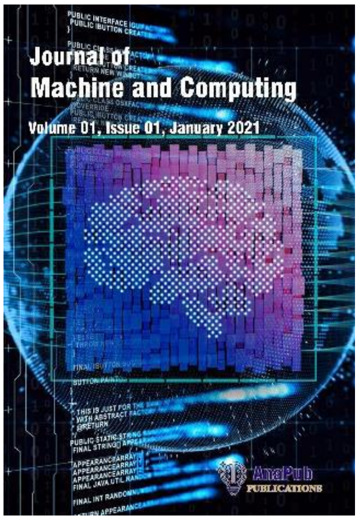
Journal Pre-proof

Adaptive Firefly Optimization Based Feature Selection and Ensemble Machine Learning Algorithm for Facial Expression Emotion Recognition



Sudha S S and Suganya S S

DOI: 10.53759/7669/jmc202505122

Reference: JMC202505122

Journal: Journal of Machine and Computing.

Received 15 January 2025

Revised form 02 April 2025

Accepted 27 May 2025

Please cite this article as: Sudha S S and Suganya S S, "An Adaptive Firefly Optimization Based Feature Selection and Ensemble Machine Learning Algorithm for Facial Expression Emotion Recognition", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505122.

Adaptive Firefly Optimization Based Feature Selection and Ensemble Machine Learning Algorithm for Facial Expression Emotion Recognition

S. S. Sudha[1*], Dr. S. S. Suganya[2]

[1*]Assistant professor, Department of Applied Mathematics and Computational Sciences, PSG College of Technology, Avinashi Road, Peelamedu – 641004, Coimbatore.

[2]Associate professor, Department of Computer Science, Dr. SNS Rajalakshmi College of Arts and Science, Thudiyalur – Saravanampatti Road, Post, Chinnavedampatti - 641049, Coimbatore, Tamilnadu, India.

E-mail Id: sudhasree2005@gmail.com[1*], suganya.annur@gmail.com[2]

Abstract

A person's emotional state can be determined from their facial expression emotion recognition (FEER). Rich emotional information can be found in FEER. One of the most crucial types of interpersonal communication is FEER. Finding computational methods to replicate facial emotion expression in a similar or identical manner remains an unresolved issue, despite the fact that it is a skill that humans naturally do. To overcome the problem, in this work, Adaptive Firefly Optimization (AFO) and Ensemble (EML) Machine Learning (EML) algorithm is proposed for FEER. In this work, initially, dataset is collected using CK+ database and KMU-FED database. In occlusion generation, occlusions around mouths and eyes are duplicated. When calculating the optical flow, we aim to preserve as much information as possible with normalized inputs that deep networks require for recognitions and reconstructions. The reconstruction is done by using Deep Q-learning (DQL) which is used for semantic segmentation (SS) based on occlusions. For Feature selection (FS), the AFO algorithm is used. From the provided database, AFO is utilised to choose more pertinent and redundant features. It generates best fitness values (FV) using objective function (OF) for higher recognition accuracy (ACC). EML algorithms including the K-Nearest Neighbour (KNN), Random Forest (RF), and Enhanced Artificial Neural Network (EANN) are used to execute FEER. EML provides faster convergence time during training and testing process. It is mainly used to classify the accurate FEER results for the given database. According to the results, the suggested AFO-EML method overtakes the current techniques by ACC, precision (P), recall (R), and f-measure.

Key words: Facial expression emotion recognition (FEER), feature selection (FS), Adaptive Firefly Optimization (AFO) and Ensemble Machine Learning (EML) algorithm

1. Introduction

In order to perceive and comprehend human emotional states, facial expressions (FEX) are essential and fundamental [1]. The development of several emotion datasets, including large-scale real-world expression datasets and laboratory-collected emotion datasets like CK+, has sped up the advancement of facial expression recognition (FER) technologies in recent years. Happiness, anger, disgust, fear, sadness, surprise, and neutral are the seven core emotion categories that current FER systems primarily used to detect a wide range of human inner states. Nonetheless, it is a basic truth that the varied human emotions in routine situations cannot be adequately expressed by a small number of emotional words. In the context of "surprise," the terms "amazement" and "astonishment" denote positive and negative feelings. Astonishment is more semantically related to fear, but amazement is more semantically related to "joy" [2]. The ability to perceive more complex emotional states in people and identify subtle FEX remains a considerable barrier, despite the potent depictions of coarse-grained basic emotions. This difficulty is especially crucial for future emotional interactions.

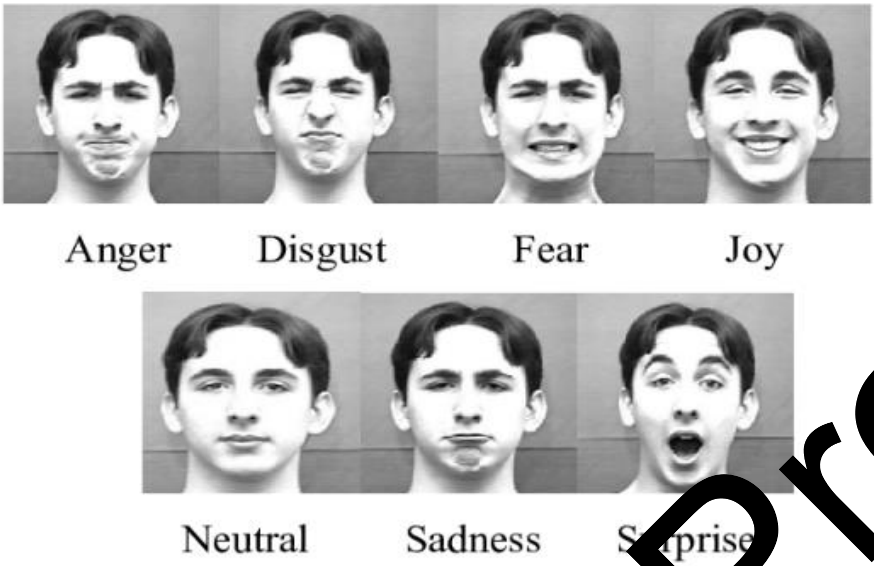


Fig 1 Various facial expressions [22]

The ability to recognise and comprehend FEX has become essential for efficient human–computer interaction as artificial intelligence (AI) advances [3]. Human facial image (HFI) expression identification has garnered a lot of research attention. It has been widely used in many different sectors, including as psychology and transportation. This increasing attention emphasises how crucial facial analysis is to comprehending human emotions [4]. HFI-based emotion analysis (EA) is quite important. Furthermore, FEX allows people to communicate their core feelings regardless of cultural variation. A discrete model and a dimensional model are the two main categories of typical face EA methods [5]. The discrete model is restricted in its capability to detect the intricacy and variation of emotions since emotion changes in a continuous and seamless process.Aside from basic emotions, the dimensional model may represent complicated emotions and quantify their intensity. The natural state of human emotions can be better captured by this dimensional approach. It is more in line with how people think. Arousal (A) and valence (V) are the most often utilised and recognised emotional dimensions. The full emotional space and shifts are well described by these A and V. V indicates if the emotion is favourable, whereas A indicates the intensity of the emotion becomes.

FER also has trouble identifying and utilising facial features in complicated backgrounds, particularly when the image is blurry or unclear. Based on the confrontation generating network, it produced facial images with particular expression tags by incorporating the features of several heterogeneous networks in various depths and regions [6] [7]. All of the processes in FER must be completed without or with the least amount of user involvement, which is a crucial criterion. Usually, this entails FEX classification, facial information extraction and tracking, and initial face detection.The specific application under this framework enforces the actual implementation and integration details. For instance, real-time (RT) performance might not be a necessary feature of the system if behavioural science is the application domain of the integrated system.

The purpose of this study is the FEER using AFO-EML. There are several research and methodologies introduced but the FEER accuracy is not achieved significantly. The existing approaches has drawback with variation in pose, illumination, and FEX. This study's primary contributions are FS, occlusion generation (OG), optical flow calculation (OFC), and FEER. The proposed method is used to provide more accurate classification results using effective algorithms for the given CK+ database & KMU-FED database

The remaining sections of the work are organised in the following order: Section 2 gives a review of some of the research being done in the FS and FEER domains. In section 3, the details of the suggested technique for the AFO-EML scheme are presented. Section 4 presents the results and discussions. At last, the findings are summarised in Section 5.

2. Related work

In [8], Liu et al. (2017) suggest a four-layer system structure for the FEER-based Human-Robot Interaction (FEER-HRI) system. Through the use of the FEERHRI system, the robots are able to recognise human emotions and generate FEX in response to them. A FER method based on 2D-Gabor, uniform local binary pattern (LBP) operator, and multiclass extreme learning machine (ELM) classifier is used to build RT FER for robots. The FEX of robots are displayed on an LED screen that is integrated into the robot. It can be represented by simple cartoon symbols that are easy enough for humans to understand. Four scenarios are used in the human-robot interaction experiment: scene simulation, residential services, fun, and guidance.FER of humans and FEX generation of robots enable smooth communication in these scenarios.

A novel multimodal emotion recognition network was presented by Cui et al. (2024) in [9]. In this suggested work, continuous FEX and EEG signals are used. In order to create amalgamated vectors containing mutual information, it integrated the cross-modal attention fusion mechanism (CMAFM) to create strong correlations between modal feature vectors. Furthermore, spatiotemporal (ST) information was extracted from FEX images using a Self-Attention Convolutional Long Short-Term Memory (SA-ConvLSTM).Using the provided datasets, the model put forward in this study is experimentally assessed. Compared to the modern research methodologies currently in use, its recognition ACC was higher. However, it continued to show good performance in the DEAP dataset-based Leave-One-Subject-Out (LOSO) experiment. The model's performance in the multimodal emotion recognition (MER) test was demonstrated by the experimental data.

For enhanced emotion detection (ED), Tao et al. (2024) introduced the multi-view (FF) feature fusion (MVFF) technique in [10], which makes use of convolutional neural networks (CNNs). First, face videos are used to extract Imaging PhotoPlethysmoGraphy (IPPG) signals. Then, to accomplish multi-view representation of emotional qualities within IPPG and facial video signals, researchers use branch CNN and heart rate variability (HRV) for feature extraction (FE). The DEAP public dataset was used to validate the suggested approach. It shows that for the A and V dimensions, the recommended method achieved ACCs of 72.37% and 70.82%.Compared to approaches that merely employ FEX, the multi-view approach improves emotion recognition (ER) ACC by 7.23% and 5.31% for A and V, respectively. This advancement demonstrates the way the suggested method could increase the P and resilience of ER approaches by capturing multimodal emotional expressions in facial videos without the need for additional sensors.
Multiple spatio-temporal FF (MSFF) was introduced by Lu et al. (2018) in [11]. In order to more accurately collect ST emotional information, this suggested method integrates two sources that are mutually complementary: the audio and the face image. Facial image and audio models are components of the framework. Three alternative architectures of spatial-temporal NN are utilized in the facial image (FI) model to extract discriminative features about various emotions from images of people's FEX. The first step involves using pre-trained CNN like ResNet-50 and VGG-Face to extract high-level (HL) spatial information from video images. The speech spectrogram images that are acquired by preprocessing audio are also modeled in a VGG-BLSTM framework for the audio model in order to more effectively define the emotional fluctuation. Lastly, to improve ED performance, a fusion strategy based on the score matrices of many spatiotemporal networks obtained from the previous framework is suggested. Our proposed MSFF has an overall accuracy of 60.64%, outperforming the winning team's performance and a considerable improvement over the baseline, according to extensive experiments
A novel approach for FEX analysis by recognising AUs from image sequences has been suggested by Pu et al. (2015) in [12] using a twofold (RF) random forest classifier. Using a Lucas-Kanade (LK) optical flow tracker to track Active Appearance Model (AAM) facial feature points and estimate feature point displacements, facial motion is measured. The displacement vectors between the neutral expression frame and the peak expression frame are the motion aspects of FEX. To ascertain the Action Units (AUs) of the matching expression sequences, they will then be converted to the first level RF. At last, the second level RF is used to classify those identified AUs using FEX.The suggested strategy can outperform a number of alternative ways on both AUs and FER, according to trials conducted on the Extended Cohn–Kanade (CK+) database.

In [13], Bhatti et al. (2021) suggested a feedforward learning paradigm and the use of FER by a teacher in the classroom. Faces are identified from gathered lecture recordings in order to achieve efficient HL FE, and once unnecessary frames are eliminated, pertinent frames are chosen. Deep features are then retrieved and fed into a classifier that employs several convolutional neural networks with parameter adjustments. In the classroom, Regularized Extreme Learning Machine (RELM) classifiers categorized unique expressions of instructors, promoting efficient learning and algorithm generalization. Three benchmark face datasets: the Cohn-Kanade, Japanese Female Face Expression (JAFFE), and FER 2013 (FER2013) datasets as well as a generated instructor FER dataset are used in the classroom experiments. Furthermore, the suggested method is compared to convolutional neural networks, traditional classifiers, and cutting-edge approaches. The trial results show a significant improvement in parameters like as recall, accuracy, and F1-score

3. Proposed methodology

In this study, AFO-EML procedure is suggested to develop the FEER results for the given CK+ database & KMU-FED database. The proposed system contains main phases are such occlusion generation, optical flow calculation (OFC), feature selection and facial expression emotion recognition. Fig. 2 shows the suggested system's overall block diagram.

3.1. Dataset collection

In terms of acquiring datasets, publicly accessible datasets are taken into account for identifying FEX, and the literature's current face occlusion techniques are utilized to establish a highly suitable testing strategy. Firstly, publicly accessible datasets pertaining to the recognition of expressions while occlusions are present are aggregated. The suggested technique can be trained on larger datasets thanks to the advantage of having an ensemble of numerous datasets. Second, different areas of the face are covered with occlusions. Normalized OFC are finally computed.

CK+ database: One popular database for FERs is Expanded CK+ [14]. 327 image sequences from 118 distinct patients are included in this collection, in addition to FE labels that depend on DFEER. These graphic sequences have the most emotional ending and a neutral beginning. Each subsequent image displays the emotion label, the FACS code, and the facial landmarks. Seven distinct feelings are categorized by the emotion labels: fear, happiness, sorrow, surprise, contempt, disgust, and rage. The experiment compares this strategy to several approaches based on the six primary expression categories using six emotions (sadness excluded). The images have 640 × 480 and 640 x 490 pixel resolutions, with grayscale values precisioned to 8 bits.

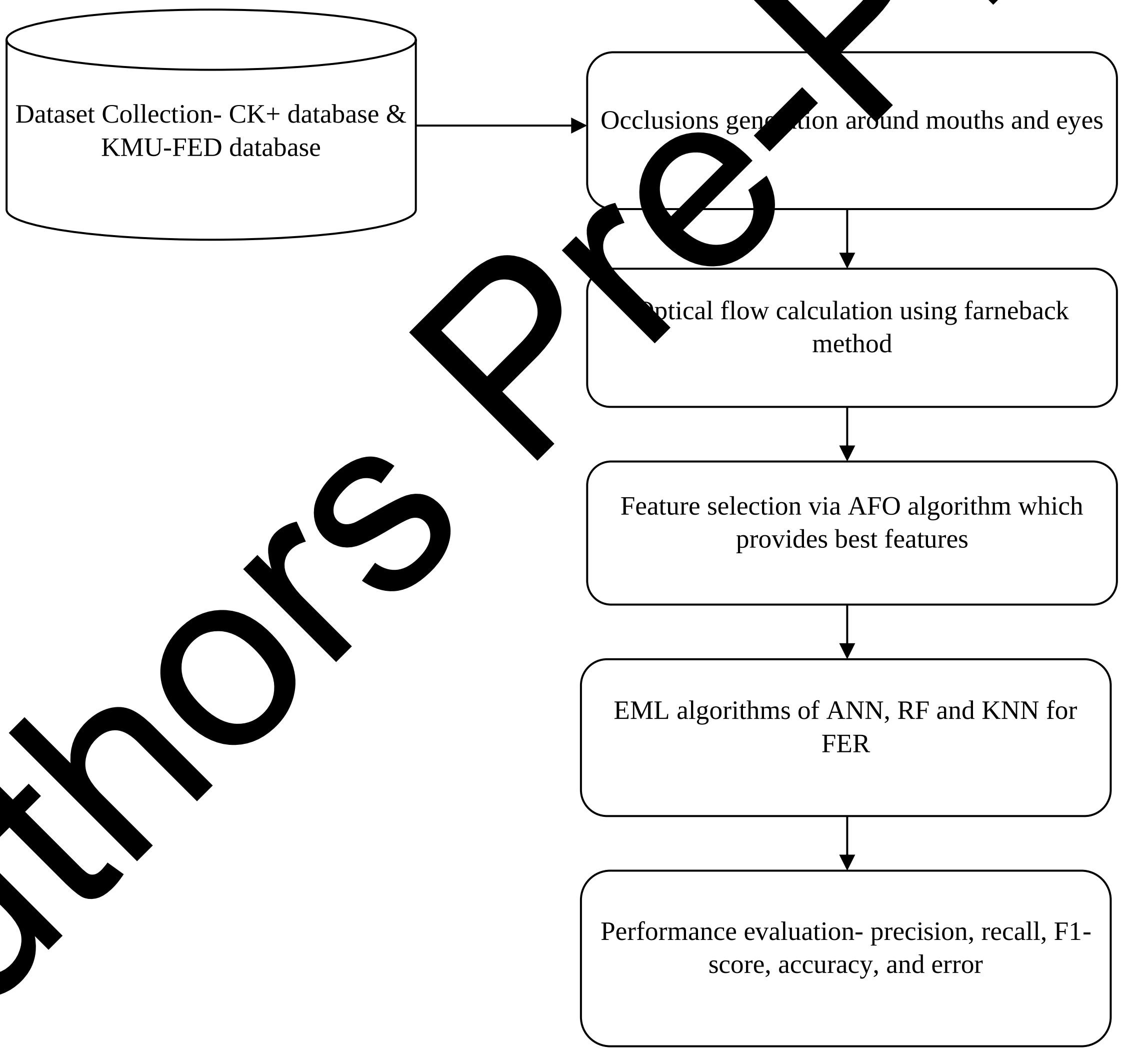


Fig 2 Overall block diagram of the suggested method

KMU-FED database: The standard dataset KMU-FED database for FERs is used which proves that the recommended approach is efficient when driving in the actual world. The dataset was created by using an NIR camera to record regular dataset series while driving in the actual world. KMUFED dataset has driver FEs obtained by NIR cameras mounted on dashboards or steering wheels. They encompass 55 image sequences with varying intensity (front, left, right, and rear light) and possible semi-occlusions of hairs or sunglasses on 12 persons. While analyzing the suggested method cross-validation approach were used on the dataset. Because there are no published findings from past research investigations utilizing the dataset accessed from the web [15].the suggested approach's accuracy values are investigated and evaluated using images resolutions of 1600 × 1200 pixels

3.2. Occlusions Generation

As shown in Fig. 3, occlusions (OCC) around mouths and eyes are replicated in order to address the majority of OCC already studied in the literature [16]. In FER, mouths and eyes are essential parts.

Fig 3 An instance of produced occlusions, utilized in our assessment, applied to an image from the CK+ dataset [23]

3.3. OFC

When calculating the optical flow, we aim to preserve as much information as possible with normalized inputs that deep networks require for recognitions and reconstructions. The original images (i.e., keeping their original resolution) are used, and optical flows are immediately scaled. First, process the images at their original resolutions. The 2nd stage involves cropping faces according to eye locations and inter-pupilar distances. Third, DQN is used to calculate optical flows from clipped faces. The optical flows from clipped faces are estimated in the third stage using Farneback approach, which is highly helpful for identifying facial expressions and can also be used to compute optical fluxes. The section on assessment discusses how to determine the best parameter size. In real flows, new x and y values are computed using sliding window components. Equation (1) is utilized to determine the new value for every coordinate (i, j).

$$resize\big(OF,(i,j)\big) = \begin{pmatrix} \mu OF[(\lfloor dx * i \rfloor], \ldots, (\lfloor dx * (i+1) \rfloor \lfloor * dx \rfloor - 1)], \\ \mu OF[(\lfloor dy * j \rfloor], \ldots, (\lfloor dy * (i+1) \rfloor \lfloor * dy \rfloor - 1)] \end{pmatrix} \tag{1}$$

where values for means of optical flows in windows, dx and dy are coefficients among actual and final sizes (dx = origSizes x/finalSizes x and dy = origSizes y/finalSizes y) where values for origSizes x and finalSizes x, reflect image's actual sizes and distances for final widths.

Then, partial occlusion segmentation is done by using Deep Q-learning. DQN—hybrid neural networks and Q-learning systems have demonstrated competitiveness in their performances. They describe relationships between state-action values and updates targets using dual network structures referred to as target networks and Q-networks. Subsequently, stochastic gradient descents update parameters of both target and Q networks. This strategy significantly reduces sample correlations and partially addresses local optimum problems. It should be mentioned that the performance of the DQN method is significantly influenced by the network design.

3.4. FS via AFO

In this study, FS is conducted using the AFO algorithm on the CK+ database & KMU-FED database. The FA was modelled after the social and biological behaviours of real firefly (FF). These fireflies emit brief and periodic flashes of light, which serve two purposes: attracting mates and signalling warnings. The Firefly Algorithm (FA) models this flashing behavior to optimize the objective function. FA works similarly to the way fireflies use their light, with a group of fireflies being guided by the intensity of the light as they move toward brighter and more desirable locations to achieve the best resolution for the target area.

At last, the system can determine the emotional states of the users and adjust to them accordingly. FEX are separated into six categories, and these categories are considered to identify people's subjective feelings and diverse cultural backgrounds. Additionally, the aforementioned six kinds of FEX are considered to be more universal when compared to other FEX classifications [17].

Thus, there are seven fundamental categories into which FEX can be classified: happy, angry, surprised, fear, disgusted, sad, and neutral. Additionally, AFO improves computing performance by reducing redundant features and data dimensions.

This method normalizes several characteristics of fireflies, as outlined

(i) Every firefly is attracted to a unique individual, regardless of its gender.

(ii) In the presence of two fireflies, the brighter light from the other influences the attraction of the initial one, establishing a clear relationship between the firefly's brightness and its level of attraction. A firefly will randomly alter its course if it cannot locate a brighter companion nearby.

The FA is selected for its effectiveness in providing optimal solutions for multi-objective (MO) problems. The brightness may be exactly proportional to the OF when maximizing is the objective [18]. The brightness of a FF, which is equivalent to the encoded OF, is assumed to be the determining factor in its attraction for its complexity. These processes are carried out frequently until the convergence requirements are satisfied.

a) The following is how the inverse square rule describes variations in light intensity: Light Intensity and Attractiveness at the Source

$$I(r) = \frac{I_0}{r^2} \quad (2)$$

Here, the light intensity of the attraction is denoted by I(r) r^2.

By randomly allocating characteristics, attractive

b) The following is the light intensity, and the following is the intermediate:

$$I(r) = I_0 \exp\,(-\gamma r) \quad (3)$$

where I_0 represents the energy absorption ability of a material.

c) The following is how the Gaussian form of the approximation is taken into consideration in order to avoid the singularity:

$$I(r) = I_0 \exp\,(-\gamma r^2) \quad (4)$$

The brightness of a FF is influenced by the number of light that fireflies in the area can detect. A new solution is generated by considering possible variations and randomly altering the pixels. As a result, β a firefly's appeal is determined as follows.

$$\beta = \beta_0 \exp(-\gamma r^m) \quad (5)$$

Here, the attraction at r=0 is denoted by β_0.

Use the following formula to get the distance among any two FF (facial features), i and j.

$$r_{i,j} = \sqrt{\sum_{k=1}^{d} (x_{i,k} - x_{j,k})^2} \quad (6)$$

x_i represents the kth component of the spatial match $x_{j,k}$. The number of dimensions in this case is denoted by d and d. To create an AF, an adaptation parameter for the random and absorption components is applied. These modifications linearly adjust this parameter during the rounds, improving the effectiveness of both local search (LS) and global search (GS). Selecting the attributes with maximum FV is how the AF determines which features are best for the CK+ database & KMU-FED database.

Determine α by using the following calculation:

$$\alpha(t+1) = \left(1 - \frac{t}{MaxG}\right) \alpha(t) \quad (7)$$

Depending on the degree of distance deviation in the optimization process, α modifies its value to improve convergence speed and solution accuracy. Furthermore, it is updated as follows to improve the adaptability of the population.

$$\alpha = \alpha_{min} + (\alpha_{max} - \alpha_{min}) \times ||x_i - x_{best}||/L_{max} \quad (8)$$

Where $L_{max} = (x_{worst} - x_{best})$ (9)

Here, α_{max} and α_{min} represent the maximum and minimum features. The worst person in the generation stands t FF is represented by x_{worst} in Eq. (9). The distance between the worst and ideal individuals overall (x_{best}) is measured by L_{max}. In the initial phases of the procedure, the FF persons are dispersed throughout the entire region, with most of them situated far from the individuals who are idealised worldwide. At the present moment, L_{max} and $(\alpha_{max} - \alpha_{min})$ are constants, and the value of $\|x_i - x_{best}\|$ is greater. As a result, Eq. (8) indicates that the early stage has a higher value of α, which has a stronger total optimisation influence. In the algorithm, individual FF are drawn to FF that are brighter than themselves and closer to the global ideal features. Now that $||x_i - x_{best}||$ is smaller, selecting the best attributes for searches from the CK+ and KMU-FED databases is easier. When the time comes, "I" will bring the FF people near the world's most ideal individuals. As the position of the optimum is considered while varying the value of during each iteration, the approach converges quickly. The step size factor "α" is based on the previously described investigations. This α fluctuates adaptively and dynamically according to the distance between the FF individuals, balancing the capacity of algorithm development and search.

The creation of a unique fitness function (FitF) that takes ACC and execution time into account is part of this work.

$$f(x) = \frac{\left({}^{I_d}/_{I_t}\right) \times \left({}^{I_f}/_{P_{init}{}^{i}}\right)}{exp^{-{}^{e_E}/_{e_M} + H_{accuracy}}} \quad (10)$$

In this instance, I_d represents the number of attributes eliminated. In terms of the total number of characteristics, m_t provided more accurate results.

The features in dataset I are denoted by I_f.

$P_{init}{}^{i}$ is the main characteristic.

e_E denotes the execution time, whereas e_M represents the maximum allowed delay.

$$x_i = x_i + \beta_0 e^{-\gamma r^2}(x_j - x_i) + \alpha(rand - \tfrac{1}{2}) \quad (11)$$

Where x_i and x_j is the separation of two firefly characteristics

Each population characteristic's FV is determined. The quantity of traits in every batch is selected at random in the 1st generation. The FV of each FF is then founded. A selection procedure is then used to choose two FF. The next generation is formed by choosing the firefly with the highest fitness value and increased brightness.

Algorithm 1: Adaptive Firefly for facial FS

Input data: CK+ database & KMU-FED database

Output: Optimal emotion features

1. OF (x), x = (x1...,) T as an OF, take greater FER ACC
2. Generate the basic FF population xi (i = 1, 2, . . ., n)
3. Ii at xi is a measure of light intensity $f(xi)$
4. Define the coefficient of light absorption γ
5. while (t < Max Generation)
6. for i=1: n all n FF (facial features)
7. for j=1: i all n FF (facial features)
8. if (Ij > Ii), Move FF i towards j in d-dimension;
9. end if
10. The distance from an object impacts how attractive it is r via exp [$-\gamma r$]
11. Find fitness level by (10) and (11)

Generate a model that is objective by (6)
Use updated light intensity computations and novel solutions to (4)
Applying, reduce the unnecessary features (9)
Improve the best features by using (8)
end for *j*
end for *i*
The present top features of the FF are ranked
end while
An attractive FF changes its appearance.
Return optimal facial features

Algorithm 1 outlines that the AFO algorithm, based on fitness, achieves better facial expression recognition accuracy. This algorithm is employed to produce optimal facial expression results. In the AFO algorithm, fireflies are sorted according to their fitness values, with the best fitness values identifying the optimal fireflies. The optimal solutions are then put to the firefly pool, and the firefly is iterated further. In this research, characteristics with more accuracy are the focus of optimum feature selection utilizing the AFO algorithm. After being applied to the AFO, the characteristics that are extracted from the test dataset are associated with the features of the CK+ database & KMU-FED database. If maximum brightness is reached, the test dataset is classified as having best facial expression features; otherwise, if brightness is minimal, the test input is categorized as non- expression feature

3.5. FER using Ensemble Machine Learning (EML) algorithm

There are two sections to the processed facial feature. One is used for testing, and the other for training. EML classifiers are used for ED since emotions can be classified into seven types. FEER is performed using Ensemble Machine Learning (EML) algorithm such as Enhanced Artificial Neural Network (EANN), RF and KNN algorithm. EML provides faster convergence time during training and testing process. It is mainly used to classify the accurate FEER results for the given database.

EANN algorithm

In this study, FEER is performed more efficiently utilising the EANN method for the provided CK+ and KMU-FED databases. The input layer (IL), hidden layer (HL) and output layer (OL) are the three stages of an ANN, which learns to collect information. In order to produce 'n' inputs, the IL gathers and processes the features of the input data. These processes adhere to a set of weights. Weights are the information used to solve neural network problems [19]. The hidden data is transferred from the IL to the OL following some useful hidden extraction. The facial expression emotion dataset is classified in this case via EANN. The chosen dataset of face features is trained using EANN, and during testing, the features are categorised by state. The EANN is a combination of an ANN and an MLP (Multilayer Perceptron). The configuration of the ANN is presented in Fig. 4.

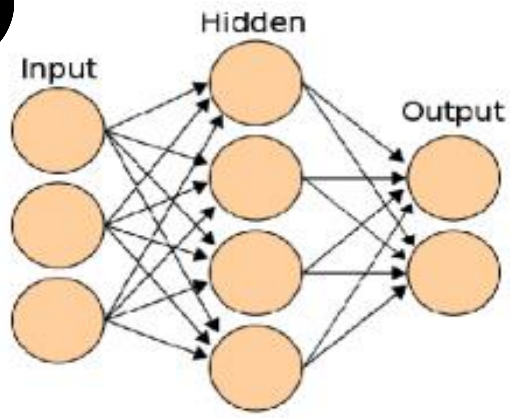


Fig 4 Architecture of ANN

IL - The selected features of the provided facial expression emotion databases are transmitted by the network's IL. This content appears to be somewhat undeveloped at first.

HL - Primary functions of these layer is to transform raw inputs into decidable outputs where EANN architecture may encompass multiple HL.

OL - After receiving data from the HL, the OL processes it to get the intended results (faster execution and a better facial emotion classifier ACC).

The most popular FNN model is the MLP FNN (Feedforward NN architecture), which arranges neurons in a cascade manner. MLP consists of at least two layers. In MLPs, inputs to neurons of i+1th levels are outputs of i^{th} layers, rather than information being transferred between the neurons that comprise a layer. Both counts of nodes in input layers and counts of nodes in output layers correspond to counts of features included in input vectors.

$$Y_n = f(\sum_{m=1}^{h}(w_{nm}, f(\sum_{l=1}^{i} v_{ml}X_l + \theta_{vm}) + \theta_{wm} \quad (12)$$

$$n = 1, \ldots\ldots, o$$

Here, the OL of the nth node is denoted as Y_n. The inputs of the l^{th} nodes in IL are supplied by X_l. Connective weights between m nodes in HL and n OL are represented by w_{nm}. Connective weights between nodes l in IL are represented by v_{ml}, while m denotes HL. The bias terms or thresholds of the transfer functions f of m nodes in HL and n OL are denoted by θ_{vm} and θ_{wm}.

The perceptron model in an EANN transmits the output 1 if the weighted sum of the inputs exceeds a programmable threshold value, otherwise referred to as an activation function (AF). The weighted sum of each neuron's inputs, including bias, is its output. "*w*" and "*x*" stand for the input neuron and weights, respectively.

$$\sum_{i=1}^{m} bias + (w^i x^i) \quad (13)$$

AF employs the Sigmoid function (SF), one of the specified functions.

$$f(x) = sigmoid = \frac{1}{1+\exp(-x)} \quad (14)$$

Each neuron's bias terms and connection weights combine to form the network weights. The most effective method of obtaining the intended output from the input is believed to be "NN training," which involves updating the network weights and figuring out the appropriate weights and biases values.

Algorithm 2: EANN for FEER

Input: Selected features (given CK+ database & KMU-FED database) are used as input.
Output: Improved FEER performance

1. EANN procedure (input, neurons, repeat)
2. Make an input database.
3. Input←database with all probable combinations
4. Execute training and testing via EANN
5. Do for input = 1 to end of input
6. Do for neurons = 1 to n
7. Do for repeat = 1 to n
8. Train EANN-storage ←store value with maximum ACC FEX features
9. End for
10. End for
11. EANN-storage←store best prediction of EANN based on inputs
12. End for
13. Return EANN-storage →Outcome with best classification of EANN for all facial feature combination

KNN Approach

In this work, KNN algorithm is introduced for FEER. The fundamental idea of KNN is to categorize a data point according to the facial feature space class labels of its closest neighbours. The parameter "k" is user-defined and sets the maximum number of nearest neighbours to be considered. The KNN algorithm is an effective tool for FEER. KNN functions based on the proximity principle, which states that a case's severity is decided by the facial feature space closeness of its closest neighbours. Genetic variations and clinical markers are examples of qualities that each case possesses, and these features form the foundation for comparing cases in terms of similarity [20]. KNN is capable of accurately classifying new instances into several severity categories based on the severity levels of nearby cases. This makes KNN an effective option for situations where subtle patterns could be presented.

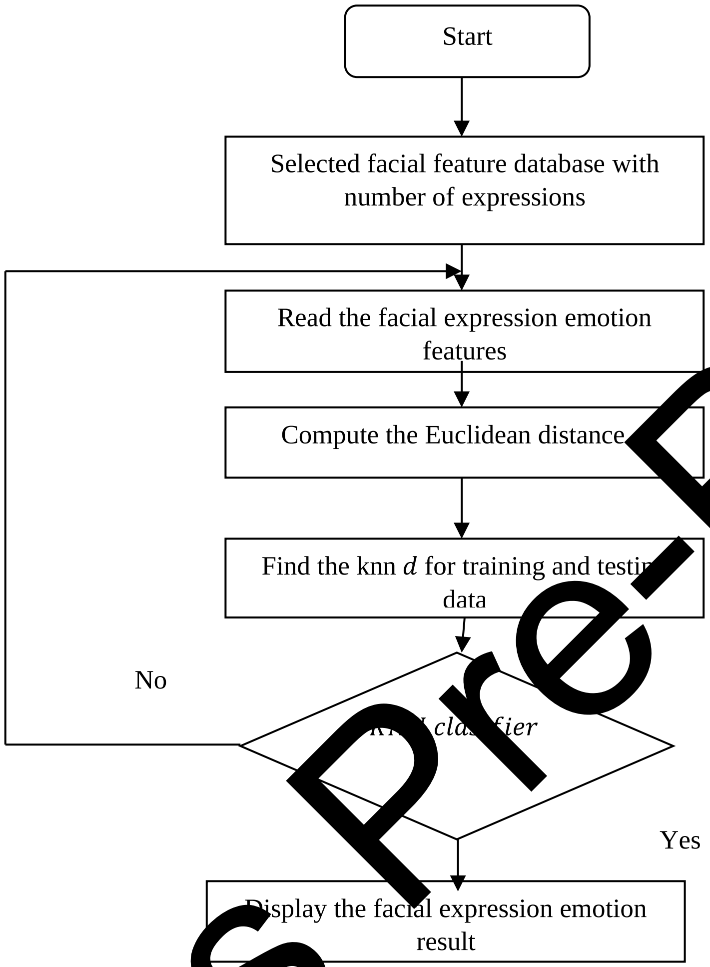


Fig 5 Flow diagram of KNN for CK+ database & KMU-FED database

To improve prediction performance, the parameter "k," which denotes the number of nearest neighbours taken into account, needs to be carefully adjusted. For facial expression emotion recognition, KNN is a promising method that, when used in given database, could contribute to better results. For the facial expression dataset, the below given hamming distance formula is used to compute the similarity score expressed in Equation (15).

$$D_H = \sum_{i=1}^{k} |x_i - y_i| \tag{15}$$

Where, $x = y$ implies $D = 0$
$x \neq y$ implies $D = 1$

The expression $|x_i - y_i|$ calculates the absolute difference between the corresponding components of the two vectors for each index i. This is done for all k components and summed up these absolute differences to get the Manhattan distance between the vectors.

If the values match, CK+ database & KMU-FED database features are expression with emotions; otherwise, they are not. After this step, the accuracy rate and error rate of the dataset are calculated. The accuracy rate indicates how many outputs from the test dataset align with the outputs of the training dataset with different features. The error rate, on the other hand, shows how many outputs from the test dataset do not match the corresponding outputs from the training dataset with varying features. Training phase: the method saves the training samples' attributes and the associated class labels. Classification phase: Depending upon the value of "k," the algorithm classifies the unlabelled test sample. The facial expression recognition is determined by calculating the similarity of features, and the final decision is made through a majority voting process

Random Forest (RF)

A potent ensemble learning method called RF is intended to improve FEER accuracy while reducing the possibility of overfitting. A random selection of training data and characteristics is employed to train each decision tree (DT), which forms the basis of a random forest model. To ensure that each tree captures a distinct facet of the underlying data distribution, this randomness is injected into the trees. An accurate and reliable result is produced during prediction when the RF aggregates the predictions of several distinct trees via voting (for classification tasks) or averaging (for regression tasks).

RF provide a reliable and efficient method for classifying activities when employed to forecast the facial expressions from CK+ database & KMU-FED database. The RF introduces variation and captures numerous aspects of the underlying data distribution by training each DT on a random subset of the training data and characteristics [21]. In facial expression emotion recognition process, the RF employs a ranking procedure to aggregate the forecasts of each tree. The severity level that is most reliably expected across all trees, that is assigned to the final result.

The number of DT generated over a subset of training data is combined using the bagging technique known as RF.To avoid one strong predictor being used by all DT and to give a chance for all the predictors a search for the splitting attribute is limited to a random subset m of the p attributes which is given as m=$\sqrt{p}$. Random Forest for the CK+ database & KMU-FED is constructed following the below algorithm.

Algorithm 3: ***Random Forest (D, N, d)***

Input: *CK+ database & KMU-FED database; Ensemble size N, subspace dimension d*

Output: *Constructed Random Forest*

$for\ i = 1\ to\ N\ do$

Take random samples $D_i\ from\ D$ data points by sampling with replacement

Randomly select the facial features of D_i and decrease the dimensionality of D_i

Construct a decision tree model M_i on D_i

End

$Return\ \{M_i | 1 <= i <= N\}$

Finally, the best facial expression emotion results are aggregated using more accurate features which increase the emotion recognition performance for the given two datasets.

4. Experimental result

CK+ and KMU-FED are the two FE related databases with the most images that are used in DFEER-relevant analysis. The objective of this journal is to ascertain the driver's FEs, which sets it apart from other investigations. A near-infrared (NIR) camera and the driver's FE are captured in a real-world driving scenario and are obtained from the KMU-FED dataset of images. An Intel Core i7 processor with 8 GB of RAM and Microsoft Windows 10 is used for each experiment. There are more than 100 iterations in this test.

4.1. Reconstruction Methods Comparison

The outcomes of the ACC and error (E) of expression identification on reconstructed optical flows for different occlusions are displayed in Tables 1 and 2, which contrast the proposed AFO-EML method with other previous methods like AE, PMVO, DQN-BSOA of the CK+ dataset. The proposed AFO-EML algorithm decreases error for MSE, Wing, and (EP) Endpoint (LF) loss functions (EPLF) for Eyes occlusion (OCC) by 1.6[illegible]%, 3.[illegible]1% and 5.77, respectively, when compared to all occlusions.

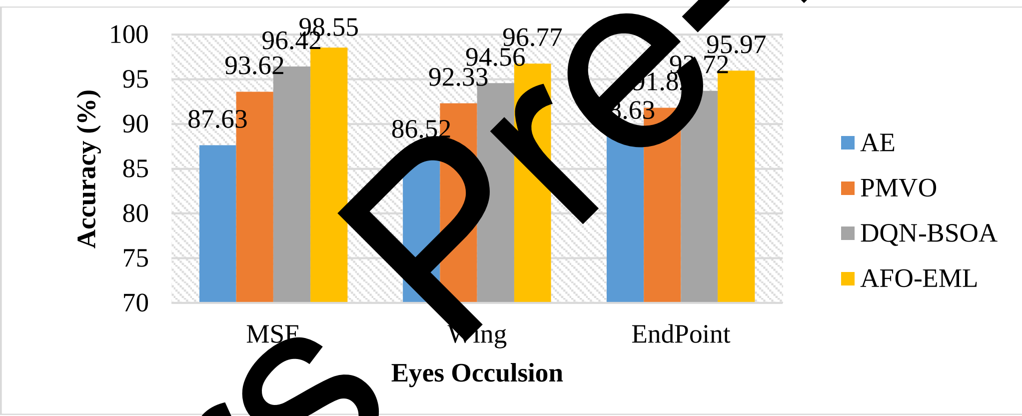
TABLE 1. RECONSTRUCTED OPTICAL FLOW ACCURACIES VS. DIFFERENTIAL LOSSES FOR (CK+ DATASET)

LF	Eyes occ. (%)				occ. (%) of Mouths				occ. (%) of Lower parts			
	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML

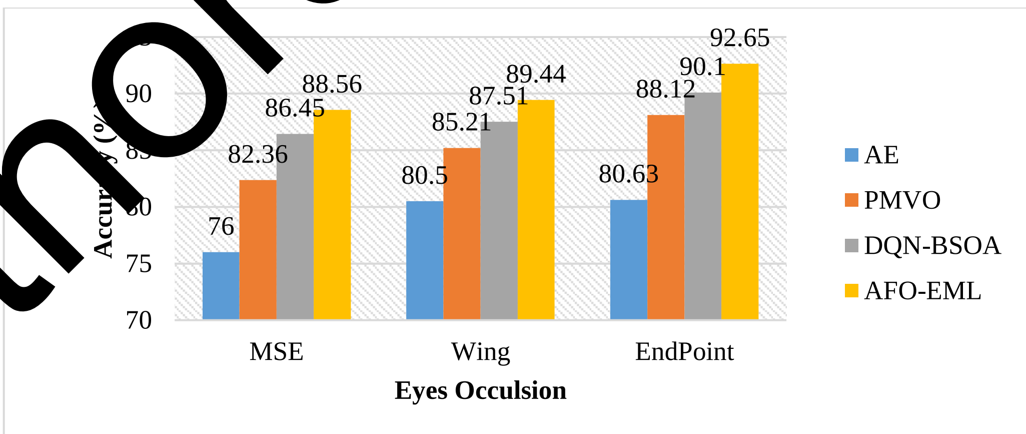
MSE	87.63	93.62	96.42	98.55	76.00	82.36	86.45	88.56	69.20	75.36	79.21	82.52
Wing	86.52	92.33	94.56	96.77	80.50	85.21	87.51	89.44	70.82	79.21	82.41	84.32
EP	88.63	91.82	93.72	95.97	80.63	88.12	90.10	92.65	71.26	82.83	87.64	89.88

TABLE 2. RECONSTRUCTED OPTICAL FLOW ERRORS VS. DIFFERENTIAL LOSSES FOR (CK+ DATASET)

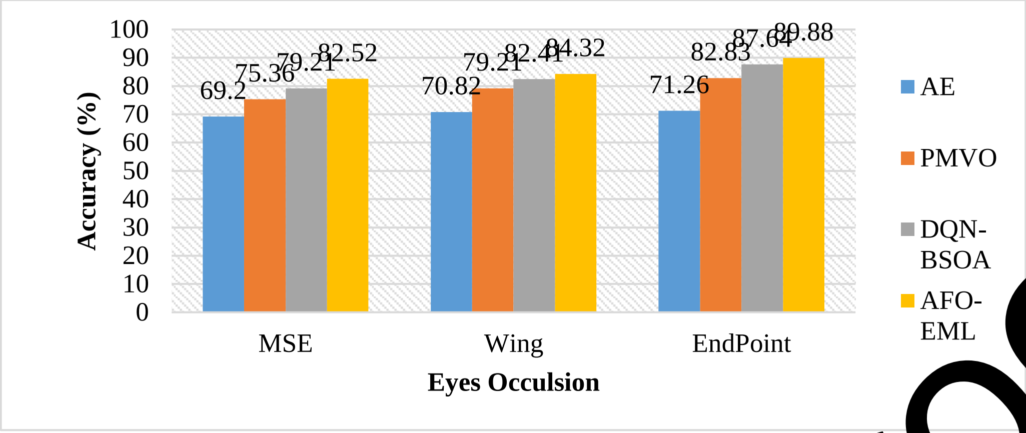
LF	**Eyes occ. (%)**				**occ. (%) of Mouths**				**occ. (%) of Lower parts**			
	AE	**PMVO**	**DQN-BSOA**	**AFO-EML**	**AE**	**PMVO**	**DQN-BSOA**	**AFO-EML**	**Flownet**	**PMVO**	**DQN-BSOA**	**AFO-EML**
MSE	12.37	6.38	3.58	1.66	24.00	17.64	13.55	11.47	30.80	[illegible]	[illegible]	18.11
Wing	13.48	7.67	5.44	3.51	19.50	14.79	12.49	10.63	[illegible]	[illegible]	17.59	15.75
EP	11.37	8.18	6.28	5.77	19.37	11.88	9.90	7.77	28.74	[illegible]	12.36	10.04



(a) Accuracy comparison of Eyes Occlusion



(b) Accuracy comparison of Mouth Occlusion



(c) Accuracy comparison of Lower Part Occlusion

FIGURE 6. COMPARATIVE ACCURCIES VS. DIFFERENT LOSSES FOR (CK+ DATASET)

Figure 6 (a-c) display the ACC outcomes of comparing three distinct OCC using three different reconstruction techniques and loss functions (CK+ dataset). The results show that the proposed AFO-EML rebuilt technique improves end point loss function accuracy by 95.97%, 92.65%, and 89.88% for lower part, mouth, and eye occlusions. The proposed AFO-EML rebuilt procedures show the accuracy gains of 98.55%, 88.56%, and 82.52% for MSE, Wing, and EPLF with eye OCC, as shown in Figure 6(a). When occlusion of the eyes occurs, MSE loses function Figure 6(c) demonstrates that the recommended AFO-EML rebuilt approach has a greater accuracy of 96.77%, 89.44% and 84.32% than the previous AE, PMVO and DQN-BSOA algorithms.

TABLE 3. RECONSTRUCTED OPTICAL FLOW ACCURACIES VS. DIFFERENTIAL LOSSES FOR (KMU-FED DATASET)

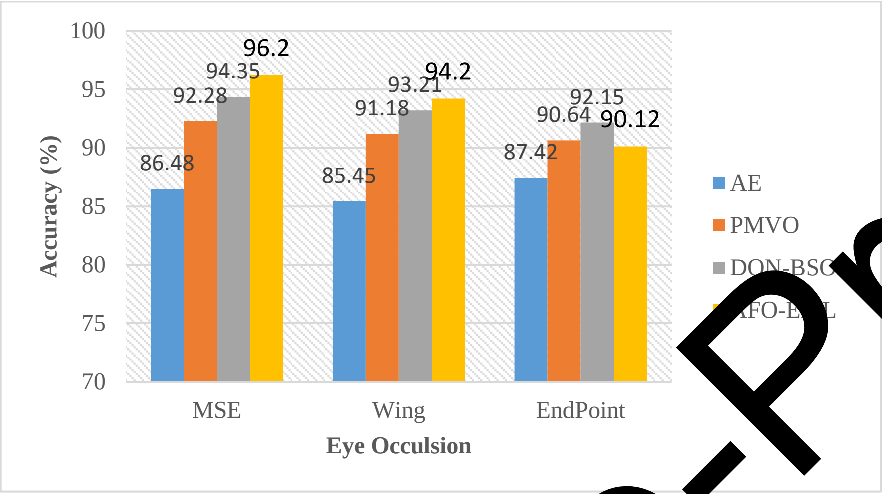
LF	Eyes occ. (%)				occ. (%) of Mouths				occ. (%) of Lower parts			
	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML
MSE	86.48	92.28	94.35	[illegible]	[illegible]	80.94	82.84	84.54	67.45	74.34	77.21	81.23
Wing	85.45	91.18	93.21	94.20	79.35	83.81	87.71	89.98	69.64	78.00	81.70	83.56
EndPoint	87.42	90.64	92.15	93.12	79.31	86.87	89.71	91.11	70.05	81.91	83.41	86.45

TABLE 4. RECONSTRUCTED OPTICAL FLOW ACCURACIES VS. DIFFERENTIAL LOSSES FOR ERRORS (KMU-FED DATASET)

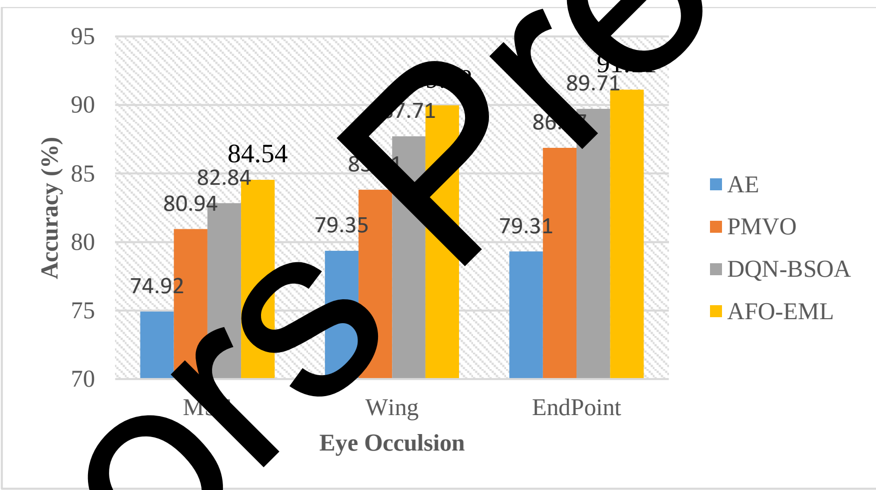
LF	Eyes occ. (%)				occ. (%) of Mouths				occ. (%) of Lower parts			
	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML	AE	PMVO	DQN-BSOA	AFO-EML
MSE	[illegible]	7.72	5.65	3.87	25.08	19.06	17.16	15.35	32.55	25.66	22.79	19.21
Wing	14.55	8.82	6.79	4.44	20.65	16.19	12.29	8.96	30.36	22.00	18.30	14.01
EndPoint	12.58	9.36	7.85	5.15	20.69	13.13	10.29	7.25	29.29	18.09	16.59	12.43

The evaluation of expression recognition accuracy and error on recovered optical flows for different occlusions is presented in Tables 3 and 4, along with a comparison with the findings of previous algorithms of DQN-BSOA, PMVO, AE of the KMU-FED dataset. These tables show how flexible the suggested approach is in this particular situation. Only the results with this loss are presented in the tests that follow since the EP loss is repaired. The suggested AFO-EML approach yields an error reduction of 3.87%, 4.44%, and 5.15% for MSE, Wing, and EPLF for Eyes OCC when compared to all occlusion.

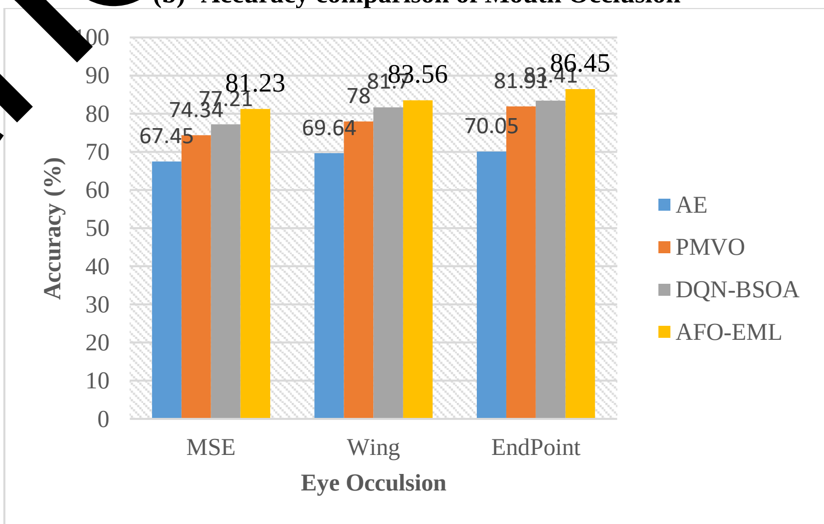
Figure 7 (a-c) display obtained accuracies of comparisons from three different occlusions using three different reconstruction techniques and loss functions (KMU-FED dataset). The results show that the EPLF having lower part occlusions is more accurate when using the proposed AFO-EML technique by 90.12%, 93.11%, and 86.45%, respectively. Figure 7(a) shows that the suggested AFO-EML technique improves accuracy by 96.2%, 84.54%, and 81.13%, respectively, for MSE loss function with eye occlusion.



(a) Accuracy comparison of Eye Occlusion



(b) Accuracy comparison of Mouth Occlusion



(c) Accuracy comparison of Lower Part Occlusion

FIGURE 7. RECONSTRUCTED OPTICAL FLOW ACCURACIES VS. DIFFERENTIAL LOSSES FOR (KMU-FED DATASET)

4.2. Recognition Methods Comparison

The performance metrics are such as P, R, F1-score, ACC, and error compared with previous DQN-BSOA, PMVO, AE and proposed AFO-EML algorithms for CK+ and KMU-FED databases

TABLE 5. COMPARATIVE RESULTS OF CK+DATASET

Approaches	**P**	**R**	**F1-Score**	**ACC**	**E**
AE	87.28	88.26	87.77	91.72	8.28
PMVO	89.82	91.45	90.63	92.18	7.82
DQN-BSOA	92.71	92.41	92.56	93.47	6.53
AFO-EML	93.57	93.32	94.60	95.78	4.2[illegible]

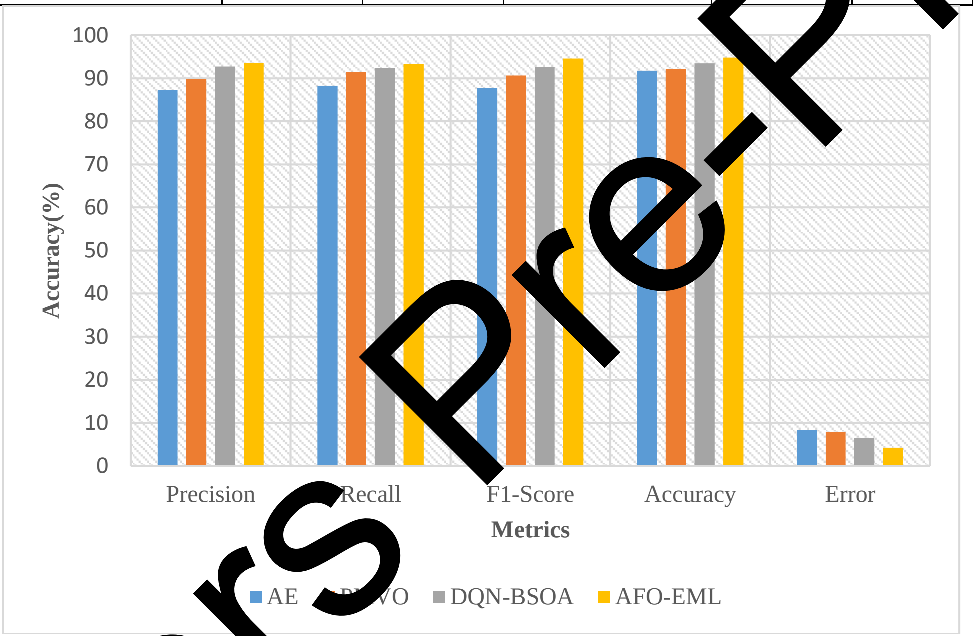


FIGURE 8. COMPARATIVE VALUES OF METRICS VS RECOGNITION TECHNIQUES ON (CK+ DATASET)

Figure 8 evaluates the performance of classifiers using metrics like precision, recall, F1-score, and accuracy previous DQN-BSOA, PMVO, AE and proposed AFO-EML algorithms for CK+ database. Based on the results, the AFO-EML classifier has the best accuracy (95.78%), whereas previous algorithm provides lower accuracy.

TABLE 6. COMPARATIVE RESULTS FOR KMU-FED DATASET

Approaches	**P**	**R**	**F1-Score**	**ACC**	**E**
AE	83.25	85.41	84.33	88.15	11.85
PMVO	84.62	87.16	85.89	92.45	7.55
DQN-BSOA	88.18	90.40	89.29	93.21	6.79
AFO-EML	90.21	91.82	91.01	94.49	5.51

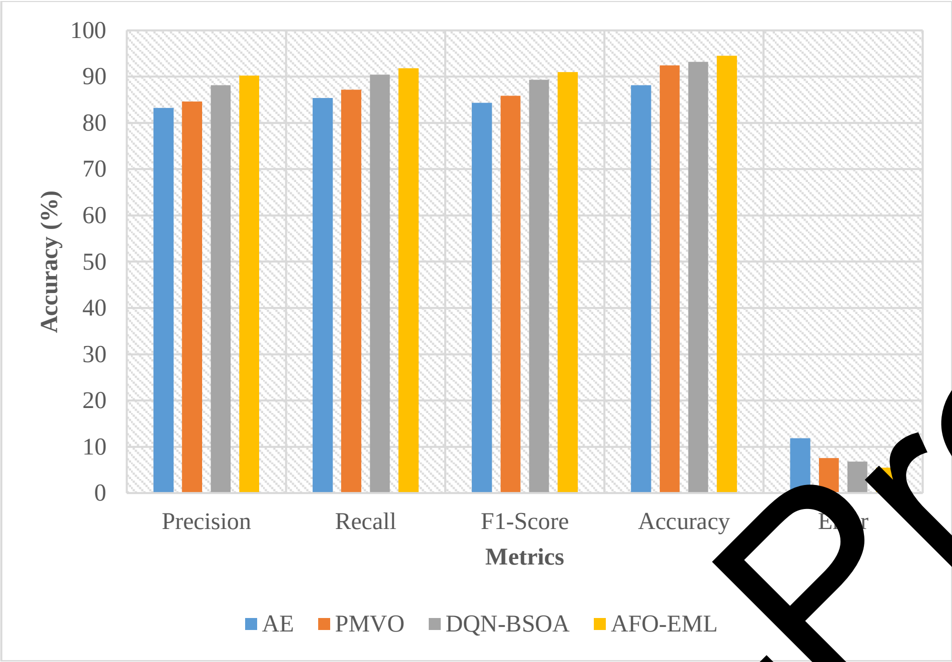


FIGURE 9. COMPRATIVE VALUES OF METRICS VS RECOGNITION TECHNIQUES ON (KMU-FED DATASET)

Figure 9 shows the outcomes of comparing the efficiency of ACC, P, R, and F1-score using previous DQN-BSOA, PMVO, AE and proposed AFO-EML algorithms for KMU-FED database. Table 6 shows that the existing algorithms provide lower accuracy and the proposed AFO-EML classifier produces the highest accuracy of 94.49% which improves the facial expression emotion recognition performance significantly

5. Conclusion

In this work, AFO and EML algorithm is proposed for facial expression emotion recognition. In this work, initially, dataset is collected using CK+ database and KMU-FED database. Occlusions around mouths and eyes are duplicated. Feature selection is done by using AFO algorithm, that can be employed for selecting more relevant and redundant attributes from the given CK+ database and KMU-FED database database. It generates best fitness values using objective function for higher recognition accuracy. Facial expression emotion recognition is performed using EML algorithm such as EANN, RF and KNN algorithm. It is focused to classify the accurate facial expression emotion recognition results for the given CK+ database and KMU-FED database. From the result, it is clear that the suggested AFO-EML procedure delivers superior efficiency by ACC, P, R and f-measure than the existing DQN-BSOA, PMVO, AE algorithms

References

1. Hussain, Shaik Asif, and Ahlam Salim Abdallah Al Balushi. "A real time face emotion classification and recognition using deep learning model." *Journal of physics: Conference series*. Vol. 1432. No. 1. IOP Publishing, 2020.
2. Wang, Xusheng, Xing Chen, and Congjun Cao. "Human emotion recognition by optimally fusing facial expression and speech feature." *Signal Processing: Image Communication* 84 (2020): 115831.
3. Abdulrazaq, Maiwan B., et al. "An analytical appraisal for supervised classifiers' performance on facial expression recognition based on relief-F feature selection." *Journal of Physics: Conference Series*. Vol. 1804. No. 1. IOP Publishing, 2021.
4. An, Heng-Yu, and Rui-Sheng Jia. "Self-supervised facial expression recognition with fine-grained feature selection." *The Visual Computer* 40.10 (2024): 7001-7013.
5. Donuk, Kenan, et al. "Deep feature selection for facial emotion recognition based on BPSO and SVM." *Politeknik Dergisi* 26.1 (2023): 131-142.
6. Aslam, Tanveer, et al. "Emotion based facial expression detection using machine learning." *Life Science Journal* 17.8 (2020): 35-43.
7. Ivanova, Ekaterina, and Georgii Borzunov. "Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions." *Procedia Computer Science* 169 (2020): 244-248.

8. Liu, Zhentao, et al. "A facial expression emotion recognition based human-robot interaction system." *IEEE CAA J. Autom. Sinica* 4.4 (2017): 668-676.
9. Cui, Rongxuan, Wanzhong Chen, and Mingyang Li. "Emotion recognition using cross-modal attention from EEG and facial expression." *Knowledge-Based Systems* 304 (2024): 112587.
10. Tao, Xue, et al. "Facial video-based non-contact emotion recognition: A multi-view features expression and fusion method." *Biomedical Signal Processing and Control* 96 (2024): 106608.
11. Lu, C., Zheng, W., Li, C., Tang, C., Liu, S., Yan, S. and Zong, Y., 2018, Multiple spatio-temporal feature learning for video-based emotion recognition in the wild. In Proceedings of the 20th ACM international conference on multimodal interaction, pp. 646-652
12. Pu, Xiaorong, et al. "Facial expression recognition from image sequences using twofold random forest classifier." *Neurocomputing* 168 (2015): 1173-1180.
13. Bhatti, Y.K., Jamil, A., Nida, N., Yousaf, M.H., Viriri, S. and Velastin, S.A., 2021, Facial expression recognition of instructor using deep features and extreme learning machine. Computational Intelligence and Neuroscience, pp.1-17
14. Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., & Matthews, I. (2010, June). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. *In 2010 IEEE computer society conference on computer vision and pattern recognition-workshops* (pp. 94-101).
15. KMU-FED. Available online: http://cvpr.kmu.ac.kr/KMU-FED.htm (accessed on 4 December 2018)
16. Allaert, B., Ward, I. R., Bilasco, I. M., Djeraba, C., &Bennamoun, M. (2019). Optical Flow Techniques for Facial Expression Analysis--A Practical Evaluation Study. arXiv preprint arXiv:1904.11592
17. Xu, Huali, et al. "An improved ReliefF algorithm for feature selection in classification." *Wireless Personal Communications* 103 (2018): 2823-2834.
18. Xie, Weidong, et al. "Improved multi-layer binary firefly algorithm for optimizing feature selection and classification of microarray data." *Biomedical Signal Processing and Control* 79 (2023): 104080.
19. Verma, Kunika, and Ajay Khunteta. "Facial expression recognition using Gabor filter and multi-layer artificial neural network." *2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC)*. IEEE, 2017.
20. Afriansyah, Yudha, Ratna Astuti Nugrahaeni, and Anggunmeka Luhur Prasasti. "Facial expression classification for user experience testing using K-nearest neighbor." *2021 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, 2021.
21. Munasinghe, M. I. N. P. "Facial expression recognition using facial landmarks and random forest classifier." *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*. IEEE, 2018.
22. Qadir, I., Iqbal, M. A., Ashraf, S., & Akram, S. (2025). A fusion of CNN And SIFT For multicultural facial expression recognition. *Multimedia Tools and Applications*, 1-19.
23. Sudha, S. S., & Suganya, S. S. (2023). On-road driver facial expression emotion recognition with parallel multi-verse optimizer (PMVO) and optical flow reconstruction for partial occlusion in internet of things (IoT). *Measurement: Sensors*, 26, 100711.