Facial Expression Detection and Recognition through VIOLA-JONES Algorithm and HCNN using LSTM Method

[1]Dinesh Kumar P, [2]Dr. B. Rosiline Jeetha

[1]Ph.D. Research Scholar, PG and Research Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India

[2]Professor and Head, PG and Research Department of Computer Science, Dr. N.G.P. Arts and Science College, Coimbatore, Tamilnadu, India

ABSTRACT

Facial expression, as one of the most significant means for human beings to show their emotions and intensions in the process of communication, plays a significant role in human interfaces. In recent years, facial expression recognition has been under especially intensive investigation, due conceivably to its vital applications in various fields including virtual reality, intelligent tutoring system, health-care and data driven animation. The main target of facial expression recognition is to identify the human emotional state (e.g., anger, contempt, disgust, fear, happiness, sadness, and surprise) based on the given facial images. This paper deals with the Facial expression detection and recognition through Viola-jones algorithm and HCNN using LSTM method. It improves the hypothesis execution enough and meanwhile inconceivably reduces the computational costs. In feature matching, the author proposes Hybrid Scale-Invariant Feature Transform (SIFT) with double δ-LBP (Dδ-LBP) and it utilizes the fixed facial landmark localization approach and SIFT's orientation assignment, to obtain the features that are illumination and pose independent. For face detection, basically we utilize the face detection Viola-Jones algorithm and it recognizes the occluded face and it helps to perform the feature selection through the whale optimization algorithm, once after compression and further, it minimizes the feature vector given into the Hybrid Convolutional Neural Network (HCNN) and Long Short-Term Memory (LSTM) model for identifying the facial expression in efficient manner.The experimental result confirms that the HCNN-LSTM Model beats traditional deep-learning and machine-learning techniques with respect to precision, recall, f-measure, and accuracy using CK+ database. Proposes Hybrid Scale-Invariant Feature Transform (SIFT) with double δ-LBP (Dδ-LBP) and it utilizes the fixed facial landmark localization approach and SIFT's orientation assignment, to obtain the features that are illumination and pose independent. And HCNN and LSTM model for identifying the facial expression.

Keywords : Face geometry, feature extraction, LBP, LSTM, SIFT, HCNN, f-measure.

I. INTRODUCTION

Facial Expression Recognition (FER) frameworks have pulled in extremely high research enthusiasm in the domain of artificial intelligence [1]. There are various methodologies generally utilized in PC vision based facial expression recognition up until this point: acknowledgment utilizing 2D still pictures and acknowledgment utilizing image sequences. Approaches utilizing image sequences frequently apply optical stream investigation to the image sequences and use design acknowledgment instruments to perceive optical stream designs related with specific facial expression. This methodology requires securing of numerous casings of pictures to perceive expressions and accordingly has constraints of real-time performance and robustness. Facial expression recognition utilizing still pictures frequently use include based techniques for acknowledgment and subsequently have genuinely quick execution [2]. In any case, the test is to improve the noise reduction, feature extraction and classification methods that function admirably, paying little heed to varieties in human subjects and natural conditions. In addition, attempting to structure the frameworks performing under various conditions (for example brightening variety of posture, diverse demeanor and so forth.) is a difficult issue in the feature extraction of the face acknowledgment. Feature extraction is a significant advance in the face of acknowledgment activity. The SIFT is here, being applied to remove the highlights over the facial pictures. The location descriptor is utilized to concentrate such highlights. A stale facial component extraction strategy utilized for facial expression identification, which depends on the distinction of nearby double example histogram arrangements [3]. To begin with, the Local Binary Pattern (LBP) method is utilized to remove features associated with facial texture. At that point the LBP Map is isolated into non-overlapping rectangle areas with explicit size, and a histogram is registered for every locale.

In [4], a feature selection network (FSN) to consequently concentrate and channel facial highlights by embedding a feature selection mechanism with the AlexNet. The planned feature selection mechanism adequately channels unessential highlights and underlines correlated features as indicated by learned feature maps. Automatic feature selection methods are significant in applications where it isn't attractive to alter the feature space, or when the feature extraction. There are numerous ways to deal with the feeling. The greater part of famous methodology is obtaining face-oriented features. In the genuine study halls, CCTV caught the facial-based features. Though in the web-based learning condition, it very well may be caught by a web-camera. The procedure further is used after the secured data is considered for getting the result of examination. Viola-Jones algorithm is one of the most ordinarily executed algorithms and it is constantly preparing for face location. It is structured by [5] through enforcing the AdaBoost algorithms for rectangular highlights classifiers in "cascade " phases. Their crude idea permits foundation pictures to be disposed of rapidly. On 15 edges, their face detection continued for every second (fps). Be that as it may, it is not quite receptive to angle variations in the face.

Face detection includes isolating picture windows into two categories: one with faces (targets), and one having the foundation (clutter). It is troublesome on the grounds that in spite of the aspect that there are traits shared between faces, they can oscillate extensively to the limit of age, skin shading and facial expression. With changing illumination, picture

features and forms, just as the plausibility of halfway impediment and mask [6,7], the issue is additionally scrambled. A right face finder would along these lines include the option to differentiate the proximity of any face under any environment of illumination, upon any foundation type. The assignment of face location can be divided into two phases. The initial step is a characterization task, which accepts a few optional pictures in the form of input and produces a paired estimation of yes or no, as the output showing if any faces exist in the picture. The face localization task forms the succeeding advance, which plans to accept a picture as the input and provides the resultant output for the area of any face or faces present inside that picture to be a bounding box having (x, y, width, height).

The significant issue of this examination work is the distinguishing proof of impediment because of enlightenment varieties after some time. Still it has issues with finding an ideal feature representation from the face thus the exhibition of FER is diminished. To conquer the previously mentioned issues, right now, image preprocessing is done and improves the nature of pictures. Face detection with impediment is performed through Viola Jones algorithm. At that point, extraction is performed by utilizing hybrid SIFT and Dδ-LBP algorithms to remove the informative features. At that point, Whale optimization algorithm is utilized to choose the ideal face features from the CK+ database. At that point the grouping is performed by utilizing HCNN-LSTM training and testing process. The commitment of this exploration is planned to improve the FER exactness and productivity utilizing proposed HCNN-LSTM algorithm impressively.

II. LITERATURE SURVEY

In [8], endeavored to study facial expression by utilizing Haar-like feature extraction technique. A lot of luminance stickers were inserted all over. Then the subject is told to make the necessary facial expressions. It utilizes a Haar-like method to remove the features and six statistical features such as variance, standard deviation, mean, power, energy and entropy are obtained from the approximation coefficients of Haar-like decomposition. Then these statistical features are utilized in the form of a contribution to the neural system for the classification of eight facial expressions. The feature variance gives superior outcome contrasted with few other measurable highlights. Anyway, it has issues with wrong outcomes.

The facial feeling of acknowledgment dependent on the Viola-Jones Algorithm in the learning condition was represented by[9]. The idea of Viola-Jones algorithm is basically a face discovery algorithm and, in any case,, we utilize facial-based features to distinguish confrontation and perceive feeling, in this manner, rectangular feature and cascading AdaBoost algorithm is applied. We distinguish the previous methods with the Viola-Jones algorithm's aspects like accuracy, precision, recall, and time-consumption, and it makes use of 50 UM's learning pictures in understudy of emotion detection. Then these aspects arrive at 0.74, 0.73, 0.76 and 15 seconds for each frame, while the past strategies[10] arrive at 0.46, 0.48, 0.52, and 42 seconds for each frame. From the result, it is confirmed that our proposed work is better when compared with our past research.

In [11], presented a completely programmed face detection algorithm and showed its exhibition on the FRGC v2.0 information. This algorithm is multimodal (2-Dimensional and 3-Dimensional) and carries out crossover (feature based and all-encompassing) coordination so as to accomplish proficiency and vigor to facial expressions. The posture exhibited by a 3-Dimensional face alongside its surface is consequently rectified utilizing a new methodology

dependent on a solitary naturally identified point and the Hotelling transform. An epic 3D spherical face representation (SFR)is utilized related to the scale-invariant feature transform (SIFT) descriptor to frame a dismissal classifier, which rapidly wipes out an enormous number of applicant faces at a beginning time for proficient acknowledgment in the event of huge exhibitions. The rest of the faces are confirmed utilizing a new region-based mechanism, which is powerful to facial expressions. This methodology naturally portions the eyes- forehead and the nose areas, which are moderately and slightly delicate to articulations and then matches them independently utilizing a changed iterative closest point (ICP) algorithm. The aftereffects of each of the coordinating motors are combined at this measurement level to accomplish greater precision.

Similarly, a hybridized feature subset selection algorithm known as the maximum Pearson maximum distance improved whale optimization algorithm was utilized by Zheng et al.[12] . To start with, in view of Pearson's correlation coefficient and correlation distance, a filter algorithm is introduced and we call it maximum Pearson maximum distance (MPMD). In order to change the loads of the pertinence and excess, two parameters are recommended in MPMD. Further, the adjusted whale optimization algorithm can go about as a wrapper algorithm. Finally, the ideal characterization exactness was identified. The proposed technique is tried on 10 standard datasets from UCI AI databases. From the experimental analysis, it is confirmed that the characterization exactness of the proposed algorithm is altogether higher when distinguished with the other three wrapper algorithms and one hybrid algorithm.

Furthermore, a feature selection system is presented [13] and it is enforced in the whale optimization algorithm (WOA) and it is a late presented meta-heuristic optimization algorithm, which follows the characteristic conduct of the humpback whales. The wrapper-based technique was applied by this proposed work in order to arrive at the ideal subset of features and further it has to get the best component subset, which expands the precision of the order when saving the base features. From UCI data repository, this proposed model is contrasted, particle swarm optimization (PSO) and genetic algorithm (GA) utilizing various appraisal markers on 16 distinct informational collections. The experimental analysis confirms that the upside of the acquainted algorithm contrasted with different enhancers.

Moreover [14], proposed a facial expression recognition utilizing nearby parallel examples. The fundamental commitment of this technique is the feature selection, in where huge difference LBP pixels were chosen to speak to faces. By choosing the pixels with high fluctuation dependent on LBPs, the acknowledgment rates got enhanced essentially. The tests are finished on the BU-3DFE database. These tests show that subsequent to using highlight determination, the rates of acknowledgement are increased by 11%.

Similarly,[15], profound learning-based spoof face recognition. Two distinctive profound learning models are utilized to accomplish this, to be specific receptive fields (LRF)-ELM and CNN. LRF-ELM. LRF-ELM is an as of late created model which comprises a convolution and a pooling layer prior to a completely associated layer, which renders the model quick. CNN, in any case, has a progression of convolution and pooling layers. Likewise, the CNN model may contain all the more completely associated layers. A progression of investigations was led on two mainstream parody face identification databases, specifically NUAA and CASIA. Then the acquired outcomes were thought about, and the LRF-ELM technique rendered superior outcomes against the two databases.

III. PROPOSED METHODOLOGY

Here the proposed HCNN-LSTM algorithm identifies the facial expressions such as sad, happy, fear, surprise, anger and neutral expressions more accurately. The proposed system comprises primary phases like image pre-processing, face detection, feature extraction, feature selection and FER using the proposed HCNN-LSTM algorithm.

A. Image preprocessing

It standardizes the picture procured from uncontrolled and uncooperative conditions. The requirement for preprocessing is: To conquer unsafe impacts of brightening, eliminates impacts rely upon angle, distance & lightning conditions against each face. An image histogram is a graphical representation of the tonal dissemination in an advanced image. It helps in plotting the quantity of pixels for every tonal worth. The tonal varieties are plotted along the horizontal axis of the chart, whereas the vertical axes speak to the quantity of pixels existing in that particular tone. Especially, a picture histogram of a gray- scale picture has in the horizontal axes the 256 brilliance levels, and in the vertical axes the occasions this level shows up in the picture Histogram Equalization [16]

For images, this method is helpful with both the background and foreground being bright or both non-illuminated. Predominantly, this process gives a remarkable view on the bone structure in the x-ray image in the biomedical domain, yielding clean and detail images

B. Viola-Jones algorithm for Face Detection with occlusion

Viola-Jones algorithm is utilized to recognize the internal essence of the square shape in the picture. At that point the complex face highlights of the square shape locale are broken down by discriminative investigation, and the hybrid face features are contributions to the classifier for achieving face acknowledgment. While building complicated features of a face in a rectangular casing, we have to utilize worldwide and nearby features separated from the face present inside the rectangular frame. Viola-Jones algorithm is a fell Adaboost classifier dependent on picture fundamentals. Each sub window moves, the indispensable picture is determined. The necessary picture is formed by the sum of the abovementioned and left pixel of x, y[17] .

The three fundamental ideas which permit it to run continuously are the vital picture, AdaBoost and the cascade structure. The Integral Image is an algorithm for financially savvy age of the total of pixel forces in a predetermined square shape in a picture. It is utilized for rapid calculation of Haar-like features. Estimation of the entirety of a rectangular region inside the first picture is very proficient, requiring just four augmentations for any self-assertive square shape size. AdaBoost is utilized for development of solid classifiers as a direct mix of weak classifiers. It is engaged to distinguish the appearances with impediment all the more successful, and furthermore it improves the computational unpredictability.

Viola Jones makes use of Haar features, for characterizing the recognized object like face, or it isn't based on the intensity level of various parts of the face. The significance of another object in front of the face hampers the detection of the acquired feature, perhaps to identify the face. More features with more data points were combined in the basic face detection method for maximizing the efficiency. Fig 1 represents few occlusions like sunglass on the face, mouth covered, eyes covered. Apart from these factors, the face detection can also be affected by hairs in front of the face or else an object in front of the face.

Fig 1 Simple occluded face images

Viola-Jones algorithm 1

1. Input: CK+ image
2. Output: state of emotion
3. for i← 1 to shift total do
4. for j ← 1 to stages total do
5. for k ← to filter total do
6. w_e = integral of sub windows
7. update weight_face
8. update weight_emotion
9. if w_e<weight_face then
10. break for k_loop
11. else
12. w_{tot} = Σ w_e
13. end_if
14. end_for
15. end_for
16. if w_{tot}< weight_ emotion then
17. output = bored
18. else
19. output = interest
20. end_if
21. end_for

where w_e refer to the weight of sub windows and w_i indicates the weight of pixel-i. The cascaded AdaBoost classifier helps to segregate the image integral for choosing the face area. Here, M filters get repeated at every stage [18]. We can extract the weak classifier (k_m), out of the classifier pool. So, the weight of classifier w_m is determined as below:

$$w_m = 1/2(ln\,ln(1 - e_m) - ln\,ln(e_m)) \quad (1)$$

where e_m refers to the normalized weight which is computed using:

$$e_m = W_e/W \quad (2)$$

Where W indicates the maximum weight. So, the threshold w_t gets updated applying:

$$w_t^{m+1} = \{w_t^m e^{\alpha_m k_m(x_l)} \quad miss \; w_t^{m_e - \alpha_m} \quad otherwise \quad (3)$$

Using the weight of the sub-windows *we distinguish* the updated threshold w_t on the next subsequent iteration (*m+1*) If the weight of the sub window is less than the threshold, the sub windows eliminate the face. If the weight of the sub window is greater than the threshold, the sub window gets processed in emotion detection.

AS like face detection, here the procedure of emotion recognition works, which comprises estimating the integral image and the cascaded AdaBoost helps divide them. But the integral image at this step is the summation of the filtered faces, expressed as below:

$$w_{tot} = \sum w_e^m \quad (4)$$

Where w_{tot} refers to the integral image achieved of the filtered faces-w_e^m. Here, by applying Equation (4), the threshold gets updated on every iteration (*m+1*). The weight of filtered faces helps to distinguish the updated threshold. In case the weight of filtered faces is lesser than the threshold, the filtered face is set as an 'interest' expression. In case the weight of filtered faces is greater than the threshold, the filtered faces are set as 'bored' expressions. Here, it is considered that when distinguished with the threshold, interest expressions exhibit a higher value, since some indicators like smile increase the texture feature's value thereby permitting a raised value.

C. Hybrid feature extraction using SIFT and double δ-LBP (Dδ-LBP) algorithm

Here, with the double δ-LBP (Dδ-LBP) method for feature extraction, the Scale Invariant Feature Transform (SIFT) algorithm is combined. In the case of machine vision, we enforce this algorithm, for extracting particular images features for applications like matching different views of an object or scenery (for binocular vision) and finding the objects. To scale and rotate the acquired features were invariant

and in lighting it exhibits partial invariance to any kind of changes and we call it as the local feature extraction technique which identifies and defines the local features present in images. The SIFT extracted features show invariance to scale of the image, orientation, variation in lighting and considerable extent of affine disturbance [18]. For object recognition purposes, we initially establish the SIFT feature extraction method.

The image information is transformed by SIFT into scale invariant directions comparative with local features. In order to change in enlightenment, the features show invariance to picture scaling and turn, and are invariant halfway. From normal pictures we can segregate huge feature quantities. For face acknowledgment, SIFT features are separated from a lot of training images and put away in the database. From the test picture to the current database, a test image is synchronized by separately seeking at each component. Depending on the Euclidean separation of their component vectors, the Best match between the removed features works. SIFT makes use of four significant phases of computation, in order to develop a set of computation. The four phases incorporate Scale-Space Extrema Detection, Key Point Localization, Orientation Assignment and Key Point Descriptor.

Key Points Extraction: Highly distinctive facial features were extracted through SIFT and further, the extraction of the key points for the complete images present in the training dataset and also the test image are done and the result key points of entire training images were controlled by the key point dataset. The steps in the key point extraction will include DoG Image Generation, Local Key Point Detection, Accurate Key Point Location and Eliminating Edge Response.

1. **Generation of Difference-of-Gaussian Images:** For key point detection, we ought to recognize the position and scale of the key point. We

confirm DoG as an effective function, for the key point location detection. Further we can estimate the DoG function, from the difference of two adjacent scales.

2. **Local Key Point Detection:** In the current image, each key point is distinguished with its eight neighbours and nine neighbours from nearby scales, which provide local maxima or minima.

3. **Accurate Key Point Location finding:** Those sensitive to noise will be rejected, which has low contrast from the number of key points and by fixing a threshold value for the key point that helps to achieve just the stabilized key points.

4. **Elimination of Edge Response:** When the edges are incorrectly decided and do not show resilience to noise, the key points will be eliminated. The principal of curvatures, poorly defined peaks. We can eliminate the unwanted key points, through fixing the extra threshold on the principle of curvatures

The Local Binary Pattern (LBP) is a generally utilized descriptor in facial expression detection because of its productivity and viability. Be that as it may, the available facial expression detection techniques dependent on LBP either disregard various types of data, for example, subtleties and the shape of countenances, or depend on the segmentation of facial images, for example, separating the face picture into squares or allowing the square fixating on tourist spots. Thinking about this issue, to utilize both detail and contour face data in facial expression detection, a new feature extraction strategy dependent on double δ-LBP (Dδ-LBP) right now. Right now, two δ-LBPs are utilized to speak to subtleties and the shape of appearances independently, which consider various types of data of facial expression [19] .

Key points Validation: This phase assists us to recognize the satisfactory key points. For handling face recognition, satisfactory key points form the

exceptionally particular key points that are separated from the picture. IN fig.2: Definite procedure associated with the Key Point Acceptance and Matching Algorithm is provided. Three stages are engaged with the procedure of approval: Calculation of Euclidean Distances, Sorting of Euclidean Distances and Key Points Acceptance.

Calculation of Euclidean Distances: When the extraction of the key points for the test and training set are done by the formula, we can estimate the Euclidean distance:

$$ED = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \qquad (5)$$

x_1, x_2, y_1, y_2 refer to x and y coordinates of the test and training image correspondingly. We can estimate the Euclidean distances, through associating each test key point with a whole set of key points in the training image and then the Euclidean distance array is formed.

- **Sorting of Euclidean Distances:** Through the comparison of the entire key points belonging to test image and the whole set of key points belonging to training image, we can estimate the array of Euclidean distances and for recognizing the first two closest neighbours, we ought to sort it out.
- **Key Points Acceptance:** Using the Euclidean distances array, we can estimate the Ratio of two closest neighbors, which, in turn, assist us to get rid of the fake key points. If this ratio is higher than 0.8, it has to be eliminated, since they were not considered as acceptable key points, else we accept it to be the valid key point (if the ratio is less than 0.8) and increase the key point counter that preserves the count of matched key points.

Here we estimate the Absolute Differences between the key points belonging to the thoroughly matched training images and the test image. More officially, I_{test} and I_{train} indicates the test and training image,

correspondingly. Unique Features for the test and training image will be expressed as:

$$K_T^{Itest} = \{K_1^{Itest}, K_2^{Itest}, K_3^{Itest}, \dots . K_N^{Itest}\} \qquad (6)$$
$$K_{TN}^{Itrain} = \{K_1^{Itrain}, K_2^{Itrain}, K_3^{Itrain}, \dots . K_N^{Itrain}\} \qquad (7)$$

N and M represents the number of key points present in the test and training images respectively. The key points of the training images that are sorted based on the MT will be denoted as:

$$K_{ST}^{Istrainset} = \{K_{1-E1}^{Istrain}, K_{1-E2}^{Istrain}, K_{1-E3}^{Istrain}, \dots . K_{1-En}^{Istrain}\}(8)$$

Where E1, E2 ….En refer to the key points belonging to the sorted training images. Key points of one of the sorted training images will be denoted as:

$$K_{ST}^{Istrain} == \{K_1^{Istrain}, K_2^{Istrain}, K_3^{Istrain}, \dots . K_E^{Istrain}\}(9)$$

D. Feature selection using whale optimization algorithm for face emotion

Here, whale optimization algorithm helps in determining the significant facial features from the given Ck+ database. This new optimization algorithm [20] can be propelled to imitate the common nature of the humpback whales. These whales normally depend upon the hunting behaviour as a method for durability. The specialty of the whale algorithm is about the capacity to utilize an arbitrary or best operator in the hunt space to pursue the prey, besides the hunting nature has been recently accustomed with address optimization issues. Recreating the bubble-net attachment techniques of the humpback whale by utilizing spirals, can also be done by this algorithm. The display of this algorithm incorporates three administrators: recreate the quest for prey (exploration step), the encircling prey, and the bubble-net foraging (exploitation step) behaviour of humpback whales.

It is utilized for face feature determination based on a wrapper-based technique. Enforcing the classification approach as a manual for feature determination

technique, dependent on some optimized component, was the key feature of the wrapper-based procedure; chose feature set, and further it enforces the WOA adaptively for recognizing the optimal face feature subset that augmenting the classification performance. The scientific definition is demonstrated and clarified as below:

Encircling of Prey: In this stage, the whale algorithm begins with a fundamental best pursuit specialist. The present arrangements were expected to be best, and it is the area of the prey or practically near it. The enduring specialists subsequently update their areas toward the best pursuit operator.

This can be defined in the expression as follows:

$$\vec{D} = |\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)|, \qquad (10)$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \qquad (11)$$

Where t represents the present iteration $\vec{A}$ and $\vec{C}$ refer to the coefficient vectors. The $\longrightarrow X*$ indicates the location vector of the best solution which is acquired till now and $\overrightarrow{X^*}$ stands for the location vector. If a better solution exists, then $X\,has$ to be updated through iteration.

The vectors $\overrightarrow{A}$ and $\vec{C}$ are computed as below:

$$\vec{A} = 2\,\vec{a}.\vec{r} - \vec{a} \qquad (12)$$

$$\overrightarrow{C} = 2.\,\vec{r} \qquad (13)$$

Where $\vec{a}$ is linearly reduced from two to zero with the number of iterations and r stands for the random vector in the range $[0, 1]$.

Any agent will be given by this modelling, in order to update its location in the field of the current best solution and reproduce the prey encircling. It further proceeds in the search space for n dimensions, and by the agents, the moves achieved in hyper cubes will get simplified, which helps to accomplish the best solution.

Exploitation phase: Also termed as bubble net attack and its operation is defined by two mechanisms given as below:

Shrinking encircling technique: We can locate the search agent, the randomness of $\vec{A}$ in [-1, 1], in

between the agent earlier position and the current best position.

Spiral updating position: Here, the movement of the whale towards its prey is got by concurrent application of the shrinking circling.

We need to estimate the distance between the locations where the whale and its prey are present, in order to pretend the movement of helix shape by humpback whales, and subsequently a spiral equation is developed between the locations taken by the whale and prey. This expression is given by:

$$\vec{X}(t+1) = \overrightarrow{D'}.e^{bl}.cos\,cos(2\pi l) + \overrightarrow{X^*}(t) \qquad (14)$$

Exploration phase: WOA helps to achieve the global optimization objective. Fig.3 indicates the whales search for their prey based on their positions to each other randomly. The $\vec{A}$ is set randomly from [-1,1] to indulge the search agent for navigating from the reference whale, which means that $\vec{A}$ has to be either greater than 1 or less than the -1. Moreover, we can perform the updated position of a search agent, by randomly choosing an agent that allows the WOA to perform global search. The modelling of this exploration mechanism is mathematically expressed as follows:

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \qquad (15)$$

Where $\overrightarrow{X_{rand}}$ refers to a random position for a random whale, which is selected from the present population.

In the case of WOA, the whales uninterruptedly correct their positions to any point in the space start, using an initial best search agent. Further, it works concurrently, in order to have their positions updated toward the best of the search agents as shown in equations (1) and (2). The individual solution is nothing but a constant vector having the same dimension as in the dataset. The solution vector values are consistent and restricted to [0, 1]. In binary representation, it is provided with the values achieved of the solution fitness evaluation. This

fitness function that is used simply matches the classification performance and the number of chosen features. This can be expressed by the equation (10) below.

$$f_\theta = \alpha.E + (1 - \alpha)\frac{\Sigma_i \theta_i}{N} \qquad (16)$$

Where f_θ refers to the fitness function provided in the form of a vector θ with size N having 1/0 elements defining the chosen/unchosen feature subset, N indicates the overall number of features present in the dataset. E stands for the classification error with the chosen feature subset provided, and α refers to a constant that controls the trade-off existing between the classification errors to the number of chosen feature subset. Here, the major objective is the classification performance; hence $\alpha = 1$ is used.

Algorithm 2

1. Initialize the population of whales $Xi(i = 1, 2, 3, ..., n)$.
2. Compute the fitness of every whale (face features).
3. Set X to be the best whale.
4. while ($t <maximumnumberofiterations$) do
5. for (eachsearchwhale) do
6. Update a, A, C, l and p.
7. if ($p <0.5$) then
8. if ($|A| < 1$) then
9. The face position is updated by the Eq. (10).
10. Else
11. if ($/A/ \geq 1$) then
12. Select the random whale Xrand
13. The whale position is updated
14. End
15. End
16. End
17. if ($p \geq 0.5$) then
18. Change the whale (face) position using the Eq.(14).
19. End
20. End
21. End
22. Verify whether any search agent goes outside the search space and amend it.
23. Compute the fitness of every search agent.
24. Update X∗ in case of existence of a better solution (15).
25. t = t+1
26. end

E. Face recognition by hybrid CNN-LSTM learning method

In this research work, Hybrid Convolutional Neural Networks (CNN) – Long Short Term Memory (LSTM) method is proposed for FER respectively. Image is the input for the system; then, it utilizes HCNN-LSTM for forecasting the facial expression label that has to be one of these labels, such as anger, happiness, sorrow, fear disgust and neutral. The complete block diagram of the newly introduced system is shown in Figure 2.
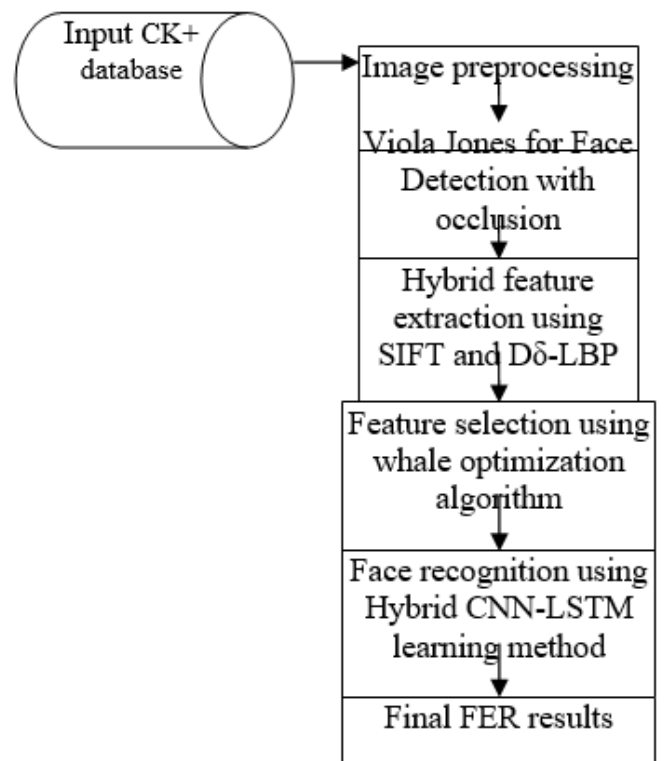

Fig 2 The entire block diagram of the novel system

The distinguished parts of the face are edited and separated and afterward utilized in the form of the contribution to the CNN's origin layer. The Training stage includes the feature extraction and classification through convolution neural systems. It is quite normal that utilizing the facial segments rather than the entire face's picture to be the contribution for the origin layer will decrease training time and increment the likelihood of elevated degree feature extraction along these lines expanding framework exactness.

CNNs utilize the ideas of open field and sharing of weight. Using these ideas, the quantity of trainable parameters gets diminished and the spread of data throughout the layers can be determined through convolution [21]. In this, a convolution of sign and a filter map is carried out, having the common loads to create a feature map.

Here, HCNN is introduced to arrange the facial appearances for the given CK+ dataset. Fundamentally CNN consists of an input and an output layer, fair as numerous hidden layers. The following were the different layers of CNN, which are the hidden layer: Convolutional layers, pooling layers and completely connected layers. The convolutional layer carries out the convolution activity of the input and the resultant convoluted output is moved to the next subsequent layer. This will follow the reaction of every single neuron to visual upgrades. Convolution systems involve neighbourhood or worldwide pooling layers that merges the output of neuron groups present in a single layer into a solitary neuron in the next subsequent layer. In the previous layer, mean pooling utilizes the normal incentive from each group of neurons. Every neuron in one layer to each neuron in the next layer were linked by methods for having layers that are completely connected. The essential guideline of CNN is the same as customary multi-layer perceptron neural system [22] .The newly introduced HCNN involves input layer, convolutional layer, sub-sampling layer and classification layer. This proposed technique has evident points of interest for breaking down high-dimensional information, which help as a parameter sharing plan through which the quantity of parameters were diminished and then constrained by the convolutional layers.

Input layer gets facial expressions features from training tests and changes the information into a brought together structure so as to convey the information into the next layer effectively. The underlying parameters, for example, the size of the neighbourhood receptive fields and different filters are characterized in this layer.

The facial expressions input data was convoluted by Convolution layer (Cx) through convolution algorithm and from the previous layer we created several layers called feature maps comprising convolution algorithm results. Extracting the key features was the main purpose here and further it reduces the computational complexity of the network. The convolution is given by the output equation given as below:

$$x_j^l = f(\sum_{i \in mj} x_i^{l-1} * k_{ij}^l + b_j^l) \qquad (17)$$

$$f(x)\frac{1}{1+e^{-x}} \qquad (18)$$

Where x refers to the output value of the convolution layer, k indicates the kernel (or known as the filter), l stands for the quantity of output layer which is chosen by the quantity of parts, indicates the stride that the kernel shifts in each progression of computation, Mj stands for the jth feature map created by various kernel, b indicates the bias and f is an activation work generally characterized in the form of a sigmoid capacity appeared in equation (18). For neurons of a similar feature map, same weights

sharing and bias is accomplished and then albeit each output neuron has distinctive responsive fields. Here, training parameters are extraordinarily diminished [23,24]

After each convolutional layer, we make use of the enactment work. Initiation work is one which helps to delineate output to a lot of data sources and it creates the system structure to exhibit non-directivity. The underlying connection weights are initialized to the whole feature esteems given. Another input design is applied next and the output is registered as

$$y(n) = f(\sum_{i=1}^{i=N} w_i(n)x_i(n)) \qquad (19)$$

$$\text{Where } f(x) = \{+1 \quad if \ x \geq 0 \ -1 \quad if \ x < 0 (20)$$

Where n refers to the iteration index
Connection weights are updated in accordance with

$$w_i(n+1) = w_i(n) + \eta(d(n) - y(n)x_i(n), \quad i = 1,2,..N \qquad (21)$$

Where η stands for the gain factor

Then standard deviation is applied

$$\sigma = \sqrt{\frac{1}{n} f_i(x_i - \underline{x})^2} \qquad (22)$$

Into the proposed HCNN network, these weighted features were provided and it acquired more accurate churn classification results. Here we sub-sampled each of the feature maps from the convolution layer in the previous stage. The subsample technique is a weighted summation computation or using the maximum value in an area of size n × n of each feature map. The output of sub-sample layer is as given

$$x_i^l = f(b_j^l downsample(x_i^{l-1})) + b_j^l \qquad (23)$$

Where x^l refers to the output estimation of the lth sub-sample layer, downsample stands for the subsample function, β indicates the bias of the sub-sample capacity, f and b stand for the actuation work and the inclination separately. With the help of the sub-sample layer and over-fitting, the quantity of training parameters, filter noises are relieved and it is maintained by a strategic distance from in the system. The LSTM is joined with a convolution layer of CNN to enhance the FER execution.

LSTM repetitive neural system is a unique and significant design of RNN that exceeds expectations at recalling esteems for either long or brief spans of time. LSTM has another structure called a memory cell. It comprises four fundamental components: an input gate, a forget gate, an output gate and a neuron unit, a focal direct unit having a constant self- repetitive connection. The memory cell settles on the choices about on data to be stored, and when to permit perusing, composing and overlooking, through the three gateways, which open and close. These three doors are actualized utilizing the activation capacity to figure an incentive somewhere in the range of zero and one. The input gate controls the degree to which another worth streams into the memory cell. The output gate regulates the degree to which esteem in the memory cell is utilized to process the resultant output. The overlook gate regulates the degree to which the worth stays in the memory cell[25,26] .
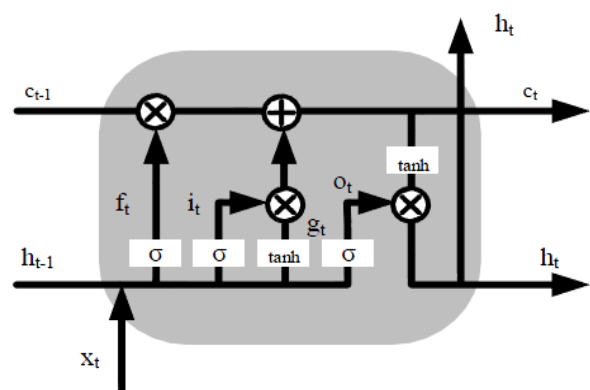

Fig 3 The design of LSTM

Classification Layer: After the facial appearance information experiences a few convolution layers and sub-sample layers, the size of the output feature maps ceaselessly diminishes. As to the classification layer, each feature map comprises of just a single neuron and turns into a feature vector. The vector and a classifier are completely associated. The time window filled in as a parameter, going from 0.5 s to 10 s (0.5, 1, 2, 5, and 10 s). At the point when the time window size will be ≤1 s, telecom information are fragmented without overlaps; in any case, information are sectioned every second with covers with past portions, e.g., a 9-second overlaps for a 10-second segment. Thus, the expense of multifaceted nature is decreased right now.

In order to train CNN models, all the sections completely inside the facial appearances states are utilized to show facial appearances class, while the portions 8 times longer than the facial appearances time frame are haphazardly chosen from inter-ictal states, characterized dependent on the partition from the outward appearances state by over 60 minutes, to demonstrate typical articulations class during CNN learning.

The trained CNN classified each portion of facial appearances information, i.e., every 0.5 second for the 0.5-second section and consistently for the remaining fragments, as sad, happy, anger, fear, neutral or surprise. To evaluate the presentation of the CNN classification, the facial appearances names by CNN were thought about. It is assessed the order execution either in leave-one-out testing or pairwise testing. In the testing stage, a model gets strained with CK+ database and then tested with CK+ information from the final subject. In the pairwise testing, a model is developed from a solitary subject's information and tested with CK+ database from each subject exclusively.

Steps in HCNN-LSTM

1. Begin
2. For all input feature ∈ CK+ dataset do
3. Transform the input into sub layers
4. Add weight factor using (19) (20) and (21)
5. Compute Standard deviation value using (22)
6. Detect customer's features using (17) and (18)
7. Do hidden layer process
8. Perform convolution process
9. Perform LSTM process
10. Extract more informative features using HCNN using (23)
11. Carry out training and testing process for CK+ dataset provided
12. Replicate the predefined facial expression information label for every feature according to the input dataset
13. Classify FER results with more accuracy
14. End

IV. EXPERIMENTAL RESULT

The CK+ database [27, 28] comprises 593 image sets acquired from 123 subjects in the age range of 18–30 years. The database comprises faces that show the seven emotions, such as, anger, fear, disgust, happiness, sorrow, astonishment, and neutral, for which 45,25, 59, 69, 28, 82, and 105 images are considered, correspondingly. The existing methods are considered as EBOSVM, Nearest Neighbour Classifier (NNC) and SVM method to evaluate CK+ databases alongside the proposed HCNN-LSTM algorithm. The performance metrics are precision, recall, f-measure and accuracy metrics.

TABLE: 1. Performance comparison result values

Methods	Precision	Sensitivity	Accuracy	F-measure
NNC	0.71	0.75	77	0.77
SVM	0.79	0.8	85	0.8

EBOSVM	0.86	0.84	92	0.84
HCNN-LSTM	0.92	0.89	94	0.9

A. Precision

The precision is computed as below:

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ positive} \qquad (24)$$

Precision can be vision in the form of a computation of accuracy or quality, while recall provides the measure of completeness or quantity. Generally, a greater precision represents that an algorithm retrieved considerably much appropriate results rather than inappropriate. With the task of classification, the precision for a class is given by the number of true positives can be segregated by the overall number of elements that are labeled to belong to the positive class.
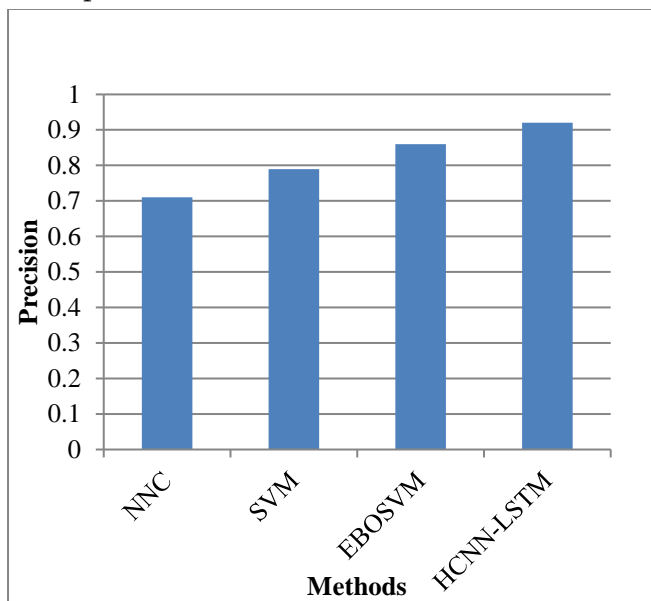
Fig 4 Precision

From Fig. 4, it very well may be seen that the correlation metric is assessed utilizing the already available and novel strategy as far as accuracy. Along the x-pivot, the strategies are plotted and the precision value is plotted along the y-axis. The current strategies are, for example, NNC,SVM and EBOSVM algorithms give lower precision though proposed HCNN-LSTM algorithm gives higher

precision for the given facial expression databases. In this way the outcome infers that the proposed HCNN-LSTM builds the facial expression acknowledgments are, for example, dismal, upbeat, outrage, dread, nonpartisan and shock articulations precisely for the given databases.

B. Recall

The recall value is computed as below:

$$\text{Recall} = \frac{True\ positive}{True\ positive + False\ negative}$$

$$(25)$$

The comparison graph is shown in Fig 5:

Recall is defined to be the number of appropriate documents recovered using a search is classified by the overall number of existing documents with relevance, whereas precision is determined as the number of relevant documents that the search has recovered is classified by the overall number of documents that the search has recovered.
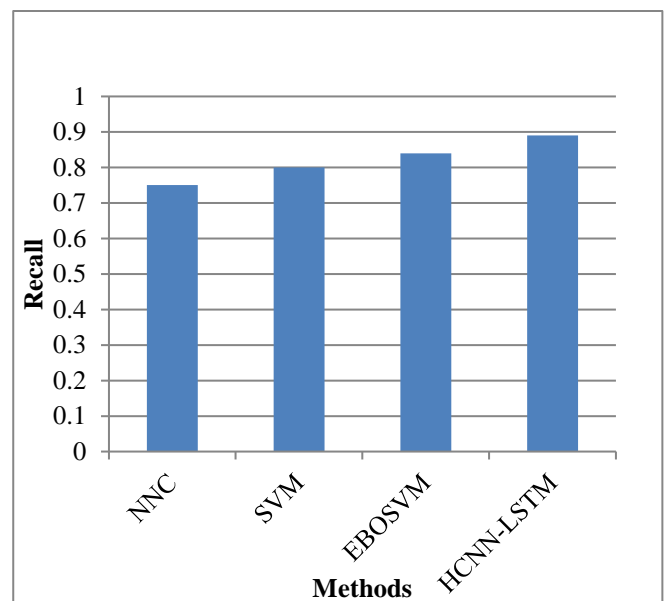

Fig 5 Recall

From the Fig 5, it very well may be seen that the examination metric is assessed utilizing the available and novel strategy as far as recall. Along the x-axis, the strategies are plotted and the recall value is plotted along the y-axis. The current strategies are, for example, NNC , SVM and EBOSVM algorithm

give lower recall though proposed HCNN-LSTM algorithm gives higher recall to the given facial appearance databases. Along these lines the outcome presumes that the proposed HCNN-LSTM expands the facial appearance acknowledgments are, for example, pitiful, upbeat, outrage, dread, impartial and shock articulations precisely for the given databases.

C. F-measure

$F1$-score is computed with the expressiongiven as below:

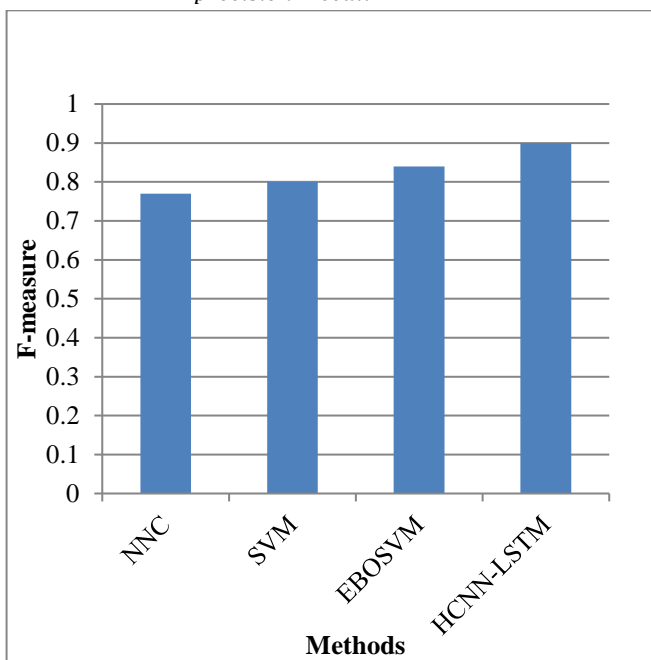$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \qquad (26)$$

Fig 6 F-measure

From the Fig 6, it very well may be seen that the correlation metric is assessed utilizing the available and the newly introduced strategy as far as F-measure. The techniques are plotted along the x-axis and the F-measure value is plotted along the y-axis. The current techniques are, for example, NNC,SVM and EBOSVM algorithm gives lower F-measure though proposed HCNN-LSTM gives higher F-measure to the given facial appearance databases. In this way the outcome presumes that the proposed HCNN-LSTM builds the outward appearance acknowledgments are, for example, tragic, glad, outrage, dread, impartial

and shock articulations precisely for the given databases.

D. Accuracy

Accuracy is determined as the overall rightness of the model and is computed as the total of original classification parameters $(T_p + T_n)$ that is segregated by the summation of the classification parameters $(T_p + T_n + F_p + F_n)$. The accuracy is computed as follows :

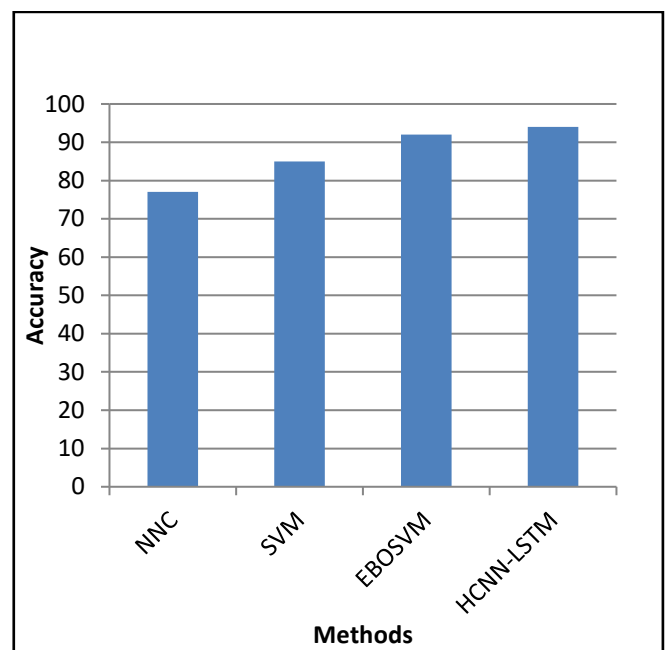$$Accuracy = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)} \qquad (27)$$

Fig 7 Accuracy

From the Fig 7, it very well may be seen that the examination metric is assessed utilizing the available and the newly introduced technique as far as accuracy. The techniques are plotted along the x-axis and the accuracy value is plotted along the y-axis. The current strategies are, for example, NNC , SVM and EBOSVM algorithm gives lower accuracy while proposed gives higher precision to the given facial expression databases. Accordingly, the outcome infers that the proposed HCNN-LSTM builds the facial appearance acknowledgments are, for example, miserable, glad, outrage, dread, impartial and shock articulations precisely for the given databases.

V.CONCLUSION

Driven by its important applications and theoretical interests, automatic facial expression recognition has attracted much attention in the last two decades. In this work, HCNN-LSTM model is proposed for logically accurate facial appearance affirmation rates. It improves the hypothesis execution enough and meanwhile inconceivably reduces the computational costs. There are six known classes of enthusiastic state which are shock, fear, fulfillment, sharpness, fair-minded and stun, related with their individual facial appearances which are detected in this work. Picture pre-preparing is done by using Histogram Equalization model which decreases the uproars for the given facial appearance information bases. At the present time, Jones calculation is used for face distinguishing proof and perceives hindered face. Filter with Dδ-LBP technique uses a fixed facial achievement limitation strategy and course undertaking of SIFT to eliminate lighting up and present invariant highlights. By then element assurance is performed using whale smoothing out estimation basically and HCNN-LSTM is applied for facial appearance affirmation. This system is solid in seeing the uncontrolled light pictures and it prompts awesome appearance planning precision for facial pictures. The result presumes that the proposed HCNN-LSTM method gives 94% accuracy which is better than the current EBOSVM, NNC and SVM methods.

V. REFERENCES

[1]. Pantic, Maja, and Ioannis Patras. "Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36.2 (2006): 433-449. Available from: DOI: 10.1109/TSMCB.2005.859075.

[2]. Ma, Liying, and KhashayarKhorasani. "Facial expression recognition using constructive feedforward neural networks." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34.3 (2004): 1588-1595. Available from: DOI: 10.1109/TSMCB.2004.825930

[3]. Liu, Wei-feng, and Yan-jiang Wang. "Expression feature extraction based on difference of local binary pattern histogram sequences." 2008 9th International Conference on Signal Processing. IEEE, 2008.pp (2082 – 2084). Available from: DOI: 10.1109/ICOSP.2008.4697555.

[4]. Zhao, Shuwen, et al. "Feature Selection Mechanism in CNNs for Facial Expression Recognition." BMVC. 2018.pp(35-41). Available from: DOI: 10.1109/EUVIP.2010.5699141.

[5]. Jia, Haipeng, et al. "Accelerating viola-jones face detection algorithm on gpus." 2012 IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems. IEEE, 2012.Available from:DOI: 10.1109/FCCM.2010.12.

[6]. Farfade, Sachin Sudhakar, Mohammad J. Saberian, and Li-Jia Li. "Multi-view face detection using deep convolutional neural networks." Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. ACM, 2015.pp.643–650. Available from:https://doi.org/10.1145/2671188.2749408.

[7]. El-Bakry, H. M., & Zhao, Q. (2004). Fast object/face detection using neural networks and fast Fourier transform. International Journal of Signal Processing, 1(3), 182-187. Available from:https://www.researchgate.net/publication/37432768.

[8]. Satiyan, M., and R. Nagarajan. "Recognition of facial expression using Haar-like feature extraction method." 2010 International Conference on Intelligent and Advanced Systems. IEEE, 2010. (pp. 1-4) . Available from:DOI: 10.1109/ICIAS.2010.5716228

[9]. Kirana, KartikaCandra, Slamet Wibawanto, and Heru Wahyu Herwanto. "Facial Emotion Recognition Based on Viola-Jones Algorithm in the Learning Environment." 2018 International Seminar on Application for Technology of Information and Communication. IEEE, 2018. (pp. 406-410). Available from: DOI: 10.1109/ISEMANTIC.2018.8549735.

[10]. Du, Cheng-Jin, and Da-Wen Sun. "Recent developments in the applications of image processing techniques for food quality evaluation." Trends in food science & technology 15.5 (2004): 230-249.Available from: https://doi.org/10.1016/j.tifs.2003.10.006.

[11]. Mian, Ajmal, Mohammed Bennamoun, and Robyn Owens. "An efficient multimodal 2D-3D hybrid approach to automatic face recognition." IEEE transactions on pattern analysis and machine intelligence 29.11 (2007): 1927-1943. Available from: DOI:10.1109/TPAMI.2007.1105.

[12]. Zheng, Yuefeng, et al. "A Novel Hybrid Algorithm for Feature Selection Based on Whale Optimization Algorithm." IEEE Access 7 (2018): 14908-14923. Available from: DOI: 10.1109/ACCESS.2018.2879848.

[13]. Sharawi, Marwa, Hossam M. Zawbaa, and EidEmary. "Feature selection approach based on whale optimization algorithm." 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2017. (pp. 163-168). Available from: DOI: 10.1109/ICACI.2017.7974502.

[14]. Ekweariri, Augustine Nnamdi, and Kamil Yurtkan. "Facial expression recognition using enhanced local binary patterns." 2017 9th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE, 2017. (pp. 43-47). Available from: DOI: 10.1109/CICN.2017.8319353.

[15]. Akbulut, Yaman, et al. "Deep learning based face liveness detection in videos." 2017 international artificial intelligence and data processing symposium (IDAP). IEEE, 2017. (pp. 1-4). Available from: DOI: 10.1109/IDAP.2017.8090202.

[16]. Sunaryo, Musthofa, and MochammadHariadi. "Preprocessing on Digital Image using Histogram Equalization: An Experiment Study on MRI Brain Image." International Journal of Computer Science and Information Technologies 7.4 (2016): 1723-1727. Available from: DOI: 10.1109/IDAP.2017.8090202.

[17]. Ren, Jianfeng, Nasser Kehtarnavaz, and Leonardo Estevez. "Real-time optimization of Viola-Jones face detection for mobile platforms." 2008 IEEE Dallas Circuits and Systems Workshop: System-on-Chip-Design, Applications, Integration, and Software. IEEE, 2008. (pp. 1-4). Available from: DOI: 10.1109/DCAS.2008.4695921.

[18]. Purandare, V., &Talele, K. T. (2014, April). Efficient heterogeneous face recognition using scale invariant feature transform. In 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA) (pp. 305-310). IEEE. Available from: DOI: 10.1109/CSCITA.2014.6839277.

[19]. Shen, Fang, Jing Liu, and Peng Wu. "Double δ-LBP: A Novel Feature Extraction Method for Facial Expression Recognition." Chinese Conference on Image and Graphics Technologies. Springer, Singapore, 2018.pp 370-379.Available from:DOI: 10.1109/CSCITA.2014.6839277.

[20]. Sharawi, Marwa, Hossam M. Zawbaa, and EidEmary. "Feature selection approach based on whale optimization algorithm." 2017 Ninth International Conference on Advanced Computational Intelligence (ICACI). IEEE, 2017. Available from:DOI: 10.1109/ICACI.2017.7974502.

[21]. Nwosu, Lucy, et al. "Deep convolutional neural network for facial expression recognition using facial parts." 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2017. Available from: DOI 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.213.

[22]. Gao, Tingwei, Yueting Chai, and Yi Liu. "Applying long short term memory neural networks for predicting stock closing price." 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2017. (pp. 575-578). Available from: DOI: 10.1109/ICSESS.2017.8342981.

[23]. Canedo, D. and Neves, A.J., 2019. Facial Expression Recognition Using Computer Vision: A Systematic Review. Applied Sciences, 9(21), p.4678. Available from: DOI: 10.3390/app9214678;

[24]. Wati, V., Kusrini, K. and Al Fatta, H., 2019, July. Real Time Face Expression Classification Using Convolutional Neural Network Algorithm. In 2019 International Conference on Information and Communications Technology (ICOIACT) (pp. 497-501). IEEE. Available from: doi: 10.1109/ICOIACT46704.2019.8938521.

[25]. Mahmood, A., Hussain, S., Iqbal, K. and Elkilani, W.S., 2019. Recognition of facial expressions under varying conditions using dual-feature fusion. Mathematical Problems in Engineering, 2019. Available from: doi:10.1155/2019/9185481.

[26]. Siddiqui, M. F., Siddique, W. A., Ahmedh, M., &Jumani, A. K. (2020). Face Detection and Recognition System for Enhancing Security Measures Using Artificial Intelligence System. INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY, 13(09), 1057-1064.Available from: DOI: 10.17485/ijst/2020/v013i09/149298.

[27]. Lucey, P., Cohn, J.F., Kanade, T., et al.: 'The extended Cohn–Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression'. IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2010, 36, (1), pp. 94–101. Available from: DOI: 10.1109/CVPRW.2010.5543262.

[28]. Siddiqi, Muhammad Hameed, et al. "Real Time Human Facial Expression Recognition System using Smartphone." INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND NETWORK SECURITY 17.10 (2017): 223-230.Available from: DOI: 10.1109/CVPRW.2014.25.

Cite this article as :