Compound Facial Emotional Expression Recognition using CNN Deep Features

Khadija SLIMANI, Yassine RUICHEK, Rochdi MESSOUSSI

***Abstract*—Most advanced researches concerning facial expressions identification generally deals with the six basic emotions with a single tag, namely happiness, anger, disgust, fear, surprise and sadness in addition to the neutral facial expression. It has recently been discovered that some affective computing analysts are more interested in delicate emotions such as a compound facial expression that deals with primary and complementary emotions (e.g. "happily-surprised" and "sadly-angry"). In addition, these emotions are more complete than the seven basic facial emotions. Therefore, current research focuses on the examination of fundamental and compound emotions, using databases of Compound Facial Expressions of Emotions (CFEE) and spontaneous Real-world Affective Faces (RAF). In this article, transfer learning based on Residual Neural Network was experimented to create a deep feature extraction model, and then different machine learning methods are used for classification. Experiments on basic and compound emotions prove that the proposed system can perfectly surpass advanced approaches. As expected, the compound emotions are more difficult to identify than the seven basic emotions.**

***Index Terms*—Compound emotions, Basic emotions, ResNet, Machine learning, Feature extraction, SVM.**

I. INTRODUCTION

THE identification of human emotions using machines is an essential, unusual, and complex artificial intelligence issue. Emotions deal specifically with interactions/communications and they help individuals to understand human attitudes and desires. There are two major types of human communication, verbal and non-verbal. Words are considered atomic information and a component of verbal interactions. On the other hand, facial expressions, body position, skeletal muscle activities, and physical responses are the atomic units of the non-verbal interaction. Usually, people reveal their emotions through facial expressions [1]. Individuals have the exceptional ability to identify facial expressions in real time. Machines struggle to decipher facial expressions. Anxiety, perceptions, and the way people from diverse cultures express themselves are some of the limitations that make this task very complex [2].

The human faces can transmit complex messages that may correspond to several communicative functions as well as a mixture of emotions of different kinds. For example, the superposition of two emotions, the masking of an emotion felt by another not felt, a bad evaluation of the emotion, or a combination of the six fundamental expressions [3]. When an individual is pleasantly surprised, his/her attitude is totally different from that of a happy but not surprised person [4]. Several parts of the face can help express some facial emotions better than other. For example, the eyes and eyebrows contain more information to describe the categories of anger and fear. Moreover, the mouth part helps to recognize the emotions of disgust and happiness, while the eyes, eyebrows, and mouth express surprise [5], [6], [7], [8]. Besides, this article aims to identify fundamental and compound facial expressions using the CNN Residual network (ResNet) to construct the deep feature extraction model and machine learning methods for the classification phase.

The rest of this paper is structured as follows: Section II provides an overview of related work on basic and compound emotions recognition using traditional and deep learning techniques. Section III describes the proposed methodology including the databases, the methods used in our experiments, and the proposed approach. The results and the discussion are presented in section IV, which includes experimental setup and results on posed and spontaneous basic and compound facial expressions. Then, section V compares our results with state-of-the-art approaches. Finally, the conclusion is presented in section VI.

II. RELATED WORK

A. Multi-label Facial Expression Definition

Different computer vision and automated learning algorithms have been employed to recognize facial expressions [9] due to the essential role played by FER (Facial Expression Recognition) in various areas like e-education, video games, human-robot interaction, distance psychotherapy, and human automated interaction [10], [11], [12], [13]. Recent developments in machine facial examination, pattern identification, semantic network, and many other machine learning procedures have made it possible to create machine face identification and facial expression identification schemes that focus on these topics. Inspired by Darwin's research [14], Ekman and Friesen analyzed the link between emotions and facial expressions. They came across a subset of emotions related to certain particular facial expressions known as the fundamental emotions i.e. happiness, sadness, surprise, anger, fear, and disgust [15], [16], [17]. Meanwhile, the Facial Action Coding System (FACS) was planned in [17] in order to study the facial expressions and the complementary emotions illustrated by all the activities of the facial muscles. Facial expressions could be examined by systemizing the facial action units of the different parts of the face into codes.

Manuscript received February 4, 2021; revised October 15, 2022. This work was supported by the National Center for Scientific and Technological Research through the scholarship of excellence no. 757UIT, Volubilis program AI 33/SI/14, and BOOSTE TON DOC 2019 program.

K. SLIMANI, Research Engineer, qassil SAS, 75001 Paris, France. Systems of Telecommunication and Decision Engineering Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco. E-mail: slimani.khadija@qassil.com

Y. Ruichek, Higher Education Professor, Distributed Knowledge and Artificial Intelligence Laboratory, UMR 7533, University of Bourgogne Franche-Comte, UTBM, Belfort, France. E-mail:yassine.ruichek@utbm.fr

R. Messoussi, Higher Education Professor, Systems of Telecommunication and Decision Engineering Laboratory, Faculty of Sciences, Ibn Tofail University, Kenitra, Morocco. E-mail: messoussi@uit.ac.ma

Previous studies on the identification of facial expressions are mainly based on the classification of the fundamental emotions [18], [19]. In fact, while studying natural human interactions, humans do not reveal the precise emotional content. Some definitive researches such as Ekman and Friesen [20], Hassin et al. [21], Izard [22], and Plutchik [23] have shown that an individual's facial description is generally not a single type of expression, but rather a mixture of various emotions. For instance, an individual might be both surprised and fearful. In this regard, it has been proven that the evaluation of facial expressions as a single-label classification issue can be very easy but deficient in effective application. According to Martinez et al.[24] the compound facial appearance is a mixture of two basic types of emotions. They came up with 15 compound expressions that were discovered in various cultures and said that human emotions can be integrated to generate new emotions. Moreover, the Facial Action Coding System (FACS) was used to illustrate the generation of these specific classifications; Hence, allowing the display of the Compound Facial Expressions of Emotion (CFEE) database. This database includes 5,060 facial images tagged with 7 basic emotions and 15 compound emotions for 230 subjects. The geometric and appearance features were mixed in a single feature space with a classifier to identify 7 basic categories and 22 classes (7 basic plus 15 compounds) of facial expressions. The authors said they achieved an accuracy of 96.96% while classifying 7 basic emotions and 76.91% for 22 classes of facial expressions.

Similarly, Aleix and Du [25] proposed a balanced model compatible with the recognition of compound facial expressions. This stable model describes how emotional expressions can be represented at different intensities. According to their research, classifications of compound emotions can also be identified by precisely blending specific continuous face spaces. Moreover, the process by which the generated model could be used for the identification of facial emotion expressions, prospective studies that need to address machine learning, and the importance of computer vision communities were illustrated. Besides, the RAFDB database (in the wild) was introduced by Li and Deng [26] where all images were tagged by 40 analysts. This database consists of 7 basic emotions and 12 compound emotions. The authors suggested a deep locality-preserving learning approach for emotion identification. They affirmed that the accuracy of the fundamental emotions classification was 84.13%. On the other hand, the accuracy of the compound emotion identification was 57.95%. This illustrates that compound emotion identification in the wild is a complex task.

Benitez-Quiroz et al. [27] introduced a method that easily explains Action Units (AUs), their intensities, and their specific emotion types for facial expression identification. This allowed the production of the EmotioNet dataset. Moreover, they have also picked out geometric and shading features. Geometric features are described as second-order demography of facial features (distance and angles between facial landmarks). On the other hand, the shading features are obtained using Gabor filters and copying the shading alterations as a result of the local distortions of some areas of the skin. This means that all the AUs are depicted with shape and geometric features. After all these steps, the Kernel Subclass Discriminant Analysis (KSDA) [28] was employed to decide whether a particular AU is active or not. Benitez-Quiroz et al. [27] got an AU annotation precision of about 81%. The 23 emotion types were described from various AU combinations.

B. Traditional Machine Learning Approaches

The conventional approaches for facial emotion identification to depend on procedures that rely primarily on how the features were obtained from the facial expression [29]. The obtained characteristics can be classified into procedures which are based on appearance features and geometric features.

Appearance features: These features outline the texture of the skin for instance crinkle and wrinkle. Some known procedures for facial expression analysis are Gabor Wavelet [30], Haar-like features [31], LBP, and its variants [19]. The recent LBP work [32] and its achievement in various computer vision issues have stimulated the advancement of a huge number of LBP variants. Due to its extensibility, the LBP descriptor can be quickly reformed to suit the needs of the different applications and unconventional texture issues. In [19], the authors examined the function of 46 state-of-the-art LBP-like descriptors and some of their current alternatives; they proposed an alternative study on the problem of identifying facial expressions using four benchmark datasets. Besides, facial expressions identification using LBP descriptor was depicted by Shan et al. [33]. It was proven that LBP features are more efficient than those of Gabor. On the other hand, the Local Ternary Pattern (LTP) [34] was employed in order to solve the weaknesses of LBP and to withstand non-uniform noise. Facial features were obtained by the active shape model by Myunghoon et al [35]. Regions that involve different facial information were used to get appearance-based features, and another traditional image processing procedure, which was applied on FER, is based on Local Binary Patterns (LBP) [33]. In fact, the human face can be devided into small areas for the calculation of the LBP histogram features. For any particular class, a sample is produced from the histogram of facial features while the closest classifier assembles a specific image.

Geometric features: Geometric features outline the shape of corners of the eyes, mouth, etc. Procedures that involve geometric features depend majorly on where the landmarks are suited like visual information, geometric displacement of the landmarks, or shape parameters of the facial component [36]-[37]. According to Ghimire and Lee [38], geometric features are obtained from the sequences of facial expression images while Ada-boost and SVM classifiers are employed for the grouping. A single frame classification of emotions using features based on geometric appearance and an SVM classifier has been suggested [39]. The aforementioned conventional methods have proven their potency on various facial expression datasets like CK+ [40], Oulu-CASIA [41], and FEED [42]. Within these datasets, the images representing the six fundamental emotions (i.e. happiness, surprise, fear, anger, disgust, sadness) were expressed in a "lab-controlled" manner and the CFEE dataset [24] was the one that dealt with compound facial expressions (happily surprised as an example). However, the mentioned datasets cannot be efficiently used to identify facial expressions. This may happen especially when the expressions are compound emotions or pictures that were taken under poor lighting, low resolution, etc. To overcome this issue, a lot of research has applied approaches based on deep learning. They have proven that these methods are more effective than conventional procedures.

C. Deep Learning based approaches

Deep learning algorithms have proven effective outcomes over the years and effective performance gain in different computer vision tasks like image selection, object identification, face identification, and facial emotion identification, etc [43], [44], [45], [46], [47]. Deep learning based neural networks were used as an "end-to-end" approach to obtain features from data. Neural networks for FER frequently use four various types of layers: convolution, max-pool, dense, and soft max layer. The training procedure can be simplified using batch normalization. The obtained features contain information concerning local spatial relation and global information. The max-pool layer helps in protecting the model from geometrical deformity while the dense and soft-max layers help in accrediting the class score. About that, Li et al [26]-[48] suggested a Deep Locality-Preserving CNN (DLP-CNN) in order to focus on the uncertainty and the multi-modality of the real-world facial expressions by including a recently directed layer on the fundamental structure of Deep Convolutional Neural Network (DCNN). They illustrated that the outlined DLP-CNN outperforms both the handcrafted features and deep learning based ones for the identification of facial expressions in the wild. To study discriminatory features for multi-label expressions, Li and Deng [49] proposed a Deep Bi-Manifold CNN (DBM-CNN). This was based on simultaneously securing the local affinity of deep features and the multiple structures of emotional descriptions. Deep Neural Networks with Relativity Learning (DNNRL) model was modified in [50] to identify emotions from face images. Accordingly, an accuracy of 70.6% was obtained on the FER-2013 dataset. On the other hand, Ng et al. [51] applied CNN with transfer learning from the ImageNet to identify emotions from motionless images and they achieved 55.6% accuracy.

Similarly, the authors in [52] suggested a method known as the Boosted Deep Belief Network (BDBN). Their method executed the three steps of learning (i.e. feature learning, feature selection, and classifier construction) constantly in a collective system. The procedure considered only the six primary emotions employing "lab-controlled" databases like CK+ and JAFFE [53]. Besides, an extreme learning machine (ELM) and an interlaced derivative patterns (IDP) suited emotion recognition system were proposed in [54]. The system reached an accuracy of 84.12% on the eNTERFACE database. In the same way, a deep neural network (DNN) based method to identify the emotion was introduced in [55]. The input of the DNN was the unprocessed facial image and an accuracy of 93.2% was obtained on the CK+ database.

A recent loss function known as the cluster loss was introduced in [56] to enact deep features compact. This method relied on a deep CNN model and obtained significant development when mixing data augmentation and cluster loss via the real-world six main emotions of the RAF-DB database. Similarly, Zeng et al. [57] introduced a histogram of oriented gradients (HoG) characteristics and a deep sparse autoencoder-based emotion recognition system. They obtained about 96% accuracy using the extended CK database (CK+). Besides, a deep network correlating many deep models, called FaceNet2ExpNet, has been proposed in [58]. Experiments on the CK+ database leaded to 96.8% of accuracy. Egede et al.[59] proposed to encrypt shape and appearance information in both types of handcrafted and deep learning feature extraction. For handcrafted features, HOG and some metrics were derived from 49 facial landmarks to depict image features and geometric features respectively. Depending on the CNN, features were studied from a blend of pixels from the prototype image (appearance) and binary masks (face shape). Besides, Yong et al. [60] introduced a Patch-Gated Convolution Neural Network (PG-CNN) which inevitably determined the occluded area of the face and focused on the most demanding un-occluded areas. PG-CNN formulates an intermediary feature map into numerous patches based on the status of joint facial landmarks to describe the likely facial region of interest. This procedure was performed on two different databases, RAF-DB and AffectNet. [61].

III. Methodology

A. Databases

The facial expression recognition experiments are performed using two various databases:

Compound Facial Expressions of Emotions (CFEE) [24] is a posed database of both standard and non-commercial facial expression images. In this database, a lot of nationalities and ethnic groups are represented. It has 5,060 images since it is composed of 1610 basic emotions and 3450 compound ones expressed by 230 individuals (130 females and 100 males with a mean age of 23). The types of emotions shown in this database can be divided into three groups (see Figure 1):

- The elementary group, which encompasses the six primary emotions, happiness, sadness, fear, anger, surprise, and disgust as well as neutral facial expressions.
- The second group deals with 12 multi-label compound emotions that are generally expressed by human beings: angrily disgusted, angrily surprised, disgustedly surprised, fearfully angry, fearfully disgusted, fearfully surprised, happily disgusted, happily surprised, sadly angry, sadly disgusted, sadly surprised, sadly fearful.
- The third group depicts a set of three single-label compound emotions: appalled, awed, hatred. Appall can be defined as a mixture of anger and disgust with an emphasis on disgust. Besides, hate deals also with emotions of anger and disgust; however, the emphasis is on anger. On the other hand, awe is made up of both the emotions of surprise and fear with an emphasis on surprise.

Real-world Affective Faces (RAF) [48] is a spontaneous huge facial expression database with about 30000 largely diversified facial illustrations transferred from the internet. The images contained in this database are of great variability in terms of race, gender, head postures, age, occlusions, lighting parameters, and post-processing activities like different filters and special effects. RAF-DB has great diversities, huge

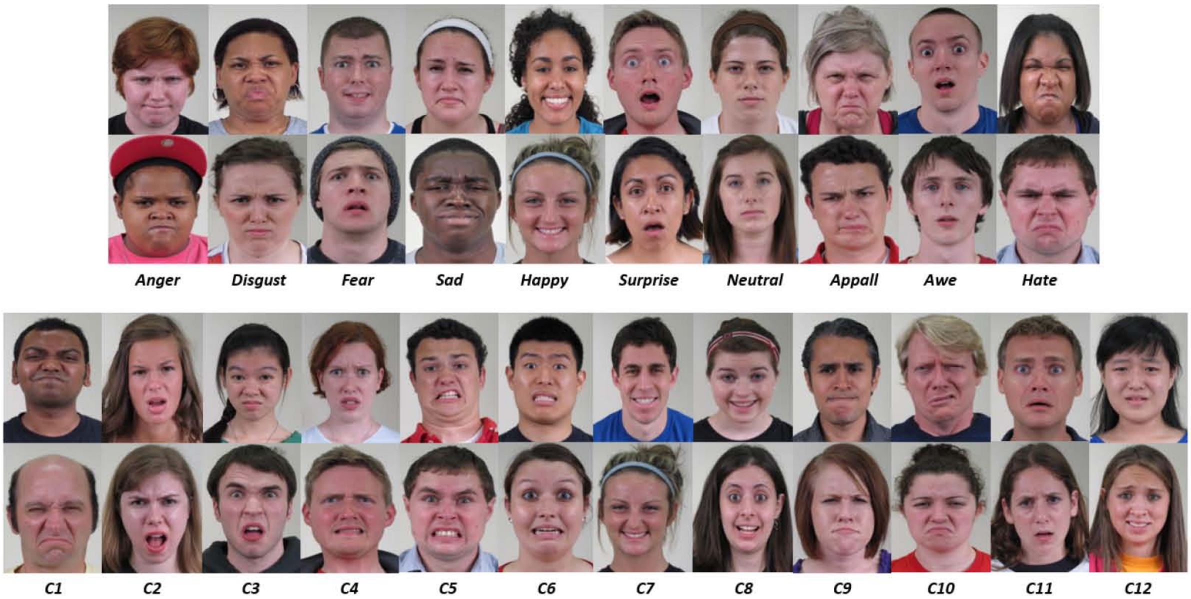

Fig. 1: Example of the CFEE basic and compound emotions. Note: **C1:** Angrily-Disgusted, **C2:** Angrily-Surprised, **C3:** Disgustedly-Surprised, **C4:** Fearfully-Angry, **C5:** Fearfully-Disgusted, **C6:** Fearfully-Surprised, **C7:** Happily-Disgusted, **C8:** Happily-Surprised, **C9:** Sadly-Angry, **C10:** Sadly-Disgusted, **C11:** Sadly-Surprised, **C12:** Sadly-Fearful

Fig. 2: Example of the RAF basic and compound emotions.

varieties, and abundant interpretations.
The RAF-DB is well known for its two subsets: basic emotions including anger, disgust, fear, happiness, sadness, surprise, neutral, and compound emotions: (C1) Angrily-Disgusted, (C2) Angrily-Surprised, (C3) Disgustedly-Surprised, (C4) Fearfully-Angry, (C5) Fearfully-Surprised, (C6) Happily-Disgusted, (C7) Happily-Surprised, (C8) Sadly-Angry, (C9) Sadly-Disgusted, (C10) Sadly-Fearful, (C11) Sadly-Surprised. (see Figure 2).

In our experiments, two datasets of CFEE are evaluated. This involves seven basic categories and the entire CFEE database, and four distinct RAF datasets have experimented; Six categories of primary emotions plus neutral expressions, 11 compound emotions' classes, 16 basic and compound emotions, and finally the 18 categories of the entire classes.

B. Used methods

Facial Emotion Recognition consists of identifying one or more faces by matching input images with database images. Therefore, the algorithms used to solve the problem of face recognition must provide sufficient recognition accuracy. The Viola-Jones algorithm is one of the most promising methods in terms of high performance, low number of false positives, and high percentage of accurately detected faces. In our experiments, the implementation of the Viola- Jones detector available in the OpenCV 2.4.8 library is used to detect the position of the face in the CFEE database images. However, for the RAF database, all detected faces are already available. Additionally, the input data has different sizes. They need to be normalized to a fixed size before being used. Thus, all CFEE and RAF datasets images are resized to 224x224. Besides, feature extraction remains one of the most preliminary steps in Machine Learning algorithms for identifying strong and weak relevant attributes. Although many feature extraction algorithms are used to extract features from basic emotions, the problem becomes more difficult for compound facial expressions classification, where each image is a combination of basic emotion categories. Hence, to build a deep feature extraction model, the Pre-trained CNN Resnet-18 architecture [62], which is composed of many ResNet blocks (a combination of convolution and identity block), is applied.

The learning rate is a hyper-parameter that controls how much the network weights are adjusted. Also, it permits to calculation of the error and updates the model for each training batch of the dataset. Using a small training rate ensures that the weights do not pass through the optimal possible solution. Hence, to optimize the Resnet-18 model, the Stochastic Gradient Descent (SGD) method with Momentum and the adaptive learning rate is used. The batch size is set to be 8, 32, or 64. Additionally, the Momentum value is set to 0.9 for all experiments. Furthermore, the Resnet-18 models are retrained around 60 iterations for the basic CFEE database and 80 iterations for the basic RAF and compound CFEE and RAF datasets. To classify basic and compound facial expressions from posed and spontaneous databases, the SVM algorithm with Radial Basis Function is used. Besides, for results comparisons, different SVM kernels

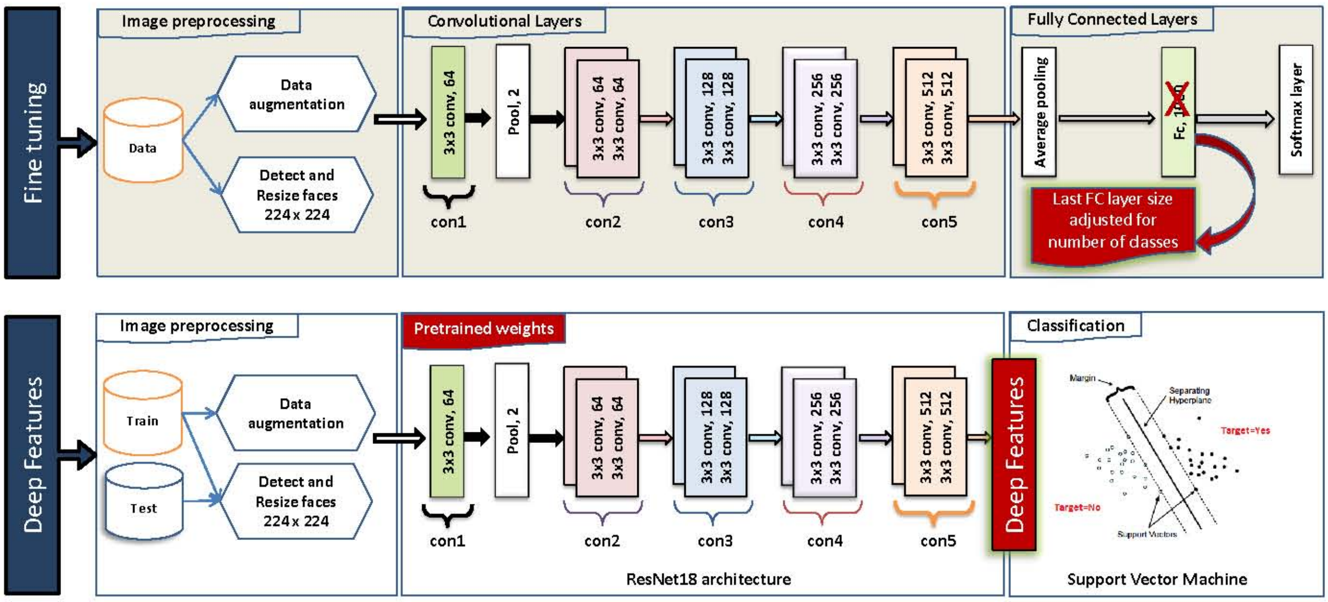

Fig. 3: Deep-feature based Facial Expression Recognition approach.

including Linear, Polynomial, and Sigmoid are trained, as well as, Random Forest and Logistic Regression algorithms.

C. Proposed approach

It is simple for complicated models such as CNNs to exceed the tiny dataset size. Hence, to deal with overfitting and a lack of database images, the suggested study reports a transfer learning approach using the pre-trained network architecture ResNet18 to find low-level pixels (like edges and textures), and SVM classifier to find out higher-level details (like eyes, mouth, etc). The model's weights are refined on the new dataset.

Overall, our adopted approach mainly consists of three steps. First, all datasets' images are preprocessed by cropping and scaling the face region to obtain a unified dimension. Subsequently, the preprocessed images are fed into the CNN Resnet-18 model to extract features of training and testing images. Finally, the Support Vector Machine is used to classify the basic and compound emotions of posed and spontaneous databases. The architecture of the proposed hybrid approach is shown in Figure 3.

The key contribution of this work is to provide a CNN-based transfer-learning approach using a pre-trained model to detect and classify basic and compound facial expressions with greater precision compared to recent works.

The underlying reasons for this choice are:

- The deep learning model involves a huge database to train it from scratch. Thus, using pre-trained weights helps to solve the issue of small databases and avoid overfitting [63], [64], [65], [66]
- Gain from accessible Deep Learning model that has attained great accuracy in identifying tasks that use images as input (i.e. ImageNet) [67].
- Networks with a large number of layers are expensive and time-consuming to train. The most complex models take weeks to train using hundreds of expensive GPUs [68].

IV. Results and Discussion

This section provides a discussion about the performance analysis of eight experimental results on CFEE and RAF databases. The whole CFEE benchmark database images are used. This includes the seven classes (i.e., six basic emotions plus neutral expressions) and the 22 categories of basic and compound emotions for the 230 identities. Moreover, the whole images of the RAF database are evaluated. This includes seven basic emotions from the single-label subset and eleven compound emotions. Moreover, sixteen and eighteen categories match spontaneous basic and compound emotions. All experiments are done using the GPU of Google collab.

A. Experimental Setup

Evaluation Procedure: To overcome the overfitting problem two well-known strategies are used; the train/test split and cross-validation.

In this work, the train/test division is used to evaluate the model of five experiments. Whereas 10-fold cross-validation is used to evaluate the deep features Resnet-18 model from three experiments on the CFEE dataset.

Assessment Metrics: Before choosing the evaluation metrics, it is necessary to highlight what type of problem we are solving: Classification, Regression, or Ranging. In our experiments, several metrics are used to evaluate the results for the classification issue. For the CFEE dataset, the confusion matrix and accuracy are used to compare the approach results with state-of-the-art. Whereas, for the RAF dataset, sensitivity is used as the primary assessment measure, due to the imbalanced distribution within each class. Overall, emotions in the real world are usually out of balance.

B. Experimental results of posed CFEE database

1) Performance Analysis of Basic Emotions: In the first experiment, 10-fold cross-validation is conducted to classify all the basic emotions of the CFEE dataset. The Resnet-18 model is fine-tuned using a subset of 980 images of the KDEF dataset [69]. The view angle 0° is the only one used, and images are picked from both sessions for the 70 subjects. Each image expresses two times the six-class of basic emotions plus neutral expressions; anger, disgust, fear, happiness, sadness, and surprise. Afterward, the adapted model is used to extract features of the basic CFEE dataset corresponding to 1,610 images and classify them using SVM-RBF. The best successful classification rate (average accuracy) achieved is **98.02%**.

For classification comparison, the SVM-Linear, SVM-Sigmoid, SVM-Poly, Random Forest (RF), and Logistic Regression (LR) are trained. Table I illustrates the different results of six classifiers using four metrics; accuracy, recall, precision, and F1-score. For each classifier, the 10-fold cross-validation test is conducted. The successful classification rates are 97.95%, 97.27%, 97.02%, and 94.22% achieved by Logistic Regression, SVM-Linear, SVM-Poly, and Random Forest respectively. Whereas, we can notice that the basic emotions classification using SVM with Sigmoid kernel does not perform well, as the lowest test accuracy obtained is 81.49%, the recall is 81.43%, the precision is 82.57%, and the F1-score is 81.86%.

TABLE I: Results of six-classifiers using KDEF to fine-tune Resnet-18 (k-fold).

Metrics	**SVM**				**RF**	**LR**
	RBF	**Linear**	**Sigmoid**	**Poly**		
Accuracy	98,02	97,27	81,49	97,02	94,22	97,95
Recall	98	97,29	81,43	97	94,57	98.14
Precision	98	97,43	82,57	97.29	94,57	97.71
F1-score	98	97,29	81,86	97	94,71	97.71

As observed in Table II, the average mean diagonal value of the confusion matrix is 98%. The happy class was perfectly-recognized (100%), followed by surprise, neutral, and fear categories with an accuracy of 99%. Meanwhile, the disgusting category reached a 97% accuracy, while anger and sadness obtained an accuracy of 96%.

The KDEF dataset is used to fine-tune the model's weights, Hence the Resnet-18 model shows its effectiveness by providing good results on basic emotions of the CFEE dataset.

TABLE II: Confusion matrix of the seven emotions using KDEF to fine-tune Resnet-18.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	**0.96**	0.01	0	0	0.01	0.02	0
Disgust	0.02	**0.97**	0	0	0	0	0
Fear	0	0	**0.99**	0	0	0	0
Happiness	0	0	0	**1**	0	0	0
Neutral	0	0	0	0	**0.99**	0.01	0
Sadness	0.03	0	0	0	0.01	**0.96**	0
Surprise	0	0	0.01	0	0	0	**0.99**

The second experiment is conducted on the six basic emotions plus the neutral expression. The whole CFEE basic emotions corresponding to 1610 are used. First, they are divided into 10 folders. Ten times the test is applied and at each time, nine folders containing a total of 1449 images are collected to fine-tune the Resnet-18 model's weights and train the SVM classifier and one different folder of 161 images is used for each testing phase. Therefore, 10 Resnet-18 models are constructed, they are used to extract features of training and testing datasets. The accuracy average of 10 splits is taken as a metric performance. Hence, the highest classification rate is 99.19% obtained by SVM-RBF. Table III shows the obtained confusion. It is clear from the results that the 7-class achieved the best accuracies. Four classes are the easiest to identify with an accuracy of 100%. This concerns neutral, happiness, surprise, and disgust categories. Moreover, fear, anger, and sad emotions are highly recognized with an accuracy of 99%, 98%, and 98% respectively.

TABLE III: Confusion matrix of the seven emotions using CFEE to fine-tune Resnet-18.

	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
Anger	**0.98**	0.01	0	0	0.01	0.01	0
Disgust	0	**1**	0	0	0	0	0
Fear	0	0	**0.99**	0	0	0	0
Happiness	0	0	0	**1**	0	0	0
Neutral	0	0	0	0	**1**	0	0
Sadness	0.01	0.01	0	0	0	**0.98**	0
Surprise	0	0	0	0	0	0	**1**

To obtain the best emotion classification, we compared the SVM-RBF result with SVM- Poly, SVM-Sigmoid, and SVM-Linear, in addition to Random Forest, and Logistic Regression classifiers. The results can be seen in Table IV. On one hand, the highest classification rates of 98.95%, 98.39%, 97.70%, and 93.91% are obtained by Logistic Regression, SVM-Poly, SVM-Linear, and Random Forest, respectively. On the other hand, the highest misclassification rate is achieved by SVM-Sigmoid (72.17%).

The results reported in this experiment are increased by 1.17% compared to the first. This is due to the use of facial expressions of the same subjects to refine the Resnet-18 model weights and train the SVM. However, for comparing the performance of our approach with [70], an experiment based on the train/test split is conducted; only seven basic facial expressions of the CFEE database are used. Therefore, a split of 70%, corresponding to 1,127 images, is applied to adapt the Resnet-18 model and train the SVM classifier and two splits of 15%, corresponding to 238 and 245 images, are selected to test the model. Using this evaluation procedure, the classification of CFEE Basic emotions has also been examined using SVM-Linear, SVM-Sigmoid, SVM-Poly, Random Forest, and Logistic Regression classifiers (see Table V). For the first split, both SVM-RBF and SVM-Linear have achieved the same highest accuracies of **81.93%** (238 images). While, for the second split, SVM-RBF achieved an accuracy of **79.18%** (245 images).

TABLE IV: Results of six-classifiers using CFEE to fine-tune Resnet-18 (k-fold).

Classifiers	SVM				RF	LR
	RBF	Linear	Sigmoid	Poly		
Split 1	**98.76**	98.14	91.30	**98.76**	95.65	**98.76**
Split 2	**100**	99.38	65.22	99.38	98.14	**100**
Split 3	**100**	**100**	96.89	**100**	98.14	**100**
Split 4	**98.76**	**98.76**	44.72	97.52	95.65	97.52
Split 5	**98.14**	95.65	50.93	96.27	89.44	**98.14**
Split 6	**98.76**	95.65	62.73	96.89	89.44	98.14
Split 7	**99.38**	96.27	73.91	**99.38**	93.17	98.14
Split 8	**100**	95.65	78.88	97.52	90.68	99.38
Split 9	**100**	**100**	98.76	**100**	98.14	**100**
Split 10	98.14	97.52	58.39	98.14	90.68	**99.38**
Average	**99.19%**	97.70%	72.17%	98.39%	93.91%	98.95%

Note: Bold accuracy concerns the best-recognized classifier.

TABLE V: Results of six-classifiers using CFEE to fine-tune Resnet-18 (train/test split).

Classifiers	SVM				RF	LR
	RBF	Linear	Sigmoid	Poly		
Split 1	**81,93%**	**81,93%**	79,41%	81,51%	76,05%	80,25%
Split 2	**79,18%**	77,55%	73,88%	75,10%	74,29%	78,37%

Note: Bold accuracies concern the best-recognized classifiers.

2) Performance Analysis of Compound Emotions: Twenty-two classes of the CFEE database are evaluated including 6 basic emotions plus neutral expressions and 15 compound emotions. Similar to the second experiment, the whole CFEE database that includes 5,060 images is divided into 10 folders. While nine folders are collected to adjust the Resnet-18 model weights and train the SVM algorithm, one folder is selected for the testing phase. Afterward, this model is used to extract features of 4,554 training images and 506 testing images. Hence, 10 Resnet-18 models are constructed, and each time, a different test folder is chosen.

The average of 10 test classification accuracies is used as a metric performance. The best average classification accuracy achieved is **80.69%** using SVM with the Radial Basis Function (RBF). Whereas, the highest accuracy split is **89.13%**.

Figure 4 illustrates the precision-recall curve, where the high scores for both metrics, show that the classifier is returning a majority of all positive results and a low false-negative rate (high recall), as well as returning accurate results (high precision). Furthermore, the average accuracies of 79.56%, 77.81%, 77,04%, 70,08% are obtained by comparing the Logistic Regression, SVM-Poly, SVM-Linear, and Random Forest classifiers, respectively. Whereas, the lowest average accuracy is reached by SVM-Sigmoid (18,05%) (see Table VI).

As illustrated in Figure 5, the confusion matrix gathers the obtained experimental results of the whole CFEE database. It can be observed that the three categories of basic emotions Neutral, Happiness, and Surprise are the easiest to identify with an accuracy between 96% and 98%. Whereas the highest accuracies of 98% and 99% are reached by two compound emotions concerning Happily Surprised and Happily Disgusted, while the highest misclassification rates are obtained by four categories. Mainly, 41% reached by Appall which is confused with Angrily Disgusted, Hate, and Disgust.

TABLE VI: Results of six classifiers by each split using 22 emotion categories.

Classifiers	SVM-RBF	SVM-Linear	SVM-Sigmoid	SVM-Poly	RF	LR
Split 1	**75.69**	66,01	07,94	66,60	65,81	71,74
Split 2	75,30	75,49	48,62	78,06	67,39	**75,69**
Split 3	84,19	81,82	61,46	**84,58**	74,30	84,19
Split 4	**85,18**	80,24	09,68	81,23	70,95	83,99
Split 5	**89.13**	85,38	07,11	86,76	77,27	87,15
Split 6	**85,97**	82,81	19,96	85,97	77,47	82,41
Split 7	81,23	80,83	04,35	79,05	73,32	**84,58**
Split 8	**78,85**	73,52	07,91	74,51	61,66	76,68
Split 9	**74,51**	71,74	04,74	68,38	67,79	72,92
Split 10	**76,88**	72,53	08,70	72,92	64,43	76,68
Average	**80,69%**	77,04%	18,05%	77,81%	70,08%	79,56%

Note: Bold accuracies concern the best-recognized classifiers.

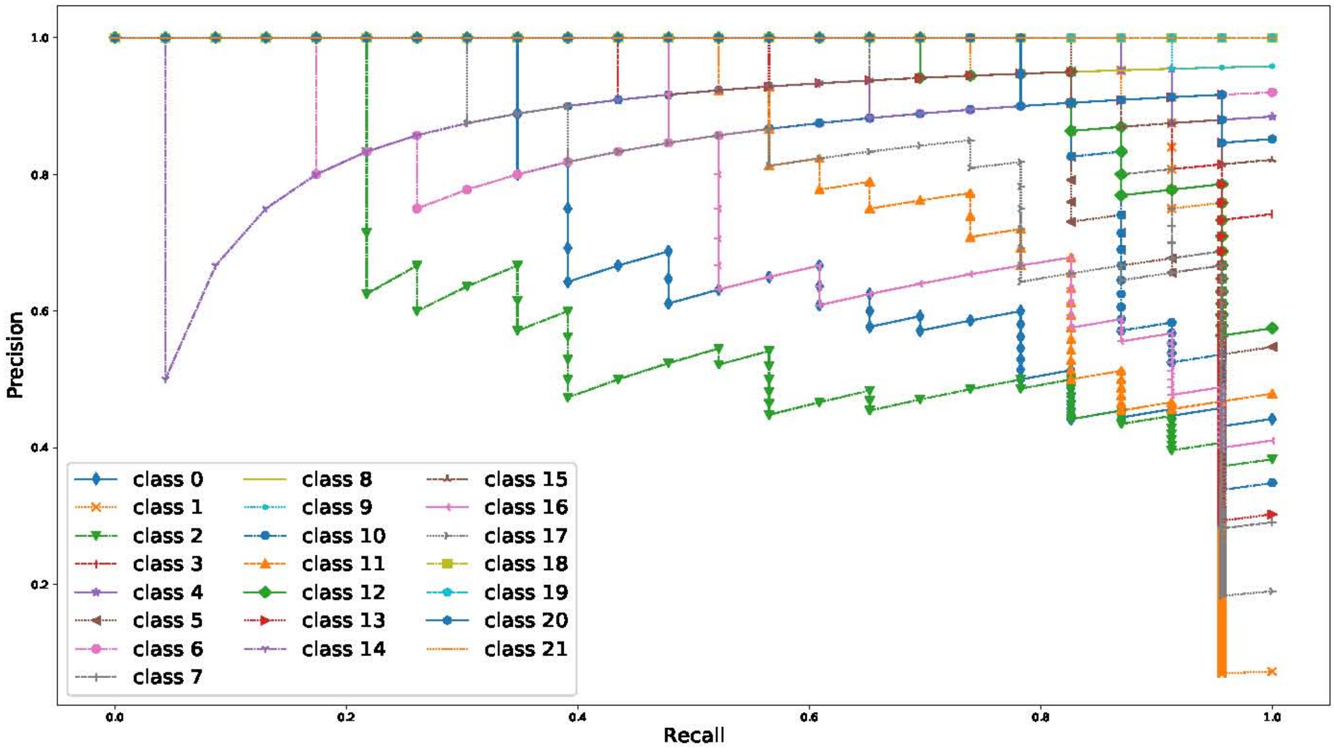

Fig. 4: Precision-Recall curves of 22-class CFEE using SVM-RBF.

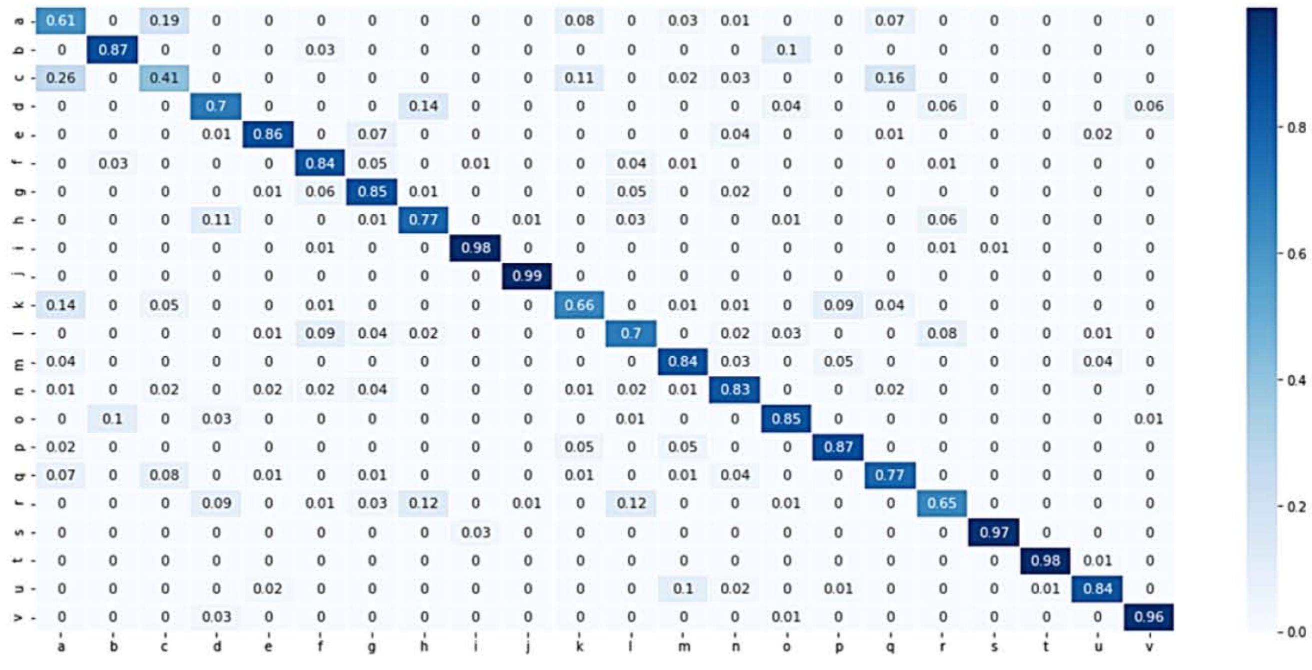

Fig. 5: Confusion matrix for the 22 categories of CFEE database. Note: **a**, angrily disgusted; **b**, angrily surprised; **c**, appalled; **d**, awed; **e**, disgustedly surprised; **f**, fearfully angry; **g**, fearfully disgusted; **h**, fearfully surprised; **i**, happily disgusted; **j**, happily surprised; **k**, hate; **l**, sadly fearful; **m**, sadly angry; **n**, sadly disgusted; **o**, sadly surprised; **p**, angry; **q**, disgusted; **r**, fearful; **s**, happy; **t**, neutral; **u**, sad; **v**, surprised. Rows, true category. Columns, recognized category.

This confusion is because of the similarity of the composition so that Appalled and Hate represent the combination of Disgust and Anger categories. The difference is just on the emphasis. This is similar to the Angrily Disgusted category that obtained an accuracy of 61%. For this category, the confusion was with Appall, Hate, and Disgusted. For the classification of the Angrily surprised category, an accuracy of 87% is achieved. The confusion was within Sadly Surprised. While the Hate category attained an accuracy of 66%. This category involves the combination of Disgust and Anger with an emphasis on Anger. Hence, a clear confusion was between the Angrily Disgusted and Angry. Furthermore, the average accuracy when using the mean diagonal value of the confusion matrix as the metric is **80.91%**.

The proposed approach also ensures the effectiveness of merging the basic and compound emotion categories. Then, a good performance is observed with the expression features extracted using Resnet-18. However, the extraction of 22 categories of basic and compound emotional features shows less improvement to the results as compared to the expression features extracted on 7 basic emotions only. The reason for these results is the confusion of the compound emotions with basic ones that compose them, as well as, the multilabel compound emotions, with single label ones, which are based on intensities and are the most difficult emotions to recognize.

C. Experimental Results of Spontaneous RAF database

1) Performance Analysis of Basic Emotions: The seven categories from the single-label subset of the RAF database are evaluated using 15,339 images. Likewise, the idea of 5-fold cross-validation is conducted where the training set was five times larger than the test set. The Resnet-18 model is adjusted using 12,271 images. Then, the built model is used to extract features of 12,271 training images and 3,068 testing images. As shown in Table VII, the accuracy, recall, precision, and F1-score metrics are calculated using six classifiers; the SVM-RBF, SVM-Linear, SVM-Sigmoid, SVM-Poly, Random Forest, and Logistic Regression. Hence, the best classification accuracy of **93.29%** is achieved by SVM with RBF kernel. As soon as the recall (mean diagonal value of the confusion matrix) is used as the metric, the classification rate decreases from 93.29% to 86%. Furthermore, since the RAF dataset is for multi-class classification, the SVM with sigmoid kernel performance was not as good as other kernels for all evaluation metrics. However, it is better suited for binary classifications. Figure 7 shows the precision-recall curves for SVM-RBF. The positive predictive values of the seven classes of the RAF database (PPV, y-axis) are plotted against the true positive rate (TPR, x-axis). Whereas the true negatives are not considered.

As shown in confusion matrix VIII, the happiness category was the easiest to identify with an accuracy of 98%, followed by neutral, sadness, and surprise with an accuracy rate of respectively 94%, 93%, and 92%. Also, accuracies of 86% and 73% are obtained for the angry and disgust categories, respectively. Moreover, considering that Fear was a challenging class, it was the hardest to recognize. Then since it experienced the highest misclassification rate (66%), it was confused with surprise.

TABLE VII: Results of the 7 basic emotions of RAF database using six classifiers.

Metrics	**SVM**				**RF**	**LR**
	RBF	**Linear**	**Sigmoid**	**Poly**		
Accuracy	93,29	90,42	54,60	91,36	85,85	92,86
Recall	86	83,57	33,14	84,29	72,14	84,14
Precision	88.57	82.71	50.14	84.57	76.86	89.14
F1-score	87.29	83,14	33,86	84,57	74	86.29

TABLE VIII: Confusion matrix of the seven single-label of RAF database.

	Anger	Disgust	Fear	Happy	Neutral	Sad	Surprise
Anger	**0.86**	0.06	0.01	0.01	0.04	0.01	0.02
Disgust	0.02	**0.73**	0.01	0.04	0.09	0.08	0.02
Fear	0.01	0.03	**0.66**	0.04	0.04	0.05	0.16
Happy	0	0	0	**0.98**	0.01	0	0
Neutral	0	0.02	0	0	**0.94**	0.03	0.01
Sad	0.01	0.02	0	0.01	0.03	**0.93**	0
Surprise	0.01	0.01	0.02	0	0.02	0.01	**0.92**

Note: Bold accuracies concern the best-recognized categories. Columns, predicted category. Rows, true category.

2) Performance Analysis of Compound Emotions: To evaluate the eleven compound emotion classes of the RAF database corresponding to 3,954 images, the Resnet-18 model's weights are adjusted using 3,162 RAF compound images. Then, the model is used to extract features of 3,162 training images and 792 testing images. Our evaluation metrics are recall (the mean diagonal values of confusion matrix), accuracy, precision, and F1-score. Table IX illustrates the best-obtained results. This corresponds to an accuracy of **76.52%** achieved by SVM with RBF kernel. Once the mean diagonal value of the confusion matrix (recall) is considered as the metric, the accuracy decreased from 76.52% to 62% using SVM-RBF and 75.38% to 63.27% using SVM-Linear. This involves also the comparison with other classifiers; SVM-Poly, SVM-Sigmoid, Random Forest, and Logistic Regression.

The confusion matrix in Figure 6 shows the recognition performance according to each compound emotion. Through this evaluation metric, the easiest and most difficult emotions to recognize are described, as well as the recognition rate of each compound emotion is analyzed. The happily surprised was perfectly recognized (91%), followed by fearfully surprised (85%) and angrily disgusted (83%), whereas disgustedly surprised, sadly fearful, and sadly surprised experienced the highest misclassification rate (37%, 32%, and 44%, respectively). Disgustedly surprised was confused with happily surprised (17%), and sadly fearful was confused with three compound emotions; fearfully surprised (23%), sadly disgusted (18%), and happily surprised (11%). Finally, the sadly surprised was confused with four classes. That involves sadly disgusted (17%), angrily disgusted (11%), fearfully surprised (11%), and sadly fearful (11%).
A Receiver Operating Characteristic curve (ROC curve) presented in Figure 8 summarizes the performance of our classification model on the positive class (SVM-RBF). Which is competitive for a great number of state-of-the-art used in our experiments. The x-axis indicates the False Positive Rate and the y-axis indicates the True Positive Rate.
The whole classes of the RAF database (eighteen categories) are evaluated including 7 basic emotions and 11 compound emotions. The same idea of 5-fold cross-validation is used. 15,433 images are selected to fine-tune the Resnet-18 model's weights. Afterward, the Resnet-18 model is used to extract features of 15,433 images that are used to train SVM and 3,860 test images. Table X shows the results of two

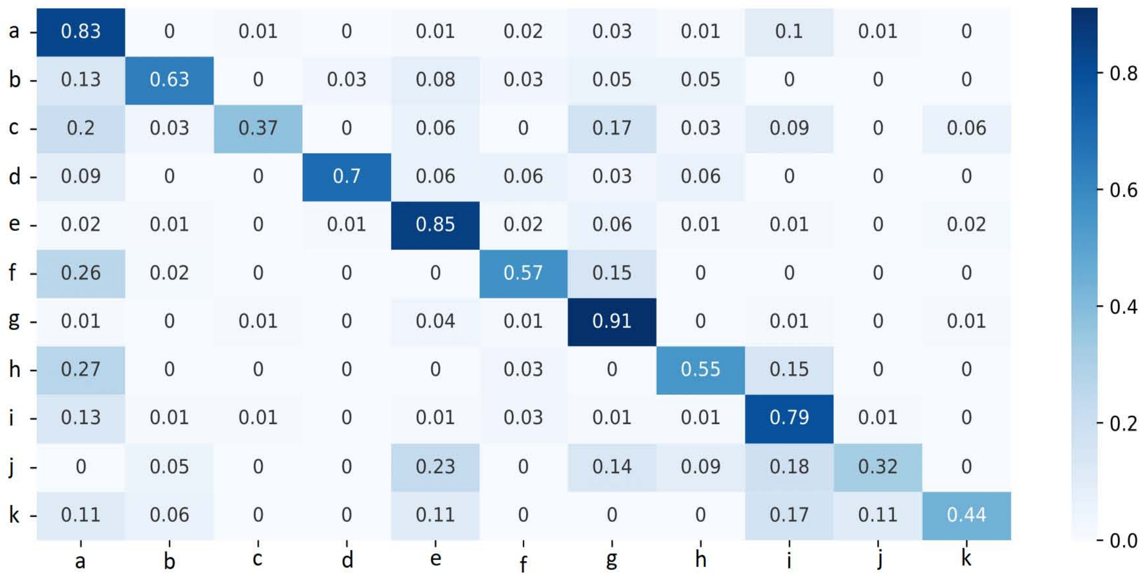

Fig. 6: Confusion matrix of 11 compound emotions of RAF database. Note: **a**, angrily disgusted; **b**, angrily surprised; **c**, disgustedly surprised; **d**, fearfully angry; **e**, fearfully surprised; **f**, happily disgusted; **g**, happily surprised; **h**, sadly angry; **i**, sadly disgusted; **j**, sadly fearful; **k**, sadly surprised.

evaluation metrics (accuracy and recall) using six classifiers; SVM (RBF, Linear, Sigmoid, and Polynomial), Random Forest, and Logistic Regression. The best-achieved accuracy is **64.15%** using SVM-RBF, while the recall decreased to 33.78%. There is a decrease in the accuracy metric when the basic and compound emotions are combined in which the difference of 12.37% decreased from the accuracy of 11 compound emotions and 29.14% of basic emotions.

TABLE IX: Results of the 11 compound emotions of RAF database using six classifiers.

Metrics	SVM				RF	LR
	RBF	Linear	Sigmoid	Poly		
Accuracy	76,52	75,38	73,36	75,88	63,51	73,48
Recall	62	63,27	57,27	61,45	46,27	59,36
Precision	79.27	73.45	79.18	81.36	61	71.36
F1-score	67	66.82	62.36	67.18	48.91	62.09

TABLE X: Results of the 18 categories of RAF database using six classifiers.

Metrics	SVM				RF	LR
	RBF	Linear	Sigmoid	Poly		
Accuracy	**64,15**	59,95	36,22	61,76	58,63	62,75
Recall	33.78	30.39	14.28	32.06	25.94	29.11

In order to compare the performance of our approach with the literature, sixteen categories of emotional expressions (9 compound classes and 7 basic classes) are evaluated with the same configuration and procedure in [71]. The dataset is randomly divided in half on the category level, 9,534 images for training, and 9,525 images for testing. Table XI illustrates the results of 10 splits, and for a comparison purpose, several classifiers are tested including SVM [RBF, Linear, Sigmoid, Poly], Random Forest, and Logistic Regression. For each classifier, 10 times test is carried out by a Random half-half division of the RAF database into train-set and test-set.

TABLE XI: Accuracy results concerning the 16-class of the RAF database.

Classifiers	SVM				RF	LR
	RBF	Linear	Sigmoid	Poly		
Split 1	61,27	56,36	36,79	59,71	56,08	60,42
Split 2	60,06	54,76	35,69	58,01	54,17	58,61
Split 3	60,41	55,78	36,16	58,82	55,36	59,39
Split 4	60,56	55,13	46,92	58,04	55	59,29
Split 5	60,63	55,57	45,01	58,91	55,78	59,67
Split 6	61,05	56,29	37,35	59,23	55,30	59,31
Split 7	60,93	55,27	42,25	58,73	55,19	59,27
Split 8	60,87	55,39	42,28	58,76	55,60	59,30
Split 9	61,39	57,18	41,07	59,69	56,15	60,33
Split 10	60,43	55,71	44,40	58,86	55,48	59,59
Average	**60,76**	55,74	40,79	**58,88**	55,41	**59,52**

The average accuracy is used as a result of the classifier performance; the highest classification rate is 60.76% using SVM-RBF, followed by an average accuracy of 59.52% attained by the Logistic Regression classifier. Whereas, the lowest classification recognition is 40.79% achieved by SVM with the sigmoid kernel.

The results outperform those presented in the state-of-the-art, while in [71] they obtained an accuracy of 44.10%. Hence, our adopted approach increases accuracy by 16.66%.

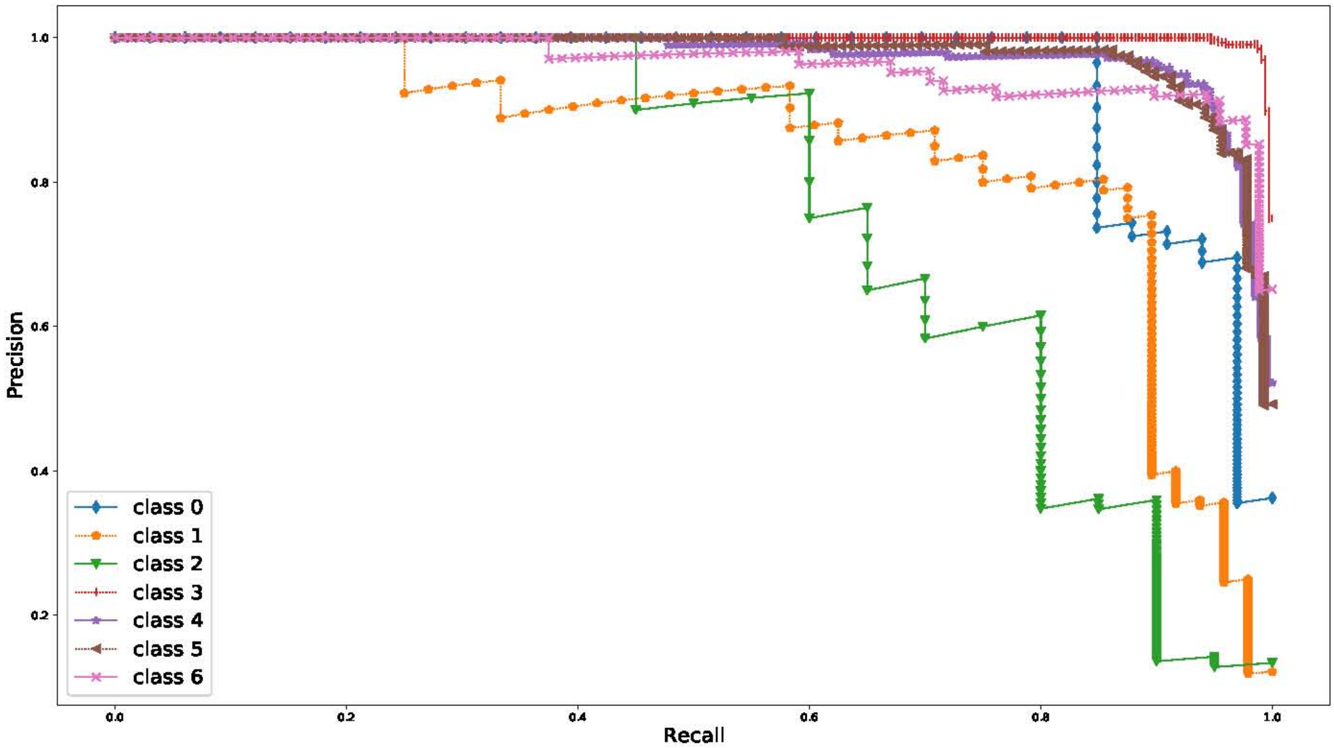

Fig. 7: Precision-Recall curves of 7-class RAF-DB using SVM-RBF.

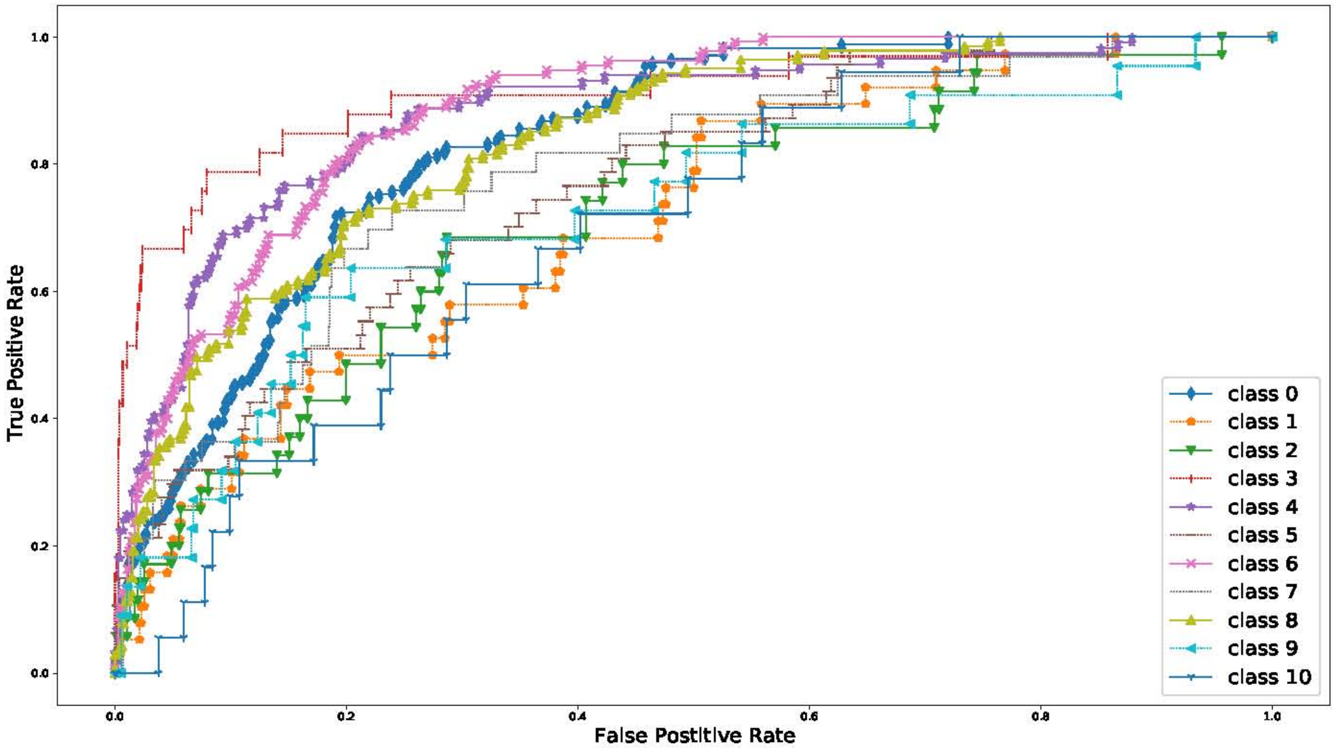

Fig. 8: ROC Curves of 11-class RAF-DB using SVM-RBF.

V. Comparison with state-of-the-art methods

This section deals with the effectiveness of our system on the basic and compound emotions of posed CFEE and spontaneous RAF databases and comparison with the existed advanced procedures. The obtained results from the reviewed state-of-the-art papers and the reached recognition rates by the best performing evaluated approaches are illustrated in Tables XII, and XIII.

The following outcomes are drawn out:

Concerning the basic emotions (7 classes) of the CFEE database, the obtained results are compared with two reviewed state-of-the-art studies (see Table XII). The test accuracy of 74.79% was achieved by [70] using AlexNet CNN architecture. They used 70% of the images for train-set, 15% for validation-set, and 15% for test-set.

TABLE XII: Comparison with state-of-the-art methods for basic and compound emotions of CFEE database.

Ref	Year	Method	Procedure	Samples	Classes	Accuracy
[24]	2014	Shape and appearance + Nearest mean	10-fold cross-validation	1610	7-class	96,96%
[70]	2017	AlexNet CNN	70% -15% -15% (train/validation/test)	1127 245 238		74,79 %
[24]	2014	Shape and appearance + Nearest mean	10-fold cross-validation	5060	22-class	76,91%
[72]	2019	Highway-CNN				52.14 %
Our results		**Resnet-18 Deep Features + SVM**	10-fold cross-validation	1610	7-class	**98,02%*** **99.19%+**
			70% train -15% test	1365		**81,93%**
			10-fold cross-validation	5060	22-class	**80,69%**

*Using KDEF dataset to fine-tune Resnet-18 model.
+Using CFEE dataset to fine-tune Resnet-18 model.

TABLE XIII: Comparison with state-of-the-art methods for RAF database.

Ref	Year	Method	Samples	Classes	Recall	Accuracy
[48]	2018	Gabor + mSVM	15339	7-class	65,12%	_
		Deep Locality-Preserving CNN + mSVM			74,20%	84,13%
[56]	2018	Augmented data and cluster loss			75.73%	_
[73]	2018	Multi-Region Ensemble -CNN (VGG-16)			76,73%	_
		Multi-Region Ensemble -CNN (AlexNet)			74,78%	_
[74]	2018	Capsule Network			77,48%	_
[75]	2018	Double Completed-LBP (Double Cd-LBP)			78,60%	_
[76]	2018	Transfer learning Resnet-18 (AffectNet database)			80%	_
[77]	2018	Covariance pooling after final convolutional layers			79,43%	87%
[60]	2018	Patch-Gated CNN (PG-CNN)			_	83,27%
[78]	2019	Conditional generative adversarial network-based EAU-Net Network (cGAN based EAU-Net)			81,83%	_
[79]	2020	Region Attention Network (RAN)			_	86,90%
[80]	2020	Pyramid With Super-Resolution (PSR) Network (VGG-16)			80.78%	88,98%
[26] [48]	2018	Gabor + mSVM	3954	11-class	33,76%	_
		Deep Locality-Preserving CNN + mSVM			44,55%	57.95%
[71]	2016	LBP + NCMML	3171	16-class	30.10%	36.70%
		HOG + NCMML			36.90%	44.10%
Our results		**Resnet-18 Deep Features + SVM**	15339	7-class	**86%**	**93,29%**
			3954	11-class	**63,27%**	**76,52%**
			19059	16-class	_	**60,76%**

This corresponds to train/validation/split of 1127, 245, and 238 images respectively. On the other hand, the highest score in the state-of-the-art (accuracy of 96.86%) was presented in [24]. This was by combining shape and appearance methods to extract features and the nearest-mean classifier to classify the seven categories, a 10-fold cross-validation test was conducted.

It is apparent from comparison Table XII that the highest accuracies are achieved by our approach, a classification rate of **99.19%** is reached when using the CFEE dataset to fine-tune the Resnet-18 and classification using SVM. Afterward, we obtained a classification accuracy of **98.02%**, using the KDEF dataset to adjust the Resnet-18 model's weights and the CFEE dataset for classification. Additionally, the basic emotions of the CFEE database are evaluated using the same configuration in [70] with 70% of images for the training phase and 15% for the testing phase. It is important to highlight that accuracy of 81.93% was obtained.

The CFEE is a novel and difficult dataset that concerns compound emotions. Thus, few studies are done on it presenting experimental results. Besides, it can be easily observed from Table XII that 76.91% was the highest accuracy in the state-of-the-art [24]. This concerns the assessment of the 22-class of CFEE dataset (5,060 images) using 10-fold cross-validation and the combination of shape and appearance method for extracting features and nearest mean for classification. Meanwhile, [72] used the Highway Convolutional Neural Network to evaluate 22 classes of the CFEE database. Test accuracy of 52.14% was obtained by 10-fold cross-validation. Our approach reached a better accuracy, which is **80.69%**.

On the other hand, it can be observed from Table XIII that all comparison references on the seven classes of the RAF database are recent (from 2016-2020) and used the same configuration based on the idea of 5-fold cross-validation. Our results are compared with eleven studies that evaluate the same number of classes and the same number of images. Most of these studies are based on CNN and deep learning approaches. Moreover, they take the mean diagonal values as a metric of comparison. Due to the imbalance- issue in the RAF database classes (Unweighted Accuracy (UA)).

Li and Deng [26] have proposed Deep Locality-Preserving CNN for feature extraction with SVM. They evaluated a 7-class and achieved a recall of 74.20% and an accuracy of 84.13%. As shown in Table XIII, the highest unweighted accuracies in the literature are 81.83% and 80.78%. The first was obtained by [78] using Conditional Generative Adversarial Network-based EAU-Net Network (cGAN based EAU-Net). Meanwhile, the second was achieved by Pyramid With Super-Resolution (PSR) Network (VGG-16) [80]. It can be easily observed that our approach achieved higher performance over the recent state-of-the-art systems with an unweighted accuracy of 86% and an accuracy of **93.29%**.

Additionally, It can be noticed from Table XIII that our results are compared with two approaches of the RAF Compound emotions dataset. Shan and Deng [26], [48] proposed the DLP-CNN method to extract features of eleven compound emotions based on the idea of 5-fold cross-validation. This method achieved an accuracy of 57.95% and a recall of 44.55%. It is clear that the result recorded by our approach, an accuracy of 76.52% and a recall of 63.27%, outperformed those obtained by both studies cited above. Nevertheless, there is one state-of-the-art study [71], which evaluated all the basic and compound emotions of the RAF database. Because of the imbalances within the number of images of categories, they manually removed some images in the huge categories and selected 3171 images corresponding to 9 compound categories and 7 basic categories. Thereafter, they constructed a sub-database with 16 categories. In this study, the authors use the traditional methods LBP and HOG with Nearest Class Mean Metric Learning (NCMML). Each method was tested 10 times by randomly dividing the sub-database in half on a category level, one for training and the other for testing. The average Accuracy as a performance metric was adopted. Moreover, they obtained a classification accuracy of 44.10% when using LBP and 36.70% when using HOG.

In our experiment, a classification accuracy of **60.76%** is achieved by testing the same categories and configuration used in [71]. The average accuracy of 10 models (half-half randomly) is used as a metric and the images of each category corresponding to a total of 19059 images are evaluated.

Concerning our experiments on basic and compound emotions of the RAF database, it can be observed from Table XIII that the classification accuracy (output by SVM) decreased when using the mean diagonal value of the confusion matrix as the case for the whole state-of-the-art approaches. Moreover, the accuracy decreased when the basic and compound categories are combined.

VI. Conclusion

Generally, Deep Learning-based approaches have reached the advanced level of performances on Facial Expression Recognition and Facial Recognition tasks compared to hand-crafted feature-based methods [81], [82], [83]. It has recently been discovered that some affective computing analysts are more interested in delicate emotions such as compound facial expression. In this research study, a compound facial expression recognition system is introduced. Therefore, experiments are conducted on the CFEE and RAF datasets, which illustrate the posed and spontaneous, primary, and compound facial expressions. The implemented model uses the functional capabilities of the CNN networks. Transfer learning based on Residual Neural Network structure is adopted for the deep feature extraction stage and SVM for classification.

It can be inferred that the small network structure Resnet-18 has perfectly-recognized all tested categories by getting the best successful classification rates. Moreover, Transfer Learning, fine-tuning, and Resnet-18 are adopted for the first time to classify compound emotions. The eight experimental results proved that the used approach offers good performances on the posed and spontaneous primary and compound emotions.

Examining this approach in various databases of compound emotions like iCV-MEFED and Affectnet and evaluating other experiments on compound emotions using a cross-databases is of great importance. This is amongst our plans for future research.

References

[1] B. C. Ko, "A brief review of facial emotion recognition based on visual information," *sensors*, vol. 18, no. 2, p. 401, 2018.

[2] R. A. Khan, A. Meyer, H. Konik, and S. Bouakaz, "Framework for reliable, real-time facial expression recognition for low resolution images," *Pattern Recognition Letters*, vol. 34, no. 10, pp. 1159–1168, 2013.

[3] C. Pelachaud, "Modelling multimodal expression of emotion in a virtual agent," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 364, no. 1535, pp. 3539–3548, 2009.

[4] S. Du and A. M. Martinez, "Compound facial expressions of emotion: from basic research to clinical applications," *Dialogues in clinical neuroscience*, vol. 17, no. 4, pp. 443–455, 2015.

[5] C. Blais, C. Roy, D. Fiset, M. Arguin, and F. Gosselin, "The eyes are not the window to basic emotions," *Neuropsychologia*, vol. 50, no. 12, pp. 2830–2838, 2012.

[6] M. G. Calvo, A. Fernández-Martín, and L. Nummenmaa, "Facial expression recognition in peripheral versus central vision: Role of the eyes and the mouth," *Psychological research*, vol. 78, no. 2, pp. 180–195, 2014.

[7] J. N. Bassili, "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of personality and social psychology*, vol. 37, no. 11, pp. 2049–2058, 1979.

[8] C. Loob, P. Rasti, I. Lüsi, J. C. Jacques, X. Baró, S. Escalera, T. Sapinski, D. Kaminska, and G. Anbarjafari, "Dominant and complementary multi-emotional facial expression recognition using c-support vector classification," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 30 May–3 June, 2017, Washington, DC., pp. 833–838.

[9] K. Shan, J. Guo, W. You, D. Lu, and R. Bie, "Automatic facial expression recognition based on a deep convolutional-neural-network structure," in *15th IEEE International Conference on Software Engineering Research, Management and Applications (SERA)*, 7-9 June, 2017, London, United Kingdom., pp. 123–128.

[10] K. Slimani, S. Bourekkadi, R. Messoussi, Y. Ruichek, and R. Touahni, "Sharing emotions in the distance education experience: Attitudes and motivation of university students," in *4th IEEE International Conference on Intelligent Systems and Computer Vision (ISCV)*, 09-11 June, 2020, Fez, Morocco., pp. 1–10.

[11] A. Mourão and J. Magalhães, "Competitive affective gaming: winning with a smile," in *21st ACM International Conference on Multimedia*, 21-25 October, 2013, Barcelona, Spain., pp. 83–92.

[12] L. D. Riek and P. Robinson, "Using robots to help people habituate to visible disabilities," in *12th IEEE International Conference on Rehabilitation Robotics (ICORR)*, 27 June-1 July, 2011, Zurich, Switzerland., pp. 1–8.

[13] M. S. Bartlett, G. Littlewort, I. Fasel, and J. R. Movellan, "Real time face detection and facial expression recognition: Development and applications to human computer interaction," in *Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, vol. 5, 16-22 June, 2003, Madison, Wisconsin, pp. 53–53.

[14] C. Darwin, S. Pozzi, and R. Benoît, *L'expression des émotions chez l'homme et les animaux*. C. Reinwald, 1874. [Online]. Available: https://books.google.fr/books?id=ildhAAAAIAAJ

[15] S. Sumpeno, M. Hariadi, and M. H. Purnomo, "Facial emotional expressions of life-like character based on text classifier and fuzzy logic," *IAENG International Journal of Computer Science*, vol. 38, no. 2, pp. 122–133, 2011.

[16] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, no. 1, pp. 88–106, 1969.

[17] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, p. 5, 1978.

[18] K. Slimani, M. Kas, Y. El Merabet, R. Messoussi, and Y. Ruichek, "Facial emotion recognition: A comparative analysis using 22 lbp variants," in *2nd ACM Mediterranean Conference on Pattern Recognition and Artificial Intelligence*, 27-28 March, 2018, Rabat, Morocco., pp. 88–94.

[19] K. Slimani, M. Kas, Y. El Merabet, Y. Ruichek, and R. Messoussi, "Local feature extraction based facial emotion recognition: A survey," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 4, pp. 4080–4092, 2020.

[20] J. F. Cohn, Z. Ambadar, and P. Ekman, "Observer-based measurement of facial expression with the facial action coding system," *The handbook of emotion elicitation and assessment*, vol. 1, no. 3, pp. 203–221, 2007.

[21] R. R. Hassin, H. Aviezer, and S. Bentin, "Inherently ambiguous: Facial expressions of emotions, in context," *Emotion Review*, vol. 5, no. 1, pp. 60–65, 2013.

[22] C. E. Izard, "Anxiety: A variable combination of interacting fundamental emotions," *Anxiety: Current trends in theory and research*, vol. 1, pp. 55–106, 1972.

[23] R. Plutchik, "Emotions: A general psychoevolutionary theory," *Approaches to emotion*, vol. 1984, no. 197-219, pp. 2–4, 1984.

[24] S. Du, Y. Tao, and A. M. Martinez, "Compound facial expressions of emotion," *Proceedings of the National Academy of Sciences*, vol. 111, no. 15, pp. E1454–E1462, 2014.

[25] A. Martinez and S. Du, "A model of the perception of facial expressions of emotion by humans: Research overview and perspectives," *Journal of Machine Learning Research*, vol. 13, no. 5, pp. 1589–1608, 2012.

[26] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 21-26 July, 2017, Honolulu, HI, USA., pp. 2584–2593.

[27] C. F. Benitez-Quiroz, R. Srinivasan, Q. Feng, Y. Wang, and A. M. Martínez, "Emotionet challenge: Recognition of facial expressions of emotion in the wild," *Computing Research Repository (CoRR)*, vol. abs/1703.01210, 2017. [Online]. Available: http://arxiv.org/abs/1703.01210

[28] D. You, O. C. Hamsici, and A. M. Martinez, "Kernel optimization in discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 631–638, 2010.

[29] Y. Zhong, L. Luo, X. Wang, and J. Yang, "Multi-factor stock selection model based on machine learning," *Engineering Letters*, vol. 29, no. 1, pp. 177–182, 2021.

[30] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *3rd IEEE International conference on automatic face and gesture recognition*, 14-16 April, 1998, Nara, Japan., pp. 200–205.

[31] P. Yang, Q. Liu, and D. N. Metaxas, "Exploring facial expressions with compositional features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 13-18 June, 2010, San Francisco, CA, USA., pp. 2638–2644.

[32] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[33] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE transactions on image processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[35] M. Suk and B. Prabhakaran, "Real-time mobile facial expression recognition system - a case study," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 23-28 June, 2014, Columbus, OH, USA., pp. 132–137.

[36] R. Shbib and S. Zhou, "Facial expression analysis using active shape model," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 1, pp. 9–22, 2015.

[37] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn, "Fera 2015 - second facial expression recognition and analysis challenge," in *11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, vol. 6, 04-08 May, 2015, Ljubljana, Slovenia., pp. 1–8.

[38] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[39] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7803–7821, 2017.

[40] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 13-18 June, 2010, San Francisco, CA, USA., pp. 94–101.

[41] Y. Rahulamathavan, R. C.-W. Phan, J. A. Chambers, and D. J. Parish, "Facial expression recognition in the encrypted domain based on local fisher discriminant analysis," *IEEE Transactions on Affective Computing*, vol. 4, no. 1, pp. 83–92, 2012.

[42] P. Ekman and H. Oster, "Facial expressions of emotion," *Annual review of psychology*, vol. 30, no. 1, pp. 527–554, 1979.

[43] Z. Hu, S. Tang, Y. Luo, F. Jian, and X. Si, "3dacrnn model based on residual network for speech emotion classification," *Engineering Letters*, vol. 29, no. 2, pp. 400–407, 2021.

[44] K. Zheng, Z. Xia, Y. Zhang, X. Xu, and Y. Fu, "Speech emotion recognition based on multi-level residual convolutional neural networks," *Engineering Letters*, vol. 28, no. 2, pp. 559–565, 2020.

[45] S. M. S. Abdullah and A. M. Abdulazeez, "Facial expression recognition based on deep learning convolution neural network: A review," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 53–65, 2021.

[46] L. Hou, H. Chen, G. Zhang, and X. Wang, "Deep learning-based applications for safety management in the aec industry: A review," *Applied Sciences*, vol. 11, no. 2, p. 821, 2021.

[47] J. M. M. Torres, T. Clarkson, K. M. Hauschild, C. C. Luhmann, M. D. Lerner, and G. Riccardi, "Facial emotions are accurately encoded in the neural signal of those with autism spectrum disorder: A deep learning approach," *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 7, no. 7, pp. 688–695, 2022.

[48] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2018.

[49] B. Santra and D. P. Mukherjee, "Local dominant binary patterns for recognition of multi-view facial expressions," in *10th ACM Indian Conference on Computer Vision, Graphics and Image Processing*, 18-22 December, 2016, Guwahati, Assam, India., pp. 1–8.

[50] H. Alshamsi and V. Kupuska, "Real-time facial expression recognition app development on smart phones," *Journal of Engineering Research and Application*, vol. 7, no. 7, pp. 30–38, 2017.

[51] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *ACM International Conference on multimodal interaction*, 9-13 November, 2015, Seattle, Washington, USA., pp. 443–449.

[52] Z. Sun, Z.-P. Hu, M. Wang, F. Bai, and B. Sun, "Robust facial expression recognition with low-rank sparse error dictionary based probabilistic collaborative representation classification," *International Journal on Artificial Intelligence Tools*, vol. 26, no. 04, p. 1750017, 2017.

[53] C. Silva, A. Sobral, and R. T. Vieira, "An automatic facial expression recognition system evaluated by different classifiers," in *X Workshop de Visão Computacional*, 06-08 October, 2014, Uberlândia/Minas Gerais, Brazil., pp. 208–212.

[54] D. Ghimire, J. Lee, Z.-N. Li, S. Jeong, S. H. Park, and H. S. Choi, "Recognition of facial expressions based on tracking and selection of discriminative geometric features," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 3, pp. 35–44, 2015.

[55] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *IEEE Winter conference on applications of computer vision (WACV)*, 7-10 March, 2016, Lake Placid, NY, USA., pp. 1–10.

[56] F. Lin, R. Hong, W. Zhou, and H. Li, "Facial expression recognition with data augmentation and compact feature learning," in *25th IEEE International Conference on Image Processing (ICIP)*, 07-10 October, 2018, Athens, Greece., pp. 1957–1961.

[57] N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, and A. M. Dobaie, "Facial expression recognition via learning deep sparse autoencoders," *Neurocomputing*, vol. 273, pp. 643–649, 2018.

[58] A. Uçar, Y. Demir, and C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," *Neural Computing and Applications*, vol. 27, no. 1, pp. 131–142, 2016.

[59] J. Egede, M. Valstar, and B. Martinez, "Fusing deep learned and hand-crafted features of appearance, shape, and dynamics for automatic pain estimation," in *12th IEEE International conference on automatic face & gesture recognition (FG 2017)*, 30 May-03 June 2017, Washington, DC, USA., pp. 689–696.

[60] Y. Li, J. Zeng, S. Shan, and X. Chen, "Patch-gated cnn for occlusion-aware facial expression recognition," in *24th IEEE International Conference on Pattern Recognition (ICPR)*, 20-24 August, 2018, Beijing, China., pp. 2209–2214.

[61] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Transactions on Affective Computing*, vol. 10, no. 1, pp. 18–31, 2017.

[62] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on computer vision and pattern recognition*, 27-30 June, 2016, Las Vegas, NV, USA., pp. 770–778.

[63] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[64] S. A. Shah, G. Ali, and R. S. Iqbal, "Cross-cultural emotion analysis a clustering approach," *International Journal of Computer Science and Network Security*, vol. 17, no. 8, pp. 35–41, 2017.

[65] K. He, Y. Fu, W. Zhang, C. Wang, Y.-G. Jiang, F. Huang, and X. Xue, "Harnessing synthesized abstraction images to improve facial attribute recognition," in *27th International Joint Conference on Artificial Intelligence*, 13-19 July, 2018, Stockholm, Sweden., pp. 733–740.

[66] S. Kornblith, J. Shlens, and Q. V. Le, "Do better imagenet models transfer better?" in *IEEE/CVF conference on computer vision and pattern recognition*, 15-20 June, 2019, Long Beach, CA, USA., pp. 2661–2671.

[67] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[68] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[69] D. Lundqvist, A. Flykt, and A. Öhman, "Karolinska directed emotional faces (kdef)," *Cognition and Emotion*, vol. 91, no. 630, pp. 2–2, 1998.

[70] V. Mavani, S. Raman, and K. P. Miyapuram, "Facial expression recognition using visual saliency and deep learning," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 22-29 October, 2017, Venice, Italy., pp. 2783–2788.

[71] Z. Liu, S. Li, and W. Deng, "Recognizing compound emotional expression in real-world using metric learning method," in *Springer Chinese Conference on Biometric Recognition*, 14-16 October, 2016, Chengdu, China., pp. 528–536.

[72] K. Slimani, K. Lekdioui, R. Messoussi, and R. Touahni, "Compound facial expression recognition based on highway cnn," in *the New Challenges in Data Sciences: Acts of the Second Conference of the Moroccan Classification Society*, 28-29 mars, 2019, Kenitra, Morocco., pp. 1–7.

[73] Y. Fan, J. C. Lam, and V. O. Li, "Multi-region ensemble convolutional neural network for facial expression recognition," in *Springer International Conference on Artificial Neural Networks*, 4-7 October, 2018, Rhodes, Greece., pp. 84–94.

[74] S. Ghosh, A. Dhall, and N. Sebe, "Automatic group affect analysis in images via visual attribute and feature networks," in *25th IEEE International Conference on Image Processing (ICIP)*, 07-10 October, 2018, Athens, Greece., pp. 1967–1971.

[75] F. Shen, J. Liu, and P. Wu, "Double complete d-lbp with extreme learning machine auto-encoder and cascade forest for facial expression analysis," in *25th IEEE International Conference on Image Processing (ICIP)*, 07-10 October, 2018, Athens, Greece., pp. 1947–1951.

[76] V. Vielzeuf, C. Kervadec, S. Pateux, A. Lechervy, and F. Jurie, "An occam's razor view on learning audiovisual emotion recognition with small training sets," in *ACM International Conference on Multimodal Interaction*, 16-20 October, 2018, Boulder, USA., pp. 589–593.

[77] D. Acharya, Z. Huang, D. Pani Paudel, and L. Van Gool, "Covariance pooling for facial expression recognition," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 18-22 June, 2018, Salt Lake City, UT, USA., pp. 367–374.

[78] J. Deng, G. Pang, Z. Zhang, Z. Pang, H. Yang, and G. Yang, "cgan based facial expression recognition for human-robot interaction," *IEEE Access*, vol. 7, pp. 9848–9859, 2019.

[79] K. Wang, X. Peng, J. Yang, D. Meng, and Y. Qiao, "Region attention networks for pose and occlusion robust facial expression recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 4057–4069, 2020.

[80] T.-H. Vo, G.-S. Lee, H.-J. Yang, and S.-H. Kim, "Pyramid with super resolution for in-the-wild facial expression recognition," *IEEE Access*, vol. 8, pp. 131 988–132 001, 2020.

[81] S. Gandhe, K. Talele, and A. Keskar, "Face recognition using contour matching," *IAENG International Journal of Computer Science*, vol. 35, no. 2, pp. 226–233, 2008.

[82] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *IEEE Conference on computer vision and pattern recognition*, 23-28 June, 2014, Columbus, OH, USA., pp. 1805–1812.

[83] X. Zhao, X. Liang, L. Liu, T. Li, Y. Han, N. Vasconcelos, and S. Yan, "Peak-piloted deep network for facial expression recognition," in *Springer European conference on computer vision*, 8-16 October, 2016, Amsterdam, The Netherlands., pp. 425–442.