Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Automatic Facial Expression Recognition and Classification Using Deep Learning Models

Sarita Sharma*, Dr. Nirupama Tiwari**
***Research Scholar, Sage University, Indore*,**
****Associate Professor, Institute of Advance Computing, Sage University,Indore**
sharma.sarita0703@gmail.com ,nirupma.tiwari1974@gmail.com

Article Info
Page Number: 9173 - 9189
Publication Issue:
Vol 71 No. 4 (2022)

Article History
Article Received: 25 March 2022
Revised: 30 April 2022
Accepted: 15 June 2022
Publication: 19 August 2022

Abstract
Computer security, neuroscience, psychology, and engineering are just few of the fields that can benefit from facial expression recognition, often known as FER. Because it does not intrude on a person's privacy, many people believe it could be an effective tool in the fight against criminal activity. Despite this, FER suffers from a number of drawbacks, the most significant of which is the lack of accuracy of its predictions when applied to severe head postures. An intriguing topic of research, automatic emotion recognition based on facial expression has been presented and implemented in a variety of fields, including safety, health, and in human machine interactions, to name a few of these domains. Researchers in this topic are interested in developing methods to understand facial expressions, create codes for them, and extract these elements so that computers can make more accurate predictions. The extraordinary success of deep learning has led to the method's various sorts of architectures being utilized in an effort to improve performance in order to compete with other approaches. An investigation into new research on automatic facial emotion recognition (FER) using deep learning is going to be the focus of this particular piece of writing. We highlight the contributions that were treated, the architecture, and the databases that were used, and we illustrate the progress that has been made by comparing the approaches that were proposed and the results that were produced. The purpose of this publication is to assist and direct future researchers by offering an overview of recent studies as well as insights that can be used to achieve advancements in this subject. Researchers have become more interested in the field of facial expression recognition over the past decade as it has a wide variety of possible applications. Expression analysis includes a number of crucial steps, one of which is feature extraction, which helps toward the goal of accurate and speedy expression recognition. Expressions of joy, surprise, contempt, sadness, rage, and terror can be shown on their faces. Facial expressions can reveal a person's emotions. The most popular method for determining a person's state of mind is by analyzing their facial expressions. There is a wide range of distinct feelings that can be divided into two groups: happy emotions and negative emotions. There are four primary categories of commonly used systems: face detection and extraction, face classification, face recognition, and face recognition. In the current system, it is not so easy to pinpoint the precise emotion that a person is feeling. in addition to the categorization of frame-based expression recognition We should work on detecting facial expressions and emotions in both good and negative pictures, and We should also create strong systems. As a result, increasing the recognition accuracy is intended to be the end result of this study.

Keywords: facial emotion recognition, deep neural networks, automatic recognition, database

I INTRODUCTION

Deep learning for the analysis of facial expressions of emotion Researchers have been gravitating toward the deep learning approach over the past decade because of its high capacity for automatic recognition. This is despite the fact that traditional facial recognition methods have achieved a notable level of success through the extraction of handcrafted features. In this regard, we will discuss some recent works in FER that reveal proposed ways of deep learning in order to get improved detection. These methods are intended to improve the accuracy of the detection. Train and test on a number of databases that are either static or sequential. [1] suggest deep CNN for FER over multiple available databases. Following the extraction of the face landmarks from the data, the images were decreased in size to be 48 pixels on each side. After that, they carried out the procedure of augmentation data. The architecture that was utilized consists of two layers of convolution pooling, followed by the addition of two modules of the inception style, each of which has convolutional layers of sizes 1x1, 3x3, and 5x5. They demonstrate the potential to utilize a technique called the network-in-network, which allows enhancing local performance due to the convolution layers that are applied locally. Furthermore, this technique makes it feasible to reduce the issue of overfitting. [2] In order to get more accurate emotion categorization, it was necessary to investigate the effect that preprocessing the data had on the dataset before training the network. Before applying CNN, which consists of two convolution-pooling layers ending with two fully linked with 256 and 7 neurons[3,] the following actions were carried out: data augmentation, rotation correction, cropping, down sampling with 32x32 pixels, and intensity normalization.

At the test stage, just the best weight that was achieved during the training stage is used. This experience was analyzed using CK+, JAFFE, and BU-3DFE, which are all publicly available databases. According to the findings of some researchers, it is more efficient to do all of these pre-processing procedures simultaneously rather than individually. They offer a new CNN for identifying AUs of the face and also incorporate these pre-processing approaches. [4] They employ two convolution layers for the network, each of which is followed by a maximum pooling layer, and they finish off the network with two completely linked layers that represent the number of AUs that are active. In 2018, in order to solve the problem of disappearance or explosion gradients, [5] proposes a new architecture for CNN that uses sparse batch normalization SBP. This network's defining characteristic is its usage of two consecutive convolution layers at the very beginning, which are subsequently followed by max pooling and SBP. Additionally, in order to mitigate the issue of overfitting, a dropout is used in the middle of three fully connected layers. In order to solve the issue of facial occlusion, the authors of [6] provide a novel approach of CNN. First, the data are input into the VGGNet network. Next, the authors use the technique of CNN using the attention mechanism ACNN. This architecture was trained and tested in three big databases, namely Affect Net, FED-RO, and RAF-DB. [7] made the suggestion that the basic features of the face may be identified. They employed three CNNs, all of which had the same architecture, and had each one detect a different area of the face, such as the eye, the mouth, or the eyebrow. The photographs have to go through the cropping process and the recognition of key-point face features before they can be used on CNN. The famous appearance achieved in conjunction with The unprocessed image was sent into a different kind of CNN that was trained to recognize facial expressions. According to the findings of the research, this method provides a higher level of accuracy than the usage of raw photos or iconize face alone. In the year 2019, [8] conduct research on the FER2013 database to

investigate the influence that varying CNN parameters have on the recognition rate. First, all of the images have a resolution of 64 by 64 pixels, and then there is some variation in the size and number of filters, as well as the type of optimizer used (adam, SGD, or adadelta) on a simple CNN. This type of CNN has two layers of successive convolution, with the second layer performing the role of maximum pooling, and then there is a softmax function for classification. According to these studies, researchers created two novel models of CNN, and they achieved an average accuracy of 65.23% and 65.77% with these models. The distinguishing feature of these models is that they do not contain fully connected layers dropout, and the same filter size is maintained within the network. [9] suggests a new kind of deep CNN that has two residual blocks, and each of those blocks has four convolution layers. After going through a pre-processing step that enables cropping and leveling the intensity of the images, this model is then trained on the JAFFE and CK+ datasets. [10] investigates the change in facial expression that occurs during different emotional states. Based on their findings, they suggest a spatiotemporal architect that combines CNN and LSTM. The CNN starts by learning the spatial features of the facial expression across all of the frames of the emotional state. Next, an LSTM is applied in order to maintain the entire sequence of these spatial properties. Also [11] Present a novel architecture that is called Spatiotemporal Convolutional with Nested LSTM (STC-NLSTM). This architecture is based on three different deep learning sub networks, such as: 3DCNN for extracting spatiotemporal features, followed by temporal T-LSTM to preserve the temporal dynamic, and then the convolutional C-LSTM for modeling the multi-level features. [12] The automatic detection of the emotional state of a human face through the use of computer-based technology is referred to as facial expression recognition, or FER for short. Because it has growing applications in a variety of fields, including psychology, sociology, health science, transportation, gaming, communication, security, and business, the topic of study is currently a hotspot for research. According to [13], facial expressions and emotions guide the lives of people in a variety of ways, and emotions are key aspects that enlighten us in how we should act, from the most elementary processes to the most intricate acts [14]. According to [13], facial expressions and emotions guide the lives of people in a variety of ways. Recent research are concentrating on human behavior and the diagnosis of mental diseases [16], and sporadic breakthroughs in the use of facial expressions in neuropsychiatric difficulties have shown more good outcomes [15].

Additionally, FER may have an impact on the process of data collection within particular research projects. For instance, [17] provided a framework for an intelligent assistant FER that might be used in e-commerce to determine the product preferences of clients. This framework could be deployed in the future. The facial data of the customer is recorded by the system while they look through the online shop for items to buy. The devices are able to make automatic recommendations for further products that may be of interest based on the face expression. It has been found that some physiological characteristics of persons can be utilized as intelligent data in the process of searching for criminals [18]. This idea is predicated on the observation that an individual who has ego is more likely to commit a high-profile crime, such as terrorism, and to display particular emotions, such as rage and fear. The correct recognition of these emotions could, as a result, lead to the implementation of additional security measures that are aimed at capturing criminals. The testing phase of video game development is another area where FER might be useful. Target groups are typically asked to participate in a game for a predetermined amount of time, during which their actions and feelings are examined both behaviorally and emotionally. Using FER technology, game

makers may be able to gain more insights and valuable deductions about the emotions recorded during game play, and they may then incorporate the feedback into the design of the game [19].

II TRAINING AND EVALUATION PHASES

If time or computing property allow, using the same hyperparameters for multiple training sessions can improve accuracy, as random initialization can affect results. When comparing hyperparameters, it is recommended to consider fixed random number generators to avoid skewing the comparison, which is also desirable. Testing more than one nature of architecture can also play an optimistic role. To obtain the same precision, it is more advantageous to choose the least complicated architecture from a prepared point of view. If related, transfer learning is suggested to recover computational time or generalize ability. After correcting all hyperparameters, the model must be retrained by combine the images previously used for preparation or validation into overall training set. In fact, previously all hyperparameters are defined; it is no longer necessary to maintain validation set. So it's worth using this global training set to try to recover exactness one last time (i.e., no post-adjustment to any hyperparameters). The retrained model can be evaluated on test apparatus. The visualization step is also imperative because it helps to better recognize what is happening in model or to ensure toughness of the results. This advance can also supply opportunities to recover routine.

Basic Training: In this process, we train all layers of complex from scratch. We initialize all the layers at random or train them from graze. This training method takes a long time to converge, but produces pretty good accuracy. We denote this preparation process as B.

Fine Tuning: In this training method, we keep the net weight of pertained image of the convolutional layer unchanged. We just aimlessly initialize weights of compactly related layers. Then we train all the layers to junction. It should be noted that the convolutional layer is trained based on net weights of the pertained image, while the dense layer is trained based on the arbitrarily initialized weights. We denote this technique as FT.

Transfer learning: In this method, we do not train convolutional layer of the CNN architecture at all. Instead, we retain the previously trained image network weights. We only train dense layers from erratically initialized weights. We denote this process as TL

III MATERIALS AND METHODS

AFR technology needs a face search database that is easy to use, reliable, and real for it to work properly. Because of this, it is important for the AFR process that the face data set has quality (such as being complete), accuracy (including states of ageing), and accurate features (such as different image file formats, color/grayscale, face resolution, limited/unlimited environment). Other facial databases have also been made for research purposes, and each of these databases is now open to the public.

In the recognising part of face recognition, classification is often used, as can be seen in In fact, it is a machine-learning technique whose goal is to first learn and then use a function that maps a person's facial features to one of two predetermined class labels: class 1 (the face of the individual) or class 2 (not the face of the individual). In this case, the result is a binary classifier. This function assigns a person's facial features to one of the predefined class labels, such as "face of the person"

or "class 2." Classifiers could be used on all of the extracted facial features or just on some of them, such as a person's gender, age, race, etc., among other face features. In recent years, techniques like neural networks have been used to sort things into groups. Face detection is the process of figuring out whether or not an image has a face. If a face is found in the image, the user is told where the face is. A person's face can be recognised by colour, a single intensity, a single picture, or a video series (with each image or frame being captured individually). Face localization is another idea that goes hand in hand with face detection. Figuring out where a single face is in a picture could be thought of as its definition. One important difference between the two is that face localization only works if there is only one face in the image. But when using face detection, you can't know ahead of time if there will be one face or more than one. Facial feature detection is the process of finding out about a person's eyes, eyebrows, nose, nostrils, mouth, lips, and ears, as well as if they have them and where they are. [19-20].

IV DATA COLLECTION

In the proposed approach, emotions will be divided into six categories, namely happiness, sadness, surprise, fear, disgust and anger. In addition, we will expand two categories: mock and neutral, which is the same as the existence in the original CK dataset. This study focused on identifying eight different emotions through the analysis of facial expressions using Different deep learning models. The method for recognizing facial expressions consists of three parts: a part of facial detection, an extraction of facial expression, and a facial expression classification part.

V PROPOSED APPROCH

The main goal of this effort is to come up with good algorithms for recognising faces. In the current body of research, the most attention will be paid to three of the biggest problems with face recognition. These problems are called the problems of different lighting in images, different expressions, and small position changes. After this step, the algorithm local binary pattern for facial expression recognition will be used to make a face descriptor. LBP is used to pull out the features of emotion-related traits by using direction information and a ternary pattern to record a fine edge in a face area where the face itself has a smooth area. This makes it possible for the face to have both smooth and rough spots. Using a method based on histograms, the face is broken up into smaller parts, and then the code for each part is sampled in the same way. After that, the categorization grid is used to make descriptions of the face, while different scales are used to get information about the expression.

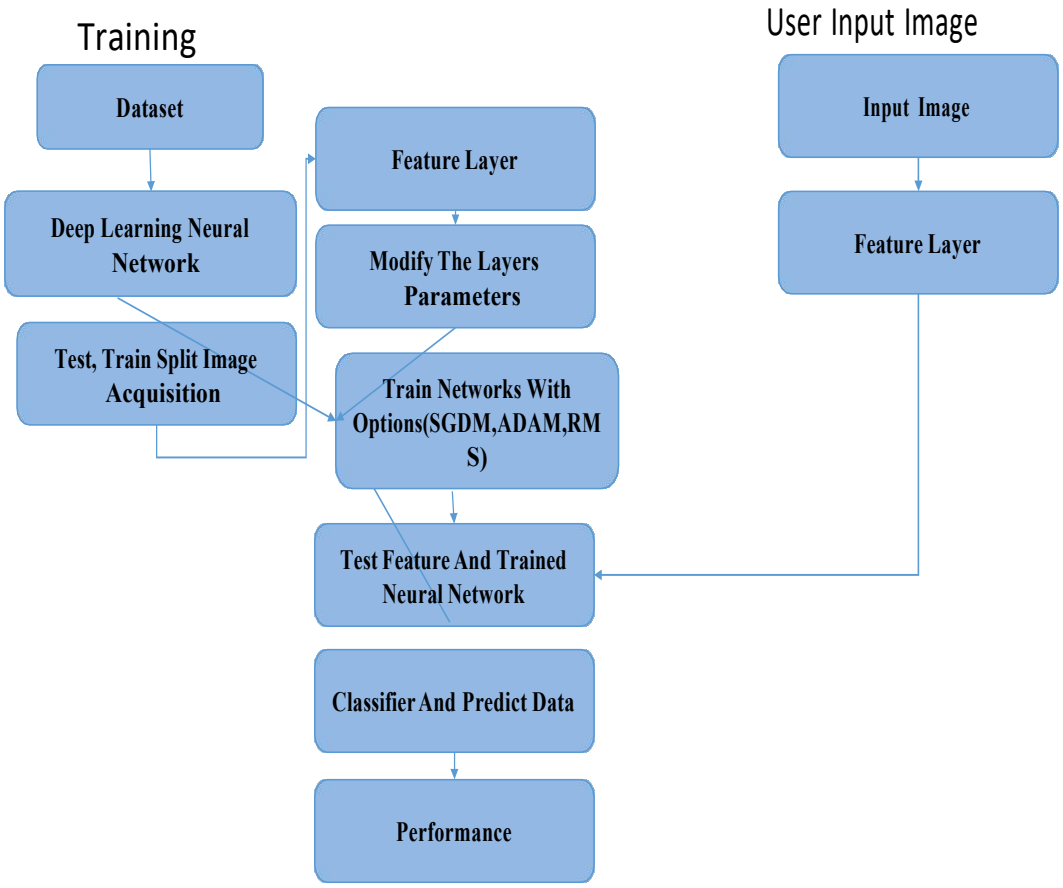


Fig.1 proposed flow diagram

In the proposed approach, emotions will be divided into six categories, namely happiness, sadness, surprise, fear, disgust and anger. In addition, we will expand two categories mock and neutral, Contempt, This study focused on identifying eight different emotions through the analysis of facial expressions using Deep leaning models. The method for recognizing facial expressions consists of three parts: a part of facial detection, an extraction of facial expression, and a facial expression classification part. Fig. 1 illustrates a method for recognizing facial expressions. Facial recognition The system will be performed a 7-way forced choice between the following emotion categories:Anger,Disgust,Fear,Happiness,Sadness,Surprise,Neutral, Contempt.

VI EXPERIMENTAL SETUP

In the current body of research, the primary focus will be placed on three of the most significant challenges associated with face recognition. These challenges are referred to as the problems of lighting variance in images, expression variance, and moderate position changes. After this step, a face descriptor will be created through the use of the algorithm local binary pattern for facial expression recognition. Hence LBP is used to extract the feature information of emotion-related characteristics by making use of directional information and a ternary pattern to record a fine edge in a face area while the face has a smooth area. This allows the face to have both smooth and textured areas. The face is broken up into smaller sections using a histogram-based face description method, and then the code for those sections is sampled in a uniform manner. After that, the

categorization grid is employed to build the face descriptions, while information regarding the expression is sampled at various scales. We take a random sample of seventy percent of all of the photos in each category and include them in the training set. In a similar fashion, the remaining 85% of the photographs in each category are placed in the test set, while the 15% of the images that are placed in the verification set are left alone. The point at which the training set and the validation set intersect. After that, we employ a variety of techniques, including intensity conversion and image enhancement, to boost the total number of images contained in the training set by a factor of 10. The model can be seen in Table 1, and it provides information regarding the amount of photos that fall into each category in the training set, validation set, and test set. All pictures are sized 224 × 224 pixels. Within the framework of the untested system, we establish the CNN architectures' performance evaluation indicators, which together comprise the entire structure. e,

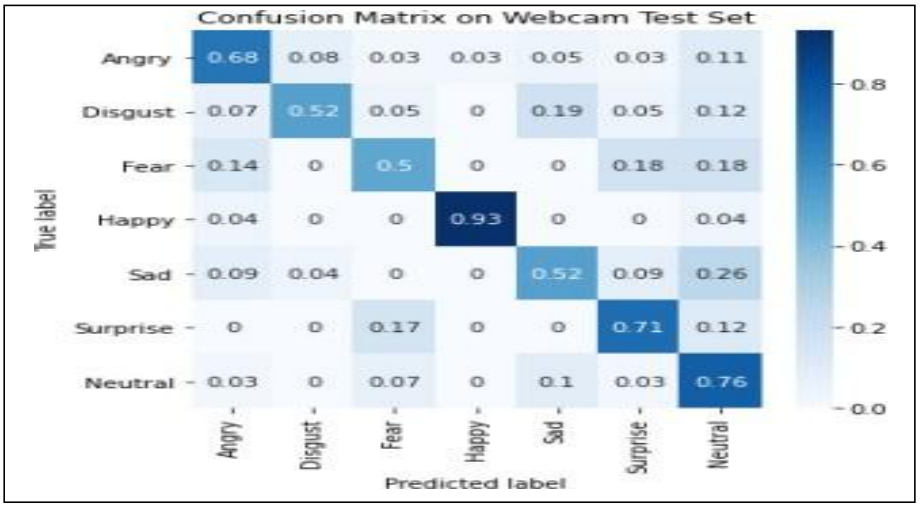


Fig.2 Confusion Matrix of Dataset

Table 1: Image collection of different classes.

Class name	No. of collected images
Fear	100
Surprise	141
Angry	110
Happy	85
Neutral	90
Sad	90
Surprise	152
Happy	90

Facial Expression- The way a person's face looks is an example of an internal difference that can cause big differences within a group. A lot of research is also being done in the fields of human-computer interaction and communication on how to recognise facial expressions. There are ways to deal with expression problems that are based on local characteristics and ways that are based on three-dimensional models.

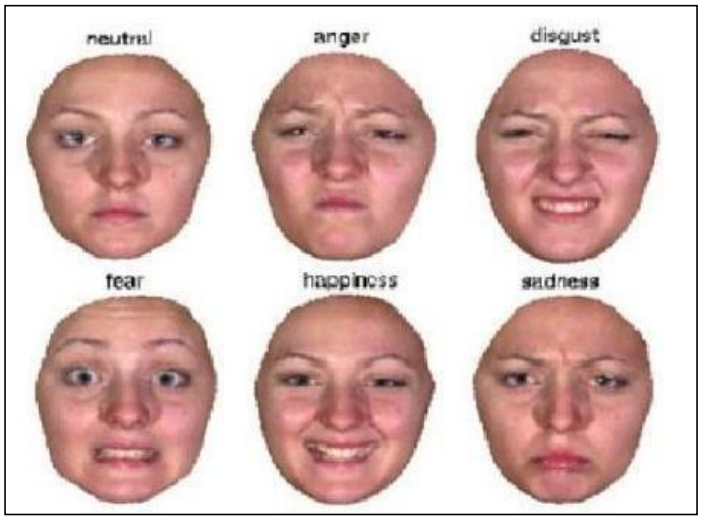


Fig.3 Facial expression variation

We have trained our dataset using six convolutional neural network architectures. They areVGG16, GoogLeNet,Inceptionv3,MobileNet,ResNet,DesNet and CNN To compare our method with the existing recognition methods of After training of each model we get the classification layeroutput for each model like weighted 1000*4096 fully connected layer output getting in VGG 19 model, .weighted 1*1*1000 fully connected layer output getting in MobileNet model. Weighted 1*1*1920 fully connected layer output getting in DenseNet model. Weighted 1000*2048 fully connected layer output getting in InceptionNet model. Weighted 1000*2048 fully connected layer output getting in ResNet model.

Weighted 1000*4096 fully connected layer output getting in InceptionNet model. The they are simple CNN models. Each of the six images contains two 16-dimensional thumbnails with a size of 222 × 222 (the end of the first coordinate field is a matrix of the same size 222 × 222 × 16). The release of the last convolutional layer of CNN's Simple. The six images contain 64 double - dimensional images with a size of 10 × 10 (the output of the last convolutional layer is a matrix with a size of 10 × 10 × 64). The first layer retains the peripheral properties of the input image, although there are some filters.Activation stores almost all of the information found in the introductory image. The exit of the last convolutional layer is not easy to understand. This representation defines a lack of visible internal information about the introductory image. Instead, this layer attempts to display information related to the image category. For different classes, the median results with different classes differ in visibility. The release of the final proposal of the first model yields a slightly less mixed image than the second phase model. Demonstrates the ability of the second step model to learn fewer features. This helps CNN Simple achieve high accuracy and precision after the second phase of training. To improve the ability of deep neural networks, the

most direct way is to increase the depth of the network. However, as the depth of the network width increases, there are too many internal dimensions, which results in more resource consumption. Therefore, in order to overcome these problems, the Inception model into the GoogLeNet architecture.

The first model consists of an upper water-supply layer and a corresponding plate. The sizes of the mixed layers were 1 × 1, 3 × 3, and 5 × 5, which were combined. Between two 1 × 1 separate layers, a max-pooling layer is used to reduce the dimensionality, and a concatenation filter is required to mix the different layers. DenseNet connects the outputs of all layers to the barriers that all layers insert into it. The thick barrier is the repetition of batch normalization, ReLU, 1 × 1 convolution, batch normalization, ReLU, and 3 × 3 convolution over a period of time, as we see the three -layer block. Each time after the thick barrier, the translation layer reduces the size as a 1 × 1.

VII RESULT AND DISCUSSION

We have qualified our dataset using the latest five CNN architectures. They are - VGG16, GoogLeNet,Inceptionv3,MobileNet,ResNet,DesNet.The presentation of these CNN architectures for baseline learning and transfer (TL). To evaluate our method with existing methods, we train the model for the data set. If the image size is set to 224 × 224, precision of the test set is 70%. The image size declare in is 512 × 512. This image size can provide less than 30% verification and proof accuracy. Here, we can see that when we adjust the previously trained ImageNet weights, all of our architectures give the best precision in the test set. On other hand, we did not train various layers of innovative convolutional planning in transfer learning. Therefore, model may not be able to capture all individuality of data set. To get the best presentation changes from each architecture (that is, when we adjust the architecture), the verification precision and the test precision are very high.

Methods like Stochastic Gradient Descent with Momentum (SGDM), Adaptive Moment Estimation (ADAM), and Root Mean Square Propagation (rmsprop) are used for training purposes. For the study, performances are looked at.

SGDM and its variations are the best way to find the best solution for many large-scale learning problems, such as deep learning [12]. Momentum is a strategy that helps move SGD in the right direction and stops it from going back and forth [17]. The value for the momentum term is 0.9. Because of this, momentum gets closer to a steady state faster and has less oscillation.

Adam is the name of a method for making things better. It can be used to change the weights of a network instead of the stochastic gradient descent [6]. This method [17] can be used to figure out the adaptive learning rates for each parameter. Also, Adam keeps an average of historical gradients that decreases in a way similar to momentum [17]. Adam is a well-known algorithm in the field of deep learning because it can give high-quality results quickly [6].

RMSProp gets learning rates for parameters that are tuned based on the average of the recent sizes of the weight's gradients. This shows that the method works well with online and non-stationary problems [6]. RMSprop is found by dividing the rate of learning by the average of squared gradients that are getting smaller and smaller exponentially [17].

Table 2 Deep learning models and number of parameters

Deep learning models	Pixels	parameter
VGG16	224, 224, 3	138 million

Deep learning Architecture	Learning Method	Training Method	Validation Accuracy
GoogLeNet	224, 224	62.3 million	
Inceptionv3	224, 224	24 Million.	
MobileNet	1, 224×224	13 million	
ResNet	224, 224, 3	23 million	
DesNet	224, 224, 3	36928M	

Process performance is precise based on concert indicators such as precision, sensitivity, specificity, or time consumption. Performance Measure: In our data set, the sample size is fairly isolated among 11 categories. When the sample size is not biased towards any particular category, precision is a good performance indicator.

TP- is total number of properly categorized prospects (true positives).

TN- is total number of poorly classified prospects (true negative numbers).

FN- is total number of false rejections, which represents the number of false pixels of foreground pixels classified as background (false negatives).

FP- is total number of false positives, which means that pixels are mistakenly classified as foreground (false positives). Calculate presentation value for each frame of input video based on overhead indicators. Accuracy = correct number of predictions, total number of predictions

Accuracy = TP + TNTP + TN + FP + FN

Specificity: Specificity is distinct as proportion of definite refusals that can be predicted as negatives (or true negatives).

Specificity = (True negative) / (True negative + False positive)

Accuracy: In field of material retrieval, accuracy is quantity of recovered documents related to the query

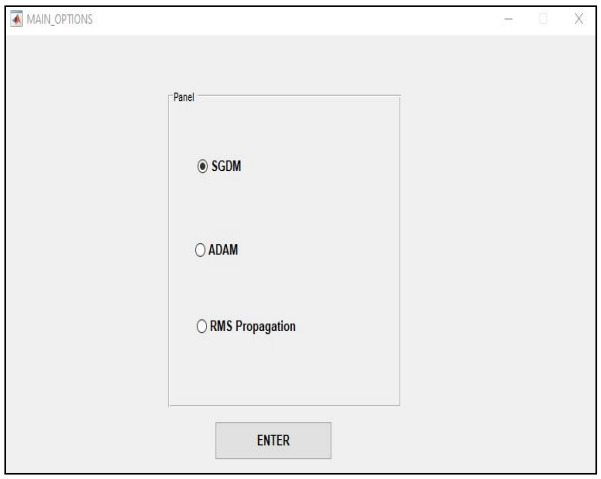
$$\Pr = \frac{(\qquad)}{(\qquad+\qquad)}$$

VGG 16	Baseline	ADAM	85.12%
		SGDM	90.25%
		RMS Propagation	75.24%
	Transfer Learning	ADAM	82.36%
		SGDM	93.36%
		RMS Propagation	94.15%
Inceptionv3	Baseline	ADAM	94.84%
		SGDM	90.48%
		RMS Propagation	91.96%
	Transfer Learning	ADAM	90.21%
		SGDM	95.31%
		RMS Propagation	93.69%
ResNet50	Baseline	ADAM	92.21%
		SGDM	93.78%
		RMS Propagation	98.23%
	Transfer Learning	ADAM	97.15%
		SGDM	99.23%
		RMS Propagation	94.89%
DenseNet	Baseline	ADAM	92.12%
		SGDM	96.32%
		RMS Propagation	90.21%
	Transfer Learning	ADAM	90.15%
		SGDM	92.23%
		RMS Propagation	86.14%
GoogLeNet	Baseline	ADAM	92.15%
		SGDM	94.23%

		RMS Propagation	90.25%
	Transfer Learning	ADAM	90.36%
		SGDM	95.12%
		RMS Propagation	93.84%
MobileNetv2	Baseline	ADAM	93.69%
		SGDM	94.39%
		RMS Propagation	86.23%
	Transfer Learning	ADAM	91.89%
		SGDM	90.23%
		RMS Propagation	96.25%

Precision is used with the retrieval rate, which is the percentage of all relevant documents returned by the search

Table 3 performance of Deep learning models



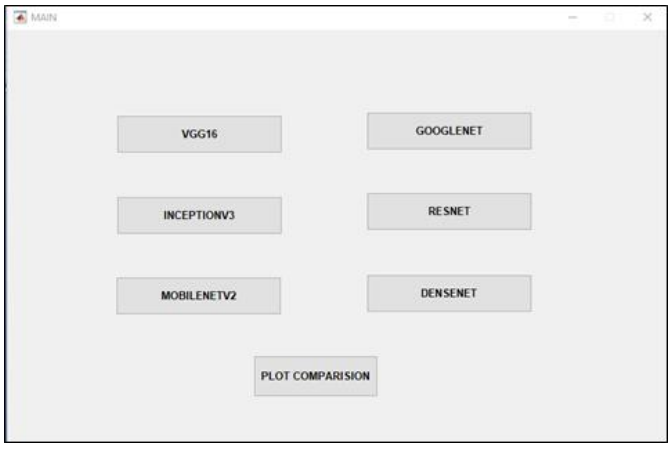


Fig.4 Learning and Training Method

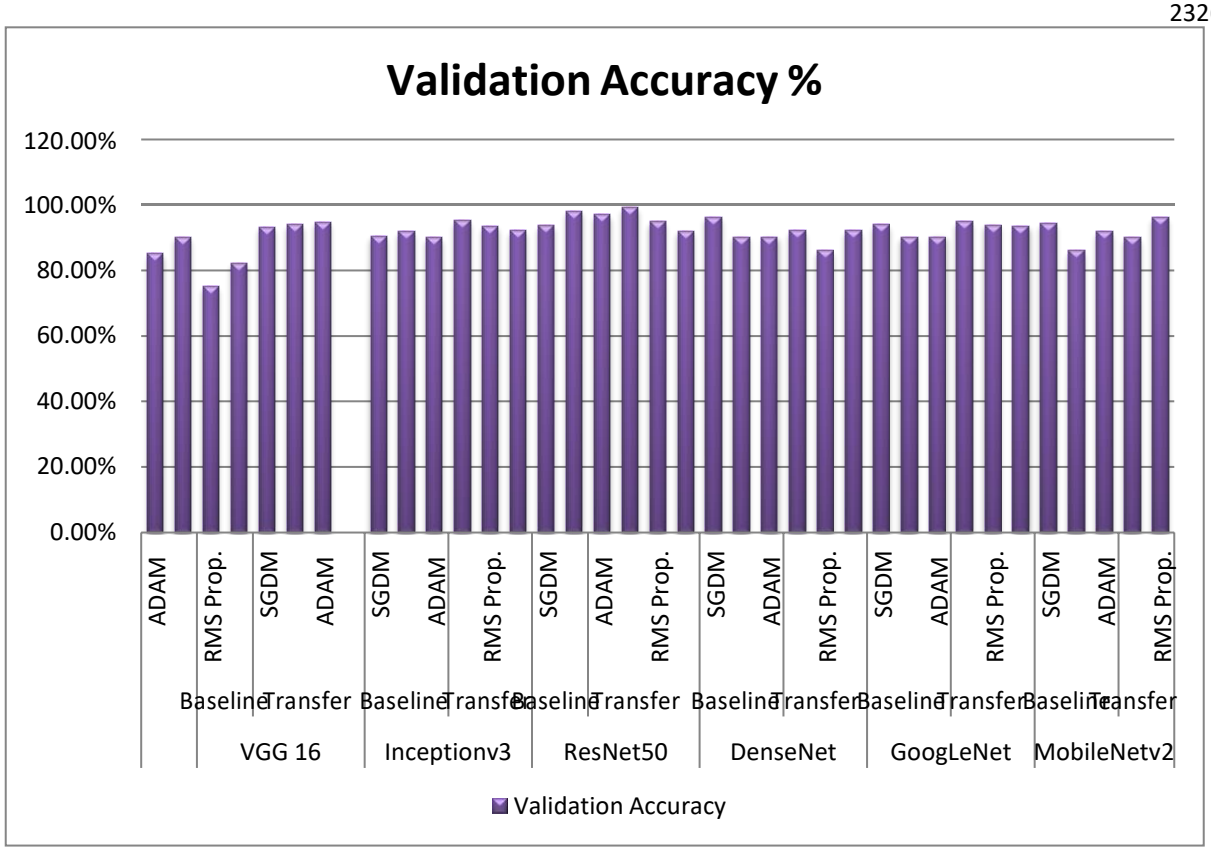


Fig.5 Different Deep learning models

Reference	**Year**	Deep Learning Architecture	Accuracy
Cai et al.[1]	**2018**	sBN-CNN	95.24%
Li et al.[2]	**2019**	ACNN	80.54%,
Yolcu et. al.[3]	**2019**	CNN	94.40%
Agrawal et Mittal.[4]	**2019**	CNN	65%
Deepak jain et al.[5]	**2020**	CNN	95.23%
Kim et al.[6]	**2021**	CNN-LSTM	78.61%
Proposed Pre-Trained Models		Deep Learning Architecture	Accuracy
		VGG16,	94.15%

	GoogLeNet,	94.32%
	Inceptionv3	95.31%
	ResNet	99.23%
	MobileNet	94%
	DesNet	96.32%

Table 4 Comparison of proposed model with existing techniques

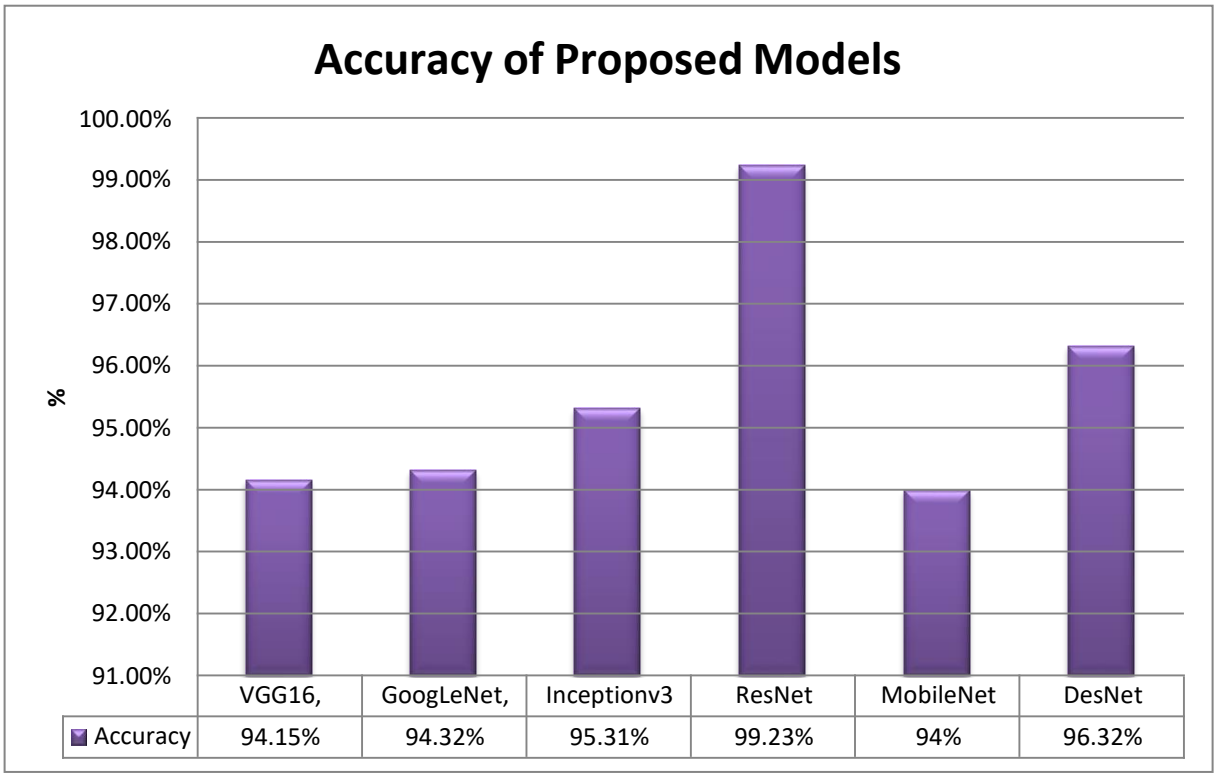


Fig.6 Comparison of proposed model with existing techniques

Table 3 showing the performance of the different deep leaning model with different leaning algorithm and training method in case of VGG16 model the accuracy for baseline learning method with SDGM training method accuracy has come 90.25% ,for ADAM accuracy has come 85.12%, RMS Propagation accuracy has come 75.24%. In case of transfer leaning method with SDGM training method accuracy has come 93.36% for ADAM accuracy has come 82.36%, RMS Propagation accuracy has come 94.15%. in case of Inceptionv3 model the accuracy for baseline learning method with SDGM training method accuracy has come 90.48% , for ADAM accuracy has come 94.84% , RMS Propagation accuracy has come 91.96 %. In case of transfer leaning method with SDGM training method accuracy has come 95.31% for ADAM accuracy has come 90.21%, RMS Propagation accuracy has come 93.69 %. in case of ResNet50 model the accuracy for baseline learning method with SDGM training method accuracy has come 93.78% , for ADAM accuracy has come 92.21%, RMS Propagation accuracy has come 98.23%. In case of transfer leaning method with SDGM training method accuracy has come 99.23 for ADAM accuracy has come 92.21%, RMS Propagation accuracy has come 94.89%.

in case of DenseNet model the accuracy for baseline learning method with SDGM training method accuracy has come 96.32% for ADAM accuracy has come 97.15%, RMS Propagation accuracy has come 90.21%. In case of transfer leaning method with SDGM training method accuracy has come 92.23% for ADAM accuracy has come 90.15%, RMS Propagation accuracy has come 90.21%. in case of GoogLeNet model the accuracy for baseline learning method with SDGM training method accuracy has come 92.15% , for ADAM accuracy has come 92.15%, RMS Propagation accuracy has come 90.25%. In case of transfer leaning method with SDGM training method accuracy has come 95.12% for ADAM accuracy has come 90.36 %, RMS Propagation accuracy has come 93.84%. in case of MobileNetv2 model the accuracy for baseline learning method with SDGM training method accuracy has come 98.69% ,for ADAM accuracy has come 93.69%, RMS Propagation accuracy has come 86.23 %. In case of transfer leaning method with SDGM training method accuracy has come 90.23% for ADAM accuracy has come 91.89%, RMS Propagation accuracy has come 96.25%.in the table 2 showing the comparison performance of proposed model with existing techniques the proposed model compare with the other deep learning model ,after the training and testing process ResNet 50 model get the higher accuracy as compare to other models,ResNet50 model get the higher accuracy if we compare with the other existing model.

Table 4 showing the comparison result with different existing techniques with proposed deep learning models that showing the resnet50 model showing higher accuracy as compare to other existing techniques

VIII CONCLUSION

Automatic identification of spontaneous and posed facial expressions has become more essential in human behavior analysis as a result of the emerging and increasingly supported hypothesis that facial expressions do not always reflect our true feelings. The most current developments in detecting facial expressions, which have taken place during the previous two decades, are discussed in this article. This article looks at a total of six different deep learning models. In particular, we have presented in-depth comments on previously conducted research on the recognition of facial expressions, looking at both the data collection and detection approach from a variety of angles. In order to achieve a more complete comprehension of this burgeoning sector, there have been recognized as well a number of complex problems. After that, a preprocessing stage is applied to these ROIs in order to resize them and partition them into blocks. This is done in preparation for the feature extraction stage, which is used to construct a face feature descriptor. Inferences regarding emotional state can then be made using a classifier. In order to compare the suggested method to others, a full experimental investigation is carried out employing a variety of various regional factors. The results of our experiments showed that our proposed strategy is effective when applied to all tested datasets and when using all tested descriptors. The face decomposition that was proposed performed better than the ones that were considered state of the art. The suggested method beats state-of-the-art methods that are based on alternative facial decompositions, according to the results of experiments conducted on two publicly available datasets.

REFERENCES

1. J. Cai, O. Chang, X. Tang, C. Xue, et C. Wei, « Facial Expression Recognition Method Based on Sparse Batch Normalization CNN », in 2018 37th Chinese Control Conference (CCC), juill. 2018, p. 9608-9613, doi: 10.23919/ChiCC.2018.8483567.
2. Y. Li, J. Zeng, S. Shan, et X. Chen, « Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism », IEEE Trans. Image Process., vol. 28, no 5, p. 2439-2450, mai 2019, doi: 10.1109/TIP.2018.2886767.
3. G. Yolcu et al., « Facial expression recognition for monitoring neurological disorders based on convolutional neural network », Multimed. Tools Appl., vol. 78, no 22, p. 31581-31603, nov. 2019, doi: 10.1007/s11042-019-07959-6.
4. Agrawal et N. Mittal, « Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy », Vis. Comput., janv. 2019, doi: 10.1007/s00371-019-01630-9.
5. D. K. Jain, P. Shamsolmoali, et P. Sehdev, « Extended deep neural network for facial emotion recognition », Pattern Recognition. Lett., vol. 120, p. 69-74, avr. 2019, doi: 10.1016/j.patrec.2019.01.008.
6. D. H. Kim, W. J. Baddar, J. Jang, et Y. M. Ro, « Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition », IEEE Trans. Affect. Comput., vol. 10, no 2, p. 223-236, avr. 2019, doi: 10.1109/TAFFC.2017.2695999.
7. Z. Yu, G. Liu, Q. Liu, et J. Deng, « Spatio-temporal convolutional features with nested LSTM for facial expression recognition », Neurocomputing, vol. 317, p. 50-57, nov. 2018, doi: 10.1016/j.neucom.2018.07.028.
8. D. Liang, H. Liang, Z. Yu, et Y. Zhang, « Deep convolutional BiLSTM fusion network for facial expression recognition », Vis. Comput., vol. 36, no 3, p. 499-508, mars 2020, doi: 10.1007/s00371-019-01636-3.
9. E. Sariyanidi, H. Gunes, et A. Cavallaro, « Automatic Analysis of Facial Affect: A Survey of Registration, Representation, and Recognition », IEEE Trans. Pattern Anal. Mach. Intell., oct. 2014, doi: 10.1109/TPAMI.2014.2366127.
10. C.-N. Anagnostopoulos, T. Iliou, et I. Giannoukos, « Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011 », Artif. Intell. Rev., vol. 43, no 2, p. 155-177, févr. 2015, doi: 10.1007/s10462-012-9368-5.
11. L. Shu et al., « A Review of Emotion Recognition Using Physiological Signals », Sensors, vol. 18, no 7, p. 2074, juill. 2018, doi: 10.3390/s18072074.
12. C. Marechal et al., « Survey on AI-Based Multimodal Methods for Emotion Detection », in High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet,
13. M. H. Alkawaz, D. Mohamad, A. H. Basori, et T. Saba, « Blend Shape Interpolation and FACS for Realistic Avatar », 3D Res., vol. 6, no 1, p. 6, janv. 2015, doi: 10.1007/s13319-015-0038-7.
14. P. V. Rouast, M. Adam, et R. Chiong, « Deep Learning for Human Affect Recognition: Insights and New Developments », IEEE Trans. Affect. Comput., p. 1-1, 2018,
15. C. Shan, S. Gong, et P. W. McOwan, « Facial expression recognition based on Local Binary Patterns: A comprehensive study », Image Vis. Comput., vol. 27, no 6, p. 803-816, mai 2009, doi: 10.1016/j.imavis.2008.08.005.

16. T. Jabid, M. H. Kabir, et O. Chae, « Robust Facial Expression Recognition Based on Local Directional Pattern », ETRI J., vol. 32, no 5, p. 694 Wafa Mellouk et al. / Procedia Computer Science 175 (2020) 689–694 6 Author name / Procedia Computer Science 00 (2018) 000–000 784-794, 2010, doi: 10.4218/etrij.10.1510.0132.
17. S. Zhang, L. Li, et Z. Zhao, « Facial expression recognition based on Gabor wavelets and sparse representation », in 2012 IEEE 11th International Conference on Signal Processing, oct. 2012, vol. 2, p. 816-819, doi: 10.1109/ICoSP.2012.6491706.
18. R. Gross, I. Matthews, J. Cohn, T. Kanade, et S. Baker, « Multi-PIE », Proc. Int. Conf. Autom. Face Gesture Recognit. Int. Conf. Autom. Face Gesture Recognit., vol. 28, no 5, p. 807-813, mai 2010, doi: 10.1016/j.imavis.2009.08.002.
19. M. Pantic, M. Valstar, R. Rademaker, et L. Maat, « Web-based database for facial expression analysis », in 2005 IEEE International Conference on Multimedia and Expo, juill. 2005, p. 5 pp.-, doi: 10.1109/ICME.2005.1521424.
20. M. F. Valstar, B. Jiang, M. Mehu, M. Pantic, et K. Scherer, « The first facial expression recognition and analysis challenge », in Face and Gesture 2011, mars 2011, p. 921-926, doi: 10.1109/FG.2011.5771374.