QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

An Efficient Deep Learning based Real Time Facial Expression Recognition System

Sharmeen M.Saleem
Department of Information Technology, College of Informatics Akre, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
Email: sharmeenbarwary2020@gmail.com

Subhi R. M. Zeebaree
Department of Energy Engineering, Technical College of Engineering, Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.
Email: subhi.rafeeq@dpu.edu.krd

Maiwan B. Abdulrazzaq
Department of Computer Science, Faculty of Science, University of Zakho, Duhok, Kurdistan Region, Iraq.
Email: maiwan.abdulrazzaq@uoz.edu.krd

ARTICLE INFO

Article History:
Received: 10/10/2023
Accepted:23/11/2023
Published: Winter2024

Keywords:
Deep Learning, CNN, Real Time, Facial Expression Recognition, RAF Dataset.

Doi:
10.25212/lfu.qzj.9.4.55

ABSTRACT

Human-computer interaction, emotion analysis, and more depend on computer vision's face expression recognition. Real-time facial expression identification enhances decision-making in human-computer interaction, healthcare, marketing, and other fields. This study introduces a fast and accurate real-time facial expression identification system utilizing Convolutional Neural Network (CNN) and Mobile Network Version 2 (MobileNetV2). The proposed methodology involves the design and implementation of a CNN architecture tailored for facial expression recognition. We strategically arrange layers, fine-tune parameters, and leverage transfer learning techniques, particularly focusing on the optimization of model depth and complexity. The model is trained and evaluated using benchmark datasets to ensure robust performance across various facial expressions. Data-driven infrastructure helps the system manage real-world changes. Our approach identifies facial emotions in real time from live footage. This paper explores facial expression recognition employing a simplified CNN

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

architecture, contrasting with more complex pre-trained networks like MobileNetV2. Despite its simplicity, the proposed CNN yields result close to the performance of MobileNetV2. The study emphasizes the viability of less intricate CNNs for facial expression recognition, offering a balance between model simplicity and competitive accuracy. High accuracy and computational efficiency make the system suitable for real-time applications and resource-constrained systems The experimental tests extracted that a very acceptable accuracy achieved when using CNN model with the AffectNet dataset with a ratio of (97.2%).

1. Introduction

Human communication is a complex interplay of verbal and non-verbal cues, with facial expressions being one of the most powerful and universally recognized means of conveying emotions and intentions[1],[2]. The ability to accurately interpret and respond to facial expressions is fundamental to our social interactions, influencing our relationships, decisions, and overall well-being[3],[4]. In the realm of artificial intelligence and computer vision, the development of real-time facial expression recognition systems represents a significant milestone with transformative applications in various domains, including human-computer interaction, mental health assessment, and human-robot collaboration[5],[6],[7]. Recognizing facial expressions in real-time is a challenging task due to the dynamic nature of human emotions and the subtleties of facial muscle movements[8],[9]. Traditional computer vision methods have made considerable progress in this domain, but their limitations in handling variations in lighting, pose, and facial expressions have fueled the exploration of deep learning techniques, which have exhibited remarkable capabilities in various image analysis tasks[10],[11].

In this research, we undertook the task of optimizing a CNN model for facial emotion extraction. By strategically rearranging layers and refining the training process, we successfully developed a CNN model that yields results comparable to pre-trained networks like MobileNetV2. Notably, our model is less complex and demands less training time compared to its more intricate counterparts. This achievement stems

from a deliberate effort to strike a balance between model efficiency and accuracy, allowing for a more streamlined and resource-efficient approach to facial emotion recognition.

This paper introduces an efficient deep learning-based real-time facial expression recognition system, designed to address the complexities of recognizing emotions from live video streams or real-time image feeds. Leveraging the power of deep neural networks, this system offers improved accuracy, speed, and adaptability, making it a valuable tool for applications ranging from sentiment analysis in market research to enhancing human-computer interaction in virtual environments. The importance of real-time facial expression recognition cannot be overstated. In clinical psychology, it aids in the diagnosis and treatment of mood disorders and mental health conditions by providing objective assessments of patients' emotional states[12]. In human-computer interaction, it enhances user experience by enabling computers to respond empathetically to users' emotions. Moreover, in fields like marketing and customer service, it allows businesses to gauge customer sentiment and tailor their responses accordingly, leading to more effective communication and customer satisfaction[13],[14],[15]. In the following sections of this paper, we will delve into the key components and methodologies of our efficient deep learning-based facial expression recognition system. We will explore the underlying neural network architecture, the dataset used for training and validation, and the real-time implementation considerations that contribute to its efficiency. Furthermore, we will discuss the system's performance evaluation, comparing it with state-of-the-art methods and highlighting its advantages in terms of accuracy and computational speed. By the end of this paper, readers will have gained a comprehensive understanding of our proposed real-time facial expression recognition system and its potential to transform various applications in both academia and industry. This research contributes to the ongoing efforts to bridge the gap between human emotions and artificial intelligence, ultimately facilitating more natural and empathetic interactions between humans and machines.

Zhou, Ning et al. [16]developed a lightweight CNN for real-time and bulk facial emotion recognition in 2020 to enhance classification. They developed a real-time vision system to test their idea. Their first facial emotions classification model uses

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

multi-task cascaded convolutional networks for face recognition and coordinate transfer. Classifying emotions follows. Multi-task cascaded convolutional networks may save memory using cascade detection alone. Their expression classification model uses Global Average Pooling instead of the deep convolution neural network's fully connected layer. The completely linked layer's black box is reduced by categorizing each feature map channel. Combining residual modules with depth-wise separable convolutions reduces parameters and makes the model portable. Finally, they test their model using FER-2013. Facial expression classification uses 3.1% of 16GB RAM, or 0.496GB. On the FER-2013 dataset, our 872.9-kilobyte model is 67% accurate. It identifies out-of-dataset figures effectively.

In 2021, Minaee, Shervin et al. [17]proposed a deep learning approach based on attentional convolutional network that can focus on important parts of the face and improves over previous models on multiple datasets. They showed that even a network with a few layers (less than 10 layers) can achieve a very high accuracy rate, including FER-2013, CK+, FERG, and JAFFE. They also employed a visualization approach to locate crucial face areas to recognize emotions, or picture sections that have the most influence on the classifier. Experiments demonstrated that various emotions are responsive to particular facial areas.

In 2023, Lee and Yoo [18]introduced a divide-and-conquer-based learning method to improve facial expression recognition (FER) by lowering intra-class distance and boosting inter-class distance MobileNet identified face regions in photos, while ResNet-18 was the facial emotion backbone deep neural network. After reviewing the confusion matrix, which comprises the trained ResNet-18 model's inference results, similar facial expression groups were aggregated and utilized to retrain the deep learning model. Thermal (Tufts and RWTH) and RGB (RAF and FER2013) datasets tested the proposed method. FER performance rose to 97.75% for Tufts, 86.11% for RWTH, 90.81% for RAF, and 77.83% for FER2013. Thus, the proposed method can classify massive facial expression data.

2. Background Theory

Facial expression recognition is a multidisciplinary field that draws from computer vision, deep learning, psychology, and neuroscience. To understand the theoretical

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

underpinnings of our efficient real-time facial expression recognition system, it is essential to explore the key concepts and theories that inform this research.

2.1 Deep learning

Deep learning is a prominent sub-discipline that can be found under the umbrella of the subject of machine learning. Over the course of the last few years, deep learning has shown tremendous growth and progress. Deep learning was established with the primary objective of bridging the gap between conventional machine learning research and the ultimate objective of building intelligent computers. This was the motivation for the development of deep learning. This should always be the primary focus of deep learning. Deep learning is a subfield of machine learning that involves representing data in the form of a conceptual tree in order to better understand it. The objective of this strategy as a whole is to develop an artificial intelligence model, and each layer of this structure is made up of ideas that are drawn from layers that came before it that were simpler. Any deep learning algorithm would be lacking a crucial component of its architectural design if it did not include the class structure as one of its constituent parts [19]. Figure (1) illustrates the relationship between intelligence Artificial, machine learning, and deep learning [20].

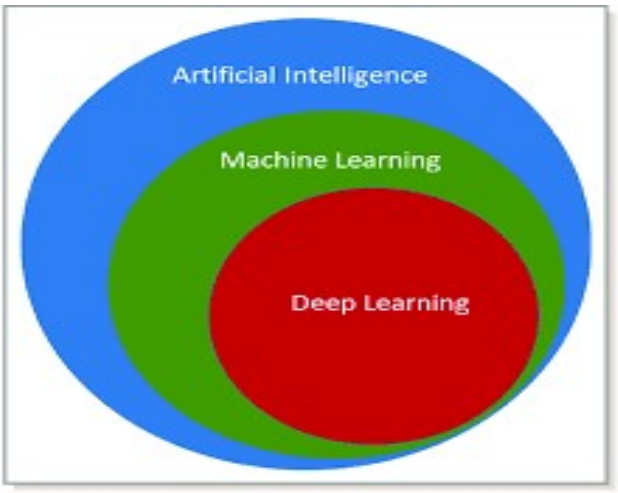

Figure (1): Relationship between AI, Machine Learning and Deep Learning[20].

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

The term "deep learning" refers to a kind of machine learning that is distinguished by a variety of features, all of which may be encapsulated in the following list of bullet points:

- As a general rule, the efficiency of deep learning algorithms tends to increase over time along with the amount of data that is continually being made available. The ideas behind distributed representation learning serve as the foundation for these methodologies.
- Deep learning is essentially a rebranding of the notion of artificial neural networks, and here is one way to think about it.
- It is now feasible to create very sophisticated deep learning models with large-scale multi-layered structures because to the availability of software frameworks like as TensorFlow, Theano, Caffe, Mxnet, and Keras, in addition to developments in hardware. This was made possible by the combination of these two factors. The truth that these frameworks are currently used by a significant number of people made this achievement conceivable.

Deep learning provides a number of benefits, some of which include the capacity to carry out supervised learning tasks and the automated extraction of a variety of attributes. These advantages have proven to be essential in providing data scientists and engineers with the capacity to effectively tackle problems that are becoming more complicated.

2.2 Types of Deep Learning Networks

There are many types of deep learning algorithms and the most used ones are: Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and Multilayer Perceptron Network Neural (MPNN).

2.3 Convolution Neural Network (CNN)

A convolutional neural network (CNN) is a layer-by-layer convolution, pooling and fully connection layer, used for pattern recognition, object detection, image classification and fragmentation ... etc. [21]. (2018., al et Arsenov).|CNN consider as

the most widely used for deep learning network especially in computer vision tasks [22].

CNN was first discovered in 1980 by (Yann LeCun). Based on the work done by Kunihiko Fukushima who invented the neural network basic principles of image recognition [23].Giant tech companies like Facebook and Google use CNN with a large number of convolutional layers for various purposes such as face recognition and image search [24].

Various structures of CNN have been developed to solve real-world problems and consider the structure of the model LeNet is the first successful implementation of CNN, developed by LeCun Yann in the 1990s and using it to read postal codes, numbers, etc., Various improvements have been made to CNN Structures (since 1989 to date) these improvements can be classified as an improvement in variables, regularization, restructuring, etc.[6].

2.3.1 CNN Models

In this section, we recall some convolutional neural network architectures which made promotion for the field of deep learning: LeNet-5, AlexNet, ZfNet, VGGNet, GoogLeNet, ResNet, NASNet Mobile, and MobileNetV2.

2.3.2 MobileNetV2

The present model is an architecture for a convolutional neural network, which was built with the objective of achieving the best possible degree of performance on mobile devices. This was the primary motivation for the development of the model. The design is founded on an inverted residual structure; the bottleneck layers are the ones who are accountable for building the residual connections with one another. The intermediate expansion layer makes use of lightweight depth-wise convolutions in order to properly filter features and introduce non-linearity. These goals are accomplished via the employment of this layer. The design of MobileNetV2 consists of an initial fully convolutional layer that has 32 filters, followed by 19 layers that are regarded as residual bottlenecks in the network [25].

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

2.4 Datasets of FER

This section deals with the famous datasets depended by recent researchers.

2.4.1 FER2013

The FER2013 [26] database was presented for the very first time at the ICML 2013 Challenges in Representation Learning session. The FER2013 database is a comprehensive and unrestricted collection of data that was gathered via the usage of the Google image search API. FER2013 stands for "Federation of European Researchers in 2013." After the images had all been registered, they were extended to have dimensions of 48 by 48 pixels, but not before eliminating incorrectly detected frames and making modifications to the area that had been cropped off. There are a total of 28,709 photos for training purposes included in the FER2013 dataset, in addition to 3,589 images for validation and 3,589 test images. These images have been classified into one of seven distinct types of emotional expression, including anger, disgust, fear, happiness, sadness, surprise, and neutrality.

2.4.2 RAF-DB

RAF-DB [27] the Real-world Affective Face collection is a comprehensive database that has 29,672 images of faces that were shot from different places all around the internet. The RAF-DB includes seven basic emotion labels: neutral, happiness, sadness, surprise, anger, disgust, and fear. The images in this data set have varying resolutions, but they are generally of good quality to capture facial expressions effectively with resolutions include 640x480 pixels and 1280x720 pixels. These pictures exhibit a wide range of styles and subjects. The samples are annotated with the assistance of a human population, which makes it possible to provide estimations that may be relied upon. The examples include a total of fourteen different label alternatives, seven of which are fundamental emotion labels and eleven of which are compound emotion labels. During the course of this specific inquiry, a total of 15,339 images were extracted from the fundamental emotion set. These photographs have been split up into two distinct categories in order to facilitate the analyzing process. The number of training samples was 12,271 in the first group, whereas the number of testing samples totaled 3,068 in the second group.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

2.4.3 AffectNet

AffectNet [28] database has over one million unique pictures that have been gathered from several websites located all over the internet with resolutions include 720p (1280x720) or higher, which corresponds to the standard high-definition (HD) resolution. However, the exact pixel dimensions may vary across the dataset, as AffectNet includes a diverse set of images captured in real-world conditions. These photos were discovered by doing searches on a number of different search engines using tags that corresponded to various states of mind. The database in question is famous for being the most thorough in terms of the facial expressions it gives, which contain not one but two different emotion models. This aspect of the database's comprehensiveness makes it stand out. The category model and the dimensional model are their respective names. There are almost 450,000 images in this specific collection, and each one has been meticulously annotated with descriptions for one of eight major emotions.

3. Research Methodology

The proposed system starts with designing a CNN model to be used for three famous datasets named (AffectNet, RAF, and FER2013). These datasets will be reduced by extracting the specific images depending on main five emotion classes (Angry, Discussed, Happy, Natural, and Surprise). These classes have been selected because they provide noticeable changes occurred to the face. After that, the MobileNetV2 will be implemented

3.1 Initiating the Dataset Preparation

In the proposed system, the dataset should include images of individuals displaying all the depended emotions, which are the five main categories (Happy, Sadness, Anger, Surprise, and Natural). In this study, the dataset was utilized from different sources as follows:

a. **Affect Net Dataset**: After filtering out irrelevant images, we selected 8,550 images within the depended five categories. As a sample, Figure (2) refers to the one Affect Dataset category which is Anger one.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Figure (2): Samples of Anger category from Affect Dataset.

b. **RAF Dataset**: After filtering out irrelevant images, we selected 6980 images within the depended five categories. As a sample, Figure (3) refers to the one RAF Dataset category which is Happy one.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

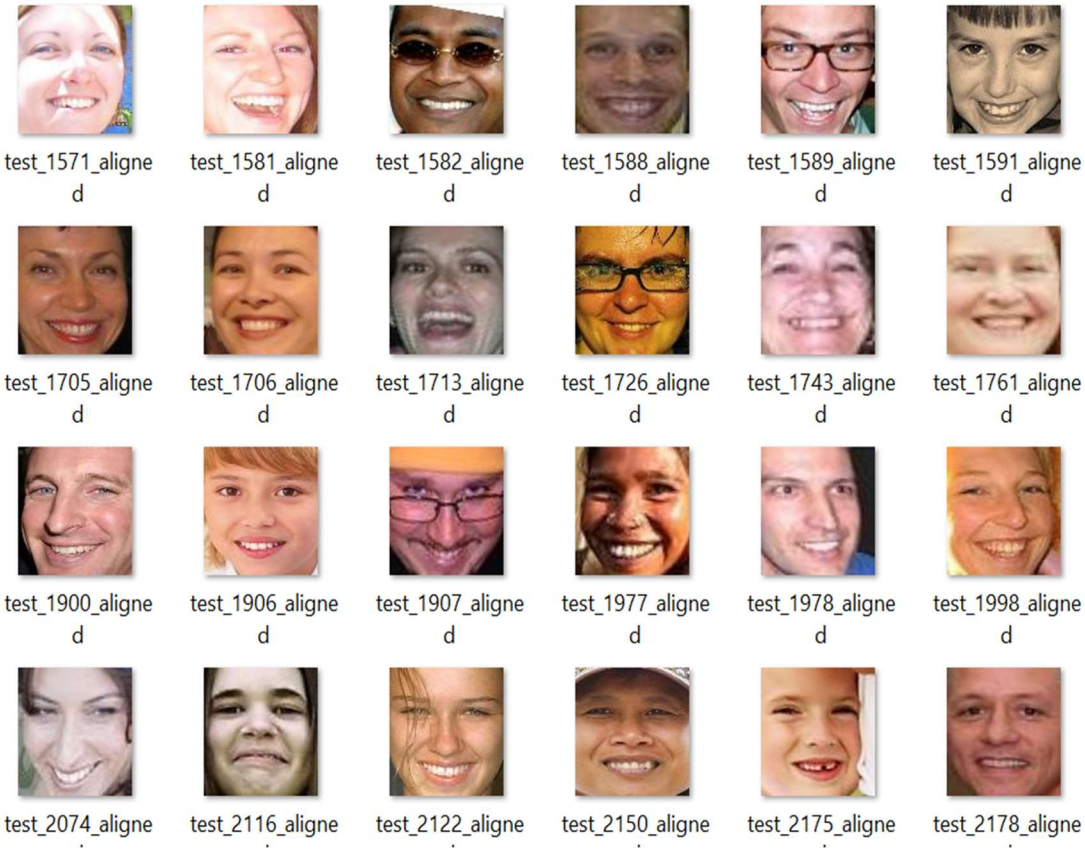

Figure (3): Samples of Happy category from RAF Dataset.

c. **FER2013 Dataset: total number of used images is** 10,935 images within the depended five categories. As a sample, Figure (4) refers to the one FER2013 Dataset category which is Surprise one.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Figure (4): Samples of Surprise category from FER2013 Dataset.

3.2 System Structure

The FER system has been built using two models:

- The first model involves using a CNN architecture.
- The second one involves utilizing the MobileNetV2 model through Transfer Learning.

The general diagram of the depended FER system is shown in Figure (5).

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

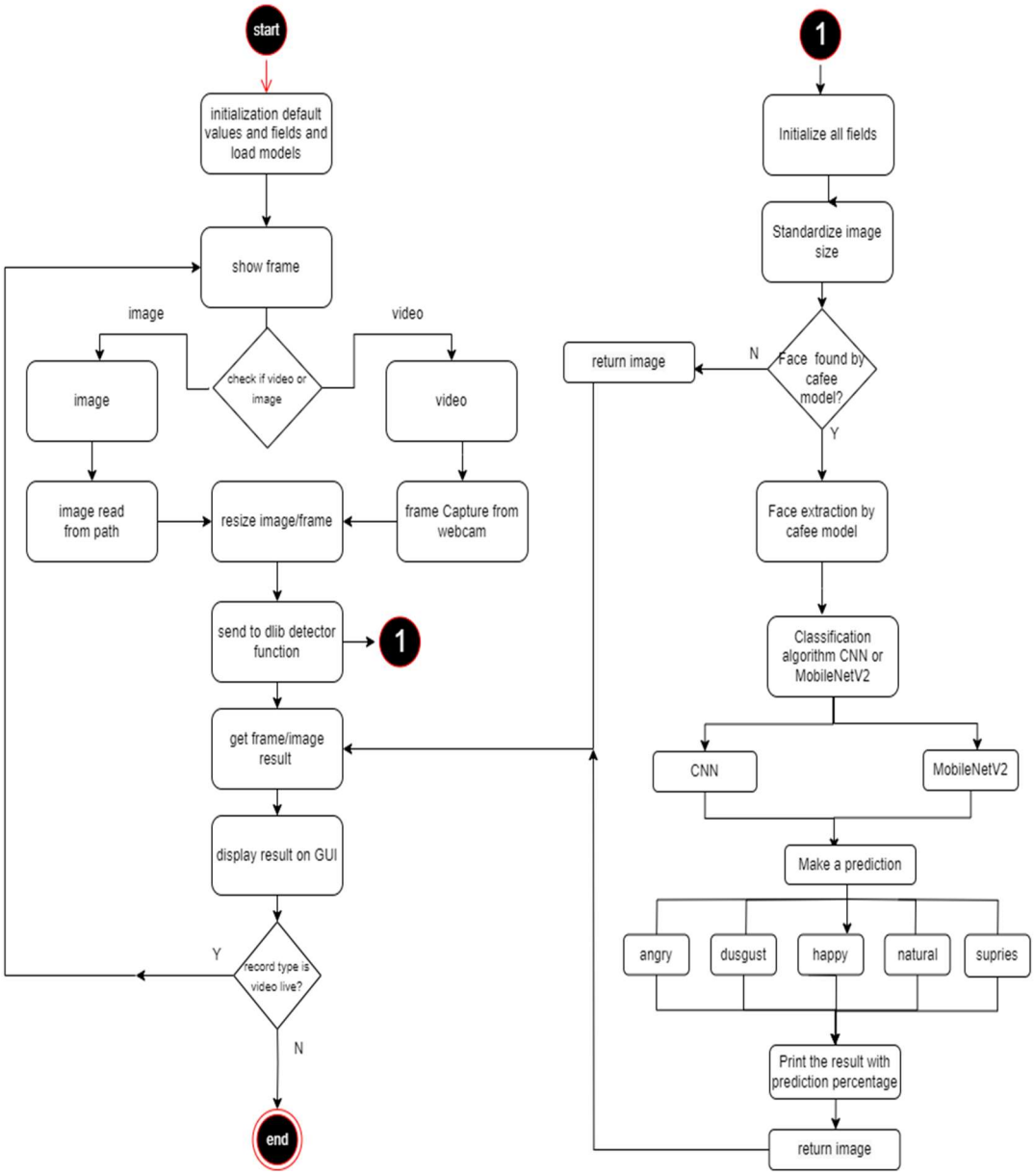

Figure (5): General diagram of the depended FER system.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

3.3 FER Model using CNN

The CNN network will be trained on the three mentioned datasets of individuals with five main depended facial emotions. The FER model was constructed using the proposed CNN architecture via following subsequent steps.

a. Dataset Preprocessing

Preprocessing the dataset is a common initial step in the workflow of deep learning. It involves preparing the raw data in a format that can be accepted by Artificial Neural Networks (ANN), making it easier to use and Meaningful. So that these data align and fit with the inputs of the ANN. This stage deals with image data using the libraries (NumPy, OpenCV, and Keras). There are five steps that were carried out in the preprocessing, and they are:

1) **Read Image:** In this step, all the images will be read and their dimensions changed to 3x180x180 pixels to standardize the color channels (RGB) and image dimensions. An image labeling process performs by assigning the names of emotion to the images such as (angry, happy…and so on), this step used to distinguish each category from the others.
2) **Image normalization:** In this step, the image will be converted into a matrix, and the pixel value range changed from (0-255) to (-1, 1) to reduce computational load during the training using Keras libraries.
3) **Split Data set**: In this step, the data will be divided into two sections: training data (80% of the dataset) and testing data (20% of the dataset). Each section contains all the classes.
4) **Data augmentation:** Image augmentation will be performed using Python libraries such as TensorFlow and Keras. This approach applies various types of transformation techniques to real images, thereby generating modified versions of the same image with alterations like rotation, zooming, flipping, and shifting, as shown in Figure (6). This step aids us in training deep learning models on a greater variety of image forms beyond what is present in the actual dataset.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Figure (6): Data augmentation step.

b. Proposed Structure

For model enabling to effectively learn the features present in the image for facial classification, the use of deep ANNs is essential. Therefore, the CNN will be employed as a key technique in this study. It focuses on data processing, including images. The CNN is utilized to extract the features from the image and reduce its size while retaining its essential properties. This is done by processing the images through multiple layers. Afterward, the derived features will be transformed into a flat, one-dimensional array before being passed to the subsequent layers. These flattened features then integrated with layers of artificial neural cells. The mathematical functions are capable of calculating the weighted sum of various inputs and outputs. The Keras library offers two methods for defining the CNN, which are: Sequential and Functional methods. The "Sequential" class will be chosen to build the suggested ANN since the model required the systematic addition of layers, one after another. This layer-by-layer approach forms the overall architecture of the model. In this study, five convolutional layers will be implemented, each followed by a pooling layer

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

(MaxPooling). Additionally, a final Fully Connected (FC) layer will be applied. This configuration is depicted in Figure (7).

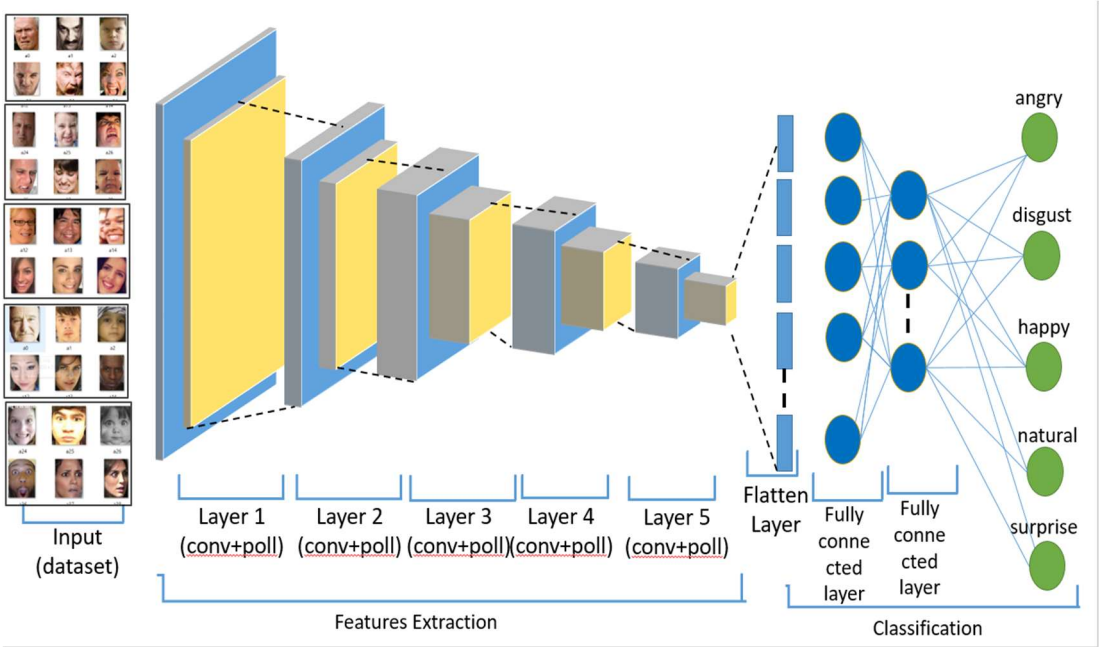

Figure (7): Architecture of the proposed CNN.

Observing Figure (7), we notice that the structure of the Convolutional Neural Network consists of two stages:

- **Feature extraction:** In this stage, five layers are added:
  - The first layer is a 2D convolutional layer (D2conv), that includes the following parameters: Input dimensions: 3x180x180, 8 Filters with a size of 5x5, Activation function of type ReLU, and Padding of type "same".
  - The second convolutional layer is similar to the first layer, with the change of the number of convolutional filters to 16.
  - The third convolutional layer is similar to the first layer, with the change of the number of convolutional filters to 32.
  - The fourth convolutional layer is similar to the first layer, with the change of the number of convolutional filters to 64.

- The Fifth convolutional layer is similar to the first layer, with the change of the number of convolutional filters to 128.

After each convolutional layer, a MaxPooling layer (D2MaxPooling) with a size of 2x2 is added. This assists in reducing the dimension of the Map Features with a larger size to smaller-sized features. As a result, it reduces the amount of information generated by the subsequent convolutional layer for each feature, while preserving the essential features.

- **Classification**

Preceding the fully connected step, a "Flatten" operation is performed to transform all the resulting two-dimensional matrices from the pooled feature maps into a continuous long linear vector. This flattened array is then fed as input to the FC layer for image classification. The FC layer, denoted as "dense," comprises 128 nodes and the output layer with 5 nodes (angry, happy ….so on). A "Sigmoid" activation function is used in the output layer. To prevent overfitting during training, a "Dropout" layer with a value of 0.5 (50%) is incorporated. The Dropout layer randomly ignores a set of neural cells, where a value of 1 means no dropout, and 0 means no output from the layer. A suitable dropout rate typically ranges between 0.5 and 1.

c. Training of FER Model

The training step is highly crucial for the model. This stage involves creating a model from the provided data. The model is trained on the training data to discover the accurate weights, which will be automatically adjusted by the designated algorithm. This aids in minimizing errors.

In this stage, the "compile" and "fit" functions is applied within the Keras library. In the "compile" function, the network optimizer (Adam - Adaptive Moment Estimation) is utilized, along with the loss function (binary cross-entropy) and the accuracy metric. The "fit" function is using to specify the training and validation data, as well as the number of training epochs. Through the experimentation, 20 epochs were chosen, which proved sufficient to optimize the weights with a learning rate of 0.0001. The batch size was set to 32. Tables (1, 2, and 3) summarize the values employed in model construction.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Table (1): Feature Extraction Layers for the proposed FER using CNN Model.

Feature Extraction Layer	Type	Layer 1	Layer 2	Layer 3	Layer 4	Layer5
	Padding	same	same	same	same	same
	Filter Number	8	16	32	64	128
	Filter Size	(5*5)	(5*5)	(5*5)	(5*5)	(5*5)
	Maxpool2D	(2*2)	(2*2)	(2*2)	(2*2)	(2*2)
	Activation	ReLU	ReLU	ReLU	ReLU	ReLU

Table (2): Classification Layers for the proposed FER using CNN Model.

Classification Layer	Concept	Dense 1 Layer	Dense2 layer	Dropout Layer	Dense3 (output)
	Nodes	128	64	-	5
	Activation	ReLU	ReLU	-	Sigmoid
	Value	-	-	0.5	-

Table (3): Functions and Values Used in the Training Stage of the proposed FER using CNN Model.

Function	Parameter	Value
Compile	optimizer	Adam
	Loss	Binary_crossentropy
	Metrics	Accuracy
Fit	Epoch	20
	Learning rate	0.0001
	Batch size	32

3.4 FER using MobileNetV2 Model

The "MobileNetV2" model can be utilized to extract features from images of the depended datasets. Normally, the convolutional layers are constructed, connected, and tuned, which employed to extract features from the images. However, If the output layers (i.e. classification layers) removed, and instead of them other

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

classifier(s) will be employed in the classification phase. This process is referred to as Transfer Learning (TL). The TL is one of the machine learning techniques that operates by reusing a pre-designed model for group of datasets. Then TL involves using this model as a starting point for a new dataset, which is typically smaller. The Softmax classifier is used to perform the classification on the extracted features and predict the results. Figure (8) shows the architecture of MoileNetV2 model.

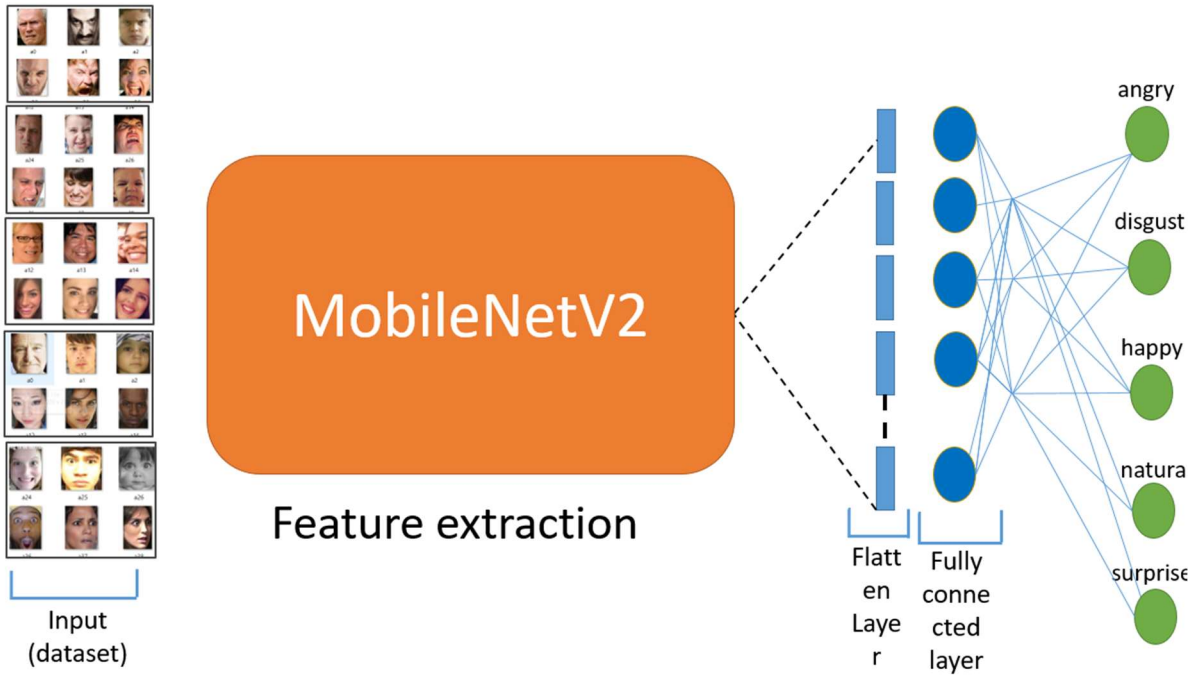

Figure (8): Architecture of the MoileNetV2 model.

The preprocessing steps for the dataset of FER model using MobileNetV2 are the same as used CNN. These steps include: image reading, image normalization, data augmentation, and data splitting. The TL is performed on the MobileNetV2 model by removing the last layers (classification layers) and replaced by Softmax to perform the classification on the extracted features and predict the outcomes.

The performance of the depended models has been evaluated in terms of (Accuracy, Precision, and Recall) metrics by applying the Confusion Matrix function. Equations (1, 2, and 3) were used for the calculations of these metrics.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Accuracy = (TP + TN) / (TP + TN + FP + FN) (1)

Precision = TP / (TP + FP) (2)

Recall = TP / (TP + FN) (3)

The Confusion Matrix is a matrix used to assess the performance of binary classification algorithms on a test dataset. It's constructed in the form of a table containing information and details about actual class labels in the first and second row, and predicted class labels in the first and second column. This aids in providing insights into the accuracy of predictions [29].

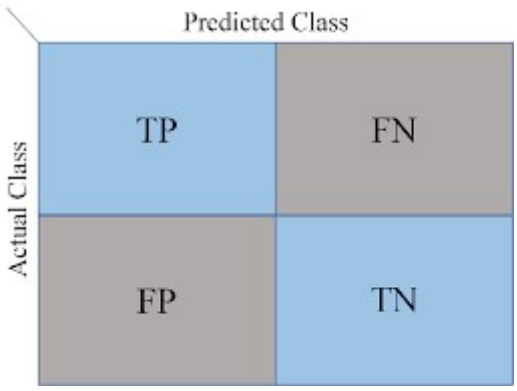

Figure (9) Confusion Matrix

Where [29]:

-(TP) True Positive: It refers to the number of cases where the model correctly predicts a positive class and the actual class is also positive.

-(TN) True Negative: It indicates the number of cases where the model correctly predicts a negative class and the actual class is also negative.

-(FP) False Positive: It represents the number of cases where the model incorrectly predicts a positive class while the actual class is negative.

-(FN) False Negative: It indicates the number of cases where the model incorrectly predicts a negative class while the actual class is positive.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

4. Experimental Results

The depended models in this research have been implemented using the three mentioned datasets (AffectNet, RAF, and FER2013). The significant metrics (Accuracy, Precision, and Recall) are calculated for each model, as illustrated in Tables (4 and 5).

Table (4): Evaluation Performance of FER CNN Model.

Dataset	Precision%	Recall%	Accuracy%
AffectNet	0.998	0.778	97.2%
RAF	0.961	0.988	95.4%
FER2013	0.962	0.798	79.2%

Table (5): Evaluation Performance of FER MobileNetV2 Model.

Dataset	Precision%	Recall%	F-Measure%	Accuracy%
AffectNet	0.95	0.99	0.967	94.2%
RAF	0.931	0.964	0.947	90.8%
FER2013	0.952	0.786	0.861	77.8%

5. Prediction Evaluation of both Model

To evaluate the performance of both models, we gathered two types of samples: static face images and Real time (video). We used common evaluation criteria in our field to analyze the results. For the static face images, we collected 50 pictures of people displaying five different emotions mentioned in our research. Each emotion had 10 images. As for the video, we recorded 50 clips featuring 10 persons. Each person individually faced a camera connected to a computer and expressed the five emotions. After running our custom-designed software, we recorded the model's output. Table (6) represents the (TP, FN, FP and TN) in Confusion Matrix of both CNN and MobileNetV2 Models.

Table (6): TP, FN, FP and TN in Confusion Matrix.

Class	TP	FN	FP	TN
Angry	C1	c2+c3+c4+c5	C6+c11+c16+c21	C(7+8+9+10+12+13+14+15+17+18+ 19+20+22+23+24+25)

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

disgust	C7	C(6+8+9+10)	C(2+12+17+22)	C (1+3+4+5+11+13+14+15+16+18+ 19+20+21+23+24+25)
Happy	C13	C(11+12+14+15)	C(3+8+18+23)	C (1+2+4+5+6+7+9+10+16+17+19+ 20+21+22+24+25)
natural	C19	C(16+17+18+20)	C(1+6+11+21)	C (1+2+3+5+6+7+8+10+11+12+13+ 15+21+22+23+25)
surprise	C25	C(21+22+23+24)	C(5+10+15+20)	C (1+2+3+4+6+7+8+9+11+12+13+ 14+16+17+18+19+21+22+23+24)

5.1 Static Images

Table (7): Static Images Confusion Matrix of CNN Model.

Confusion Matrix							
	Predicted class						
Actual class	**Class**	**Angry**	**Disgust**	**Happy**	**Natural**	**Surprise**	**Total**
	Angry	10	0	0	0	0	**10**
		C1	**C2**	**C3**	**C4**	**C5**	
	disgust	3	7	0	0	0	**10**
		C6	**C7**	**C8**	**C9**	**C10**	
	Happy	0	0	10	0	0	**10**
		C11	**C12**	**C13**	**C14**	**C15**	
	natural	0	0	1	9	0	**10**
		C16	**C17**	**C18**	**C19**	**C20**	
	surprise	0	0	0	0	10	**10**
		C21	**C22**	**C23**	**C24**	**C25**	

Table (8): Static Images Confusion Matrix of MobileNetV2 Model.

Confusion Matrix							
	Predicted class						
Actual class	**Class**	**Angry**	**Disgust**	**Happy**	**Natural**	**Surprise**	**Total**
	Angry	8	2	0	0	0	**10**
		C1	**C2**	**C3**	**C4**	**C5**	
	disgust	2	7	0	1	0	**10**
		C6	**C7**	**C8**	**C9**	**C10**	
	Happy	0	0	9	1	0	**10**
		C11	**C12**	**C13**	**C14**	**C15**	
	natural	0	0	0	10	0	**10**

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

		C16	**C17**	**C18**	**C19**	**C20**	
	surprise	0	0	1	0	9	**10**
		C21	**C22**	**C23**	**C24**	**C25**	

Table (9): Static Images Prediction Evaluation Performance of CNN Model.

Expressions	Precision	Recall	Accuracy
Anger	0.769	1.00	94%
Disgust	1.00	0.70	94%
Happy	1.00	0.769	98%
Normal	1.00	0.75	98%
Surprise	1.00	1.00	100%
Avg. Rate	95.38	84.38	96.8%

Table (10): Static Images Prediction Evaluation Performance of MobileNetV2 Model.

Expressions	Precision	Recall	Accuracy
Anger	0.80	0.80	92%
Disgust	0.77	0.70	90%
Happy	0.90	0.90	96%
Normal	0.83	1.00	96%
Surprise	1.00	1.00	100%
Avg. Rate	0.86	0.88	94.8%

5.2 Real Time (Video)

Table (11): Real Time Confusion Matrix of CNN Model.

Confusion Matrix							
Predicted class							
Actual class	**Class**	**Angry**	**Disgust**	**Happy**	**Natural**	**Surprise**	
	Angry	8	2	0	0	0	**10**
		C1	**C2**	**C3**	**C4**	**C5**	
	disgust	2	7	0	1	0	**10**
		C6	**C7**	**C8**	**C9**	**C10**	
	Happy	0	0	10	0	0	**10**
		C11	**C12**	**C13**	**C14**	**C15**	
	natural	0	0	0	10	0	**10**

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

		C16	**C17**	**C18**	**C19**	**C20**	
	surprise	0	0	0	0	10	**10**
		C21	**C22**	**C23**	**C24**	**C25**	

Table (12):

Real Time Confusion Matrix of MobileNetV2 Model.

Confusion Matrix							
	Predicted class						
Actual class	**Class**	**Angry**	**Disgust**	**Happy**	**Natural**	**Surprise**	
	Angry	7	2	0	1	0	**10**
		C1	**C2**	**C3**	**C4**	**C5**	
	disgust	2	6	0	2	0	**10**
		C6	**C7**	**C8**	**C9**	**C10**	
	Happy	0	0	9	1	0	**10**
		C11	**C12**	**C13**	**C14**	**C15**	
	natural	0	0	1	9	0	**10**
		C16	**C17**	**C18**	**C19**	**C20**	
	surprise	0	0	1	0	9	**10**
		C21	**C22**	**C23**	**C24**	**C25**	

Table (13): Real Time Prediction Evaluation Performance of CNN Model.

Expressions	**Precision**	**Recall**	**Accuracy**
Anger	0.80	0.80	92%
Disgust	0.777	0.70	90%
Happy	1.00	1.00	100%
Normal	0.90	1.00	98%
Surprise	1.00	1.00	100%
Avg. Rate	89.5	0.90	96%

Table (14): Real Time Prediction Evaluation Performance of MobileNetV2 Model.

Expressions	**Precision**	**Recall**	**Accuracy**
Anger	0.80	0.72	90.1%
Disgust	0.60	0.50	84.6%
Happy	0.81	0.90	94%
Normal	0.692	0.90	90%

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

Surprise	1.00	0.90	98%
Avg. Rate	0.78	0.784	91.34%

6. Graphical User Interface of Proposed Models Implementation

The Graphical User Interface (GUI) was designed using the functions available in the Python (Tkinter) library, which is used to build GUI in the Python programming language. The GUI of the proposed system consists of number of main parts which are (select image, Live video, parts predict, complete predict, and close). The prediction results of main three parts of the face are appeared in the GUI, which are (Left Eye, Right Eye, Nose, and Mouth). The Final prediction results represent by the predicted Emotion with its ratio value, which is appeared as a Result in the Left Bottom of the GUI. Figures (11 and 13) represent the GUIs using CNN model for Surprise Emotion prediction using (Static Images and Real Time) respectively. While, Figures (12 and 14) represent the GUIs using MobileNetV2 model for (Disgust and Natural) Emotions prediction using (Static Images and Real Time) respectively.

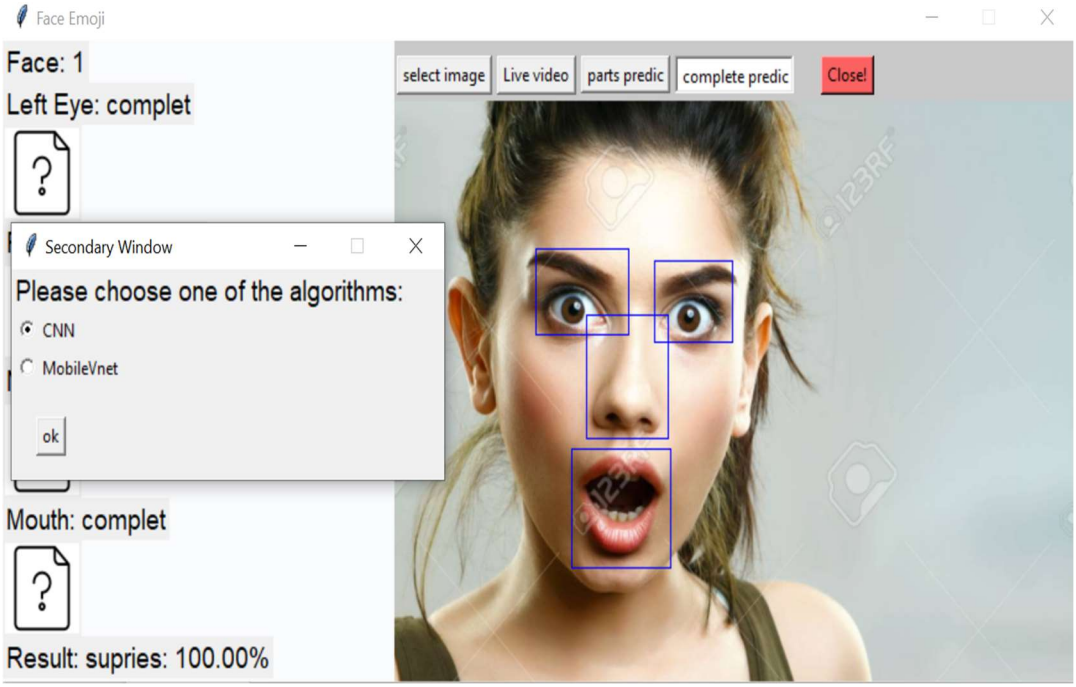

Figure (11): Using CNN model.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

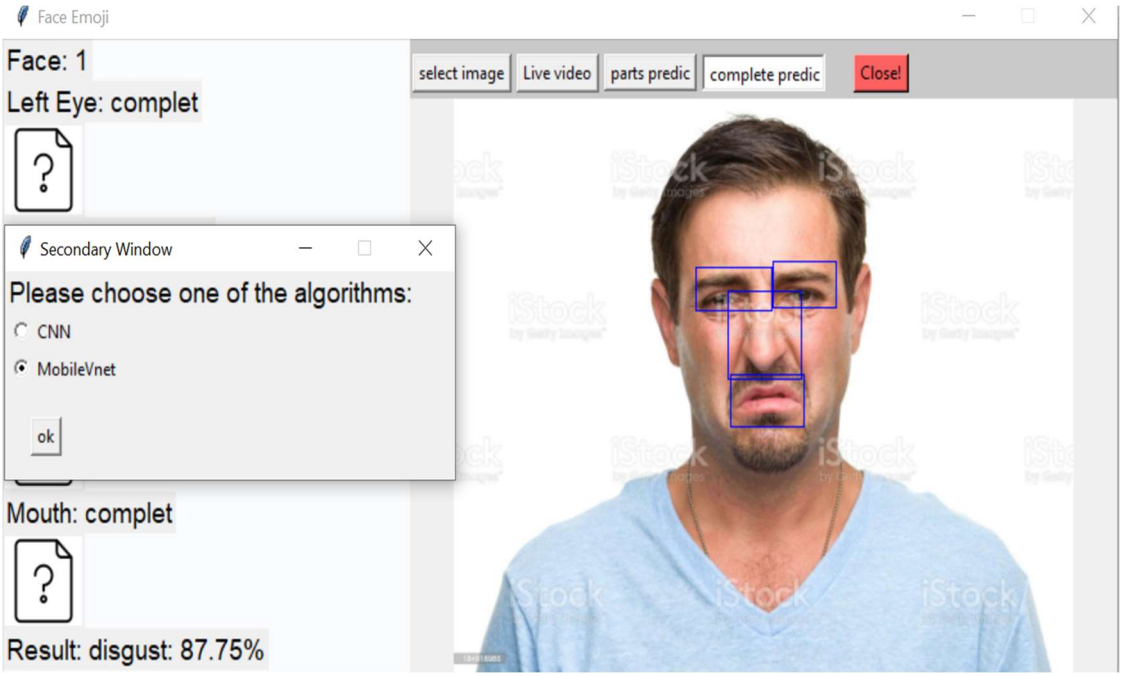

Figure (12): Using MobileNetV2 model.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

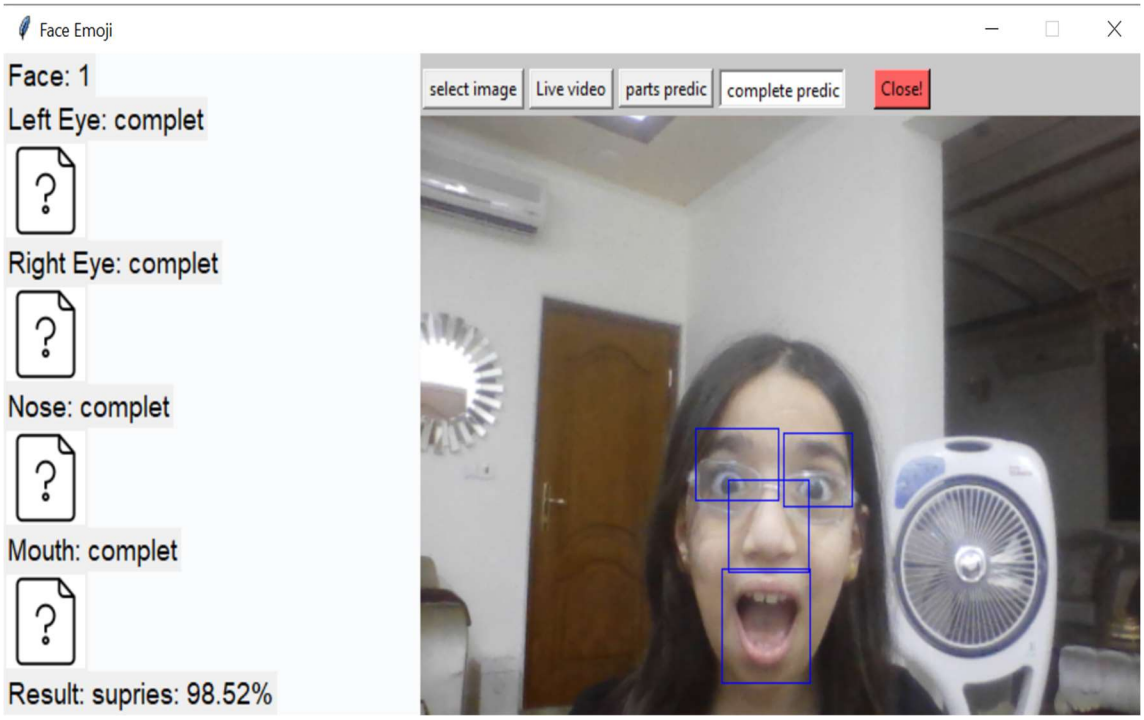

Figure (13): Using CNN model.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

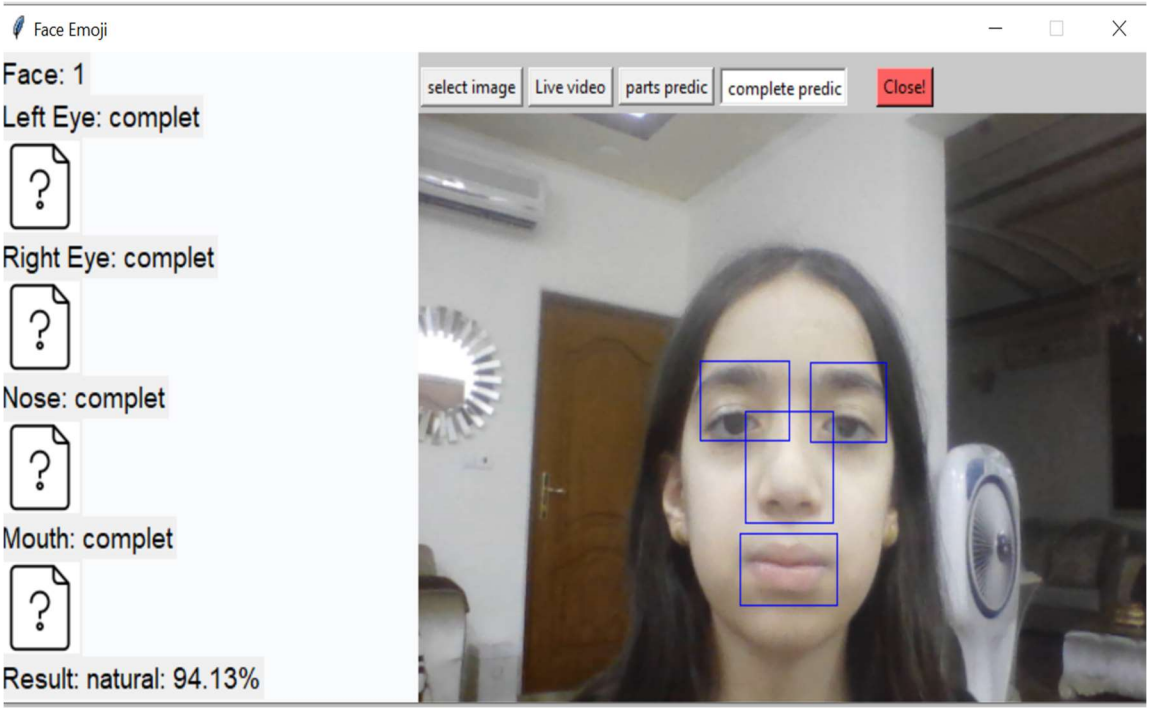

Figure (14): Using MobileNetV2 model.

7. Discussion and Comparison

The proposed system consists of two main models (CNN and MobileNetV2). Each one has been implemented for three famous datasets (AffectNet, RAF, and FER2013). Best accuracy has been obtained with using AffectNet dataset. For the CNN model it was (97.2%), while for the MobileNetV2 got (94.2%).

The prediction stage divided into two main parts (Static Images and Real Time Videos). For Static Images: the CNN model produced best accuracy for (Surprised, Happy and Normal) Emotions, which are (100%,98% and 98%) respectively. While, for the MobileNetV2 the best accuracy obtained with Surprised Emotion (100%). However, for Real Time Videos: the best accuracy was obtained for (Happy and Surprised) Emotions (100% and 100%) when using CNN model. While, both (Happy and Surprised) Emotions got best accuracy (98% and 94%) respectively when using MobileNetV2 Model.

QALAAI ZANISTSCIENTIFIC JOURNAL
A Scientific Quarterly Refereed Journal Issued by Lebanese French University – Erbil, Kurdistan, Iraq
Vol. (9), No (4), Winter 2024
ISSN 2518-6566 (Online) - ISSN 2518-6558 (Print)

When comparing our proposed system and the obtained results with those of previous related works, the following points can be observed:

- Zhou, Ning et al. [16],worked on the FER2013 dataset, the obtained accuracy was (67%). While our system got accuracy of (79.2%) for the same dataset.
- Minaee, Shervin et al.[17], showed that even by using a network with a few layers (less than 10 layers), it is able to achieve an acceptable accuracy rate, including FER-2013. Our system reached a very good accuracy ratio with only 5 layers.
- Lee and Yoo[18], used MobileNetV2 model with RAF and FER2013 datasets. They achieved accuracy ratios of (90% and 77%) respectively. While, our system produced accuracy ratios of (95.4% and 79.2%) respectively for the same model with the same datasets.

8. Conclusion

Depending on the architectures and obtained results of both models (CNN and MobileNetV2) when using the three datasets (AffectNet, RAF, FER2013), it can be concluded that this research produced an efficient FER system based on famous five classes (Angry, Happy, Discussed, Surprised, and Normal). As evaluation of the obtained results, it can be observed that the obtained results of the proposed CNN model are close to those of the modern model which is MobileNetV2. The reason of this achievement is due to the good selection of the necessary changes done on the CNN architecture suitable number of (layers, filters), filters-sizes, and suitable parameters, to be used efficiently with the extracted datasets. The reduction of datasets sizes has been done via removing irrelevant images from the original datasets. Another factor affected on the efficiency of our proposed system is the selection of specific five emotion classes that cause accurate face recognition. From the obtained results, it can be concluded that best accuracy has been obtained when using CNN model with the AffectNet extracted dataset which is (97.2%).

References

1. *Kong, Y., et al., Real-time facial expression recognition based on iterative transfer learning and efficient attention network. IET Image Processing, 2022.* ***16****(6): p. 1694-1708.*
2. *Hassouneh, A., A. 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, 2020.* ***20****: p. 100372.*
3. *Ekman, P., Cross-cultural studies of facial expression. Darwin and facial expression: A century of research in review, 1973.* ***169222****(1).*
4. *Ekman, P. and W.V. Friesen, Constants across cultures in the face and emotion. Journal of personality and social psychology, 1971.* ***17****(2): p. 124.*
5. *Alom, M.Z., et al., A state-of-the-art survey on deep learning theory and architectures. electronics, 2019.* ***8****(3): p. 292.*
6. *Khan, A., et al., A survey of the recent architectures of deep convolutional neural networks. Artificial intelligence review, 2020.* ***53****: p. 5455-5516.*
7. *Li, S. and W. Deng, Deep facial expression recognition: A survey. IEEE transactions on affective computing, 2020.* ***13****(3): p. 1195-1215.*
8. *Gill, R. and J. Singh. A deep learning approach for real time facial emotion recognition. in 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART). 2021. IEEE.*
9. *Ozdemir, M.A., et al. Real time emotion recognition from facial expressions using CNN architecture. in 2019 medical technologies congress (tiptekno). 2019. IEEE.*
10. *Pathar, R., et al. Human emotion recognition using convolutional neural network in real time. in 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT). 2019. IEEE.*
11. *Dukić, D. and A. Sovic Krzic, Real-time facial expression recognition using deep learning with application in the active classroom environment. Electronics, 2022.* ***11****(8): p. 1240.*
12. *Talegaonkar, I., et al. Real time facial expression recognition using deep learning. in Proceedings of international conference on communication and information processing (ICCIP). 2019.*
13. *Singh, S.K., et al. Deep learning and machine learning based facial emotion detection using CNN. in 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom). 2022. IEEE.*
14. *Dwijayanti, S., M. Iqbal, and B.Y. Suprapto, Real-time implementation of face recognition and emotion recognition in a humanoid robot using a convolutional neural network.* IEEE Access, 2022. **10**: p. 89876-89886.

15. Saleem, S.M., S.R. Zeebaree, and M.B. Abdulrazzaq. *Real-life dynamic facial expression recognition: a review*. in *Journal of Physics: Conference Series*. 2021. IOP Publishing.
16. Zhou, N., R. Liang, and W. Shi, *A lightweight convolutional neural network for real-time facial expression detection.* IEEE Access, 2020. **9**: p. 5573-5584.
17. Minaee, S., M. Minaei, and A. Abdolrashidi, *Deep-emotion: Facial expression recognition using attentional convolutional network.* Sensors, 2021. **21**(9): p. 3046.
18. Lee, D.-H. and J.-H. Yoo, *CNN Learning Strategy for Recognizing Facial Expressions.* IEEE Access, 2023.
19. Tiwari, T., T. Tiwari, and S. Tiwari, *How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?* International Journal of Advanced Research in Computer Science and Software Engineering, 2018. **8**(2): p. 1.
20. Tiwari, T., T. Tiwari, and S. Tiwari, *How Artificial Intelligence, Machine Learning and Deep.*
21. Arsenov, A., et al. *Evolution of Convolutional Neural Network Architecture in Image Classification Problems*. in *ITS*. 2018.
22. Yamashita, R., et al., *Convolutional neural networks: an overview and application in radiology.* Insights into imaging, 2018. **9**: p. 611-629.
23. Sakib, S., et al., *An overview of convolutional neural network: Its architecture and applications.* 2019.
24. Bambharolia, P. *Overview of Convolutional Neural Networks*. in *Proceedings of the International Conference on Academic Research in Engineering and Management, Monastir, Tunisia*. 2017.
25. Howard, A.G., et al., *Mobilenets: Efficient convolutional neural networks for mobile vision applications.* arXiv preprint arXiv:1704.04861, 2017.
26. Goodfe*llow, I.J., et al. Challenges in representation learning: A report on three machine learning contests. in Neural Information Processing: 20th International Conference, ICONIP 2013, Daegu, Korea, November 3-7, 2013. Proceedings, Part III 20. 2013. Springer.*
27. *Li, S., W. Deng, and J. Du. Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*
28. *Mollahosseini, A., B. Hasani, and M.H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Transactions on Affective Computing, 2017.* ***10****(1): p. 18-31.*
29. *Demir, F., Deep autoencoder-based automated brain tumor detection from MRI data*, in *Artificial Intelligence-Based Brain-Computer Interface*. 2022, Elsevier. p. 317-351.