Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Deep learning framework for real-time face recognition using multimodal facial expression with feature optimization and hybrid classification

[1]Avishek Dey, [2]Dr.Satish Chander

[1]Research Scholar, Department of Computer Science Engineering,
BIT, Mesra, INDIA
[2]Assistant Professor,Department of Computer Science Engineering,
BIT, Mesra, INDIA

Abstract: Smile detection plays key role in recognizing human facial expressions, especially in inter-personal relations. In modern biometrics and security applications, face recognition is widely used. In real-life applications, automated face recognition has always been challenging because people express themselves differently. Real-world scenarios present many challenges, such as low-quality images, temporal variations, and disguises that alter facial characteristics. Face recognition has been accomplished using deep learning frameworks, but there are still several challenges to overcome, such as expression variations and lack of training data. In these methods, margins are used to enforce intra-class compactness and inter-class discrepancy to prevent overfitting. Multimodal facial expression detection is used for real-time face recognition to solve the aforementioned problems. First, we utilizes the Prewitt edge detector for eyes, nose, mouth, and brow detection, and design a deep transfer learning model to extract deep features from face images. Next, feature optimization is done by harpy eagle search optimization (HESO) to selects optimal best among multiple features which reduce data dimensionality problem. Then, we employ a semi-region based convolutional neural network (SR-CNN) to analyzes the facial expressions from multimodalities such as eyes (blinking), nose (shapes), mouth (smiling), and brow (shapes). The analyzed facial expressions and corresponding input images are fed into the benchmark SVM classifier for face recognition. The facial expression based face recognition framework is most suited for real-time scenario like video surveillance. Finally, we validate the performance our framework using benchmark multi-modal Cohn-Kanade and real-time 5671 person images datasets. The simulation results show that the proposed framework can effectively suppress data dimensionality issues and achieve high recognition rate compare to state-of-art frameworks.
Keywords: face recognition, facial expression, deep feature extraction, feature optimization, facial expression detection

1. Introduction

Lately, the rise of current innovative assets has sped up, making a requirement for exact client distinguishing proof frameworks to control admittance to innovation. Biometric validation frameworks are by a wide margin the most grounded choice accessible [1][2]. Biometrics is the study of laying out an individual's personality through semi-computerized or completely mechanized innovation in light of social qualities like voice and mark or potentially actual qualities like iris and finger impression. Client confirmation has become perhaps of the hardest test in the present advanced world. Conventional confirmation techniques [3] that depend on tokens, passwords, and individual ID numbers (PINs) are gradually becoming out of date. Obviously, tokens and PINs/passwords have restricted security and can without much of a stretch be lost, taken, neglected, speculated or split the difference. In such manner, as per a new report by the World Economic Forum, 80% of safety breaks in 2020 will be brought about by powerless or taken passwords. Furthermore, the report affirms that half of IT help work area spending is dedicated to secret word resets for organizations, with a typical expense of $1 million for representatives alone [char]. These deficiencies have made biometrics the focal point of the exploration local area throughout the last year. It utilizes exceptional conduct and organic elements, for example, fingerprints, hand math, nerves, faces, irises, voices, hands and DNA[5][6] to rapidly and safely recognize individuals.

A biometric framework regularly comprises of four modules: sensor, include extraction, coordinating and choice modules. There are two kinds of biometric frameworks: Unimodal [7] and multimodal [8]. Uni

modular frameworks utilize a solitary biometric component to distinguish a client. Albeit the coordinated framework [9] has shown to be solid and better than traditional strategies utilized previously, it has constraints. These remember clamor for acknowledged information, non-all inclusiveness issues, weakness to caricaturing assaults, and between and intra-class similitude issues. Fundamentally, multimodal biometric frameworks [10] require various highlights to distinguish the client. Because of their capacity to beat issues experienced in comparative biometric frameworks, they are generally utilized in true applications. A multimodal biometric framework can utilize the data accessible in one module of a biometric framework to consolidate various elements. There are four kinds of combination: sensor-level combination, include level combination, score-level combination and result-level combination [11]. The benefits of multimodal biometrics over Unimodal frameworks make them a more appealing and secure ID strategy. Contrasted with single biometric innovation, multimodal biometric innovation enjoys benefits like higher verification exactness, higher security, and more extensive relevance. For sure, multimodal biometrics might incorporate numerous single biometric screening cycles and host all classes of single biometric elements to give a system and embrace different execution procedures. As of late, multimodal biometrics has arisen as a significant testing strategy while calibrating biometric exactness [12] . For instance, utilizing various sensors to distinguish different biometric highlights, utilizing biometric cycles to assemble data, and utilizing extra restrictive specialized strategies to acquire novel cutthroat scores, there is extension for development in multimodal biometrics utilized by different classes today. Biometric data has turned into an alluring technique that is broadly utilized with a definitive objective of individual verification and control.

Recently, many researchers are exploring the combination of two or more biometrics to improve identity and virtual security [13]-[20]. Multimodal biometrics based on integrated characteristics of camera captures are becoming increasingly popular due to their convenience and low acquisition costs. An information base of 250 virtual areas created from Face FRGC and Palmprint PolyU data set was contrasted and different combination plans at match score level and element levels. Molecule Swarm Optimization (PSO) [13] altogether diminishes the quantity of elements while keeping up with a similar degree of execution. The meager coding blunder rate (SCER) [14] is utilized to quantify the certainty contrast among face and ear questions. A versatile element weighting plan is utilized to decrease the adverse consequences of untrustworthy strategies progressively. An improvement based block-mode dynamic learning technique [15] is utilized for face acknowledgment. The adaptability of the system is stressed by its capacity to consolidate accessible data. Multimodal conduct biometric frameworks join ECG and sound signs for verification [16]. Gabor-Wigner Transform (GWT) [17] is utilized to remove include vectors from different biometric strategies. An advancement method, molecule swarm improvement, is utilized to choose prevailing aspects from the element vector. A multimodal biometric framework in view of contactless multispectral finger pictures is planned to beat the restrictions of Unimodal biometrics [18]. A combination method called no fixed element combination, in which the design of the interleave grid is built utilizing neighborhood highlights extricated from two modalities, face picture and palm print picture [19]. Fluffy least talk improvement likewise utilizes Euclidean distance (EED) for brain based recognition frameworks [20]. The info picture is passed to the area of interest (ROI) and the following stage is to acquire the suitable picture.

Our contributions. Multimodal facial expression detection is used for real-time face recognition to solve the aforementioned problems. The primary commitments of our proposed work are given as follows:

1. Prewitt edge detector for eyes, nose, mouth, and brow detection, and design a deep transfer learning model to extract deep features from face images.
2. Feature optimization is done by harpy eagle search optimization (HESO) to selects optimal best among multiple features which reduce data dimensionality problem.
3. The analyzed facial expressions and corresponding input images are fed into the benchmark SVM classifier for face recognition.
4. A semi-region based convolutional neural network (SR-CNN) is used to analyze the facial expressions from multimodalities such as eyes (blinking), nose (shapes), mouth (smiling), and brow (shapes).
5. To validate the performance our framework using benchmark multi-modal Cohn-Kanade and real-time 5671 person images datasets. The simulation results show that the proposed framework can

effectively suppress data dimensionality issues and achieve high recognition rate compare to state-of-art frameworks.

The remaining tests are performed as follows. Recent work related to our collaboration is presented in Section 2. The problem statement and the proposed working model are described in Area 3. An itemized portrayal of the proposed multi-biometric framework is introduced in Section 4 and exploratory outcomes are introduced. , Segment 5. At long last, complete Area 6.

2. Related works

Abozaide et al [21] have proposed a multidimensional biometric authentication system capable of skilled human identification through face detection, speech recognition, brain segmentation and statistical data. Extract individual faces and principal component analysis (PCA) face recognition features using different extraction methods and compare the results. Based on the combined scores, the score combination produced the lowest EER and was considered an appropriate multivariate combination. Regid et al. [22] proposed a multimodal biometric framework in the form of ECG-ear-iris biometrics at the functional level. Pre-processing strategies, for example, normalization and division are utilized for ECG, ear and iris biometrics. We utilized nearby form descriptors, 1D neighborhood paired models (1D-LBP), Shift-1D-LBP and 1D-MR-LBP to extract features from ECG signals and analyze ear and iris images. KNN and RBF are used to classify anonymous users as real or fake. This configuration was approved utilizing the ID-ECG, USTB1, USTB2, AMI EAR and CASIA v1 iris information bases. A 100 percent exact acknowledgment rate (CRR) is accomplished with a blunder rate (EER) of 0.5%. Kugura et al [23] presented a flexible time acquisition method to capture the neural structure of the finger using a variety of inexpensive cameras. The cognitive system relies on the use of brain circuits and repetitive interactions. According to the results obtained, the framework guarantees up to 99.13% validation for the collected data set. Toygar et al [24] have presented a deep learning-based brain convolution (CNN) for two models using layer combination to select hand, back and wrist biometric features from brain images. Considering the accuracy and computation time of FYO dataset, the results show consistent results in different datasets such as Dongji uncorrelated neural dataset, VERA, PUT, Badawi and Bosphorus datasets. Using 3-stream biometrics improves overall performance and can be used to secure against multiple attacks. Omara et al [25] have studied collaborative learning distance estimation and coordinated non-cyclic diagram SVM model (LDM-DAGSVM) for two novel approaches to heterogeneous biometric system, learning distance measurement and query problem management for DAGSVM. Preliminary results on graph datasets, for example, the AR face and Wonderment ear data sets show that the framework accomplishes tremendous contrasts in execution and independent methods and outperforms a variety of methods. The multimodal model has an accuracy of about 99.85% when extracting face and ear images 5 times.

Casanovan et al. [26] have proposed two wellsprings of conduct biometric information are examined for the improvement of this web client recognizable proof model, contact elements and the qualities removed from the periocular region connected with the understudies, flickers and obsessions. The outcomes acquired show the commitment of these two unique biometric qualities and, most importantly, of their combination. The combination approach permits getting acquiring exactness higher than that of individual biometrics, arriving at a precision of more than 92%. Alkeem et al. [27] have proposed a hearty and dependable distinguishing proof method grounded on multimodal biometrics, which uses profound figuring out how to join unique mark, ECG and facial picture information, especially valuable for ID and orientation grouping purposes. The multimodel approach allows the model to handle multiple data spaces by eliminating the need to independently prepare each method and further improves the ability to predict these functions across spatial connectivity. Removing complexity from a project while working on different tasks helps organize others and leads to better overall performance. This approach integrates multimodality using a combination of labeling levels and scoring levels. This model outperforms the reference data set, element-level fusion and other fusion techniques used.

Most traditional fingerprint authentication processes are designed by hand and require prior knowledge. Therefore, it is important to research and develop a suitable feature expression and fusion method for multi-biometric authentication. A joint discriminative feature learning (JDFL) framework [28] is used to combine finger wing (FV) and finger joint axis (FKP) models for multimodal finger recognition. Saeed [29] investigated the use of microscopic imaging as soft biometrics. The system is based on a combination of traditional facial

features for modeling facial expressions and soft biometrics for modeling isolated emotions in visual scenes. Experiments are performed on a commonly used micro-representational database, and the combination of soft biometric features with face recognition increases the recognition rate by ~5% at the decision stage. Huang et al. [30] have presented multi-view discriminant assessment approach in the possibility of test assortment for ECG biometric affirmation. This strategy is used for delivering alternate points of view by using single lead ECG signal. A multi-sees learning framework is applied to consider test assortment to create a more discriminative subspace. A denoising impediment is used to get comfortable with the associations between different viewpoints, which make a consistent depiction against ECG disturbance. Kavipriya et al. [31] have proposed a procedure to see ear with online flighty heading. After evaluation, an ear image is generated according to the structure and a raw image frame with corresponding edges appears as output. In any case, a person's throat is open, because unauthorized persons act accordingly. This improves accuracy and reduces the size of the segment vector. This frame is useful if the picture has hair and nails. Table 1 shows the survey results for selected wells in previous work.

Table 1: Summary of research gaps

Ref.	Methodology	Technique used	Modalities	Parameter improved	Research Gaps
[21]	Face reorganization	GMM, ANN and SVM	Eyes, nose, mouth and brow	Accuracy, precision, Recall and F-measure	It suffers considerably in unconstrained environments
[22]	Facial expression detection	1D-LBP, Shifted-1D-LBP and 1D-MR-LBP	Eyes and nose	Accuracy, precision, Recall and F-measure	Cannot fully utilize the sample-based diversity information
[23]	Facial expression detection	CNN and RNN	Nose	Accuracy, precision, Recall and F-measure	Huge feature dimensions and time complexity
[24]	Face reorganization	CNN	Mouth and brow	Accuracy, precision, Recall and F-measure	Image and audio data are highly dimensional
[25]	Multi-modal facial expression	LDM-DAGSVM	Eyes, nose, mouth and brow	Accuracy, precision, Recall and F-measure	Not efficient and sufficient to predict the right subject
[26]	Facial expression detection	CCA and PCA	Eyes, nose, mouth and brow	Accuracy, precision, Recall and F-measure	Complex design architecture
[27]	Face reorganization	CNN-ResNet50	Eyes and mouth	Recall and F-measure	Out-of-plane rotations are still an unsolved problem
[28]	Face reorganization	Joint discriminative feature learning (JDFL)	Eyes and nose	Recall and F-measure	Every separate biometric still faces the problems of improving accuracy
[29]	Face reorganization	Support vector machine (SVM)	Mouth and brow	Accuracy	Suffering from spoofing attacks
[30]	Facial expression detection	Multi-view construction	Eyes, nose, mouth and brow	Recall and F-measure	Not suitable for high density data's

3. Problem methodology and System model

3.1 Research gap

Feeling acknowledgment in view of look is a difficult exploration subject and has drawn in a lot of consideration in the beyond couple of years. A technique, using multi-modular system to remove feeling

highlights from look pictures [31]. The essential thought is to join the low-level exact component and the undeniable level self-learning highlight into a multi-modular element. The 2-layered direction of facial central issues are extricated as low-level experimental component and the significant level self-learning highlight are removed by the Convolutional Brain Organizations (CNNs). To decrease the quantity of free boundaries of CNNs, little channels are used for all convolutional layers. Inferable from different little channels are likeness an enormous channel, which can decrease the quantity of boundaries to successfully learn. Furthermore, name safeguarding change is utilized to amplify the dataset falsely, to address the over-fitting and information awkwardness of profound brain organizations. Then, at that point, two sorts of modular highlights are intertwined straightly to frame the look include. Broad examinations are assessed on the lengthy Cohn-Kanade (CK+) Dataset.

According to many studies in the literature [21-32], disadvantages of single-module biometric systems include poor recognition mainly due to intra-class variability and intra-class similarity and disadvantages of face recognition and detection. It increases the complexity of biometrics. Systems with two or more sensors are expensive and difficult to use with large amounts of biometric data. Therefore, among many security issues in our society, identifying people with high accuracy and low system complexity is important. Iris and fingerprint biometrics are simpler, more accurate and more reliable than other options. Combining these features makes it a particularly promising solution to today's authentication challenges. Additionally, eyes and brows are more secure than other facial metrics. However, iris biomaterials have higher specificity, stability and penetration resistance than fingerprint biomaterials. However, since the conventional fusion method uses the same weight for each organism, the optimal error rate is not accurate. The false acceptance rate indicates how often legitimate users misidentify fraudsters and the false rejection rate indicates how often legitimate users misidentify fraudsters as genuine users. In a real system, FAR and FRR should be as close to zero as possible, but they are not.

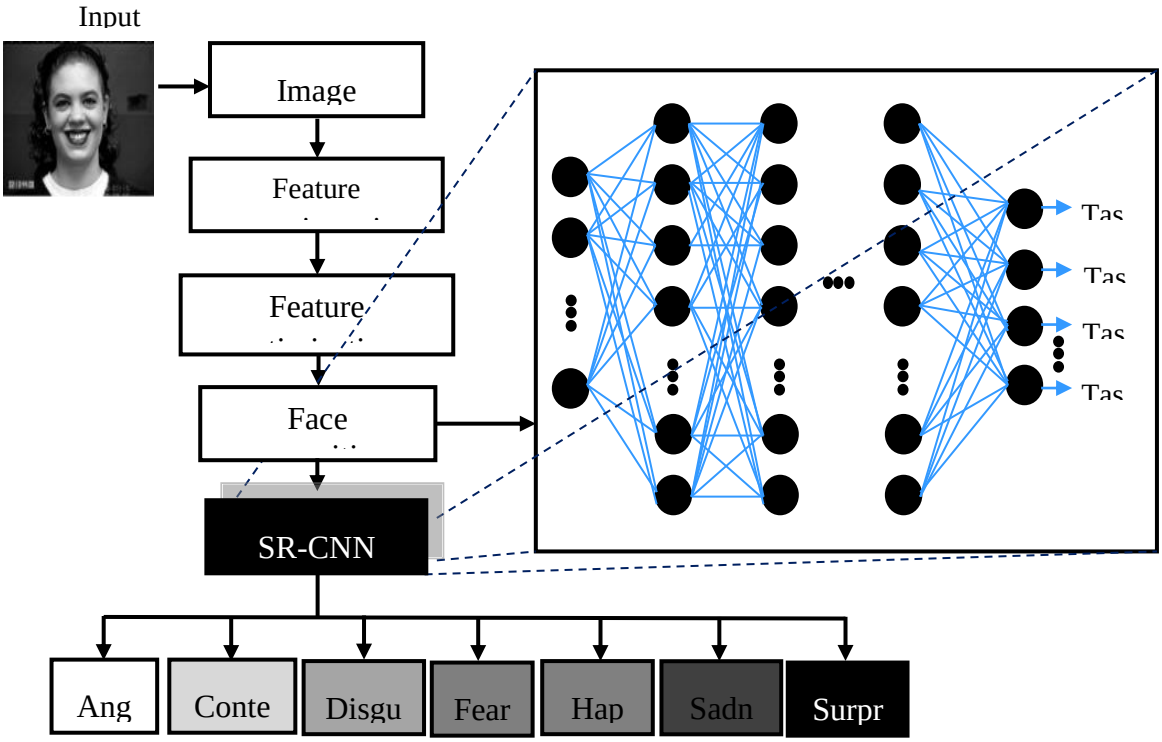

Fig 1: overall system architecture of proposed work

3.2 System architecture of proposed work

Here, we describe the working process of proposed real-time face recognition and multimodal facial expression detection using hybrid deep learning technique. The system architecture of our proposed work is shown in Fig. 1. Prewitt edge detector for eyes, nose, mouth, and brow detection, and design a deep transfer

learning model to extract deep features from face images. Based on that gathered features, we fed those features into the benchmark SVM classifier for face recognition. HESO algorithm is utilized to chooses ideal best among various elements which reduce data dimensionality problem. SR-CNN classifier is then used to analyze the facial expressions from multimodalities such as eyes (blinking), nose (shapes), mouth (smiling), and brow (shapes).

4. Proposed methodology

4.1 Deep feature extraction

First, we utilizes the Prewitt edge detector for eyes, nose, mouth, and brow detection, and design a deep transfer learning model to extract deep features from face images. Highlight examination is a quantitative technique to measure and recognize primary irregularities in various cerebrum tissues. As the tissues present in the mind are truly challenging to arrange utilizing the shape or the dark level powers, surface component extraction is vital for additional characterization. Surface is that natural property of all surfaces that depicts visual example that contains significant data about the primary game plan of the surface likewise described by the spatial dissemination of dark levels in an area. The spatial elements are removed by utilizing the curvelet deterioration and furthermore Prewitt edge locator. While handling a picture, a variety histogram demonstrates the quantity of shots of a specific tone. Variety guides can be made in any variety space. For the most part, variety spaces are separated into relative number ranges, each with a similar variety esteem. It likewise shows a smooth capability that determines the number of pixels and the given color range.

The ideal solution is denoted by y∗. Here g is equal to 0.5. To increase the population diversity, the search efficiency as follows:

$$Y_{NEW} = y_{s1}^{T} + G(Y_{s2}^{T} - Y_{s3}^{T}) \tag{1}$$

where g is a variance weighting coefficient uniformly distributed over random numbers. In the second case, when a butterfly fails to detect the scent of another butterfly in its search area, a stochastic evolution called the local search period occurs.

$$qg_{j} = dJ^{b} \tag{2}$$

Butterfly consists of an eagle and a layer vector that can be updated when received.

$$y_{j}^{(T+1)} = y_{j}^{T} + G_{j}^{(T+1)} \tag{3}$$

The IBO algorithm consists of two main phases: a global search phase and a local search phase. These steps are decomposed into Eq. (6) and (7) are:

$$G_{j}^{(T+1)} = (s^{2} \times f^{*} - y_{j}^{T}) \times qg_{j} \tag{4}$$

$$G_{j}^{(T+1)} = (s^{2} \times y_{i}^{T} - y_{z}^{T}) \times qg_{j} \tag{5}$$

where are the I and Z butterflies of the association. Local is a random gate. The probability of changing q is used to change from an existing search to a local search. This confirms the function of Eq. (eight):

$$R(G_{j}^{z}(T)) = \frac{1}{1+e^{-G_{j}^{z}(T)}} \tag{6}$$

where $G_{j}^{z}(T)$) is the scent of the jth butterfly of size z in T . Use the random threshold described in Eq. (9) Find the binary precision for the sigmoid process. A function maps an infinite number of inputs to a finite number of outputs :

$$Y_{j}^{z}(T+1) = \begin{cases} 0, & if \;\; Rand < R(G_{j}^{z}(T)) \\ 1, & otherwise \end{cases} \tag{7}$$

Repeat this for the second stage and record the position and scent of the Jth butterfly. Feature selection can also be viewed as a multi-optimization problem. The best solution is low mileage with best-in-class accuracy. Therefore, the execution function of this method can be seen as an expression. (ten):

$$fitness = \alpha\gamma_S(C) + \beta\frac{|S|}{|M|} \tag{8}$$

Statistical features involve the circulation of dim qualities in a spatial space by assessing the neighborhood qualities of every pixel in the given picture and deducting the arrangement of factual boundaries utilizing nearby qualities. The quantity of pixels in a pixel is gotten by adding the worth of 'j' to the worth of 'I' close to the pixel worth, and afterward partitioning the entire lattice by the all out number of such correlations. Each record has a pixel value close to the "i" pixel value, which is a fourteen statistical term defined using a parallel event matrix to define a structure:

$$angularSecond\ moment = \sum_j\sum_i q(j,i)^2 \tag{9}$$

$$contrast = \sum_{N=0}^{n_h-1} n^2 \left\{ \sum_{j=1}^{n_h}\sum_{i=1}^{n_h} q(j,i)^2 \right\}, |j-i| = N \tag{10}$$

$$Correlation = \frac{\sum_j\sum_i (ji)q(j,i) - \mu_y\mu_x}{\sigma_y\sigma_x} \tag{11}$$

$$Variance = \sum_j\sum_i (j-\mu)^2 q(i,j) \tag{12}$$

$$Inverse\ diference\ method = \sum_j\sum_i \frac{1}{1+(j-i)^2} q(j,i) \tag{13}$$

$$sum\ average = \sum_{j=2}^{2n_h} jq_{y+x}(j) \tag{14}$$

$$sum\ Variance = \sum_{j-2}^{2n_h} (j-F_t)^{2q_{y+x}}(j) \tag{15}$$

$$sum\ entropy = \sum_{j=2}^{2n_h} q_{y+x}(j)\log\{q_{y+x}(j)\} = F_t \tag{16}$$

$$entropy = -\sum_j\sum_i q(i,j)\log(q(j,i)) \tag{17}$$

$$Diff\ Variance = \sum_{j=0}^{n_h-1} j^2 q_{y+x}(j) \tag{18}$$

$$Diff\ entropy = -\sum_{j=0}^{n_h-1} q_{y+x}(j)\log\{q_{y-x}(j)\} \tag{19}$$

$$Info.Measure\ of\ Correlation1 = \frac{GYX - GYX1}{Max\{GY, Gx\}} \tag{20}$$

$$Info.Measure\ of\ Correlation2 = (1-\exp[2(GYX2 - GYX)])^{1/2} \tag{21}$$

$$MaxcorrelationCoeff = Squareroot\ of\ the\ Second\ largest\ Eigenvalues \tag{22}$$

Neighborhood paired designs (LBP) are visual portrayals used to group PC vision. The blend of LBP and histogram of arranged angles (Hoard) descriptors has been displayed to further develop discovery effectiveness essentially. Each pixel inside the cell is contrasted and its eight neighbors. State "0 on the off chance that the middle pixel esteem is more noteworthy than the neighbor pixel esteem, in any case, express "1" Track down the histogram of information picture and saw as 256-layered include vector.

4.2 Feature optimization

Next, feature optimization is done by harpy eagle search optimization (HESO) to selects optimal best among multiple features which reduce data dimensionality problem. To justify the parallelism of each hunting phase, the HESO algorithm reproduces the hunting behavior of the peregrine falcon. It can be alienated into three parts: choosing exploration area, viewing the selected search area, and diving. During the evaluation phase, vixen raptors will search for and select the most suitable area for hunting within the selected search area. To install the knock hack, do the following:

$$Pnew, x = P_{best} + \alpha * U(P_{mean} - P_x) \tag{23}$$

α is a boundary that controls the change in magnitude and takes a worth among 1.5 and 2, while r is an arbitrary amount that receipts a worth among 0 and 1. In the assortment phase, Eagle selects zones based on available zones. Eagles will randomly select another search area, but choose one of the previous search areas. Eagles indicate P_{best} the currently selected search position based on the best position found in the previous search. Birds randomly visit all centers near the recently selected search area. At the same time, the eagle indicates that all information has been used previously. Copying newly learned data with α does not set funny birds in stone. During the search phase, the harpy eagle searches for its prey and moves in any direction to speed up the hunt. Suitable cleaning conditions are calculated as follows.

$$P_{x,new} = P_x + b(x) * (P_x - P_{x+1}) + A(x) * (P_x - P_{mean}) \tag{24}$$

$$a(x) = \frac{au(x)}{\max(|\ au\ |)}, b(x) = \frac{au(x)}{\max(|\ au\ |)} \tag{25}$$

$$au(x) = u(x) * \sin(\theta(x)),\ bu(x) = u(x) * \cos(\theta(x)) \tag{26}$$

$$\theta(x) = b * \pi * uand \tag{27}$$

$$u(x) = \theta(x) + u * uand \tag{28}$$

Here, a is between 5 and 10, to estimate the midpoint, focus on the center opinion, and R takes values of 0.5 and 2 to determine the amount of exploration cycles. During the push phase, the hawk moves from its best position in the pursuit zone toward the hunting target. Below all foci is the shape of the corresponding foci:

$$P_{x,new} = uand * P_{best} + b1(x) * (P_x - c1 * P_{mean}) + x1(x) * (P_x - d2 * P_{best}) \tag{29}$$

$$a1(x) = \frac{au(x)}{\max(|\ au\ |)}, b1(x) = \frac{au(x)}{\max(|\ au\ |)} \tag{30}$$

$$au(x) = u(x) * \sin(\theta(x)), \quad bu(x) = u(x) * \cos(\theta(x)) \tag{31}$$

$$\theta(x) = b * \pi * uand \tag{32}$$

$$u(x) = \theta(x) \tag{33}$$

where $c1, c2 \in [1, 2]$ hay processing involves various structures. We use polar positions to address these bird migrations. Also, we calculate the ideal opinion by increasing the alteration among the stem present and the focal point on the x-axis by the change among the following present and the ideal opinion on the y-hub. In the previous Segment, the main components of BES are presented, including the prototyping, development and refinement phases. The operating period is the number of cycles set to zero. Note that the data are generated arbitrarily for each setting of P . At that time the usage of each atom is evaluated. For each solution in group P, perform the following steps. Lines 4-12 are used to find the region that gives the best solution, estimate the new region, and divide the region using a spiral motion. Find and select the field that generates the random number in both cases. The solution moves to the center of the next point. After three steps, the loop counts k steps from 2 to 31. Algorithm 1 illustrates the steps involved in latent feature optimization using the HESO algorithm.

Algorithm 1: Feature optimization using HESO algorithm

Input : features, threshold condition

Output : optimal best feature selection	
1	Create an initial set of points
2	Calculate the fitness values of the starting points
3	Define boundary conditions $Pnew, x = P_{best} + \alpha * U(P_{mean} - P_x)$
4	Because (each population point I)
5	$P_{best} = P_{new}$
6	All points lead to corresponding points $P_{x,new} = rand * P_{best} + b1(x) * (P_x - c1 * P_{mean}) + a1(x) * (P_x - c2 * P_{best})$
7	End if
8	Set $K := K + 1$
9	Update the final values of b
10	End

4.3 Face recognition and detection

Optimal layered kernel-based support vector machine (OK-SVM) classifier is used for face recognition and detection. A kernel function is a method that takes data as input and converts it into the format required for data processing. An optimal kernel is used because of a set of arithmetic operations used in a SVM, which provides a window for controlling information. In this manner, the piece capability commonly changes the preparation informational index so the subsequent nonlinear surface turns into a direct condition in higher layered spaces. The standard of SVM order is to segment the information into a preparation set and test/forecast set. Every occasion in the preparation set has $\{(y_1, x1), (y_2, x2).....(y_m, x_m)\}$ an objective worth and a few credits, which can be highlights and indicator factors. Let Y and X be the info and result informational collections, and let the preparation set be general, and try to find the corresponding x ∈ X given the value of y ∈ Y seen earlier:

$$X = F(y, \alpha) \tag{34}$$

where α are the kernel function parameters, which should be calibrated for more exact picture order results. An ideal portion SVM (alright SVM) classifier can perform picking up utilizing various kinds of piece capabilities like straight, polynomial and spiral premise capability for effective picture arrangement. In alright SVM, all hyper planes are defined by an ordinary vector w and a constant b perpendicular to the hyper plane can be expressed as

$$F(y) = W.Y + a \tag{35}$$

Where h is the expressible ideal hyper plane and h1 and h2 are the expressible planes.

$$h : y_j.w + a = 0 \tag{36}$$

$$h1 : y_j.w + a = -1 \tag{37}$$

$$h2 : y_j.w + a = +1 \tag{38}$$

A situation where a straight part model isn't reasonable, a multinomial portion technique ought to be utilized, which can successfully order more complicated information. The statement of the polynomial portion strategy is as per the following:

$$F(Y) = w.\Phi(y) + a \tag{39}$$

where Φ(y) is a planning capability in a high-layered space. A string is opened if it is a prefix of other long strings or if it is a character. For the remaining strings, the frequency of occurrence in the training data set is calculated, representing their F_n and F_b frequency in normal and unusual layers, respectively. Next is the relative ratio

$$q_F = \frac{F_n}{F_b} \tag{40}$$

Here it is used as a code to identify malicious strings. Here ε percent strings are selected for dictionary construction. Then the resulting constant length is the property vector

$$v = \frac{1}{m}\sum_{m=1}^{m} z_M \tag{41}$$

$$M = N - K + 1 \tag{42}$$

The size of the sliding window K is predetermined. The K value plays an important role in model discovery as it represents the problem of feature extraction and model training. Let y be the input for each y and let y1 be the support vector. The core can be defined as

$$k(y, y1) = sum(y * y1) \tag{43}$$

The measure of the istance between the information and the help vectors is known as the bit. A straight mix of information sources is known as a direct bit. Part types, for example, polynomial and outspread change the contribution to its higher aspects. The polynomial ideal bit written as follows:

$$k(y, y1) = 1 + sum(y * -y1)^D \tag{44}$$

$$k(y, y1) = \exp(-gamma * sum((y - y1^2)) \tag{45}$$

The default esteem embraced for gamma is 0.1, yet the gamma worth can be somewhere in the range of 0 and 1. It very well may be utilized to make complex areas in two-layered space. The working process of Face recognition and detection using OK-SVM illustrates in Algorithm 2.

Algorithm 2: Face recognition and detection using OK-SVM classifier

Input: optimal best features
Output: face detection

1. Initialize the random population
2. Generate a SVM model $X = F(y, \alpha)$
3. The polynomial kernel method expression $F(Y) = w.\Phi(y) + a$
4. If j=0 and i=1
5. Then the achieved fixed-length feature vector $v = \frac{1}{m}\sum_{m=1}^{m} z_M$
6. Define radial kernel and update solution
$$k(y, y1) = \exp(-gamma * sum((y - y1^2))$$
7. Update the final best solution
8. End

4.4 Facial expression classification

In this subsection, semi-region based convolutional neural network (SR-CNN) is used to analyze the facial expressions from the multimodalities. First, we formulate precise theory behind SR-CNN an input $w \times g$, grey scale double, $\mathrm{l} \in \mathrm{R}^{w \times g}$, signified as follows:

$$l = \{y(M, N) \mid 1 \le M \le w, 1 \le N \le G\} \tag{46}$$

Given $y(M, N)$ the pixel intensity, M, N, and $w_K \times g_K$ filters (or kernels), k, using the filter, k, convolution Y, produces a map of image features. A filter K with a step of l and a zero-padding T_K value slides over the image.

$$(l \otimes K)_{M,N} = \sum_{u=-w_K}^{w_K} \sum_{v=-g_K}^{g_K} K_{u,v} I_{M+u,i+v} \tag{47}$$

At each convolutional layer, denoted by *l*, a convolution function and an additive bias are applied to the input, indexed $F \in \{1,...,F(l)\}$. So, to map the j-th function to the output of the jth layer, $x_j^{(l)}$, From the output of the previous layer, $x_j^{(l)}$, by:

$$x_j^{(l)} = \phi\left(A_j^{(l)} + \sum_{i=1}^{F^{(l-1)}} k_{j,i}^{(l)} * y_j^{(l-1)} \right) \tag{48}$$

where F is the rectified linear unit (ReLU) activation function $A_j^{(l)}$ and the dependence matrix $k_{j,i}^{(l)}$ of the magnitude filter $2W_K + 1 \times 2g_K + 1$. Thus, the output elements of the layer, l, $x_j^{(l)}$ for the function map, i, to position M, N

$$(x_j^{(l)})_{M,N} = \phi\left(\left(A_j^{(l)}\right)_{M,N} + \sum_{i=1}^{F^{(l-1)}} \left(k_{j,i}^{(l)} \otimes y_j^{(l-1)}\right)_{M,N} \right) \tag{49}$$

$$= \phi\left(\left(A_j^{(l)}\right)_{M,N} + \sum_{i=1}^{F^{(l-1)}} \sum_{u=-w_K^{(l)}}^{F^{(l-1)}} \sum_{i=1}^{F^{(l-1)}} \left(k_{j,i}^{(l)}\right)_{M,N} \left(y_j^{(l-1)}\right)_{M+u,N+v} \right) \tag{50}$$

where the $x_j^{(l)}$ pooling process with steps , T_q, and $w_q \times g_q$ the pooling window.

$$Q(x_j^{(l)})_{M,N} = \max\left(x_j^{(l)}\right)_{M,N} \tag{51}$$

Whenever a dimension has a maximum pooling window, the maximum function is applied

$$DQ(x_j^{(l)})_{M,N} = \left((w - w_K)/T_q + 1\right) \times \left((G - g_K)/T_q + 1\right) \tag{52}$$

In the fully connected layer of the SR-CNN, the result of the past pooling layer is ventured into a segment vector and turns into the contribution of this layer. All neurons in the past layer are associated with all neurons in the following layer in the FC layer. $F_1^{(l)}$ the number of size feature maps, $F_2^{(l)} \times F_3^{(l)}$ and the map of feature i in layer l is computed as:

$$(x_j^{(l)})_{M,N} = \phi\left(\sum_{q=1}^{F_1^{(l-1)}} \sum_{p=1}^{F_2^{(l-1)}} \sum_{r=1}^{F_3^{(l-1)}} w_{j,q,p,r}^{(l)} \left(y_j^{(l-1)}\right)_{p,r} \right) \tag{53}$$

In the $w_{j,q,p,r}^{(l)}$ feature map, layer j contains the weights M, N that associate the entity at position (p,r) in the feature map, $(l-1)$ layer q.

$$\hat{y}_j = \varphi(y_j) \sum_{K=1}^{k} F_K(y_j), \quad F_K \in f \tag{54}$$

To improve the final result, control is used as follows:

$$L(\varphi) = \sum_j l(\hat{x}_j, x_j) + \sum_K \Omega(F_K) \tag{55}$$

which estimates the distinction of target and determined results are x_j and $\hat{x}_j$. Ω characterizes the intricacy of the model as follows:

$$\Omega(F) = \gamma S + 1/2\lambda \sum_{i=1}^{S} w_i^2 \tag{56}$$

Standard terms have been deleted to create a simplified scope as follows:

$$\tilde{L}^{(s)} \approx \sum_{j=1}^{N}\left[h_j F_j(y_j) 1/2 h_j F_j^2(y_j)\right]+\Omega(F_s) \tag{57}$$

where $i_i = \{j \mid p(y_j) = i\}$ signify the occurrence set of leaf s, and

$$h_j = \frac{\partial l\left(\hat{y}_i^{(s-1)}, x_j\right)}{\partial \hat{y}_i^{(s-1)}} \tag{58}$$

$$h_j = \frac{\partial^2 l\left(\hat{y}_i^{(s-1)}, x_j\right)}{\partial\left(\hat{y}_i^{(s-1)}\right)^2} \tag{59}$$

where $p(y_j)$ is calculate as follows:

$$w_i^* = -\frac{\sum_{j\in i_j} h_j}{\sum_{j\in i_j} h_j + \lambda} \tag{60}$$

$$\tilde{L}^{(s)}(p) = -\frac{1}{2}\sum_{i=1}^{S}\frac{(\sum_{j\in i_j} h_j)^2}{\sum_{j\in i_j} h_j + \lambda} + \gamma S \tag{61}$$

In practice, using the set score of the events of the left J_l and right J_R nodes after the split, the $J = J_r \bigcup J_l$ evaluation for the split candidates is as follows, minimizing the loss after the split:

$$L_{split} = \frac{1}{2}\left[\frac{\left(\sum_{j\in i_l} h_j\right)^2}{\sum_{j\in i_l} h_j + \lambda} + \frac{\left(\sum_{j\in i_l} h_j\right)^2}{\sum_{j\in i_l} h_j + \lambda} + \frac{\left(\sum_{j\in i_l} h_j\right)^2}{\sum_{j\in i_l} h_j + \lambda}\right] - \gamma \tag{62}$$

Finally, Algorithm 3 describes the steps involved in the facial expression classification using SR-CNN classifier.

Algorithm 3: Facial expression classification using SR-CNN classifier

Input	: $w \times g$, features, threshold condition
Output : facial expression	
1.	Initialize the random population
2.	Define discrete convolution $(l \otimes K)_{M,N} = \sum_{u=-w_K}^{w_K}\sum_{v=-g_K}^{g_K} K_{u,v} I_{M+u,i+v}$
3.	If i=0 , j=1
4.	While **Do**
5.	Define Ω penalizes the complexity $\Omega(F) = \gamma S + 1/2\lambda\sum_{i=1}^{S} w_i^2$
6	Define final prediction $\hat{y}_j = \varphi(y_j)\sum_{K=1}^{k} F_K(y_j), \quad F_K \in f$
7.	If not discard **then**
8.	Compute quality of a tree structure $\tilde{L}^{(s)}(p) = -\frac{1}{2}\sum_{i=1}^{S}\frac{\left(\sum_{j\in i_l} h_j\right)^2}{\sum_{j\in i_l} h_j + \lambda} + \gamma S$
9.	Update the tree values
10.	Compute fitness
11.	end

5. Results and Discussion

In this part, we approve the exhibition of our proposed ongoing face acknowledgment and multimodal look location by utilizing two distinct benchmark datasets are expanded Cohn-Kanade (CK+) and constant 5671 man pictures. Furthermore, we use python programs in view of Google Colab open source programming bundles. The information handling stage is a PC with Windows 7, Intel(R) Core(TM) i3-2120 computer processor (3.30 GHz,) and 4.00 GB Slam. The consequences of our proposed SVM+ SR-CNN classifier is contrasted and the current condition of-craftsmanship classifiers are LBP-TOP [32], CNN+SVM [32], LBP-SP+SVM [33], AUDN+SVM [34], AlexNet+SVM [35] and CNN [31].

5.1 Dataset description

CK+ dataset: Patrick Lucey et al. introduced the CK+ dataset containing 593 arrangements from 123 subjects, who are of the two sexual orientations and have different social and training foundations. Every one of the successions integrates pictures from beginning to the pinnacle articulation. In any case, just 327 of the 593 groupings were found to meet one of the measures for discrete feelings. Along these lines, 327 pinnacle outlines were chosen and named. In spite of its prevalence, CK+ data set contains predetermined number of pictures and subjects, so we augment the dataset falsely through mark safeguarding change. The picture pixel is brought together to 160× 120. Four patches with the size of 128×96 are trimmed from the upper left, upper right, base left and base right corner of the picture to broaden information. At long last, 1592 pictures together form the beginning look picture dataset. Fig. 2 shows the test samples from the CK+ dataset with different facial expressions.

Table 2: Summary of dataset description

Facial expressions	CK+ dataset		Kaggle 5671 person image dataset	
	Training	Testing	Training	Testing
Angry	115	65	4589	450
Contempt	115	29	3155	450
Disgust	115	121	3578	450
Fear	115	85	2746	450
Happy	115	161	5378	450
Sadness	115	109	2498	450
Surprise	115	217	3784	450
Neutral	NA	NA	2981	439
Total samples	805	787	28709	3589

Kaggle 5671 person image dataset: The information comprises of 48x48 pixel grayscale pictures of appearances. The appearances have been consequently enlisted with the goal that the face is pretty much focused and possesses about a similar measure of room in each picture. The assignment is to sort each face in view of the inclination displayed in the look into one of seven classes (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The preparation set comprises of 28,709 models and the public testing set comprises of 3,589 models. Table 2 portrays the detail respects both datasets.

Fig 2: Test samples from CK+ dataset with different facial expressions (top to bottom) are Happy, Anger, Contempt, Disgust, Fear, Sadness and Surprise

5.2 Comparative analysis

5.2.1 Performance comparison with respect to CK+ dataset

In this subsection, we validate our proposed SVM+ SR-CNN classifier with respect to the CK+ dataset using different performance measures. Table 3 describes the comparative analysis of our planned and existing state-of-art video retrieval frameworks with respect to CK+ dataset. The accuracy of planned SVM+ SR-CNN classifier is 21.3%, 20.485%, 18.054%, 17.13%, 12.143%, 6.699%, 8.668%, 6.672% and 3.563% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. The precision of proposed SVM+ SR-CNN classifier is 21.553%, 20.728%, 18.268%, 17.333%, 12.287%, 6.779%, 8.771%, 6.752% and 3.605% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. The sensitivity

of proposed SVM+ SR-CNN classifier is 21.449%, 20.629%, 18.181%, 17.25%, 12.228%, 6.746%, 8.728%, 6.719% and 3.588% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+ SVM and CNN, respectively. The specificity of proposed SVM+ SR-CNN classifier is 21.473%, 20.652%, 18.2%, 17.269%, 12.242%, 6.754%, 8.738%, 6.726% and 3.592% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. F-measure of proposed SVM+ SR-CNN classifier is 21.501%, 20.679%, 18.225%, 17.292%, 12.258%, 6.763%, 8.75%, 6.735% and 3.597% efficient than the existing state-of-art video retrieval frameworks are NB, k-NN, RF, LSTM, SVM, CNN CPU [35], CPU serial [34], and CUDA-GPU parallel [33], respectively. Fig. 3 shows the comparative analysis of quality measures with respect to CK+ dataset. The false positive rate (FPR) of proposed SVM+ SR-CNN classifier is 26.051%, 25.218%, 22.731%, 21.786%, 16.686%, 11.119%, 13.132%, 11.091% and 7.911% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively.

Table 3: Comparative analysis with respect to CK+ dataset

Facial expression classifiers	In %							In seconds
	Accuracy	Precision	Sensitivity	Specificity	F-measure	FPR	FAR	Detection time
LBP-TOP	79.917	78.792	79.249	79.145	79.02	73.786	71.441	125
CNN+SVM	84.698	83.573	84.03	83.926	83.801	78.567	76.222	120
LBP-SP+SVM	89.917	88.792	89.249	89.145	89.02	83.786	81.441	105
AUDN+SVM	88.03	86.905	87.362	87.258	87.133	81.899	79.554	95
AlexNet+SVM	89.943	88.818	89.275	89.171	89.046	83.812	81.467	60
CNN	92.924	91.799	92.256	92.152	92.027	86.793	84.448	52
SVM+ SR-CNN	96.34	95.215	95.672	95.568	95.443	94.209	95.944	41

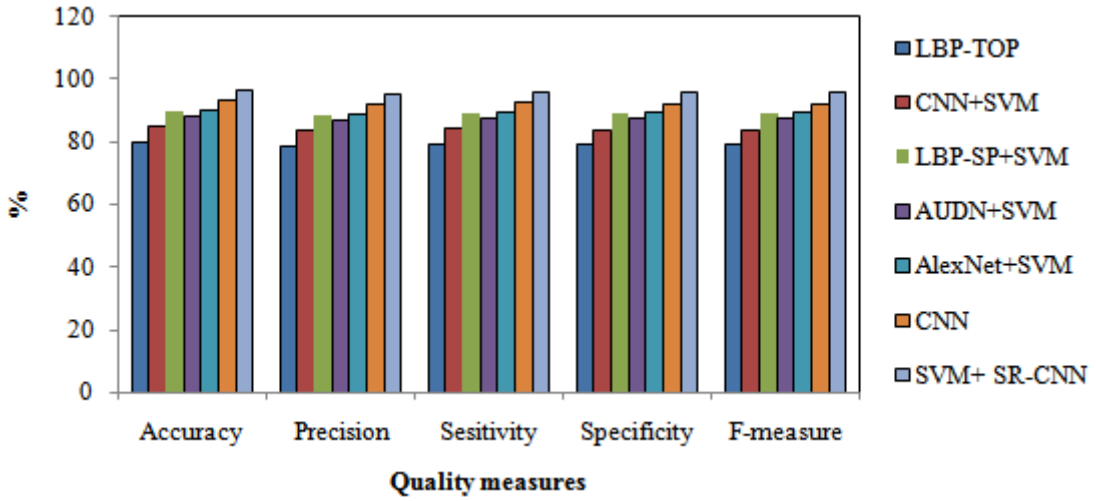

Fig 3: Quality measures for CK+ dataset

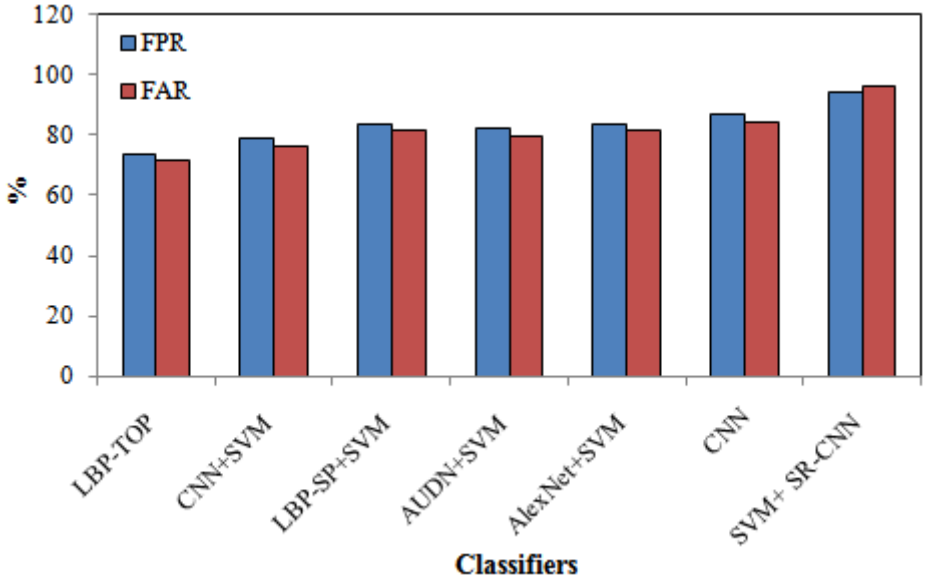

Fig 4: Error measures for CK+ dataset

False acceptance rate (FAR) of proposed SVM+SR-CNN classifier is 29.851%, 29.033%, 26.592%, 25.664%, 20.656%, 15.19%, 17.166%, 15.163% and 12.041% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. Fig. 4 shows the comparative analysis of error measures with respect to CK+ dataset. The average retrieval time of proposed SVM+ SR-CNN classifier is 66.23%, 67.32%, 61.4%, 67.54%, 66.33%, 61.41%, 57.37%, 32.47% and 22.07% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. Fig. 5 shows summary of timing measures are analyzed with respect to CK+ dataset. Fig. 6 shows the facial expression classification results that achieved from our proposed SVM+ SR-CNN classifier for CK+ dataset.

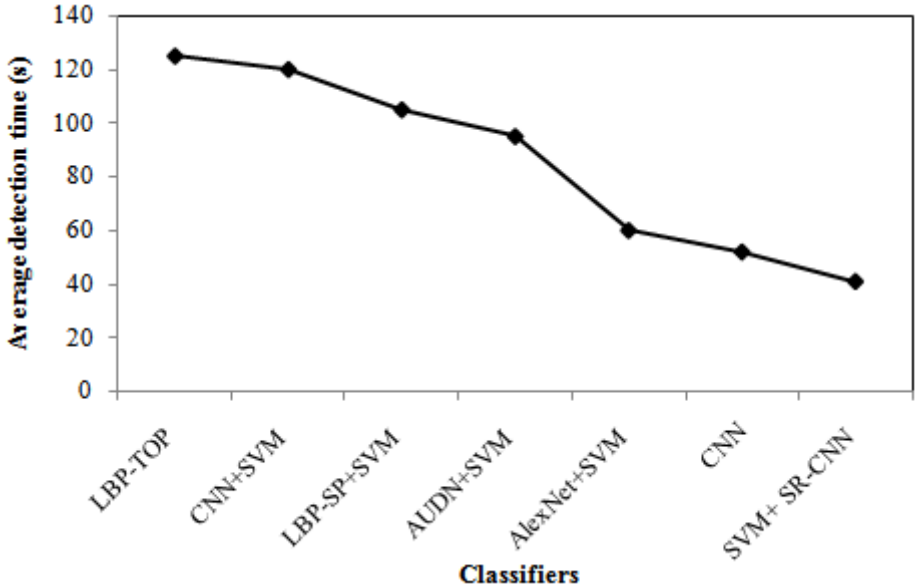

Fig 5: Timing measures for CK+ dataset

Fig 6: Results of proposed SVM+ SR-CNN classifier for CK+ dataset

Table 4: Comparative analysis with respect to Kaggle 5671 person image dataset

Facial expression classifiers	In %							In seconds
	Accuracy	Precision	Sensitivity	Specificity	F-measure	FPR	FAR	Detection time
LBP-TOP	83.451	82.326	82.783	82.679	82.554	77.32	74.975	129
CNN+SVM	88.232	87.107	87.564	87.46	87.335	82.101	79.756	124
LBP-SP+SVM	93.451	92.326	92.783	92.679	92.554	87.32	84.975	109
AUDN+SVM	91.564	90.439	90.896	90.792	90.667	85.433	83.088	99
AlexNet+SVM	93.477	92.352	92.809	92.705	92.58	87.346	85.001	64
CNN	96.458	95.333	95.79	95.686	95.561	90.327	87.982	56
SVM+ SR-CNN	99.874	98.749	99.206	99.102	98.977	97.743	99.478	45

5.2.2 Performance comparison with respect to Kaggle 5671 person image dataset

In this subsection, we validate our proposed SVM+ SR-CNN classifier with respect to the Kaggle 5671 person image dataset using different performance measures. Table 4 describes the comparative analysis of our planned and existing state-of-art video retrieval frameworks with respect to Kaggle 5671 person image dataset. The accuracy of planned SVM+ SR-CNN classifier is 21.3%, 20.485%, 18.054%, 17.13%, 12.143%, 6.699%, 8.668%, 6.672% and 3.563% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+ SVM, AlexNet+SVM and CNN, respectively. The precision of proposed SVM+ SR-CNN classifier is 21.553%, 20.728%, 18.268%, 17.333%, 12.287%, 6.779%, 8.771%, 6.752% and 3.605% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. The sensitivity of proposed SVM+ SR-CNN classifier is 21.449%, 20.629%, 18.181%, 17.25%, 12.228%, 6.746%, 8.728%, 6.719% and 3.588% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+ SVM and CNN, respectively. The specificity of proposed SVM+ SR-CNN classifier is 21.473%, 20.652%, 18.2%, 17.269%, 12.242%, 6.754%, 8.738%, 6.726% and 3.592% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively.

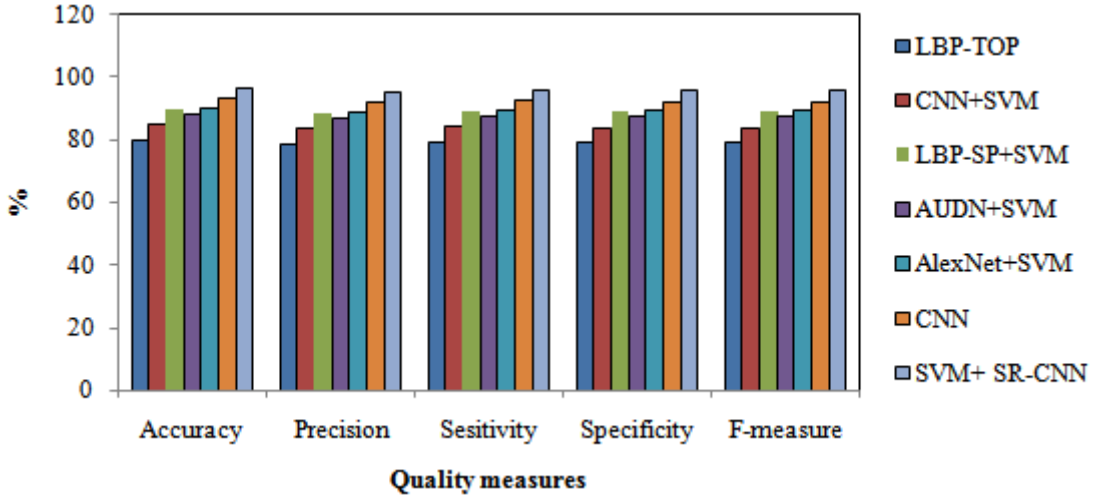

Fig 7: Quality measures for Kaggle 5671 person image dataset

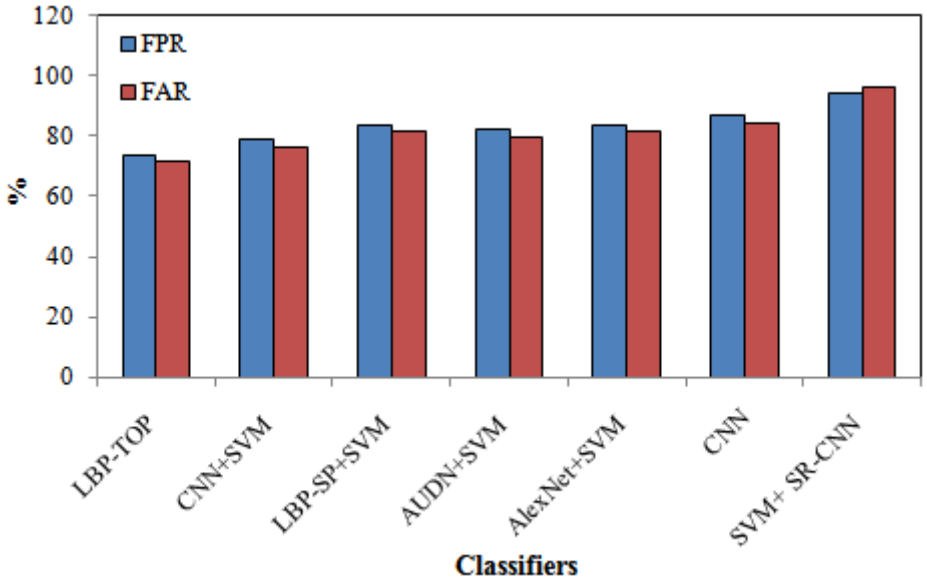

Fig 8: Error measures for Kaggle 5671 person image dataset

F-measure of proposed SVM+ SR-CNN classifier is 21.501%, 20.679%, 18.225%, 17.292%, 12.258%, 6.763%, 8.75%, 6.735% and 3.597% efficient than the existing state-of-art video retrieval frameworks are NB, k-NN, RF, LSTM, SVM, CNN CPU [35], CPU serial [34], and CUDA-GPU parallel [33], respectively. Fig. 7 shows the comparative analysis of quality measures with respect to Kaggle 5671 person image dataset. The false positive rate (FPR) of proposed SVM+ SR-CNN classifier is 26.051%, 25.218%, 22.731%, 21.786%, 16.686%, 11.119%, 13.132%, 11.091% and 7.911% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. False acceptance rate (FAR) of proposed SVM+SR-CNN classifier is 29.851%, 29.033%, 26.592%, 25.664%, 20.656%, 15.19%, 17.166%, 15.163% and 12.041% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. Fig. 8 shows the comparative analysis of error measures with respect to Kaggle 5671 person image dataset. The average retrieval time of proposed SVM+ SR-CNN classifier is 66.23%, 67.32%, 61.4%, 67.54%, 66.33%, 61.41%, 57.37%, 32.47% and 22.07% efficient than the existing state-of-art classifiers are LBP-TOP, CNN+SVM, LBP-SP+SVM, AUDN+SVM, AlexNet+SVM and CNN, respectively. Fig. 9 shows the summary of timing measures are analyzed with respect to Kaggle 5671 person image dataset. Fig. 10 shows the facial expression classification results that achieved from our proposed SVM+ SR-CNN classifier for Kaggle 5671 person image dataset.

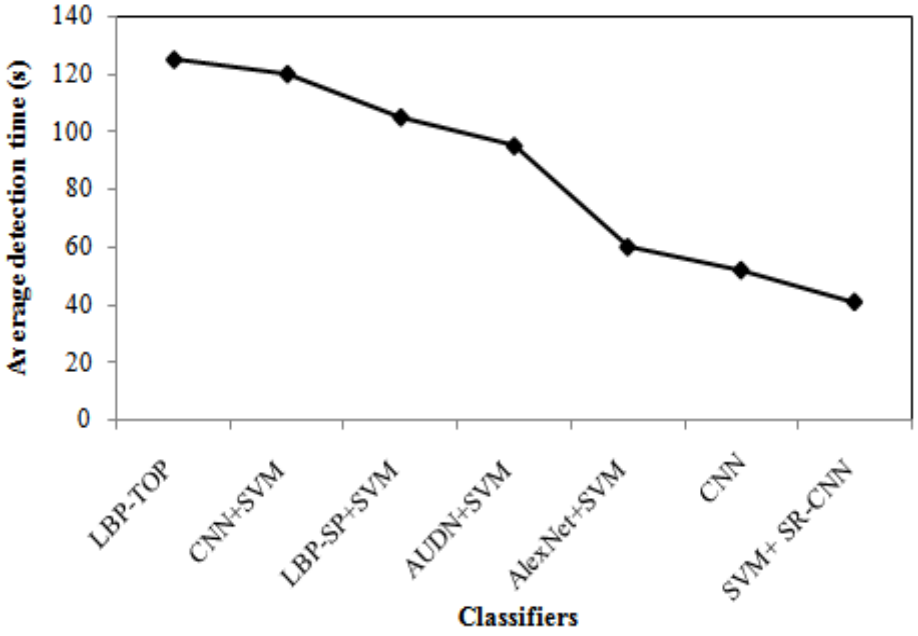

Fig 9: Timing measures for Kaggle 5671 person image dataset

Fig 10: Results of proposed SVM+ SR-CNN classifier for Kaggle 5671 person image dataset

6. Conclusion

In this paper, we have proposed real-time face recognition and multimodal facial expression detection using hybrid deep learning techniques. Prewitt edge detector for eyes, nose, mouth, and brow detection, and design a deep transfer learning model to extract deep features from face images. The feature optimization is done by harpy eagle search optimization (HESO) to selects optimal best among multiple features which reduce data dimensionality problem. SVM classifier is used for face recognition and detection. SR-CNN classifier is then utilized here for the multi-class detection of facial expressions from multimodalities. Finally, we validate the performance our framework using benchmark multi-modal Cohn-Kanade and real-time 5671 person images datasets. From the simulation results, we observed that the effectiveness of our proposed SR-CNN classifier is very high with respect to the both datasets. The proposed SR-CNN classifier achieved maximum accuracy, precision, sensitivity, specificity, f-measure of 99.874%, 98.749%, 99.206%, 99.102% and 98.977%, respectively which is efficient associated to the existing state-of-art classifiers. The average retrieval time of our proposed SR-CNN classifier is 45 seconds which is 45.57% efficient than the existing state-of-art classifiers.

References

[1] Panksepp J (2005) Affective consciousness: Core emotional feelings in animals and humans. Conscious Cogn 14(1):30-80.

[2] Plutchik R (2001) The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Amer Scient 89(4):344-350.

[3] Zautra AJ (2003) Emotions, stress, and health. Oxford University Press, Oxford.

[4] Kohler CG, Martin EA, Stolar N, Barrett FS, Verma R, Brensinger C et al (2008) Static posed and evoked facial expressions of emotions in schizophrenia. Schizophr Res 105(1-3):49-60.

[5] Ambron E, Foroni F (2015) The attraction of emotions: irrelevant emotional information modulates motor actions. Psychon Bull Rev 22(4):1117-1123.

[6] Kumari J, Rajesh R, Kumar A (2016) Fusion of features for the effective facial expression recognition. Paper presented at the international conference on communication and signal processing, IEEE, Melmaruvathur, 6–8 June 2016.

[7] Shergill GS, Sarrafzadeh A, Diegel O, Shekar A (2008) Computerized sales assistants: the application of computer technology to measure consumer interest-a conceptual framework. J Electron Commer Res 9(2):176-191.

[8] Tierney M (2017) Using behavioral analysis to prevent violent extremism: Assessing the cases of Michael Zehaf-Bibeau and Aaron Driver. J Threat Assessm Manag 4(2):98-110.

[9] Nonis F, Dagnes N, Marcolin F, Vezzetti E (2019) 3D approaches and challenges in facial expression recognition algorithms - A literature review. Appl Sci 9(18):3904.

[10] Sandbach G, Zafeiriou S, Pantic M, Rueckert D (2011) A dynamic approach to the recognition of 3D facial expressions and their temporal models. Paper presented at the ninth IEEE international conference on automatic face and gesture recognition, IEEE, Santa Barbara, 21–25 March 2011.

[11] E. Paul, D. Matsumoto, and V. F. Wallace, “Facial expression in affective disorders,” What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), vol. 2, pp. 331–342, 1997.

[12] T. Kanade, J. F. Cohn, and Y. Tian, “Comprehensive database for facial expression analysis,” in Proceedings of the the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), pp. 46–53, IEEE, Grenoble, France, March 2000.

[13] J. Mei, K. Li, and K. Li, “Customer-satisfaction-aware optimal multiserver configuration for profit maximization in cloud computing,” IEEE Transactions on Sustainable Computing, vol. 2, no. 1, pp. 17–29, 2017.

[14] V. F. Wallace, E. Paul et al., “Emfacs-7: emotional facial action coding system,” University of California at San Francisco, vol. 2, no. 36, p. 1, 1983.

[15] G. Schubert, “Human ethology and evolutionary epistemology: the strange case of dmEibesfeldthuman eieeyaPp. xvi, 848, $69.95,” Journal of Social and Biological Systems, vol. 13, no. 4, pp. 355–387, 1990.

[16] E. S. Jaha and L. Ghouti, “Color face recognition using quaternion pca,” in Proceedings of the the 4th International Conference on Imaging for Crime Detection and Prevention 2011, pp. 1–6, ICDP, 2011.

[17] P. Ekman, Di Perrett, and H. D. Ellis, “Facial expressions of emotion: an old controversy and new findings: Discussion,” Philosophical Transactions of the Royal Society of London,Series A B, vol. 335, p. 69, 1992.

[18] X. Zhou, K. Li, G. Xiao, Y. Zhou, and K. Li, “Top f probabilistic products queries,” IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 10, pp. 2808–2821, 2016.

[19] M. Bichsel and A. P. Pentland, “Human face recognition and the face image Set′s topology,” CVGIP: Image Understanding, vol. 59, no. 2, pp. 254–261, 1994.

[20] K. Li, W. Ai, Z. Tang et al., “Hadoop recognition of biomedical named entity using conditional random fields,” IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 11, pp. 3040–3051, 2014.

[21] Kommineni, J., Mandala, S., Sunar, M.S. and Chakravarthy, P.M., 2021. Accurate computing of facial expression recognition using a hybrid feature extraction technique. The Journal of Supercomputing, 77(5), pp.5019-5044.

[22] Vedantham, R., 2022. Adaptive increasing-margin adversarial neural iterative system based on facial expression recognition feature models. Multimedia Tools and Applications, 81(3), pp.3793-3830.
[23] Dubey, A.K. and Jain, V., 2022. An accurate recognition of facial expression by extended wavelet deep convolutional neural network. Multimedia Tools and Applications, pp.1-31.
[24] Said, Y. and Barr, M., 2021. Human emotion recognition based on facial expressions via deep learning on high-resolution images. Multimedia Tools and Applications, 80(16), pp.25241-25253.
[25] Juneja, K. and Rana, C., 2021. An extensive study on traditional-to-recent transformation on face recognition system. Wireless Personal Communications, 118(4), pp.3075-3128.
[26] Bellamkonda, S., Gopalan, N.P., Mala, C. and Settipalli, L., 2022. Facial expression recognition on partially occluded faces using component based ensemble stacked CNN. Cognitive Neurodynamics, pp.1-24.
[27] H. Li, K. Li, J. An, and K. Li, "Msgd: a novel matrix factorization approach for large-scale collaborative filtering recommender systems on gpus," IEEE Transactions on Parallel and Distributed Systems, vol. 29, no. 7, pp. 1530–1544, 2017.
[28] Atsawaraungsuk, S., Katanyukul, T., Polpinit, P. and Eua-anant, N., 2022. Fast and robust online-learning facial expression recognition and innate novelty detection capability of extreme learning algorithms. Progress in Artificial Intelligence, 11(2), pp.151-168.
[29] da Silva, E.P., Costa, P.D.P., Kumada, K.M.O. and De Martino, J.M., 2022. Facial action unit detection methodology with application in Brazilian sign language recognition. Pattern Analysis and Applications, 25(3), pp.549-565.
[30] India, T. (2023, October 26). Discussing About Artificial Intelligence (AI) in Data Science with Damodarrao Thakkalapelli -Data Solutions Architect. Tribuneindia News Service. Retrieved November 10, 2023, from https://www.tribuneindia.com/news/impact-feature/discussing-about-artificial-intelligence-ai-in-data-science-with-damodarrao-thakkalapelli-data-solutions-architect-556765
[31] Machine Learning Patents and Patent Applications (Class 706/12) - Justia Patents Search. (n.d.). Retrieved November 10, 2023, from https://patents.justia.com/patents-by-us-classification/706/12
[32] Desk, O. W. (2023, October 25). Discussing Real world Data Processing Problems and Solutions with Damodarrao Thakkalapelli -Data Solutions Architect. https://www.outlookindia.com/. Retrieved November 10, 2023, from:
[33] https://www.outlookindia.com/business-spotlight/discussing-real-world-data-processing-problems-and-solutions-with-damodarrao-thakkalapelli-data-solutions-architect-news-326551
[34] Thakkalapelli, D. (2023). Cloud Migration Solution: Correction, Synchronization, and Migration of Databases. Tuijin Jishu/Journal of Propulsion Technology, 44(3), 2656-2660.
[35] Y. Xu, K. Li, L. He, L. Zhang, and K. Li, "A hybrid chemical reaction optimization scheme for task scheduling on heterogeneous computing systems," IEEE Transactions on Parallel and Distributed Systems, vol. 26, no. 12, pp. 3208–3222, 2014.
[36] Wei, W., Jia, Q., Feng, Y., Chen, G. and Chu, M., 2020. Multi-modal facial expression feature based on deep-neural networks. Journal on Multimodal User Interfaces, 14(1), pp.17-23.
[37] Randhi, V. R., Thakkalapelli, D., Kavali, R. V. S., & Dabbiru, R. (2022). U.S. Patent Application No. 17/830,849.
[38] Kavali, R. V. S., Randhi, V. R., Thakkalapelli, D., Vegulla, V. K., & Maramreddy, R. (2023). U.S. Patent Application No. 17/576,539.
[39] Grandhye, N. B., Randhi, V. R., Vegulla, V. K., Kavali, R. V. S., & Thakkalapelli, D. (2023). U.S. Patent No. 11,716,278. Washington, DC: U.S. Patent and Trademark Office.
[40] https://doi.org/10.1016/j.techfore.2017.10.002
[41] Vegulla, V. K., Kavali, R. V. S., Randhi, V. R., & Thakkalapelli, D. (2023). U.S. Patent Application No. 17/680,561.
[42] Talluri, S., Randhi, V. R., Thakkalapelli, D., & Kavali, R. V. S. (2022). U.S. Patent Application No. 17/307,173.
[43] Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. IEEE Trans AfectComput 6(1):1–12

[44] Gupta, S., Kumar, P. and Tekchandani, R.K., 2022. Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimedia Tools and Applications, pp.1-30.

[45] Liu MY, Li SX, Shan SG, Chen XL (2015) AU-inspired deep networks for facial expression feature learning. Neurocomputing159:126–136

[46] Grandhye, N. B., Randhi, V. R., Vegulla, V. K., Kavali, R. V. S., & Thakkalapelli, D. (2023). U.S. Patent Application No. 17/583,634.

[47] Kavali, R. V. S., D'silva, L., Randhi, V. R., & Thakkalapelli, D. (2023). U.S. Patent No. 11,604,691. Washington, DC: U.S. Patent and Trademark Office.

[48] Randhi, V. R., Thakkalapelli, D., Kavali, R. V. S., & Dabbiru, R. (2022). U.S. Patent Application No. 17/830,849.

[49] Dr. Sourabh Sharma (2023), "The Recognition of Women Justice And Equality", Journal of Survey in Fisheries Sciences 10(1) 2953-2963.

[50] Vo DM, Sugimoto A, Le TH (2016) Facial expression recognition by re-ranking with global and local generic features. In: 23rd international conference on pattern recognition, IEEE, New York, 2016, pp 4118–4123