INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

Patch-SIFT: Enhanced feature descriptor to learn human facial emotions using an Ensemble approach

D Hema[1]*, **S Kannan**[2]

1 Assistant Professor, Department of Computer Science, Lady Doak College, Madurai
2 Professor, Department of Computer Applications, School of Information Technology, Kamaraj University, Madurai

Abstract

Background: Having experienced more than a year of pandemic, a variety of applications such as online classrooms, virtual office meetings, conferences, online games, Social media & Networks, Mobile applications, and many other infotainment areas have made humans live with gadgets and respond to them. However, all these applications have an impact on human behavioral transformation. It is very significant for employers to understand the emotions of their employees in the era of online office & work from home concept to increase productivity. Learning and identifying emotions from the human face has its application in all online portals when physical contact could not be achieved. **Ojbective:** Human Facial emotions can be learned using enormous feature descriptors that extract image features. While local feature descriptors retrieve pixel-level information, global feature descriptors extract the overall image information. Both of the feature descriptors quantify the image information, however, they don't provide complete and relevant information. Hence, this research work aims to improve the existing local feature descriptor to perform globally for emotion recognition. **Method:** Our proposed feature descriptor, Patch-SIFT collects features from multiple patches within an image. This strategy is evolved to globally apply the local feature descriptor as a hybridization paradigm. The extracted features are trained and tested on an ensemble model. **Findings:** The Proposed Feature descriptor (Patch-SIFT) performance with ensemble model is found to produce an improved accuracy of 98% compared with existing feature descriptors and Machine learning classifiers. **Novelty:** This research work tries to evolve a new Feature descriptor algorithm based on SIFT algorithm for an efficient emotion recognition system that works without the need for any additional GPU or huge dataset.

Keywords: Classification; Ensemble; Feature descriptor; PatchSIFT

1 Introduction

Features play a significant role in object recognition. If relevant and sufficient information is not retrieved, there are possibilities of images being classified to irrelevant classes. While applying local feature descriptor, the features or information might be missed in certain areas of images where it doesn't spot any corners/key points/ blobs. In case of global feature descriptor, features or information are not extracted at pixel level. However, either of the feature extraction methods doesn't extract complete information at the granular level. So, in this research work, we have focused on implementing local feature descriptor globally. The conventional SIFT features are utilized for training and classifying Human facial emotions and it is found to produce an accuracy of 84% which is not sufficient for many real time applications. In the era of big data and deep learning models, achieving accuracy close to 100% with small amount of data using machine learning techniques will have its appreciation. Hence, we tried to improve the existing SIFT descriptor that can be used for recognizing human emotions without the need for huge amount of data. Based on the research work [1] on the analysis of various Feature descriptors, it is found that the Scale Invariant Feature Transform (SIFT) descriptor performance is poor. SIFT is a key point based descriptor that finds the key points in an Image. It is identified that SIFT doesn't match the key points on original image to key points in scaled/rotated image as experimented in [1]. SIFT local descriptor extracts key points only on limited areas of the image. It doesn't recognize the key points where relevant information is present. The original SIFT is applied on the entire image whereas the improved SIFT (Patch-SIFT) extracts key point features in small patches. Real time objects can be detected and learnt. Once the object is being detected, its behavior or state or nature can be learnt in the next level. The learning about objects is based on the object category. For instance, if the object detected is a water bottle, object learning involves identifying the liquid level inside the water bottle like partially filled, empty or full. In the proposed work, object learning involves emotions present in human face using Enhanced SIFT descriptor using an Ensemble classifier. The performance of the enhanced feature descriptor on recognizing human facial emotions is compared with the existing research work and is found efficient with an accuracy of 98%.

SIFT features are implemented along with ORB features to improve the object recognition task [2]. Bag of visual words technique is used for recognizing objects which build a visual vocabulary of extracted local features using SIFT/ ORB/ SURF feature descriptor algorithm etc. However, the bag of visual vocabulary strategy as in [2] is complex and time-consuming. FER (Facial Expression Recognition) algorithm is integrated with online course platforms [3] due to the rapid development of online education. A variant of Local Binary Pattern, Angled local directional pattern as a texture feature is used for recognizing emotions [4] on CK+ dataset that conveyed an accuracy of 89.79%. Based on the contractions and expansions in facial muscles, high frequency edges is combined with LBP features to develop a new feature extraction algorithm called digital signature [5] for facial expression recognition. The framework in [5] incorporated existing machine learning classifier, Support Vector Machine (SVM) and reported an accuracy of 96.25% on CKFED dataset. Emotion recognition can be affected by several factors and the study by [6] proves that recognition accuracy can vary when influenced by age and gender. This research work also reports an accuracy of 83.73% for recognizing emotions which is not still satisfying. 3 Feature extraction methods using 2D orthogonal sub space (2DPCA, 2DFDA, 2DSVD) 2DPCA along with CNN is designed for emotion recognition in [7]. Results of coupling a CNN model with well-tuned parameters in [7] achieves 97.6% and 94.5% accuracy rates respectively in original samples and 2DPCA based features in Yale dataset. Local Binary Pattern (LBP) and Oriented Fast and Rotated Brief (ORB) are combined as one of the feature descriptor to recognize facial expressions in [8]. The study in [8] reports a Subject Independent accuracy of 88.5% on JAFFE dataset & 93.2 % on CK+ dataset. Gabor features, LBP features, CNN deep features, Joint geometric features and mixed features are implemented in [9] for emotion recognition with an average recognition rates of 94.75% and 96.86%, JAFFE and CK+ respectively. An Appearance Network and Geometric Network is combined to form a Deep Joint Spatiotemporal Network (DJSTN) [10] in which the authors have applied a 3D convolution on Face images to extract spatial and temporal features. Another deep learning strategy evolved using Novel dynamic weighting technique is SIFE-Net [11]. However, the framework in [11] requires secondary expression information alongside single label information which may not be reliable in most cases. Appearance Network using LBP processed images and Geometric Network with Action Unit Landmarks facial images are fused to predict facial emotions [12]. The fused Network [12] is also capable of generating Neutral images in human faces using auto encoder but still it delivers an accuracy of 96.46% on CK+ and 91.27% in JAFFE dataset. Facial Expressions are recognized in videos by a framework proposed in [13] that uses ResNet CNN. The framework in [13] is faster and delivers 96.45% accuracy on CK+ dataset but requires Nvidia GPU. In case of Image retrieval, Traditional SIFT keypoints along with edge information is extracted; out of which the Top-32 Keypoints are selected using RootSIFT in [14]. CBIR (Content Based Image Retrieval) using ORB, SIFT feature combination [15] and using BayesNet and K-NN as in [16] is also gaining popularity in addition to Facial Expression Recognition. A comprehended summary of various Face detection techniques, Features and dataset used is presented in [17] which is helpful to understand the challenges and applications of the same in HCI. To extend face detection under occlusion and Non-uniform illumination conditions [18], YcbCr,HSV and Lxaxb Color models are combined which is applied on customized dataset. Several feature descriptors such as SIFT,SURF,ORB and Shi Tomasi corner detector algorithm are

combined to recognize objects using eXtreme Gradient Boosting Classifier [19] that reports an accuracy of 88.36% for Caltech-101 dataset. A performance comparison of 3 feature extraction methods is investigated in [18], out of which hybridizing SIFT, SURF, ORB Features with Random Forest classifier yields good object recognition results. Famous deep learning models for Object detection such as YOLO (You Only Look Once) and Faster R-CNN are deployed for detecting masks [20] in Human face during the COVID-19 period. From the review of several related articles, it is evident that all the emotion recognition research works were carried on existing feature descriptors and machine learning classifiers. Moreover, the existing frameworks are not robust and could not reach accuracy above 97% on JAFFE and CK+ dataset. To achieve accuracy close to 100%, a huge amount of data and deep learning model is required. It also requires higher end GPU such as Nvidia which is not affordable in many situations. Hence, a robust framework for detecting emotions without the need for huge amount of data and complex hardware should be developed to achieve high accuracy.

2 Methodology

To create an efficient facial emotion recognition system with limited resources and high accuracy, the proposed work is designed. In this research, we tried to provide a solution for facial emotion recognition considering the limitations of the hardware and dataset. From the survey, it is understood that hybridizing 2 or more features yield good results. Hence, the proposed method combines the strategy of Global and local feature extraction. The classification process is accomplished using a customized ensemble classifier.

A. Dataset

In our proposed work, the improved version of SIFT called Patch-SIFT is applied for emotion recognition in human faces. The experiment is done using JAFFE dataset and CK+ dataset (approximately 300 images/ class).

JAFFE- The Japanese Female Facial Expression (JAFFE) Dataset is used in this research work to recognize and learn emotions, The JAFFE dataset was developed using facial expressions provided by 10 Japanese female subjects. There are nearly 7 Posed Facial Expressions that includes 6 basic facial expressions (anger, disgusted, happy, sad, surprise) plus 1 neutral expression. The dataset has 213 images in total. The resolution of the images in this dataset was 256x256 pixels. However, all the images are 8-bit gray scale. The images have an extension of Tiff format. The advantage of this dataset is that the images are not compressed. The image dataset was planned, constructed and assembled by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba, at Kyushu University, Japan. The JAFFE images are permitted to be used for non-commercial scientific research under certain terms of use, which should be accepted to access the data.

Cohn kanade(CK) Dataset- Along with JAFFE dataset, this research work has made use of the Cohn-Kanade (CK) (T. Kanade, 2000) database. CK dataset was developed and released for promoting research in automatically detecting independent facial expressions. Subsequently, the CK database has become one of the most popular test-beds for algorithm development and evaluation. The CK database has been used for both AU (Action Unit) codes and emotion detection. An Extended Cohn-Kanade (CK+) database was released with revised emotion labels. CK and CK+ dataset was released in Kaggle community for researchers with an agreement to use and download.

B. Feature Extraction using Patch-SIFT algorithm

In the proposed work, if object detected is human, the learning is done to extract facial emotions like happy, sad, angry, disgusted and surprised. To perform emotion learning in human faces, the enhanced SIFT descriptor called Patch-SIFT is used to extract facial key point features. Patch-SIFT is an improvement of the existing SIFT local feature descriptor. SIFT contemplates corners as key points and retrieves 128 dimension feature vector. The proposed algorithm is enhanced by dividing a whole image into multiple patches of size 10X10. In each of the patch/window, the SIFT algorithm is implemented to extract the features and store the feature vectors. It steps to the next patch/window and applies the SIFT algorithm to retrieve features in that patch. In a sliding window fashion, the patch SIFT covers the global information existing in the image. Hence, maximum number of relevant and significant features are extracted compared to a conventional SIFT. By using Patch-SIFT, the features are mined and stored as a feature vector of size 128 dimension for each key point. It is then aggregated into a single cluster to get a feature vector of size 1 X 128.After retrieving the key points, features are stored as feature vectors each having a size 128 dimensions. However, different training images have different feature vector sizes. For instance, an image with 100 key points has a feature vector of size 100 X 128 and an image with 200 key points the feature vector size would be 200 X 128. Hence, the Feature vectors of all training images are converted to same dimensions of size 1 X 128. Successively, Feature vectors of individual images are clustered into a single feature vector of 1 X 128 dimensions. The next step is to encode the labels of the images that should range

between 0 and n-1 class labels for n number of classes. A Fit() function fits the label encoder whereas fit_transform() is used to fit the label encoder and returns the labels of the encoded class. In our study, 5 emotion classes are trained for classification and therefore the class labels are 0 (angry), 1 (disgust), 2(happy), 3(sad) and 4(surprise). Features are then regularized by removing the mean and scaling it to unit variance. To accomplish this, a Standardscaler function is used. There are 1000 images (200 per class) trained in this investigation for emotion recognition.

C. Existing Machine Learning Classifiers for learning Emotions

Dataset Standardization and regularization plays an important role in most of the Machine Learning classifiers and models. When the features are on 0 or 1, no useful classification can be done and hence feature scaling is done to recognize the exact variance. The scaled features and labels are then stored for the purpose of training and validation. The stored features and labels are fed for training and testing phase with a split of 80:20. The Features and labels are trained and tested on various Machine learning classification models like K-Nearest Neighbors, Support Vector Machine, Decision Trees, Random Forest, Naïve Bayes, Neural Network, Linear Discriminant Analysis, Linear Regression and CART. These Models are fit for the training phase with the encoded labels and scaled features. It is observed that most of the machine learning models gives good performance in predicting emotions in human face with accuracy above 90%. Hence, the performance of those machine learning models is combined to form an ensemble model.

D. Ensemble Classifier for emotion recognition

An Ensemble Machine learning model is designed to classify emotions (angry, disgust, happy, sad and surprise) in this research work. This ensemble model is a stacked hybrid model that includes Linear Discriminant Analysis (LDA) algorithms and Random Forest (RF) Classifiers. The ensemble model deployed here is a hard voting classifier that picks the majority of predictions made by the contributing stacked classifiers. A voting ensemble integrates the predictions from several machine learning models. Voting ensembles are used for both regression and classification tasks in many research investigations. For regression tasks, predictions are done by the average predictions of the contributing machine learning models. For classification tasks, predictions are done by the majority predictions or vote of contributing machine learning models. Further, for classification ensemble there are two categories such as hard voting and soft voting ensemble classifier models. Hard voting predicts and returns the class label that has maximum sum of votes from the machine learning models. However, soft voting predicts and returns the class label that has the maximum summed probability from the models. Hard voting includes voting for crisp class labels while soft voting includes the predicted probabilities for the class labels.
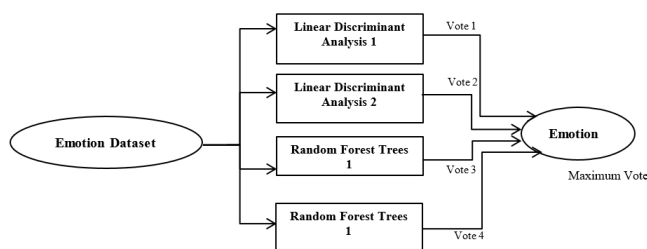
Fig 1. Hard voting Ensemble classifier to recognize emotions

Normally, a voting ensemble is a model of models known as meta model. The extended model of any voting classifier is a stacked generalization model which learns when and how much to trust each machine learning model while making predictions. In this study, a hard voting ensemble model (as shown in Figure 1) using a combination of Random Forest and Linear Discriminant Analysis algorithms is utilized.

K-Fold Cross Validation

In order to embrace all the observation from the original dataset for training in this study, K-Fold cross validation is implemented. This validation is done by randomly splitting the set of observation dataset into K groups or folds of similar size. The first group/fold is considered as validation set and the machine learning model fits on the remaining K-1 groups/folds. The process is repeated until every K-fold functions as the test set. Performing the K-Fold cross validation will boost the accuracy of the Machine learning models. In few instances, Leave one out cross validation is also utilized to validate the observation as

individual points and not as folds/groups. In leave one out cross validation technique, among N observation data points only one data point is set aside for testing and rest of the data points (N-1) is used for training. This process is repeated for N number of iterations. Assume, during 1st iteration, 1st data point will be set for testing and 2nd till Nth data points are used for training. While proceeding for 2nd iteration, 2nd data point is set for testing and 1st and 3rd till N^{th} data points are used for training. In this proposed research work, K-Fold cross validation is used with 10 splits of dataset and a random seed of 5.

3 Results and Discussion

This Research work had its inception from [21] where human is detected in Phase-1 and facial emotions are recognized using SIFT features. The Experiment for emotion recognition using enhanced SIFT feature descriptor is performed on JAFFE and CK+ human facial emotion dataset. There are 5 emotion classes identified such as angry, disgust, happy, sad and surprise. The features are extracted using enhanced SIFT (Patch-SIFT) algorithm and trained using existing machine learning models and proposed ensemble model. The models used were Linear Regression (LR), K-Nearest Neighbor (KNN), Random Forest (RF), Naïve Bayes (NB), Classification and Regression Trees (CART), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Neural Network (NN). The proposed machine learning model is a customized voting ensemble model in this research work.

A. Accuracy

The accuracy of the machine learning models used for emotion recognition as a part of object (human) learning is summarized in Figure 2. It is observed that the customized Ensemble model yields a high accuracy of 98% while using the Patch-SIFT feature descriptor for emotion recognition. Subsequently, Random Forest classifier produces an accuracy of 97%. The Neural Network implemented in this research work is rated next to Random Forest with an accuracy of 94% and then comes the Linear Discriminant Analysis (LDA) with an accuracy of 91%. The CART, NB and SVM perform poor compared to other machine learning models with an accuracy accounted below 80%.The accuracy of each machine learning model is calculated using confusion or error matrix.
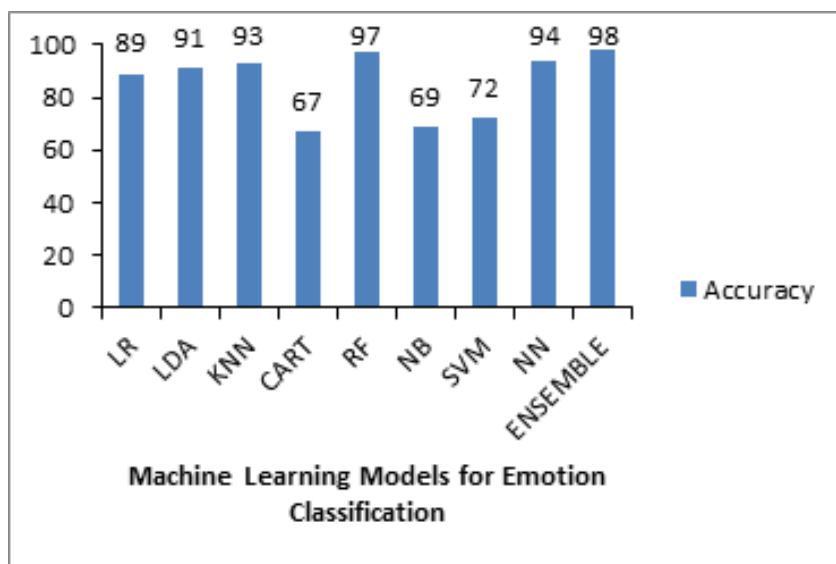
Fig 2. Accuracy of the machine learning models on Patch-SIFT feature descriptor

B. Confusion Matrix

Confusion Matrix reveals the classified emotions with actual emotions for each class. Figure 3 (A-I) shows the confusion matrix for each of the machine learning models implemented in this research work on Patch-SIFT features for 100 images. The customized Ensemble Model (Figure 3-I) reveals that 98 out of 100 test images are correctly classified.
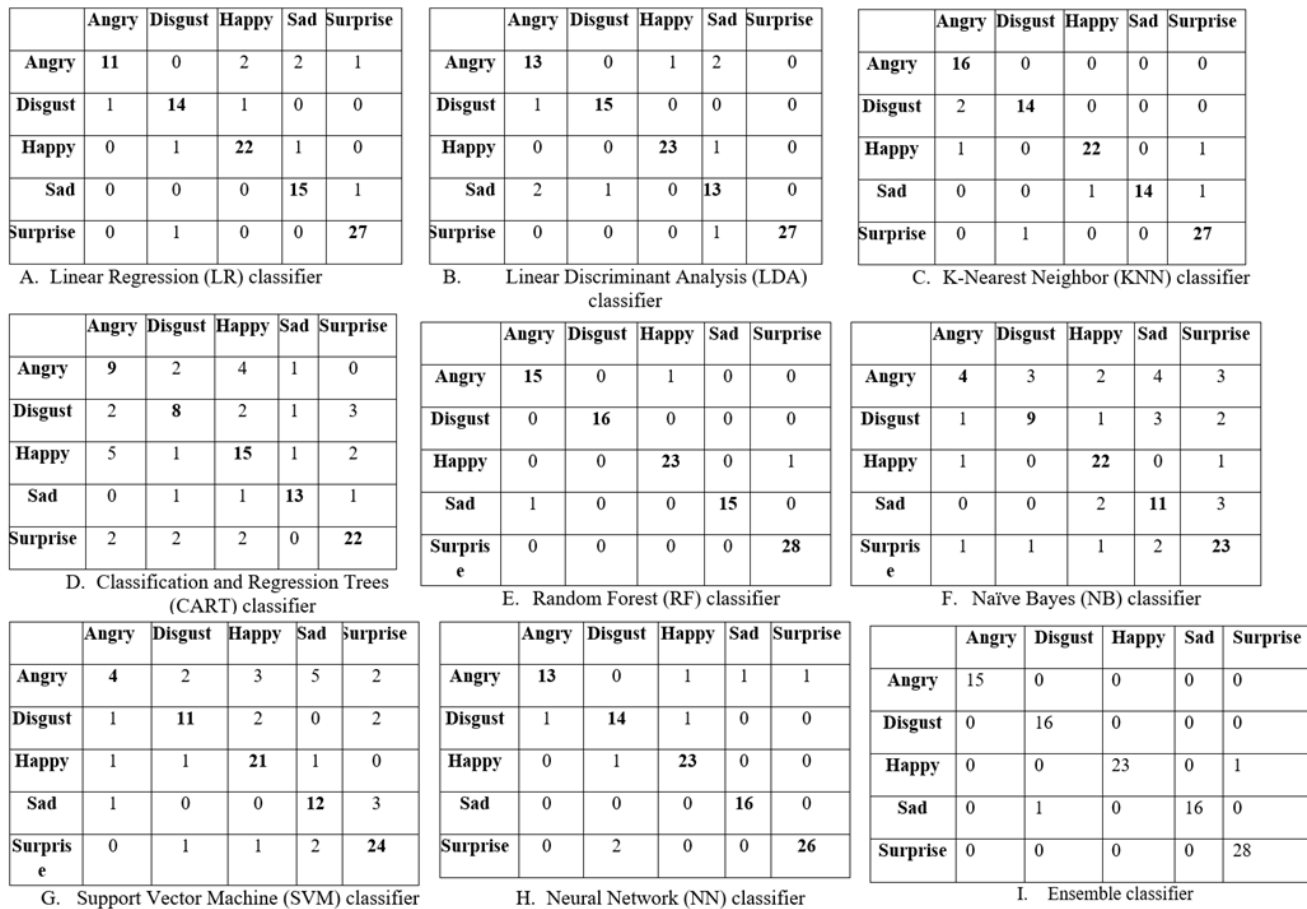
	Angry	Disgust	Happy	Sad	Surprise
Angry	11	0	2	2	1
Disgust	1	14	1	0	0
Happy	0	1	22	1	0
Sad	0	0	0	15	1
Surprise	0	1	0	0	27

A. Linear Regression (LR) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	13	0	1	2	0
Disgust	1	15	0	0	0
Happy	0	0	23	1	0
Sad	2	1	0	13	0
Surprise	0	0	0	1	27

B. Linear Discriminant Analysis (LDA) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	16	0	0	0	0
Disgust	2	14	0	0	0
Happy	1	0	22	0	1
Sad	0	0	1	14	1
Surprise	0	1	0	0	27

C. K-Nearest Neighbor (KNN) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	9	2	4	1	0
Disgust	2	8	2	1	3
Happy	5	1	15	1	2
Sad	0	1	1	13	1
Surprise	2	2	2	0	22

D. Classification and Regression Trees (CART) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	15	0	1	0	0
Disgust	0	16	0	0	0
Happy	0	0	23	0	1
Sad	1	0	0	15	0
Surprise	0	0	0	0	28

E. Random Forest (RF) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	4	3	2	4	3
Disgust	1	9	1	3	2
Happy	1	0	22	0	1
Sad	0	0	2	11	3
Surprise	1	1	1	2	23

F. Naïve Bayes (NB) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	4	2	3	5	2
Disgust	1	11	2	0	2
Happy	1	1	21	1	0
Sad	1	0	0	12	3
Surprise	0	1	1	2	24

G. Support Vector Machine (SVM) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	13	0	1	1	1
Disgust	1	14	1	0	0
Happy	0	1	23	0	0
Sad	0	0	0	16	0
Surprise	0	2	0	0	26

H. Neural Network (NN) classifier

	Angry	Disgust	Happy	Sad	Surprise
Angry	15	0	0	0	0
Disgust	0	16	0	0	0
Happy	0	0	23	0	1
Sad	0	1	0	16	0
Surprise	0	0	0	0	28

I. Ensemble classifier

Fig 3. Confusion Matrices of various Machine learning classifiers using enhanced SIFT features

C. Evaluation Metrics: Accuracy, Precision, Recall, F1 Score & Support

From the Confusion Matrix of each machine learning model, evaluation metrics like Recall, Precision, F1-Score and Support are calculated. Accuracy is the proportion of correct cases predicted to the total number of cases. Precision is the proportion of correctly predicted positive cases to the total number of predicted cases. High precision and accuracy are the indicators of a good machine learning model. It also reveals that the model has predicted more true positive cases than false positives in the actual data. Recall is the proposition of the data points that the model says was relevant were actually relevant. The optimal integration of precision and recall is the F1 Score which is the harmonic mean of precision and recall. The support is the total samples of the true response that fit to that image class. The Weighted average precision, F1 Score and recall for each of the machine learning model is summarized in Table 1. Even when accuracy of a machine learning model is high, sometimes precision, recall and F1 score might be low which means the model performs poor. It is observed in Table 1 that the ensemble model for emotion recognition customized in this research work performs efficiently with an accuracy of 98 % along with recall, precision and F1 score having 0.98. The ensemble model is built using Random Forest (RF) and Linear Discriminant Analysis (LDA) classifier with a hard voting technique stacked one above the other. It is also observed that this ensemble model outstand the performance of the existing machine learning models.

After validating the models using K-Fold technique, the models are tested with individual images. Different emotions are being correctly identified and recognized using those models on different human faces irrespective of the age, gender and race. The emotions learnt and recognized on human face include angry, disgust, happy, sad and surprise which is illustrated in Figure 4 .

Table 1. Weighted average accuracy, precision, recall and F1 score of the machine learning classifiers implemented for human facial emotions.

Model	Accuracy	Precision	Recall	F1 Score
LR	89	0.89	0.89	0.89
LDA	91	0.91	0.91	0.91
KNN	93	0.93	0.93	0.93
CART	67	0.67	0.67	0.67
RF	97	0.97	0.97	0.97
NB	69	0.68	0.69	0.67
SVM	72	0.71	0.72	0.70
NN	94	0.92	0.92	0.92
Ensemble	98	0.98	0.98	0.98

Fig 4. Emotions learnt and recognized on human face using Patch SIFT features on Ensemble classifier a. Angry, b. Disgust, c. Happy, d. Sad, e. Surprise

D. Performance Evaluation

The Performance of the proposed Patch SIFT features along with the Ensemble classifier is compared with the existing research works for emotion recognition as summarized in Table 2 . The proposed work is found to produce a highest accuracy of 98% for emotion recognition. Table 2 shows the list the features, dataset used and accuracies of the existing research works and the proposed work for emotion recognition. The existing research work on emotion recognition has reached a maximum of 97.6% on YALE dataset while the proposed Patch-SIFT consistently report an accuracy of 98% for a combination of JAFFE & CK dataset. It is also evident from Figure 3-I that the only 2 images out of 100 are misclassified due to some similar features between happy-surprise and sad-disgust pairs.98 testing images in the confusion matrix(Figure 3-I) provides user and producer accuracy at the maximum. The proposed research work is executed on a normal Intel(r) core(tm) i5-6200u CPU (2.30 ghz) and 4 GB RAM without the need for any Nvidia GPU. Therefore, this framework is very much adaptable and suitable for any real time applications with limited resources.

Table 2. Performance evaluation of the proposed feature descriptor (on Ensemble classifier) with existing research for human emotion recognition

Research work on human emotion recognition	Feature /Feature Descriptor	Dataset used	Reported Accuracy
A.M.M, Shabat and Tapamo J., 2018 [4]	LBP,LDP	CK+	89.79%
Talele, K., Tuckley, K., 2020 [5]	Edge +LBP	CKFED	96.25%
Ergin, S., Isik, S., Gulmezoglu, M.B., 2021[7]	2DPCA, 2DFDA, 2DSVD	YALE	97.6%
Ben Niu, Zhenxing Gao, Bingbing Guo, 2021[8]	LBP+ORB	JAFFE	88.5%
Ben Niu, Zhenxing Gao, Bingbing Guo, 2021[8]	LBP+ORB	CK+	93.25
Hao Meng, Fei Yuan, Yue Wu, Tianhao Yan,2021 [9]	Gabor features, LBP features, CNN deep features, Joint geometric features and mixed features	JAFFE	94.75%
Hao Meng, Fei Yuan, Yue Wu, Tianhao Yan,2021 [9]	Gabor features, LBP features, CNN deep features, Joint geometric features and mixed features	CK+	96. 86%
Proposed	Enhanced SIFT (Patch SIFT-Proposed)	JAFFE+CK	98%

4 Conclusion

The existing traditional feature descriptors don't report efficient and consistent accuracy for emotion recognition. To quantify and retrieve maximum features in an image, an enhanced Feature descriptor (Patch-SIFT) is proposed in this work for learning human facial emotions. An ensemble classifier consisting of Linear Discriminant Analysis and Random Forest is also built in this study. The Ensemble model is compared with existing machine learning classifiers. The proposed work is also evaluated using accuracy, precision, recall and f1 score computed from the confusion matrix. The performance of the proposed emotion recognition framework is found to be reliable and robust while comparing with the existing research works reporting an increased accuracy of 98%. Without any additional GPU and limited dataset, the proposed Patch-SIFT feature reports accuracy close to deep learning model and hence it can be applied to detect other objects as well. In future, similar feature descriptors like SURF, ORB, KAZE etc., can also be enhanced or hybridized for emotion recognition. Further, these feature descriptors can also be used with Convolutional Neural Networks and the performance can be improved. The proposed research work is also designed for static individual images. It can be extended to work on images present in live videos and camera footage.

References

1) Hema D, and SK. Performance study of Prevalent Feature Descriptors for Object detection. *International Journal of Computer Engineering and Applications*. 2018;p. 294–301. Available from: http://www.ijcea.com/wp-content/uploads/2018/02/NCDCM_2017_paper_49.pdf.

2) Gupta S, Kumar M, Garg A. Improved object recognition results using SIFT and ORB feature detector. *Multimedia Tools and Applications*. 2019;78(23):34157–34171. Available from: https://dx.doi.org/10.1007/s11042-019-08232-6.

3) Wang W, Xu K, Niu H, Miao X. Emotion Recognition of Students Based on Facial Expressions in Online Education Based on the Perspective of Computer Simulation. *Complexity*. 2020;p. 1–9. Available from: https://dx.doi.org/10.1155/2020/4065207.

4) Shabat AMM, Tapamo JR. Angled local directional pattern for texture analysis with an application to facial expression recognition. *IET Computer Vision*. 2018;12(5):603–608. Available from: https://dx.doi.org/10.1049/iet-cvi.2017.0340.

5) Talele K, Tuckley K. Facial expression recognition using digital signature feature descriptor. *Signal, Image and Video Processing*. 2020;14(4):701–709. Available from: https://dx.doi.org/10.1007/s11760-019-01595-1.

6) wu S, Wang D. Effect of subject's age and gender on face recognition results. *Journal of Visual Communication and Image Representation*. 2019;60:116–122. Available from: https://dx.doi.org/10.1016/j.jvcir.2019.01.013.

7) Ergin S, Isik S, Gulmezoglu MB. Face Recognition by Using 2D Orthogonal Subspace Projections. *Traitement du Signal*. 2021;38(1):51–60. Available from: https://dx.doi.org/10.18280/ts.380105.

8) Niu B, Gao Z, Guo B. Facial Expression Recognition with LBP and ORB Features. *Computational Intelligence and Neuroscience*. 2021;p. 1–10. Available from: https://dx.doi.org/10.1155/2021/8828245.

9) Meng H, Yuan F, Wu Y, Yan T. Facial Expression Recognition Algorithm Based on Fusion of Transformed Multilevel Features and Improved Weighted Voting SVM. *Mathematical Problems in Engineering*. 2021;p. 1–17. Available from: https://dx.doi.org/10.1155/2021/6639598.

10) Jeong D, Kim BG, Dong SY. Deep Joint Spatiotemporal Network (DJSTN) for Efficient Facial Expression Recognition. *Sensors*. 2020;20. Available from: https://doi.org/10.3390/s20071936.

11) Tian Y, Cheng J, Li Y, Wang S. Secondary Information Aware Facial Expression Recognition. *IEEE Signal Processing Letters*. 2019;26(12):1753–1757. Available from: 10.1109/LSP.2019.2942138.

12) Kim JH, Kim BG, Roy PP, Jeong DM. Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. *IEEE Access*. 2019;7:41273–41285. Available from: 10.1109/ACCESS.2019.2907327.

13) Koujan MR, Alharbawee L, Giannakakis G, Pugeault N, Roussos A. Anastasios Roussos, "Real-time Facial Expression Recognition "In The Wild" by Disentangling 3D Expression from Identity". *15th IEEE International Conference on Automatic Face and Gesture Recognition*. 2020. Available from: https://arxiv.org/pdf/2005.05509.pdf.

14) Jeong D, Kim J, Lee Y, Kim B. Robust Weighted Keypoint Matching Algorithm for Image Retrieval. *Proceedings of the 2018 the 2nd International Conference on Video and Image Processing*. 2018. Available from: https://doi.org/10.1145/3301506.3301513.

15) Chhabra P, Garg NK, Kumar M. Content-based image retrieval system using ORB and SIFT features. *Neural Computing & Applications*. 2020;32:2725–2733. Available from: https://doi.org/10.1007/s00521-018-3677-9.

16) Kumar M, Chhabra P, & Naresh Kumar Garg. An Efficient Content Based Image Retrieval System using BayesNet and K-NN. *Multimedia Tools and Applications*. 2018. Available from: 10.1007/s11042-017-5587-8.

17) Kumar A, Kaur A, Kumar M. Face detection techniques: a review. *Artificial Intelligence Review*. 2019;52(2):927–948. Available from: https://dx.doi.org/10.1007/s10462-018-9650-2.

18) Kumar M, Bansal M. 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimedia Tools and Applications*. 2021. Available from: 10.1007/s11042-021-10646-0.

19) Monika, Kumar M. XGBoost: 2D-Object Recognition Using Shape Descriptors and Extreme Gradient Boosting Classifier. *International conference on Computational Methods and Data Engineering*. 2021. Available from: 10.1007/978-981-15-6876-3_16.

20) Singh S, Ahuja U, Kumar M. Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimedia Tools and Applications*. 2021;80:19753–19768. Available from: https://doi.org/10.1007/s11042-021-10711-8.

21) Hema D, Kannan S. Hybridizing Local and Global Features by Sequential Fusion Technique for Object Detection and Learning. *International Journal of Advanced Science and Technology*. 2020;29(05):2401–2407. Available from: http://sersc.org/journals/index.php/IJAST/article/view/11023.