



Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review

Sharmeen M. Saleem Abdullah^{1*}, Adnan Mohsin Abdulazeez¹

¹Duhok Polytechnic University, Duhok, Kurdistan Region, IRAQ

*Corresponding Author

DOI: <https://doi.org/10.30880/jscdm.2021.02.01.006>

Received 26 January 2021; Accepted 17 March 2021; Available online 15 April 2021

Abstract: Facial emotional processing is one of the most important activities in effective calculations, engagement with people and computers, machine vision, video game testing, and consumer research. Facial expressions are a form of nonverbal communication, as they reveal a person's inner feelings and emotions. Extensive attention to Facial Expression Recognition (FER) has recently been received as facial expressions are considered. As the fastest communication medium of any kind of information. Facial expression recognition gives a better understanding of a person's thoughts or views and analyzes them with the currently trending deep learning methods. Accuracy rate sharply compared to traditional state-of-the-art systems. This article provides a brief overview of the different FER fields of application and publicly accessible databases used in FER and studies the latest and current reviews in FER using Convolution Neural Network (CNN) algorithms. Finally, it is observed that everyone reached good results, especially in terms of accuracy, with different rates, and using different data sets, which impacts the results.

Keywords: Facial expression recognition, deep learning, convolutional neural networks, feature extraction

1. Introduction

Recognition of emotions is a dynamic process that targets the person's emotional state, which means that the feelings corresponding to each individual's actions are different [1]. Human beings, in general, communicate their feelings in different ways. To ensure meaningful communication, accurate interpretation of these emotions is essential [2]. Facial communication is the primary way to express people's feelings [3],[4], [5],[6]. Mehrabian observed that 7% of knowledge moves between people through writing, 38% through voice, and 55% through facial expression [7], [8]. The Ekman & Friesen Facial Action Coding Scheme (FACS), released in 1978, describes seven primary facial expressions expressed by people without language, including terror, coldness, surprise, disgust, good luck, truth, and neutrality. This system is known as the threshold for Facial Expression Recognition (FER) [9].

Various applications involve the comprehension of human emotion by facial expression, including human-computer interaction, robotics, and health care. [10], [11], [12]. However, emotional recognition in our everyday lives is important for social contact, and emotions play an important role in deciding human actions [13]. However, the affective status of the pupil can be instantly interpreted by FER in the area of schooling. It lets teachers recognize their students' academic interests to include appropriate instructional methods to increase teaching effectiveness [14]. It is essential to monitor for some time the process of analyzing people's gestures, such as visitors who enjoy exhibitions in museums and have a method of thinking and exploring what they see. "neutral, surprising, fear" and other adjustments will be made. This move offers essential semantic details and time structure to determine categories of speech [15]. FER also has various social life uses, such as smart protection, lie detection, and intelligent medical practice [8], [16].

Many challenges still face the FER that the retrieved research has addressed, some of which have been addressed, and others are still being studied by researchers, FER in 3D face form models, emotion recognition in occlusion videos, and so on. Real-time FER is also a difficult challenge [17]. In fact, identifying facial expressions correctly under

uncontrollable situations remains a difficult task because of varying facial positions/view angles, sophisticated light levels, face wrinkles, partial occlusions, and so on [18], [19].

This paper aims to analyze recent advancements in detecting emotions with the aid of numerous methods, algorithms, and architectures to detect facial expressions using CNN and Deep (CNN). Observations are discussed in our review of the topics and contributions.

This paper is organized into four sections. A brief overview of the implementation of (FER) is given in the first section, followed by a brief theory of CNN architecture. FER Systems is presented in the second section. A detail of 22 previous studies about expression recognition using CNN techniques with a table of comparison is shown in the third section. A general conclusion is provided in section 4.

2. Facial Expression Recognition (FER)

This specific area involves systems to classify the fundamental human emotions with current artificial intelligence algorithms, particularly neural networks FACS [20]. The FER general architecture comprises three phases: pre-processing, extraction of features, and classification [14], [21], [22].

- **Pre-processing:** Data preparation is commonly used in deep learning experiments focused on image and signal processing to improve performance. The pieces of this section include facial orientation, gray image transfer, 2-D noise-removal adaptive filtering, image sharpening using sharp masking, and data increase [23]. In pre-processing, input data quality (image) is improved, and redundancy is reduced or eliminated [21]. Next, an image input RGB of $m \times n$ size is read and converted into a gray image with the standard equation [24]. The circumference of the face was detected with Haar [25], [26], [27] Cascade pictures library. Those rectangular facial expressions were then cut off and reported to the same scale. The pictures' pixel values have also been transformed into 64×64 gray images to be put in neural networks. Also, this is required to prevent the excess density of the neural networks [28]. In a real-life situation, data collection for images can be captured in various conditions, such as different directions, locations, sizes, and visibility. Thus in such raw images, the conventional pre-processing technique such as standardization, cutting, and centralization enhances images' recognition during any experimental period [29].
- **Feature extraction:** The primary step is to extract facial features from the image or video input [28], [30]. Current research's primary focus is on extracting compelling features to differentiate between the six fundamental expressions (anger, disgust, fear, sadness, happiness, and surprise) [31]. The feature extraction method extracts new features from the original dataset. It is very beneficial to decrease the number of resources required for processing without missing relevant feature datasets [32]. During the feature extraction step, pre-processing data is transformed to obtain the best representative features [33]. The extracted data should provide valuable knowledge that could contribute to a proper estimation for the classification technique. The product of the extraction process affected the system's output [21]. The traditional algorithms for facial extraction can be divided into two categories: 1) geometric approaches such as Active Appearance Models (AAM); and 2) appearance-based methods like Gabor wavelet representation and Local Binary Pattern (LBP) [34]; in a geometric approach, various geometrical parameters such as position, angle, points of reference, etc. are considered. The entire input image is considered in an appearance-based process, and features are extracted from the picture that best represents the input image [22], [35], [36]. As a deep extractor for extracting image functionality. These features are then used to train conventional classifications, including Support Vector Machine (SVM), Discriminant Analysis classification (LDA), and K-Nearest Neighbors (K-NN) [25], [37]. In the extraction process, several complicated properties are formed to clarify the expression of facial form or texture modifications. For example, the SDM points to the facial organ shape changes and the features of texture like Local Directional Ternary Pattern (LDTP), Histogram of Oriented Gradients (HOG), Spatio-Temporal Texture Map (STTM), and LBP, to define changes in facial organ texture. Several works use the characteristics to clarify the differences in shape and texture [14]. The extraction characteristics were then used to classify the face expression [38].
- **Classification:** This is the final stage of the FER method, and the actual mapping for labeled emotion of action units happens here [21]. During the classification point, different methods are applied to train the classifier using the extracted functions. [14]. Deep learning (DL) algorithms have recently been seen to be an effective technique in the field of image classification [29], [39]. A series of architectures of deep neural networks such as CNN's, Recurrent Neural Networks (RNNs), and Deep Neural Networks (DNNs) have been studied to help identify face emotions [30]. This system uses an end-to-end model to train a deep network structure with millions of parameters that automatically learn accurate features from massive data without the hand-made features [14], [40]. The in-depth features or activations may input conventional classification schemes [34], [37], [41]. Any special classification algorithms accomplish classification [21], SVM, K-NN, and LDA are most commonly used to increase accuracy in identification [22], [36], [42]. After describing the features, the features are applied to a classifier to recognize various facial expressions. Therefore, the classifier's efficiency primarily

depends on the consistency of the features extracted [34], [43]. Each of the conventional classifiers has its own features. For instance, SVM [25]. To distinguish several data groups, kernels can transform several feature representations into higher dimensional space. Furthermore, SVM performs well in target recognition and face detection applications. On the other hand, the LDA method will find the right way to differentiate various groups [37]. In different extraction methods, combined with various classification algorithms, the best possible combination for identifying emotional strength was presented [34].

3. Convolutional Neural Network (CNN)

This section explained the theory of CNN, which consists of the following:

- CNN structures are designated to be the convolution's core operation [36], [44]. CNN is a powerful model commonly used in computer vision [29]. CNN's is successfully extended to the identification of facial expression and achieved good performance [8], [15], [45], [46]. Deep learning is one of the most effective approaches in recent years for machine learning [47]. CNN's are a form of a deep learning network that involves less pre-processing than other standard image classification algorithms [37]. CNN has been the most prevalent image recognition technique [48]. Convolution is carried out for any input filter, followed by non-linearity [49], [50]. They have many applications in many fields, including computer vision and natural language processing [37]. There are currently several versions of CNN architecture, but their underlying structure is quite similar. CNN's basic architecture uses three layers: convolution layer, pooling layer, and complete connection layer [23], [51].
- The convolution layer attempts to learn the input feature representation [51]; this is a simple building block. The primary purpose of the Convolution layer is to extract features from the input signal. The input picture is given as a group of learning neurons. [52]. Then, supply the following layer with these features; this layer includes a series of neurons with trained weights and distortions. Neuron weights are changed by activation map when any new feature must be added[53]. For one, the convolution layer consists of many characteristics. In the previous layer, each neuron of a function map is related to its neighboring neurons. In the last layer, such a neighboring area is called the local visual region of the neuron. For computing a new feature map, the feature map is first combined with a learning kernel (also referred to as filter and feature detector), and the output is passed to a non-linear activation function [51]. Several filters (kernels) in the convolution layer are convoluted with the entry picture to extract some features like vertical or horizontal lines (e.g., edges)[36], [13], [54]. Generally, the number of kernels is the number of feature graphs obtained by filtering the front layer feature graph. The stronger the kernels, the stronger the network learning function, and the higher the recognition accuracy. The number of kernels should be calculated according to particular specifications[52], [55].
- In the pooling layer, the computer screening scheme for discriminating primary visual features and assembling them into abstract visual features of variation can be realized. The number of output characteristic graphs and the characteristic graphs' dimensions become smaller after sampling in the pooling layer [52]. Applying the pooling layer eliminates the input's dimensionality; the selection is primarily achieved by complete or average pooling [36]. The system randomly accesses a set of rectangles without overlap and uses a non-linear method for query purposes in the space between convolutional layers [49]. Regular pooling is the most common variety of pooling. It can better represent the data by stacking several convolutional layers on top of each other [51].
- After many overlaps and pooling layers, a complete layer of connection is in the previous strata. Both neurons are linked to each neuron in the current strata, and spatial information is not retained in the complete connection layer[51], [56]. The fully connected layer will incorporate local knowledge in conjunction layer or pooling layer, with category discrimination, in synthesizing the features derived from the front. The total complete linked layer still has the most parameters because of its fully connected existence. [52]. Some types of fully connected layers are available, including a fully connected input layer, the first fully connected layer, and a fully connected output layer. Often the flattening layer may be considered a connected input layer. The output of the previous layers is taken from this layer and transformed into a single vector. The ultimately linked output layer gives the mark probabilities [23]. The network is connected to the top connection layer after extraction of features from the convolutional layer, which increases the network extraction capacity and restricts the network's scale [52], [57].

4. Facial Expression Recognition Survey

This literature explores the latest research on the recognition of facial expression. It provides insight into the face, identification methods, model architecture for extraction and classification features, and precision obtained by the following researchers in FER's field.

Kaviya et al. [24] used two separate datasets, namely FER-2013 and customized datasets. The RGB image is pre-processed by turning it into a gray image for emotional recognition. The faces are then identified in real-time

photographs or static images of Haar. If the face is detected, the face characteristics can be resized and processed. CNN is used to train the facial traits acquired to identify them according to the five emotions. Community emotion is measured using a weighted average feeling. Finally, the expected group emotion is moved to the speech synthesizer to get audio output. The result showed that the proposed CNN would gain facial expression and the accuracy of testing of the model is 65% for FER-2013 and 60% for custom datasets.

Nie [51] learned a framework focused on deep learning of facial expression, which uses the OpenCV computer vision class library to complete face recognition; Opencv, based on the AdaBoost algorithm, is used for facial recognition. The neural network construction model's CNN architecture adopts Keras as the open-source artificial neural network library written by python. The model uses the Stochastic algorithm of Descent Gradient (SGD). He used the FER2013 database as CNN neural network training data.

Hussein et al. [36] recommended a CNN model to understand face emotions with three continuum emotions. This model uses residual blocks and depth-separable convolutions inspired by Xception to minimize the sum of parameters to 33k. They use a convolutional neural FER network for emotional stability identification. CNN uses convolution operations to learn extract features from the input images, which reduces the need to extract features from images manually. The proposed model offers 81 percent total precision for invisible results. It senses negative and positive emotions, respectively, with a precision of 87% and 85%. However, the accuracy of neutral emotion detection is just 51%.

Ravi et al. [58] made a fair contrast between two of the most widely used FER techniques and shed some light on their accuracy. LBP and CNN are the approaches being used here. The LBP is intended only to extract features to classify the extracted features from LBP using the SVM classifier. The results indicate that CNN is better than LBP for its integrated classifier (softmax).

Ganapathy et al. [22] attempted to identify emotional conditions with Electrodermal Activation (EDA) signals and the CNN's learned features. The EDA signals are received from the public DEAP database and are standardized and used to decompose into tonic and phasic sections by the cvxEDA process. In a short time, the phasic part will be subject to the change in the Fourier. 38-time, frequency, and time-frequency functions are extracted from the phasic signal. CNN uses these extracted properties to learn robust and powerful features. Five algorithms are being used to classify machine learning, including LDA, Multi-Layer Perceptron (MLP), SVM, DT, and Extreme Learning Machine (ELM). The findings demonstrate that the method presented can classify the emotional states with the dimensions of enthusiasm. The suggested technique used seems to help evaluate both the ordinary and clinical conditions' different emotional circumstances.

Fei et al. [37] proposed a new emotional intelligence system focused on a new deep CNN to promote mental state identification and diagnosis. The proposed method will process face images and analyze emotional, temporal creation via a new approach. Deep characteristics are excluded from the fully connected AlexNet 6 sheet, using a typical linear LDA to achieve the final classification results. The machine contains three parts: the input of videos of facial expressions, the pre-processing procedure of photographs, and the facial expression's predictive interpretation. Experiments presented show that the proposed method exceeds the other precise and efficient techniques, which indicate that it can serve as an intelligent, low-cost cognitive assist for identifying, tracking, and diagnosing the patient's mental health using an automated analysis of facial expression.

Ameur et al. [59] proposed an approach based on Monogenic Binary Pattern (MBP) and CNN to increase the face's detection rate. The proposed method firstly extracts salient local features by MBP, a robust local descriptor compared to the ubiquitous Gabor filters-based LBP models. Besides, they use DCNN, which is one of the best technologies for improving large-scale image recognition. MBP-CNN has strong robustness to variations of lighting, occlusion, facial expression, texture, and facial form since it incorporates MBP and neural network. MBP-CNN is even more reliable when combinations of global and local knowledge were used.

Bargshady et al. [29] suggested a model CNN standard Bidirectional Long Short-Term Memory (BiLSTM) hybrid method for the deep learning of facial images for four-stage pain recognition. Maximize the proposed algorithm's total calculation efficiency. The PCA (VGG) face's completely connected layers have been improved with a fully linked additional layer and the extracted features' dimensionality. The reduced extracted features, which were the most helpful patterns to determine pain severity, feed into the newly established Enhanced Joint Hybrid (EJH)-CNN-BiLSTM classification portion. Experimental findings found that the proposed EJH-CNN-BiLSTM system increases efficiency considerably without the use of the standard solution. The improved algorithm reached 98.4 percent Area under Curve (AUC) and 90 percent test precision in the UNBC-McMaster Shoulder Pain database. The artificial intelligence methods built in their study may have applicable repercussions for the diagnostic fields, especially in support of clinicians and other medical scientists' automated pain control practices.

Ozcan and Basturk [23] developed An optimized FER method. The fundamental aspects of this approach are data processing and hyperparameter tuning for CNN-supported transfer learning. The picture is transferred by the face alignment method as first for the data preparation portion. Then the sample is transformed from RGB to gray format, sound reduction, and picture sharpening steps. The data increase is the last step in the data planning section. The Particle Swarm Optimization (PSO) algorithm with high global search potential is then used to collect hyperparameters. The testing of this proposed optimized approach has been tested on the JAFFE data collection, and the best accuracy

score available in the literature was reached. The latest Erciyes University FER (ERUFER) dataset has been introduced. This dataset consists of 9005 samples in 10 groups. The ERUFER groups are six simple phrases, plus neutral, disdain, concern, and enthusiastic. The dataset can be commonly used because of the large number of participants and many samples.

Li et al. [34] suggested a new approach for understanding facial expression using an attention mechanism. Not only raw images but also LBP features are applied to the network attention layers. LBP features provide texture details, representing delicate skin textural changes that may help to discern gestures with slight distinctions. They also compiled and branded a new dataset to identify facial expressions called Nanchang University Facial Expression (NCUFE). The dataset consists of 490 photographs obtained from 35 subjects with seven facial expressions (i.e., anger, disgust, fear, happiness, sadness, surprise, and neutral). They captured both RGB images and depth images for each subject. Substantial studies on five separate datasets are carried out. Datasets, such as CK+, JAFFE, Oulu-CASIA, and NCUFE, are obtained in the real world and those collected under laboratory conditions such as FER2013. They also equate model output with state-of-the-art algorithms for speech recognition. The model findings demonstrate that it is superior to many of the current data sets approaches. The process is, therefore, only suitable for 2D images.

Meryl et al. [49] tested the efficacy of CNN for expression recognition with (RBF). The patients' psychiatric problems can quickly be assessed, and remedial steps accelerated. They focus on studying how people feel, cope, and gain ill health by analyzing their facial expressions. (FER) has essential steps in the extraction and classification of features. The designation is one of the critical mechanisms by which terms such as pleasure, sadness, anger, hate, surprise, and fear are identified. There are three kinds of signals on the face: static, sluggish, and heavy. The experimental results indicate that the suggested mix of methods gives the (FER) 2013 data set comparatively improved precision.

Jiang et al. [60] introduced a new loss feature called the advanced softmax loss to eradicate imbalanced training expressions. The proposed losses guarantee that any class would have a level playing field and potential using fixed (unlearnable) weight parameters of the same size and equally allocated in angular space. The research shows that proposed (FER) methods are better than specific state-of-the-art FER methods. The proposed loss can be used as an isolated signal or used simultaneously with other loss functions. To sum up, detailed studies on FER2013 and the real-world practical face (RAF) databases have shown that ASL is considerably more precise and effective than many state-of-the-art approaches.

Agrawal et al. [61] introduced two novel CNN architectures based on a FER-2013 dataset. These are both main and unique for the collection of hyperparameters across network layers. The two Model 1 and Model 2 network architectures achieve human accuracy with a FER-2013 dataset. The Model2 edition is streamlined. Model 1 architecture is unique because it uses a set kernel size and specifies the number of filters across the network depth. The number of filters decreases with network depth in this design. Model2 is compacter to Model1. Both architectures use a particular kernel size of 8—compare the proposed models with the state-of-the-art signs of the most fitting architecture for data set FER-2013.

Chen et al. [14] suggested a two-stage social signal analysis method focused on DCNN to understand facial expression. In the context of the face language's non-state nature, the proposed system comprises two stages: the first stage is taken immediately out of the sequence of facial expression by the SoftMax score for a binary CNN. In the second stage, the chosen neutral expression framework and the fully-expressed frame are then fed to the corresponding DCNN. The suggested approach will essentially remove the individual discrepancy with the neutral language system's variations and the entire expression frame. The accuracy achieved was 96,28% in an e-learning context for the student's affective condition study.

Mohan et al. [25] created a new DCNN system to categorize facial expressions using holistic features of features. In this way, a biological feature extractor is used to retrieve the local low-level features before using the proposed Deep Belief network (DCNN) model. The GF feature extractor will generate two local intermediate features (M and D). At the end of the proposed DCNN model, a Softmax classifier is used to determine the likelihood values of 7 expressions. This process combines the outputs using measured data and discrete data based on the proposed model. Empirical findings suggest that both local and holistic elements will boost the FER mission together. Nevertheless, in general, the efficiency of FER in a laboratory-controlled environment is not as good. Besides, the proposed model should be investigated in specific real-life applications.

Cheng and Zhou [52] suggested an expression reconnaissance model for a DCNN enhanced VGG (CNN). The model optimizes network structure and network parameters based on the VGG-19. They used migration learning methods to address the lack of picture samples. In the CK+ Database, Shallow CNN, Alex-Net, and enhanced VGG-19 deep CNN train and evaluate the facial expression data and compare the findings collected for experiments. The test results showed that in the extraction of image features, the DCNN is superior to the shallow CNN, but a database should calculate the number of layers, and the number should be moderate. Otherwise, the over-fit phenomenon occurs. The combination of convolution layers with tiny filters will also display more efficient input data features and use fewer parameters than a convolution layer with large filters with the same receptivity. The enhanced VGG-19 network model will achieve the performance of other network models.

Zou et al. [62] suggested the inclusion of batch regularization and the ReLU activation function convolution neural network compared to the original CK+ neural network to overcome the gradient disappearing problem. Dropout technology is added to solve the issue of network fitting. To increase identification accuracy, they suggest the extracted features be used to classify the expressions. They compared their findings with the CK+ data collection AlexNet and Visual Geometry Group (VGG-19) algorithms. The results indicate that the enhanced neural network algorithm's average recognition score is 6.9% higher than that of the traditional AlexNet algorithm. The identification rate also increased by 4.07 percent compared to the VGG-19 algorithm.

Wang et al. [63] suggested a system combined with features derived from CNN and the C4.5 classification to identify facial expressions, which can fix the incompleteness of hand-made features and prevent the high hardware configuration in the deep learning model. In the meantime, they proposed several changes to the C4.5 classification and the conventional random forest in experiments. The C4.5 classification is chosen for the appreciation of speech. Since a few concerns are taken into account, such as overfitting and poor generalization abilities of the single classifier, ensemble learning is used to boost classification precision in the Decision Tree (DT) algorithm. The random forest algorithm is then chosen as the facial expression classification. A significant number of studies have demonstrated the reliability and viability of the approach proposed.

Wang et al. [64] decreased due to lack of image knowledge and noise interference under occlusion, the conventional machine learning process. The classical LeNet-5 model has been improved, and the multi-layer cross-connected LeNet-5 CNN model has been created and introduced. Various parameters, including the learning rate, convolution kernel size, and the CNN model's excitation function, are calculated using experimental study. In all cases of deterministic occlusion and uncertain occlusion, the CK+ expression database based on data increase is used. A random occlusion 1, a random occlusion 2, a random occlusion 3, etc., can be tested with various network models and learning algorithms. The model is stable and has a high detection rate under occlusion through careful review of the data. To solve the issue of occlusion-robust detection.

Wang et al. [65] suggested a new approach for hybrid transfer learning based on the advantages of two models CNN and Convolution Restricted Boltzmann Machine (CRBM), in the hope of reducing the material gaps of the two separate data sets before using the transmission learning algorithm. When the goal settings are treated as the deep CNN transfer model's input, the enhanced CRBM is used to replace the conventional neural network model's entire connecting layer. The enhanced CRBM model's feedback is obtained by integrating all sorts of characteristic maps with general structural characteristics. Based on the current hypotheses, the benefits of the latter model are thoroughly applied. The special higher-order statistical features of the target data set are obtained by retraining the CRBM model using efficient methods for ensuring an effective image classification. The variation in the content of data sets does not influence the role recognition capability.

Otberdout et al. [66] addressed deep covariance descriptors and (FER) significant covariance trajectories in response to static and dynamic performance, respectively. Our concept is of a compact matrix of global and local DNN features. After noiseless flattening, a softmax layer is used to get a confusion matrix from each feature vector. They coded in the compact covariance matrices all linear associations between in-depth facial features taken from the last convolutional level. They labeled these static descriptors with the Symmetric Positive Definite (SPD) multiple-defined gaussian kernel to respect the covariance matrices' non-linear form in an SPD multiplier. The findings show that this classification method is productive than the traditional classification of fully linked and softmax layers.

Ozdemir et al. in [28] suggested the low-cost and practical approach of classifying in real-time seven different emotions based on LeNet CNN architecture (happy, sad, shocked, furious, hideous, terrified, and neutral). Where facial expression pictures, of which a limited number can be said, have been educated on CNN and have attained a high degree of accuracy. The effect of unimportant pixels beyond facial expressions has been minimized with the Haar Cascade library. Moreover, one-proof positioning of the pixels in the images in networks resulted in decreased training time and the number of networks. They combined three separate datasets (KDEF, JAFFE, and their custom dataset). The evaluation and test performance of a custom database are better than training on established databases. The real-time test model has the feature to scan in any second picture.

5. Discussion

Table 1 shows that all studies used CNN's different techniques/algorithms and architecture to achieve a higher rate of deeper understanding to improve precision, trust, and performance. It has been noticed that seven researches ([28] [29], [37], [52] [61], [64] and [65]) focused on the architecture of the CNN, as some used old architectures such as LeNet-5, as in [28] and [64]. In [64], the author introduced a convolution layer and a LeNet-5 pooling layer to resolve the robust occlusion ER issue. Simultaneously, in [28], the author suggested the low-cost and practical approach of classifying seven distinct emotions in real-time dependent on LeNet CNN architecture to accomplish a high-level model to understand feelings. From here, it is noticed that both of them obtained a good result and were close to each other using a different dataset. In [64], the accuracy was 97.23 % using the data set CK+. Simultaneously, in [28], it combined three types of data set together (JAFFE, KDEF, and custom), and the accuracy obtained was 96.43%.

In [37] and [52], the author's use of Alex-Net architecture is noted, which is more recent than LeNet. In [37], the author used DCNN with the entire Alexnet architecture to create user-friendly, inexpensive, and reliable systems. The average accuracy was reached 88% using JAFEE and KDEF dataset. While in [52], the writer used Alex-Net, Shallow CNN, and enhanced VGG-19 DCNN for data processing to boost the machine learning algorithm's accuracy and performance, with the accuracy reached 96 %. After observing the results from these two previous research types, the accuracy was higher with DCNN [52]. The findings revealed that the DCNN was better than the shallow CNN at extracting image features that allowed more layers. Still, too many layers would result in overfitting of the training data, and output may decline.

Some researchers have used the VGG architecture in the FER, a more contemporary architecture than Alex-net and LeNet, as in [29], the author used VGG architecture and the PCA dimension reduction process to extend effective facial recognition to binary classes of gestures. The proposed method can significantly impact medical research areas with an accuracy of 90 % to detect pain.

Besides, some studies have combined two architectures at once, as indicated in [61], where the author used both VGG net architecture and Alexnet architecture to create a reduced model for state of the art, precision results proved that the size of the kernel and the number of layers influence network accuracy.

To avoid the negative impact of the transition of learning characteristics between various data sets, the author in [65] replaced the fully connected layer in CNN by CRBM. It is used to identify facial expressions for learning transmission where the accuracy differed according to the dataset used, which reached the highest level 99.2% with the JAFFE and the lowest 73.75% with the FER2013.

On the other hand, many feature extraction methods have been reviewed in this study. Among them, two researchers ([34] and [58]) depended on LBP. In [34] and the attention mechanism, the writer used LBP to develop the attention model to achieve better outcomes. CNN and LBP were used in research [58]. Results show that CNN, with its built-in classifier (softmax), is better than LBP in terms of accuracy, there was 97.32% with the CK+ dataset, but with YALE FACE, it was low, about 31.8 %. MBP is used in [59] to improve FER system performance, a robust local descriptor for feature extraction compared to the widespread Gabor filters-based LBP models.

Moreover 6 researchers ([25], [28], [61], [62], [64] and [65]) relied on DCNN with its built-in extraction and classification features. In [62] the author used maximum pooling method for feature extraction, and softmax as classification method. In [25], they proposed DCNN that has two branches. The first branch examines geometric features, such as edges, curves, and arcs, while the second branch extracts holistic features. The results showed that both local and holistic elements would effectively improve the FER mission.

In [24], the author used the Haar filter to extract facial features. As observed, the accuracy ratio was low compared to other methods, taking into account the used dataset, where the accuracy was 65% using the FER2013 dataset. In [29], the PCA was used for the dimension reduction process rather than extracting the features.

Also, some reviewed FER methods were based on the optimization approaches. These methods used several optimized algorithms such as PSO in [23], SGD in [14], and [51]. In [23], PSO is used in hyperparameter selection. In contrast, SGD in [14] and [51] is used for training a neural network for increasing the precision of recognition.

Typically, the accuracy of the methods examined differs from one methodology to another. However, the optimization-based methods of selection of features obtained higher efficiency than traditional methods.

On the other hand, most of the reviewed methods were based on classification approaches. Where these methods used several classification algorithms such as SVM in [66] and [58], SDM in [14], in some reviewed papers, more than one classifier was used such as SVM, LDA, and KNN in [37], also LDA, MLP, SVM and DT in [22]. Nevertheless, this article observed that the SVM algorithm's use is the best among all the algorithms used to increase accuracy. Even SVM generates different accuracy on various posed databases, to approximately 97% when dealing with the static image in laboratories such as CK+ data set. Simultaneously, with a wild data set such as (SFEW, AFEW), the accuracy is less than 49%, 49.5%, respectively, up to half of it.

However, in [63] decision tree algorithm (C4.5 classifier) is used to reduce the machine learning model's hardware. These new methods offer considerable advantages over similar approaches in other fields. The results showed that C4.5 had an accuracy of 99.9%, 98.9% with CK+, and JAFFE, respectively, while FER-2013 had an accuracy of 84.3%.

The two researchers listed above use the Softmax function for the CNN parameters and the Adam optimization algorithm. According to the test findings, the Adam Optimization technique proposed by [36] comparing to the softmax function was used by [60], accomplishing a higher precision score of 87 percent accuracy. It is observed that everyone reached good results, especially in terms of accuracy, with different rates, and using different data sets, which impacts the results. Using the CK+ database, everyone reached high results with more than 95% and recorded the lowest rate with the (FER) 2013 database.

Table 1 - A table of comparison.

Ref.	Objectives	Technique	Significant results	Dataset	Accuracy
Kaviya et al. [24]	To improve the recognition accuracy.	Haar filter to extract facial features.	Prove the facial expression features of the proposed CNN that can be learned.	FER-2013 custom dataset	65% 60%
Nie [51]	To improve the recognition accuracy.	SGD algorithm is used for training a neural network.	Accuracy increased via the evaluation.	FER-2013	—
Hussein et al. [36]	Improve (FER) to detect Emotional Stability.	1-Adam optimizer 2-loss function	The identification of positive and negative emotions is slightly more significant than a neutral emotion.	CK+ FER2013	Overall accuracy 81%
Ravi et al. [58]	To show the equal difference in terms of accuracy between CNN and LBP.	1-LBP for feature extraction 2-SVM as a classifier	The results indicate that CNN is better than LBP for its integrated classifier (softmax).	CK+, YALE FACE	97.32% 31.82%
Ganapathy et al. [22]	Offers more essential and stable characteristics for dynamic EDA signals.	LDA, MLP, SVM, DT, and ELM are known as characterizing EDA signals on arousal valence scale.	Prove the CNN based attribute education can help classify common feelings under different health circumstances.	DEAP database	75% for arousal 72% for valence
Fei et al. [37]	To create user-friendly, inexpensive, and reliable systems for mental health FE analysis.	1-DCNN for the entire AlexNet 2- SVM, LDA, and KNN classifier algorithms.	This is highly accurate to understand compared with many other algorithms for DL.	JAFFE+ Patient Dataset KDEF+ Patient Dataset	89.90% 89.33%
Ameur et al. [59]	To improve the performance of the FR system.	MBP	Offered an example of relative success on the LFW scheme, relative to other approaches.	Labeled Faces in the Wild (LFW)	98.90%
Bargshady et al. [29]	To extend effective facial recognition to binary classes of gestures, such as pained or non-pain.	VGG-Face and Principal Component Analysis (PCA) dimension reduction process.	The proposed method can have a significant impact on medical research areas.	human-face images, UNBC (McMaster Shoulder Pain archive)	90 %.
Ozcan and Basturk [23]	To increase the success of DL methods.	PSO approach to... the hyperparameter tuning operation.	Increasing the number of training samples which impacts the positive predictive result.	ERUFER JAFFE.	92,56% 99.53 %
Li et al. [34]	To avoid overfitting and prevent the generalization ability of the training data.	LBP is combined with an attention mechanism.	Model findings demonstrate that it is superior to many of the current data sets approaches. The process is, therefore, only suitable for 2D images.	Nanchang University of Facial Expression (NCUFE) CK+ FER2013	94.33% 98.68% 75.82%
Meryl et al. [49]	To provide better accuracy.	Radial basis function (RBF)	The efficacy of CNN in the RBF is	FER 2013	95.4%.

			investigated.		
Jiang et al. [60]	To increase accuracy and reliability.	Advanced ASL (softmax loss) to minimize data imbalance-induced bias	The proposed ASL methodology is more accurate and precise than existing systems.	FRE2013 and Real World Affective Facial Coding (RAF).	70.80% for a private test 68.56 % for public test.
Mohan et al. [25]	To highly reliably differentiate feelings	DCNN method to classify a person's facial expression.	Experimental studies suggest that both local and holistic characteristics will strengthen the FER mission together.	FER2013 JFFE CK+ KDEF RAF	78% 98% 97.9% 96.36% 82.33%
Agrawal et al. [61]	To build and customize a network architecture (fast to train and easy to implement).	VGGnet and the Alexnet architecture.	Results proved that the size of the kernel and the number of layers influence network accuracy.	FER2013	65%
Chen et al. [14]	To develop the characteristics of human facial expressions.	1-End-to-end DCNN 2-The (SGD) algorithm for optimization. 3-Supervised Descent Method (SDM) as a classifier	The proposed method can effectively eliminate individual differences.	CK+ BU-4DFE	95.4% + 77.4%
Zou et al. [62]	To overcome low-recognition challenges and difficulty of conventional (FER) approaches.	Maximum pooling method for feature extraction. Softmax as a classification method.	Enhanced CNN can increase image recognition accuracy of face speech relative to standard methods.	CK+	99.14%
Wang et al. [63]	To reduce the hardware in the machine learning model.	C4.5 for DT to detect and classify FE. Random forest for DT to solve the bottleneck problem.	These new methods offer considerable advantages (reliability and viability) over similar approaches in other fields.	JAFFE CK+ FER-2013 RAF-DB	98.9% 99.9% 84.3% 92.3%
Wang et al. [64]	To resolve the issue of robust occlusion ER. To improve the recognition accuracy.	Introduces a layer of convolution and a pooling layer based on LeNet-5	A stable model with a high occlusion recognition score	CK+	97.23%
Wang et al. [65]	Develop the ability to identify facial expressions for learning transmission.	The fully connected layer in CNN mode is to be replaced by CRBM.	Effectively avoid the negative impact of the transition of learning characteristics between various data sets.	JAFFE, FER2013, SFEW and RAF-DB	99.2% 73.75% 96.09% 92.97%
Cheng and Zhou [52]	To facilitate accurate classification and boost the accuracy and performance of a machine learning algorithm.	Shallow CNN, Alex-Net and improved VGG-19 DCNN.	Showed that the DCNN was superior to the shallow CNN at extracting image features and allowed more layers. Still, too many layers would result in overfitting of the training data,	CK+	96%

			and output may decline.			
Ozdemir et al. in [28]	To accomplish a high-level model to understand feelings through FE.	LeNet architecture.	CNN	1-It has provided data with higher quality than can be found in existing databases. 2-The real-time test model offers images of the subjects in real-time.	Merged three dataset (JAFFE, KDEF, and custom)	96.43%
Otberdout et al. [66]	To increase the efficiency of FER analysis.	SVM classifier using a suitable Gaussian kernel in the SPD manifest.		The researchers suggested a new method for (FER) by using deep covariance function descriptors.	Oulu-CASIA CK+ (static facial expression in the wild) SFEW (acted facial expressions in the wild) AFEW	87% 98% 49% 49.5%

6. Conclusion

Facial Expression Recognition (FER) is one of the most effective ways of delivering emotional knowledge but is often confined to six-basic emotional learning plus neutral learning. This article has addressed the latest FER analysis. Numerous CNN architectures have been identified that have recently been proposed. They have provided various databases of random photographs obtained from the actual world and other laboratories to detect human emotions accurately. Researchers have learned and checked their patterns in many datasets to verify the neural network architecture's feasibility. The recognition rates differ from one database to another in compliance with the same deep learning pattern. The growing volume of data, processing time, and computational resources needed for the research processes will influence the algorithms used in classification, adequate memory, and model performance.

Acknowledgement

The authors would like to thank the Duhok Polytechnic University, Duhok, Kurdistan Region, Iraq.

References

- [1] Hu, L., Li, W., Yang, J., Fortino, G. and Chen, M. (2019). A Sustainable Multimodal Multi-layer Emotion-aware Service at the Edge. *IEEE Transactions on Sustainable Computing*, doi: 10.1109/TSUSC.2019.2928316
- [2] Shrivastava, V., Richhariya, V. and Richhariya, V. (2018). Puzzling Out Emotions: A Deep-Learning Approach to Multimodal Sentiment Analysis, 2018 International Conference on Advanced Computation and Telecommunication (ICACAT), Bhopal, India, doi: 10.1109/ICACAT.2018.8933684
- [3] Perveen, N., Roy, D. and Chalavadi, K. M. (2020). Facial Expression Recognition in Videos Using Dynamic Kernels. *IEEE Transactions on Image Processing*, 29, 8316-8325
- [4] Zhi, R., Zhou, C., Li, T., Liu, S. and Jin, Y. (2020). Action Unit Analysis Enhanced Facial Expression Recognition by Deep Neural Network Evolution. *Neurocomputing*, 425, 135-148
- [5] Li, S. and Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.2981446
- [6] Mahmood, M. R. and Abdulazeez, A. M. (2017). A Comparative study of a new hand recognition model based on line of features and other techniques, *International Conference of Reliable Information and Communication Technology*. 420-432
- [7] Mellouk, W. and Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science*, 175, 689-694
- [8] Wen, G., Chang, T., Li, H. and Jiang, L. (2020). Dynamic Objectives Learning for Facial Expression Recognition. *IEEE Transactions on Multimedia*, 22, 2914-2925
- [9] Ekman, P., Friesen, W. V. and Hager, J. (1978). *Investigator's guide to the facial action coding system*. Paolo Alto, CA: Consulting Psychologists Press

- [10] Bailly, K. and Dubuisson, S. (2017). Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. *IEEE Transactions on Affective Computing*, 10, 167-181
- [11] Liu, D., Ouyang, X., Xu, S., Zhou, P., He, K. and Wen, S. (2020). SAANet: Siamese action-units attention network for improving dynamic facial expression recognition. *Neurocomputing*, 413, 145-157
- [12] Zhi, R. and Wan, M. (2019). Dynamic Facial Expression Feature Learning Based on Sparse RNN. 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 1373-1377
- [13] Wu, M., Su, W., Chen, L., Pedrycz, W. and Hirota, K. (2020). Two-stage Fuzzy Fusion Based-Convolution Neural Network for Dynamic Emotion Recognition. *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2020.2966440
- [14] Chen, J., Lv, Y., Xu, R. and Xu, C. (2019). Automatic social signal analysis: Facial expression recognition using difference convolution neural network. *Journal of Parallel and Distributed Computing*, 131, 97-102
- [15] Chen, L., Ouyang, Y., Zeng, Y. and Li, Y. (2020). Dynamic facial expression recognition model based on BiLSTM-Attention. 15th International Conference on Computer Science & Education (ICCSE), 828-832
- [16] Lasri, I., Solh, A. R. and El Belkacemi, M. (2019). Facial Emotion Recognition of Students using Convolutional Neural Network. 2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS), 1-6
- [17] Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N., et al. (2020). Facial sentiment analysis using AI techniques: state-of-the-art, taxonomies, and challenges. *IEEE Access*, 8, 90495-90519
- [18] Sun, X., Xia, P., Zhang, L. and Shao, L. (2020). A ROI-guided deep architecture for robust facial expressions recognition. *Information Sciences*, 522, 35-48
- [19] Samadiani, N., Huang, G., Cai, B., Luo, W., Chi, C.-H., Xiang, Y., et al. (2019). A review on automatic facial expression recognition systems assisted by multimodal sensor data. *Sensors*, 19, 1863
- [20] Babajee, P., Suddul, G., Armoogum, S. and Foogooa, R. (2020). Identifying Human Emotions from Facial Expressions with Deep Learning. 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), 36-39
- [21] Ekundayo, O. and Viriri, S. (2019). Facial expression recognition: a review of methods, performances and limitations. 2019 Conference on Information Communications Technology and Society (ICTAS), 1-6
- [22] Ganapathy, N., Veeranki, Y. R. and Swaminathan, R. (2020). Convolutional Neural Network based Emotion Classification using Electrodermal Activity Signals and Time-Frequency Features. *Expert Systems with Applications*, 113571
- [23] Ozcan, T. and Basturk, A. (2020). Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization. *Multimedia Tools and Applications*, 79, 26587-26604
- [24] Kaviya, P. and Arumugaprakash, T. (2020). Group Facial Emotion Analysis System Using Convolutional Neural Network. 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), 643-647
- [25] Mohan, K., Seal, A., Krejcar, O. and Yazidi, A. (2020). Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-12
- [26] Sulaiman, D. M., Abdulazeez, A. M., Haron, H. and Sadiq, S. S. (2019). Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization. 2019 International Conference on Advanced Science and Engineering (ICOASE), 82-87
- [27] Othman, G. and Zeebaree, D. Q. (2020). The Applications of Discrete Wavelet Transform in Image Processing: A Review. *Journal of Soft Computing and Data Mining*, 1, 31-43
- [28] Ozdemir, M. A., Elagoz, B., Alaybeyoglu, A., Sadighzadeh, R. and Akan, A. (2019). Real Time Emotion Recognition from Facial Expressions Using CNN Architecture. *Medical Technologies Congress (TIPTKNO)*, 1-4
- [29] Bargshady, G., Zhou, X., Deo, R. C., Soar, J., Whittaker, F. and Wang, H. (2020). Enhanced deep learning algorithm development to detect pain intensity from facial expression images. *Expert Systems with Applications*, 149, 113305
- [30] Lee, J. R. H. and Wong, A. (2020). TimeConvNets: A Deep Time Windowed Convolution Neural Network Design for Real-time Video Facial Expression Recognition. 17th Conference on Computer and Robot Vision (CRV), 9-16
- [31] Xue, M., Mian, A., Duan, X. and Liu, W. Learning Interpretable Expression-sensitive Features for 3D Dynamic Facial Expression Recognition. 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), 1-7
- [32] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J. (2020). A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction. *Journal of Applied Science and Technology Trends*, 1, 1, 56-70
- [33] Le, T. T. Q., Tran, T.-K. and Rege, M. (2020). Dynamic image for micro-expression recognition on region-based framework. 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), 75-81

- [34] Li, J., Jin, Zhou, K., Kubota, N. and Ju, Z. (2020). Attention Mechanism-based CNN for Facial Expression Recognition. *Neurocomputing*, 411, 340-350
- [35] Chinchankar, N. A. (2019). Facial Expression Recognition Using Deep Learning: A Review. *International Research Journal of Engineering and Technology (IRJET)*, 6, 3274-3281
- [36] Hussein, E. S., Qidwai, U. and Al-Meer, M. Emotional Stability Detection Using Convolutional Neural Networks, 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 136-140
- [37] Fei, Z., Yang, E., Li, D. D.-U., Butler, S., Ijomah, W., Li, X., et al. (2020). Deep convolution network based emotion analysis towards mental health care. *Neurocomputing*, 388, 212-227, 2020
- [38] Baddar, W. J., Lee, S. and Ro, Y. M. (2019). On-the-Fly Facial Expression Prediction using LSTM Encoded Appearance-Suppressed Dynamics. *IEEE Transactions on Affective Computing*, doi: 10.1109/TAFFC.2019.2957465
- [39] Najat, N. and Abdulazeez, A. M. (2017). Gene clustering with partition around mediods algorithm based on weighted and normalized Mahalanobis distance. 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 140-145
- [40] Zeebaree, D. Q., Abdulazeez, A. M., Hassan, O. M. S., Zebari, D. A. and Saeed, J. N. (2020). Hiding Image by Using Contourlet Transform, *The Mattingley Publishing Co., Inc* 83, 16979 - 16990
- [41] Sulaiman, M. A. (2020). Evaluating Data Mining Classification Methods Performance in Internet of Things Applications. *Journal of Soft Computing and Data Mining*, 1, 11-25
- [42] Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H. and Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access*, 8, 203097-203116
- [43] Deng, L., Wang, Q. and Yuan, D. (2019). Dynamic Facial Expression Recognition Based on Deep Learning. 2019 14th International Conference on Computer Science & Education (ICCSE), 32-37
- [44] Zeebaree, D. Q., Haron, H. and Abdulazeez, A. M. (2018). Gene selection and classification of microarray data using convolutional neural network. 2018 International Conference on Advanced Science and Engineering (ICOASE), 145-150
- [45] Lopes, A. T., de Aguiar, E., De Souza, A. F. and Oliveira-Santos, T. (2017). Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. *Pattern Recognition*, 61, 610-628
- [46] Kong, F. (2019). Facial expression recognition method based on deep convolutional neural network combined with improved LBP features. *Personal and Ubiquitous Computing*, 23, 531-539
- [47] Dong, J., Zheng, H. and Lian, L. (2018). Dynamic facial expression recognition based on convolutional neural networks with dense connections, 2018 24th International Conference on Pattern Recognition (ICPR), 3433-3438.
- [48] Sharma, A., Sharma, Y. and Mansotra, V. (2017) Study and Analysis of Students Emotions with Respect to Perception of Learning in a Class. *International Journal of Engineering Science Invention*, 6,10-26
- [49] Meryl, C. J., Dharshini, K., Juliet, D. S., Rosy, J. A. and Jacob, S. S. (2020). Deep Learning based Facial Expression Recognition for Psychological Health Analysis. 2020 International Conference on Communication and Signal Processing (ICCSP), 1155-1158
- [50] Abdulazeez, A., Salim, B., Zeebaree, D. and Doghramachi, D. (2020). Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol, 2020
- [51] Nie, Z. (2020). Research on facial expression recognition of robot based on CNN convolution neural network. 2020 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), 1067-1070
- [52] Cheng, S. and Zhou, G. (2020). Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network. *International Journal of Pattern Recognition and Artificial Intelligence*, 34, 2056003
- [53] Verma, M., Vipparthi, S. K., Singh, G. and Murala, S. (2019). LEARNet: Dynamic imaging network for micro expression recognition. *IEEE Transactions on Image Processing*, 29, 1618-1627
- [54] Zebari, N. A., Zebari, D. A.D., Zeebaree, Q. and Saeed, J. N. (2021). Significant features for steganography techniques using deoxyribonucleic acid: a review. *Indonesian Journal of Electrical Engineering and Computer Science*, 21, 338-347
- [55] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M. and Zebari, D. A. (2019). Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer. 2019 International Conference on Advanced Science and Engineering (ICOASE), 106-111
- [56] Bargarai, F., Abdulazeez, A., Tiryaki, V. and Zeebaree, D. (2020).. Management of Wireless Communication Systems Using Artificial Intelligence-Based Software Defined Radio. *International Journal of Interactive Mobile Technologies (iJIM)*, 14, 107-133
- [57] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M. and Zebari, D. A. (2019). Machine learning and Region Growing for Breast Cancer Segmentation. 2019 International Conference on Advanced Science and Engineering (ICOASE), 88-93
- [58] Ravi, R. andYadhukrishna, S. (2020). A face expression recognition using CNN & LBP. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 684-689

- [59] Ameer, B., Belahcene, M., Masmoudi, S. and Hamida, A. B. (2020). Unconstrained Face Verification Based on Monogenic Binary Pattern and Convolutional Neural Network. 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 1-5
- [60] Jiang, P., Liu, G., Wang, Q. and Wu, J. (2020). Accurate and Reliable Facial Expression Recognition Using Advanced Softmax Loss with Fixed Weights. *IEEE Signal Processing Letters*, 27, 725-729
- [61] Agrawal, A. and Mittal, N. (2020). Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy. *The Visual Computer*, 36, 405-412
- [62] Zou, J., Cao, X., Zhang, S. and Ge, B. (2019). A Facial Expression Recognition Based on Improved Convolutional Neural Network. 2019 IEEE International Conference of Intelligent Applied Systems on Engineering (ICIASE), 301-304
- [63] Wang, Y., Li, Y., Song, Y. and Rong, X. (2019). Facial Expression Recognition Based on Random Forest and Convolutional Neural Network. *Information*, 10, 375
- [64] Wang, G. and Gong, J. (2019). Facial expression recognition based on improved LeNet-5 CNN, 2019 Chinese Control and Decision Conference (CCDC), 5655-5660
- [65] Wang, Y., Li, Y., Song, Y. and Rong, X. (2019). The Application of a Hybrid Transfer Algorithm Based on a Convolutional Neural Network Model and an Improved Convolution Restricted Boltzmann Machine Model in Facial Expression Recognition. *IEEE Access*, 7, 184599-184610
- [66] Otherdout, N., Kacem, A., Daoudi, M., Ballihi, L. and Berretti, S. Automatic analysis of facial expressions based on deep covariance trajectories. *IEEE Transactions on Neural Networks and Learning Systems*. 31, 3892-3905