Emotion Recognition from Facial Expression Using Machine Vision Approach

ᵃ Muhammad Zubair, ᵃ Aqib Ali, ᵃ Samreen Naeem, ᶜ * Farrukh Jamal, ᵈ Christophe Chesneau

a: Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur,06300, Pakistan.
b: Department of Statistics, Govt S.A Post Graduate College Dera Nawab Sahib, Bahawalpur, 63351, Pakistan.
c: Department of Mathematics, Université de Caen, LMNO, Campus II, Science 3, 14032 Caen, France.
* Corresponding author. E-mail: drfarrukh1982@gmail.com

Abstract--In this study emotion-based face expression recognition framework has been proposed using a machine vision (MV) approach. The face emotion dataset has been collected local survey in Bahawalpur city dataset divide into 3 classes happy, sad, and angry. A total of 600 images of size (256 x 256) were transformed into a gray level format and employed a median filter for noise removal. Three non-overlapping regions of interest (ROIs) of size (50 x 50) have been taken and analyze 1800 (600 x 3) ROIs on the overall dataset. Total 45 Statistical features named as texture, histogram, and binary features were extracted. Select optimize features using the correlation-based feature section technique. The optimized dataset employed of MV classifiers namely random forest (RF), logistic (Lg), and J48 are obtained very promising accuracy 96.33%, 95.67%, and 95.33% respectively.

Keywords-Machine Vision, Regions of Interest, Random Forest, Emotion, Face Expression.

I. INTRODUCTION

IN human communication, facial emotions play a very important role that helps in intention understanding [1]. Generally, people conclude other people's emotional states, such as happy, sad, and angry, using facial expressions. Oral components transmit one-third of human communication and non-verbal components transmit two-thirds [2]. Among the various non-verbal components, with an emotional meaning, facial expressions are one of the main channels of information in interpersonal communication [3]. Facial expression is an important part of nonverbal communication; because the facial gesture has an important part in explaining human actions. To obtain the required evidence, the facial expression data is divided into several sections which are applied in different methods to machine vision. The zone of the nose, lips, and eyes is transformed during the dialog [4].

The researcher proposed automated face analysis using a multi-feature approach. The FASSEG dataset is used for experimentation. Segmentation performs on the bases of nose, eyes, and skin, and random forest, support vector machine (SVM) classifiers used for classification and obtained 96.25% accuracy [5].

The researcher proposed conventional facial emotion recognition (FER) approaches are described along with deep learning (DL) based FER approaches and obtain very promising accuracy result of 72.65% [6]. The researcher proposed an automatic assessment of gender and FER. The idea based on image segmentation in six parts and label it. Random decision forest classifier and SVM classifier used for classification and obtain 96.25% accuracy [7].

The researcher proposed Male and female adolescents with conduct disorder associated with impairments FER. Concurrent eye tracking was used to relate categorization performance to participants' allocation of overt attention [8].

The researcher described the promising sectors in the field of FER, processing images streamed in real-time from a mobile device, which is health care. The proposed approach based on convolutional neural networks (CNN) and achieved better classification results [9].

The researcher described facial expression datasets such as eyes, nose, lips, and chin, etc. The overall accuracy is 90% using K- Nearest Neighbor (KNN) algorithms [10].

The researcher described edge detection algorithms for eyes and lips variation during human communication. canny edge detection (CED) provides better results on following facial expressions such as Sad 91.33%, Smile 99.44%, and Surprise 95.71% [11].

II. IMAGE DATASET

In this study digital image dataset has been collected from the local survey in Bahawalpur using a digital camera with 12 Megapixels. Three types of classes consider during the data collection phase first one is happy, the second is sad and the last one is angry. All the images based on real human emotions. For

this study 200 humans consider and divide into 3 classes, so a total 600 (200 x 3) images size (256 x 256) were collected for experimentation. The collected dataset sample is shown in Figure 1.
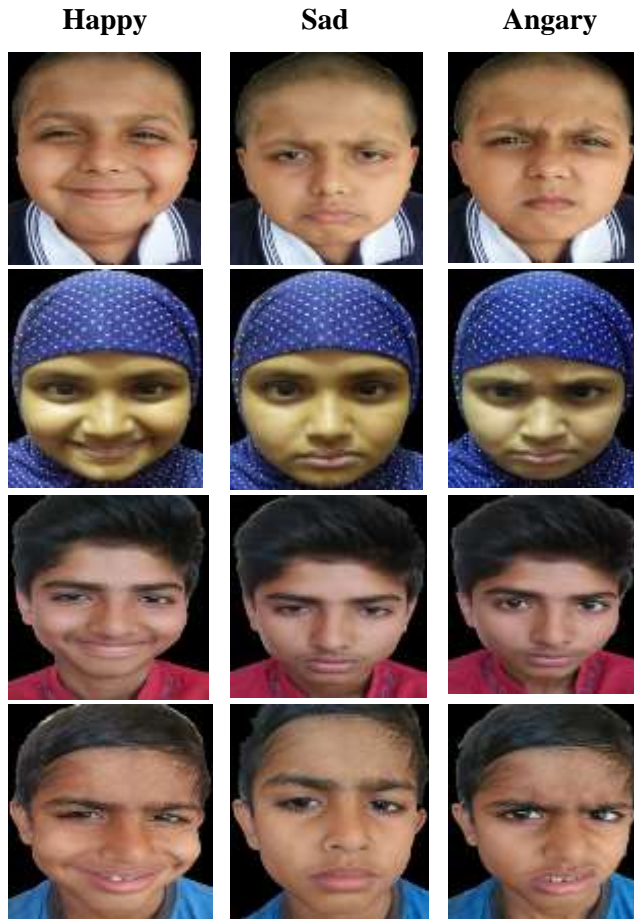
Happy	Sad	Angary

Figure 1: Human Face Image Dataset

III. PROPOSED METHODOLOGY

In this study, first, we collected an image dataset for experimentation. All the acquired image datasets standardized on the behave of ground truth. In the second step, color digital images transformed into gray level format [12]. After that using the CVIP tool select non-overlapping region of interest (ROIs) size (30 x 30) and (50 x 50). At the third step, texture-based features extracted named as texture (T), binary (B), and histogram (H) features. A fourth step correlation-based feature selection (CFS) applied for feature selection [13]. At last step machine learning classifiers [14] named as random forest (RF), logistic (Lg), and J48 applied on selected features dataset. All these experiment phases organized using Intel(R) core i7 processor 3.4Giga Hertz (GHz), 4 GB RAM, and 32-bit window 8.1 operating system.

Now we describe proposed algorithm then proposed diagram in Figure 1.

A. Proposed Algorithm

Start main ()
  {
   Input: digital image dataset

 For
{
Step 1 to 4
 Step 1: Transformed color to gray level and select non over lapping ROIs.
Step 2: Extract texture-based features (T + H + B).
Step 3: CFS based feature selection.
End For
 }
Step 4: Apply ML classifiers (RF, Lg, and J48)
Output: Facial Emotions Classification.
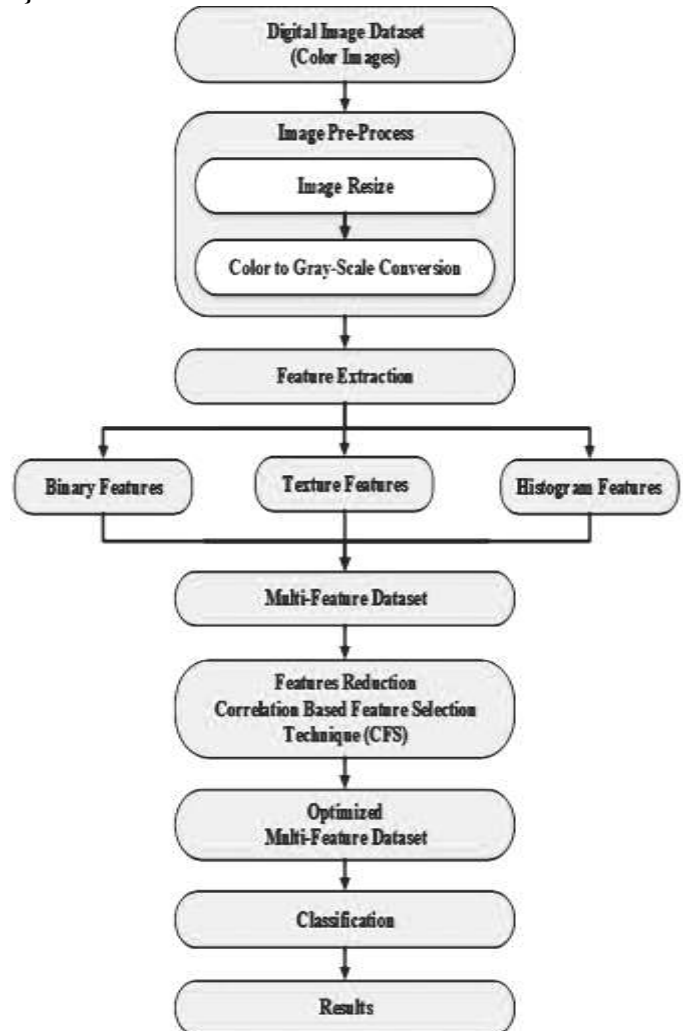**End main**
 }

Figure 2: Proposed Framework for Emotion Recognition

B. Image Preprocessing

Image preprocessing is a very important part of image analysis during image dataset collection some noisy image is the part of the dataset. So, for data standardization, we deal with the preprocessing stage. Firstly, crop all the images and set stander size 254 x 254, and all the image transformed into a gray level format. Some noise removal filters were applied named as a median filter [15], after that selected non-overlapping ROIs.

The ROIs selection is the main part of this study our focus is

on eyes, nose, and lips these are the more dominant portion which effected based on human mood. We select multiple ROIs with different size (30 x 30) and (50 x 50).
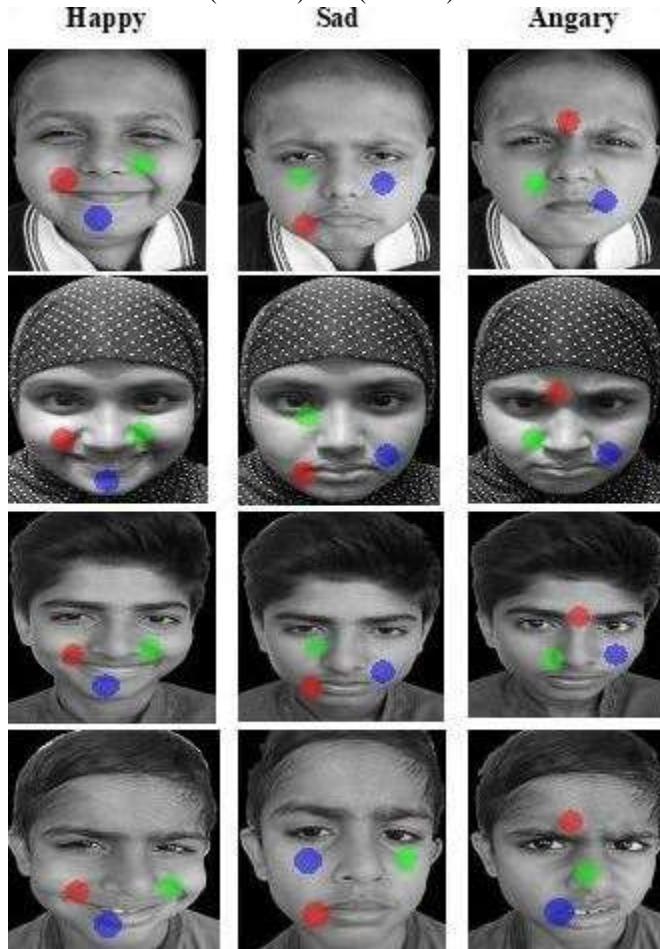

Figure 3: Selected 3 Non-Overlapping ROI's Image

C. Feature Extraction

For texture feature extraction we used three type of features named as binary (B), histogram (H), and texture (T) features. Statically dimension of feature extraction B, H and T are debated given below:

1) Binary Feature

Binary features recognize the substances and the shape of the entity in image processing based on Area, Second Axis, Euler number, Central Area and Projection [16].

2) Histogram Feature

Histogram shapes have been calculated the information about image on the basis of gray level and number of pixels [17]. The 1st order histogram W(h) as follows in equation (1).

$$W(h) = \frac{M(h)}{K} \tag{1}$$

Here N is total pixel in the image and M(h) is total pixel at gray level of K.

The 1st order histogram are used following method that are mean, standard deviation, skewness, energy and entropy for statistical calculation.

Mean define as follow in equation (2).

$$\bar{h} = \sum_{h=0}^{W-1} hW(h) = \sum_x \sum_y \frac{I(y,z)}{K} \tag{2}$$

Standard deviation (SD) describes in equation (3).

$$\sigma_h = \sqrt{\sum_{h=0}^{W-1}(h - \bar{h})^2 W(h)} \tag{3}$$

Skewness define as follows in equation (4).

$$Skew = \frac{1}{\sigma_h^3}\sum_{h=0}^{W-1}(h - \bar{h})^3 W(h) \tag{4}$$

Energy is defined in equation (5).

$$Energy = \sum_{h=0}^{W-1}[W(h)]^2 \tag{5}$$

Entropy is defined in equation (6).

$$Entropy = -\sum_{h=0}^{W-1} W(h)\log_2[W(h)] \tag{6}$$

3) Texture Feature

Energy evaluated smoothness by calculating the distribution between gray levels [18] are define in equation (7).

$$Energy = \sum_x \sum_y (B_{xy})^2 \tag{7}$$

Here B_{xy} are the values in the co-occurrence matrix by distribution values of pixel.

Correlations method defines as follows in equation (8).

$$Co = \frac{1}{\sigma_i \sigma_j}\sum_x \sum_y (x - \mu_a)(y - \mu_b)B_{xy} \tag{8}$$

Here μ_a and μ_b are the means of a and b respectively.

$$\mu_a = \sum_y y \sum_y B_{xy} \tag{8.1}$$

$$\mu_b = \sum_x n \sum_x B_{xy} \tag{8.2}$$

$$\sigma_i^2 = \sum_x (y - \mu_a)^2 \sum_x B_{xy} \tag{8.3}$$

$$\sigma_j^2 = \sum_y (x - \mu_b)^2 \sum_y B_{xy} \tag{8.4}$$

Entropy is measuring the content information of image. Entropy defines as follow in equation (9).

$$Entropy = -\sum_x \sum_y B_{xy} \log_2 B_{xy} \tag{9}$$

Inverse difference method is measuring the local homogeneity of image that defines as follow in equation 10.

$$\text{Inverse Diff} = \sum_x \sum_y \frac{B_{xy}}{|x-y|} \quad (10)$$

Inertia method is measuring the contrast that defines as follow in equation (11).

$$\text{Inertia} = \sum_x \sum_y (x-y)^2 \, B_{xy} \quad (11)$$

A total of 45 extracted statistical features (B + H + T) has been examined for each ROIs. It has been observed that the total feature vector space is 81000 (45 x 1800) on the basis of an employed dataset.

D. Classification

Correlation-based feature selection (CFS) approach used for select 9 optimize features for the classification. Three machines vision classifiers named random forest (RF), logistic (Lg), and J48 were employed using a cross-validation (10-fold) approach.

IV. EXPERIMENTS AND RESULTS

In the first step, machine vision classifiers employed of emotion dataset were ROIs size (30 x 30) and observe that impressive classification accuracy is random forest (RF), logistic Lg, and J48 were 95.33%, 94.67%, and 94.33% respectively as shown in Table 1.

Table 1: Machine vision classifiers employed of face dataset ROIs size (30 x 30)

Classifiers	Kappa Statistic	TP Rate	FP Rate	ROC	MAE	RMSE	Time (Sec)	OA
RF	0.9067	0.953	0.047	0.973	0.0832	0.2059	0.11	95.33%
Lg	0.8933	0.947	0.053	0.975	0.0530	0.2275	0.08	94.67%
J48	0.8867	0.943	0.057	0.953	0.0777	0.2294	0.03	94.33%

In the second step for improving classification accuracy result we increase the size of ROIs (50 x50) and observe very promising classification accuracy is RF, Lg, and J48 were 96.33%, 95.67%, and 95.33% respectively as shown in Table 2.

Table 2: Machine vision classifiers employed of face dataset ROIs size (50 x 50)

Classifiers	Kappa Statistic	TP Rate	FP Rate	ROC	MAE	RMSE	Time (Sec)	OA
RF	0.9267	0.963	0.037	0.970	0.0817	0.2015	0.18	96.33%
Lg	0.9133	0.957	0.043	0.979	0.0563	0.1906	0.04	95.67%
J48	0.9067	0.953	0.954	0.974	0.0841	0.2075	0.03	95.33%

A very promising accuracy obtain we MV classifiers deployed on ROIs size (50 x 50) as shown in Figure 4 and observe that random forest classifier performs outstanding among all of these a confusion matrix shows in Table 3.
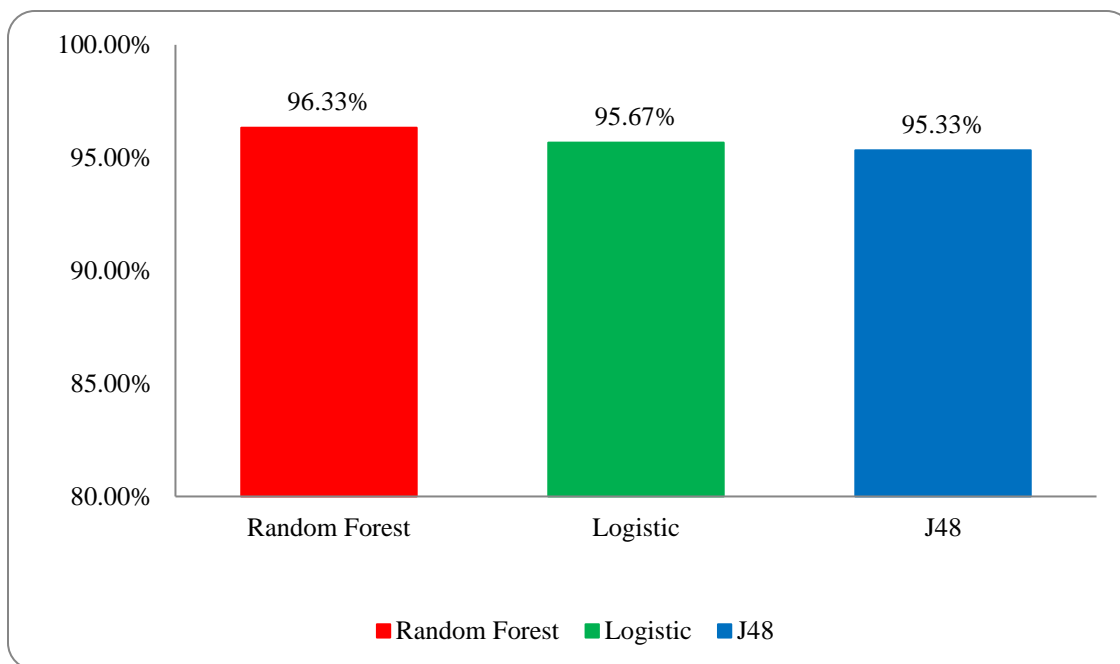
Figure 4: Comparison Graph of employed MV classifiers on ROIs (50 x 50)

Table 3: Confusion Matrix for Tree Random Forest Classifier

Classes	Happy	Sad	Angry	Accuracy
Happy	583	7	10	97.16%
Sad	10	581	9	96.83%
Angry	20	10	570	95%

A comparative analysis between MV classifiers employed on ROIs size (30 x 30) and (50 x 50) are shown in Figure 5.
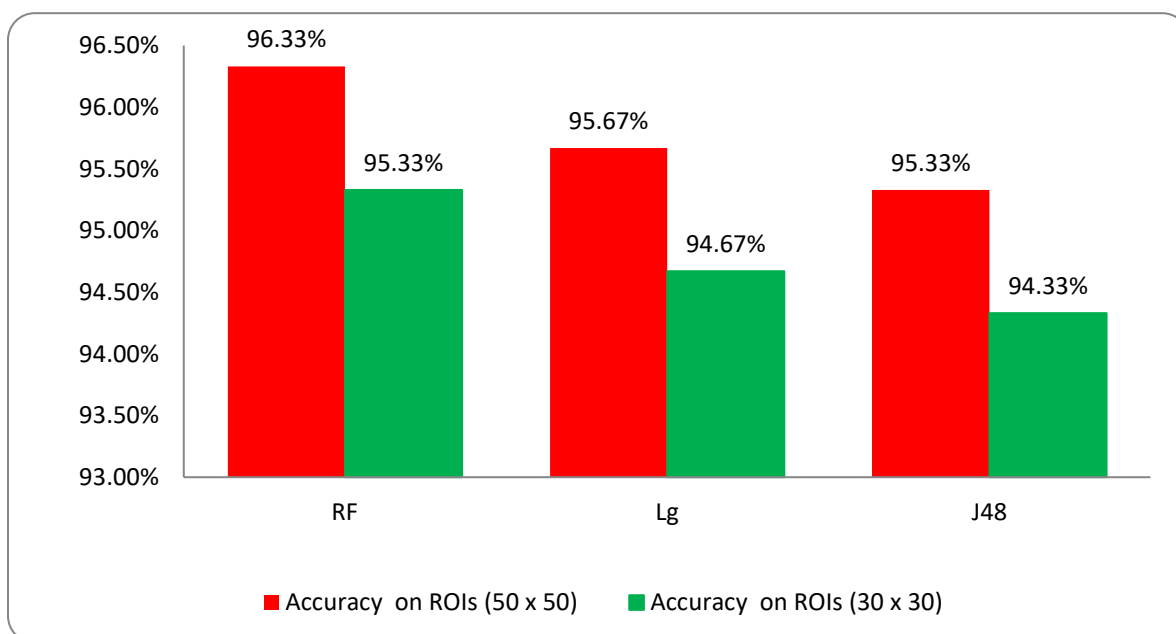
Figure 5: Comparative analysis graph of employed MV classifiers on ROIs (50 x 50) and (30 x 30)

V. CONCLUSION

In this study MV classifiers employed on human face digital images for recognizing the emotions. Dataset divides into three classes happy, sad, and angry. The various image processing technique is employed and three MV classifiers (RF, Lg, and J48) are tested using cross-validation (10-fold) approach and obtained very promising accuracy 96.33%, 95.67%, and 95.33% respectively. It has been observed RF classifier perform very well among all of these. The obtained accuracy for happy, sad, and angry were 97.16%, 96.83%, and 95% respectively.

VI. ACKNOWLEDGMENT

VII. REFERENCES

[1] Everingham, M., Van Gool, L., Williams, C.K., Winn, J. and Zisserman, A., 2010. The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), pp.303-338.

[2] Ranjan, R., Patel, V.M. and Chellappa, R., 2017. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1), pp.121-135.

[3] Schroff, F., Kalenichenko, D. and Philbin, J., 2015. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823).

[4] Happy, S.L. and Routray, A., 2014. Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing, 6(1), pp.1-12.

[5] Benini, S., Khan, K., Leonardi, R., Mauro, M. and Migliorati, P., 2019. Face analysis through semantic face segmentation. Signal Processing: Image Communication, 74, pp.21-31.

[6] Ko, B.C., 2018. A brief review of facial emotion recognition based on visual information. sensors, 18(2), p.401.

[7] Khan, K., Mauro, M., Migliorati, P. and Leonardi, R., 2017, September. Gender and expression analysis based on semantic face segmentation. In International Conference on Image Analysis and Processing (pp. 37-47). Springer, Cham.

[8] Martin-Key, N.A., Graf, E.W., Adams, W.J. and Fairchild, G., 2018. Facial emotion recognition and eye movement behaviour in conduct disorder. Journal of Child Psychology and Psychiatry, 59(3), pp.247-257.

[9] Gervasi, O., Franzoni, V., Riganelli, M. and Tasso, S., 2019, January. Automating facial emotion recognition. In Web Intelligence (Vol. 17, No. 1, pp. 17-27). IOS Press.

[10] Perveen, N., Ahmad, N., Khan, M.A.Q.B., Khalid, R. and Qadri, S., 2016. Facial Expression Recognition Through Machine Learning. International Journal of Scientific & Technology Research, 5(03).

[11] Chen, X. and Cheng, W., 2015. Facial expression recognition based on edge detection. International Journal of Computer Science and Engineering Survey, 6(2), p.1.

[12] Tarel, J.P. and Hautiere, N., 2009, September. Fast visibility restoration from a single color or gray level image. In 2009 IEEE 12th International Conference on Computer Vision (pp. 2201-2208). IEEE.

[13] Hall, M.A., 1999. Correlation-based feature selection for machine learning.

[14] Frank, E., Hall, M., Trigg, L., Holmes, G. and Witten, I.H., 2004. Data mining in bioinformatics using Weka. Bioinformatics, 20(15), pp.2479-2481.

[15] Eng, H.L. and Ma, K.K., 2001. Noise adaptive soft-switching median filter. IEEE Transactions on image processing, 10(2), pp.242-251.

[16] Ren, S., Cao, X., Wei, Y. and Sun, J., 2014. Face alignment at 3000 fps via regressing local binary features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1685-1692).

[17] Zhang, H., Gao, W., Chen, X. and Zhao, D., 2006. Object detection using spatial histogram features. Image and Vision Computing, 24(4), pp.327-341.

[18] Manjunath, B.S. and Ma, W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Transactions on pattern analysis and machine intelligence, 18(8), pp.837-842.