RECOGNITION AND CLASSIFICATION OF FACIAL EXPRESSION THROUGH MULTIPLE FEATURES

M. Binthu Kumari[1] **Dr. B. Sivagami**[2]

[1] **M. Binthu Kumari**, Research Scholar, Reg. No.: 18223152162009, Department of Computer Science & Applications, South Travancore Hindu College, Nagercoil, Affiliated with Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India.

[2] **Dr. B. Sivagami**, Associate Professor, Department of Computer Science & Applications, South Travancore Hindu College, Nagercoil, Affiliated with Manonmaniam Sundaranar University, Abishekapatti ,Tirunelveli-627012, Tamil Nadu, India.

[1]Corresponding Author: M. Binthu Kumari (saibithu@gmail.com)

Abstract

Recognizing human facial expressions is one of the most important and inspiring tasks in social communication.Face expressions are often a simple and obvious way for people to express their feelings and intentions.Facial expressions are the essential components of nonverbal communication.This study developed a model for detecting facial expressions that used both deep learning and handcrafted features. Preprocessing, feature extraction, and recognition are the three divisions of the suggested technique. The face recognition dataset is first preprocessed to reduce the noise using the Unsharp Masking Laplacian Non-linear Filter model.The shape, texture, and deep features are then retrieved from the preprocessed image using the Histogram of Oriented Gradients (HOG), Local Binary Pattern (LBP), and VGG16 network, respectively.These three features are also integrated to provide a more precise recognition. Finally, the K-Nearest Neighbour classifier is used to recognise the face expressions. The FER, CK+, and KDEF datasets are used to evaluate experiments. The suggested framework is successful and accurate, according to experimental results of seven emotion states (neutral, joy, sorrow, surprise, anger, fear, and disgust).

Keywords: Deep Learning, Feature extraction, Facial expression recognition, Laplacian Non-Linear Filter, K-Nearest Neighbour.

DOINumber: 10.48047/nq.2022.20.2.NQ22380 **NeuroQuantology2022;20(2):800-818**

1. INTRODUCTION

Human facial expressions are regularly utilised to interpret emotional states of people, which is very important in social communication. Both verbal and nonverbal cues are typically used in conversation. Facial expressions are used to convey nonverbal messages. The subtle indications of the bigger conversation are expressed in a person's face. Human communication that takes place through gestures, facial expressions, body language, and paralanguage is known as nonverbal communication.

To successfully connect with another person while employing service robots, intelligent agents, or smart phones to assist patients' requirements in hospitals and to provide a user-friendly service [1], it is crucial to comprehend human emotional states.Most cultures exhibit the fundamental facial expressions of anger, fear, contempt, sorrow, happiness, and surprise, and many facial expression recognition algorithms employ these expressions even though additional research indicates that cultural differences exist in facial manifestations of emotion [2]. By assessing the accurate creation of facial expressions for certain emotions and the proper comprehension of expressions made by others, a facial expression recognition system may also be beneficial for performance-based human emotional intelligence evaluation [3].

This present study aims to develop an automatic facial expression recognition framework using deep features as well as hand crafted features. Face boundary coordinates and edge-specific information limited the effectiveness of the facial emotion detection techniques that were already in use. To address this issue, with the help of the Laplacian Non-linear Filter and the Unsharp Masking function, this model emphasises the picture borders, defining the coordinates of the face's boundaries and minimising inaccuracy in the detection of facial expressions. Then the incorporation of deep and hand crafted features extracted from the preprocessed images are fed into the K-Nearest Neighbour classifier for accurate facial expression recognition.

Contribution of this Study

- Proposed an efficient preprocessing model to improve noise ratio by employing Unsharp Masking Laplacian Non-linear Filter.
- Incorporation of Deep VGG16 features along with HOG and LBP features will improve the accuracy of facial expression recognition.

2. RELATED WORKS

A few studies on facial expression recognition have been reported in the literature.A texture-based LDN feature descriptor was presented by Rahul et al. [4] in conjunction with PCA, which makes it easier to discern face emotion. Two masks, such as the Robinson and Kirsch mask, are used to compare the effectiveness of obtaining directional information. Ryu et al. [5] introduced a novel local pattern called LDTP that effectively stores the forms of emotion-related characteristics such the eyebrows, eyes, mouth, and upper nose by employing the directional information. Using ternary patterns, LDTP's robust encoding can discriminate between directed patterns on edges and smooth areas where random, meaningless, and noise-sensitive patterns are formed. Active patterns and subregions are used to efficiently characterise face images based on LDTP for strong facial expression identification, helping to assign additional spatial information to facial characteristics associated with emotions.Using uniform Extended Local Binary Pattern (ELBP) patterns combined with the K-L transform's covariance matrix, Guo et al [6] introduced the unique face expression identification technique known as BK-ELBP (KLT). In order to extract the major feature vectors from the expression pictures, the covariance matrix transform is employed on the ELBP matrix in the first stage to extract the feature matrix of the expression images. SVM is also used for classification, and this yields the best recognition results.

Chang et al. [7] introduce a new support vector machine-based facial expression recognition (FER) method (SVM). The Haar-like features approach and the self quotient image (SQI) filter are used in the first step of the FERS methodology to create a face detection method. The design of the feature extraction for facial expression in the FERS approach is then carried out simultaneously using three schemes: the angular radial transform (ART), the discrete cosine transforms (DCT), and the Gabor filter (GF). Using local binary patterns (LBP) for feature extraction, Jain et al. [8] developed an ensemble-based facial expression detection system. The global shape and local texture of face photos are represented by the retrieved feature histogram. For classification, three distinct classifiers Euclidian distance, neural network, and support vector machine are employed. Comparing the performance of the suggested ensemble classifier technique to that of individual classifiers, it has been shown to perform better.

In order to help computers communicate with people more intelligently, Kumar et al. [9] describe a framework for facial expression detection that infers emotional states in real-time. The system is made resistant to size and position fluctuations by the histograms of oriented gradients (HOG) characteristics that are derived from the active face patches. The feature vectors are then given to a support vector machine (SVM) classifier, which divides them into six universal expressions or neutral categories.

To identify facial emotion, Jiang et al. [10] introduced a salient facial patches (SFP) technique dubbed Profile Salient Facial Patches (PSFP). The facial patches are picked utilising the identified facial landmarks in order to acquire the prominent facial patches from profile face photos. To extract profile facial expression characteristics, three traditional methods the histogram of oriented gradients (HOG), local binary pattern (LBP), and Gabor were used. According to experimental findings, PSFP with HOG characteristics can obtain greater accuracy under the majority of head rotations.

Facial expression recognition technology was introduced by Islam et al. [11]. A facial picture is first divided into many sub-images, such as the left eye, right eye, mouth, nose, etc. The features are then extracted from each sub-image individually using a transformation based on two-dimensional principal component analysis (2DPCA). Last but not least, a classification operation is done to classify facial expressions using a minimum distance classifier (MDC).

Using computer-generated markers, Murugappan et al. [12] developed a triangulation approach for deriving a fresh set of geometric characteristics to categorise six emotional expressions. Haar-like traits are used to identify the subject's face. By moving the locations of eight markers to create each triangle's edge, five triangles are created. Subsequently, the Lucas-Kanade optical flow algorithm tracks these eight markers continuously while people make facial expressions. To categorise the facial emotions, the area of the triangle (AoT), inscribed circle circumference (ICC), and inscribed circle area of a triangle (ICAT) are retrieved as characteristics. Six different face emotions are distinguished using these characteristics and a variety of machine learning methods.

To increase the robustness of facial emotion detection, Wu et al [13] suggested an approach based on Local Binary Pattern (LBP) in conjunction with enhanced deep belief networks (DBNs). The upgraded deep belief networks are used as the detector and classifier in this technique after LBP is used to extract the feature.Two components of convolutional neural networks were suggested as part of a face emotion identification method by Mehendale et al. [14]. The backdrop of the image is removed in the first step, and the face feature vector extraction is the focus of the second. The five basic forms of regular facial expression are found using the expressional vector (EV) in the FERC model.

A unique feature fusion dual-channel expression recognition system was put out by Song et al. [15] and is based on both machine learning theory and philosophical thought. Particularly, the issue of minor variations in facial emotions is disregarded by the feature derived using convolutional neural network (CNN). The ROI area's Gabor feature is entered into the proposed algorithm's initial route. The second path proposes an efficient channel attention network based on depth separable convolution to enhance linear bottleneck structure that combines the depth of the feature map with spatial information. In [16], the frequency neural network (FreNet), a deep learning-based technique was implemented with the goal of achieving the best possible computing efficiency in facial expression identification. The FreNet concentrated on examining the frequencies found in the picture, which were more effective in eliminating redundancies in facial features. But it was not able to effectively capture the nuances in the spatial domain (pixel-to-pixel relationships). Computers may be able to produce more precise predictions about a person's mental state and more individualised replies by recognising or detecting facial expressions that evoke human emotions.

In order to accomplish automated recognition, Shan et al. [17] used a deep convolutional neural network (CNN), which is able to find more detailed feature representations of facial emotion. The suggested system consists of an input module, a pre-processing module, a recognition module, and an output module. To make the results more compelling, a K-nearest neighbour (KNN) method is contrasted with CNN.

A design for an artificial intelligence (AI) system that can recognise emotions from facial expressions is presented by Jaiswal et al. [18]. The three primary phases of the proposed technique are face detection, feature extraction, and emotion categorization. In order to recognise emotions from photos, this research developed a deep learning architecture based on convolutional neural networks (CNN).

Several automated approaches have been reported for face expression recognition with the low-level handcrafted features as well as deep features. Even though the feature representations available in literature are accepted, none have reported performance to a satisfactory level. Hence there is still a need for a good and an effective feature descriptor which could better aid in the recognition process. So this present study incorporate deep features and hand crafted features for more reliable and robust features. With this combination, it will be feasible to address the shortcomings of the Machine Learning (ML) and Deep Learning (DL) methodologies in the context of both big and small datasets.

3. Methodology

This section gives a description of the suggested method for identifying facial expressions. Figure 1 presents a complete map of the intended work. The suggested work is divided into three stages: pre-processing, feature extraction, and face expression recognition.

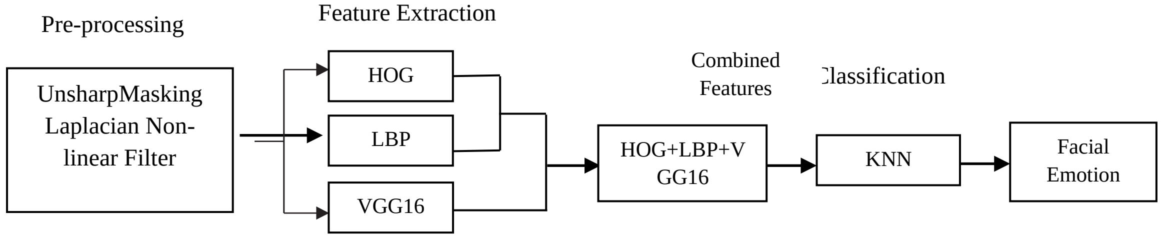


Figure 1: Complete map of the proposed methodz

As a first step, Unsharp Masking Laplacian Non-linear Filter model is employed to preprocess the face image. Then the series of shape, texture and deep features are extracted from the preprocessed image using the HOG, LBP and VGG16 methods. These features are combined and fed into the K-Nearest Neighbors (KNN) classifier for precise recognition.

3.1 Preprocessing

Laplacian Non-linear Filter model

In this article, the Unsharp Masking Laplacian Non-linear Filter model is employed to improve the noise ratio. In order to identify the coordinates for the face boundary, this filter model highlights the borders in the picture while masking the unsharp areas.

In order to emphasise picture borders, the Unsharp Masking Laplacian Non-linear Filter subtracts a blurred face image from the original face image. This increases boundary information in images. Figure 2 displays the outcome of using the Unsharp Masking Laplacian Non-linear Filter as a preprocessing step.

Figure 2: (a) Input Image (b) Pre-processed image with Unsharp Masking Laplacian Non-linear Filter

The above figure showsthe resultant pre-processed images that can improved noise ratio with application of non-linear filter.

As shown in the above figure, the resultant preprocessed face images both with and without application of Unsharp Masking Laplacian Non-linear Filter-based preprocessing shows an improved noise ratio with application of non-linear filter.

3.2 Feature Extraction

Feature extraction is a method used to extract crucial data about face pictures from pre-processed photos, which facilitates high accuracy recognition. A crucial necessity for the process of recognising face expressions is an effective feature extraction approach for extracting information from an input image.

3.2.1 HOG descriptor

Dalal and Triggs, researchers at the French National Institute for Research in Computer Science and Control, develop HOG descriptors [19]. HOG is an improvement over STIPs detection.Even without exact information of the associated gradient or edge coordinates, the distribution of local intensity gradients or edge orientations may frequently represent the look and structure of local objects rather effectively.The HOG descriptor is created on the accumulation of gradient axes across the pixels of a small spatial area known as a "cell," and in the subsequent building of a one-dimensional histogram whose concatenation provides the features vector to be taken into consideration for later uses.Unless the entire form is stable, HOG features are more sensitive to even minor variations in the object shape. Let I represent a gray scale intensity function that describes the image to be studied. The extraction of HOG features is seen in Figure 3.

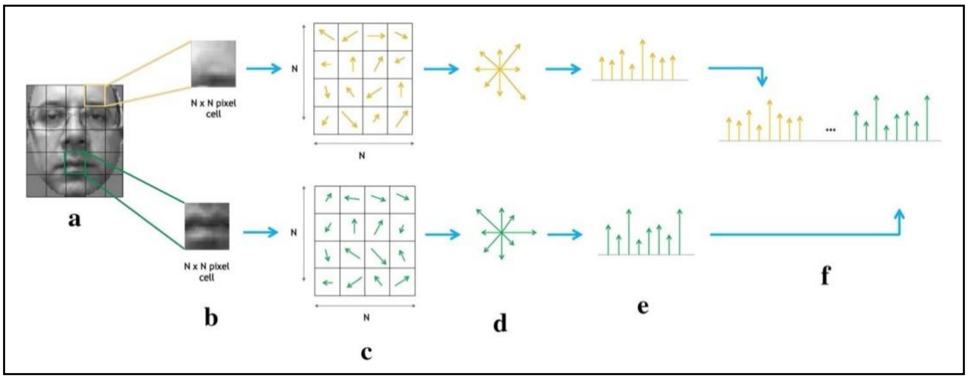


Figure 3: HOG Feature Extraction Method

As seen in Figure 3a, the picture is segmented into cells of size N pixels.Using Eqn.1 the orientation $\theta_{x,y}$of the gradient in every pixel is calculated, which is shown in Fig. 3b and 3c.

$$\theta_{x,y} = \tan^{-1}\frac{I(x,y+1)-I(x,y-1)}{I(x+1,y)-I(x-1,y)} \quad (1)$$

The orientations $\theta_i^j \, i = 1 \ldots\ldots N^2$ that are part of the same cell j are successively quantized and collected into an M-bins histogram, which is displayed in Figs. 3d and 3e. The end result of this algorithmic phase, i.e. the features vector to be taken into consideration for the next processing, is a unique HOG histogram (Fig. 3f), which is created by ordering and concatenating all the produced histograms.The example states that the cell histograms have a size of 4 pixels and 8 orientation bins.

3.2.2 Local Binary Patterns (LBP)

Texture features are critical for picture pattern analysis because only the textural aspects of images contain meaningful information for discriminating [20]. To extract texture information from photos, local binary patterns (LBP) were provided. LBP was initially defined in 1994 [21,22], and it has since been discovered to be a dominant feature for texture image.

The LBP operator is specified for 3 x3 neighbourhoods, uses each pixel as the centre pixel, and evaluates the eight pixels around the selected pixel depending on a predefined threshold. A local image descriptor is formed by the resultant binary-valued picture patch. The LBP operator is represented by the following Eqn.2:

$$LBP(X_c, Y_c) = \sum_{n=0}^{7} 2^n S(i_n - i_c) \quad (2)$$

For the centre pixel c's eight neighbours, the grey values of c and n are i_c and i_n respectively, and if $u \geq 0$, $s(u)$ is 1 otherwise 0. In a binary code, a uniform pattern is present if an LBP operator comprises no more than one 0-1 and one 1-0 transition.Primitive structural details for corners and edges are contained in the uniform pattern. The feature vector's length may be decreased using this information, and a straightforward rotation-invariant descriptor can be put into practise.In this work, the length of the feature vector for a single cell may be decreased by applying a uniform-pattern LBP descriptor to extract features from faces.Building a histogram represents the entire image after locating the local binary pattern.

The LBP procedure is illustrated in Figure 4.

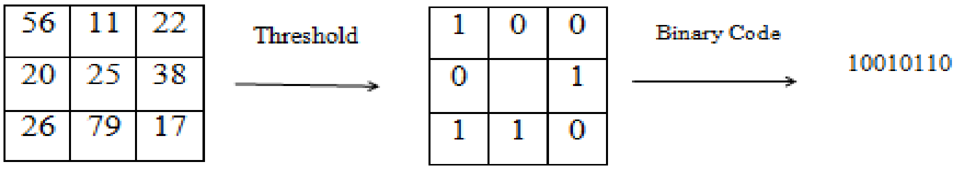


Figure 4: An example of LBP operation

3.2.3 Deep Features (VGG-16)

Both feature extraction and picture categorization are possible with the convolutional neural network. A CNN-based pre-trained VGG16 network is used in this study to extract deep features. Figure 5 depicts the VGG 16 network's design.A common neural network design called VGG-16 was suggested by Andrew Zisserman and Karen Simonyan of Oxford University [23].

Convolutional neural network model is the foundation of the VGG-16 network. It has five max-pooling layers with a 2x2 window size and 16 convolutional layers with a 3x3 filter size. Three completely linked layers come after a stack of convolutional layers.

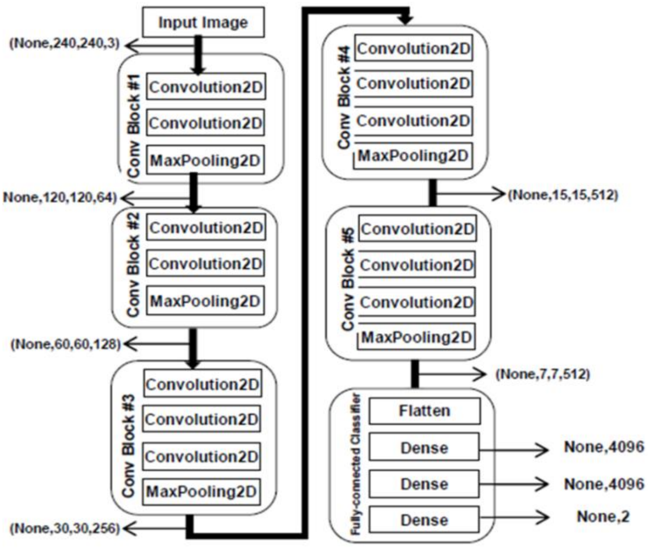


Figure 5: Structural design of VGG16

Five convolutional blocks are present. Two convolutional layers and one max-pooling layer are present in the first two blocks. Three convolutional layers and one max-pooling layer follow the first three blocks.The first block's input measures 240 × 240 x 3, and the RGB image's channel size is 3. The output of the first block is 64 channels, 120x120x64. The second block input receives this feature map, and so forth.The size of the channel doubles and the output of the next blocks are cut in half. The sixth block, measuring 7x7x512, created the final feature map. After each hidden layer comes the ReLu activation function.

As no one feature type can adequately represent all aspects of a picture, feature fusion is frequently used. In this study the extracted texture and shape features using LBP and HOG are incorporated to form a new feature.Moreover, deep VGG16 characteristics are used with these included features to improve classification accuracy. These new features have gained a lot of popularity since they frequently exceed hand-crafted features and deep features when combined.

4.1 Face expression recognition

After feature extraction, the recognition of facial emotion is the final step. K-Nearest Neighbour classifier is used to test these combined characteristics.

K-nearest neighbour (KNN) classifier

KNN is regarded as the most straightforward but effective algorithm and the research area have effectively used it to address both supervised and unsupervised learning issues.An input sample is categorised by a majority vote of its neighbours, with the image being allocated to the class that is most popular among its k closest neighbours.It is said to be quicker than the majority of other methods for finding correlation.To get better performance, the distance measure utilised for classification should be properly chosen.One of the distance measures, such as Euclidean distances, Canberra distances, Chi-square distances, or the Manhattan distance measure, can be used to determine the distance between the training set and the test set.Let $T = \{(s_1, c_1), (s_2, c_2), \ldots, (s_n, c_n)\}$, be the representation of the training set, where n is the number of input feature vectors and $c_i (i = 1, 2, \ldots, n)$ is the number of output class labels. With the majority of k-neighbours having the shortest distance, the test sample t_i as feature map is given the class label. This study utilized Euclidean distance using the Eqn.3. The number of feature sets utilised for training in this case iss_i, while the number of feature sets used for testing ist_i.

$$E_d(s,t) = \sqrt{\sum_i^l (s^2(i), t^2\,(i))} \quad (3)$$

The closest neighbour classifier uses this method to estimate similarity. As it calculates the distance between two histograms' nearest neighbours, this metric draws inspiration from instinct. It simply includes simple operations; hence its computational complexity is relatively minimal.

4. Experimental Analysis

In this section the details of the experiments carried out in this study are provided. To test the efficacy of the suggested facial expression recognition technique, two experiments using the FER, Chon-Kanade CK+ [24], and KDEF databases were carried out.The testing set in the first experiment included the participants who were utilized for training. The training individuals in the first experiment were also in the testing group.The subjects from the first experiment were not employed in the second experiment.The participants utilized for training were not employed for testing in the second experiment.This indicates that cross-validation strategies, dependent variables (DV), and independent variables (IV), are employed in Experiment 1 and Experiment 2, respectively. K-Nearest Neighbor classifier is employed in the experiments.

4.1 Dataset description

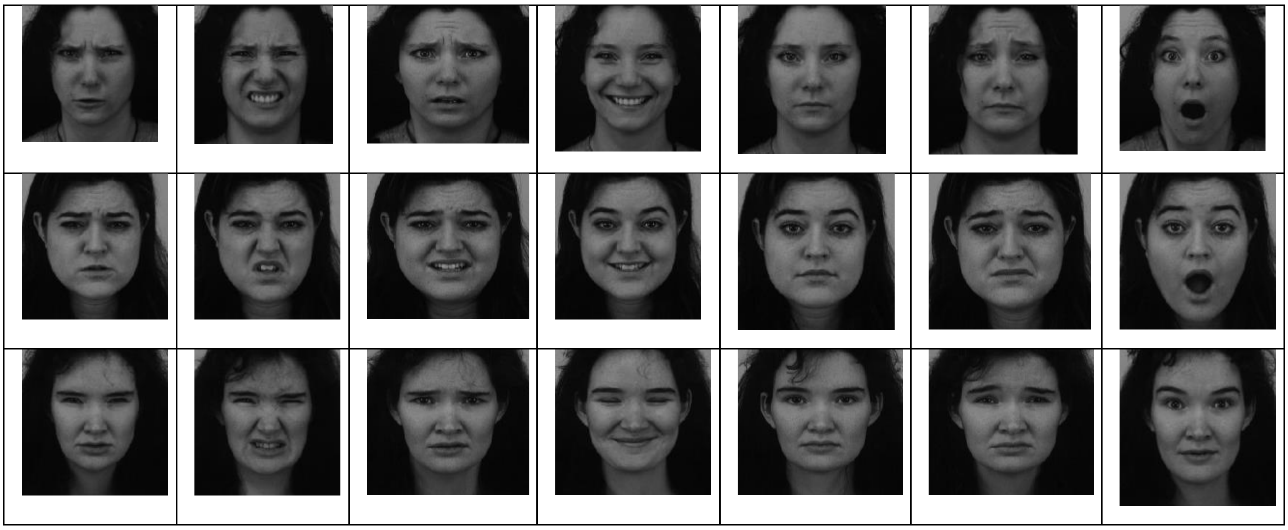
A database of face corrective re-annotation of FER, Chon-KanadeCK+, and KDEF is used to conduct facial expression recognition. The standard datasets had several photos that were discovered to be mismatched despite human review. Corrective hand re-annotation of the photos is done in order to accomplish better model training.

There are 32,900+ total photos, divided into 8 kinds of emotions, including anger, disgust, happiness, sadness, fear, contempt, neutrality and surprise. The photos in the collection are all PNG files with 224 by 224 gray scale resolutions, and they are all of human faces in gray scale. The following Table1 lists the total number of hand annotated photos in the training and validation sets for each category of phrases.

Table 1Accuratenewly annotated images

S. No	Image Categories	Total
1	Anger	4725
2	Disgust	795
3	Happiness	9049
4	Sadness	5403
5	Fear	3454
6	Contempt	130
7	Neutrality	5072
8	Surprise	4226

The following figure 6 displays the many facial expressions that were extracted from the input face dataset.



(a) anger (b) disgust (c) fear (d) happy (e) neutral (f) sad (g) surprise

Figure 6:Various expressions on the face

4.2 Results and Discussion

4.2.1 Experiments based on FER dataset

The experimental findings and analyses produced by the LBP, HOG, and VGG16 feature descriptors are provided in this part. The FER dataset's face photos were first pre-processed by the Laplacian Non-linear Filter model, which produces a clear image, in order to gauge the classification performance.Next, shape (HOG), texture (LBP), and VGG16 features are extracted from the pre-processed image. In order to identify which extracted feature performed the best, each extracted feature is then independently put into the k-Nearest Neighbour classifier. Moreover, the K-Nearest Neighbour classifier is used to assess the integrated characteristics.

The results of proposed face recognition method using the suggested features and the K-Nearest Neighbour (KNN) classifier on FER dataset are displayed in Table 4.1. The combined (HOG+LBP+VGG16) features exceed the separate shape, texture, and deep features, as shown by the recognition results in this table.

The accuracy of experiments with different features areshown in Table 1. Results of combined features (HOG+ LBP + VGG16) were better than theresults of individual features in Experiment 1 and Experiment 2. The K-Nearest Neighbor classifier with combined features achieves accuracy of 88.57% for experiment 1, and accuracy of 77.43% for experiment 2. Experiment 1 gives better accuracy of +11.14% than experiment 2.

The inference from Table 2 is that hybrid features with KNN classifier on experiment 1 of FER dataset have resulted in improvement in accuracy by 7.79%, 10.01%, and 11.87% respectively than VGG16, LBP, and HOGfeatures with KNN classifier. Moreover, HOG-LBP-VGG16 with KNN classifier on experiment 2 has resulted in improvement in accuracy by 3.38%, 6.83%, 12.93% respectively than VGG16, LBP, and HOGfeatures with KNN classifier.

Table 2: The accuracy of different methods on FER Dataset

Experiment 1		**Experiment 2**	
Features	Accuracy (%)	Features	Accuracy (%)
HOG	76.7	HOG	64.5
LBP	78.56	LBP	70.6
VGG16	80.78	VGG16	74.05
HOG+LBP+VGG 16	**88.57**	HOG+LBP+VGG 16	**77.43**

Figure 7 illustrate the accuracy of experiment 1 and experiment 2 with different features on FER dataset.

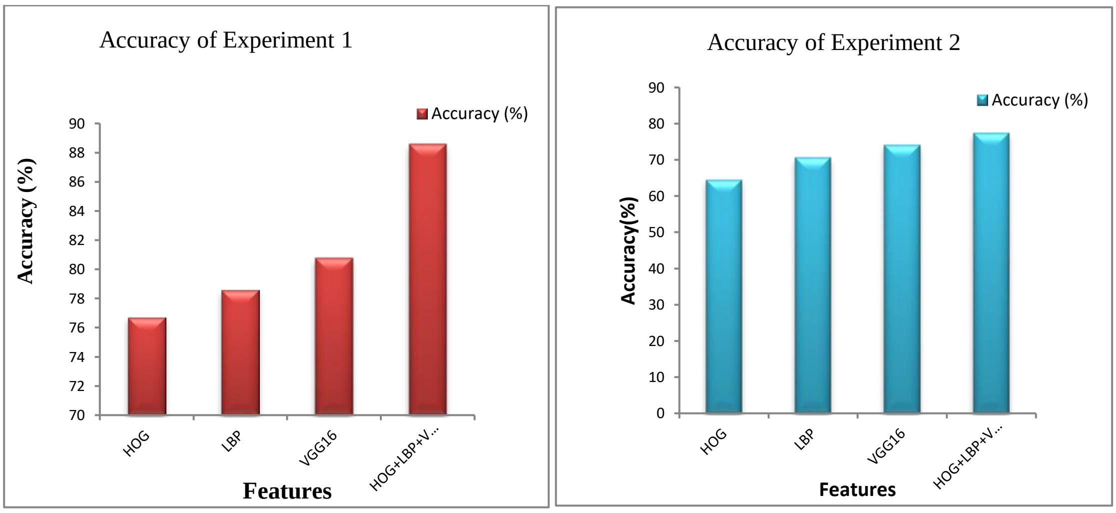


Figure 7: Accuracy of different features for Experiment 1 and Experiment 2

4.2.2 Experiments based on CK+ dataset

The tests based on the CK+ database followed the same process as those based on the FER database.The results of proposed face recognition method using the suggested features and the K-Nearest Neighbour (KNN) classifier on CK+ dataset are displayed in Table 4.2. According to the results of the recognition in this table, the combined (HOG+LBP+VGG16) features outperform the individual shape, texture, and deep features in this dataset.

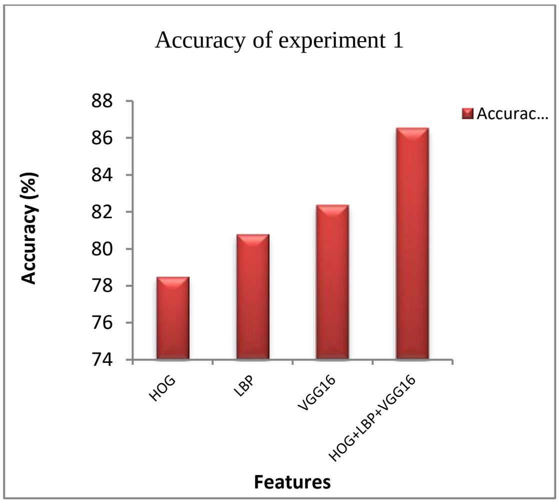
The accuracy of experiments with different features areshown in Table 2. In Experiment 1 and Experiment 2, combined feature findings (HOG + LBP + VGG16) outperformed individual feature results. The K-Nearest Neighbor classifier with combined features achieves accuracy of 90.5% for experiment 1, and accuracy of 84.5% for experiment 2. Experiment 1 gives better accuracy of +6% than experiment 2.

The inference from Table 3 is that hybrid features with KNN classifier on experiment 1 of CK+ dataset have resulted in improvement in accuracy by 3.8%, 9.05%, and 11.07% respectively than VGG16, LBP, and HOGfeatures with KNN classifier. Moreover, HOG-LBP-VGG16 with KNN classifier on experiment 2 has resulted in improvement in accuracy by 7.8%, 9%, 16.03% respectively than VGG16, LBP, and HOGfeatures with KNN classifier.

Table 3:The accuracy of different methods on CK+ Dataset

Experiment 1		**Experiment 2**	
Features	Accuracy (%)	Features	Accuracy (%)
HOG	79.43	HOG	68.47
LBP	81.45	LBP	75.5
VGG16	86.7	VGG16	76.7
HOG+LBP+VGG16	**90.5**	HOG+LBP+VGG 16	84.5

Figure 8 illustrate the accuracy of experiment 1 and experiment 2 with different features on CK+ dataset.



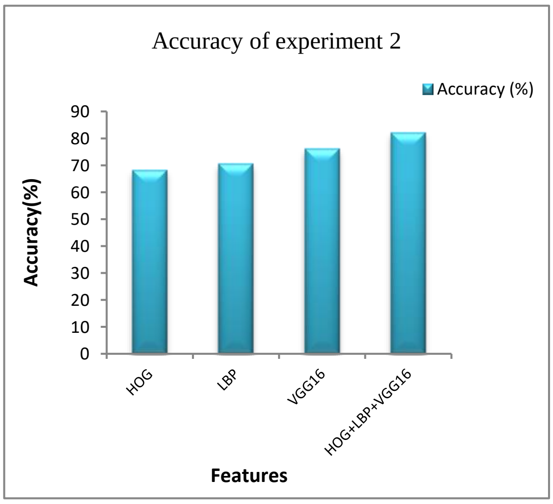


Figure 8: Accuracy of different features for experiment 1 and experiment 2.

4.2.3 Experiments based on KDEF dataset

Table 4 shows the results of the proposed face recognition approach utilising the K-Nearest Neighbour (KNN) classifier and proposed features on the KDEF dataset. The recognition results in this table demonstrate that the combined (HOG+LBP+VGG16) features outperform the individual shape, texture, and deep features.

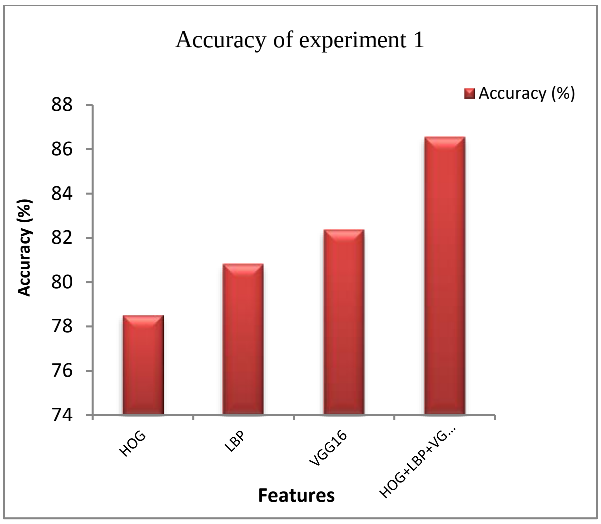
The accuracy of experiments with different features areshown in Table 3.Results of combined features (HOG + LBP + VGG16) were better than theresults of individual features in Experiment 1 and Experiment 2. The K-Nearest Neighbor classifier with combined features achieves accuracy of 86.53% for experiment 1, and accuracy of 82.22% for experiment 2. Experiment 1 gives better accuracy of +4.31% than experiment 2.

The inference from Table 4 is that hybrid features with KNN classifier on experiment 1 have resulted in improvement in accuracy by 4.17%, 5.74%, and 8.03% respectively than VGG16, LBP, and HOGfeatures with KNN classifier. Moreover, HOG-LBP-VGG16 with KNN classifier on experiment 2 has resulted in improvement in accuracy by 6.14%, 11.77%, and 14% respectively than VGG16, LBP, and HOGfeatures with KNN classifier.

Table 4:The accuracy of different methods on KDEF Dataset

Experiment 1		**Experiment 2**	
Features	Accuracy (%)	Features	Accuracy (%)
HOG	78.5	HOG	68.22
LBP	80.79	LBP	70.45
VGG16	82.36	VGG16	76.08
HOG+LBP+VGG16	86.53	HOG+LBP+VGG 16	82.22

Figure 9 illustrate the accuracy of experiment 1 and experiment 2 with different features on KDEF dataset.



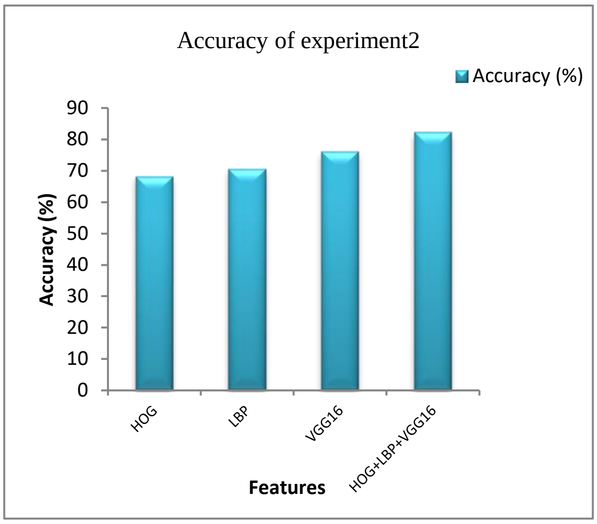


Figure 9: Accuracy of different features for experiment 1 and experiment 2.

5. Conclusion

An approach for recognising facial expressions using hybrid characteristics was proposed in this study. HOG, LBP and VGG16 feature descriptors were used for feature extraction, and KNN classifier was used to perform facial expression recognition. There are two experiments were conducted. In experiment 1, the testing set in the first experiment included the participants who were utilized for training. In experiment 2, the training individuals in the first experiment were also in the testing group. According to the experimental findings, the combinedfeature descriptorfor experiment1 outperformed various commonly used methodologies based on the FER database, CK+ database, and KDEF database.Because these combined features exceed both hand-crafted and deep features individually, they have gained a lot of popularity. Future research will take into account face expressions captured in video sequences.

Conflict of Interest

The authors declare that is no conflicts of interests.

References

[1] Song I, Kim H.J,Jeon P.B.,(2014). “Deep Learning for Real-Time Robust Facial Expression Recognition on a Smartphone”. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 10–13 pp. 564–567.

[2] Ekman P, (1994).“Strong evidence for universals in facial expressions: A reply to Russell’s mistaken critique”. Psychol. Bull. 115, 268–287.

[3] Mayer J.D, Roberts R.D, Barsade, S.G., (2008).“Human Abilities: Emotional Intelligence”. Annu. Rev. Psychol. 59, 507–536.

[4] Rahul R.C, Cherian M. (2016). “Facial Expression Recognition Using PCA and Texture-Based LDN Descriptor”. Proceedings of the International Conference on Soft Computing Systems. Advances in Intelligent Systems and Computing, vol 398. Springer. doi.org/10.1007/978-81-322-2674-1_11

[5] BRyu, Rivera, A. R, KimJ., &Chae, O. (2017). "Local Directional Ternary Pattern for Facial Expression Recognition", *IEEE Transactions on Image Processing*, 26(12), pp. 6006-6018, doi: 10.1109/TIP.2017.2726010.

[6] Guo, M., Hou, X., Ma, Y. *et al.* (2017). “Facial expression recognition using ELBP based on covariance matrix transform in KLT”. Multimed Tools Appl **76**, 2995–3010. doi.org/10.1007/s11042-016-3282-9.

[7] Chang, H.T.Y., (2017). “Facial expression recognition using a combination of multiple facial features and support vector machine”. Soft Comput. 22, 4389–4405. doi.org/10.1007/s00500-017-2634-3.

[8] Jain, S., Durgesh, M., Ramesh, T., (2016). “Facial Expression Recognition Using Variants of LBP and Classifier Fusion”. Int. Conf. ICT Sustain. Dev. 725–732. doi.org/10.1007/978-981-10-0129-1.

[9] KumarP,Happy S. L, &RoutrayA. (2016). "A real-time robust facial expression recognition system using HOG features," International Conference on Computing, Analytics and Security Trends (CAST), pp. 289-293, doi: 10.1109/CAST.2016.7914982.

[10] Jiang, B., Zhang, Q., Li, Z. (2012). “Non-frontal facial expression recognition based on salient facial patches”. J Image Video Proc. 15 (2021). doi.org/10.1186/s13640-021-00555-5

[11] Islam, D.I, Anal, S.R.N, Datta, A., (2018). “Facial Expression Recognition Using 2DPCA on Segmented Images”. Adv. Comput. Commun. Paradig. 289–297. doi.org/10.1007/978-981-10-8237-5

[12] Murugappan M, Mutawa A. (2021). “Facial geometric feature extraction based emotional expression classification using machine learning algorithms”. PLoS One. 18;16(2):doi: 10.1371/journal.pone.0247131.

[13] Wu, Y., Qiu, W., (2017). “Facial Expression Recognition based on Improved Deep Belief

Networks". AIP Conf. Proc. 1864. doi.org/10.1063/1.4992947

[14] Mehendale, N. (2020). "Facial emotion recognition using convolutional neural networks (FERC)". SN Appl. Sci. **2**, 446. doi.org/10.1007/s42452-020-2234-1

[15] Song Z. (2021)."Facial Expression Emotion Recognition Model Integrating Philosophy and Machine Learning Theory". Front Psychol.12:759485. doi: 10.3389/fpsyg.2021.759485.

[16] Tang, Yan, et al. "Facial Expression Recognition Using Frequency Neural Network." IEEE Transactions on Image Processing, 2021. DOI: 10.1109/TIP.2020.3037467

[17] Shan, K., Guo, J., You, W., Lu, D., Bie, R., (2017). "Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural-Network Structure". IEEE 15th Int. Conf. Softw. Eng. Res. Manag. Appl. 123–128.

[18] A.Jaiswal, A. KrishnamaRaju and S. Deb, (2020). "Facial Emotion Detection Using Deep Learning,"International Conference for Emerging Technology (INCET), pp. 1-5, doi: 10.1109/INCET49848.2020.9154121 .

[19] Dalal N, Triggs B (2005). Histograms of oriented gradients for human detection. IEEE IntConfComput Vis Pattern Recognit (CVPR), pp 886-93.

[20] Haralick, R. M., Shanmugam, K., &Dinstein, I (1973)."Textural features for image classification". IEEE Transactions on Systems, Man, and Cybernetics. 6:610–21.

[21] Ojala, T. Pietikainen, M. Harwood, D. (1994). "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions". Proc. 12th IAPR Int. Conf. Pattern Recognit. Conf. A: Comput. Vis. Image Process, Vol. 1, pp. 582–585.

[22] Ojala, T.; Pietikäinen, M.; Harwood, D.(1996)."A comparative study of texture measures with classification based on featured distributions". Pattern recognition 29, 51–59.

[23] Simonyan, K., &Zisserman, A., (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[24] Lucey P,+ Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010)."The extended cohn-kanade dataset: A completed dataset for action unit and emotion-specified expression". In: IEEE

Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 94–101.