International Journal of Scientific Trends- (IJST)
ISSN: 2980-4299
Volume 2, Issue 6, June, 2023
Website: https://scientifictrends.org/index.php/ijst
Open Access, Peer Reviewed, Scientific Journal

Face Expression Recognition in Grayscale Images Using Image Segmentation and Deep Learning

Ghena Zuhair Abdualghani1
Asst. Prof. Dr. Sefer Kurnaz 2
1,2 Altinbas University, Electrical and Computer Engineering, Türkiye
203720314@ogr.altinbas.edu.tr1
Sefer.kurnaz@altinbas.edu.tr2

Abstract

Face recognition relies heavily on the ability to read emotional states conveyed by the face. Face recognition enables a computer to identify individuals in a photograph or video. In contrast, facial expression recognition aids computers in analyzing the emotional state of a single human being, leading to enhanced human-computer interaction. There are several obvious traits, such as the eyes and the shape of the lips that can be used to decipher an individual's emotions. People's lips curve upward and their eyebrows descend as they grin. The same holds true for other emotions, such as anger, grief, surprise, and so on. This study proposed an approach based on transfer learning and deep learning techniques for human facial expression recognition. The Extended Cohn-Kanada (CK+) dataset is used in this study for the experiments. The proposed approach consists of four outstanding deep learning models named; VGG199, Conv2D, VGG16 and DenseNet201. Along with the CNN features the VGG16 model outperforms all the existing approaches as well as the DL models used in this research with the accuracy of 99.10%. The proposed approach is able to efficiently identify the human emotions from a gray-scale image in a very short time.

Keywords: DL, CNN, Machine Learning, Face recognition, VGG199.

Introduction

There have been many technological breakthroughs in the recent decade, but artificial intelligence is one of the most notable. Artificial intelligence (AI) is the study of creating machines with intellect similar to that of humans. Facial expression recognition is a significant barrier to many scientists working in this area because the same expression can have multiple meanings depending on a person's age, race, or gender. One person's expression can be read in a variety of ways based on lighting, context, and body language.

Expressions of emotion communicated through the face are powerful tools in human interaction. Physical, morphological changes in the face are an essential part of facial expression [1]. Human-computer interaction (HCI), healthcare, surveillance, vehicle safety, dishonesty detection, etc. can all benefit from the social information provided by these facial expression variations [2]. Facial expression recognition could be used, for instance, to gauge customers' happiness with a commercial offering. It would allow the vendor to undertake a live market analysis. Recognizing when a driver is nodding off or becoming upset is another intriguing use case. Recognition of facial expressions could assist avoid both of these potentially dangerous scenarios.

Since emotions play such a significant part in human-robot interaction, there has been a recent uptick in interest in building facial expression detection technology into social robots. The study of human-robot interaction (HRI) draws on a wide range of disciplines, including human-computer interaction (HCI), artificial intelligence (AI), robotics, NLP, design, and the social sciences [3].

The ability of robots to identify, analyse, and appropriately respond to social cues from humans is crucial for productive interactions between the two species. Love is a multifaceted attribute that includes a person's emotional state, their demeanour in social situations, their outlook on life, and their character [4]. If a robot can read human emotions, it will be better equipped to make decisions and provide aid to humans [5]. Better user acceptability could result from these robots' ability to foster more productive and enjoyable relationships [6].

In the software development process, interface design is of great importance and has become one of the main tasks of this process. In recent decades, with the popularization of computers, interfaces have been sought that achieve the most “friendly” human-computer interaction possible One interface is part of a computer system with which a person comes into contact ; either physically, perceptively or conceptually [7] Seeking to help the analysis of problems in the interfaces that can negatively impact the user's interaction with the system, it is common to record, through video, the user's interaction with the system during a usability test, an affectivity analysis, among others applications. However, the use of image processing techniques in the generated records to reduce the analysis time is still little explored, mainly when it comes to face recognition and facial expressions. Computer Vision (CV) is the science responsible for the way the machine sees the environment around it. By extracting significant information from video cameras, scanners and other devices

Object recognition is a main area of CV, recognizing faces is not a trivial task for the computer, because faces are complex and multidimensional visual stimuli, making recognition a high-level task there are several ways to recognize faces such as: extracting feature vectors from the basic parts of the face based on

Convolutional Neural Networks (CNN) for being inspired by the human visual system, and also for being characterized as a particular type of deep neural network much easier to train, has been widely adopted by the Computer Vision community for the recognition of expressions. What characterizes this type of network is that it is basically composed of convolutional layers, which process the inputs considering local receptive fields.

Research Objective and Scope

Processing and cropping face parts, modelling facial behaviours, performing categorization, and predicting actions are all tasks that fall under the umbrella of facial analysis, which is a vast area of study. The primary goal of this thesis is to provide solutions to the most pressing problems in face and expression recognition, which may pave the way for the widespread adoption of robust and reliable models for these applications in their various forms. We narrowed our focus to these goals:

- General objective: Develop a tool to support usability tests, which uses facial expression recognition techniques and associate these expressions with the user's interaction with the system at a given moment.
- Specific Objectives: Develop an algorithm capable of inferring emotions such as joy, disgust, anger, surprise. Find ways to associate expressions with their context of use in order to facilitate analysis.

LITERATURE REVIEW

In this section, we discuss the research that has been done on the topic of facial expressions that have been specifically addressed in this thesis. The various face detection techniques are briefly discussed. Techniques for recognizing facial expressions are discussed. Human-machine interactions like those found in security surveillance, video games, and social robots are just a few of the places where FER has found a home. In addition to its use in medicine for the monitoring of pain, sadness, anxiety, and the treatment of mental retardation, FER is used in the field of behavioral research to learn about social facts (such as origin, gender, and age). Despite the fact that human beings can read most facial emotions with relative ease, accurate expression detection by machines remains challenging.

Yet, computer vision systems have yet to deliver on their promise of being able to precisely and quickly identify individuals by their faces in everyday settings. Over the past two decades, there has been a rise in the pursuit of efficient strategies for the automated coding of faces and expressions. Business, national security, robotics, consumer applications, education, mental and physical health, and even automobiles are just some of the many potential uses for such systems. The key to ushering in the long-awaited new era in AI, where machines communicate, predict, and plan seamlessly with people, is in enabling machines too accurately and instantly recognize faces and facial emotions in an unrestricted manner.

Facial Expression Recognition

The study of how computers can detect and interpret human emotions is a hot topic in several disciplines, including IT, psych, and medical. It's been put to use in HCI [8] and, more recently, in HRI, to boost performance.

In the last couple of decades, numerous methods have been presented for identifying facial expressions.

In [9], the authors employ Bayesian networks, support vector machines, and decision trees to rank the best emotion detection machine learning algorithms. To classify facial expressions, a SVM was used in [10]. Gauss-Laguerre wavelets, which have extensive frequency extraction capabilities, are studied by the authors of [11] for their potential to extract texture information from a wide range of face expressions. The facial region is first isolated in each input image. Following feature extraction using GL filters, expression recognition is accomplished using the KNN classification method. The authors of [12] used a hierarchical support vector machine (HMM) using principal component analysis and independent component analysis to extract global and local features, respectively, for facial emotion recognition. As seen in [13], hundreds of facial features were extracted using Gabor feature extraction methods.

To speed up classification, a well-designed three-layer neural network classifier is fed thousands of extracted features that were selected using an AdaBoost-based hypothesis and trained using a back-propagation technique. A face emotion detection system based on a radial basis function (RBF) hidden node and the online sequential extreme learning machine (OSELM) was proposed in [14].

Deep learning methods

Recently, deep learning techniques have helped to advance the state of the art in facial emotion identification, as evidenced by publications like [15, 16, 17, 18]. Single DNNs, which include convolution layers and deep residual blocks, were proposed as a model in [43]. In [44], the authors propose using CNN in tandem with a tailored picture pre-processing phase for the emotion detection job, and in [15], they provide a Hybrid Convolution-Recurrent Neural Network approach to facial expression recognition (FER) in images. Comparing the results of the pre-trained VGG-Face architecture for facial recognition with those of the Inception and VGG architectures for object recognition is what [16] does. In [17], a probability-based fusion scheme is described for an ensemble of convolutional neural networks (CNNs) for facial emotion detection. The architecture of each CNN is altered by employing the convolutional rectified linear layer as the first layer and several hidden maxout layers. Despite the impressive advances in facial expression recognition, the vast majority of articles only address how to enhance outcomes in a single dataset, rather than addressing the issue of cross-dataset evaluation. This issue has been the subject of research in some recent publications, such as [18]. To solve the issue of face expression recognition (FER) using numerous well-known standard face datasets, a deep neural network architecture was developed in [19]. The authors conducted two experiments to gauge the efficacy of the proposed deep neural network architecture: one without human participants and another that compared data from diverse sources. The effect fine-tuning has on performance while using a cross-dataset strategy was studied in [20]. This analysis relied on a modified version of the VGGFace Deep Convolutional Network model (originally trained for face recognition), which was then tweaked to identify emotions. One dataset was designated as the test set, while the remaining datasets served as the training set, and several runs of each experiment were conducted to establish the reliability of the findings.

We base our work on these most recent emergent studies to determine if and how much using various sources during CNN training improves performance during testing. We employ a convolutional neural network in tandem with some tailored pre-processing of images.

PROPOSED METHODOLOGY

The biometrics community as a whole is aware of the growing interest in face recognition for more difficult practical uses. Constructing reliable face recognition systems that can account for a wide range of individual facial characteristics is a difficult topic that has not yet been fully resolved. In a normal surveillance setting, the inability to identify or confirm the identity of individuals being monitored is problematic. Many academics have developed different face recognition algorithms [21] to allow for fully automatic subject detection based on facial features. In such setups, the signal-to-noise ratio represented human face variables such

Identify applicable funding agency here. If none, delete this text box.

expressions, positions, and lighting. Because of the significant impact that variations in the face can have on face recognition accuracy, it is crucial to address both the face itself (the identity component) and the variations in the face (the variation component). It has been found that identity and variation components (which might include things like expression, position, illumination, etc.) are typically encoded independently from the human face in the suggested joint techniques [22, 23] (joint face and facial variation).

The proposed technique for this study is laid out in great depth in this section. Once a research method has been chosen, the following procedures must be carried out in strict accordance for the best possible outcomes to be achieved. In this study, the problem is addressed by employing Convolutional Neural Networks (CNN) and its various variants. The versatility of this method suggests it could be useful for pictures.

By mining the database, we can train deep learning models to detect and categorise facial expression of human faces into seven distinct classes. In this research, we use a Convolutional Neural Network (CNN) to learn what features may be extracted from images. Without collecting additional data, data augmentation can increase the breadth of information available for use in training models. One could classify this as a hybrid structure.

Our entire research endeavour is built on this CNN-based method, and its workflow is depicted in Figure 3.1. Because model training relies so heavily on the available data, the initial stage is the collection of dataset of gray-scale images with most number of classes. Once we had gathered a dataset from various resources, we go on the next stage of methodology. This study used Cohn-Kanada (CK+) datasets for the experiments. Furthermore four very important deep learning models were employed (Conv2D, Vgg16, vgg19 and DensNet201).

Additionally, we split the dataset, placing the majority (80%) of the data/images in the train folder and the remainder (20%) in the test folder. Following this, we use OpenCV to determine the frequency with which a given frame. Each gray scale image has 64 by 64 by 3 pixels. We used a CNN model to extract features from the images. We used a pre-trained network to derive meaningful information from the incoming images.

Our final target is the model implementation. We've created a four distinct networks based on transfer learning concept. The four pre-trained models are: Conv2D, Vgg16, vgg19 and DensNet201. Different parametric settings is employed using the transfer learning as discussed in this chapter. The goal is to achieve the human facial expression from one of these (sad, angry, surprise, disgust, fear, happy and contempt/natural). The following steps explain in depth methodology.

Dataset Selection

Initially, we test our hypotheses on the CK+ dataset. The CK+ dataset includes 981 photos; the last three frames from each sequence are taken. The 981 pictures cover a wide range of emotions, from anger (135) to disgust (177) to fear (75) to happiness (207), sadness (84), surprise (249), and contempt/natural (54). The experiment (80%) of the data/images in the train folder and the remainder (20%) in the test folder. If there isn't enough information in the database, the network could over-fit too quickly; in this scenario, we employ data augmentation to add new information and strengthen the network. Before adding them to the training set, the 64x64 photos are randomly sliced and rotated. During the testing step, we randomly select a corner of each image and perform a variety of cutting and flipping operations, including the top left, top right, left, bottom, right, and center. These procedures multiply the database by a factor of 10 and drastically cut down on classification errors.

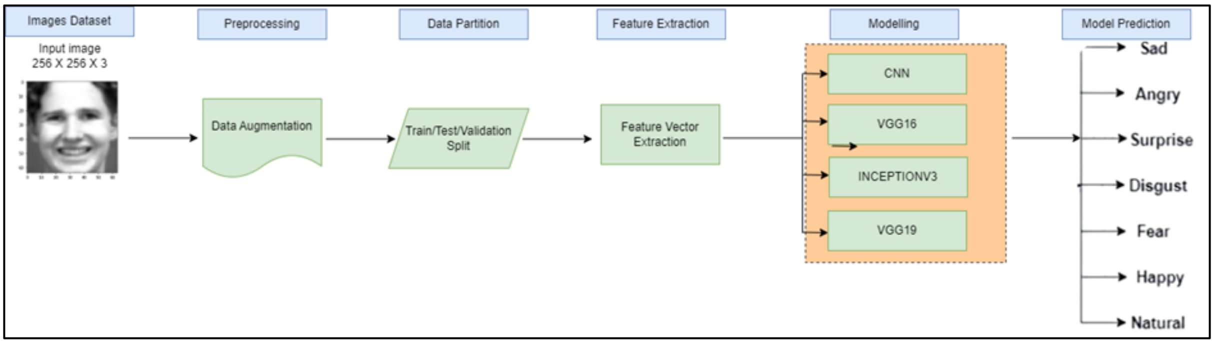


Figure 3.1: The proposed methodology.

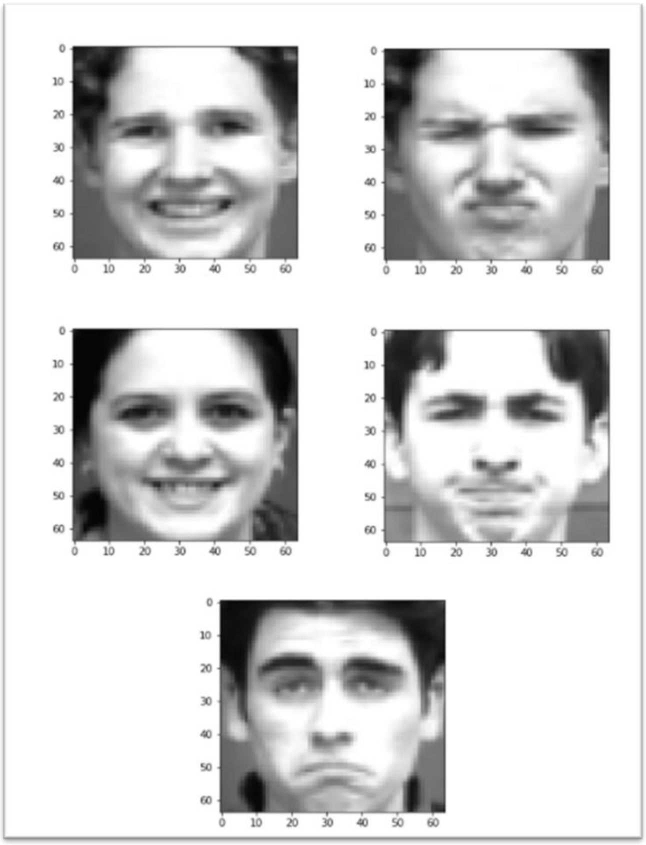


Figure 3.2: Dataset Images Visualization.

Data Pre-processing

In this stage, the database is refined before moving on to the implementation phase. Any research effort that intends to use data for modelling must first undergo the process of data pre-processing. This research project employs image scaling in an effort to expedite the modelling procedure. The original sizes of the photographs are to blame for this disparity.

One major advantage of image scaling is that it makes it possible to train models more quickly. Due to the fact that smaller visuals help models learn quicker, smaller sizes are preferred. Images are scaled down from their original dimensions to 64 x 64 x 3 pixels.

Furthermore, we transform the class vector to integer class matrix using the keras to categorical function.

For performing operations on NumPy arrays, we use the tools provided by the keras.utils.to categorical package. Using the to_categorical () method, a NumPy array containing a vector of integers representing the categories was transformed into a numpy array whose binary elements represent the number of categories in the data. Using the provided tools, we may develop a deterministic and practically applicable version of the karas programmed.

Data Augmentation

As data is transformed, an enhanced image is also created. One of the most common techniques of data augmentation is using scaled photographs from the dataset to increase the size of the training dataset. To begin the process of picture augmentation, the images used for training are modified. In this research, an image data generator is employed to improve photos for categorization using a deep learning model. To do this, a number of transformations need to be performed. Images are often resized as part of data pre-processing, but in order to use them in subsequent modelling, further changes must be applied through image augmentation.

In this study, we examine the process of image enhancement for deep learning models at many phases. At first, the batch size is set for the entire data set. When fine-tuning deep learning models, batch size is one of the most important hyper factors to consider. We've decided to make our batches 32 images in size. The batch size for each model must be appropriately defined. Conclusions regarding the entire population can be skewed when the sample size fluctuates. substantially. After settling on a batch size, all of the images within it are flipped. You can rotate images by up to 30 degrees in the clockwise direction. Even though it's possible that certain pixels will be out of range after an image has been rotated, the rotation process guarantees that every pixel will be retrieved. The mirror image is another thing to think about. In this case, the images are flipped vertically. The vertical flip of an image is achieved by first inverting the image horizontally. At the conclusion, I'll get the revised information. All of these adjustments are made to training data, which is subsequently used to refine the deep learning model's initial settings.

Classification Model

This section details the methodology and model employed during the image classification process. In this research, we demonstrate deep machine learning techniques. Both the model and the procedures followed are described. Deep learning has seen remarkable success in many computer vision applications during the past few years. Since then, this emerging area of ML has shown spectacular expansion. New uses for image identification, computer vision, speech transcription, medical disease detection, and many more are made possible by the superior performance of deep learning over more standard machine learning methods.

This section discusses the model used for image classification, specifically using deep machine learning techniques. The success of deep learning in various computer vision applications is highlighted, and the use of VGG19, 2D Convolutional Network, VGG16, and DenseNet201 models is detailed. The VGG19 model has 19 weight layers, and the VGG16 model has 16 weight layers, and both utilize the transfer learning concept. The 2D Convolutional Network model has six convolutional layers, six max pooling layers, a global average pooling layer, eight dropout layers, and two dense layers. The DenseNet201 model is a 201-layer convolutional neural network. Each model's architecture, activation functions, loss function, optimizer/learning rate, and other experimental settings are described. The training and validation process for each model is also explained, including the plotting of training loss, training accuracy, validation loss, validation accuracy, and confusion matrix. The text also briefly touches on activation functions, such as ReLu and softmax, and loss function, such as categorical cross-entropy.

RESULTS & DISCUSSION

The images in this study included 981 grayscale images, which included seven distinct emotions/sentiments/facial expressions. This study employed four deep learning models for classification. Selecting a model and applying it to a dataset for training is merely the beginning. When developing a machine learning model, it is essential to use a variety of evaluation metrics to gauge the accuracy of the model on untried data sets. We'll check how successfully our suggested method identifies human expressions in this section. Let's take a look at the methodological underpinnings of the experiment before we go into the findings.

EVALUATION MATRIX

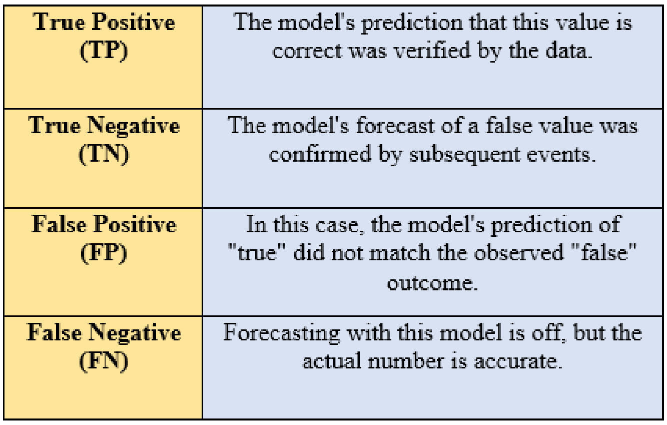
It could be tricky to pick a classifier and then train it using the available data. Every DL model needs to be put through rigorous testing, where its efficiency is evaluated in terms of a number of parameters using a specially crafted test dataset. The model's ability to categorise people into groups is a key metric for gauging its usefulness. Two basic accuracy matrices are taken into account when determining a model's performance on data categorization tasks:

Confusion Matrix

After collecting information in various ways, the data is fed into a top-notch model, which then produces trustworthy probabilities and results. Hold on for the time being! If we're going to compare models, how fully will we compare them? Waste minimization is essential if improved performance is to be achieved. Right now, it is absolutely crucial that the Confusion matrix be made available. To evaluate the efficacy of a ML/DL classification system, the Confusion Matrix is a helpful tool.

The CM is an effective assessment tool for problems with classification. Computational values are used to classify both negative and positive forecasts. The confusion matrix shows the total number of rows where incorrect labels were assigned as a result of overgeneralization. There are four main parts to the confusion matrix, as depicted in the following Figure 4.1.

Figure 4.1 Confusion Matrix



True Positive (TP)	The model's prediction that this value is correct was verified by the data.
True Negative (TN)	The model's forecast of a false value was confirmed by subsequent events.
False Positive (FP)	In this case, the model's prediction of "true" did not match the observed "false" outcome.
False Negative (FN)	Forecasting with this model is off, but the actual number is accurate.

Classification Report

Determining a model's CR can tell you how effectively it can foretell the future. It demonstrates the striking contrast between the model's accurate and inaccurate forecasts. Multiple measures exist for determining a CR's trustworthiness, but TP, TN, FP, and FN are the most widely utilized. The results of machine learning performance tests can be analyzed with the help of the categorization report. Its purpose is to offer hard evidence that the trained classification model is superior in terms of (accuracy, precision, recall, F1 score, and ROC curve) metrics. An effective measure of a model's usefulness is its precision, which reveals how well the true positive observations match the observed positive data.

Simply divide the proportion of correctly labelled classes by both the overall classes to get recall. If the model recall is below 0.5, it is not satisfactory. The F1-score is calculated by combining the results for accuracy and recall. Harmonic means are displayed, along with weights for accuracy and recall. It has been hypothesized that when the F1 score gets closer to 1, the performance of the model will improve.

The likelihood of finding a supporting class in the data provides an indication of how solid a hypothesis is. It takes a more global look at performance instead of focusing on model-specific details.

Area under the receiver operating characteristic curve (AUC)

You can tell if an algorithm reliably classifies data or if it randomly selects labels by looking at the ROC curve. The ROC curve is made by putting the True Positive Rate (TPR) on the y-axis and the False Positive Rate (FPR) on the x-axis. Different cutoffs on a curve depict the likelihood ratio (TPR) vs the false positive rate (FPR). Figure 6 displays the receiver operating characteristics (ROC) and area under the curve (AUC). The area under the curve (AUC) is a measure of how well a classification system does its job. The AUC measures how well a model can predict class labels, and a larger value indicates better performance. The area under the curve (AUC) can range from 0 to 1, representing the classifier's ability to accurately separate positive and negative examples. If the AUC is 0, the classifier misidentifies positive data as negative, but if it's 1, the classifier correctly classifies both positive and negative data.

DATASET SPLITTING

The construction of a training dataset and a test sets is crucial to data mining analysis. Validating a model's predictions is straightforward if the testing set contains real-world examples of the feature in question. Following data cleaning, the dataset will be partitioned into a (training set and a testing set) data. Models are "trained" with this technique by using the data that was collected during the training process. If you have testing data, you can see how well the training data holds up. We have effectively halved the original dataset. To begin, we'll split the current training dataset in two. Nearly eighty percent of the data collected was used for training purposes. In the second split, we used roughly 20% of the full training dataset for testing and validation.

EXPERIMENTAL SETTINGS

The simulations in this investigation are done using the programming language Python. This study utilized the online Python editor Kaggle platform. The high-level, interpreted programming language Python is strong, flexible, and extensively used. Regardless of the complexity of the system, the syntax design and capabilities enable programmers to create code that is straightforward to comprehend and follow. This is an experimental use of Python 3.7.

MODEL CLASSIFICATION RESULTS

This section presents the classification results as shown in Table 4.2. The deep learning models are used in this research. Various DL models (VGG19, Conv2D, VGG16 and DenseNet201) are employed in this research. The models are evaluated using accuracy, precision, recall, F1-score and AUC-score.

Table 4.1: Results of proposed approach

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC score (%)
VGG19	95.05	97.40	93.15	96.10	98.60
Conv2D	95.01	97.37	92.50	95.01	98.31
VGG16	99.10	99.09	99.11	99.08	99.63
DenseNet201	90.01	91.89	85.05	90.01	99.53

The VGG19 model exhibited 95.05% accuracy in this research. In addition, VGG19 obtained 93.15% precision, 93.15% recall and 96.10% f1-score. The Auc-Roc score of the VGG19 model is 98.60%, shows that the model training was outstanding. The confusion matrix and model training and validation loss and accuracy were also plotted in this research. The VGG19 model training and validation accuracy is shown in the Figure 4.2 while the model training and validation loss is plotted in Figure 4.3. The confusion matrix of proposed VGG19 model is shown in Figure 4.4, showing the true positive and miss classified instances of the dataset.

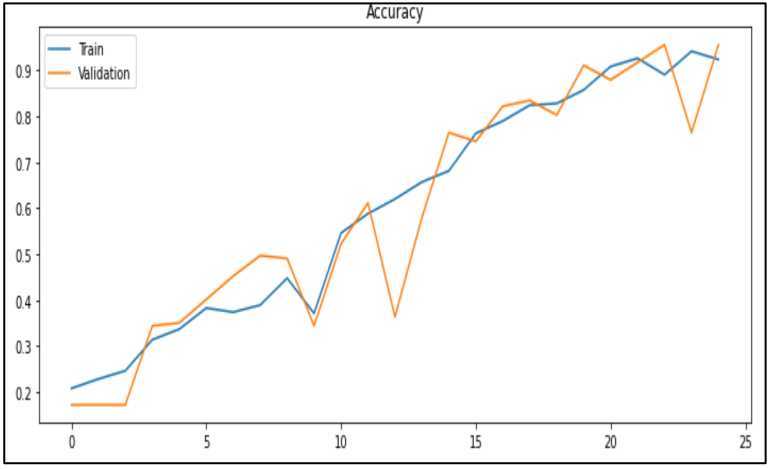


Figure 4.2: VGG19 Model Training and Validation Accuracy

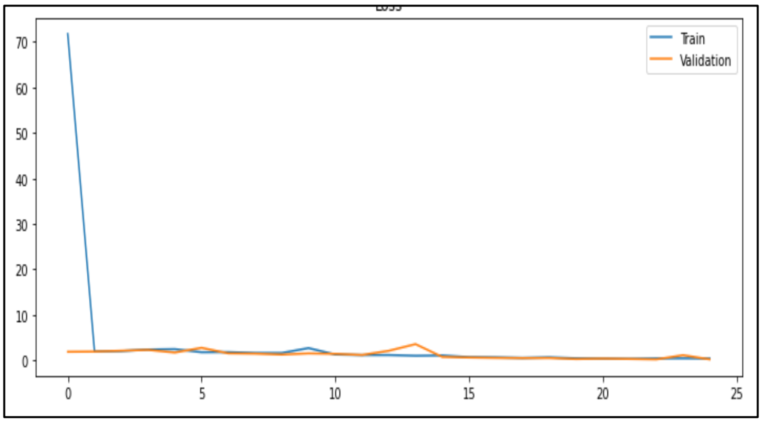
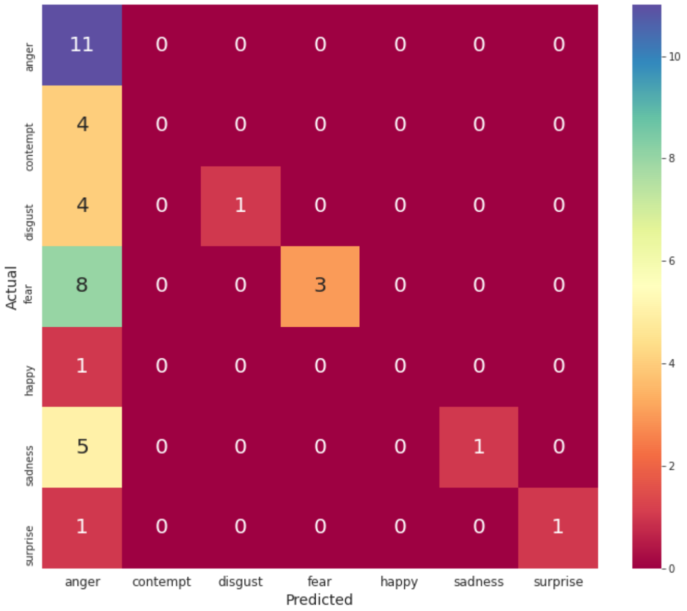


Figure 4.3: VGG19 Model Training and Validation Loss



	anger	contempt	disgust	fear	happy	sadness	surprise
anger	11	0	0	0	0	0	0
contempt	4	0	0	0	0	0	0
disgust	4	0	1	0	0	0	0
fear	8	0	0	3	0	0	0
happy	1	0	0	0	0	0	0
sadness	5	0	0	0	0	1	0
surprise	1	0	0	0	0	0	1

Figure 4.4: VGG19 Model Confusion Matrix

In the next step, the Conv2D model is employed for the experiments and it showed very promising results. The Conv2D model obtained the best results of 95.01% accuracy, 97.37% precision, 92.50% recall, 95.01% f1-score and 98.31% AUC-ROC curve score. The model loss and accuracy for both training and validation is shown in Figure 4.5 and 4.6, whereas the model confusion matrix is presented in Figure 4.7.

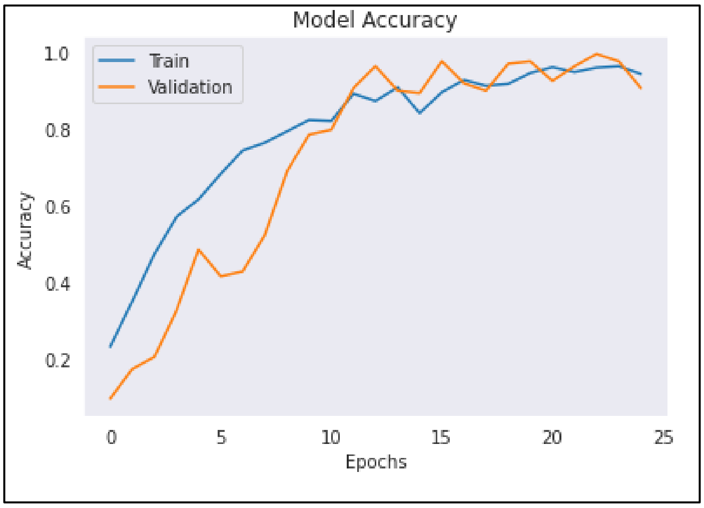


Figure 4.5: Conv2D Model Training and Validation Accuracy

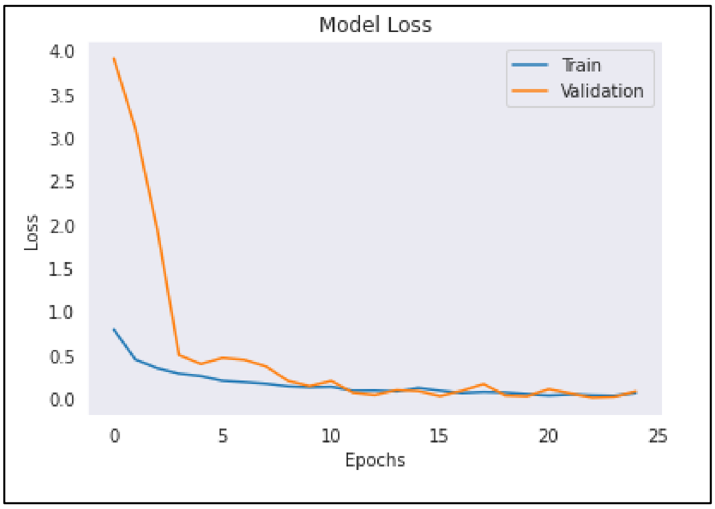


Figure 4.6: Conv2D Model Training and Validation Loss

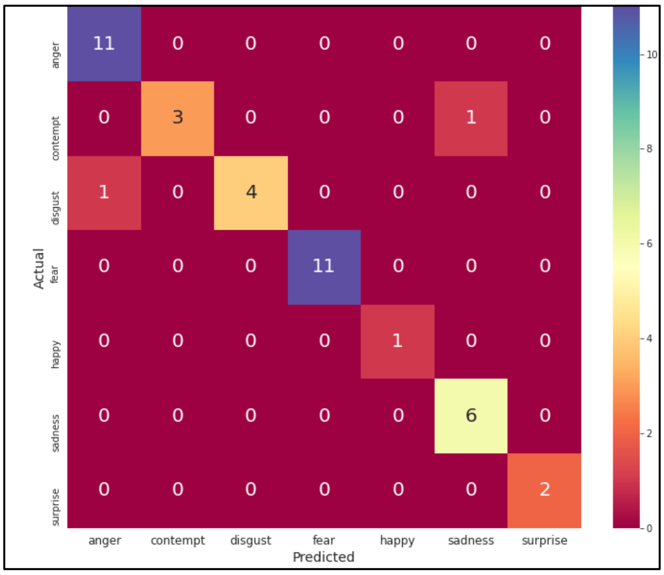


Figure 4.7: Conv2D Model Confusion Matrix

We also employed VGG16 model to evaluate the effectiveness of the proposed method. When compared to other DL models, the VGG16 model that employed transfer learning showed the best performance. With a 99.10% accuracy score, 99.09% precision, 99.11% recall, 99.08% f1-score, and 99.63% area under the receiver operating characteristic curve, the model outperformed all others. The visualization of the proposed VGG16 model's results are shown in below figures. The training and validation accuracy and loss of the VGG16 model is displayed in the Figure 4.8 and 4.9. The confusion matrix of the VGG16 model is also plotted and presented in the Figure 4.10.

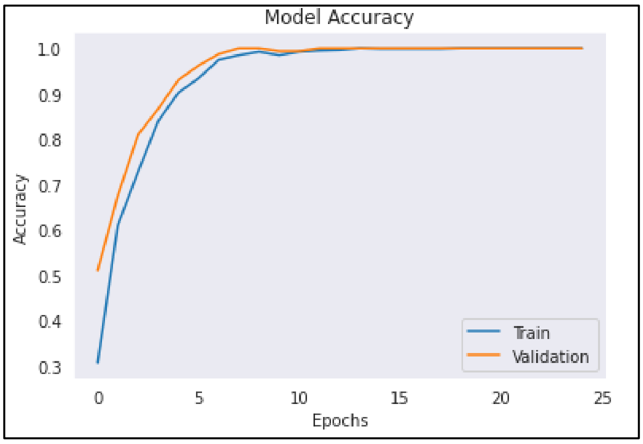


Figure 4.8: VGG16 Model Training and Validation Accuracy

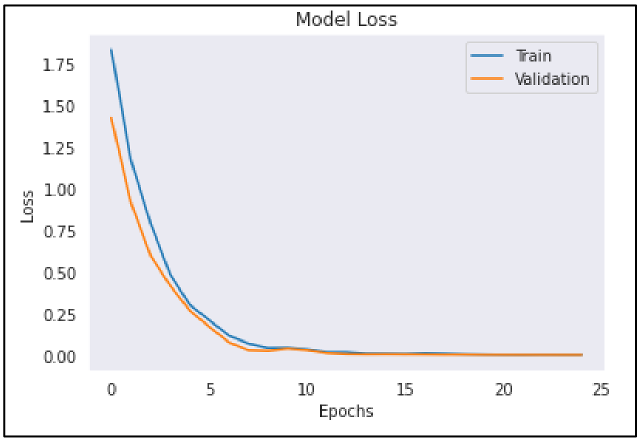
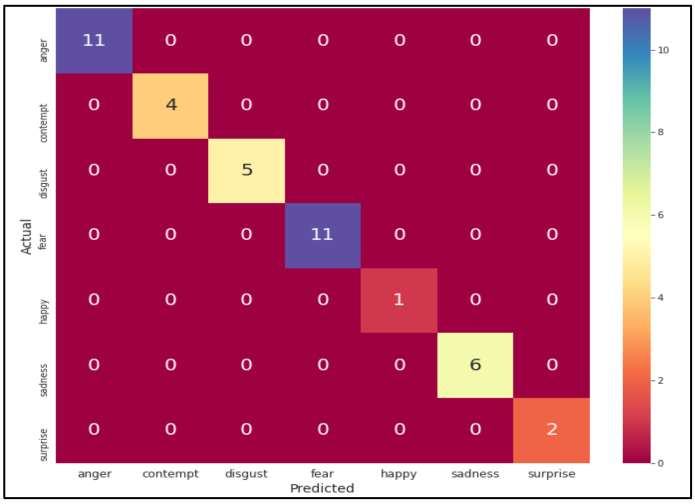


Figure 4.9: VGG16 Model Training and Validation Loss



	anger	contempt	disgust	fear	happy	sadness	surprise
anger	11	0	0	0	0	0	0
contempt	0	4	0	0	0	0	0
disgust	0	0	5	0	0	0	0
fear	0	0	0	11	0	0	0
happy	0	0	0	0	1	0	0
sadness	0	0	0	0	0	6	0
surprise	0	0	0	0	0	0	2

Figure 4.10: VGG16 Model Confusion Matrix

In the end, the final DL model used in this research is DenseNet201. It is the most newly introduced DL model for image classification and other purposes. The results of the transfer learning based DenseNet model is quite impressive. It achieved the accuracy score of 90.01%, 91.89% precision, 85.05% recall, 90.01% f1-score and 99.53% AUC-ROC curve score. The accuracy and loss for the DenseNet201 during the training and validation stage is shown in Figure 4.11 and 4.12. Finally, the confusion matrix of the proposed approach is presented in the Figure 4.13.

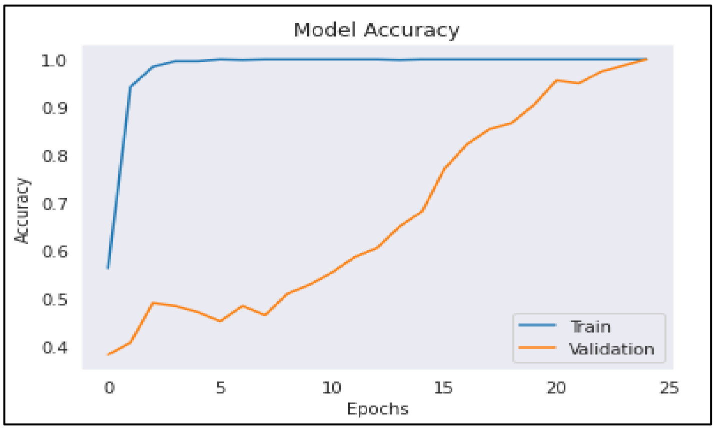


Figure 4.11: DenseNet201 model training and validation accuracy

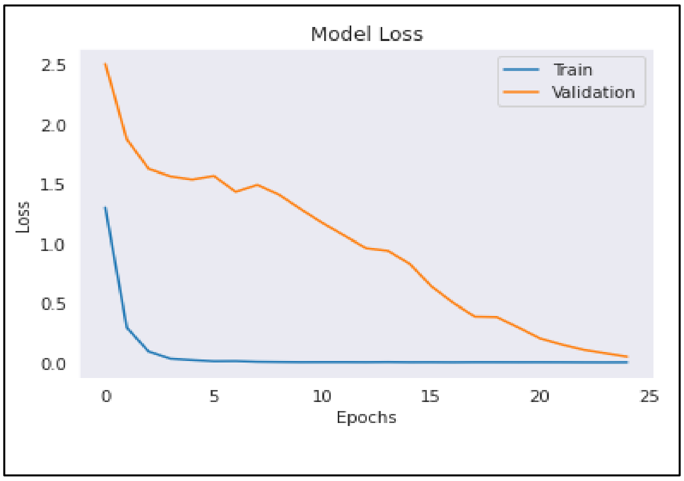


Figure 4.12: Densenet201 Model Training and Validation Loss

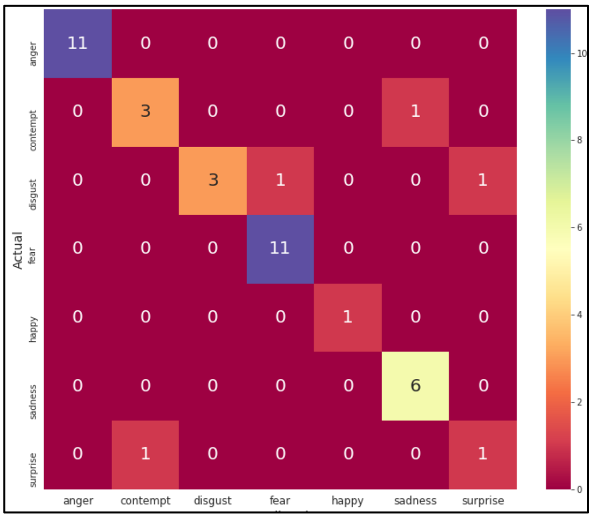
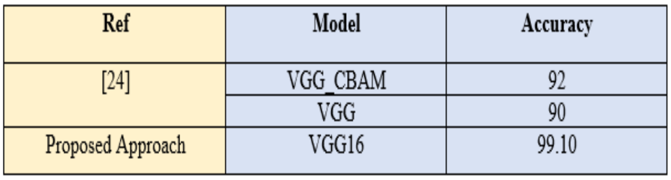


Figure 4.13: DenseNet201 model confusion matrix

Comparative Analysis

The proposed approach is compared with the state-of-the-art approach [24]. The existing approach employed two deep learning models (VGG_CBAM, VGG). They obtained 92% accuracy using VGG_CBAM and 90% using the simple VGG model. The proposed approach obtained the highest accuracy using the VGG16 model and exhibited 99.10% accuracy. The increase in the accuracy by7.10% obtained using the proposed approach when compared with the existing approach. The comparison of both the approaches is shown in Figure 4.14.

Figure 4.14: Comparative Analysis



Ref	Model	Accuracy
[24]	VGG_CBAM	92
	VGG	90
Proposed Approach	VGG16	99.10

Conclusion

Over the past few decades, a plethora of tools and studies have been developed for examining the myriad ways in which people's faces move and change. While much progress has been made in face recognition and analysis, unconstrained face and facial expression analysis has been hampered by the difficulty of dealing with massive visual stimuli, spontaneously expression data, and video-based data. Misclassification occurs in face and expression recognition due to multiple factors in spontaneous real-world applications. In the field of biometrics, obtaining an intrinsic facial feature representation is necessary for achieving human-level accuracies for face identification in the wild. Meanwhile, in the field of computer vision, modelling a system for real-time facial expression analysis requires massive amounts of data from videos of people naturally expressing themselves. In this thesis, we make several novel advances to the field of face and expression recognition. Researchers will benefit immensely from studying and developing novel frameworks for concurrent facial expression detection because expression variation is a challenging issue in present state-of-the-art face recognition systems. Yet, in practical contexts, existing methods fall short of reliably identifying individual differences in people's faces. Therefore, it is evident that the scientific and industrial communities can greatly benefit from this research. i. Develop a tool to support usability tests, which uses facial expression recognition techniques and associates these expressions with the user's interaction with the system at a given moment. ii. Develop an algorithm capable of inferring emotions such as joy, disgust, anger, surprise. Find ways to associate expressions with their context of use in order to facilitate analysis. iii. Implementation of the deep learning approaches based on the transfer learning method for early detection of various human facial expressions from gray-scale images.

Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

References

1. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use "Ref. [3]" or "reference [3]" except at the beginning of a sentence: "Reference [3] was the first ..."
2. Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.
3. Unless there are six authors or more give all authors' names; do not use "et al.". Papers that have not been published, even if they have been submitted for publication, should be cited as "unpublished" [4]. Papers that have been accepted for publication should be cited as "in press" [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.
4. For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].
5. Calvo, M. G., & Nummenmaa, L. (2016). Perceptual and affective mechanisms in facial expression recognition: An integrative review. Cognition and Emotion, 30(6), 1081-1106. doi: 10.1080/02699931.2015.1049124.
6. Happy, S. L., & Routray, A. (2014). Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing, 6(1), 1-12
7. Furrer, F., Burri, M., Achtelik, M., & Siegwart, R. (2016). RotorS—A modular Gazebo MAV simulator framework. In Robot Operating System (ROS) (pp. 595-625). Springer, Cham.
8. Scherer, K. R. (2000). Psychological models of emotion. The neuropsychology of emotion, 137(3), 137-162
9. Picard, R. W. (2000). Affective computing. MIT press
10. Sorbello, R., Chella, A., Calí, C., Giardina, M., Nishio, S., & Ishiguro, H. (2014). Telenoid android robot as an embodied perceptual social regulation medium engaging natural human–humanoid interaction. Robotics and Autonomous Systems, 62(9), 1329-1341.
11. C. Ruoxuan, L. Minyi, and L. Manhua Facial Expression Recognition Based on Ensemble of Multiple CNNs, CCBR 2016, LNCS 9967, pp. 511-578, Springer International Publishing AG 2016.M. F. Valstar, M. Mehu, B. Jiang, M. Pantic, and K. Scherer.
12. Happy, S. L., & Routray, A. (2014). Automatic facial expression recognition using features of salient facial patches. IEEE transactions on Affective Computing, 6(1), 1-12.
13. Sebe, N., Lew, M. S., Sun, Y., Cohen, I., Gevers, T., & Huang, T. S. (2007). Authentic facial expression analysis. Image and Vision Computing, 25(12), 1856-1863.
14. Siddiqi, M. H., Ali, R., Sattar, A., Khan, A. M., & Lee, S. (2014). Depth camera-based facial expression recognition system using multilayer scheme. IETE Technical Review, 31(4), 277-286.
15. Trujillo, L., Olague, G., Hammoud, R., & Hernandez, B. (2005). Automatic feature localization in thermal images for facial expression recognition. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)- Workshops (pp. 14-14). IEEE.
16. Poursaberi, A., Noubari, H. A., Gavrilova, M., & Yanushkevich, S. N. (2012). Gauss– Laguerre wavelet textural feature fusion with geometrical information for facial expression identification. EURASIP Journal on Image and Video Processing, 2012(1), 17.

17. Owusu, E., Zhan, Y., & Mao, Q. R. (2014). A neural-AdaBoost based facial expression recognition system. Expert Systems with Applications, 41(7), 3383-3390.
18. Uçar, A., Demir, Y., & Güzeliş, C. (2016). A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering. Neural Computing and Applications, 27(1), 131-142
19. Jain, D. K., Shamsolmoali, P., & Sehdev, P. (2019). Extended deep neural network for facial emotion recognition. Pattern Recognition Letters, 120, 69-74. ISSN 0167-8655.
20. Lopes, A. T., de Aguiar, E., De Souza, A. F., & 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. ISSN 0031-3203.
21. Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition. Pattern Recognition Letters, 115, 101-106. ISSN 0167-8655
22. Sajjanhar, A., Wu, Z., & Wen, Q. (2018). Deep learning models for facial expression recognition. In 2018 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-6). IEEE. doi: 10.1109/DICTA.2018.8615843.
23. Wen, G., Hou, Z., Li, H., Li, D., Jiang, L., & Xun, E. (2017). Ensemble of deep neural networks with probability-based fusion for facial expression recognition. Cognitive Computation, 9(5), 597-610.
24. Zavarez, M. V., Berriel, R. F., & Oliveira-Santos, T. (2017). Cross-database facial expression recognition based on fine-tuned deep convolutional network. In 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 405-412). IEEE.
25. A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, “2D and 3D face recognition: A survey,” Pattern Recognition Letters, vol. 28, no. 14, pp. 1885 – 1906, 2007. Image: Information and Control.
26. X. Li, G. Mori, and H. Zhang, “Expression-invariant face recognition with expression classification,” in Computer and Robot Vision, 2006. The 3rd Canadian Conference on, pp. 77–77, June 2006.
27. M. Liang and X. Hu, “Recurrent convolutional neural network for object recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3367–3375, 2015.
28. Cao, W., Feng, Z., Zhang, D., & Huang, Y. (2020). Facial expression recognition via a CBAM embedded network. Procedia Computer Science, 174, 463-477.