and low accuracy. This study proposed a model for improving the accuracies for fear and sadness and to extract features to distinguish between fear and sadness. Images of different people of varying ages were extracted from two datasets got from Japanese female facial expression (jaffe) dataset and Cohn cade got from Kaggle. In other to achieve an incremental development, classification was done using Linear Support Vector Machine (LSVM) and Random Forest Classifier (RFC). The accuracy rates for the LSVM models, LSVM1 and LSVM2 were 88% and 87% respectively while the RFC models, RFC1 and RFC2, were 81% and 82% respectively.

_____
*Corresponding author: E-mail: adebola_ojo@yahoo.co.uk;

1 Introduction

Over the last decade, facial expression recognition has been increasingly attracting attention and has become an important issue in the scientific community. Today, Facial expression recognition systems have great potential for different applications, such as human computer interaction (HCI) and medical analysis. Facial expression recognition involves identification of cognitive activity [1], deformation of facial feature and facial movements. It is central to several leading theories of emotion and has been the focus of, at times, heated debate about issues in emotion science. Facial expression figures prominently in research on almost every aspect of emotion, including psychophysiology, neural correlates development, perception, addiction, social processes, depression and other emotion. In recent years, it has become a useful scheme for computers to effectively understand the emotional state of human beings. One critical step for facial expression recognition (FER) is to accurately extract emotional features. Certain expressions such as fear and sadness have lower recognition accuracy in most FER systems when compared with other expressions and this is due to the fact that sometimes, the same action units are activated for both expressions. Different methods of extracting features have been developed and used with different classifiers but most of them, based on research, have had low accuracies for classifying fear and sadness because these tend to form similar features on the face. Some methodologies have also combined feature extraction techniques which have yielded somewhat better accuracies for these expressions. In the conventional mechanisms of human computer interaction, we have voice or speech being recorded and recognized as well as images and gestures. Expressions such as fear and sadness, are lagging behind in terms of accuracy level or recognition rate when compared with the recognition rate or accuracy level of other expression. This study proposes a model for improving the facial expression classification for fear and sadness using the strength of two algorithms (Linear Support Vector Machine and Random Forest Classifier) to establish a better accuracy for fear and sadness. Histogram Oriented Gradients was used for face detection.

2 Literature Review

2.1 A conceptual review of face recognition classification

Facial expression recognition is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neutral networks. Facial expression has been a focus of research in human behaviour for over a hundred Years. It is central to several leading theories of emotion and has been the focus of, at times [2], heated debate about issues in emotion science. Facial expression figures prominently in research on almost every aspect of emotion, including psychophysiology, neural correlates development, perception, addiction, social processes, depression and other emotion.

2.2 Stages in facial expression recognition system

A basic automatic facial expression recognition system generally consists of three steps [3]: Face acquisition, facial feature extraction and representation, and facial expression classification.

- *Face acquisition:* Is a processing stage to automatically find the face region for the input image or sequences. It can be a detector to detect face for each frame or just detect face in the first frame and then track the face in the remainder of the video sequence. To handle large head motion, the head finder, head tracking, and pose estimation can be applied to a facial expression analysis system.
- *Facial feature extraction extract:* In facial feature extraction for expression analysis, there are mainly two types of approaches: Geometric feature-based methods and appearance-based methods. The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, and nose). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. With appearance-based methods, image filters, such as Gabor wavelets, are applied to either the whole-face or specific regions in a face image to extract a feature vector. Depending on the different facial feature extraction methods, the effects of

in-plane head rotation and different scales of the faces can be eliminated by face normalization before the feature extraction or by feature representation before the step of expression recognition.

- *Facial expression recognition* is the last stage of FER systems. The facial changes can be identified as facial action units or prototypic emotional expressions. The goal of face recognition is to gather data and analyse images- to make appropriate detection possible. Such data may be obtained from different Action unit. However, learners' emotions can be divided into six kinds of categories: sadness, happiness, surprise, fear, anger and disgust. Over the last decade, automatic facial expression recognition has been increasingly attracting attention and has become an important issue in the scientific community, since facial expressions are one of the most powerful, nature and immediate means for human beings to communicate their emotions and intentions [3]. However, recognition of facial expressions is a complex task as physiognomies of faces vary from one individual to another quite considerably due to age, ethnicity, gender, facial hair. Since the 1990s, most of the efforts have been made to develop theorems and methods for automatic facial expression recognition. Some of these works have attempted to extract the features of interest from still and dynamic facial images for representing facial expression.

Sometimes the facial expression analysis has been confused with emotion analysis in the computer vision domain. For emotion analysis, higher level knowledge is required. Although facial expressions can convey emotion, they can also express intention, cognitive processes, physical effort, or other intra- or interpersonal meanings. Interpretation is aided by context, body gesture, voice, individual differences and cultural factors as well as by facial configuration and timing.

2.3 Review of related works

It is a challenging and difficult task for a computer vision system to recognize an individual across different expressions. Automatic emotion recognition in faces is a hard problem, requiring a number of pre-processing steps which attempt to detect or track the face, to locate characteristic facial regions. [4] developed for face detection using neural network classifier to classify the emotion, Haar Cascades was used to detect the input image, a face, as the basis for the extraction of eyes and mouth and then through the Sobel Edge Detection to obtain the characteristic value. Experiments using JAFF database, the results show a low accuracy level for fear and sadness emotion as compared with the accuracy level for other emotions. [5] presented an automatic facial expression recognition system based on self-organizing feature maps. Firstly, Viola and jones Algorithm was used to detect a face from an image. After a human face is detected, a composite method was proposed to locate pupils so that the located face image can be rotated, trimmed and facial features and a multi-layer perceptron (MLP) was adopted for the classification.

Navdeep and Varinderjit [6] proposed KNN Regression algorithm with SURF feature for facial expression detection, Eigenfaces are constructed and the most relevant Eigenfaces have been selected using Principal Component Analysis (PCA). Kumari et al. [7] presented a new approach to facial expression recognition, which uses Wavelet for reducing the high dimensional data of facial expression images into a relatively low dimensional data and then uses K Nearest Neighbor (KNN) as the classifier for the expression classification afterwards. Kai [8] presented a temporal-reinforced approach to enhancing emotion recognition from facial images. Shape and texture models of facial images were computed by using active appearance model (AAM), from which facial feature points and geometrical feature values were extracted. The extracted features were used by relevance vector machine to recognize emotional states. Urvashi and Singhal [9] introduced a new technique to recognize human face artificially using Discrete Cosine Transform (DCT), PCA and Self-Organizing Map (SOM) neural network. PCA is a classical and successful method of dimension reduction. DCT is a well-known compression technique and Self-Organizing Map (SOP) act as a classifier and has been used for face space representation. Jizheng et al. [10] proposed a novel FER algorithm by exploiting the structural characteristics and the texture information hiding in the image space. Firstly, the feature points were marked by an active appearance model. Secondly, three facial features, which are feature point distance ratio coefficients, connection angle ratio coefficients and skin deformation energy parameter, were proposed to eliminate the differences among the individual, finally, a radical basis function neural network was utilized as the classifier or the FER.

Expression Fear and Sadness are lagging behind in terms of accuracy level or recognition rate when compared with the recognition rate or accuracy level of other expression. On this premise, this study is interested in collapsing the strength of two algorithms to establish a better accuracy for fear and sadness.

3 Methodology

3.1 Model overview

This study presents a model for facial expression recognition that performs relatively better at classifying fear and sadness emotions. The model makes use of the textural features of the face images to classify the emotion on the face. Fig. 1 shows an overview of the methodology and it starts by first detecting and isolating the face part of the images. It uses a histogram of oriented gradients to detect the facial landmarks and the images cropped to adjust the facial regions alone. The local binary pattern for the isolated faces is then extracted and the histograms calculated from the Local Binary Pattern. This is then used to train a classifier to recognize the different facial expressions.
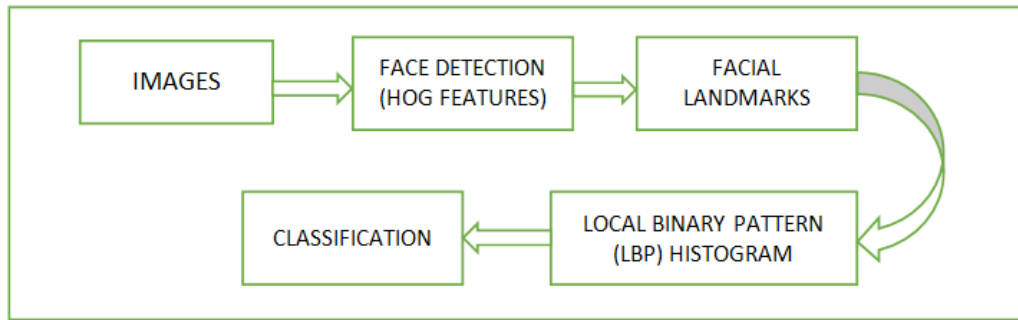

Fig. 1. The overview of the proposed facial expression methodology

3.2 Feature extraction

As with most image processing models, the first step is to reduce noise. This is done by passing the image through a Gaussian filter which smoothens the images out. A Gaussian filter is implemented by first creating a kernel (a square array of pixels) and multiplying the image by this kernel. To do this, the center pixel of the kernel is placed on the image pixel and the values in the original image multiplied with the pixels in the kernel that overlap. The values resulting from these multiplications are added up and that result is used for the value at the destination pixel. After this preliminary step, the images are then converted to grayscale and we proceed to detect and the face part of the image is detected and isolated using a Histogram of Oriented Gradient (HOG). A Histogram of Oriented Gradients (HOG) is a feature descriptor, generally used for object detection. A HOG relies on the property of objects within an image to possess the distribution of intensity gradients or edge directions. Firstly, the gradient of each pixel in the image, consisting of a magnitude and direction, is calculated. Then, the feature descriptors are calculated over blocks of pixels with 8 x 8 dimensions. These descriptor values for each pixel over 8 x 8 block are quantized into 9 bins, where each bin represents a directional angle of gradient and value in that bin, which is the summation of the magnitudes of all pixels with the same angle as shown in the equations 1 and 2.

$$g = \sqrt{g_x^2 + g_y^2} \tag{1}$$

$$\theta = \arctan\frac{g_y}{g_x} \tag{2}$$

g_x and g_y are respectively the horizontal and vertical components of the change in the pixel intensity. A window size of 128 x 144 is used for face images since it matches the general aspect ratio of human faces.

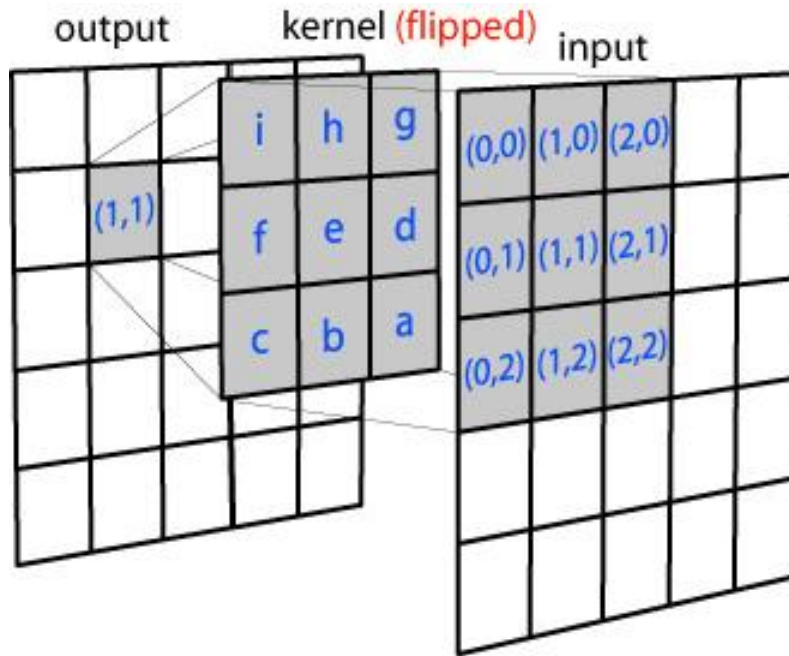Fig. 2 shows output of the image after applying HOG to it.

Fig. 2. Gaussian filter

Fig. 3. Isolating face using HOG

3.3 Local binary pattern

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighbourhood of each pixel and considers the result as a binary number [11,12,13]. First an empty LBP Image which is a 2D array the same size as the original face image is created and then for each pixel in the grayscale image of the isolated face, a neighbourhood is selected around the current pixel and the LBP value for the pixel calculated using the neighbourhood. The intensity of these surrounding pixels is considered and then compared to the value of the central pixel. If the intensity of the pixel is lower than or equal to that of the central pixel, it is computed as a 1 otherwise it is a 0 as shown in Fig. 4. Each of the binary values of the neighboring pixel is then put together to form a single binary value which is then converted to decimal and set as the value of the center pixel in the LBP Image. We start at the top-right point and work our way ***clockwise*** accumulating the binary string as we go along. We can then convert this binary string to decimal as described in Fig. 4.
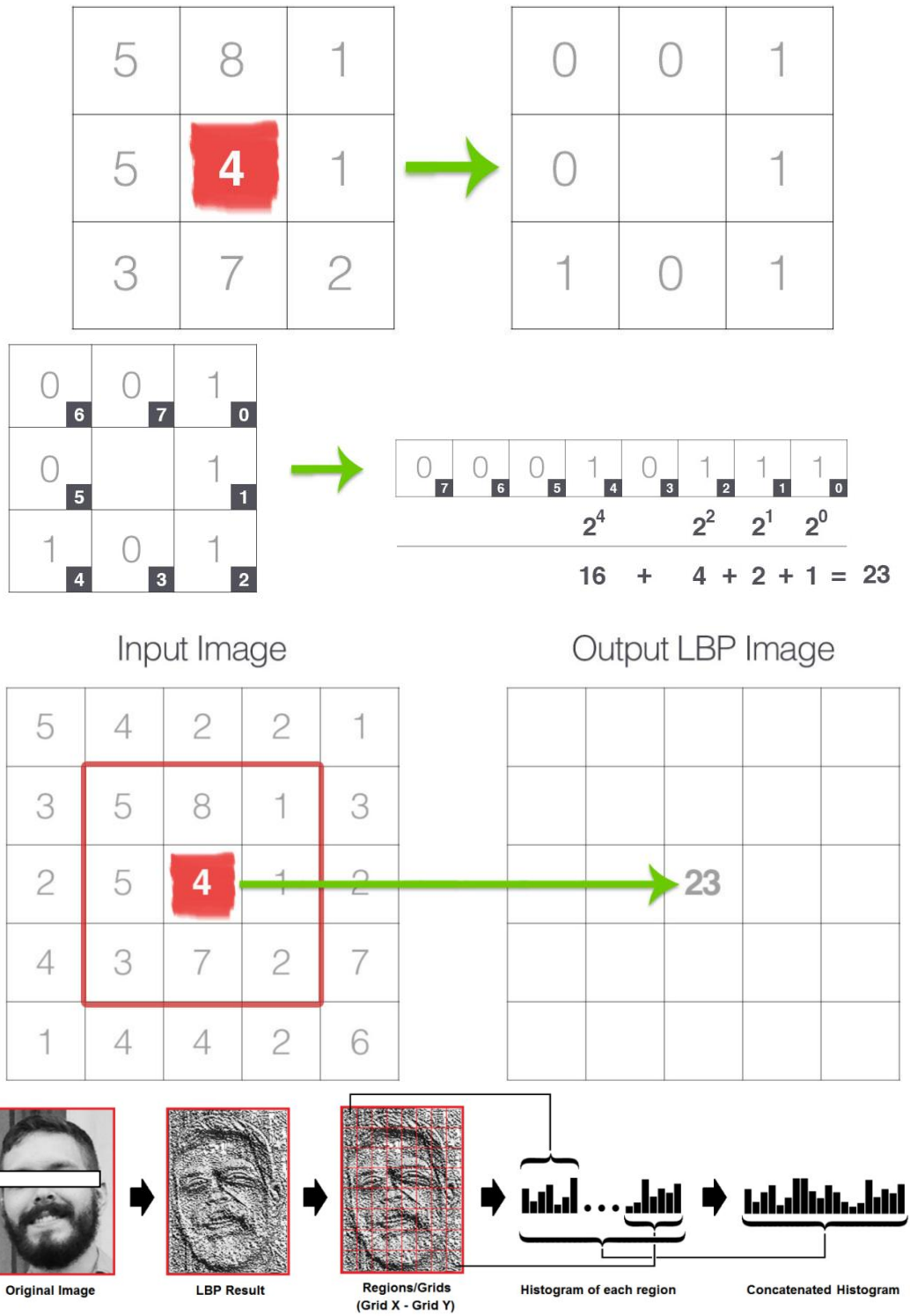
Fig. 4. Extracting local binary pattern

After computing the LBP value for each pixel and updating the corresponding value in the LBP image we get an output like the one shown in Fig. 4. Next, we calculate the histogram for the LBP Image. We use a 3x3 block which would yield 256 possible patterns hence we would have minimum value of 0 and maximum of 256. The histograms from each block are then combined together to form the feature vector for classification.

3.4 Classification

The resulting histogram from the local binary pattern is now used to train a classifier. In this study, two different classifiers are employed: the Support Vector Machine and the Random Forest Classifiers.

Support Vector Machine: Instead of the regular Support Vector Machine algorithm typically used for classification, a different variation called the Linear SVC is employed. The objective of a Linear SVC (Support Vector Classifier) is to fit to the data provided, returning a "best fit" hyperplane that divides, or categorizes, the data. From there, after getting the hyperplane, you can then feed some features to your classifier to see what the "predicted" class is. A regular SVM uses a radial basis function (RBF) as the SVM kernel. This is basically a Gaussian kernel also known as a bell-curve. Meaning that the no man's land between different classes is created with a Gaussian function. The Linear SVC on the other hand, uses a liner kernel as its basis function which allows for much faster convergence.

Random Forest: Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. Decision tree concept is more to the **rule-based system**. Given the training dataset with targets and features, the decision tree algorithm will come up with some **set of rules**. The same set rules can be used to perform the prediction on the test dataset [14].
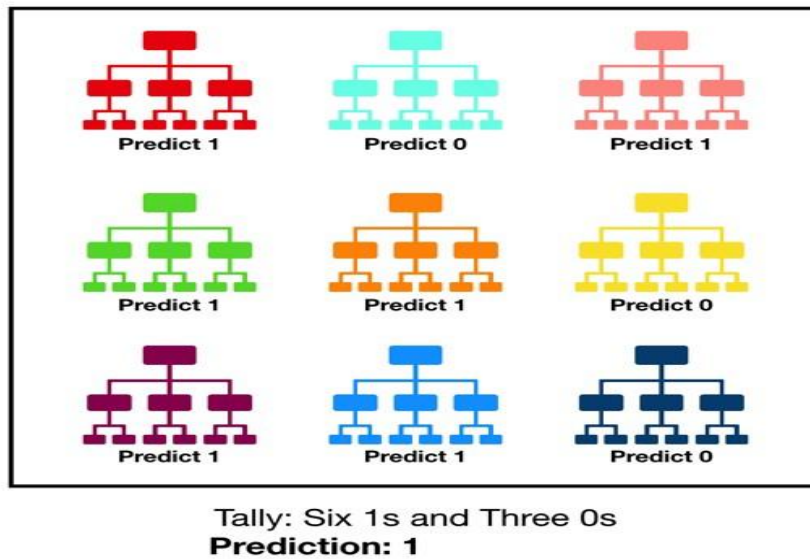

Fig. 5. Random forest [14]

4 Results and Discussion

This work was implemented using python programming language installed via the anaconda distribution. The experiments were carried out using a number of python libraries some of which include

- OpenCV Library – For reading images into python and splitting images into the three colour channels.
- Scikit-Image Library – For denoising images using a wavelet denoiser.
- Pandas Library – For calculating and viewing the correlation between images in our dataset. It was also used to hold the flattened versions of our images in a table like structure called a Data frame
- Scikit-Learn Library – For performing Hierarchical clustering on the flattened images
- Matplot Library – For visualizing our clustering algorithms as well as a dendrogram Jupyter notebook which is a part of the anaconda distribution was also used to run scripts in this research which allows for the python scripts to be run piece by piece. This enables easier debugging and incremental development.

4.1 Datasets used

For this study, two datasets were used: the first was the Cohn Kanade Plus dataset got from Kaggle, which comprised of 981 images of the different people of varying ages showing 7 different facial expression. The other was from Japanese female facial expression (jaffe) dataset [15]. The images were labelled and grouped based on the facial expression. The expressions captured in this dataset Anger, Disgust, Sadness, Fear, Surprise, Contempt and Happiness.

4.2 Features

A quick look at the histograms for the facial expression of fear and sadness in Fig. 6 and Fig. 7, after the feature extraction process described, we can easily see some kinds of similarity between the two of them.
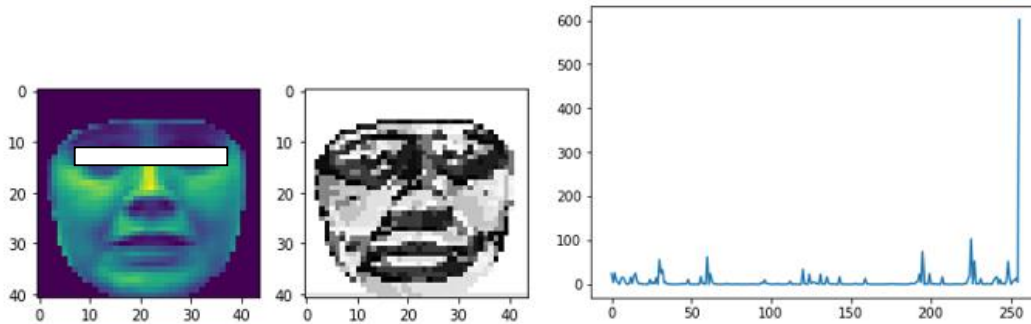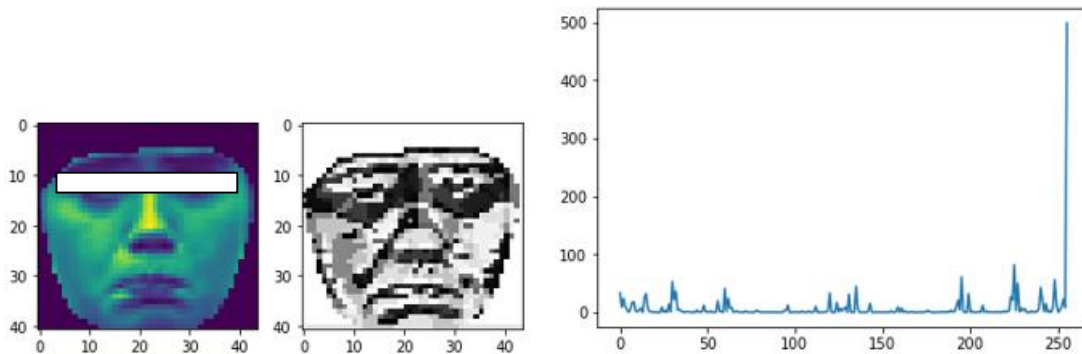

Fig. 6. Histogram for fear

Fig. 7. Histogram for sadness

4.3 Results

The extracted features were split into two parts, with 70% percent of them used to train two classifiers namely, SVM and Random Forest and the remaining 30% used to test and the performance of the model. The accuracy of the model is compared to that gotten from the model developed by [4]. To evaluate the performance of our model we use the common classification evaluation metrics which are Accuracy, Precision, Recall, F1-score on our dataset:

- Precision - (also called positive predictive value) is the fraction of relevant instances among the retrieved instances. It is calculated in equation (3):

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{3}$$

- Recall - (also known as sensitivity) is the fraction of relevant instances that have been retrieved over the total amount of relevant instances and it is calculated as shown in equation (4).

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{4}$$

- F1 Score - The F1 score is the harmonic mean of precision and recall taking both metrics into account and it is calculated in equation (5).

$$F_1 = 2 \times \frac{precicion \times recall}{precision + recall} \tag{5}$$

All of which are shown as a percentage.

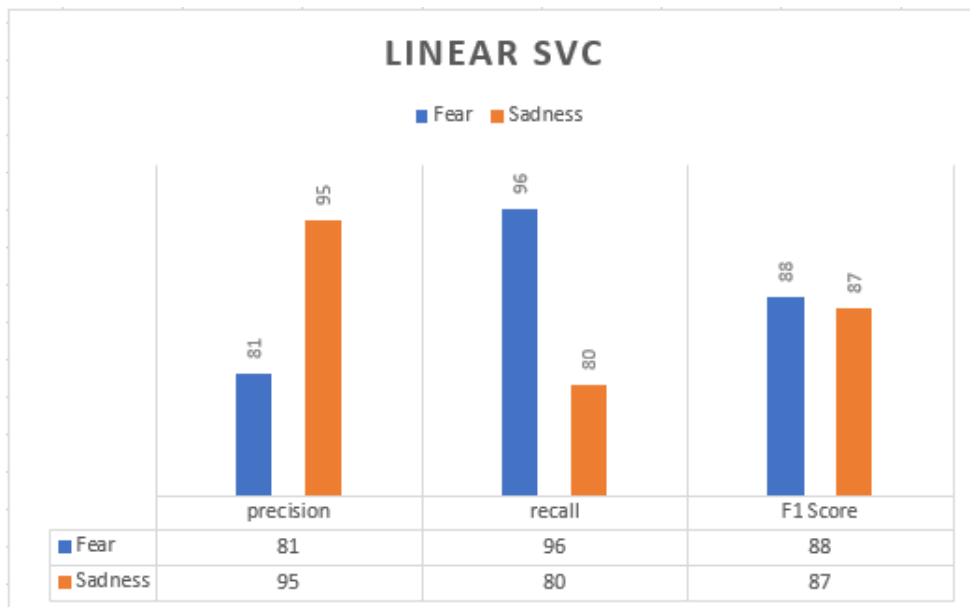## 4.4 Classification using linear SVC

Fig. 8. Classification using the linear SVC algorithm

4.5 Classification using random forest

Comparing the results from Linear SVC with the results from Random Forest in Fig. 9, we observe that the Linear SVM performs relatively better than the Random Forest Algorithm.
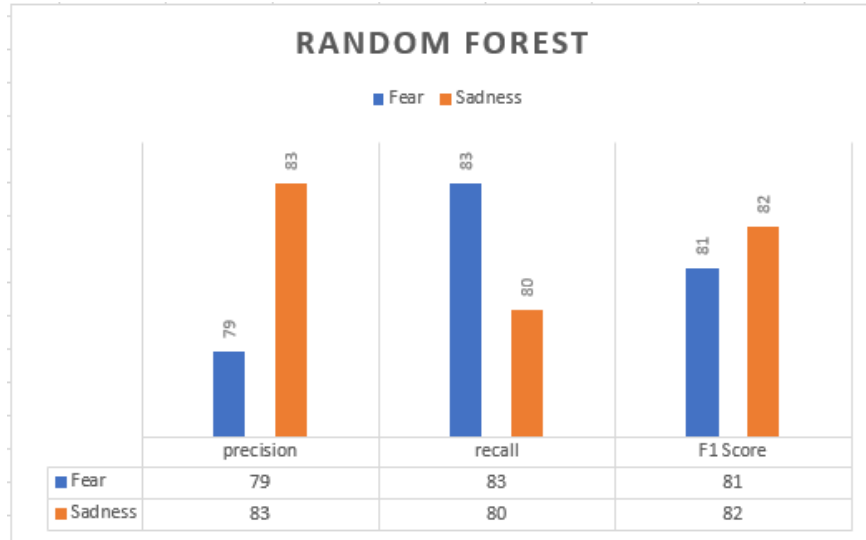
RANDOM FOREST

■ Fear ■ Sadness

	precision	recall	F1 Score
■ Fear	79	83	81
■ Sadness	83	80	82

Fig. 9. Random forest evaluation

4.6 Overall accuracy

To measure the overall accuracy of both classification algorithms, the ratio of true positives to the total number of data points in the testing data is calculated. The results in Fig. 10 show that the Linear SVC gives a better classification accuracy than the Random Forest (as shown in Fig. 9). Comparing the Accuracy of classification for fear and sadness got from our model with that of [4], we see a significant improvement, as shown in Fig. 11.
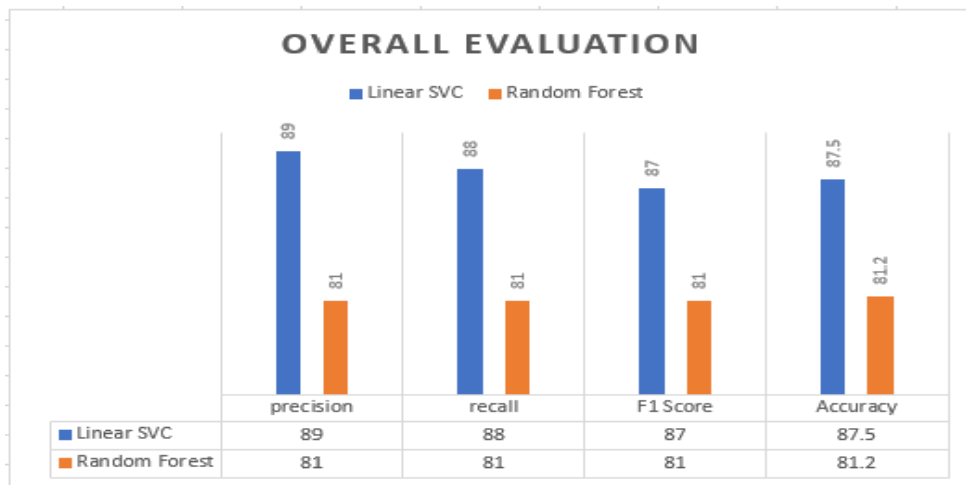
OVERALL EVALUATION

■ Linear SVC ■ Random Forest

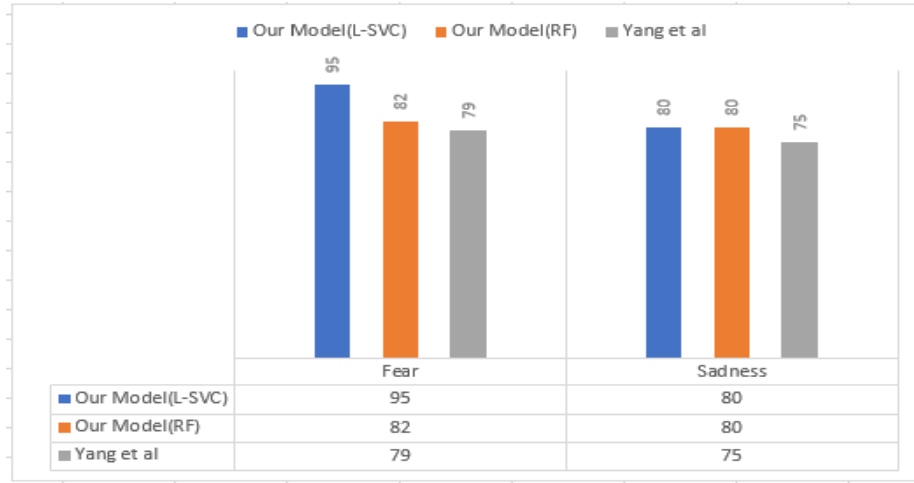|	precision	recall	F1 Score	Accuracy
■ Linear SVC	89	88	87	87.5
■ Random Forest	81	81	81	81.2

Fig. 10. Overall evaluation

Fig. 11. Comparison with the existing methodology

5 Conclusion

This study presents an approach that combines the Histogram of Oriented Gradients algorithm with the local binary pattern of an image in order to extract the features that can be used to classify the expressions. It uses a Linear Support Vector Classifier and a Random Forest Classifier to classify the different expressions. It was observed that the Linear SVC algorithm did a better job of classifying the facial expressions compared to the random forest classifier. Comparing this result to an existing methodology that uses edge detection for feature extraction, we see that our model performs significantly better. For future work, from a practical viewpoint, in many applications of human-computer interaction (HCI), it is important to be able to detect the emotional states of the person in a natural situation. Measuring the intensity if spontaneous facial expression is, of course more difficult than measuring acted facial expression due to the complexity, subtlety and variability of natural expression. Hence there is still room for improvement and extension to spontaneous facial expression in order to make a dynamic facial expression.

Consent and Ethical Approval

As per university standard guideline, participant consent and ethical approval have been collected and preserved by the author(s).

Competing Interests

Authors have declared that no competing interests exist.

References

[1]   Kaustubh Kulkarni, Ciprian Adrian Corneanu, Ikechukwu Ofodile, Sergio Escalera, Xavier Baro, Sylwia Hyniewska, Juri Allik, Gholamreza Anbarjafari. Automatic recognition of facial displays of unfelt emotions. Journal of IEEE Transactions on Affective Computing; 2018.

[2]   Fernando De la Torre, Jeffrey F. Cohn. Facial expression analysis. In Visual Analysis of Humans, London; 2011.

[3]   Shiqing Zhang, Xiaoming Zhao, Yuelong Chuang, Wenping Guo, Ying Chen. Learning discriminative dictionary for facial expression recognition. IETE Technical Review. 2017;35(3):1-7.

[4]   Yang Alsadon A, Prasad PWC, Singh AK, Elchouemi A. An emotion recognition model based on facial recognition in virtual learning environment. Science Direct Procedia Computer Science. 2018;125(2009):2-10.

[5]   Mu-Chun Su, Chun-Kai Yang, Shih-Chieh Lin, De Yuan Huang, Yi-Zeng Hsieh, Pa-Chun Wang. An SOM-based automatic facial expression recognition system. International Journal on Soft Computing, Artificial Intelligence and Application (IJSCAI). 2013;2(4).

[6]   Navdeep Kaur, Er. Varinderjit Kaur. Improved emotion detection by regression algorithm with SURF feature and SVM. International Journal of Engineering and Computer Science. 2014;3(10):8639-8642.

[7]   Kumari J, Rajesh R, Pooja KM. Facial expression recognition: A survey. In Procedia - Procedia Computer Science; 2015.

[8]   Kai Tai Song, Shuo-Cheng. Facial expression recognition based on basic expressions. IJSR. 2016;5(2).

[9]   Urvashi Bakshi R. Singhal. A new approach of face recognition using DCT, PCA and neural network in Matlab. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). 2014;3(3).

[10]   Jizheng Yi, Xia Mao, Lijiang Chen, Yuli Xue. Facial expression recognition considering individual differences in facial structure and texture. IET Computer Vision. 2014;8(5):429-440.

[11]   Caifeng Shan, Tommaso Gritti. Learning discriminative LBP-Histogram bins for facial expression recognition, 36, Eindhoven 5656 AE The Netherlands.

[12]   Prakasa. Texture feature extraction by using local binary pattern. Jurnal INKOM. 2016;9(2):45-48.

[13]   Ahonen Timo, Pietikäinen, Matti. Soft histograms for local binary patterns. In Proc. Finnish Signal Processing Symposium (FINSIG); 2007.

[14]   Yiu T. Understanding Random Forest, 12 June. [Online]; 2019.
Available:https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[15]   Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, Jiro Gyoba. Jaffe dataset: Coding facial expressions with Gabor wavelets. In Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara Japan; 1998.