Iraqi Journal of Science, 2026, Vol. 67, No. 3, pp: 1698-1714
DOI: 10.24996/ijs.2026.67.3.33

ISSN: 0067-2904

Facial Expression Recognition Using Deep Learning EfficientNetB0

Amal Sufiuh Ajrash*, Wildan Jameel Hadi, Ammar Hussein Jassim, Nada Khaleel Kareem, Rasha Mohamed Jaafar Sadiq
Department of Computer Science, College of Science for Women, University of Baghdad, Baghdad, Iraq.

Received: 28/7/2024 Accepted: 10/ 3/2025 Published: 30/3/2026

Abstract

Natural settings make it challenging to identify facial expressions since head position, illumination level, and occlusion vary. Thus, developing a more generic model without front-facing images alone is quite crucial. This research proposes a facial expression recognition model based on pre-trained deep convolutional neural networks with transfer learning. The model was trained on several cases to classify face expressions into seven classifications efficiently. The proposed system used the EfficientNetB0 model that has one dense dropout layer. The model first rescales and norms the input dataset in the input layer that takes images of a larger resolution to get better results. After entering 7 blocks sequential in each one, the data convolution two times, then speeding up training and avoiding overfitting by adding a dropout layer and batch normalization layer. The model achieves an accuracy of 70.60% when features are frozen, and the classifier is unfrozen. In contrast, the Fine Tune model achieves the highest accuracy, 72.69%, by unfreezing the feature extractor and training the entire model.

Keywords: Augmentation, Convolutional neural network, Deep learning, EfficientNetB0 model, Face expression recognition.

التعرف على تعابير الوجه باستخدام خوارزمية EfficientNetB0في التعلم العميق

امال سفيح عجرش*, ولدان جميل هادي , عمار حسين جاسم, ندى خليل كريم, رشا محمد جعفرصادق
قسم علوم الحاسوب، كلية العلوم للبنات، جامعة بغداد، بغداد، العراق

الخلاصة

تجعل الإعدادات الطبيعية من الصعب التعرف على تعبيرات الوجه نظرًا لاختلاف وضع الرأس ومستوى الإضاءة والانسداد. وبالتالي، فإن تطوير نموذج أكثر عمومية بدون صور مواجهة للأمام وحدها أمر بالغ الأهمية. يقترح هذا البحث نموذجًا للتعرف على تعبيرات الوجه يعتمد على شبكات عصبية ملتوية عميقة مدربة مسبقًا مع التعلم الانتقالي. قم بتدريب النموذج على عدة حالات لتصنيف تعبيرات الوجه بكفاءة إلى سبعة تصنيفات. استخدم النظام المقترح نموذج EfficientNetB0 الذي يحتوي على طبقة تسرب كثيفة واحدة. يقوم النموذج أولاً بإعادة قياس ومعايرة مجموعة البيانات المدخلة في طبقة الإدخال التي تلتقط صورًا بدقة أكبر للحصول على نتائج أفضل، بعد إدخال 7 كتل متتالية في كل منها، يتم التفاف البيانات مرتين ثم تسريع التدريب وتجنب الإفراط في الملاءمة عن طريق إضافة طبقة تسرب وطبقة تطبيع الدفعة. يحقق النموذج دقة 70.60٪ عندما يتم تجميد الميزات وإلغاء تجميد المصنف. في المقابل، يحقق نموذج الضبط الدقيق أعلى دقة بنسبة 72.69% من خلال إلغاء تجميد مستخرج الميزات وتدريب النموذج بأكمله.

*Email: amalsa_comp@csw.uobaghdad.edu.iq

1. Introduction

A person can express his psychological state using his facial expressions. This allows a person to convey approximately 55% of the information in a non-verbal form and the rest through speech. From this, we conclude that distinguishing facial expressions is one of the most important tasks in the computer field under many circumstances [1] [2]. Understanding facial expressions is an important thing between humans and for human interaction with machines. The importance of interpreting facial expressions lies in the fact that they differ from one person to another and what their mind interprets at that moment [3]. On the other hand, for the natural interaction between a person and a machine to succeed, there must be basic standards based on strong models to distinguish the facial expressions of humans [4] [5]

One of the important research areas of human-machine interaction is Facial Expression Recognition (FER) which is capable of detecting human emotions by analyzing facial expressions [1]. Also, the detection of human emotion plays an important role in many areas like multimedia, robotics, etc. [6]. The FER has many applications for investigating human-computer interaction, computer vision, and non-human behavior. This is considered one of the difficult tasks due to the presence of backgrounds, the unclear accuracy of the faces in these images, and the similarity of the procedure in different head positions [7]. The complexity lies in interpreting emotions, anger, and the effects of behavior. Therefore, in this research, a model was designed to distinguish different emotions from facial expressions, such as anger, happiness, sadness, fear, disgust, surprise, and neutrality.

Many techniques support the FER field, but in this work we used one of the frequently employed deep neural networks the Convolutional Neural Network (CNN) [8] [9] [10] [11]. A CNN consists of a convolution layer, a pooling layer, and a dense (fully connected) layer. The convolution layer represents the core of CNN, which is responsible for data processing. The pooling layer is responsible for reducing dimensions using max-pooling, average-pooling, and sum [12] [13].

2.Related Work

Much research has been discussed for either classical or deep learning-based methods for FER. This section will display a summary of the most modern approaches.

In [14], the research proposes using a Standalone-Based Neural Network (SBNN) and Ensemble-Based Neural Network (EBNN) approaches. The proposed network is classified as SBNN with 16 (VGG-16) as the classification model pre-trained on the ImageNet dataset and fine-tuned for emotion classification, where the model was modified into using 13 convolutional layers and GAP as the last pooling layer. The classification is performed on the FER-2013 dataset, and the proposed model results are 69.40% accurate.

In [15], the researchers presented two techniques; the first one, Convolutional Neural Networks (CNNs), classify facial expression without taking the temporal features into consideration. This model used images and video. The second technique, which used temporal information to classify the facial emotions in contrast to full images, is fed to the Long Short Term Memory (LSTM) network to use the good advantage of temporal information. Also, this method uses the good advantage of different techniques like the Product Fusion Method (PFM), Average Fusion Method (AFM), and Multimodal Compact Bilinear Pooling (MCBP). This system uses the FABO dataset that presents the bi-modal face and body data that gets 77.7 accuracy in facial expressions and 76.8 in upper body movement, and the FER-2013 dataset gets 90.42 in facial expressions and 79.27 in upper body movement. The author used more across database training network parameters to get better generalization capability.

In [16], a proposed model for face expression identification combines elements of both hybrid and deep learning. The suggested CNN model successfully identified primary and secondary expressions, such as sadness and happiness, as well as secondary expressions such as astonishment and anger. The highest level of precision was attained using a dropout rate of 0.2. The model achieved 97.07% and 94.12% accuracy on the FER2013 and JAFFE datasets, respectively. The authors need to test the model on real time video to ensure it gets a high accuracy result.

In [17], the authors proposed a FER method based on a two-stream convolutional neural network. The model feeds the input weights at every level of convolution input and then uses soft attention mechanism modules on the space-time features of the combination of static and dynamic streams. Then, it applies a lighting preprocessing chain algorithm to remove most lighting effects. The recognition rate of this method on the AFEW6.0 dataset is 95.05%, and on the Multi-PIE dataset is 61.40%. The disadvantage of this model is that the sample size of existing natural scene expression databases is relatively limited, especially when it comes to natural scene video databases. Deep learning methods generally require large amounts of training data that can't be found in the datasets the researcher used.

In [18], the researchers introduce fast EfficientNetV2 models. In this model, they increase the image size during training, adaptively adjusting regularization (data augmentation) and image size, and optimizing with training aware neural architecture search and model scaling where it trains fast up to 11x. They test the system on ImageNet and CIFAR/Cars/Flowers datasets. The accuracy reaches 87.3%

In [19], A system using different pre-trained models is implemented. The proposed model has two pre-trained models applied after changing classifiers and constructing the hybrid model. The system includes a hybrid deep neural network system of EfficientNetB0 and MobileNetV2 in the extraction part and two Dense-Dropout layers in the classifier part, which apply the model to the FER2013 database. The proposed system has a good accuracy result of 74.39% for the hybrid model and 73.33% for fine-tuning the single EfficientNetB0 model.

In [20], the authors enhanced FER performance on the FER2013 database by combining pre-trained deep learning architectures AlexNet, ResNet50, and Inception V3. Combining multi-pre-trained architectures can yield a perfect recognition outcome where the system's accuracy rate is 73.56%.

In [21], The authors propose a new approach for the FER system that uses Local Binary Patterns (LBP) and Histograms of Oriented Gradients (HOG) as feature extraction and Quantum Gazelle Optimization algorithm (QGO) as feature selection, then an optimized deep neural network as a classifier. The system adopted FER2013 dataset to test the work with accuracy of 93.26 .

3. The EfficientNetB0 Architecture

In general, the CNN network relies on feeds with a large amount of data to discover the hidden properties during training, which takes a long time and may take several days [22]. Therefore, instead of training such networks from scratch, it is possible to use pre-trained networks with approved weights instead of relying on random weights. This behavior reduces training time and avoids many problems by freezing the weights of specific layers [23] [24]. This technique is known as transfer learning, which works on image classification [25] [26] [27].

The EfficientNet architecture has been proven to be highly efficient regarding both computational resources and accuracy, making it a popular choice for various image recognition tasks. EfficientNet possesses remarkable capabilities in the area of feature

extraction. It has a lower number of parameters and a higher level of accuracy [28]. EfficientNetB0 has an efficient and accurate architecture of FER, leading to better recognition performance. The first thing is any network is its stem, after which all the experimenting with the architecture starts which is common in the EfficientNet B0 model, and the final layers [29] [30], the layers shown in Figure 1.

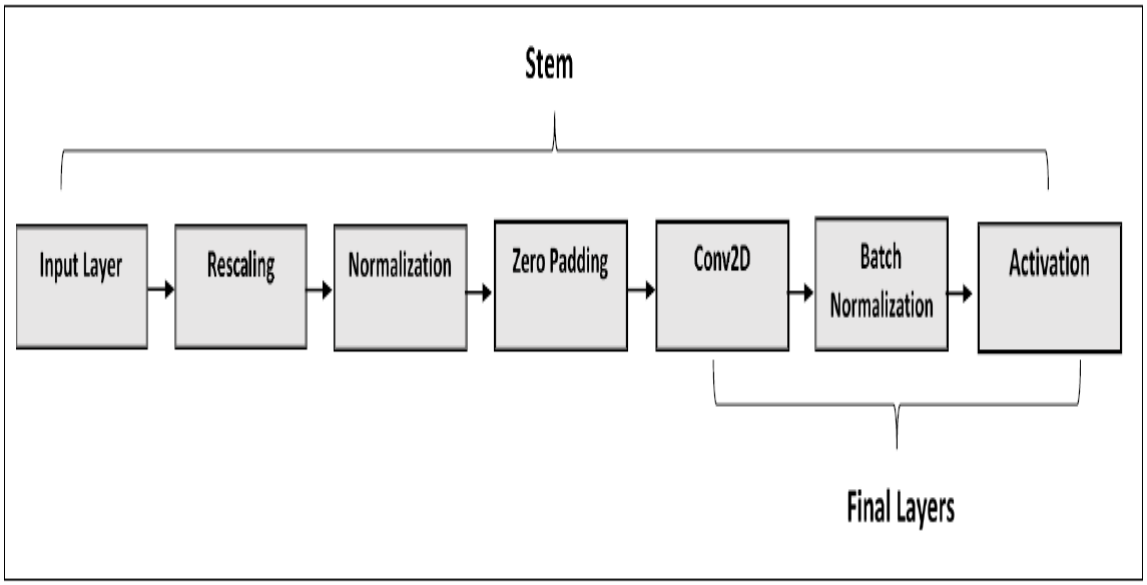

Figure 1: EfficientNetB0 model layers [30]

4. Evaluation Metrics

Many metrics that evaluate the accuracy of CNN based systems, including accuracy, the simple and best metric, F1-Score, recall, and precision. These systems used to measure the accuracy of the model depending on four basic concepts, which include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), another way through which it is possible to measure the accuracy of the system is based on the confusion matrix [31] [32] [33].

4.1 *Accuracy*

Accuracy is a measure of the accuracy of the CNN model based on the percentage of expectations for the classes used. This measure provides accurate results if the observations used to measure the system's accuracy are balanced. The following Eq. (1) shows how the measure is computed:

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (1)$$

4.2 *Precision*

Positive Predictive Value (PPV or precision) It is used to determine whether the observations that were classified within a certain category belong to that category, meaning that this model is suitable for this category or not. The following Eq. (2) shows how the measure is computed:

$$Precision = \frac{TP}{(TP+FP)} \quad (2)$$

4.3 *Recall*

Recall is used to know the percentage of positive observations that were correctly predicted. This measure is of great importance in some places. Eq. (3) shows how the measure is computed

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

4.4 *F1-score*

This measurement is very important in models where the data is not evenly distributed within the categories used. This measure is based on recall and precision. Eq. (4) shows how the measure is computed:

$$F1 - score = \frac{Precision * Recall}{(Precision + Recall)} \quad (4)$$

5. Problem Statement

The FER problem is a significant challenge, necessitating the development of a highly effective system that identifies expression for individuals accurately. The dataset's facial images exhibit a wide range of pose variations , making it challenging to establish robust feature representations invariant to expression and effect on face changes. Low-resolution images and varying sizes contribute to information loss and hinder extracting crucial facial features for recognition. The dataset also contains diverse facial expressions demanding the recognition system to handle dynamic facial variations accurately. The ultimate goal is to create a reliable feature in the expression recognition system qualified for performing effectively in real-world scenarios, even with limited data and varied conditions.

6. Proposed Methodology

The structure of the proposed FER system is outlined here, as shown in Figure 2. This system preprocesses the images before feeding them into a selected CNN model that employs the EfficientNetB0 algorithm to predict facial expressions: neutral, happiness, sadness, anger, disgust, surprise, and fear.

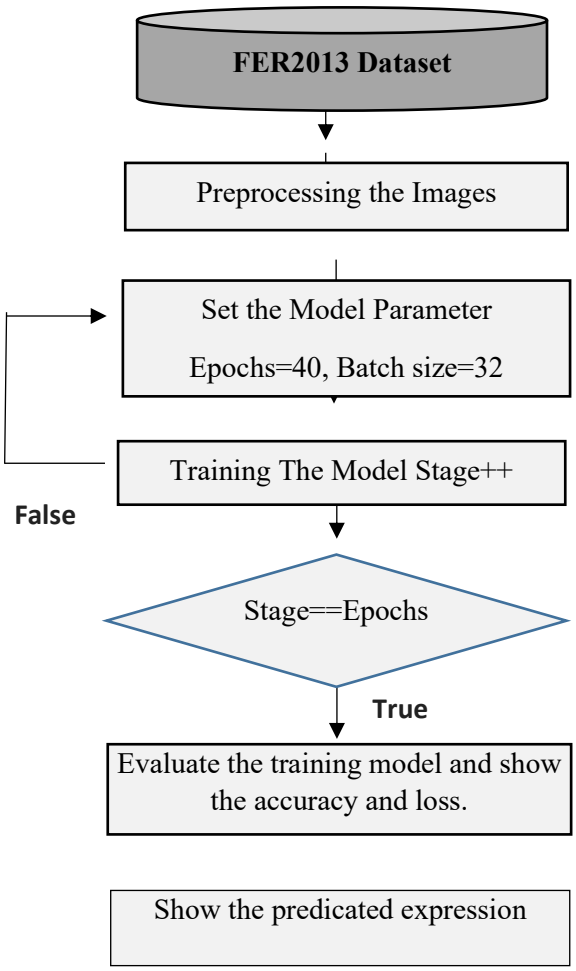

Figure 2: The structure of the proposed FER system

6.1 Dataset and Preprocessing.

The dataset used in this study is FER2013 [34]. The dataset comprises 35,887 instances of 48x48 pixel grayscale images of cropped faces. It consists of 28709 images in the training set and 7178 images in the public testing set. These images depict seven different expressions, including neutral, happy, sorrow, anger, disgust, surprise, and fear. The FER2013 dataset exhibits a significant level of imbalance due to variations in the number of images for each expression, as well as differences in rotation and poor illumination. These factors provide challenges for the model's performance across all expressions. Illustrations of these images are shown in Figure 3.

Figure 3: Samples of images for different expressions from FER2013 Dataset [1][8].

The allocation of test classes (facial expressions) in the FER 2013 dataset is illustrated in Figure 4. In addition to the original data, an augmentation process is used to generate new training samples from the current training set to achieve generality in the model and prevent it from overfitting the training outcomes. The data augmentation used (Zoom, Horizontal_ Flip, Rotate, Width_Shift, and Height_Shift). Randomly flip the image right and left, the range of shifting the image vertically and horizontally (translation) is set to 0.5 value, randomly rotating the image max ten percentages, and the range of zoom to the image is set to 0.9, this process increasing images size and converting it into RGB. After that, the preprocessing is responsible for resizing the image (200,200) pixels.

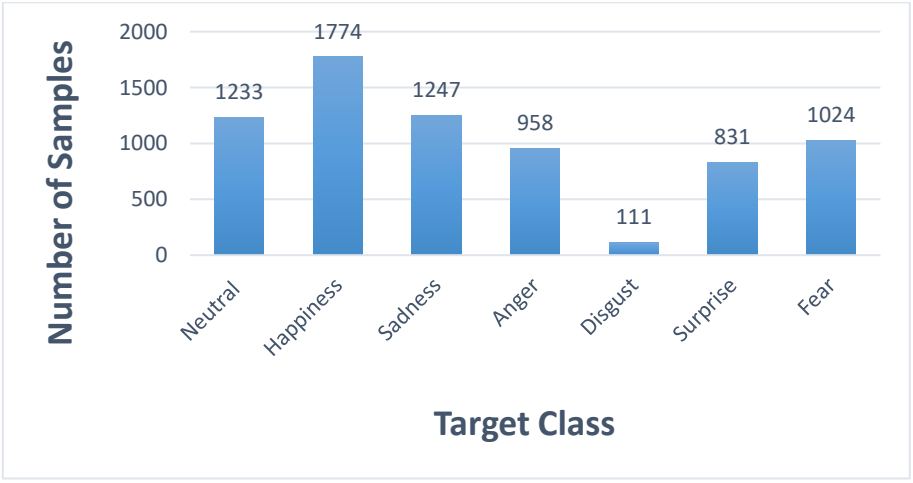

Figure 4: Distribution of Target Classes (facial expressions) in the FER2013 Dataset.

6.2 Convolution Neural Network

CNNs can handle large volumes of data, making them suitable for the vast and diverse FER2013 dataset. This allows for the training of robust models capable of generalizing across the wide content in the different datasets. The fundamental configuration of the CNN is depicted in Table 1, which displays the model parameters. The embedding layer takes features and is mapped to a dimensional embedding vector. The Convolutional Layer uses the Rectified Linear Unit (ReLU) Activation Function that denotes the convolutional operations. The Average Pooling Layer (APL) reduces the dimensionality of the output from the dropout layer by applying operation over specified regions, effectively halving the size of the feature maps after each Conv2D layer. In dropout layers, apply dropout to the output of each Conv2D layer and before the dense layer, setting a fraction of input units to 0 to help prevent overfitting is 0.1 for Conv2D layers and 0.5 before the dense layer. In the output layer, use softmax activation function. An embedding layer with 61,080,283 parameters indicates a large vocabulary size and embedding dimensionality of 100. This layer is crucial for converting input images into their expression, dense vectors that can be processed by the neural network.

Table 1: Model parameter setting

Layer Type	Output Shape
EMBEDDING	(200, 200, 3)
CONVOLUTION 2D	(100, 100, 32)
AVGPOOLING2D1	(100, 100, 32)
CONV2D_2	(100, 100, 512)
AVGPOOLING2D2	(50,50,96)
DROPOUT_1	(50, 50, 512)
AVGPOOLING2D3	(7,7,1280)
DROPOUT_2	(24, 24, 512)
FLATTEN	(None,1280)
DROPOUT	(None, 1280)
DENSE	(7)

Seven convolutional layers are designed to extract features from the embedded input with five hundred and twelve filters in each layer to capture various patterns in the sequence data. The convolutional layers are interspersed with dropout layers to prevent overfitting. Used after each Conv2D layer and before the dense layer at the end, dropout layers help reduce overfitting by randomly ignoring a subset of neurons during training, making the network more robust and preventing it from relying too much on any one neuron. Global average pooling of 2D layers lowers the spatial representation size, reducing network parameters and processing. This procedure extracts key features and reduces overfitting. The flattening layer transforms the output of the APL from a multi-dimensional tensor to a one-dimensional tensor, making it possible to feed it into the final dense layer, which outputs 7 neurons, each representing a class. Therefore, the ReLU function is connected at the end of the output layer. That layer is a mathematical function that converts a vector of numbers into a vector of probabilities, where the probabilities of each value are proportional to the exponentials of the input numbers.

6.3 The EfficientNetB0 model Architecture

In this work, an EfficientNetB0 was used as the smallest CNN model from the EfficientNet family with one dense dropout layer. The model first rescales and norms the

input dataset in the input layer that takes images of a larger resolution to get better results. After entering 7 blocks sequential in each one, the data convolution two times then speed up training and avoid overfitting by adding a dropout layer and batch normalization layer. After each output layer used the activation ReLU function, the last layer of the model is a fully connected dense layer where the output is passed through a dense layer with 7 units (that correspond to the 7 different facial expressions), using the softmax activation function. This layer serves as the feature extractor for recognizing facial expressions, as it outputs the probability distribution of the input image belonging to each of the 7 facial expression classes. During training, the model learns to minimize the loss between the predicted probabilities and the true labels of the facial expressions in the training dataset, leading to a model that can accurately classify facial expressions in real world scenarios. Table 2 displays EfficientNetB0 model parameters settings.

The proposed model aims to get the best result in a lower number of epochs, so it tries to improve the standard structure of the EfficientNetB0 work and get the best result for FER classification. Fine-tuning taking the weights of a trained neural network and using it as initialization for a new model being trained on data from the same domain.

Table 2: Training Parameters

Parameter	Value
INPUT IMAGE SIZE	200x200
IMAGE COLOR SPACE	RGB
DROPOUT RATE	0.3
LEARNING RATE	0.001
BATCH SIZE	32
EPOCHS	40
OPTIMIZER	Adam
ACTIVATION	ReLU
OUTPUT CLASSES	7

The model was trained and fine-tuned in the two cases, as shown in the following steps:
Step 1: The model was trained by freezing the feature extraction components and unfreezing the classifier for 40 epochs. The highest accuracy achieved was 0.70. The accuracy and loss of the model were presented in Figure 5 when training started using the frozen feature extractor.
Step 2: Implement the Fine Tune model by allowing the feature extractor to be modified and training the entire model for 40 epochs and get a higher accuracy value of 0.72. The results are visualized in Figure 7.

7. Results and Discussion

The proposed model was executed on high-end device specs, as indicated in Table 3. Upon completing the model's training, its performance was assessed by measuring accuracy, loss, and the confusion matrix.

Table 3: The device specifications

Device	Specifications
CPU	Ryzen p9 3900x
RAM	DDr4 64GB 3200MHz
GPU	Dalx NVidia 2080Ti 11GB
STORAGE	Nvme SSD 1 TB
OS	Windows10 v2004
ENVIRONMENT	Python3.8, Tensorflow2.9.0

As the suggested model EfficientNetB0 is being trained to classify whether an image contains one of the standard expressions. It's important to know the total number of layers in EfficientNetB0 architecture; the total is 237. Table 4 displays the number of layers for each stage, the input resolution, and displays the output channels.

Table 4: EfficientNetB0 baseline Network

Stage	Resolution(Hxw)	Channels/Features	Layers
1	100x100	32	1
2	100x100	96	2
3	50x50	24	2
4	13x13	80	3
5	13x13	112	3
6	7x7	192	4
7	7x7	1280	2
FINAL	-	7	1

The model was assessed on the FER2013 dataset. The training of the model consisted of freezing the feature extraction components and then unfreezing the classifier for a total of 40 epochs. During the forty times, the maximum accuracy that was attained was 0.70. When starting training the feature extractor was frozen, as shown in the training results in Table 5 and Figure 5 that illustrates the model's accuracy and the loss that occurred when training began using the frozen feature extractor and note that the learning rate was still static in all epochs.

Table 5: Start Training with Feature Extractor Frozen

Epoch Number	Loss Training Data	Accuracy Training Data	Loss Valid Data	Accuracy Valid Data	Learning Rate
1	1.1825	0.5517	1.0522	0.6141	0.0010
2	0.9957	: 0.6278	1.0360	0.6205	0.0010
3	0.9280	0.6543	0.9495	0.6436	0.0010
4	0.8804	0.6722	0.9672	0.6392	0.0010
5	0.8278	: 0.6951	0.9620	0.6498	0.0010
6	0.7867	0.7074	0.9189	0.6645	0.0010
7	0.7498	0.7205	0.8981	0.6790	0.0010
8	0.7017	0.7410	0.8907	0.6826	0.0010
9	0.6664	0.7524	0.8956	0.6890	0.0010
10	0.6198	: 0.7715	0.9795	0.6654	0.0010
11	0.5799	0.7864	0.9222	0.6846	0.0010
12	0.5418	0.8013	0.9537	0.6838	0.0010
13	0.5049	0.8145	1.0883	0.6723	0.0010
14	0.4708	0.8258	0.9655	0.6832	0.0010
15	0.4392	0.8386	1.0271	0.6904	0.0010
16	0.4082	0.8503	1.0727	0.6918	0.0010
17	0.3795	0.8594	1.2174	0.6838	0.0010
18	0.3545	0.8700	1.2029	0.6824	0.0010
19	0.3326	0.8787	1.1874	0.6840	0.0010
20	0.3145	0.8858	1.1814	0.6985	0.0010
21	0.2960	0.8929	1.2285	0.6879	0.0010
22	0.2737	0.9010	1.2198	0.7005	0.0010
23	0.2664	0.9041	1.2605	0.6999	0.0010
24	0.2566	0.9060	1.2885	0.6974	0.0010
25	0.2375	0.9126	1.3411	0.6854	0.0010
26	0.2432	0.9102	1.3553	0.6840	0.0010
27	0.2229	0.9194	1.2929	0.7003	0.0010
28	0.2177	0.9219	1.3357	0.7008	0.0010
29	0.2121	0.9230	1.4262	0.7013	0.0010
30	0.2004	0.9276	1.3983	0.7047	0.0010
31	0.2015	0.9275	1.3600	0.7058	0.0010
32	0.1809	0.9342	1.4964	0.6960	0.0010
33	0.1892	0.9329	1.4693	0.6840	0.0010
34	0.1806	0.9341	1.4490	0.7013	0.0010
35	0.1749	0.9383	1.4990	0.6890	0.0010
36	0.1751	0.9359	1.5051	0.6963	0.0010
37	0.1652	0.9411	1.4181	0.7038	0.0010
38	0.1623	0.9410	1.4861	0.7041	0.0010
39	0.1579	0.9446	1.4624	0.7052	0.0010
40	0.1526	0.9461	1.5213	0.7060	0.0010

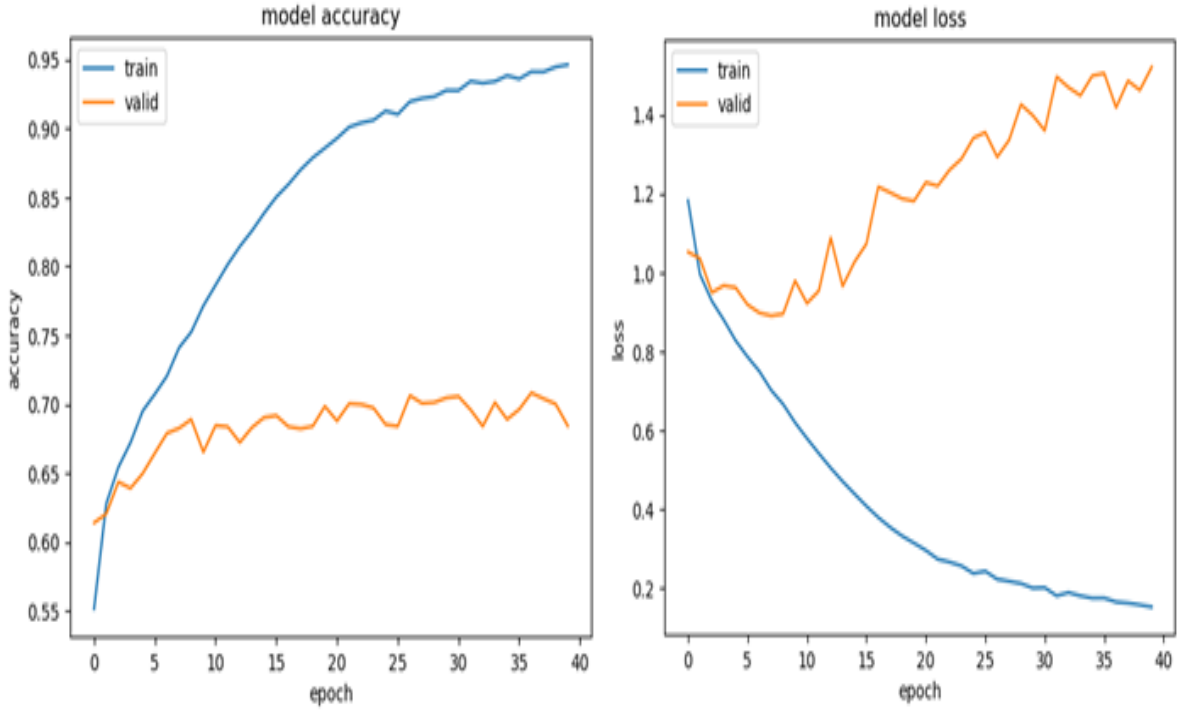

Figure 5: The proposed EfficientNet-B0 model the accuracy and loss result in freezing the feature extraction.

Figure 6 shows the confusion matrix, and Table 6 shows the classification report on a training set of FER2013 of EfficientNet-B0 Model for step 1, where freezing the feature extraction components and unfreezing the classifier got a good result in recognition.

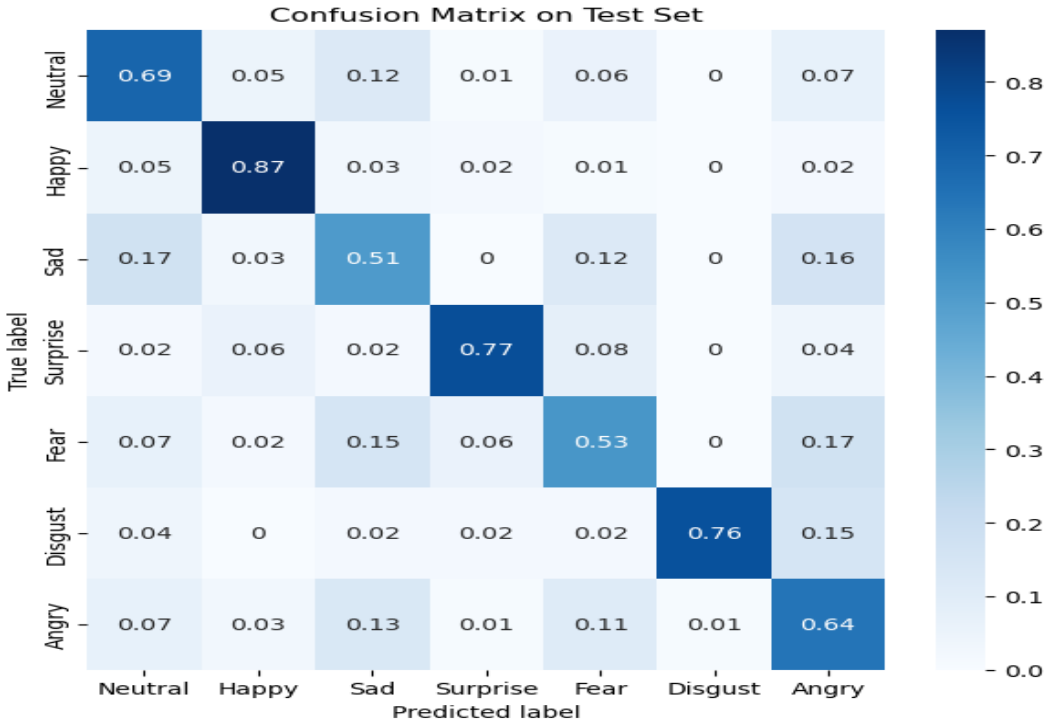

Figure 6: Confusion matrix of classification result of freezing the feature extraction.

Table 6: Classification report of freezing the feature extraction.

Expression	Precision	Recall	F1 -Score	Support
Neutral	0.65	0.69	0.67	626
Happy	0.88	0.87	0.88	879
Sad	0.55	0.51	0.53	594
Surprise	0.84	0.77	0.80	416
Fear	0.57	0.53	0.55	528
Disgust	0.78	0.76	0.77	55
Angry	0.53	0.64	0.58	491
Accuracy			0.70	3589
Macro Avg	0.69	0.68	0.68	3589
Weighted Avg	0.69	0.68	0.68	3589

The next step in the proposed system is allowing the feature extractor to be updated and training the entire model, which are both necessary steps in the implementation of the Fine Tune model. Table 7 and Figure 7 state that the number of epochs reaches 40 when the loss values stop changing and the accuracy remains stable at 0.72.

Table 7: Unfreezing the feature extractor and training the whole model

Epoch Number	Loss Training Data	Accuracy Training Data	Loss Valid Data	Accuracy Valid Data	Learning Rate
1	0.1672	0.9397	1.4836	0.7035	0.0010
2	0.1633	0.9427	1.4443	0.7130	0.0010
3	0.1596	0.9421	1.4625	0.7033	0.0010
4	0.1474	0.9478	1.5156	0.7052	0.0010
5	0.1555	0.9443	1.4838	0.7164	0.0010
6	0.1458	0.9468	1.5087	0.7019	0.0010
7	0.1438	0.9489	1.5594	0.6938	0.0010
8	0.1421	0.9485	1.5980	0.7027	0.0010
9	0.1404	0.9504	1.5587	0.7041	0.0010
10	0.1324	0.9533	1.7045	0.6941	0.0010
11	0.1338	0.9532	1.6024	0.7094	0.0010
12	0.1354	0.9523	1.5751	0.7060	0.0010
13	0.1230	0.9571	1.6044	0.7133	0.0010
14	0.1241	0.9560	1.6696	0.7086	0.0010
15	0.1192	0.9576	1.6037	0.7161	0.0010
16	0.0788	0.9729	1.6922	: 0.7264	5.0000e-04
17	0.0674	0.9754	1.7239	0.7214	5.0000e-04
18	0.0616	0.9770	1.7398	0.7230	5.0000e-04
19	0.0604	0.9778	1.7825	0.7233	5.0000e-04
20	0.0575	: 0.9793	1.8513	0.7197	5.0000e-04
21	0.0575	0.9786	1.7175	0.7278	5.0000e-04
22	0.0541	0.9802	1.8735	0.7239	5.0000e-04
23	0.0561	0.9794	1.8575	0.7225	5.0000e-04
24	0.0540	0.9802	1.8914	0.7233	5.0000e-04

25	0.0481	0.9825	1.9980	0.7233	5.0000e-04
26	0.0527	0.9807	1.8816	0.7164	5.0000e-04
27	0.0555	0.9798	1.7844	0.7183	5.0000e-04
28	0.0536	0.9813	1.8245	0.7214	5.0000e-04
29	0.0511	0.9806	1.9361	0.7236	5.0000e-04
30	0.0514	0.9825	1.9461	0.7219	5.0000e-04
31	0.0497	: 0.9820	1.8675	0.7180	5.0000e-04
32	0.0370	0.9857	1.9520	0.7228	2.5000e-04
33	0.0303	0.9881	2.0019	0.7311	2.5000e-04
34	0.0305	0.9880	2.0180	0.7269	2.5000e-04
35	0.0299	0.9887	2.0920	0.7281	2.5000e-04
36	0.0284	0.9894	2.0739	0.7250	2.5000e-04
37	0.0270	0.9894	2.1518	0.7261	2.5000e-04
38	0.0286	0.9890	2.1297	0.7272	2.5000e-04
39	0.0266	0.9898	2.1209	0.7236	2.5000e-04
40	0.0253	0.9901	2.1687	0.7242	2.5000e-04

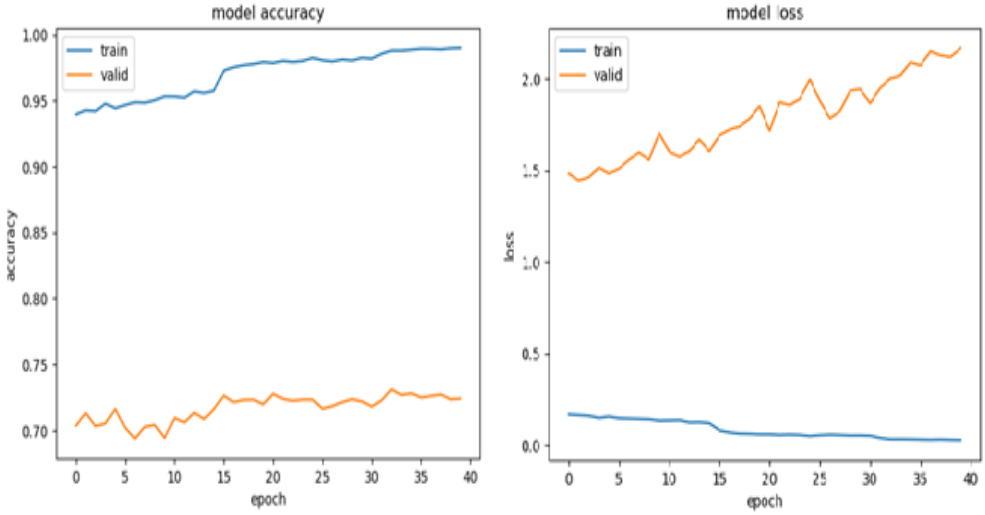

Figure 7: The proposed EfficientNet-B0 model shows the accuracy and loss result when unfreezing the feature extraction.

The confusion matrix is shown in Figure 8, and the classification report in Table 8 on fine tune model of a training set of FER2013 of step 2 where unfreezing the feature extractor.

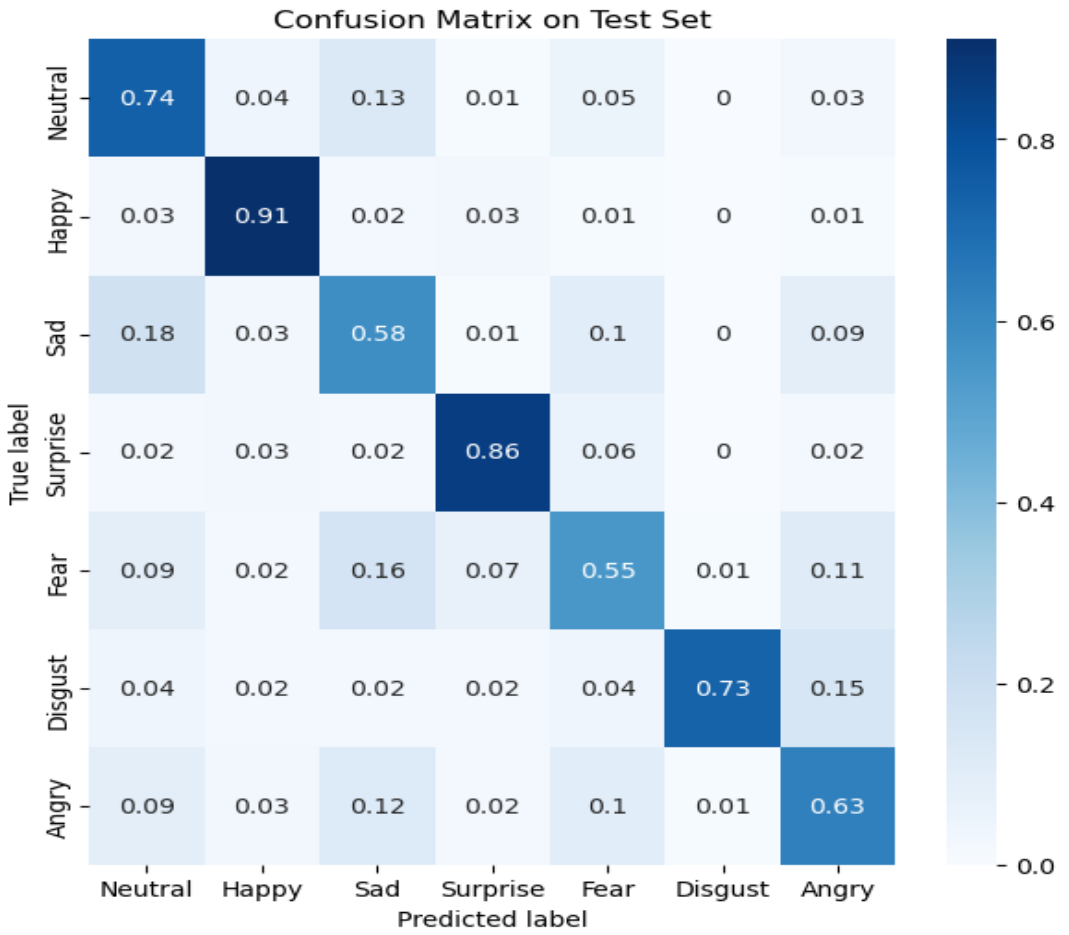

Figure 8: Confusion matrix of classification result of unfreezing the feature extractor.

Table 8: Classification report of unfreezing the feature extraction.

Expression	Precision	Recall	F1-Score	Support
Neutral	0.66	0.74	0.70	626
Happy	0.91	0.91	0.91	879
Sad	0.58	0.58	0.58	594
Surprise	0.81	0.86	0.84	416
Fear	0.62	0.55	0.58	528
Disgust	0.83	0.73	0.78	55
Angry	0.67	0.63	0.65	491
Accuracy			0.72	3589
Macro Avg	0.73	0.71	0.72	3589
Weighted Avg	0.72	0.72	0.72	3589

A comparison is made between the suggested model and the other studies based on the findings and evaluations that were performed using the FER2013 dataset as shown in Table 9.

Table 9: Comparing proposed technique to existing models

Study	CNN Model	Achieved Accuracy (%)
[6]	CNN without LBP….without attention	67…73
[19]	Hypered Deep NNW	74,39 … 73,28
[20]	AlexNet, ResNet50, and Inception V3	73.56
[21]	Deep NNW	93.26
[28]	CNN	65
Our System	EfficientNetB0 model when freezing model	70.60%
	EfficientNetB0 model when unfreezing model	72.69%

8.Conclusions

A model called EfficientNetB0 was suggested to recognize facial expressions from images that include human faces. Training a single pre-trained model with additional classifiers involves two directions, freezing and unfreezing the feature extraction part in the first stage. The second stage involves the classification of the expression. A dropout rate of 0.3 was implemented to achieve optimal accuracy. The FER2013 dataset was used to assess the model's performance, resulting in accuracy of 70.60% and 72.69%.

Most FER models are subjected to difficulties during their work, especially when some of the face expressions are similar and a mix happen between them for the same person or with another person. Also, all the datasets were taken from different and various environments, which may have various lighting, obstructions for some facial features, etc., which made it difficult to classify them. The proposed model can overcome most of these problems found in the datasets.

Below are the list of conclusions from this work:

- The model gets a good result and high performance if trained on a database similar to which training is required. When the database is different from the one that the model was trained on, the model's performance will not be very effective.
- The proposed model recognized most of the images in the FER2013 dataset but some of the images the model recognized incorrectly .
-The model got the best results with a minimum number of training epochs.

In the future, the model will be expanded to classify the expression in real-time video data and images by increasing the number of epochs.

References

[1] S.Saeed, A.A.Shah, M.K. Ehsan , M.R. Amirzada, A. Mahmood and T.Mezgebo, “Automated Facial Expression Recognition Framework Using Deep Learning,” *Journal of Healthcare Engineering*, vol. 2022, no. 1, pp. 1-11, 2022. Available: https://doi.org/10.1155/2022/5707930.

[2] M.A.V. Rajasimman, R. K. Manoharan, N.Subramani,M. Aridoss and M.G.Galety, “Robust Facial Expression Recognition Using An Evolutionary Algorithm with a Deep Learning Model,” *Applied Sciences*, vol.13, no.1, pp.468- 488, 2023. Available: https://doi.org/10.3390/app13010468

[3] A. T. KABAKUS, "PyFER: A Facial Expression Recognizer Based on Convolutional Neural Networks,"*IEEE Access*, vol.8, pp.142243-142249,2020.Available: https://doi.org/10.1109/ACCESS.2020.3012703

[4] D. K. Jain, P. Shamsolmoali and P. Sehdev, "Extended deep neural network for facial emotion recognition,"*Pattern Recognition Letters*, vol. 120, pp. 69-74, 2019. Available: https://doi.org/10+.1016/j.patrec.2019.01.008

[5] A. R. Khan, "Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges," *Information*,

vol.13, no.6, pp.268-275, 2022. Available: https://doi.org/10.3390/info13060268

[6] J. Li, K. Jin , D. Zhou, N. Kubota and Z. Ju, "Attention mechanism-based cnn for facial expression recognition," *Neurocomputing*, vol.411, pp.340-350, 2020. Available: https://doi.org/10.1016/j.neucom.2020.06.014

[7] I. N. Alam, I. H. Kartowisastro and P. Wicaksono, " Transfer Learning Technique with EfficientNet for Facial Expression Recognition System,"*Revue d'Intelligence Artificielle,* vol.36,no.4, pp.543-552,2022. Available: https://doi.org/10.18280/ria.360405

[8] M. Khalaf and B. N. Dhannoon, "MSRD-Unet: Multiscale Residual Dilated U-Net for Medical Image Segmentation," *Baghdad Science Journal*, vol. 19, no. 6, pp.1603-1611, 2022. Available: https://doi.org/ 10.21123/bsj.2022.7559

[9] H. Nayyef and M. S.H. Al-Tamimi, "Deep Learning Techniques For Skull Stripping Of Brain Mr Images," *International Journal on Technical and Physical Problems of Engineering (IJTPE)*, vol. 15, no. 4, pp. 321-327, 2023. https://www.iotpe.com/

[10] W. F. Kamil and I. J. Mohammed, " Adapted CNN-SMOTE-BGMM Deep Learning Framework for Network Intrusion Detection using Unbalanced Dataset," *Iraqi Journal of Science*, vol. 64, no. 9, pp. 4846-4864, 2023. Available: https://doi.org/ 10.24996/ijs.2023.64.9.43.

[11] S.Porcu, A. Floris and L. Atzori, " Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems," *Electronics,* vol.9, no. 11, pp.1892-1904, 2020. Available: https://doi.org/10.3390/electronics9111892

[12] K. Li, Y. Jin,M. W. Akram,R. Han and J. Chen, "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy," *The Visual Computer* , vol. 36, pp. 391-404, 2019. Available: https://doi.org/ 10.1007/s00371-019-01627-4

[13] H.B.Ul haq,W. Akram, M.N. Irshad, A. Kosar, M. Abid, " Enhanced Real-Time Facial Expression Recognition Using Deep Learning," Acadlore Transactions on AI and Machine Learning,vol. 3, no.1, pp. 24-35, 2024. https://doi.org/10.56578/ataiml030103

[14] G. P. Kusuma, Jonathan and A.P. Lim, "Emotion Recognition on FER-2013 Face Images Using Fine-Tuned VGG-16," Advances in Science, *Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 315-322, 2020. Available: https://doi.org/ 10.25046/aj050638

[15] C.M. A. Llyas, R. Nunes, K. Nasrollahi, M.Rehm and T.B. Moeslund, "Deep Emotion Recognition through Upper Body Movements and Facial Expression," in *Proceedings of the 16th International Conference on Computer Vision Theory and Applications,* vol.5, pp.669-679, 2021. Available: https://doi.org/ 10.5220/0010359506690679

[16] G. Verma and H.Verma, "Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions," *The Review of Socionetwork Strategies*, vol. 14, pp. 171-180, 2020. Available: https://doi.org/ 10.1007/s12626-020-00061-6

[17] L. Zhao, "A facial expression recognition method using two-stream convolutional networks in natural scenes," *Journal of Information Processing Systems*, vol.17, no.2, pp.399-410, 2021. Available: https://doi.org/ 10.3745/JIPS.01.0070

[18] T. Quoc and V. Le,"EfficientNetV2: Smaller Models and Faster Training Mingxing," International Conference on Machine Learning, pp.1-11, 2021.https://doi.org/10.48550/arXiv.2104.00298

[19] W. Rashid,N. K. Elabbadi and A. M. Gaber, "Hybrid Deep Neural Network for Facial Expressions," *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, vol. 9, no. 4, pp. 993-1007, 2021. Available: https://doi.org/10.52549

[20] R.K.Reghunathan, V.K. Ramankutty, A. Kallingal, V. Vinod, " Facial Expression Recognition Using Pre-trained Architectures," *Eng. Proc*. Vol. 62, no. 22, pp.1-6,2024. Available: https://doi.org/10.3390/engproc2024062022 .

[21] O. Askria,G. Manitac, M. A. Hajjaji, "Efficient Facial Emotion Recognition Using An Optimized

Deep Learning Model Based On Quantum Gazelle Optimization Algorithm," *28th International Conference on Knowledge Based and Intelligent information and Engineering Systems, Procedia Computer Science*. Vol. 246, pp.2772–2781, 2024. Available: https://doi.org/10.1016/j.procs.2024.09.394

[22] D. T. Long, T. T. Tung and T. T. Dung, "A Facial Expression Recognition Model using Lightweight Dense-Connectivity Neural Networks for Monitoring Online Learning Activities," *International Journal of Modern Education and Computer Science*, vol. 14, no. 6, p.53-64,2022. Available: https://doi.org/ 10.5815/ijmecs.2022.06.05

[23] J. A. Alhijaj and R.S. Khudeyer, "Integration of EfficientNetB0 and Machine Learning for FingerprintClassification," *Informatica*, vol. 47, pp. 49-56, 2023. Available: https://doi.org/10.31449/inf.v47i5.4527

[24] M. F. Alsharekh, "Facial Emotion Recognition in Verbal Communication Based on Deep Learning," *Sensors* , vol. 22, no. 16, p. 6105-6113, 2022. Available: https://doi.org/10.3390/s22166105

[25] M. Aly, A. S. Ghallab, and I. S. Fathi, "Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model," *IEEE Access*, vol.11, no.1, pp.121419-121433, 2023. Available: https://doi.org/ 10.1109/ACCESS.2023.3325407

[26] M. Khalaf and B. N. Dhannoon, "Skin Lesion Segmentation based on U-Shaped Network," *Karbala International Journal of Modern Science*, vol. 8, no. 3, pp. 493-502, 2022. Available: https://doi.org/ 10.33640/2405-609X.3248

[27] M.Sari, A. Moussaoui and A. Hadid, "Automated Facial Expression Recognition Using Deep Learning Techniques: An Overview," *International Journal of Informatics and Applied Mathematics*, vol. 3, no. 1, pp. 39-53, 2020. https://dergipark.org.tr/en/download/article-file/1125402

[28] A. Agrawal and N. Mittal, " Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy," *The Visual Computer*, vol.36, no.2, pp.405-412,2020. Available: https://doi.org/ 10.1007/s00371-019-01630-9.

[29] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images," *Pattern Recognition Letters*, vol. 133, pp. 232-239, 2020. Available: https://doi.org/ 10.1016/j.patrec.2020.03.011

[30] A.Babisha, A.Swaminathan, D.Anuradha,C.Gnanaprakasam, T.Kalaichelvi, "Advancements in Facial Expression Recognition: State-of-the-Art Techniques and Innovations," IJISAE, 2024, vol.12, no.19s, pp.538–546, 2024. https://ijisae.org/index.php/IJISAE/article/view/5097

[31] E. Churaev and A. V. Savchenko ,"A standalone software for real-time facial analysis in online conferences and e-lessons," Software Impacts, vol.16, pp.100507-100514, 2023. Available: https://doi.org/ 10.1016/j.simpa.2023.100507.

[32] S. Pundkar, P. Gulhane, H. Haewani, D. Satpute and R. Mokhale, "The State Of The Art In Facial Emotion Recognition: A Review," Int. Research Journal of Modernization in Engineering Technology and Science, vol. 5, no. 4, pp. 5927-5933, 2023. Available: https://doi.org/10.56726/IRJMETS37433

[33] W.J. Hadi, A.S. Ajrash, S.M. Salman, M. j. jasim and M.T. Ibrahim," Densenet Model for Binary Glaucoma Classification Performance Assessment with Texture Feature," bsj,vol.21,no.12,pp.3903-3913,Dec.2024. Available: https://doi.org/ 10.21123/bsj.2024.9857

[34] [Online]. Available: https://www.kaggle.com/datasets/msamb are/fer2013.