I.J. Intelligent Systems and Applications, **2024, 3, 47-58**
Published Online on June 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2024.03.04

Hybrid Deep Optimal Network for Recognizing Emotions Using Facial Expressions at Real Time

Rakshith M. D.*
Department of Computer Science and Engineering, Canara Engineering College, Benjanapadavu, Visvesvaraya Technological University, Belagavi, 590018, India
E-mail: rakshithmd@canaraengineering.in
ORCID iD: https://orcid.org/0000-0002-0196-531X
*Corresponding author

Harish H. Kenchannavar
Department of Information Science and Engineering, Gogte Institute of Technology, Belagavi, Visvesvaraya Technological University, Belagavi, 590018, India
E-mail: harishhk@git.edu
ORCID iD: https://orcid.org/ 0000-0001-7369-0565

Received: 09 December 2023; Revised: 14 January 2024; Accepted: 09 February 2024; Published: 08 June 2024

Abstract: Recognition of emotions by utilizing facial expressions is the progression of determining the various human facial emotions to infer the mental condition of the person. This recognition structure has been employed in several fields but more commonly applied in medical arena to determine psychological health problems. In this research work, a new hybrid model is projected using deep learning to recognize and classify facial expressions into seven emotions. Primarily, the facial image data is obtained from the datasets and subjected to pre-processing using adaptive median filter (AMF). Then, the features are extracted and facial emotions are classified through the improved VGG16+Aquila_BiLSTM (iVABL) deep optimal network. The proposed iVABL model provides accuracy of 95.63%, 96.61% and 95.58% on KDEF, JAFFE and Facial Expression Research Group 2D Database (FERG-DB) which is higher when compared to DCNN, DBN, Inception-V3, R-152 and Convolutional Bi-LSTM models. The iVABL model also takes less time to recognize the emotion from the facial image compared to the existing models.

Index Terms: Facial Emotion Recognition, Adaptive Median Filter, Aquila Optimizer, Feature Extraction.

1. Introduction

Facial and emotional expression contributes significantly in expressing internal emotions and intention by means of non-verbal manner [1]. The emotion is one which the nervous system is associated with the behavioral changes, consciousness and dis-gratification [2]. In past few decades, facial action coding system (FACS) is considered widely for classifying the human emotions using action units (FU) [3]. In AUs, more than 46 minor visual facial motions are contemplated to read the facial emotions efficiently. But, the FACS based automated system fails to code the dynamic movements and is highly limited due to different facial expression movements [4]. Recently, recognizing emotion using facial expression is the problem of interest in computer vision field. For the recognition of facial emotion of humans, two basic techniques such as image sequence based and still image based techniques are used [5]. In this, the image sequence based technique performs well in classifying the emotion recognition more accurately. In some of the critical circumstance like hospitalized patients restricts their emotions, therefore an effective recognition technique needs to be established for better communication [6]. Automatic recognition of emotion using facial expression has grown a lot of attention towards smart environments such as hospitals, cities, smart homes etc. [7]. One of the popular techniques called intelligent personal assistant (IPAs) uses natural language processing (NLP) for communicating with the human. Using this application, the communication and human intelligence is highly enhanced [8]. However, this technique cannot recognize all the emotions and high time complexity arises. Recently, AI based techniques actively participate in many applications by efficiently recognizing the facial features like eyes brows, mouth, voice tunes, body language etc. [9].

Of this, ML based technique can compromise the emotion recognition problems for human computer interaction (HCI). Human interaction and communication are involved in finalizing the human emotions [10]. Low time

complexity, low cost and robust recognition are the advantages of this approach. Recognition of facial expression remains rigorous problem in several applications like HCI and medical related applications [11]. In [12, 13], a total of 68 features are used to recognize the three emotions like blank, positive and negative in real time platform. Also, in [14], emotions are recognized with 79 features and 26 geometric features that achieve low accuracy of 71%. Some of the alternative approaches use still images. But these techniques lack to recognize the fear and disgusting emotional states from the facial expression. However, hybridizing the facial expression with other components such as gestures, speech signals cannot able to recognize the important emotions like dumb, deaf etc. This results in inconvenient for helping physically challenged people with the special needs. The previous research is not reliable and cannot be applicable for real time systems for many applications. Nowadays, DL based techniques play an integral role in accurate emotion recognition with more facial states. Existing DL techniques such as CNN [15], hybrid CNN-LSTM [16] and DNN [17] shows accurate recognition of emotion states using facial expression. But these techniques cannot be processed with huge number of training datasets.

Over the recent years, face emotion analysis is emerged as a substantial research area in predicting the emotions in different applications. Rapid and accurate recognition of human face emotions plays pivotal role in security, medical and various organizations. Inappropriate recognition models and nonexistence of enhanced recognition approaches leads to degraded results. The manual recognition cannot be accurate most of the time and it consumes more time for recognizing the face emotions. Even though, diverse models are conducted but the processing overhead tends to be high, accuracy gets degraded and exact recognition cannot be attained because of noise in images. Even though several benefits are offered by deep learning models they suffer from over fitting issue and recognition error. Hence, the imaging models needs to be integrated with novel deep learning approach for outcome enhancement. Thereby, motivating to establish a reliable, robust and automated face emotion recognition system, an efficient approach to overcome the existing shortcomings is exhibited in the presented research.

The contributions of the presented research are to:

- Provide novel improved VGG16+Aquila_BiLSTM (iVABL) model for recognizing different facial emotions accurately.
- Remove the noise in the facial image using AMF filtering approach.
- Retrieve features using CNN based improved VGG16 model to obtain improved accuracy rate and minimized error rate.
- Classify the facial emotions using Aquila_BiLSTM network with enhanced classification accuracy.

The sections of the paper are structured as follows: Section 2 illustrates the existing works in FER, Proposed methodology is described in Section 3, and Result analysis is illustrated in Section 4. Finally, conclusion and future work is discussed in Section 5.

2. Literature Review

Some current research works based on emotion recognition in classifying the facial expressions are surveyed as follows.

M. A. H. Akhand et al. [18] suggested a system by adapting transfer learning in deep CNN (DCNN) with pipeline tuning policy using facial images. The pre-trained DCNN technique was used by switching its dense upper layer well-matched with facial emotion recognition. In pipeline tuning policy, the dense layer training had followed by successive tuning of every block of pre-trained DCNN. This could lead to gradual enhancement of recognition accuracy. However, the suggested system would need to reduce the time complexity.

Hasan Deeb et al. [19] presented an extreme learning machine (ELM) with enhanced black hole algorithm (EBHA) for recognizing facial emotions. Here, the universal approximation characteristics of ELM and the global search ability of EBHA were integrated to distinguish the facial emotion features and determine their classes. The linear discriminant analysis (LDA) and principal component analysis (PCA) were adopted to preserve the features of interest and minimize the dimensions of facial images. Although attaining better performance, this model had exhibited drawbacks like minimum accuracy and high computational time.

Naveen Kimari & Rekha Bhatia [20] offered an effective DL based facial emotion recognition scheme by employing modified joint trilateral filter (MJTF) and DCNN classifier. At first, the DL model had used MJTF to eliminate the noise in the datasets. Then, the filtered image was undertaken with CLAHE (contrast-limited adaptive histogram equalization) for enhancing the image visibility. Lastly, DCNN was employed to classify the facial emotions. As a result, better outcomes had identified with performance metrics but it not met the expected performance.

Xiao Sun et al. [21] suggested a deep neural network (DNN). Attention mechanism was used to combine features that are deep and shallow for detecting facial expressions. Here, attention shallow model (ASM) had applied to obtain the shallow features by using relative location of facial landmarks. With the benefit of DCNN, an attention deep model (ADM) was adopted for mining the deep features. At last, ASM and ADM were combined to multi-attention shallow and deep scheme to achieve recognition. Also, this model had enhanced the effectiveness in recognizing the facial

expression. However, accuracy performance was not satisfied, and recognition rate needs to be improved.

Ramachandran Vedantham et al. [22] introduced optimized Deep Belief Network (DBN) along with Spider Monkey Optimization (SMO) for robust feature extraction and classification of facial expressions. At first, pre-processing was carried out then texture descriptors were utilized to extract features. Discrete Cosine Transform (DCT) features were extracted to minimize the processing cost and improve recognition rate. Then, PCA was applied for reducing dimensionality. Subsequently DBN with SMO was utilized for classifying the expressions. While analyzing the outcomes, this network had gained better outcomes however it would need to enhance the rate of recognition.

Dandan Liang et al. [23] recommended a deep convolutional bidirectional long short term memory (BiLSTM) to increase the performance of recognizing facial expressions. This approach focuses on jointly learning temporal dynamics and spatial features for recognition. Here, the deep network was utilized to extract spatial features and convolutional network was applied to model temporal dynamics. Eventually, by using BiLSTM networks, the suggested model accumulates clues from fused features. As a result, better outcomes had obtained and excelled in modeling temporal data with high complexity.

Hussain et al. [24] proposed an efficient method for identification and detection of facial expressions accurately. The proposed method's main intention is to recognize and authenticate facial features. A CNN model with the KDEF dataset and VGG16 was employed for face recognition and categorization. An accuracy level of 88% was attained by the suggested strategy and this method's drawback is high time computation.

Jain et al. [25] recommended an efficient method called Multi-Angle Optimal Pattern-based Deep Learning (MAOP-DL) technique. The main goal of this strategy is to overcome the problem of rapid lighting changes. The proposed approach involves DL and optimization. The proposed methods were applied on CK+ and MMI database. An accuracy level of 96.72% was obtained. The main drawbacks of the presented work are poor performance in handling involuntary expressions and low intensity.

Khan et al. [26] proposed an efficient method called Long Short Term Memory (LSTM). This article deals with increased computational complexity. The suggested technique utilizes feature-based weight updating to maximize classification accuracy. Testing classification accuracy for the suggested method is 96.00% and 83.30%. The suggested method utilizes Bonn dataset. This method's drawback was that it required a lot of training time.

Dastider et al. [27] suggested efficient methods called the CNN and LSTM techniques. The suggested network effectiveness is confirmed using a five-fold cross-validation approach. The incorporation of the Long Short-Term Memory (LSTM) layers after thorough result analysis reveals a promising improvement in the classification performance by an average of 7 12%, which is around 17% more than the existing techniques.

Chavali, S.T., et al. [28] presented techniques for estimating age and gender from human faces, also a system for identifying an individual's emotional state based on their facial expression. Determining the effects of human age and gender on facial expressions is another goal of this research. The architectures of sixteen pre-trained models were used for experimental study.

Bhangale, K., et al. [29] developed a model on the basis of VGG16. The model used a real-time database of 50 subjects' facial images. Face detection is accomplished using Multitask Cascaded Convolutional Neural Networks (MTCNN), and our model is trained using a real-time database using VGGNet. A real-time database with images that have different pose, illumination, and background variations to deal with pose variation, view variation, illumination changes, occlusion, and background variations has been presented.

3. Methodology

Emotion recognition from human faces is the most significant and challenging task in the domain of social communication. Facial expressions are extensively required to analyze the emotions of the human. Different forms of expressions like smile, sad, surprise and fear can be analyzed from human faces. The accurate recognition of facial emotions is needed to mimic human interventions and can be used in several applications like disease diagnosis, physiological detection of interactions. Hence in the proposed work, the facial emotions can be recognized by using a prominent model with the input of face expression images. The proposed model undergoes various steps including image pre-processing and a single model including feature extraction and classification of images.

Figure 1 depicts the overall face emotion recognition architecture. A novel methodology is established for recognizing real time face emotions of human and this can dynamically target over the prominent features of images through the training process. Initially, the data are collected from the public sources KDEF, JAFFE and FERG-DB. The gathered data are pre-processed through the adoption of AMF filtering approach [30]. The pre-processed data are fed to improved VGG16+Aquila_BiLSTM (iVABL) deep optimal network for extracting the features and classifying the facial emotions. The VGGNet-16 is exploited as features extractor for representing the images as dimensional feature vector [31]. The novelty of this paper is to increase accuracy by using improved VGG-16, and BiLSTM by concentrating long term dependencies. Also, the proposed model consumes less time to test the images during real-time. The description of every process is given in detail as follows.

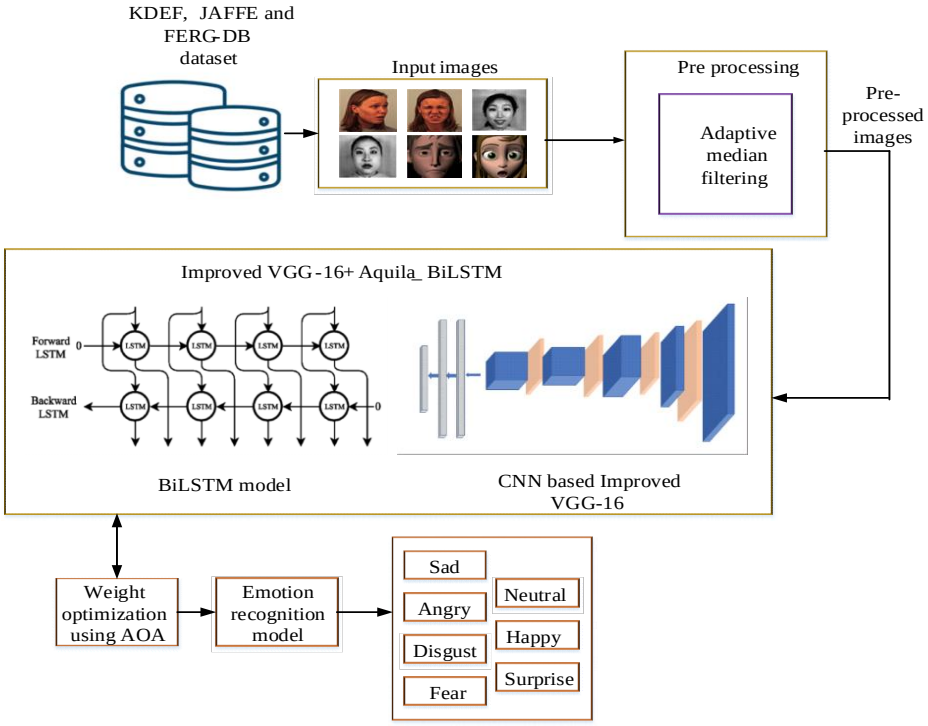

Fig.1. Proposed architecture for facial emotion recognition

3.1. Image Pre-processing

Pre-processing of images tends to be a prominent section undertaken in diverse image processing applications that acts as a key characteristic in enhancing the classification accuracy. The chief objective of pre-processing approach is to improve the image quality for the effective performance analysis. The presence of external and internal noise can adversely affect the image quality, generates deterioration over spatial resolution of images, distorts the significant image features and image details can be lost. The distorted images tend to highly concerned especially in image processing applications as it degrades the general performance. Hence in the proposed technique, the pre-processing step is carried through the adoption of AMF filtering approach.

The median filtering approach is the order statistic filter that comes under the group of non-linear filters. The median founded filters are generally applied for minimizing the levels of impulse noises from the corrupted images. Through the simplicity and edge preserving ability, the median filter promotes better performance. The general median filter employs in spatial field, depends upon windowing procedure and the filter size $F_U \times F_V$ is used. For input X, the filtered image Y can be obtained as,

$$Y(m,n) = \underset{(a,b)\in A_{mn}}{Median}\{X(a,b)\} \tag{1}$$

From the above expression, the pixel coordinates are represented as *(m, n)* at the center of contextual region A_{mn} and is signified as $F_U \times F_V$. The pixel coordinates belong to the contextual region is denoted as *(a, b)*. It represents that the filtered pixel value is the data median value over particular region. The conventional median filter improves the image quality through the eradication of thin lines, blurring, corners or distortion from the images. But the working of median filter tends to be highly complex and consumes more time in processing the images. Hence, to conquer the limitations of conventional median filter, AMF is employed. In AMF, switching based median filter is employed to eradicate the corrupted image pixels. The AMF approach helps in preserving the local details of image that is separated into two stages as noise detection and cancellation process. In the stage of noise detection, the noisy pixels are identified at first and then a rough approximation of noise level is made on the image. The assumption is made as two intensities which generate the impulse noise are given as maximum, minimum dynamic range like 0 and L-1. At this phase, for every pixel position *(a, b)*, the mask β is marked through the below expression.

$$\beta(a,b) = \begin{cases} 1 : X(a,b) = L-1 \\ 1 : X(a,b) = 0 \\ 0 : Otherwise \end{cases} \tag{2}$$

The noise pixel is represented as 1 whereas the noise free pixel is denoted as 0. After the classification of pixels using equation (2), the total number of noisy pixels is calculated as,

$$S = \sum_{a=0}^{U-1} \sum_{b=0}^{V-1} \beta(a, b) \tag{3}$$

By using S value, the level of μ impulse noise that degrades the image can be roughly estimated. The value of μ can be defined as the noise pixel ratio to the overall image pixels.

$$\mu = S/(UV) \tag{4}$$

The μ value lies in between 0 and 1 whereas the noise mask β will be applied in the noise cancellation stage for eradicating the noises. In noise cancellation phase, the input image X is filtered and the filtered image Y is produced. The output can be expressed as,

$$Y(a, b) = [1 - \beta(a, b)]\, X(a, b) + [\beta(a, b)\, M(a, b)] \tag{5}$$

From the above expression, β denotes the noise mask, median value is denoted as M. This takes the value either 0 or 1 and the output value $Y(a, b)$ is both equal to $X(a, b)$ or $M(a, b)$. The value of $M(a, b)$ is calculated when there are noisy pixels and in case of noise free pixel, $X(a, b)$ is moved straight as the value of $Y(a, b)$. The AMF approach improves the processing speed as not all pixels required to be filtered. The $Y(a, b)$ can be rearranged as,

$$Y(a, b) = \begin{cases} X(a, b): \beta(a, b) = 0 \\ M(a, b): Otherwise \end{cases} \tag{6}$$

The square filters possessing odd dimensions were considered over filtering procedure which can be expressed as,

$$F = F_U = F_V = 2R+1 \tag{7}$$

The value of R considered any positive integer value. The AMF approach ensures that the value is not influenced by noise and is extra biased to real data. The value of $M(a, b)$ can be described through the subsequent algorithm. For every pixel position (a, b) with $\beta(a, b) = 1$, initialize the filter size $F=2R_{MIN}+1$ where R_{MIN} represents the minor integer value. Evaluate the noise free pixel amount and if they are fewer than eight pixels, then maximize the filter size by two and the compute the noise free pixels again also it reduces more noises. Evaluate $M(a, b)$ on the basis of noise free pixels present in $F \times F$ window and update $M(a, b)$ using equation (5) or (6). To reduce the trail number to determining the exact filter size, R_{MIN} has to be approximated as,

$$R_{MIN} = \frac{1}{2}\sqrt{\frac{7}{1-\beta}} \tag{8}$$

By using the above expression, better convergence and less looping can be obtained which in turn it speeds up the process. Hence, the preprocessing of images is carried effectively through the adoption of AMF approach in the removal of impulse noise over the gathered images. The proposed preprocessing technique is used to reduce extra noise for promoting accurate classification performance.

3.2. Feature Extraction Using Improved VGG16

The VGGNet-16 is exploited as features extractor for representing the images as dimensional feature vector. The number of feature maps generated by the VGG16 network occupies more space in the graphics card. This is because of multiple occurrences of convolution cores in every layer. Implementation results of VGG16 illustrates that the foremost layer which is fully connected generates huge parameters. This leads to complex calculations and more memory resources are consumed. Also, the VGG16 model will not perform accurately when medium and small sized datasets are considered for experimentation. This reduces generalization ability of the model which in turn causes overfitting. The solution for this problem is to minimize the depth of network thereby reducing the number of parameters. In the improved version of VGG16, the dimensions of the feature maps and parameters are reduced by using the large convolution kernel which also reduces the model's complexity. The output of convolution layer in improved VGG-16 model is described as,

$$f_u^s = C\left(\sum_{v=1}^{M} f_v^{s-1} * a_{vu}^s + b_u^s\right) \tag{9}$$

In equation (9), the matrix f_v^{s-1} denotes the v^{th} feature map of the preceding (s-1)th layer, the u^{th} feature map of the current s^{th} layer is signified f_u^s as and M represents the amount of input feature maps. Random initialization is supported over a_{vu}^s and b_u^swhich is set to zero. The non-linear function is described as C (.) that denotes ReLU function and $*$ represents the convolution operation. The ReLU layer comes after the convolutional layer and it performs nonlinear transformation over the input image. It is a piecewise linear function which provides the output for positive input or else it generates zero output.

$$C(m) = max\ (0, m) \tag{10}$$

The pooling layer progressively decreases the feature size illustration to lessen the parameter amount, computational difficulty and to tackle over fitting issues. The pooling layer compresses the features produced by a convolution layer and it limits the feature resolution to enhance steadiness. The few final layers are structured by the fully connected layer which gathers the pooling layer output and extracts the significant features. Fully connected layer is a fundamental form of feed forward neural networks whereas the output of every section is fed over the energized unit of subsequent layer. The fully connected layer represents that each node in the leading layer is linked over every node in second layer.

3.3. Classification Using Aquila_BiLSTM Network

The efficient features required for classifying the facial emotions are extracted using CNN based improved VGG-16 which are then fed into the BiLSTM network model. The major aim of choosing BiLSTM network is the learning ability can be enhanced and more data can be gathered for better recognition. The LSTM model is a variant of recurrent neural networks (RNN) and is integrated with special units called memory chunks.The memory chunk in LSTM is comprised three multiplicative gates and a memory cell. The basic gates of LSTM model are input gate, forget gate and output gate. The input gate fetches the input, the information forgotten before is identified by the forget gate and an output gate produces the output. The system's state is reserved by the memory cells whereas the activation information flow is directed by these three gates. Feature vectors attained from the feature extraction network are denoted as $f_1, f_2, f_3, f_4 ..... f_n$ whereas the hidden vector are denoted as $h_1, h_2, h_3, h_4 .....h_n$. The cell state in case of single directional LSTM cell can be estimated as follows.

$$H = \left[\frac{h_v - 1}{p_v}\right] \tag{11}$$

The final hidden vector h_v produced from the BiLSTM layer can be mathematically expressed as follows.

$$h_v = [Mh_v, Ah_v] \tag{12}$$

Due to the random selection of weights in BiLSTM model, there leads to degradation in the overall recognition performance and generates a loss function. Hence the loss functions are minimized by updating the weights using Aquila optimization algorithm (AOA) [32]. The AOA is simple to use algorithm that works well with large amounts of noisy training data. The cross entropy loss obtained in the outcome of BiLSTM can be expressed as,

$$CE_L = -\frac{1}{M}\sum_{p=1}^{M}\sum_{q=0}^{M} O_{pq}\ log\ \hat{O_{pq}} \tag{13}$$

In the above expression, batch size is denoted as M, O_{pq} represents the actual and predicted value is signified as $\hat{O_{pq}}$.The particular loss function is optimized through AOA which improves the convergence rate and accuracy by balancing the exploration and exploitation stages. The loss function of BiLSTM model can be optimized through the updated position of Aquila.

4. Results

The proposed Hybrid Deep Optimal Network model is examined with different stages like pre-processing, feature retrieval and classification. The model's performance is evaluated and compared with the existing models. The dataset details, performance metrics considered and its mathematical formulation, performance analysis and comparison are illustrated in the sub-sections below. Table 1 exemplifies the hyper parameter settings of proposed model.

Table 1. Hyperparameters details

Sl. No	Hyper-parameters	Particulars
1	Initial learning rate	0.0001
2	Batch size	32
3	Learning algorithm	Aquila optimizer
4	Maximum epoch size	300
5	Activation function	ReLU
6	Drop out factor	0.2
7	Iterations	100

4.1. Dataset Details

The facial emotions are recognized through the consideration of data from KDEF [33], JAFFE [34] and real time FERG-DB dataset [35]. Here, for training 80% and testing 20% of data are considered. The images from considered datasets are used for demonstrating the real-time emotion recognition which involves reading the facial image through OpenCV, then cropped face is resized to 224×224 and normalization is performed. The details of the datasets are mentioned in Table 2.

Table 2. Characteristics of dataset

Characteristics	**JAFFE**	**KDEF**	**FERG**
Image resolution	256×256 pixels	562×762 pixels	256×256 pixels
Expressions	7	7	7
Samples	213	4900	27,881
Expressions nature	Posed	Posed	Posed
Number of subjects	10	70	6

4.2. Performance Assessment

The performance of proposed deep optimal iVABL model can be assessed through the consideration of performance metric like Accuracy, precision, recall, specificity and F1-Score. To analyze model's performance, the experimentations are contrasted with the existing models to examine the model effectiveness of facial emotion recognition. Better outcomes of proposed model shows that they are superior in recognizing facial emotions exactly.

The performance outcomes of proposed iVABL model is related with the existing approaches and the outcomes are analyzed with respect to KDEF, JAFFE and FERG-DB dataset. Table 3, 4 and 5 represents the proposed performance outcomes of three datasets in terms of diverse metrics.

Table 3. Performance comparison for KDEF dataset

Techniques	**Accuracy (%)**	**Precision**	**Recall**	**Specificity**	**F1-score**
DCNN[18]	89.54	0.87	0.85	0.86	0.86
DBN[22]	90.22	0.89	0.87	0.87	0.88
Inc-V3[36]	91.24	0.90	0.89	0.90	0.89
R-152[37]	93.21	0.92	0.91	0.92	0.91
CBi-LSTM[23]	94.23	0.93	0.92	0.94	0.92
iVABL (Proposed)	**95.63**	**0.95**	**0.94**	**0.95**	**0.94**

Table 4. Performance comparison for JAFFE dataset

Techniques	**Accuracy (%)**	**Precision**	**Recall**	**Specificity**	**F1-score**
DCNN[18]	90.87	0.87	0.85	0.86	0.86
DBN[22]	92.45	0.88	0.86	0.88	0.87
Inc-V3[36]	93.22	0.90	0.88	0.90	0.89
R-152[37]	94.12	0.92	0.90	0.91	0.91
CBi-LSTM[23]	95.54	0.94	0.92	0.93	0.93
iVABL (Proposed)	**96.61**	**0.96**	**0.95**	**0.95**	**0.95**

Table 5. Performance comparison for FERG-DB dataset

Techniques	**Accuracy (%)**	**Precision**	**Recall**	**Specificity**	**F1-score**
DCNN[18]	86.78	0.85	0.84	0.82	0.84
DBN[22]	88.76	0.88	0.86	0.86	0.87
Inc-V3[36]	90.34	0.89	0.88	0.89	0.88
R-152[37]	92.11	0.91	0.90	0.92	0.90
CBi-LSTM[23]	93.21	0.94	0.93	0.93	0.93
iVABL (Proposed)	**95.58**	**0.95**	**0.94**	**0.94**	**0.94**

The above tables illustrate the comparative examination of proposed technique for KDEF, JAFFE and FERG-DB datasets. From the above table, it is clear that the proposed method attains better performance than existing techniques. Figure 2 (a)-(c) demonstrates the graphical performance comparison for KDEF, JAFFE and FERG-DB datasets.

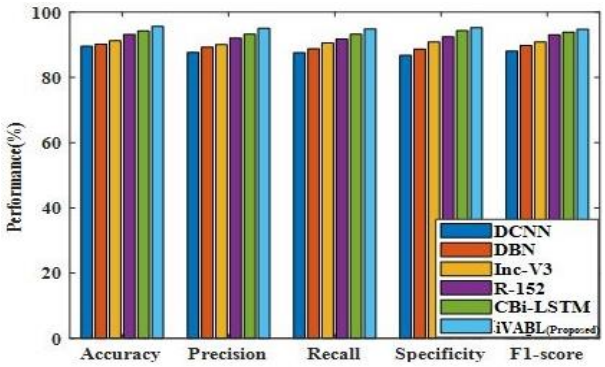

Fig.2.(a). Performance comparison for KDEF dataset

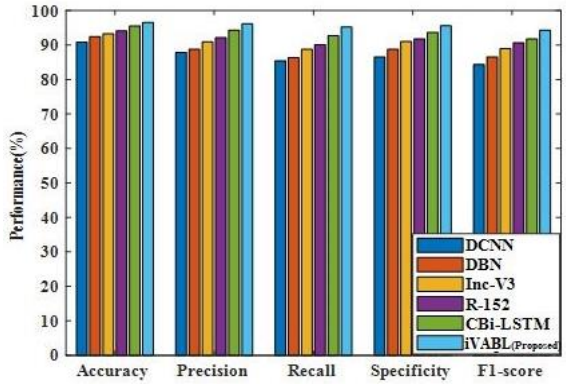

Fig.2.(b). Performance comparison for JAFFE dataset

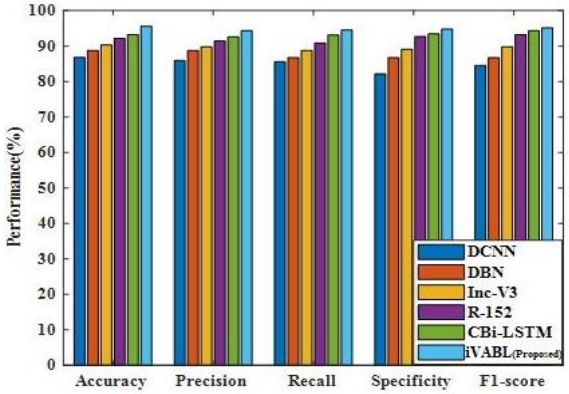

Fig.2.(c). Performance comparison for FERG-DB dataset

The existing techniques like DCNN, DBN, Inception V3 (Inc-V3), Resnet-152 (R-152), and deep convolutional Bi-LSTM (CBi-LSTM) are considered for comparison. The existing techniques attains poor outcome due to high misclassification of emotion states from the facial expression. In addition to this, the existing technique shows poor performance when huge number of datasets are trained. But the presented model attains better outcome due to accurate classification of emotions with low error. Improved accuracy is obtained in the proposed model because of high feature training capability, less overfitting issues and better convergence. Due to increase time complexity and degraded model learning, the existing models obtained only lesser accuracy. Figure 3 (a)-(c) illustrates the confusion matrix of proposed technique for KDEF, JAFFE and FERG-DB datasets.

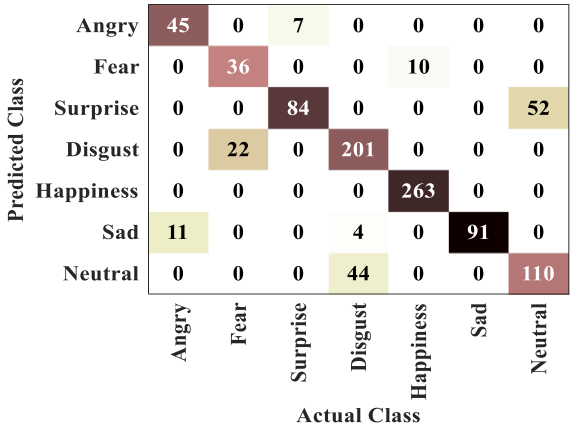

Fig.3.(a). Confusion matrix for KDEF dataset

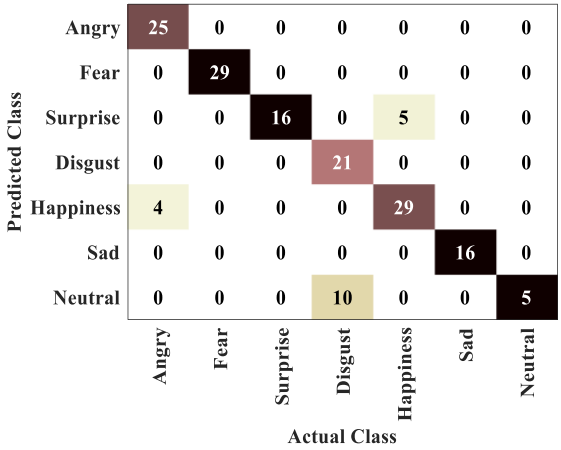

Fig.3.(b). Confusion matrix for JAFFE dataset

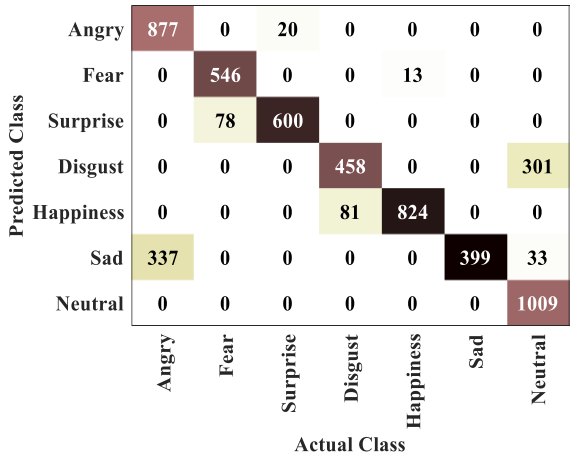

Fig.3.(c). Confusion matrix for FERG-DB dataset

The confusion matrix obtained by considering the batch size of 32 from the validation set of KDEF indicates 95.63% of accuracy. 45 images of angry, 36 images of fear, 84 images of surprise, 201 images of disgust, 263 images of happiness, 91 images of sad and 110 neutral images are effective predicted with the exact classes. In JAFFE dataset, 25 angry classes, 29 fear classes, 16 surprise classes, 21 disgust classes, 29 happiness classes, 16 sad classes and 5 neutral classes are accurately classified. For FERG-DB dataset, 877 angry classes, 546 fear classes, 600 surprise classes, 458 disgust classes, 824 happiness classes, 399 sad classes and 1009 neutral classes are accurately classified. Figure 4 (a)-(c) indicates the performance of computational time for testing the data.

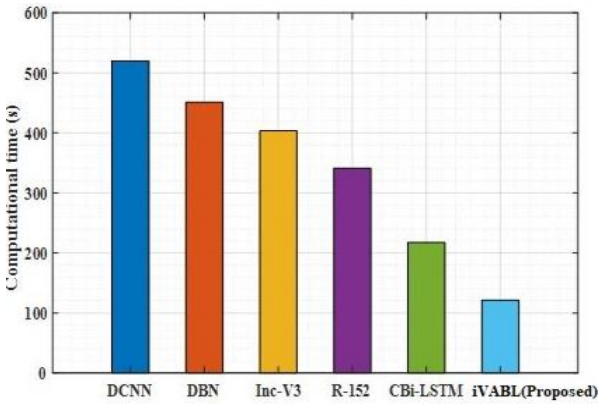

Fig.4.(a). Computational time for KDEF dataset

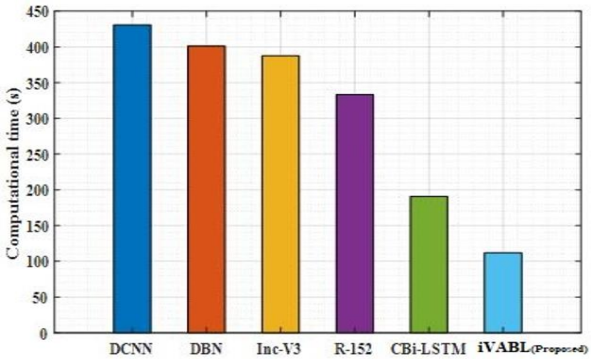

Fig.4.(b). Computational time for JAFFE dataset

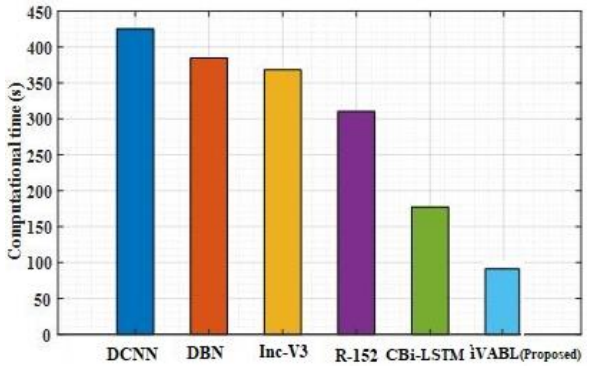

Fig.4.(c). Computational time for FERG-DB dataset

The computational testing time of proposed model has been compared with the existing models like DCNN, DBN, Inc-V3, R-152, and CBi-LSTM. The testing time of the single image input has been described in case of three datasets. The computational time of DCNN is obtained to be 520 seconds, DBN as 451.2 seconds, Inc-V3 as 403.54 seconds, R-152 as 341.24 seconds and CBi-LSTM as 217.54 seconds respectively. The iVABL model has consumed 121.34 seconds which is highly lesser when compared to the existing models for KDEF dataset. For JAFFE dataset, the proposed model attained 112.154 seconds and for FERG-DB dataset, 98.975 seconds is attained that are comparatively lesser than the existing methodologies.

5. Conclusions and Future Scope

The iVABL network is presented in this paper. Initially, the data are collected from KDEF, JAFFE and FERG-DB. The gathered images are pre-processed using AMF approach for removing the pixel noise. Followed by this, the pre-processed images are fetched to iVABL model where CNN based improved VGG-16 is incorporated for extracting features of interest and BiLSTM is used for classification. The weights of BiLSTM network are optimized through AOA approach for enhancing the overall efficiency and minimize the loss function. The advantage of iVABL network in classification is, promoting more dependency and increasing the classification accuracy. The proposed model performances are analyzed and compared with existing models like DCNN, DBN, Inc-V3, R-152 and CBi-LSTM. The accuracy of 95.63% is obtained on KDEF dataset, 96.61% is attained on JAFFE dataset and FERG-DB dataset obtained

the classification accuracy of 95.58%. Real time test results demonstrate the model's efficiency. The research presented is highly limited over facial images rather than videos and in future, emotions can also be detected through frames extracted from video data and by applying the appropriate deep learning model.

References

[1] Isha Talegaonkar, Kalyani Joshi, Shreya Valunj, Rucha Kohok, and Anagha Kulkarni, "Real time facial expression recognition using deep learning", in Proceedings of International Conference on Communication and Information Processing (ICCIP), 2019. DOI:10.2139/ssrn.3421486.

[2] M. Kalpana Chowdary, Tu N. Nguyen, and D. Jude Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications", Neural Computing and Applications, pp. 1-18, 2021. DOI: 10.1007/s00521-021-06012-8.

[3] Syed Aley Fatima, Ashwani Kumar, and Syed Saba Raoof, "Real time emotion detection of humans using mini-Xception algorithm", In IOP Conference Series: Materials Science and Engineering, Vol. 1042, No. 1, pp. 012-027. IOP Publishing, 2021. DOI:10.1088/1757-899X/1042/1/012027.

[4] Yahia Said, and Mohammad Barr, "Human emotion recognition based on facial expressions via deep learning on high-resolution images," Multimedia Tools and Applications, Vol. 80, No. 16, pp. 25241-25253, 2021.DOI: 10.1007/s11042-021-10918-9.

[5] Luu-Ngoc Do, Hyung-Jeong Yang, Hai-Duong Nguyen, Soo-Hyung Kim, Guee-Sang Lee, and In-Seop Na, "Deep neural network-based fusion model for emotion recognition using visual data", The Journal of Supercomputing, Vol. 77, No. 1, pp. 10773-10790, 2021.DOI: 10.1007/s11227-021-03690-y.

[6] Edeh Michael Onyema, Piyush Kumar Shukla, Surjeet Dalal, Mayuri Neeraj Mathur, Mohammed Zakariah, and Basant Tiwari, "Enhancement of patient facial recognition through deep learning algorithm: ConvNet", Journal of Healthcare Engineering, pp.1-8, 2021. DOI: 10.1155/2021/5196000.

[7] Shrey Modi, and Mohammed Husain Bohara, "Facial emotion recognition using convolution neural network", in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1339-1344, IEEE, 2021. DOI: 10.1109/ICICCS51141.2021.9432156.

[8] James Ren Lee, Linda Wang, and Alexander Wong, "Emotionnet Nano: An efficient deep convolutional neural network design for real-time facial expression recognition", Frontiers in Artificial Intelligence, Vol.3, pp. 609-673, 2021. DOI:10.3389/frai.2020.609673.

[9] Moutan Mukhopadhyay, Aniruddha Dey, RabindraNath Shaw, and Ankush Ghosh, "Facial emotion recognition based on textural pattern and convolutional neural network", in 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1-6. IEEE, 2021.

[10] Jun Liu, Yanjun Feng, and Hongxia Wang, "Facial expression recognition using pose-guided face alignment and discriminative features based on deep learning", IEEE Access, Vol. 9, pp. 69267-69277, 2021. DOI: 10.1109/ACCESS.2021.3078258.

[11] P.K. Das, A. Pradhan, S. Meher, "Detection of Acute Lymphoblastic Leukemia using Machine Learning Techniques", in book: Machine Learning, Deep Learning and Computational Intelligence for Wireless Communication, Springer, pp. 425.437, 2021. DOI:10.1007/978-981-16-0289-4_32.

[12] Binh T.Nguyen, Min H.Trinh, Tan V.Phan, Hien D.Nguyen, "An efficient real-time emotion detection using camera and facial landmarks." In: 2017 seventh international conference on information science and technology (ICIST); 2017. DOI: 10.1109/icist.2017.7926765.

[13] C. Loconsole, C.R. Miranda, G. Augusto, A.Frisoli, V. Orvalho, "Real-time emotion recognition novel method for geometrical facial features extraction." Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), Vol.1, pp.378-385, 2014. DOI: 10.5220/0004738903780385.

[14] Giuseppe Palestra, Adriana Pettinicchio, Marco Del Coco, Pierluigi Carcagn, Marco Leo, Cosimo Distante, "Improved performance in facial expression recognition using 32 geometric features." In: Proceedings of the 18th international conference on image analysis and processing. ICIAP, pp. 518–528, 2015. DOI:10.1007/978-3-319-23234-8_48.

[15] Aya Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," Informatics in Medicine Unlocked, vol. 20, 2020. DOI: 10.1016/j.imu.2020.100372.

[16] Mehmet Akif Ozdemir, Berkay Elagoz, Aysegul Alaybeyoglu, Reza Sadighzadeh, and Aydin Akan, "Real time emotion recognition from facial expressions using CNN architecture." In 2019 medical technologies congress (tiptekno), pp. 1-4. IEEE, 2019. DOI:10.1109/TIPTEKNO.2019.8895215.

[17] M. Jeong, B.C. Ko, "Driver's facial expression recognition in real-time for safe driving", Sensors ,18:4270, pp.1-17, 2018.

[18] M.A.H Akhand, S. Roy, N. Siddique, M.A.S. Kamal, M.A.S. and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," Electronics, Vol.10, No.9,1036, pp.1-19, 2021. DOI:10.3390/electronics10091036.

[19] H. Deeb, A. Sarangi, D. Mishra and S.K. Sarangi, "Human facial emotion recognition using improved black hole based extreme learning machine", Multimedia Tools and Applications, Vol.81, No.17, pp. 24529-24552, 2022. DOI:10.1007/s11042-022-12498-8.

[20] N. Kumari and R. Bhatia, "Efficient facial emotion recognition model using deep convolutional neural network and modified joint trilateral filter", Soft Computing, pp.1-19, 2021.

[21] X. Sun, P. Xia and F. Ren, "Multi-attention based deep neural network with hybrid features for dynamic sequential facial expression recognition", Neurocomputing, Vol. 444, pp. 378-389, 2021. DOI:10.1016/j.neucom.2019.11.127.

[22] R. Vedantham and E.S. Reddy, "A robust feature extraction with optimized DBN-SMO for facial expression recognition", Multimedia Tools and Applications, Vol. 79, No.29, pp.21487-21512, 2020. DOI: 10.1007/s11042-020-08901-x.

[23] D. Liang, H. Liang, Z. Yu and Y. Zhang, "Deep convolutional BiLSTM fusion network for facial expression recognition", The Visual Computer, Vol. 36, No.3, pp.499-508, 2020. DOI: 10.1007/s00371-019-01636-3.

[24] Shaik Asif Hussain 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. DOI: 10.1088/1742-

6596/1432/1/012087.
[25] Deepak Kumar Jain, Zhang Zhang, and Kaiqi Huang. "Multi angle optimal pattern-based deep learning for automatic facial expression recognition", Pattern Recognition Letters, Volume 139, pp.157-165, 2020. DOI:10.1016/j.patrec.2017.06.025.
[26] Pritam Khan, Yasin Khan, Sudhir Kumar, Mohammad S. Khan, Amir H. Gandomi, "HVD-LSTM based recognition of epileptic seizures and normal human activity", Computers in Biology and Medicine, 136, 104684, 2021. DOI: 10.1016/j.compbiomed.2021.104684.
[27] A. G. Dastider, F. Sadik, and S.A. Fattah, "An integrated autoencoder-based hybrid CNN-LSTM model for COVID-19 severity prediction from lung ultrasound", Computers in biology and medicine, Vol.132, 104296, 2021. DOI: 10.1016/j.compbiomed.2021.104296.
[28] S.T. Chavali, C.T. Kandavalli, T.M. Sugash, and R. Subramani, "Smart Facial Emotion Recognition with Gender and Age Factor Estimation", Procedia Computer Science, Vol. 218, pp.113-123, 2023. DOI: 10.1016/j.procs.2022.12.407.
[29] K. Bhangale, P. Ingle, R. Kanase, and D. Desale, "Multi-view multi-pose robust face recognition based on VGGNet", in book: Second International Conference on Image Processing and Capsule Networks: ICIPCN, pp. 414-421, 2021, Springer International Publishing. DOI: 10.1007/978-3-030-84760-9_36.
[30] Bhaskara Rao Jana, Haritha Thotakura, Anupam Baliyan, Majji Sankararao, Radhika Gautamkumar Deshmukh, and Santoshachandra Rao Karanam, "Pixel density based trimmed median filter for removal of noise from surface image", Applied Nanoscience, Vol. 13, No. 2, pp. 1017-1028, 2023. 10.1007/s13204-021-01950-0.
[31] G.S. Nijaguna, Ananda Babu J, B. D. Parameshachari, Rocío Pérez de Prado, and Jaroslav Frnda, "Quantum Fruit Fly algorithm and ResNet50-VGG16 for medical diagnosis", Applied Soft Computing, Vol. 136, 2023: 110055. DOI: 10.1016/j.asoc.2023.110055
[32] B. Gao, Yuan Shi, Fengqiu Xu, and Xianze Xu. "An improved Aquila optimizer based on search control factor and mutations." Processes, Vol. 10, No. 8, 2022: 1451. DOI: 10.3390/pr10081451
[33] D.Lundqvist, A. Flykt & A. Öhman, "The Karolinska Directed Emotional Faces - KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, ISBN 91-630-7164-9,1998.
[34] Michael J. Lyons, Miyuki Kamachi, Jiro Gyoba, "Coding Facial Expressions with Gabor Wavelets (IVC Special Issue)", DOI: 10.5281/zenodo.4029679.
[35] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, B. Mones, "Modeling stylized character expressions via deep learning," Asian Conference on Computer Vision, pp. 136–153, 2016. DOI:10.1007/978-3-319-54184-6_9.
[36] Nimra Bukhari, Shabir Hussain, Muhammad Ayoub, Yang Yu, and Akmal Khan, "Deep learning based framework for emotion recognition using facial expression", Pakistan Journal of Engineering and Technology, Vol.5, No. 3, pp.51-57, 2022. DOI:10.51846/vol5iss3pp51-57.
[37] W. Xu and Rayan S. Cloutier, "A facial expression recognizer using modified ResNet-152", EAI Endorsed Transactions on Internet of Things, Vol. 7, No. 28, 2022. DOI:10.4108/eetiot.v7i28.685.

Authors' Profiles

Rakshith M. D. is working as an Assistant Professor with 11 years of experience in the Department of Computer Science and Engineering at Canara Engineering College, Bantwal which is affiliated to Visvesvaraya Technological University, Belagavi. His research interests include computer vision and deep learning. He is the author of more than 6 papers in refereed national and international journals.

Harish H. Kenchannavar is working as Professor in the Department of Information Science and Engineering, Gogte Institute of Technology, Belagavi. He has 21 years of teaching experience and 10 years of research experience. His research contributions are in the field of Wireless sensor network, Quality of Service (QoS), Computer Vision and system modeling. He has vast experience in Academic, research and coordination at University, College and Departmental level. He is Life member to computer society of India and ISTE professional bodies. He has published 20 + research papers at national and international conference in India as well as outside country. He was college-level National Accreditation Board (NBA) coordinator.

How to cite this paper: Rakshith M. D., Harish H. Kenchannavar, "Hybrid Deep Optimal Network for Recognizing Emotions Using Facial Expressions at Real Time", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.3, pp.47-58, 2024. DOI:10.5815/ijisa.2024.03.04