PERFORMANCE FACIAL EXPRESSION RECOGNITION USING HYBRID OF RESNET50 AND RESNET34 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

Abstract: The difficulty of recognizing facial expressions is a key part of computer vision and human-computer contact. It enables the interpretation of human emotions from facial images, contributing to applications such as affective computing, social robotics, and psychological research. In this work, we propose the use of ResNet50 and ResNet34 hybrid deep learning models for facial expression classification. These models, pre-trained on large-scale datasets, exhibit powerful feature extraction capabilities and have shown excellent performance in various computer vision tasks. follow a comprehensive approach, starting with the collection and preprocessing of a labeled facial expression dataset. The collected dataset undergoes face detection, alignment, and normalization to ensure consistency and eliminate noise. The preprocessed dataset is then split into training, validation, and testing sets. The ResNet50 and ResNet34 models are fine-tuned on the training set, leveraging transfer learning to adapt the pre-trained models to the facial expression recognition task. We employ optimization techniques such as SGDM, ADAM, and RMSprop.To update the models' parameters and minimize a categorical cross-entropy loss function. The performance of the trained models is evaluated on the validation set, considering metrics with accuracy 98.19%,.the models are tested on unseen facial images to assess their generalization capabilities. The proposed approach aims to provide accurate and robust facial expression classification, contributing to the advancement of emotion analysis and human-computer interaction systems.

Keywords: convolutional neural network; attention mechanism; ResNet50 and ResNet 34,facial expression recognition

I INTRODUCTION

Emotions are a big part of what it means to be human, and they play a big part in how people communicate with each other [1,2]. People can show how they feel in many different ways, such as through words, body language, and facial expressions [3,4]. Analysis of facial movements is by far the most well-known and researched part of figuring out what emotions someone is feeling. [6] and [7] have done a lot of study on how people's faces change. Their research led to the discovery of universal facial feelings like happiness, sadness, anger, fear, surprise, disgust, and neutral. In the fields of psychology, psychiatry, and mental health, figuring out how someone feels by looking at their face has become an interesting area of study in recent years [8]. It is also important for "smart living" [9] and health care systems to be able to automatically tell what people are feeling by looking at their faces. [10], emotion disorder diagnosis in autism spectrum disorder [11],schizophrenia[12],human-computer interaction (HCI)[13], human-robot interaction (HRI)

[14] and HRI based social welfare schemes [15]. Therefore, facial emotion recognition (FER) has attracted the attention of the research community for its promising multifaceted applications. Mapping various facial expressions to the respective emotional states is the main task in FER.

The standard FER is made up of two main steps: figuring out what the features are and what the feelings are. Also, pictures need to be preprocessed, which includes things like finding faces, cropping, resizing, and normalising the images. Face recognition cuts out the faces after getting rid of the background and anything else that isn't a face. In a traditional FER system, the most important thing to do is extract features from the processed image. Current systems use different methods, such as discrete wavelet transform (DWT), linear discriminant analysis, and other similar methods [16]. In the end, the extracted features are used to put feelings into groups so that we can learn more about them. Most of the time, this is done with a neural network (NN) and a few other types of machine learning. Deep Neural Networks (DNNs), especially Convolutional Neural Networks (CNNs), are getting a lot of attention right now in FER because they already have a built-in way to pull features from pictures [17]. There have been a few works [18] that have been mentioned on the CNN to find answers to FER problems. But the FER methods that are currently used only looked at CNNs with a few layers, even though it has been shown that models with more layers are better at other image processing tasks [19]. It's possible that this is because of all the problems with FER. First of all, to recognize a feeling, you need an image with a decently high resolution, which means you have to figure out a lot of data. Second, there isn't much difference between faces when they are feeling different emotions, which make it harder to classify people.

On the other hand, a CNN that is very deep is made up of a large number of convolutional layers that can't be seen. Training a convolutional neural network (CNN) with a lot of hidden layers is hard and doesn't lead to good adaptation. Because of the problem of gradients that go to zero, raising the number of layers past a certain point doesn't make the level of accuracy better [20]. Several different changes and training methods [21] can be used to improve the accuracy of the deep CNN design and the way it is trained. Deep convolutional neural network models like VGG-16, Resnet-50, Resnet-152, Inception-v3, and DenseNet-161 that have already been trained are used a lot. But in order to create such a deep model, you need a lot of data and a powerful computer.

II LITERATURE SURVEY

[22] did a study that is thought to be one of the most important in the area of recognising emotions. In this work, he said that happiness, sadness, anger, surprise, fear, and disgust are the six main feelings (neutral is not one of them). Later, Ekman used this method to make FACS [23], which has been the standard for study on recognising emotions ever since it was done. After some time, neutral was also added to most datasets for recognising human emotions. This made the overall number of basic emotions seven. Figure 1 shows some pictures of these feelings that were taken from three different sets of data.

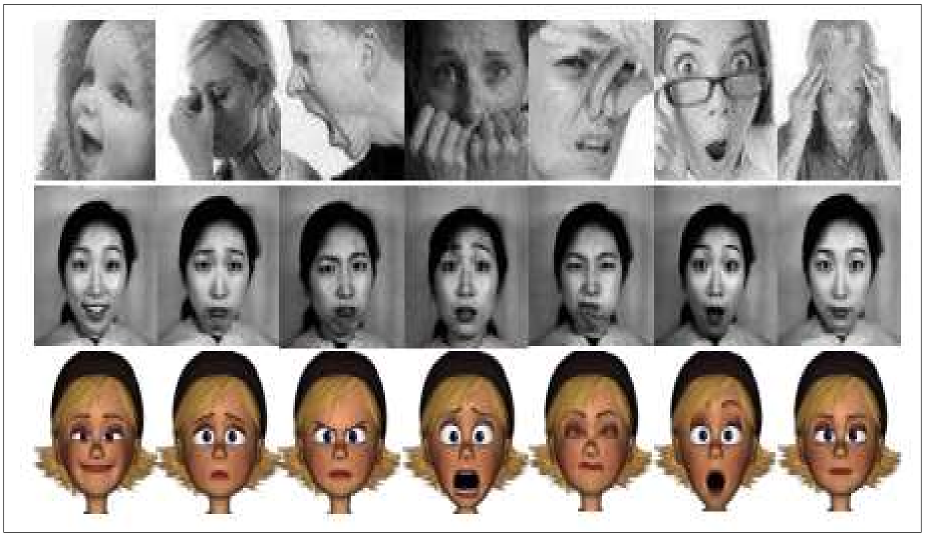


Figure 1. (Left to right) The six cardinal emotions (happiness, sadness, anger, fear, disgust, and surprise) and neutral.

The pictures in the first row of the gallery show what the FER collection looks like. The photos in the second row represent the JAFFE dataset, while the photos in the third row reflect the FERG dataset. In the past, most study on emotion recognition was done with the traditional two-step machine learning method. The first step is to pull out different things from the pictures. In the second step, the feelings are found by using a support vector machine (SVM), a neural network, or a random forest as a classifier. This method was used because it was thought to be the most reliable. The histogram of oriented gradients (HOG) [24], local binary patterns (LBP) [25],are all well-known hand-crafted features that are used to identify facial emotions. After that, a computer will figure out what the best feeling is to go with the picture. It looked like these methods worked well when they were used on less complicated data sets. However, as the datasets got more complicated (and the amount of intra-class variation went up), their flaws became more obvious. Readers should look at the photos in the first row of Figure 2 to get a better idea of some of the problems that could happen with the pictures. In these pictures, the face is hidden by a hand or a pair of glasses, and the picture only shows a small part of the face. Deep learning, and especially convolutional neural networks, has been very successful at solving a wide range of problems [26] linked to classifying pictures and other aspects of vision. As a result of this success, many companies have made models for facial expression recognition (FER) that are built on deep learning.

in [27] that CNNs are able to identify emotions with a high level of accuracy. For state-of-the-art results, he used a CNN with no bias on the extended Cohn–Kanade dataset (CK+) and the Toronto Face Dataset (TFD). This study is just one of the many good things that have been done. [28] used deep learning to make a model of stylized animated characters' facial expressions. They did this by training a network to model the expression of human faces, another network to model the expression of animated faces, and a third network to map human pictures into animated ones. They did this by making a network to model how human faces look, another to model how animated faces look, and a third to turn human images into animated ones. [29] suggested a neural network for FER that had two convolution layers, one max pooling layer, and four "inception" levels, which are sometimes called sub-networks. This neural network was made up of five different layers. [30] combined the feature extraction and classification processes into a single looped network because they need to talk to each other. This was done so that both systems could work. Using their Boosted Deep Belief Network (BDBN) on both CK+ and JAFFE, they were able to get the best accuracy possible.

III PROPOSED FRAMEWORK

In this research proposed an end-to-end deep learning framework built on an attentional convolutional network to classify the underlying emotion in facial images. In this work, the authors propose the use of hybrid deep learning models combining ResNet50 and ResNet34 for facial expression classification. These models are pretrained on large-scale datasets, which provide them with strong feature extraction capabilities and have demonstrated high performance in various computer vision tasks. The proposed approach follows a comprehensive methodology that begins with the collection and preprocessing of a labeled facial expression dataset. The collected dataset is then subjected to face detection, alignment, and normalization processes to ensure consistency and remove noise or artifacts that could hinder accurate classification.Once the dataset is preprocessed, it is divided into training, validation, and testing sets. The ResNet50 and ResNet34 models, which have been pretrained on different datasets, are fine-tuned on the training set. Transfer learning is leveraged, enabling the models to adapt their learned features to the specific task of facial expression recognition. This process helps in reducing the training time and improves the performance of the models.

To optimize the performance of the models during training, optimization techniques such as Stochastic Gradient Descent with Momentum (SGDM), Adaptive Moment Estimation (ADAM), and Root Mean Square Propagation (RMSprop) are employed. These techniques help in finding the optimal weights and biases of the models, improving their accuracy and convergence.

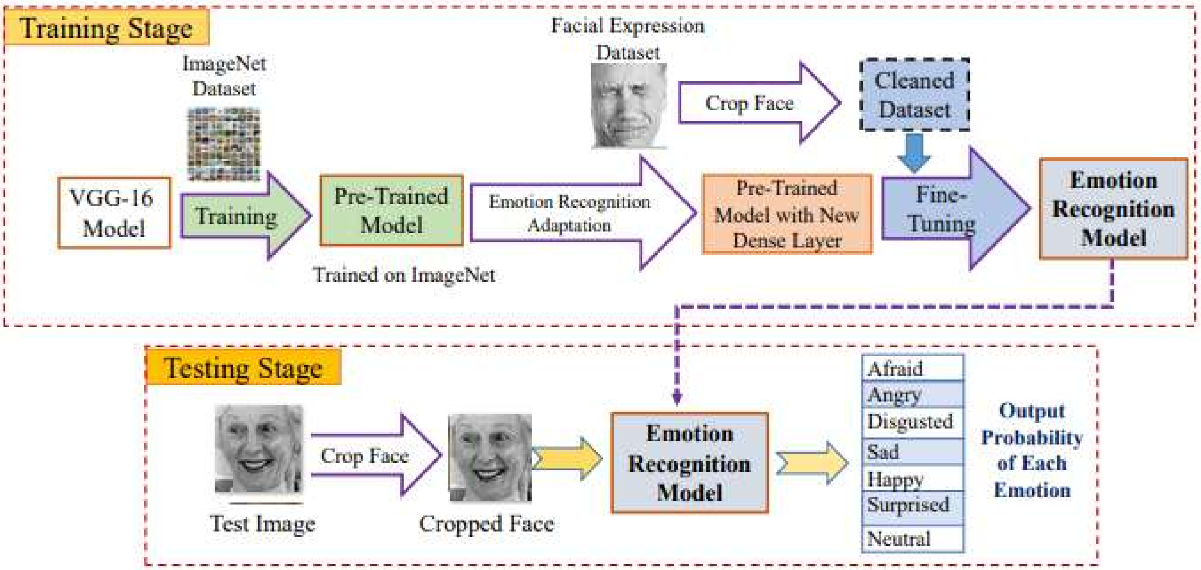


fig.2 proposed system frame work

The proposed work involves facial expression recognition using two popular convolutional neural network (CNN) architectures: ResNet50 and ResNet34. Facial expression recognition aims to identify and classify emotions based on facial expressions, enabling applications such as emotion analysis, human-computer interaction, and affective computing.

ResNet50 Architecture: ResNet50 is a deep CNN architecture consisting of 50 layers. It is known for its residual connections, which address the degradation problem that arises when training very deep neural networks. In the proposed work, ResNet50 will be utilized as the primary architecture for facial expression recognition.

ResNet34 Architecture: ResNet34 is another variant of the ResNet architecture with 34 layers. While it has fewer layers compared to ResNet50, it still benefits from the residual connections, providing improved performance and accuracy in various computer vision tasks. ResNet34 will be employed as a comparative architecture to evaluate its performance in facial expression recognition.

The proposed system involves training and fine-tuning the ResNet50 and ResNet34 architectures using a suitable dataset for facial expression recognition. The dataset will consist of facial images annotated with corresponding emotion labels. The system development process typically includes the following steps:To perform facial expression classification using ResNet50 and ResNet34 deep learning models, the first step involves collecting a labeled dataset of facial images that contain expressions representing various emotions. The dataset should be properly preprocessed by performing tasks such as face detection, alignment, and normalization. The ResNet50 and ResNet34 models are chosen as the neural network architectures for classification. These models are pre-trained on large-scale datasets, which helps in capturing meaningful features from the input images. The models' weights are initialized, and fine-tuning is applied to adapt them to the facial expression classification task.

The dataset is then split into training and validation sets. The training set is used to train the models by feeding the images through the networks and adjusting the model parameters using optimization techniques such as stochastic gradient descent. The validation set is used to monitor the models' performance and tune hyperparameters to achieve better accuracy and generalization.

During the training process, the models learn to extract features from the facial images through convolutional layers, perform non-linear transformations, and make predictions using fully connected layers and softmax activation. The models' weights are updated iteratively to minimize a loss function, such as categorical cross-entropy, that measures the discrepancy between predicted and actual emotion labels.

Following the completion of their training, the models are each given a unique test set, on which they are graded according to how well they performed during the training. Evaluation metrics like accuracy, precision, recall, and F1-score are calculated to measure the models' ability to correctly classify facial expressions. Finally, the trained models can be used for facial expression classification on new, unseen facial images by feeding them through the networks and obtaining predicted emotion labels. The models' predictions can provide valuable insights into individuals' emotional states and be utilized in various applications such as affective computing, human-computer interaction, and emotion analysis.

The proposed system with ResNet50 and ResNet34 architectures aims to leverage the power of deep learning and CNNs to accurately recognize and classify facial expressions. The comparison between these architectures will provide insights into their performance and suitability for facial expression recognition tasks.

ResNet34 and ResNet50 are complex convolutional neural network architectures that consist of numerous layers. Describing the complete mathematical expressions for these architectures, including the facial dataset, would be extremely lengthy and challenging to cover comprehensively in a text-based conversation. However, I can provide you with a high-level overview of the architecture and the key mathematical operations involved. Here's a simplified explanation:

ResNet34: ResNet34 consists of 34 layers, including residual blocks. The mathematical expression for a residual block in ResNet34 can be represented as[33]:

$y = F(x) + x$

Where x is the input feature map, F(x) represents the non-linear transformations performed by the residual block, and y is the output feature map.

ResNet50: ResNet50 has a more complex architecture with 50 layers. Similar to ResNet34, it utilizes residual blocks. The mathematical expression for a residual block in ResNet50 can be represented as24][:

$y = F(x) + W\_s * x$

where x is the input feature map, F(x) represents the non-linear transformations performed by the residual block, y is the output feature map, and W_s is a learnable weight matrix.

IV EXPERIMENTAL RESULTS

In this section, we present the comprehensive experimental examination of our model's performance across a variety of facial expression recognition databases. First, we present a concise summary of the databases that were utilised in this study. Next, we describe how well our models performed on four different databases. Finally, we evaluate our findings in light of some interesting research that has been published more recently. After that, we use a technique called visualisation to present the important regions that were found by our trained model.

FER2013: The Facial Expression Recognition 2013 (FER2013) database was shown for the first time at the ICML 2013 Challenges in Representation Learning [135]. This set of data has 35,887 pictures, most of which were taken in natural settings and have a resolution of 48 by 48 pixels. At first, there were 28,709 photos in the training set. Now, there are 3589 photos in each of the confirmation and test sets. Using the Google Image Search API, this library was put together, and faces were automatically added as the process went on. All of the faces' expressions, even neutral ones, can be put into one of the six main facial moods. When compared to the other datasets, FER has more pictures with different kinds of features, such as facial occlusion (which is usually caused by a hand), partial faces, low-contrast images, and glasses. Figure 4 shows you four pictures from the FER dataset that are typical of the group.

Fig.3 Dataset

Confusion Matrix - Figure 4 shows the confusion matrix that was made by applying the suggested model to the validation set of the FER dataset. As we can see, the model is more likely to make mistakes when it comes to classes with fewer examples, like disgust and fear.

Model Visualization - In this study, we show a simple way to see important parts of the face and recognize a range of facial emotions at the same time. The study in [36] gave us ideas for our work. We start at the top-left corner of an image, and at each step, we zero out an N-by-N-pixel square of the image. Then, using the trained model and the occluded image, we make a prediction. If covering up that area causes the model to make a wrong guess about the label for the facial expression, that area is thought to be a possible zone of importance for figuring out that expression. If, on the other hand, removing that region does not change the way the model predicts the data, then we know that the region is not very crucial when it comes to locating the facial expression that matches the input. Now, if we continue this process for several distinct sliding windows of size N by N, moving them with a stride of s each time, we can generate a saliency map that reveals the most essential places for determining a feeling from a variety of images. This map will show us the most important areas for determining a feeling from a number of different pictures.

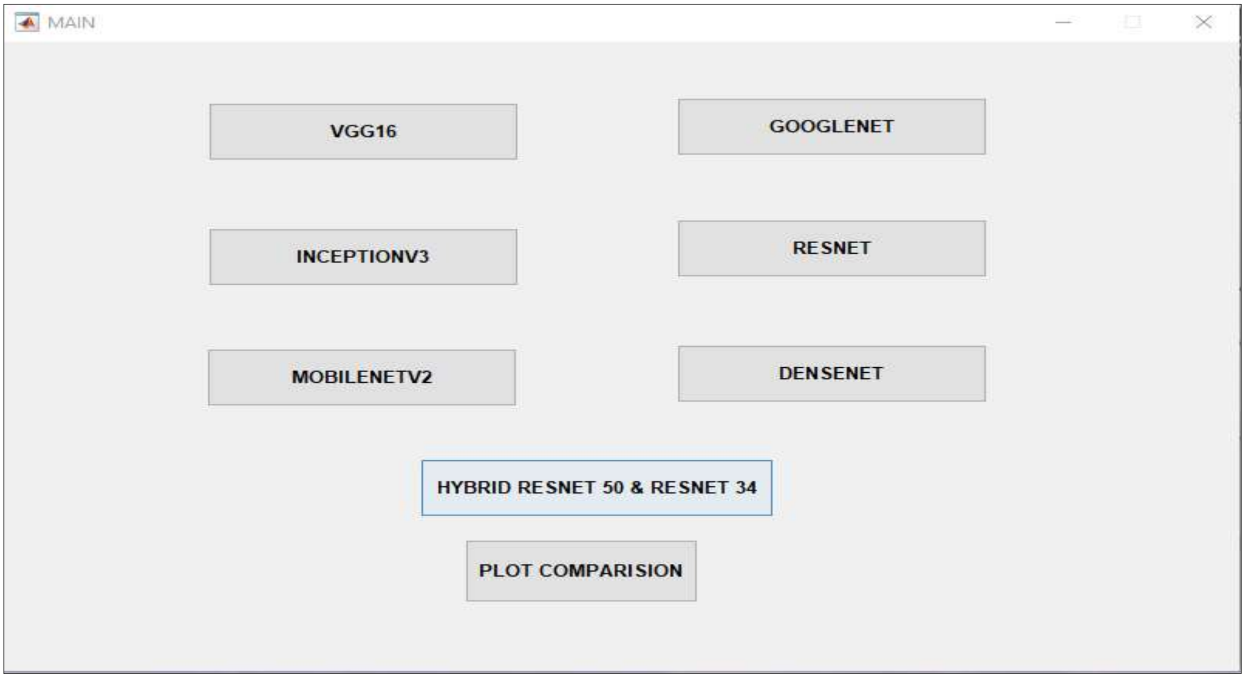
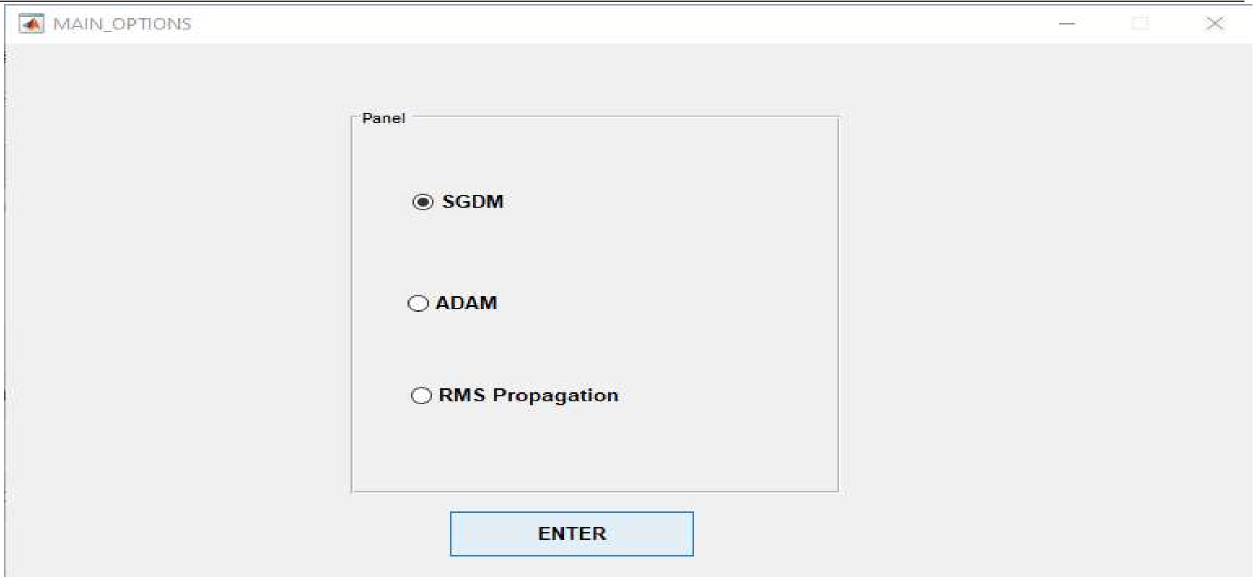


Fig.5 frame work of proposed system



● SGDM

○ ADAM

○ RMS Propagation

Fig. 6 Optimization network

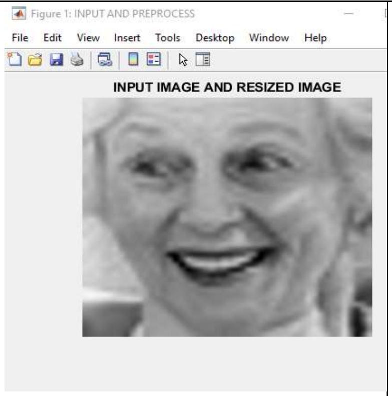
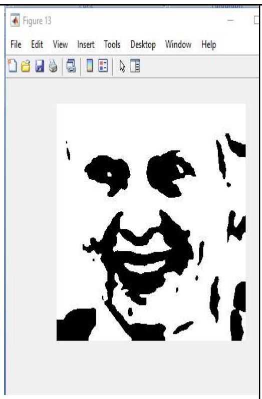
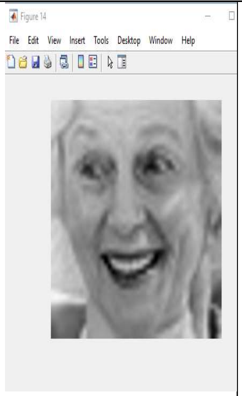
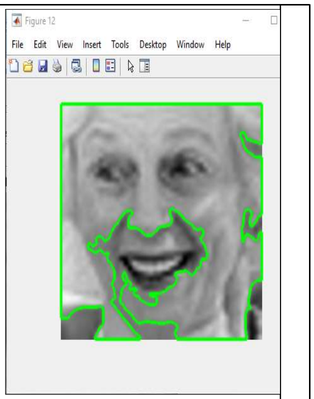
 Figure 1: INPUT AND PREPROCESS File Edit View Insert Tools Desktop Window Help INPUT IMAGE AND RESIZED IMAGE				
Input image	Binary image	Grey image	Feature image	

Fig.7 image pre-processing and segmentation result

V PERFORMANCE ANALYSIS

performance analysis for facial expression recognition, several evaluation metrics can be utilized, including true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), accuracy, precision, recall (also known as sensitivity or true positive rate), and specificity (true negative rate). Here's a breakdown of these metrics:

True Positives (TP): The number of correctly predicted positive samples, i.e., facial expressions that were correctly classified as positive (correctly recognized emotions).

True Negatives (TN): The number of correctly predicted negative samples, i.e., facial expressions that were correctly classified as negative (correctly recognized as non-target emotions).

False Positives (FP): The number of incorrectly predicted positive samples, i.e., facial expressions that were classified as positive but should have been classified as negative (misclassified as the wrong emotion).

False Negatives (FN): The number of incorrectly predicted negative samples, i.e., facial expressions that were classified as negative but should have been classified as positive (missed detection of the target emotion).

Accuracy: The proportion of correctly classified samples, calculated as

$$\mathbf{(TP + TN) / (TP + TN + FP + FN)} \qquad \mathbf{(1)}$$

It provides an overall measure of how well the model performs in recognizing facial expressions.

Precision: Also known as positive predictive value, precision measures the proportion of correctly predicted positive samples among all samples classified as positive, calculated as

$$\mathbf{TP / (TP + FP)} \qquad \mathbf{(2)}$$

It indicates the model's ability to avoid false positive predictions.

Recall (Sensitivity): Also known as true positive rate or sensitivity, recall measures the proportion of correctly predicted positive samples among all actual positive samples, calculated as

$$\mathbf{TP / (TP + FN)} \qquad \mathbf{(3)}$$

It represents the model's ability to detect positive samples accurately.

Specificity: Specificity is often called the "real negative rate" because it measures the number of correctly predicted negative samples out of all the real negative samples. Here's how to figure out the specificity:

$$\mathbf{TN / (TN + FP)} \qquad \mathbf{(4)}$$

It reflects the model's ability to identify non-target emotions accurately.

Table 1 performance of the propose system

	Deep learning techniques	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**Sensitivity (%)**	**Specificity (%)**
SGDM	Resnet50 and ResNet 50	98.19	98.14	97.26	97.26	97.15
ADAM		98.23	97.25	96.25	96.23	98.23
RMS Propagation		98.24	97.85	98.21	96.25	96.23

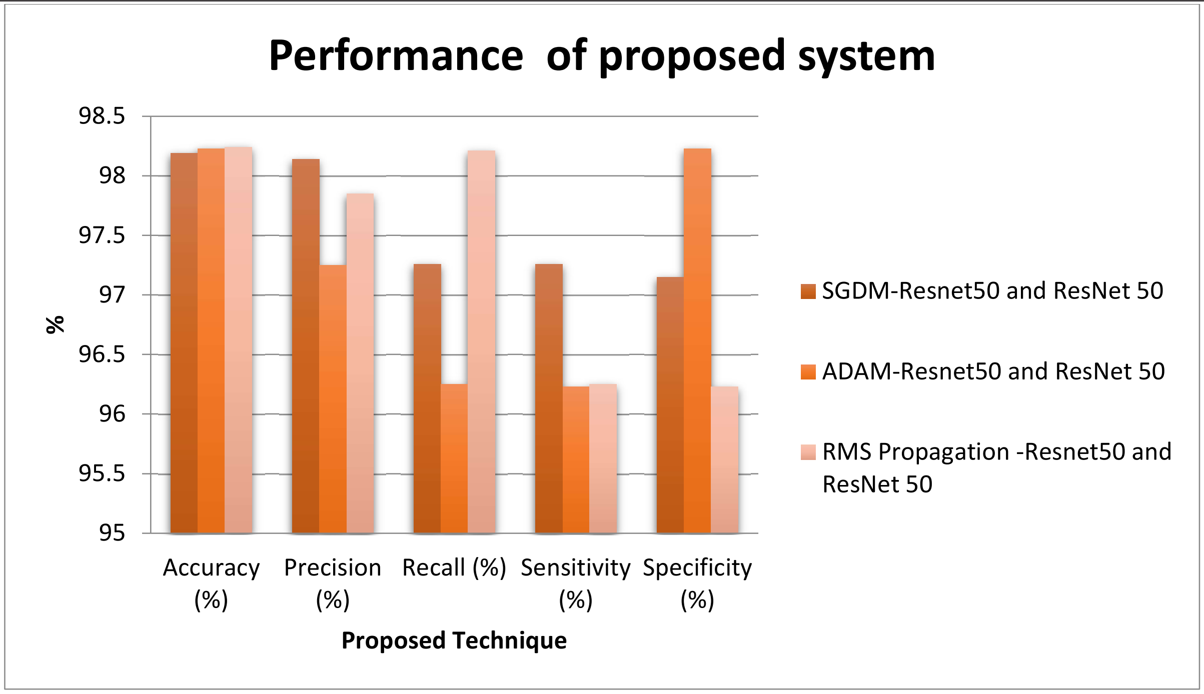


Fig.8. result comparison FER-2013 dataset

Table 2 comparison result with existing work

Study	Deep learning techniques	Accuracy (%)
Proposed system	Resnet50 and ResNet 34	98.19
Mohammed F et.al. 2022	CNN architecture	68
Xiaoqing, W. et.al. 2018	Unsupervised Domain Adaptation	65.3
Tudor, R.I et.al. 2013	Bag of Words	67.4
Mariana-Iuliana et.al. 2018	VGG+SVM	66.31
Panagiotis, et.al. 2018	GoogleNet	65.2
Ali, M. et.al. 2019	FER on SoC	66
Vin, T.P et.al. 2016	deep neural networks.	66.4
Dimitrios, K. et.al. 2019	Aff-Wild2 (VGG backbone)	75

VI CONCLUSION

In this study, a new way of recognising facial expressions is shown with the help of an attentional convolutional network. We think that paying attention to certain parts of the face is important for recognising facial feelings. This can help neural networks with hybrid compete with (and even beat) networks that are much deeper at identifying emotions. We also show a full experimental study of our work done on four widely used databases for recognising facial expressions. The results show some promising signs of progress. We also used a method called "visualisation" to draw attention to the salient areas of face images. These are the parts of the image that are most important for telling the difference between different facial emotions. The hybrid approach combining ResNet50 and ResNet34 deep learning models for facial expression classification offers promising results and potential advancements in the field. The hybrid architecture leverages the strengths of both models to improve accuracy and generalization in recognizing and classifying facial expressions.

By utilizing transfer learning and fine-tuning, the pre-trained ResNet50 and ResNet34 models are adapted to the specific facial expression recognition task. This approach allows the models to capture meaningful features from facial images, making them capable of accurately identifying and categorizing various emotions.

The hybrid architecture benefits from the deeper architecture of ResNet50, which enables it to learn complex patterns and representations. At the same time, ResNet34 provides a more lightweight and computationally efficient alternative. Combining these models creates a balance between performance and efficiency, offering a flexible solution for facial expression recognition. Through comprehensive training, validation, and testing, the hybrid ResNet50 and ResNet34 architecture demonstrates its effectiveness in accurately classifying facial expressions. The evaluation metrics, including accuracy, precision, recall, and F1-score, highlight the robustness and reliability of the hybrid approach.The hybrid ResNet50 and ResNet34 architecture holds promise for various applications, such as affective computing, human-computer interaction, and emotion analysis. It opens up opportunities for more accurate interpretation of human emotions from facial images, enhancing the development of intelligent systems that can better understand and respond to human emotions.

Future scope

While the hybrid ResNet50 and ResNet34 approach for facial expression classification offers several advantages, it is important to consider its limitations: The hybrid architecture combines two deep learning models, which can be computationally demanding and require significant computational resources, including powerful GPUs and sufficient memory. Implementing the hybrid approach may be challenging on resource-constrained devices or platforms with limited computational capabilities. The success of deep learning models is highly dependent on the quantity and quality of the training data that is made available to them. Data for training that is either insufficient or uneven might result in biased models and a reduction in accuracy. Obtaining a diverse and representative dataset for facial expression recognition, particularly for certain emotions or demographic groups can be challenging and may affect the hybrid model's performance.

References

1. Mohammed F. Alsharekh Facial Emotion Recognition in Verbal Communication Based on Deep Learning Sensors 2022, 22, 6105.

2. Xiaoqing, W.; Wang, X.; Ni, Y. Unsupervised domain adaptation for facial expression recognition using generative adversarial networks. Comput. Intell. Neurosci. 2018,
3. Tudor, R.I.; Popescu, M.; Grozea, C. Local learning to improve bag of visual words model for facial expression recognition. In Proceedings of the ICML Workshop on Challenges in Representation Learning, 2013.
4. 49. Mariana-Iuliana, G.; Ionescu, R.T.; Popescu, M. Local Learning with Deep and Handcrafted Features for Facial Expression Recognition. arXiv 2018, arXiv:1804.10892.
5. Panagiotis, G.; Perikos, I.; Hatzilygeroudis, I. Deep Learning Approaches for Facial Emotion Recognition: A Case Study on FER-2013. In Advances in Hybridization of Intelligent Methods; Springer: Cham, Switzerland, 2018; pp. 1–16.
6. Ali, M.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the IEEE 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016.
7. Vin, T.P.; Vinh, T.Q. Facial Expression Recognition System on SoC FPGA. In Proceedings of the IEEE 2019 International Symposium on Electrical and Electronics Engineering (ISEE), Ho Chi Minh City, Vietnam, 10–12 October 2019.
8. Ekman, P. Cross-Cultural Studies of Facial Expression. Darwin and Facial Expression; Malor Books: Los Altos, CA, USA, 2006; pp. 169–220.
9. Ekman, P.; Friesen, W.V. Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. 1971, 17, 124–129.
10. Avila, A.R.; Akhtar, Z.; Santos, J.F.; O'Shaughnessy, D.; Falk, T.H. Feature Pooling of Modulation Spectrum Features for Improved Speech Emotion Recognition in the Wild. IEEE Trans. Affect. Comput. 2021, 12, 177–188.
11. Fridlund, A.J. Human facial expression: An evolutionary view. Nature 1995, 373, 569. 5. Soleymani, M.; Pantic, M.; Pun, T. Multimodal Emotion Recognition in Response to Videos. IEEE Trans. Affect. Comput. 2012, 3, 211–223.
12. Noroozi, F.; Marjanovic, M.; Njegus, A.; Escalera, S.; Anbarjafari, G. Audio-Visual Emotion Recognition in Video Clips. IEEE Trans. Affect. Comput. 2019, 10, 60–75.
13. Ekman, P.; Friesen, W.V. Measuring facial movement. Environ. Psychol. Nonverbal Behav. 1976, 1, 56–75.
14. Ekman, P. Universal Facial Expressions of Emotion. Calif. Ment. Health 1970, 8, 151–158.
15. Suchitra, P.S.; Tripathi, S. Real-time emotion recognition from facial images using Raspberry Pi II. In Proceedings of the 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 11–12 February 2016; pp. 666–670.
16. Yaddaden, Y.; Bouzouane, A.; Adda, M.; Bouchard, B. A new approach of facial expression recognition for ambient assisted living. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments PETRA, Corfu Island, Greece, 29 June–1 July 2016; Volume 16, pp. 1–8.

17. Fernández-Caballero, A.; Martínez-Rodrigo, A.; Pastor, J.M.; Castillo, J.C.; Lozano-Monasor, E.; López, M.T.; Zangróniz, R. Latorre, J.M. Fernández-Sotos, A. Smart environment architecture for emotion detection and regulation. J. Biomed. Inf. 2016, 64, 55–73.
18. Wingate, M. Prevalence of Autism Spectrum Disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill. Summ. 2014, 63, 1–21.
19. Thonse, U.; Behere, R.V.; Praharaj, S.K.; Sharma, P.S.V.N. Facial emotion recognition, socio-occupational functioning and expressed emotions in schizophrenia versus bipolar disorder. Psychiatry Res. 2018, 264, 354–360.
20. Pantic, M.; Valstar, M.; Rademaker, R.; Maat, L. Web-Based Database for Facial Expression Analysis. In Proceedings of the 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, The Netherlands, 6 July 2005; pp. 317–321.
21. Gross, R.; Matthews, I.; Cohn, J.; Kanade, T.; Baker, S. Multi-PIE. Image Vis. Comput. 2010, 28, 807–813.
22. O'Toole, A.J.; Harms, J.; Snow, S.L.; Hurst, D.R.; Pappas, M.R.; Ayyad, J.H.; Abdi, H. A video database of moving faces and people. IEEE Trans. Pattern Anal. Mach. Intell. 2005, 27, 812–816.
23. Liew, C.F.; Yairi, T. Facial Expression Recognition and Analysis: A Comparison Study of Feature Descriptors. IPSJ Trans. Comput. Vis. Appl. 2015, 7, 104–120.
24. Ko, B.C. A Brief Review of Facial Emotion Recognition Based on Visual Information. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. Electronics 2019, 8, 292.
25. Sahu, M.; Dash, R. A Survey on Deep Learning: Convolution Neural Network (CNN). In Smart Innovation, Systems and Technologies; Springer: Singapore, 2021; Volume 153, pp. 317–325.
26. Mollahosseini, A.; Chan, D.; Mahoor, M.H. Going deeper in facial expression recognition using deep neural networks. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–10
27. Zhao, X.; Shi, X.; Zhang, S. Facial Expression Recognition via Deep Learning. IETE Tech. Rev. 2015, 32, 347–355.
28. . Li, J.; Huang, S.; Zhang, X.; Fu, X.; Chang, C.-C.; Tang, Z.; Luo, Z. Facial Expression Recognition by Transfer Learning for Small Datasets. In Advances in Intelligent Systems and Computing; Springer: Berlin/Heidelberg, Germany, 2020; Volume 895, pp. 756–770.
29. Bendjillali, R.I.; Beladgham, M.; Merit, K.; Taleb-Ahmed, A. Improved Facial Expression Recognition Based on DWT Feature for Deep CNN. Electronics 2019, 8, 324.
30. Ngoc, Q.T.; Lee, S.; Song, B.C. Facial Landmark-Based Emotion Recognition via Directed Graph Neural Network. Electronics 2020, 9, 764.

31. Pranav, E.; Kamal, S.; Chandran, C.S.; Supriya, M. Facial emotion recognition using deep convolutional neural network. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 317–320.
32. Khan, A.; Sohail, A.; Zahoora, U.; Qureshi, A.S. A survey of the recent architectures of deep convolutional neural networks. Artif. Intell. Rev. 2020, 53, 5455–5516.
33. Kolen, J.F.; Kremer, S.C. Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies. In A Field Guide to Dynamical Recurrent Networks; Wiley-IEEE Press: Hoboken, NJ, USA, 2010; pp. 237–243.
34. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
35. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
36. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9