Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 01, PP. 141-153, 2025
DOI: https://doi.org/10.54216/JISIoT.140111

Quasi Oppositional Jaya Algorithm with Computer Vision based Deep Learning Model for Emotion Recognition on Autonomous Vehicle Drivers

Rajesh .D[1], S. Thenappan[2,*], Prachi Juyal[3], Thiyagarajan .V .S[4], D. M. Kalai Selvi[5], J. Rajeswari[6], M. Hema Kumar[7], V. Saravanan[8]

[1]Faculty in DICT&TD, New Delhi, India
[2]Department of Electronics and Communication Engineering, Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India
[3]Department of Allied Science (Mathematics), Graphic Era (Deemed to be University), Dehradun, Uttarakhand, India
[4]Department of CSE, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, Tamilnadu, India
[5]Department of CSE, R. M. D Engineering College, R.S.M Nagar Kavaraipettai, 601206, Chennai, India
[6]Department of ECE, Agni College of Technology, Thazhambur, Chennai, India
[7]Department of ECE, Sona College of Technology, Salem, TamilNadu, India
[8]Department of Nano Electronics Materials and Sensors, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai-602105, Tamilnadu, India

Emails: extrajesh@gmail.com; drthenappans@veltech.edu.in; Prachijuyal@yahoo.com; thiyagu.cse86@gmail.com; dmkalai@gmail.com; rajeswari.ece@act.edu.in; hemakumarbeece@gmail.com; saravananv.sse@saveetha.com

Abstract

Facial emotion recognition (FER) technology in autonomous vehicle drivers can considerably strengthen the efficiency and safety of the driving experience. The system can analyze facial expressions in real-time by employing advanced computer vision (CV) techniques, which identify emotions such as stress, fatigue, or distraction. This enables the vehicle to adapt its behavior, triggering interventions or alerts where applicable to alleviate possible threats. Ensuring the emotional well-being of the driver promotes a safer road environment, improving overall road safety and diminishing the possibility of accidents in the era of autonomous vehicles. FER using (Deep Learning) DL is an advanced technique that leverages deep neural network (DNN) to automatically interpret and identify emotions from facial expressions. DL algorithms, especially Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) have attained outstanding results in this field since they allow us to learn temporal dependencies hierarchy and features within the data. This research develops a novel Computer Vision with Optimal DL-based Emotion Recognition (CVODL-ER) model for Autonomous Vehicle Drivers. The CVODL-ER method concentrates on the automated classification of various sorts of emotions of autonomous vehicle drives. To accomplish this, the CVODL-ER technique makes use of the SE-ResNet model for learning intrinsic patterns from the driver's facial images. Besides, the hyper parameter tuning of the SE-ResNet model takes place via a quasi-oppositional Jaya (QO-Jaya) algorithm. For the recognition of driver emotions, the CVODL-ER system executes the deep belief network (DBN) algorithm. The performance analysis of the CVODL-ER technique takes place using a benchmark facial image database. The obtained results underline the improved efficiency of the CVODL-ER technique over other models.

Received: February 15, 2024 Revised: April 25, 2024 Accepted: July 14, 2024

Keywords: Facial Emotion Recognition; Autonomous Vehicle Drive; Computer Vision; Jaya Algorithm; Deep Learning

1. Introduction

Emotions are the features that affect a driver's abilities near even the negative and positive sides of driving. To observe the emotions of driver's, face expression recognition (FER) technology has been employed to identify the facial expressions of the being [1]. Face expression study will have positive effects on the emergence of human-machine communication for secure driving performance and road security [2]. Conveying emotions commonly arises in dual communication approaches non-verbal and verbal communication. Verbal communication is simple for understanding and communicating among persons in many cases, while nonverbal communication like demonstration of sentiments, is complex to realize in any circumstances [3]. Camera capture is an example of the IoT devices in which various facial reactions of drivers for Autonomous driving methods. In entire independent vehicles, drivers have concern about takeover handovers, as they will be separated from certain factors of driving. Factors are affect efficiency like response time and the contribution of non-driving relevant challenges could be focused [4]. However, taking into account the crucial function of emotions, manual driving and human communication impact the driver's cognitive effectiveness. The identity of the face is very crucial for emotional sentiment identification to drivers in autonomous vehicles [5]. Automatic and intelligent face identification devices are extremely precise in an essential state and unrestricted, with poorer dependability in independent vehicles.

Modern technological developments like wearable devices have allowed the analysis of emotions in real-time sceneries, resulting in an increasing number of studies exploring the negative effect of specified emotions while driving (for example, sadness, anger, or worry) [6]. Facial expression is a major prevailing indication for humans to convey emotional conditions. Moreover, facial expression can be a simple one to achieve and necessitate only easy tools, and several research workers have analysed FER and accomplished efficient accuracy [7]. Furthermore, the noise and body movements than the fMRI or EEG signals will minimally affect the gathering of driver facial emotion data in driving. Therefore, facial expression-based emotion identification is a highly proper and appropriate emotional reaction identification for an automatic emotional human-machine technique [8]. Computer vision (CV)-based deep learning (DL) techniques have been widely implemented for emotion monitoring and FER. Experimental analysis of deep facial recognition recapitulates facial emotions databases and gathers surroundings like laboratories or the internet [9]. However, the present researches are frequently executed on lab-captured databases because of the lack of actual databases. There is no on-road driver facial database available for the driver FER task, and the driving tasks will defeat the facial emotions. Owing to this issue, a shortage of on-road driver FER studies that can be important for automatic human-machine systems [10].

This research paper develops a new Computer Vision with Optimal DL based Emotion Recognition (CVODL-ER) model for Autonomous Vehicle Drivers. The CVODL-ER method concentrates on the automated classification of numerous types of emotions the autonomous vehicle drives. To accomplish this, the CVODL-ER technique makes use of the SE-ResNet model for learning intrinsic patterns from the driver's facial images. Besides, the hyperparameter tuning of the SE-ResNet model takes place via quasi-oppositional Jaya (QO-Jaya) algorithm. Eventually, the CVODL-ER technique executes the deep belief network (DBN) approach for the detection of driver emotions. The performance analysis of the CVODL-ER technique takes place using a benchmark facial image database.

2. Related Works

Jain et al. [11] presented a new Squirrel Search Optimization with DL-assisted FER (SSO-DLFER) algorithm that follows 2 main methods such as emotion recognition and face detection. Primarily, the RetinaNet system was utilized. Secondarily, the SSO-DLFER method implemented the NASNet massive feature extractor with a gated recurrent unit (GRU) architecture. The SSO-based hyperparameter tuning method has been implemented. In [12], a face-sensitive convolutional neural network (FS-CNNs) has been designed that includes two phases, patch cropping, and CNNs. The primary phase could be utilized for identifying faces in higher-resolution imageries and producing the faces for more processing. The secondary phase defines CNNs that could be implemented for predicting facial expression dependent upon landmark analytics; it can be implemented under pyramid images to perform the scale invariance. In [13], an automatic model was designed for the driver's real emotion recognizer (DRER) employing a DL. Transfer learning (TL) has been executed in the NasNet huge CNN architecture. Besides, a custom driver emotion identification image database was presented.

Saranya et al. [14] provided the development of Automatic Facial Emotion Detection employing an Arithmetic Optimizer Algorithm with Deep-CNNs (AFED-AOADCNNs) method. The system implemented the DCNN technique for making the feature extractor. Subsequently, the AOA was employed for best hyperparameter tuning of the DCNNs technique. Lastly, the quantized neural networks (QNNs) technique was employed for recognizing and classifying the various types of FER. In [15], a multi-modal fusion-enabled FER method was developed, employing a structured light imaging camera that offers 3 categories of images - Depth Maps RGB, and Near-infrared (NIR). The system was executed in two stages wherein the primary phase removes features from particular

modalities individually by employing 3D-ResNet whereas the secondary stage integrates the multi-modal features and categorizes the emotions. Sahoo et al. [16] projected a DL-based FER system. This study employs a TL-based model for FER that can support in design of an in-vehicle driver assistance model. This employs TL SqueezeNet 1.1 for classifying various FERs. Data augmentation and data preprocessing methods like image resizing were used to increase the efficiency.

In [17], an innovative technique for FER was introduced. The technique was employed neural networks convolutionary (FERC). The FERC must be dependent upon a CNN model of 2 measures: the primary part extracted the backdrop of the image; and another portion extracted the face vectors. The expressional vector was applied in the FERC system. The double-level CNN will be constant as well and the weight and exponent values fluctuate with every iteration. Chand and Karthikeyan [18] proposed an innovative technique employing CNN accompanied by analyzing emotions. These driving patterns have been examined and are dependent upon the Revolutions per Minute (RPM), vehicle's speed, acceleration model, and FER of the drivers. The driver's facial pattern could be developed with 2D-CNN method for identifying the driver's expressions and behavior.

3. The Proposed Method

In this research, we have developed a novel CVODL-ER system for autonomous vehicle drivers. The CVODL-ER method concentrates on the automated classification of numerous types of emotions the autonomous vehicle drives. To accomplish this, the CVODL-ER technique comprises SE-ResNet-based feature extractor, QO-Jaya algorithm-based hyperparameter tuning, and a DBN-based classification process. Figure 1 shows the workflow of CVODL-ER technique.

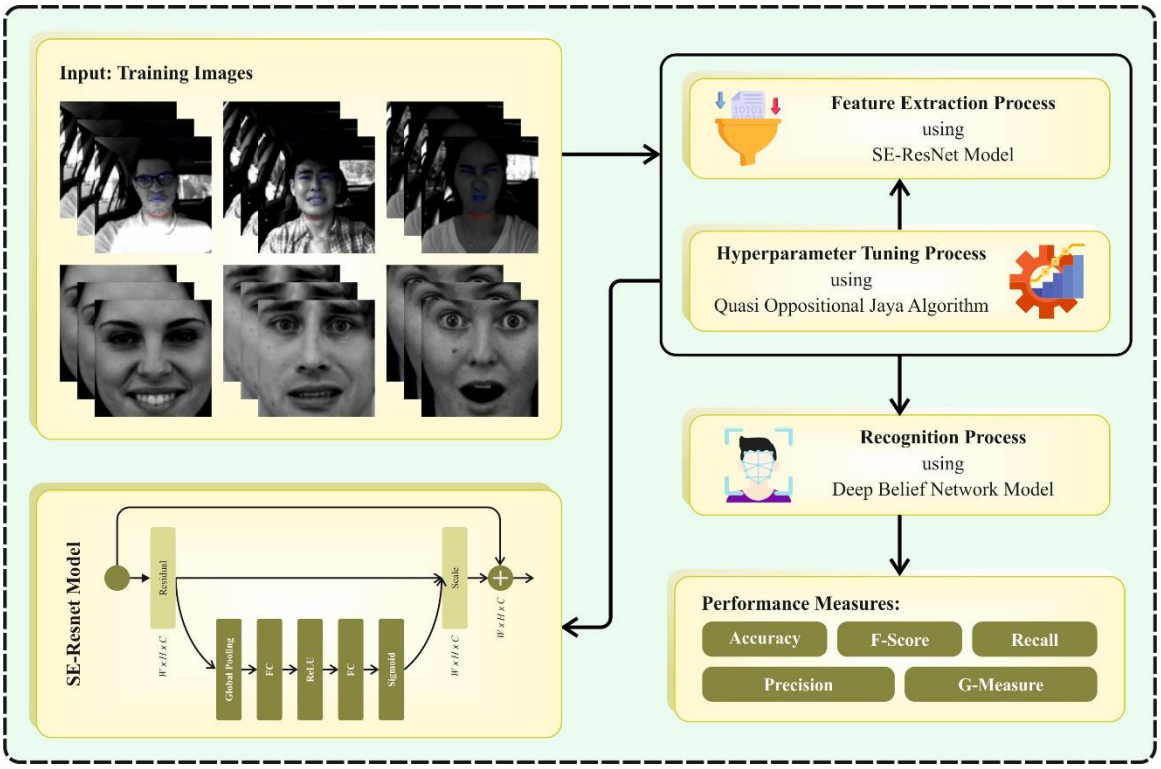

Figure 1. Workflow of CVODL-ER technique

A. SE-ResNet Feature Extractor

The CVODL-ER technique makes use of the SE-ResNet model for learning intrinsic patterns from the driver's facial images. ResNet is used to add shortcut association branch outside the convolution (Conv) layer to form the basic residual learning unit and to carry out constant mapping, and then resolves the performance degradation problems that are challenging to train once the CNN model is extremely deeper with stacked residual learning unit sequentially, enables to train deep CNN (DCNN) [19]. The basic residual learning unit either increases the computational complexity or introduces new parameters. The ResNet is expressed as:

$$x_l = h(x_l) + F(x_l, W_L) \quad (1)$$

$$x_{l+1} = f(y_1) \quad (2)$$

Where $f()$ indicates the activation function and $h()$ refers to the direct mapping.

The residual block has been represented by:

$$x_{l+1} = x_l + F(x_l, W_1). \tag{3}$$

The relationships among the l layers and the deeper L layers are:

$$x_L = x_l + \sum_{i=1}^{L-1} F\left(x_{i,} W_i\right). \tag{4}$$

The loss gradient function ε with esteem to χ_l in Eq. (5), according to the chain rule for the derivative applied in backpropagation,

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\left(1 + \frac{\partial}{\partial x_l}\sum_{i=1}^{L-1} F\left(x_{i,} W_i\right)\right) = \frac{\partial \varepsilon}{\partial x_L} + \frac{\partial \varepsilon}{\partial x_L}\frac{\partial}{\partial x_l}\sum_{i=1}^{L-1} F\left(x_i, W_i\right). \tag{5}$$

$\frac{\partial}{\partial x_l}\sum_{i=1}^{L-1} F\left(x_i, W_i\right)$ cannot be -1 all the time during the training, so it does not pose a gradient disappearance problem in the ResNet, and $\frac{\partial \varepsilon}{\partial x_l}$ indicates the grade of L layer passed openly to any l layer. It does not need deep network, the amount of samples in the data reduces and after the conversion of mathematical features to image features, so ResNet18 is adopted in this study.

The extracted features through CNN by stacked Conv layer are represented as high-dimensional features. The ResNet residual block skips the feature extractor through Conv layer and combines the feature before n layers, with the Conv feature after n layers, so that higher and lower features of dimensional can be retained, and the performance model can be enhanced. In addition, global average pooling is used for replacing the full connection (FC) layer in the traditional classical CNN, it reinforces the correlation between categories, and feature maps on the FC layer, which is superior suited for Conv model. Moreover, there is no parameter to be enhanced in the global pooling, which prevents over-fitting. Regarding the spatial transformation of input, global average pooling is more robust and aggregates spatial data. SE-ResNet focuses on the interdependency among the feature channel convolved by 1D Conv. The SE block is accomplished by using a squeeze function that summarizes the available data and an excitation operator used to scale the importance of the feature map. Thus, the squeeze operator only removes significant data, and the excitation operator estimates the dependency between networks with the FC layer using non-linear function.

B. Hyperparameter tuning using QO-Jaya Algorithm

In this phase, the hyperparameter tuning of the SE-ResNet algorithm takes place via the QO-Jaya algorithm. Similar to the general heuristic method, the Jaya algorithm has popular control parameters like generation number, size of population, and so on [20]. The initial population will be produced arbitrarily in the series of variables. But, this varies from alternative heuristic methods in that the Jaya algorithm has only one upgrade function. This was more rapid for upgrading the population. From each generation, the population updates dependent upon Eq. (6):

$$W'_{g,p,j} = W_{g,p,j} + r_{1,g,j}\left(W_{g,best,j} - \left|W_{g,p,j}\right|\right) - r_{2,g,j}\left(W_{g,worst,j} - \left|W_{g,p,j}\right|\right), \tag{6}$$

$$g = 1,2,\ldots,G; p = 1,2,\ldots,Pop;\ j = 1,2,\ldots,n,$$

Now, Pop refers to the population size, G refers to the max generation amount; p represents the pth individual, $and\ g$ denotes the gth generation. n denotes the variables count; j represents the jth variable. In Eq. (7), $r_{1,g,j}$ and $r_{2,g,j}$ defines the random number in [0, 1], $W'_{g,p,j}$ is the upgraded new individual; $W_{g,p,j}$ is the previous individual. $W_{g,worst,j}$ and $W_{g,best,j}$ have been correspondingly the worst and best solutions, respectively. $\left(W_{g,best,j} - \left|W_{g,p,j}\right|\right)$ and $\left(W_{g,worst,j} - \left|W_{g,p,j}\right|\right)$ describe the trend of the solution to the best and worst solution, correspondingly. During the operation, the random number performs as a scaling parameter to confirm the better exploration effectiveness. The worst and best solutions assure the optimistic way. At the final iteration, each satisfactory result should be maintained.

Meanwhile, in the VBHF optimizer, we investigate a multi-objective complexity, the Jaya algorithm requires any approaches included to transform to the multi-objective algorithm of Jaya. The non-dominance categorizations, notions of constraint-dominance, and crowding distance calculation have been employed for identifying the position of the solution. These values are the key factors for directing the Pareto frontier.

The comprehensive solving stages of the multi-objective Jaya algorithm are defined as given below.

Step 1: Pop solutions have been created arbitrarily as an initial populace. It will be organized dependent on the non-dominance and constraint-dominance principles.

Step 2: Primarily, the constraint-dominance was employed for first determining the dominance between solutions. Next, the non-dominance and crowding distance have been executed for additional determining the significance of the solution. The outcome with a top position will be higher than the alternative one. If the solution has a similar position, then the performance with a greater crowding distance will be measured than the other.

Step 3: Select the solution with the lowest position as the worst solution and with the top rank (rank =1) as optimum one. Next, the subsequent generation result will be upgraded by Eq. (7).

Step 4: Later the performances are upgraded, and the novel solutions integrate with the previous result as 2 Pop. Rearrange such solutions by the values of non-dominance, constraint-dominance categorizing, and crowding distance calculation. Choosing the Pop solution as the novel populace depends on the novel categorization.

Step 5: Go to Step 3 in order to upgrade generation until the stop principle to fulfil.

The notion of constraint-dominance promises that possible solutions have a superior position than infeasible solutions. In such a possible solution, the higher solution positions are greater than the dominated solution. From the impossible solution, a greater position has been assigned to the solution with decreased complete constraint conflicts.

QO-Jaya algorithm describes a QO-based Jaya system that will include a notion of opposition-based learning. A populace contrary to the existing populace should be produced for more expand the populace and increase the convergence rate of Jaya. We enhance the QO populace generation approach for higher realism in encoding. Eqs will produce the reverse populace. (7)-(9):

$$a = \frac{W_j^L + W_j^U}{2}, \tag{7}$$

$$b = W_j^L + W_j^U - W_{g,p,j}, \tag{8}$$

$$W_{g,p,j}^q = a + (b - a) * r_3, \tag{9}$$

Here a refers to the middle point of the variable, b defines the mirror point, W_j^L and W_j^U denote the upper and lower limits, $W_{g,p,j}$ describes the existing population, and $W_{g,p,j}^q$ means the opposite populace, r_3 denote the randomly generated number at $[0\ and\ 1]$.

The QO-Jaya technique develops an FF to achieve improved classifier accuracy. It expresses a positive number to epitomize the higher efficiency of the candidate outcomes. At this point, the classifier error rate minimization has been regarded as FF.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{No.of\ misclassified\ instances}{Total\ no.of\ instances} * 100 \tag{10}$$

C. Emotion Detection using DBN Classification

Finally, the CVODL-ER technique applies the DBN model for the recognition of driver emotions. As a representative DL model with the remarkable ability of feature expression, DBN has been used extensively in natural language processing, fault detection, and image recognition [21]. The classical DBN includes multiple hidden layer (HL), an input layer, and an output layer, where the neuron from the similar layer is non-connected and the neuron from the adjacent layer is fully connected (FC).

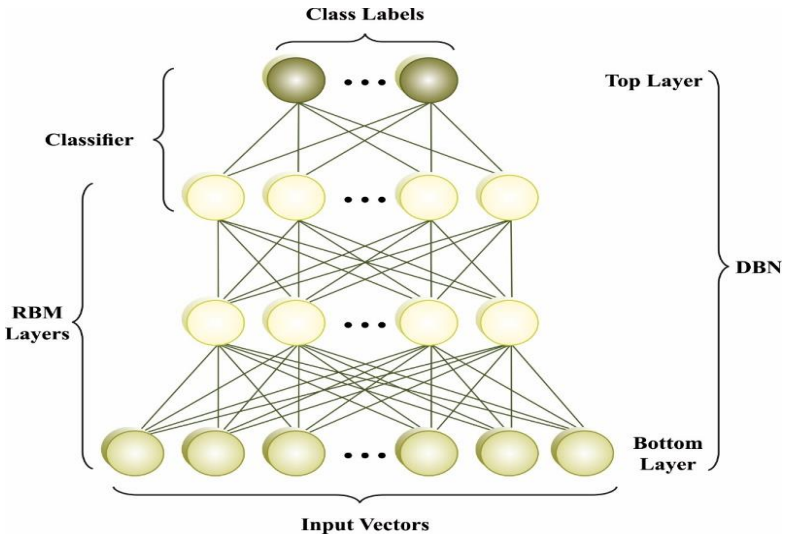

Figure 2. DBN architecture

Figure 2 depicts the infrastructure of DBN. However, unlike the traditional FFNN model trained by the error backpropagation model, DBN is a generalization method with the stack of multiple RBM and exploits the unsupervised layer wise model training to improve the learning ability, avoid the local minimization, and enhance the training efficiency. RBM is a main branch of ANN, involving a HL and a visible layer (VL). The HL is utilized for the fundamental expression of data, and the VL is applied for data input and output. RBM is used to define the system as energy-based. Once the sample $s = (s_1, s_2, \dots, s_n)$ is inputted to RBM, then network energy is:

$$E(s) = -\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij}\, s_i s_j - \sum_{i=1}^{n} \theta_i\, s_i \tag{11}$$

In Eq. (11), the weight connecting the i^{th} and j^{th} neurons is w_{ij}; θ_i denotes the threshold of i^{th} neurons. However, each neuron in RBM is fully connected, which extremely increases unnecessary complexity. Therefore, Hinton developed RBM. In RBM, the weight connection exists between the HL and VL.

In general, DL has more hyper parameters; hence, the training efficacy is low. Therefore, the unsupervised layer wise model training is used in DBN to improve this situation. The contrastive divergence model is adopted for training the first-layer RBM in DBN; then, the HL of the first-layer RBM is applied as the VL of the second-layer RBM, and likewise, the CD model is adopted for training the second-layer RBM, etc. This procedure is known as pre-training. Lastly, the error backpropagation model is used for training the entire DBN, called fine-tuning. "Pre-training+fine-tuning" significantly saves the cost of the training model.

4. Performance Validation

The performance evaluation of the CVODL-ER technique takes place using the KDEF [22] database and KMU-FED [23] database. The KDEF database comprises 4900 samples and the KMU-FED database includes 1106 samples. Tables 1 and 2 represent the detailed description of dual databases.

Table 1: Details of KDEF database

KDEF Database	
Classes	No. of Images
Afraid	701
Angry	700
Disgusted	700
Happy	700
Sad	699
Surprised	700
Neutral	700
Total No. of Images	4900

Table 2: Details of KMU-FED database

KMU-FED Database	
Classes	No. of Images
Afraid	200
Angry	196
Disgusted	120
Happy	210
Sad	180
Surprised	200
Total No. of Images	1106

The classifier outcomes of the CVODL-ER methodology at the KDEF database are demonstrated in Figure 3. The confusion matrix provided by the CVODL-ER model on 70%TRAPH: 30%TESPH is portrayed in Figures. 3a-3b. The outcome shows that the CVODL-ER technique has identified and classified all 7 classes accurately. Afterward, the PR outcome of the CVODL-ER method is established in Figure 3c. The outcome indicates that the CVODL-ER model has attained high PR values at 7 classes. At last, the ROC investigation of the CVODL-ER model is exemplified in Figure 3d. The outcome displays that the CVODL-ER model has led to proficient results with higher ROC rates below various classes.

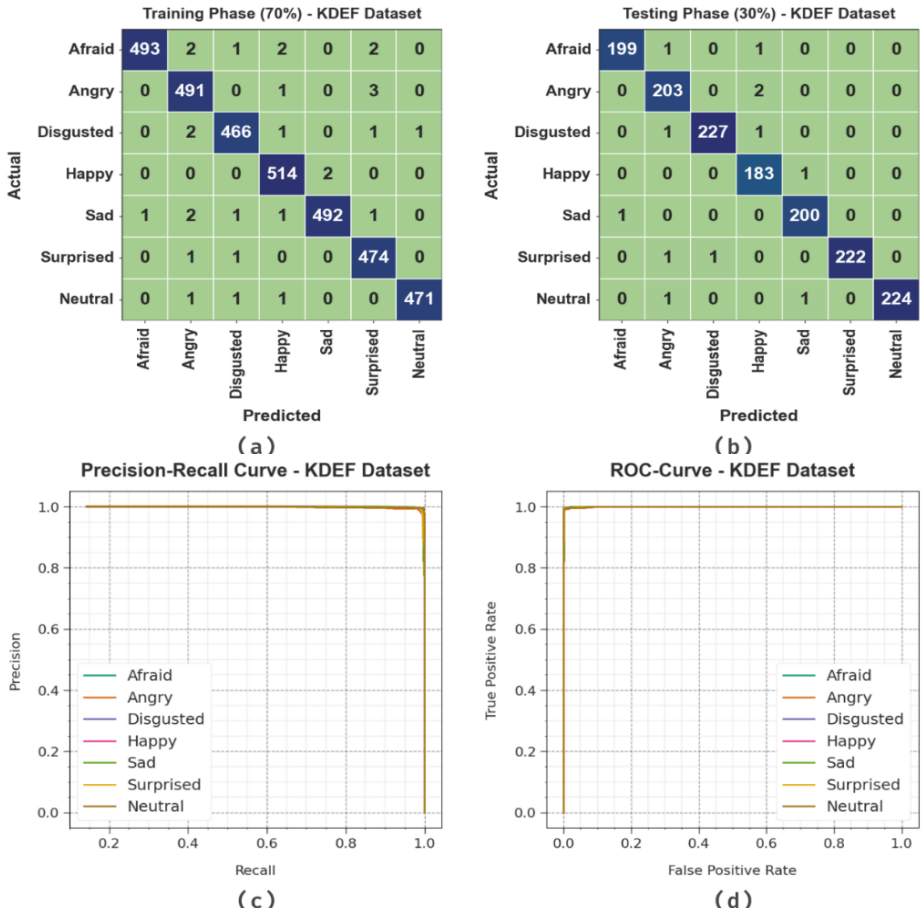

Figure 3. KDEF database (a-b) 70%TRAPH: 30%TESPH of confusion matrices and (c-d) PR and ROC curves

The recognition results of the CVODL-ER method on the KDEF database are given in Table 3 and Figure 4. The outcomes emphasized that the CVODL-ER model appropriately recognized various emotions. With 70% of TRAPH, the CVODL-ER system offers an average $accu_y$ of 99.76%, $prec_n$ of 99.16%, $reca_l$ of 99.15%, F_{score}

of 99.16%, and $G_{measure}$ of 99.16%. Moreover, with 30% of TESPH, the CVODL-ER method provides an average $accu_y$ of 99.77%, $prec_n$ of 99.14%, $reca_l$ of 99.19%, F_{score} of 99.16%, and $G_{measure}$ of 99.17%.

Table 3: Detection outcome of CVODL-ER technique on the KDEF database

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	$G_{Measure}$
TRAPH (70%)					
Afraid	99.77	99.80	98.60	99.20	99.20
Angry	99.65	98.40	99.19	98.79	98.79
Disgusted	99.74	99.15	98.94	99.04	99.04
Happy	99.77	98.85	99.61	99.23	99.23
Sad	99.77	99.60	98.80	99.19	99.19
Surprised	99.74	98.54	99.58	99.06	99.06
Neutral	99.88	99.79	99.37	99.58	99.58
Average	99.76	99.16	99.15	99.16	99.16
TESPH (30%)					
Afraid	99.80	99.50	99.00	99.25	99.25
Angry	99.59	98.07	99.02	98.54	98.54
Disgusted	99.80	99.56	99.13	99.34	99.34
Happy	99.66	97.86	99.46	98.65	98.66
Sad	99.80	99.01	99.50	99.26	99.26
Surprised	99.86	100.00	99.11	99.55	99.55
Neutral	99.86	100.00	99.12	99.56	99.56
Average	99.77	99.14	99.19	99.16	99.17

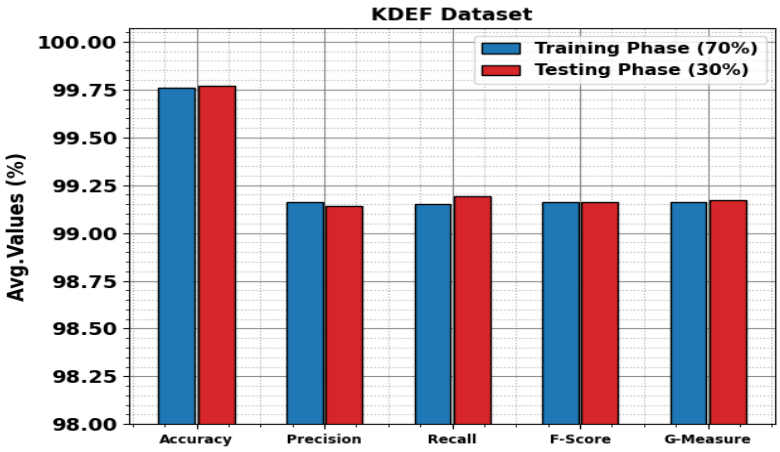

Figure 4. Average outcome of CVODL-ER technique under KDEF database

Table 4 reports a detailed comparative result of the CVODL-ER technique on the KDEF database [11]. In Figure 5, a comparative investigation of the CVODL-ER model on the KDEF database takes place in terms of $accu_y$. The figure implies that the MULTI-CNN and ALEXNET-LDA techniques have attained the least $accu_y$ values of 89.34% and 88.84%. Meanwhile, the HDNN, DL-FER, and GLFCNN-SVM techniques have reported closer $accu_y$ values of 96.81%, 95.98%, and 98.65%. Although the SSO-DLFER model has led to a considerable $accu_y$ of 99.69%, the CVODL-ER technique offers a maximum $accu_y$ of 99.77%.

Table 4: Comparative outcome of CVODL-ER methodology with existing methods under the KDEF database

KDEF Database		
Methods	Accuracy (%)	Computational Time (sec)
CVODL-ER	99.77	1.62
SSO-DLFER	99.69	2.45
MULTI-CNN	89.34	2.95

HDNN	96.81	2.32
ALEXNET-LDA	88.84	3.80
DL-FER	95.98	3.64
GLFCNN-SVM	98.65	5.59

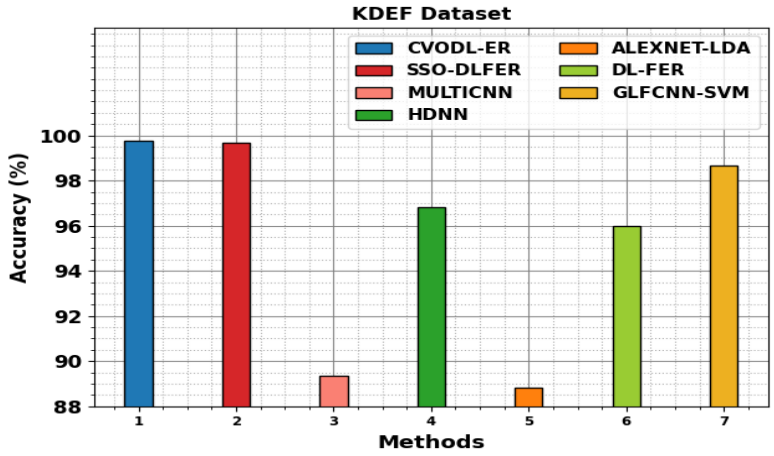

Figure 5. $Accu_y$ Analysis of CVODL-ER technique under the KDEF database

A comparative investigation of the CVODL-ER technique on the KDEF database takes place in terms of computational time (CT) is shown in Figure 6. The figure indicates that the GLFCNN-SVM and ALEXNET-LDA techniques have attained highest CT values of 5.59s and 3.80s. Meanwhile, the DL-FER, MULTI-CNN, and HDNN approaches have reported closer CT values of 3.64s, 2.95s, and 2.32s. Even though the SSO-DLFER method has resulted in a considerable CT of 2.45s, the CVODL-ER method provides a minimum CT of 1.62s.

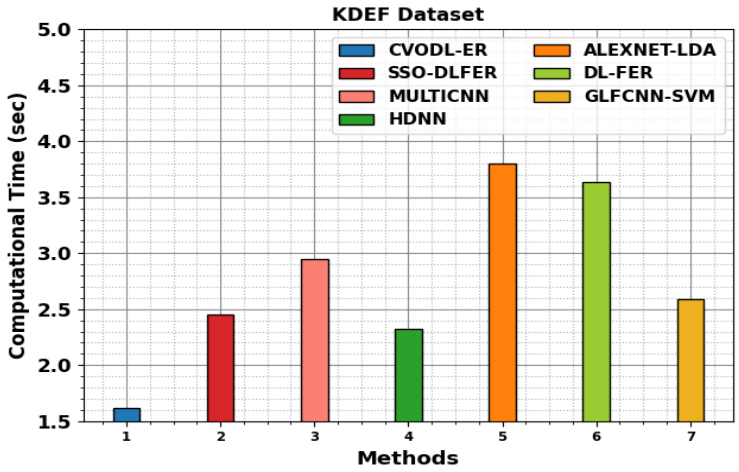

Figure 6. CT analysis of CVODL-ER technique under KDEF database

The classifier outcomes of the CVODL-ER system on the KMU-FED database are shown in Figure7. The confusion matrix offered by the CVODL-ER system on 70%TRAPH: 30%TESPH is represented in Figs 7a-7b. The outcome stated that the CVODL-ER approach has identified and classified all 6 classes. In addition, the PR outcome of the CVODL-ER method is revealed in Figure 7c. The figure inferred that the CVODL-ER algorithm has attained high PR performance below 6 classes. Eventually, the ROC investigation of the CVODL-ER model is demonstrated in Figure 7d. The outcome indicates that the CVODL-ER algorithm has led to proficient outcomes with the highest ROC rates under numerous classes.

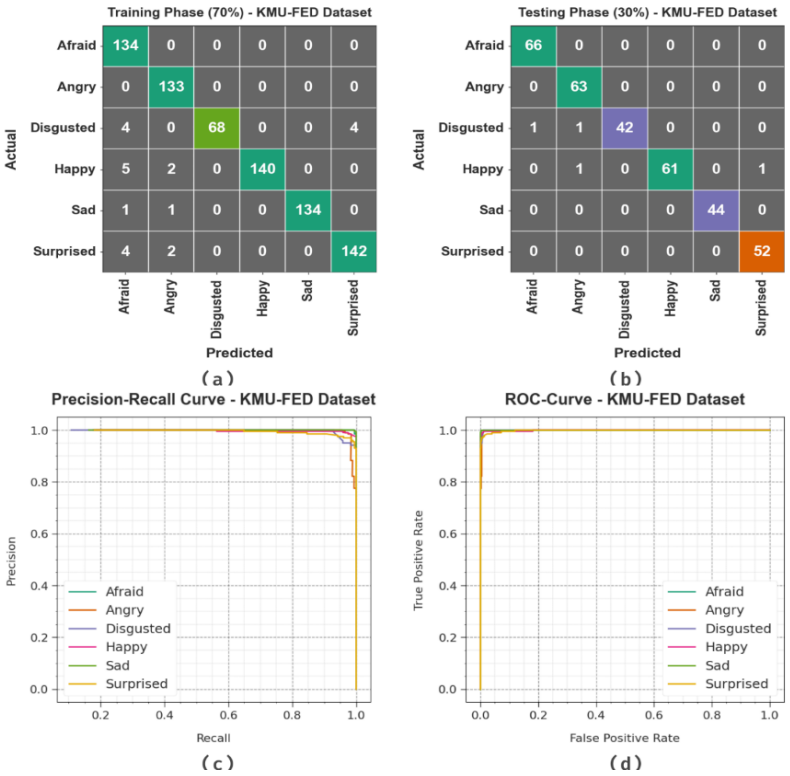

Figure 7. KMU-FED database (a-b) 70%TRAPH: 30%TESPH of confusion matrices and (c-d) PR and ROC curves

The detection outcomes of the CVODL-ER algorithm on the KMU-FED database are shown in Table 5 and Figure 8. The outcomes highlighted that the CVODL-ER system suitably detected different emotions. With 70% of TRAPH, the CVODL-ER method provides an average $accu_y$ of 99.01%, $prec_n$ of 97.36%, $reca_l$ of 96.53%, F_{score} of 96.84%, and $G_{measure}$ of 96.89%. Furthermore, with 30% of TESPH, the CVODL-ER method provides an average $accu_y$ of 99.60%, $prec_n$ of 98.92%, $reca_l$ of 98.71%, F_{score} of 98.80%, and $G_{measure}$ of 98.81%.

Table 5: Detection outcome of CVODL-ER methodology on KMU-FED database

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	$G_{Measure}$
TRAPH (70%)					
Afraid	98.19	90.54	100.00	95.04	95.15
Angry	99.35	96.38	100.00	98.15	98.17
Disgusted	98.97	100.00	89.47	94.44	94.59
Happy	99.10	100.00	95.24	97.56	97.59
Sad	99.74	100.00	98.53	99.26	99.26
Surprised	98.71	97.26	95.95	96.60	96.60
Average	99.01	97.36	96.53	96.84	96.89
TESPH (30%)					
Afraid	99.70	98.51	100.00	99.25	99.25
Angry	99.40	96.92	100.00	98.44	98.45
Disgusted	99.40	100.00	95.45	97.67	97.70
Happy	99.40	100.00	96.83	98.39	98.40
Sad	100.00	100.00	100.00	100.00	100.00
Surprised	99.70	98.11	100.00	99.05	99.05
Average	99.60	98.92	98.71	98.80	98.81

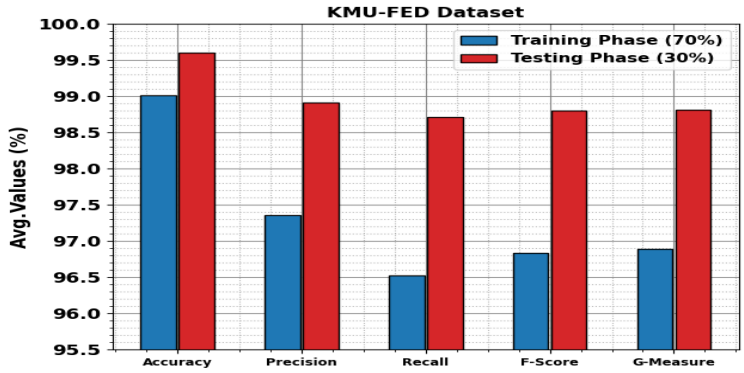

Figure 8. Detection outcome of CVODL-ER technique under KMU-FED database

A detailed comparative outcome of the CVODL-ER method on the KMU-FED database is reported in Table 6. In Figure 9, a comparative investigation of the CVODL-ER algorithm on the KMU-FED database takes place in terms of $accu_y$. The figure shows that the FTDRF and MobileNetV3 techniques have obtained minimum $accu_y$ values of 93.17% and 94.28%. Meanwhile, the VGG16, CCNN, and GLFCNN-SVM techniques have demonstrated closer $accu_y$ values of 94.53%, 97.79%, and 98.83%. Even though the SSO-DLFER model has resulted in a considerable $accu_y$ of 99.50%, the CVODL-ER method provides a higher $accu_y$ of 99.60%.

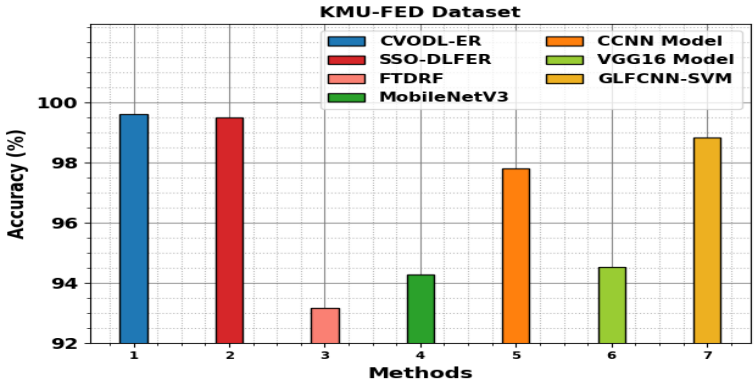

Figure 9. $Accu_y$ Analysis of CVODL-ER technique under KMU-FED database

Table 6: Comparative analysis of CVODL-ER technique with existing methods under KMU-FED database

KMU-FED Database		
Methods	Accuracy (%)	Computational Time (sec)
CVODL-ER	99.60	1.55
SSO-DLFER	99.50	2.78
FTDRF	93.17	5.46
MobileNetV3	94.28	5.33
CCNN Model	97.79	4.15
VGG16 Model	94.53	5.67
GLFCNN-SVM	98.83	4.63

A comparative investigation of the CVODL-ER technique on the KMU-FED database takes place in terms of CT is shown in Fig. 10. The figure indicates that the VGG16 and FTDRF techniques have attained high CT values of 5.67s and 5.46s. Meanwhile, the MobileNetV3, GLFCNN-SVM, and CCNN techniques have shown closer CT values of 5.33s, 4.63s, and 4.15s. Even though the SSO-DLFER approach has resulted in a considerable CT of 2.78s, the CVODL-ER method provides a minimum CT of 1.55s. These results ensured the enhanced FER outcomes of the CVODL-ER technique.

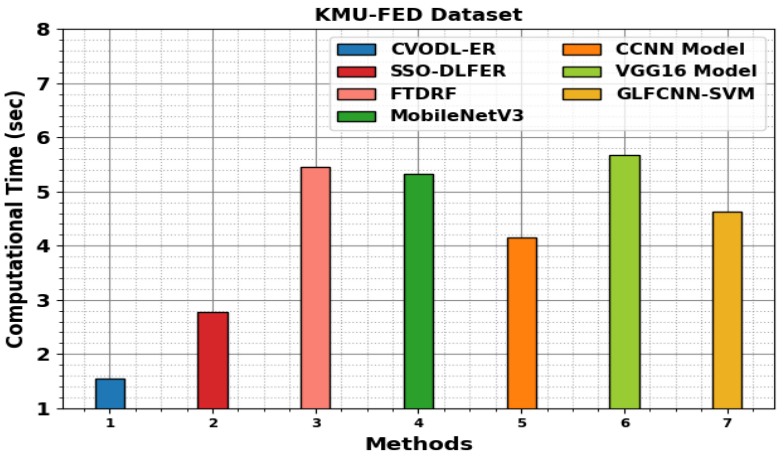

Figure 10. CT analysis of CVODL-ER technique under KMU-FED database

5. Conclusion

In this study, we have developed a novel CVODL-ER system for autonomous vehicle drivers. The CVODL-ER approach concentrates on the automated classification of numerous classes of emotions the autonomous vehicle drives. To accomplish this, the CVODL-ER technique comprises a SE-ResNet-based feature extractor, QO-Jaya algorithm-based hyperparameter tuning, and a DBN-based classification process. Moreover, the CVODL-ER technique makes use of the SE-ResNet model for learning intrinsic patterns from the driver's facial images. Besides, the hyperparameter tuning of the SE-ResNet algorithm takes place via the QO-Jaya algorithm. For the recognition of driver emotions, the CVODL-ER technique applies the DBN model. The performance analysis of the CVODL-ER technique occurs using a benchmark facial image database. The obtained results underline the improved efficiency of the CVODL-ER technique over other models.

Funding: "This research received no external funding"

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] Sini, J., Marceddu, A.C. and Violante, M., 2020. Automatic emotion recognition for the calibration of autonomous driving functions. Electronics, 9(3), p.518.

[2] Lee, K.H., 2020, October. Design of a convolutional neural network for speech emotion recognition. In 2020 International Conference on Information and Communication Technology Convergence (ICTC) (pp. 1332-1335). IEEE.

[3] Kandeel, A.A., Abbas, H.M. and Hassanein, H.S., 2021, January. Explainable Model Selection of a Convolutional Neural Network for Driver's Facial Emotion Identification. In International Conference on Pattern Recognition (pp. 699-713). Springer, Cham.

[4] Izquierdo-Reyes, J., Ramirez-Mendoza, R.A., Bustamante-Bello, M.R., Pons-Rovira, J.L. and Gonzalez-Vargas, J.E., 2018. Emotion recognition for semi-autonomous vehicles framework. International Journal on Interactive Design and Manufacturing (IJIDeM), 12(4), pp.1447-1454.

[5] Lu, H., Liu, Q., Tian, D., Li, Y., Kim, H. and Serikawa, S., 2019. The cognitive internet of vehicles for autonomous driving. IEEE Network, 33(3), pp.65-73.

[6] Meshram, H.A., Sonkusare, M.G., Acharya, P. and Prakash, S., 2020, November. Facial Emotional Expression Regulation to Control the Semi-Autonomous Vehicle Driving. In 2020 IEEE International Conference for Innovation in Technology (INOCON) (pp. 1-5). IEEE.

[7] Shafaei, S., Hacizade, T. and Knoll, A., 2018, December. Integration of driver behavior into emotion recognition systems: a preliminary study on the steering wheel and vehicle acceleration. In Asian Conference on Computer Vision (pp. 386-401). Springer, Cham.

[8] Garg, M., Ubhi, J.S. and Aggarwal, A.K., 2022. Neural style transfer for image steganography and destylization with supervised image-to-image translation. Multimedia Tools and Applications, pp.1-18.

[9] Chen, L., Lin, S., Lu, X., Cao, D., Wu, H., Guo, C., Liu, C. and Wang, F.Y., 2021. Deep neural network based vehicle and pedestrian detection for autonomous driving: a survey. IEEE Transactions on Intelligent Transportation Systems, 22(6), pp.3234-3246.

[10] Arefnezhad, S., Eichberger, A., Frühwirth, M., Kaufmann, C., Moser, M. and Koglbauer, I.V., 2022. Driver Monitoring of Automated Vehicles by Classification of Driver Drowsiness using a Deep Convolutional Neural Network Trained by Scalograms of ECG Signals. Energies, 15(2), p.480

[11] Jain, D.K., Dutta, A.K., Verdú, E., Alsubai, S. and Sait, A.R.W., 2023. An automated hyperparameter tuned deep learning model enabled facial emotion recognition for autonomous vehicle drivers. Image and Vision Computing, 133, p.104659.

[12] 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.

[13] Zaman, K., Sun, Z., Shah, S.M., Shoaib, M., Pei, L. and Hussain, A., 2022. Driver Emotions recognition based on improved faster R-CNN and neural architectural search network. Symmetry, 14(4), p.687.

[14] Saranya, R., Vijayaragavan, M., Kalilulah, S.I., Satyanarayana, K.N.V., Suresh, M. and Srimathi, S., 2023, December. Automated Facial Emotion Detection using Arithmetic Optimization Algorithm with Deep Convolutional Neural Network for Autonomous Vehicle Drivers. In 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 823-828). IEEE.

[15] Chen, J., Dey, S., Wang, L., Bi, N. and Liu, P., 2021, July. Multi-modal fusion enhanced model for driver's facial expression recognition. In 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-4). IEEE.

[16] Sahoo, G.K., Das, S.K. and Singh, P., 2022, May. Deep learning-based facial emotion recognition for driver healthcare. In 2022 National Conference on Communications (NCC) (pp. 154-159). IEEE.

[17] Sarvakar, K., Senkamalavalli, R., Raghavendra, S., Kumar, J.S., Manjunath, R. and Jaiswal, S., 2023. Facial emotion recognition using convolutional neural networks. Materials Today: Proceedings, 80, pp.3560-3564.

[18] Chand, H.V. and Karthikeyan, J., 2022. CNN Based Driver Drowsiness Detection System Using Emotion Analysis. Intelligent Automation & Soft Computing, 31(2).

[19] Chen, T., Qin, H., Li, X., Wan, W. and Yan, W., 2023. A Non-Intrusive Load Monitoring Method Based on Feature Fusion and SE-ResNet. Electronics, 12(8), p.1909.

[20] Jiang, X., Hong, Z., Feng, Y. and Tan, J., 2024. Multi-Objective Optimization of VBHF in Deep Drawing Based on the Improved QO-Jaya Algorithm. Chinese Journal of Mechanical Engineering, 37(1), p.5.

[21] Yin, X., Huang, X., Pan, Y. and Liu, Q., 2023. Point and interval estimation of rock mass boreability for tunnel boring machine using an improved attribute-weighted deep belief network. Acta Geotechnica, 18(4), pp.1769-1791.

[22] Lundqvist, D.; Flykt, A.; Öhman, A. Karolinska directed emotional faces. Cogn. Emot. 1998. https://kdef.se/

[23] KMU-FED. Available online: http://cvpr.kmu.ac.kr/KMU-FED.htm