Pertanika J. Sci. & Technol. 32 (5): 1963 - 1977 (2024)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Facial Emotion Recognition with Deep Neural Network: A Study of Visual Geometry Group-16 (VGG16) Technique with Data Augmentation for Improved Precision

Sarthak Kapaliya[1], Debabrata Swain[1], Ritu Sharma[2]*, Kanishka Varyani[2] and Jyoti Thakar[2]

[1]School of Technology, Pandit Deendayal Energy University Raisan, Gandhinagar- 382426, Gujarat, India
[2]School of Liberal Studies, Pandit Deendayal Energy University, Raisan, Gandhinagar- 382426, Gujarat, India

ABSTRACT

Emotions play a significant role in both verbal and nonverbal communication. Facial emotion recognition has applications in various sectors where we can get real-time feedback about student activeness by detecting their expression. In this paper, we aim to provide an improved deep-learning technique to detect emotions by using publicly available datasets to perform this detection. To get more data for the well-being of the Machine Learning Model, we have used data augmentation using the TensorFlow framework. Visual Geometry Group-16 (VGG16) is a convolutional neural network of 16 layers deep. There has been an alteration to the default VGG16 structure to get better classification results. Various optimization algorithms and loss functions increase the model's accuracy. We have used many evaluation parameters from the technical side, like precision, accuracy, recall, Area Under the Receiver Operating Characteristic Curve (AUC), and F1 Score. The proposed model has an accuracy of 89% while having a precision of 81 percent for classification. We have achieved an F1 Score of 0.42 and an area under the ROC curve (AUC) of 0.734. Overall, it would be beneficial for analyzing and categorizing positive and negative emotions, which would aid in detecting signs of stress, anxiety, and burnout, as well as taking preventive actions to enhance well-being.

Keywords: Deep learning, emotion, facial emotion, human-computer interaction, image classification, neural networks

ARTICLE INFO

Article history:
Received: 26 July 2023
Accepted: 10 January 2024
Published: 08 August 2024

DOI: https://doi.org/10.47836/pjst.32.5.02

E-mail addresses:
kapaliyasarthak@gmail.com (Sarthak Kapaliya)
Debabrata.Swain@sot.pdpu.ac.in (Debabrata Swain)
dr.sharmaritu@gmail.com (Ritu Sharma)
kanishkavaryani@gmail.com (Kanishka Varyani)
jyoti.thakar85@gmail.com (Jyoti Thakar)
* Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

INTRODUCTION

The last decade has seen a significant increase in the integration of emotions and technology. It has led to a multidisciplinary study of computer science, cognitive science, psychology, and neuroscience (Shank, 2014). Emotions play a significant role in both verbal and nonverbal communication in humans. It can be most effectively understood through facial affect, which refers to the feelings and emotions humans experience. Face expressions are widely regarded as the universal language of internal emotional states since people of all cultural backgrounds understand them, and they also help in understanding the moods of individuals. Facial experiences are regarded as a universal language to understand emotions and are recognized and interpreted similarly across all cultures. The universality hypothesis by Darwin (1872) suggests that humans can express six internal emotions, i.e., happiness, surprise, fear, disgust, anger, and sadness (Ekman, 2009).

Further research by Ekman and Friesen (1971) suggested a similar concept. He proposed that there are six universal primary emotions in humans: anger, disgust, fear, happiness, sadness, and surprise, and argued that these emotions are expressed with the same facial expressions regardless of culture or language. Therefore, the terms facial expressions or emotional expressions are often used to describe the various emotions and moods of individuals universally (Ortony, 2022).

While talking about facial emotions, the issues of universality and cultural specificity have often been considered (Barrett et al., 2019). There has been enough empirical evidence to prove that emotions in facial expressions are recognized similarly in different cultures. The facial expressions of basic emotions are considered innate and universal (Mandal, 2004). Facial expressions represent a person's state of mind, which informs other people how a person is feeling and is most useful when communicating with others. Facial expressions can be analyzed in various fields, such as psychology (Song, 2021).

Recently, with the growth of digital technologies, various social media platforms have introduced emoticons and animated graphics interchange formats representing emotions. These visual elements provide a more detailed and expressive way to communicate emotions and sentiments. Many tech companies like Microsoft Azure (2018) and Google Inc. have developed software to read emotions objectively by understanding facial expressions. Additionally, facial expressions, which can indicate emotional states, help identify and manage psychiatric illnesses. Treatment plans for people with Autism Spectrum Disorder (ASD) and other Neurological disorders also include emotional expression and facial recognition. A study recently observed that emotion recognition technologies teach children with ASD (Autism Spectrum Disorder) how to express and identify emotions (Garcia-Garcia et al., 2022). Rapid human-computer interaction development, such as Artificial intelligence (AI), Cloud computing, and various machine-driven learning, has enhanced their application in Psychology.

With the increase in suicide rates in India, with 12 deaths per 100,000 people reported in 2021 (Salve, 2022) and also over the world (https://ourworldindata.org/grapher/death-rate-from-suicides-ghe), this technology can be utilized by counselors to comprehend and understand the emotional state of their clients, especially in a virtual setting. By analyzing subtle changes in their facial expressions, counselors can tailor their approach toward their clients and provide more effective treatment.

Facial emotion recognition technology can identify potential threats and suspicious behavior in law enforcement and crime. By analyzing the emotions expressed in a suspect's facial expressions, new ways of identifying individuals involved in criminal activity can be evolved. It can also analyze the emotions expressed in surveillance footage, providing valuable insights into criminal behavior and aiding the investigation. In the field of forensics, it can be used to identify individuals in photographs when their faces are partially obscured, or images are of poor quality, supplementary to traditional methods of solving crimes.

Overall, facial emotion recognition technology has the potential to revolutionize a wide range of fields and applications. This technology can help improve outcomes and enhance our understanding of the complex nature of human emotions.

The environment today generates an enormous amount of data, which is accumulated to improve the situation in various settings such as education, organizations, research, and more. However, when a large amount of data is generated, it becomes crucial to remove subjectivity and inaccuracy in interpreting emotional cues to make the data more meaningful, personalized, and objective (Saini et al., 2023).

However, accurately interpreting emotional cues could be challenging. Subjectivity and inaccuracy can arise due to differences in cultural expressions, state of mind, and interpretation. For instance, people with social anxiety have been found to interpret facial expressions of emotions based on prior knowledge and expectations (Song et al., 2022).

Human-computer interaction can help to mitigate these issues by using facial recognition technology to recognize emotional cues accurately. The significance of facial recognition of emotions has been documented in various studies. The study done by Pabst et al. (2023) on facial recognition technology helped us to discover how healthy individuals and individuals with severe alcohol use disorder reacted to various socio-emotional cues.

Saini et al. (2023) recognized the importance of technology and media in communication. They conducted a study to compare three speech emotion recognition machine learning methods to determine feelings and sentiments in speeches. Studying emotions in the workplace becomes imperative, as an employee's state of affect can lead an organization to flourish or fail. Such technology can help take preventive measures by tracking employees' moods and emotional states (Subhashini & Niveditha, 2015).

Siam et al. (2022) used machine learning techniques to detect emotion. They used principal component analysis with basic models like support vector machine and logistic regression. The model achieved a validation accuracy of 70% on the face emotion dataset. Jaymon et al. (2021) proposed the detection of emotion on actual feeds that grasp the information about personalities. To build the model, they used TensorFlow, the framework of a Convolutional Neural Network, which gave them an accuracy of 65%

In addition, as the number of studies that concentrate on enhancing or increasing facial emotion detection has increased, the problem of a restricted database of photos for model training has also been brought to light (Devi & Preetha, 2023). Data augmentation, which increases the image count by modifying it at various levels, can increase the image count, thereby training the model for better precision and accuracy and contributing to refining human-computer interaction. It is accomplished by increasing the levels at which the data is modified.

This research highlights the need for extensive application of this technology. Language is full of emotions, which help in decoding the underlying meaning behind the intent of communication. Facial recognition of emotion in human-computer interaction is imperative to create more personalized, meaningful experiences for users in various industries. By incorporating facial recognition technology, we can better understand and respond to the user's emotional needs, leading to increased productivity, innovation, and well-being.

RESEARCH OBJECTIVES

The main contributions of the work can be summarized as follows:

1. A new fast and robust emotion detection framework for cyber-physical vision applications is proposed.
2. A comparison between traditional psychological and modern AI methods is evaluated to recognize human emotion.
3. Exploration of data augmentation to help us increase the data and introduce variability in the datasets.
4. Performing emotion classification using deep learning techniques.
5. A brief model implementation uses a live camera feed and images.

This approach system is based on Convolutional Neural Networks for facial expression recognition. A new architecture is proposed in which the system's input is an image; then, a Convolutional Neural Network (CNN) is used to predict the facial expression labels.

METHOD

Data Preprocessing System

Emotions are the most important part of a human being. Humans can recognize and differentiate between faces. Recognizing facial emotions and differentiating between them is believed to be achieved by computers nowadays. Recognizing facial expressions that communicate fundamental emotions like fear, happiness, and disgust is known as facial emotion recognition. A highly accurate emotion identification model has been developed thanks to the development of computer vision techniques.

Data Collection

We have collected different types of facial emotion data from many online dataset repositories. We have used the Facial Expression Recognition 2013 dataset for training purposes. It contains approximately 30000 facial images in Red, Blue, and Green (RGB) form of differential expression with a size restricted to 48 × 48 pixels. It contains mainly seven types of emotion labels: angry (0), disgusted (1), fear (2), Happy (3), Sad (4), Surprise (5), and Neutral (6).

We used different data for model testing. CK+48 is a small dataset containing seven classes: fear, sadness, anger, disgust, happiness, contempt, and surprise. The images are 48 × 48 and have a grey-scaled color palette. Good variations and feature distributions can be used in testing to obtain good results. It has a frontal view with clear images of faces.

Data Augmentation

Data augmentation (Alamsyah et al., 2022) is a group of methods for creating additional data points from previously collected data to artificially increase the amount of data. It includes making minor adjustments to the data or creating new data points using deep learning models. By creating new and varied examples, this has a wide range of applications for enhancing the performance and results of machine learning models.

We have used the Tensorflow framework for the deep learning model. Tensorflow provides an image preprocessing technique for data augmentation by generating batches of tensor images. We have done the following data augmentation operation:

1. Rotation: In this, we rotate the image to a certain degree. If the rotation degree is set to 40, the new image will be 40 degrees and rotate to the original one.
2. Shearing: It is also used to transform the orientation of the image. Additionally, it implies that the image will be warped along a particular axis, typically to alter or modify the perception angles.
3. Zooming: It allows us to either zoom in or zoom out. The specified zoom-in range allows us to get a different image, which can help train the ML model.
4. Flipping allows us to flip the orientation of the image. We can use horizontal or vertical flips. This operation can be misleading for a model. If the image is flipped along the wrong axis, it can make no sense during the training of the deep learning model. So, in face detection, we do not need vertical flips.
5. Rescale: We rescale the image pixel in the range 0 to 255.
6. Shifting: We shift the image by a certain length, making it different from the real image. It has a height and width shift.

For instance, here demonstrates how Data Augmentation changes the Face image. It creates different types of images similar to the original ones. It is applied to the whole dataset to increase the dataset. We can see the implementation of these techniques on an image, as shown in Figure 1.

Figure 1. Image augmentation

Data augmentation plays an important role in this research. By using augmentation techniques, we increase the number of data samples from 30,000 to 50,000, which helps achieve better accuracy. Many machine learning techniques include data augmentation, as it increases the amount of data and helps achieve more useful information from the images. It makes Data Augmentation necessary for image-based multi-classification. This technique has been utilized extensively for various purposes and is useful for resolving data generation and precision issues. Several techniques, including posture transfer, hairdo transfer, expression transfer, cosmetics transfer, and age transfer, have been suggested to change the appearance of a genuine face image. In the meantime, the simulated virtual faces can also be improved to match the realism of the genuine ones. One can increase the training data's variety and strengthen the trained model's resilience using any augmentation technique.

Model Training

A neural network with three or more layers is what machine learning essentially is, which includes deep learning. Although these neural networks try to emulate how the human brain works, they cannot match it, allowing the computer to "learn" from enormous amounts of data. Even if a neural network with only one layer can still provide approximate predictions, more hidden layers can help to tune and improve accuracy.

Here, we have used VGG16 as a deep-learning model. There are several reasons for the choice of VGG16 as a primary architecture for this study, as mentioned below:

1. Facial emotion involves capturing special details and unique features, which can be done using VGG16 hierarchical representations, capturing both low-level and high-level features in facial expressions.
2. VGG16 leverages the power of pre-trained weights using the ImageNet Large Scale Visual Recognition Challenge Dataset; this helps the model acquire general knowledge during training and enhances detection accuracy.
3. With transfer learning capabilities, VGG 16 uses small filter sizes and contains several convolutional layers, which provide a balanced receptive field. This receptive field helps the model capture local and global features from the images.

VGG16 is a convolutional neural network that is 16 layers deep. It is a pre-trained model that has been trained on the ImageNet database. The pre-trained model can categorize 1000 different types of objects. The network has, therefore, acquired rich feature representations for various images. The network can accept images up to 224 × 224 in size. This model has achieved 92% in the ImageNet Challenge for 14 million images belonging to 1000 classes. It has a fixed input size of 224 × 224 and RGB channels, resulting in (224, 224, 3) tensor. Here, it calculates the probabilities of different classes. After every prediction, we get probabilities associated with different classes based on similarity. The classification vector has to ensure that these probabilities add to 1 and check it. We use the Softmax function.

The 16 in VGG16 refers to 16 layers with weights (Figure 2). VGG16 has 13 convolutional layers, five max-pooling layers, and three dense layers, i.e., learnable parameters layer. It contains a 3x3 filter with Stride 1 and the same padding and max pool layer of a 2x2 filter with Stride 2. The convolutional and max pool layers are consistently arranged throughout the whole structure.

The VGG16 model is a deep convolutional neural network trained on the Image Net dataset. It consists of 13 convolutional layers and three fully connected layers. By using the VGG16 model as the base model, the pre-trained weights are used as initial weights for the model, which can improve performance and reduce training time. After the VGG16 base model, a dropout layer with a rate of 0.5 is added to prevent overfitting. Then, a flattened layer is added to flatten the output from the base model to a 1-dimensional tensor. A batch normalization layer is added to normalize the output. Then, three dense layers are added with 32 units, each followed by a batch normalization layer, the rectified linear activation unit (ReLU) activation function, and a dropout layer with a rate of 0.5. The ReLU activation function is used to introduce non-linearity into the model. Finally, a dense layer with seven units is added, and the softmax activation function is used to output a probability distribution over the seven classes.

This model is suitable for classification tasks on images with seven classes, and the use of a pre-trained VGG16 base model can lead to improved performance and faster

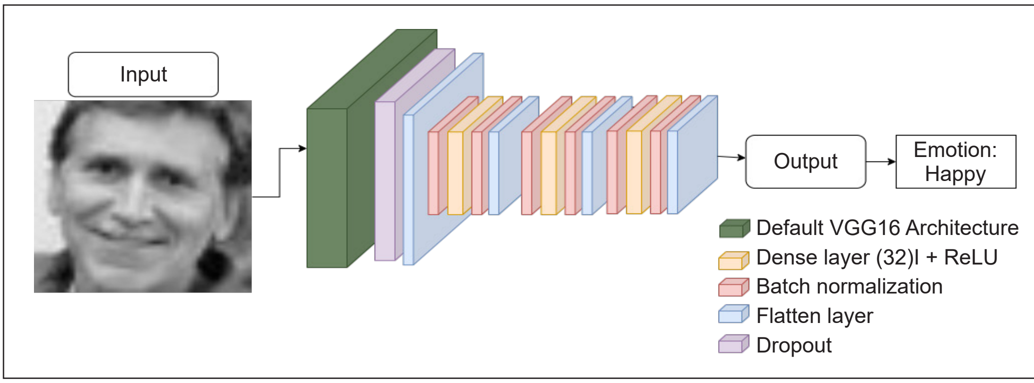

Figure 2. Improved VGG 16 Block Diagram
Source: Input image taken from the public dataset

training time. A good optimizing algorithm can help a deep learning model train by getting differences in results in minutes, hours, and days. In this case, we will apply the Adam optimization method. In addition to stochastic gradient descent, the Adam optimizer is used.

Adam Optimization

Adam optimization is a gradient descent algorithm for optimizing the parameters in a neural network. It uses momentum and adaptive learning rates to converge to the minimum of the cost function faster than traditional gradient descent methods.

The algorithm shows the working of Adam Optimization

Initialize $V_{dw}=0, S_{dw}=0, V_{db}=0, S_{db}=0$

On iteration t:

1. Computer dw, db using current mini-batch gradient descent
2. $V_{dw} = \beta_1 V_{dw} + (1- \beta_1)dw$, $V_{dw} = \beta_1 V_{db} + (1- \beta_1)db$
3. (momentum β_1 exponentially weighted average)
4. $S_{dw} = \beta_2 S_{db} + (1- \beta_2)dw^2$, $S_{db} = \beta_2 S_{db} + (1- \beta_2)db$
5. (RMSProp β_2)
6. $V_{dw}^{corr} = V_{dw}/(1- \beta_1^t)$, $V_{db}^{corr} = V_{db}/(1- \beta_1^t)$
7. $S_{dw}^{corr} = S_{dw}/(1- \beta_2^t)$, $S_{db}^{corr} = S_{db}/(1- \beta_2^t)$
8. $W = W - \alpha \ V_{dw}^{corr} / \sqrt{}(S_{dw}^{corr} + €)$,$b = b - \alpha V_{db}^{corr} / \sqrt{}(S_{db}^{corr} + €)$

Where Hyper parameter choices:

α: Learning rate. It needs to be tuned.

β_1: beta 1 moving average weight of dw. Default value 0.9

β_2: beta 2 moving average weight of dw^2 and db^2. Default value 0.999

€: 10^{-8}

The algorithm starts by initializing two variables, Vdw and Sdw, and Vdb and Sdb, to zero. On each iteration, the gradients dw and db are computed using the current mini-batch of data. Then, it updates the moving average of the first and second moments of the gradients (Vdw, Vdb, Sdw, Sdb) using exponential decay rates (beta1 and beta2). It also corrects these moving averages to prevent bias towards zero at the beginning of the training by dividing them by (1-beta1^t) and (1-beta2^t).

Finally, it updates W and B parameters using the corrected moving averages and a learning rate alpha. The learning rate determines the step size of the parameter updates and needs to be tuned. The beta1 and beta2 values have default values of 0.9 and 0.999, respectively. € is a small value added to the denominator to avoid division by zero.

These days, deep learning applications for computer vision and natural language processing are commonly regarded as dependable. Adam is a shortened version of adaptive moment estimation. Using estimated values for the first and second moments of the gradients, it calculates individual adaptive learning rates for various parameters. It uses momentum and sealing terms for the gradient cost function. Adam additionally uses the average of the second moments of the gradients, as opposed to adjusting the parameter learning rates based on the average first moment as in root mean squared propagation (RMSProp). It combines the idea of moment optimization with RMSProp and exponential decay.

How Adam Works

Adam optimization combines the Moment's and RMSP methods, which accelerate the gradient descent algorithm with exponentially weighted averages of the gradient.

Moment's Method

Generally, the main aim is to accelerate the gradient descent algorithm with an exponentially weighted gradient average. We use averages to converge faster toward minima, as shown in Equations 1 and 2.

$$w_{t+1} = w_t - \alpha_t . m_t \quad [1]$$

Where

$$m_t = \beta . m_t - 1 + (1 + \beta) \, [\, \partial L / \partial w_t] \quad [2]$$

Where m_t = total gradients at time t [present]. (mt initially equlas 0); $m_t - 1$ = the sum of gradients at time $t - 1$; w_t = weights at time t; $w_t + 1$ = weights at time $t + 1$; at = learning rate at time t; ∂L = derivative of Loss Function; ∂w_t = derivative of weights at time t; β = moving average parameter (constant = 0.9)

RMSP Method

It avoids the early stopping problem by accumulating recent gradients through exponential decay. Here, we take the exponential moving average, as shown in Equations 3 and 4.

$$v_t = \beta . v_{t-1} + (1 - \beta) * [\partial L / \partial w_t] \quad [3]$$

$$w_{t+1} = w_t - \alpha_t / (v_t + \epsilon) 1/2 * [\partial L / \partial w_t] \quad [4]$$

Where w_t = weights at time t; $w_t + 1$ = weights at time $t + 1$; a_t = learning rate at time t; $\#L$ = derivative of Loss Function; $\#w_t$ = derivative of weights at time t; v_t = sum of the square of past gradients. [i.e sum($\partial L / \partial Wt - 1$)] (initially, $Vt = 0$); β = moving average parameter (constant = 0.9); ϵ = a small positive constant (10 – 8)

Adam Optimizer inherits the strength or the positive attributes of the above two methods and builds upon them to give optimized results, as shown in Equation 5.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)[\left[\frac{dL}{dw_t}\right], v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left[\frac{dL}{dw_t}\right]^2 \quad [5]$$

Parameters used:

1. $\epsilon = 10 - 8$
2. $\beta_1 + \beta_2$ = average of gradients in the above methods ($\beta_1 = 0.9$ & $\beta_2 = 0.999$)
3. α = learning rate (0.001)

Cross Entropy Loss Function

The loss function measures how far the algorithm's current output deviates from the desired output. This information theory-derived function aims to compare two averages of the distribution's bit count. Cross-entropy is used as the Log Loss function to compare two probability distribution functions (not the same, but they measure the same thing).

We have employed binary and multiclass problems; categorical cross-entropy is utilized; the label must be encoded as a categorical, one-hot encoding representation (for three classes: [0, 1, 0], [1, 0, 0]).

Model Evaluation

The most crucial phase, model evaluation, allows us to assess and enhance our model. Validation and testing methodologies were the primary assessment criteria that we employed. Twenty-eight thousand photos from seven distinct classes/emotions were used to train the model. Eight thousand photos were used to validate the model as it was being improved. Only the Facial Expression Recognition (FER) 2013 dataset was utilized for training and validation. For testing, we used photos from the CK+48 dataset, which contains various images from the FER2013 data. The CK+48 data includes 981 pictures.

We used different evaluation metrics. We used five classification metrics: (1) Accuracy, (2) Precision, (3) Recall, (4) AUC, and (5) F1 Score.

RESULTS AND DISCUSSION

In this paper, we have proposed an Emotion detection model. Deep neural networks are used to precisely predict emotion from facial images. Features are extracted using deep learning methods. Classification metrics like Precision Recall F1 Score and Accuracy evaluate the effectiveness of the Deep learning Pre-trained model. We have used two different types of datasets that are available publicly. The FER2013 dataset contains over 30000 images, of which we have used 28000 for training the VGG16 Pre-Trained Model. The Rest of the Images were used to validate the model while training, which uses transfer learning methods. For the testing of the model, we used the CK+48 dataset. The model performs well in detecting all emotions provided in the dataset.

Instead of using cross-validation, we used a different strategy for dividing the data into training, validation, and testing sets. The dimensions and properties of the employed datasets impacted this choice. Due to the FER2013 dataset's size, we could assign a sizeable fraction for training and validation, creating a solid base for model learning. The CK+48 dataset, used only for testing, guarantees a fair assessment of the model's performance on omitted data. Even while cross-validation is a useful approach, its use may be impacted by computing limitations and dataset size, so we decided to use a partitioned training-validation-testing strategy in this situation.

We have evaluated it using the classification metrics. The graphs in Figure 3 represent the status of the metrics during the model training. We can see that validation Accuracy is greater than training accuracy for the 30 epochs. It shows that the model is not overfitting.

We have also practically implemented the model using the Open Source Computer Vision Library and Tensorflow. The model can detect facial emotion through live video and image feeds. Some of the implementation examples are shown in Figure 4.

The proposed model's classification accuracy is 89%, while its precision is 81% (Table 1). We have achieved an F1 Score of 0.42 and an AUC of 0.734.

The model performs well in terms of overall accuracy, categorizing dataset instances correctly 89% of the time, as shown in Table 2. When the model predicts a positive class, a precision of 81% shows a pretty good level of accuracy. A balanced performance in terms of precision and recall is indicated by the F1 Score of 0.42. The model's ability to differentiate between positive and negative classes is modest, according to the AUC of 0.734. Additionally, we are working on a research trend where the proposed approach is being implemented on hardware. Additionally, this problem can be resolved using additional machine learning techniques, including dictionary learning and semi-supervised learning.

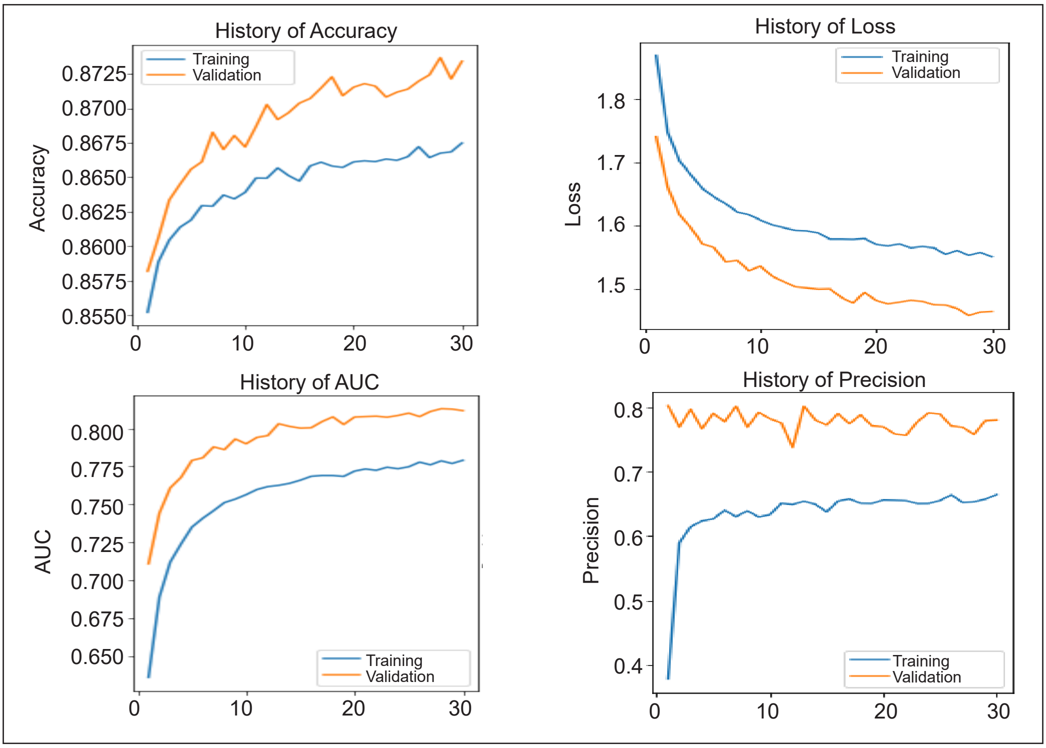

Figure 3. Graphical representation of model performance

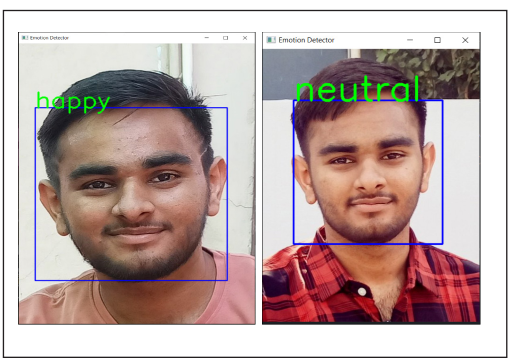

Figure 4. Test result of the model
Source: Personal images of the first author

Table 1
Result analysis of VGG16

Metrics	Score
Accuracy	89%
Precision	81
F1 Score	0.42
AUC	0.734

In Figure 3, the interaction of a sizable dataset and the intrinsic complexity of the VGG16 model may be responsible for the observed difference in loss between training and validation after 30 epochs. Given the complexity of VGG16, attaining convergence in a constrained number of epochs may present difficulties. However, various overfitting mitigation strategies, one of which is using regularization techniques like weight regularization and extra dropout layers to provide a more balanced output, can overcome this issue.

It is important to understand that assessing the model's performance only in terms of the loss graph might not give a complete picture. Despite the observed variation in loss, the model demonstrates

Table 2
Result comparison with our proposed results

Topic	Methodology	Framework	Validation Accuracy	Reference
Deploying machine learning techniques for human emotion detection. Computational Intelligence and Neuroscience	Reducing Dimension of Data Using Principal Component Analysis	Support Vector Machine, Logistic Regression	70%	Siam et al. (2022)
Real-time emotion detection using deep learning. 2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies	Convolutional Neural Network (TensorFlow)	CNN	65%	Jaymon et al. (2021)
Proposed VGG16 Model (our Result)	Applied Data Augmentation techniques on the dataset and evaluated the model using separate real-time test data.	Modified VGG16 Transfer Learning Model	89%	

promising classification results, achieving an overall precision of 81% and an AUC of 73%. These metrics affirm the model's efficacy in classification tasks. Another reason this solution has not been employed is that this research is computationally expensive. We need better computing resources to train a model using a better hyperparameter.

We did not employ certain alternatives, such as adding extra regularization, due to their high computational cost. We require more potent computational resources to improve the model's hyperparameters. Although the loss graph shows possible overfitting, we must weigh this against our actual practical constraints. The model still does well in categorization, demonstrating the importance of considering metrics and resource limitations when evaluating a system.

CONCLUSION AND THE WAY FORWARD

Emotions are an innate part of human beings. They direct and impact communication, relationships, and social distance. Emotions involve three distinct components: a subjective experience, a physiological response, and a behavioral or expressive response. The need to accurately decipher emotions has arisen with the advent and advancement of technology. The study mentioned above assists in doing so. It has collected data from different types of facial emotion expression data from many digital repositories of around 30,000 facial images. It has used the deep neural networks model VGG16 technique to decipher the seven emotional expressions: happiness, sadness, anger, fear, disgust, surprise, and neutral expressions. The data gathered was augmented to produce more data and enrich the modeling technique for more precision. The model has achieved 89% precision and 81% emotion classification.

This study will help to achieve better accuracy in decoding, comprehending, and classifying facial emotional expressions across various media such as images, media, and virtual reality environments and thus benefit several industries such as organization, education, entertainment, and research. Due to the high computational cost of certain alternatives, such as adding extra regularization, we did not employ them. We require more potent computational resources to improve the model's hyperparameters. It can be improved in the future with better resources. Therefore, we conclude that human emotions are innate and directly impact their lives. The new advanced technologies, such as virtual and augmented reality, may impose novel challenges to people's mental well-being, hence making it imperative to employ such approaches to detect emotional changes in a person's conduct. Overall, it would be beneficial to analyze and classify positive and negative emotions, which would help detect signs of stress, anxiety, and burnout and take preventive measures to enhance well-being.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to all the participants and stakeholders of Pandit Deendayal Energy University, India, for their support in this research work.

REFERENCES

Alamsyah, T. M. S. N., Abidin, T. F., Ferdhiana, R., Dirhamsyah, M., & Chaidir, M. (2022, December 8-9). *Analysis of face data augmentation in various poses for face recognition model*. [Paper presentation]. Seventh International Conference on Informatics and Computing (ICIC), Bali, Indonesia. https://doi.org/10.1109/ICIC56845.2022.10006997

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest, 20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray. https://doi.org/10.1037/10001-000

Devi, B., & Preetha, M. M. S. J. (2023). A descriptive survey on face emotion recognition techniques. *International Journal of Image and Graphics*, *23*(1), Article 2350008. https://doi.org/10.1142/S0219467823500080

Ekman, P. (2009). Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3449–3451. https://doi.org/10.1098/rstb.2009.0189

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377

Garcia-Garcia, J. M., Penichet, V. M. R., Lozano, M. D., & Fernando, A. (2022). Using emotion recognition technologies to teach children with autism spectrum disorder how to identify and express emotions. *Universal Access in the Information Society, 21*(4), 809–825. https://doi.org/10.1007/s10209-021-00818-y

Jaymon, N., Nagdeote, S., Yadav, A., & Rodrigues, R. (2021, February 19-20). *Real-time emotion detection using deep learning*. [Paper presentation]. International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India. https://doi.org/10.1109/ICAECT49130.2021.9392584

Mandal, M. K. (2004). *Emotion*. Affiliated East-West Press.

Ortony, A. (2022). Are all "basic emotions" emotions? A problem for the (basic) emotions construct. *Perspectives on Psychological Science, 17*(1), 41–61. https://doi.org/10.1177/1745691620985415

Pabst, A., Bollen, Z., Masson, N., Billaux, P., Timary, P. D., & Maurage, P. (2023). An eye-tracking study of biased attentional processing of emotional faces in severe alcohol use disorder. *Journal of Affective Disorders, 323*, 778–787. https://doi.org/10.1016/j.jad.2022.12.027

Salve, P. (2022, October 3). *India's suicide rate has increased. But is it because of better reporting or rising distress?* Croll.in. https://scroll.in/article/1034045/indias-suicide-rate-has-increased-but-is-it-because-of-better-reporting-or-rising-distress

Saini, A., Khaparde, A. R., Kumari, S., Shamsher, S., Joteeswaran, J., & Kadry, S. (2023). An investigation of machine learning techniques in speech emotion recognition. *Indonesian Journal of Electrical Engineering and Computer Science, 29*(2), Article 875. https://doi.org/10.11591/ijeecs.v29.i2.pp875-882

Shank, D. B. (2014). Technology and emotions. In J. E. Stets & J. H. Turner (Eds.), *Handbook of the sociology of emotions: Volume II* (pp. 511–528). Springer. https://doi.org/10.1007/978-94-017-9130-4_24

Siam, A. I., Soliman, N. F., Algarni, A. D., El-Samie, F. E. A., & Sedik, A. (2022). Deploying machine learning techniques for human emotion detection. *Computational Intelligence and Neuroscience, 2022*, Article 8032673. https://doi.org/10.1155/2022/8032673

Song, S., Zhao, S., Gao, Z., Lu, M., Zhang, M., Gao, S., & Zheng, Y. (2022). Influence of affective verbal context on emotional facial expression perception of social anxiety. *International Journal of Psychophysiology, 181*, 141–149. https://doi.org/10.1016/j.ijpsycho.2022.09.002

Song, Z. (2021). Facial expression emotion recognition model integrating philosophy and machine learning theory. *Frontiers in Psychology*, *12*, Article 759485. https://doi.org/10.3389/fpsyg.2021.759485

Subhashini, R., & Niveditha, P. R. (2015). Analyzing and detecting employee's emotions for amelioration of organizations. *Procedia Computer Science, 48*, 530–536. https://doi.org/10.1016/j.procs.2015.04.131