Journal of Computer Science

Original Research Paper

Advanced Facial Emotion Recognition Using DCNN-ELM: A Comprehensive Approach to Preprocessing, Feature Extraction and Performance Evaluation

[1]Boopalan K., [2]Satyajee Srivastava, [3]K. Kavitha, [4]D. Usha Rani, [5]K. Jayaram Kumar,
[6]M. V. Jagannatha Reddy and [7]V. Bhoopathy

[1]School of Computing, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
[2]Department CSE, M.M. Engineering College, Maharishi Markandeshwar (Deemed to Be University),
Mullana, Ambala, Haryana, India
[3]Department of CSE-AI&ML, GMR Institute of Technology, Rajam Andhra Pradesh, India
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India
[5]Department of Electronics and Communication Engineering, Aditya University, Surampalem, India
[6]Department of AIML, M S Engineering College, Navarathna Agrahara, Sadahalli, Bengaluru, India
[7]Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India

Article history
Received: 31-05-2024
Revised: 24-09-2024
Accepted: 07-10-2024

Corresponding Author:
V. Bhoopathy
Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupathi, India
Email: v.bhoopathy@gmail.com

Abstract: As a subfield of affective computing, Facial Emotion Recognition (FER) teaches computers to read people's facial expressions to determine their emotional state. Because facial expressions convey 55% of an individual's emotional and mental state in the whole range of face-to-face communication, Facial Emotion Recognition is crucial for connecting humans and computers. Improvements in the way computer systems (robotic systems) interact with or assist humans are another benefit of advancements in this area. Deep learning is key to the highly advanced research being conducted in this area. Recently, FER research has made use of Ekman's list of fundamental emotions as one of these models. Anger, Disgust, Fear, Happy, Sad, Surprise, and Neutral are the seven main emotions mapped out on Robert Plutchik's wheel. Opposite to each of the main emotions is its polar opposite. There are four steps to the suggested method: Preprocessing, feature extraction, model performance evaluation, and finalization. The preprocessing step makes use of the kernel filter. The proposed approach uses SWLDA for feature extraction. Facial Emotion Recognition (FER) is critical for improving human-computer interactions, particularly in educational settings. This study presents a novel hybrid approach combining Deep Convolutional Neural Networks (DCNN) with Extreme Learning Machines (ELM) to enhance emotion recognition accuracy. The proposed model demonstrates superior performance compared to traditional DCNN and standalone ELM approaches, offering real-time emotion detection in online learning environments. The effectiveness of the model is validated using publicly available datasets, setting a new benchmark for FER. This study makes major contributions to the field of Facial Emotion Recognition (FER) by offering a robust architecture that combines Deep Convolutional Neural Networks (DCNN) with Extreme Learning Machines (ELM). The methodology's efficacy is proven with publicly available datasets, establishing a new standard in FER, particularly in educational settings.

Keywords: Facial Emotion Recognition (FER), Linear Discriminant Analysis (LDA), Extreme Learning Machine (ELM), Deep Convolutional Neural Network (DCNN)

Introduction

Interactions between people allow for verbal, nonverbal, and emotional expression. Systems that can detect these kinds of things are in great demand across many sectors. It will be much easier for AI systems to connect with humans naturally if they can detect and react to human emotions. It has broad 6healthcare and

© 2025 Boopalan K., Satyajee Srivastava, K. Kavitha, D. Usha Rani, K. Jayaram Kumar, M. V. Jagannatha Reddy and V. Bhoopathy. This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

counseling applications. The method of presenting in an online learning system is dependent on the student's present level of proficiency. However, static emotion detection isn't always effective. It is vital to be able to understand the user's emotional state as it happens. Therefore, a model for real-time face emotion recognition is proposed in the study. Several real-time tasks are performed by facial emotion recognition systems. Additional layers of security could be achieved by integrating facial emotion recognition with existing technology. Part of having emotional intelligence is being able to read people's expressions on their faces. An individual outside look plays a significant role in two-way communication, as the term "interface" suggests. Observing and responding appropriately to nonverbal clues, such as facial expressions, can alter the trajectory of a discussion and even the intended meaning of spoken words, according to studies. For effective communication, the ability to recognize and interpret emotions is fundamental, since they govern up to 92% of the language used in a normal discussion. For that reason, this research aims to enhance computer emotion identification by analyzing data collected from sensors. Reading people's facial expressions has been one use of this experiment. The study of human emotions has attracted a huge number of researchers since its inception with Darwin's groundbreaking work. We all experience the same seven basic emotions. Contempt, anger, disgust, afraid, happiness, sadness, and fear are the seven basic emotions that can be discerned from a person's facial expressions. The learning process has been enhanced due to the increased accessibility of educational resources. This causes students to either drop out of the course completely or not complete it at all. That is why it is so important to monitor how actively students are participating in an online class. Consequently, there has been a significant uptick in research on solutions to the most current problems with online education. Having an emotional stake in what one is learning serves as a reliable surrogate for engagement level. There is a direct correlation between a student's emotional state and their learning performance. Students' focus and engagement in problem-solving tasks are enhanced when they are eager and engaged because they are better able to control their emotions. In contrast, students lose interest in their studies when they are bored or irritated since it drains their attention and cognitive resources. Therefore, at the present moment, the study focuses on the area of automated facial expression identification for online learners. Using an online learning environment to read students' emotions in real time is a huge challenge taking ownership of the oversight that was lacking in human supervision. It is possible to read the mood of a class in this setting just by looking at their faces. Physical muscle movements that are translated from emotional impulses are known as facial expressions. Some examples of these motions include raising the eyebrows, furrowing the forehead, and curling the lips. Computer vision and artificial intelligence researchers are making great strides in the field of face emotion recognition. Researchers have struggled to make progress on this topic because of its intricacy, which has persisted over many decades. The fields of emotion detection and recognition have substantial practical contributions as well. Both verbal and nonverbal cues can be used to convey a wide range of emotions. Examples of non-verbal ways of expressing emotion include a person's facial expressions and the tone of their voice, whilst examples of verbal ways include written words. The vast array of facial expressions used by humans can be broadly categorized into seven primary emotional states and fifteen more nuanced states. The basic emotions include happiness, sadness, surprise, anger, fear, disgust, and neutrality. In addition, angry surprise, horrified surprise, angry disgust, sad anger, disgusted surprise, and glad anger are all ways to communicate compound emotions. My mind conjures up feelings of shock, dread, rage, disgust, surprise and terror.

The potential for prejudice and misuse of face recognition technology has sparked demands for further research into the ethical and privacy consequences of these technologies. With the help of advanced image processing, an emotion recognition system can read a person's expressions on their face. As an example of a new technology that has gained traction, consider facial recognition software. Facial emotion recognition software is now one option for accessing a program that can analyze and understand a person's facial expressions. One of the features of this program that uses complex image distribution to simulate the human brain is emotion recognition. In order to combine them with other data, "Artificial Intelligence," or A.I., can identify and analyze various facial expressions. The ability to identify a person's emotional state has several potential uses for law enforcement, such as in interviews and investigations. The ability of facial emotion recognition technology to improve the year is remarkable. The use of emotion detection to learn how consumers feel about a company's goods, services, promotions, personnel, and in-store experiences is becoming increasingly common. Interpersonal interactions enable people to express themselves verbally, nonverbally, and emotionally. Systems that can recognize these expressions are in high demand in a variety of industries, including education, healthcare, and human-computer interaction. If AI systems can sense and respond to human emotions, they will be able to engage with humans more organically. Facial Emotion Recognition (FER) is important in this context since it improves the interaction between students and online learning systems. The technique of presentation in an online learning system depends on the student's current degree of ability. However, static emotion detection does not always work. It is critical to understand the user's emotional state as it unfolds. As a result, this study provides a model for facial expression identification that is specifically built for online learning environments. Facial emotion detection systems do several real-time activities, such as monitoring student involvement and delivering feedback to educators.

Literature Survey

An individual's emotional state can be better understood or protected with the help of human emotion detection. A two-pronged security system that can identify both emotional states and facial expressions may be required if this is viewed as an expansion of face detection. It's feasible to verify that the target is only a flat image of the person (Sowmya *et al*., 2024). Emotions have been the subject of much research, although the term has yet to be defined in a consensus among experts. Fatima *et al*. (2021) emotions can be seen as the outward expression of internal sentiments. Putting emotions aside, it may or may not be real. Fear, disdain, rejection, wrath, surprise, sadness, happiness, and neural are some of the human emotions. These emotions are so trivial. They are difficult to differentiate because their distinct expressions are the product of subtle shifts in the muscular contortions of their faces. Riyantoko *et al*. (2021) Since emotions are highly situational, it's possible for multiple people to express the same experience in various ways. Because they can learn from start to finish with physics-based modeling in real-time and other preprocessing methods, FER systems based on deep learning significantly reduce training time. Online education, whether at universities or training institutions, has rapidly increased in recent decades (Allen and Seaman, 2017). This presents new opportunities for FER. Online courses differ from traditional face-to-face courses in terms of constraint and communication, leading to faculty concerns (Shea *et al*., 2016). However, studies suggest that students' learning outcomes may be comparable to traditional face-to-face courses (Cason and Stiller, 2011), with the exception of skills that require optimal performance. It is undeniable that the rapid growth of online education can effectively provide convenience and flexibility for more students, so it has a large development space in the future; therefore, ensuring that students maintain the same level of concentration and learning efficiency as traditional courses during online education is critical to promoting the further development of online education (Dolan *et al*., 2015). According to Chen *et al*. (2012), employing Gabor wavelet and shape features, machine learning techniques such as linear SVM outperform other methods for distinguishing relevant and irrelevant facial expressions. Combining HAAR cascades with a neural network improves emotion detection (Yang *et al*., 2018a). Yang *et al*. (2018b) created a DNN model that uses vectorized facial characteristics. The vector representation of human face expressions enables high-accuracy DNN training. Convolution Neural Networks (CNN) outperform other sophisticated machine learning algorithms in terms of automated feature extraction, reduced input, and improved classification accuracy (Dachapally, 2017). Yu and Zhang, 2015) applied a nine-layer CNN to classify sentiment on the SFEW2.0 dataset. The convolutional network's output layer accurately classified emotions into seven categories using the SOFTMAX classifier at a rate of 61.29%. Lopes *et al*. (2015) used CK+ as their expression recognition dataset and utilized a CNN. First, the dataset was pre-processed to extract expression recognition-related features. After pre-processing the dataset, the network achieved 97.81% accuracy, leading to improved recognition accuracy and reduced training time. Deep learning algorithms automatically extract distinct features. Deep learning algorithms use a tiered architecture to represent data, with the final levels serving as high-level feature extractors and lower layers as low-level feature extractors (Fu *et al.*, 2014). Recurrent Convolution Networks (RCNs) (Donahue *et al.*, 2015) analyze temporal information by applying convolutional neural networks to video frames and feeding them to a Recurrent Neural Network (RNN).These models are effective for complicated concepts with limited training data but have difficulties when applied to deep networks. To address this issue, a model called DeXpression (Li *et al.*, 2018) was developed for strong face recognition. The system comprises two parallel feature extraction blocks with layers like convolution, pooling, and ReLU. This approach improves performance by combining numerous features rather than relying on single ones. Deep Belief Network (DBN) (Ebrahimi *et al.,* 2015) is a suggested graphical model based on unsupervised learning technologies such as autoencoders. When someone smiles broadly, for instance, their lips may expand in surprise, causing the lower half of the mask to rise while the upper half expands vertically. The research on Facial Emotion Recognition (FER) has evolved significantly, leveraging deep learning techniques to enhance the accuracy and efficiency of recognizing human emotions from facial expressions. In summary, the main contribution of this study is as follows. This study proposes a framework that combines existing online education platforms with a facial expression recognition model based on convolutional neural networks. This allows for real-time monitoring of students' emotions in online courses and timely feedback to teachers, allowing for flexible adjustment of teaching programs and ultimately improving the quality and efficiency of online learning. Some authors focused on obtaining temporal and geographical features from their models. Zhang *et al*. (2019) employed deep and classical learning to address

FER in video sequences. They refer to their method as a hybrid deep learning model. Two Convolutional Neural Networks (CNNs) were utilized to extract spatial and temporal data. The features were then included in the deep belief network. After average pooling, the linear support vector machine was used to classify the results. Experiments were carried out using BAUM-1s, RML, and MMI data sets. The authors claim to outperform state-of-the-art outcomes on these datasets. Kim *et al*. (2018) conducted research on emotion recognition using a dimensional framework. The authors used more than simply a person's face to predict underlying emotions in the valence-arousal space. Researchers discovered that an image's background can predict its overall feeling. The model used was a feedforward deep neural network. Lee *et al*. (2020) developed recurrent attention networks that predict emotions in the valence-arousal space using image color, depth, and thermal videos as inputs. The authors claim that their method can deliver cutting-edge results in dimensional FER on RECOLA, SEWA, and AFEW data sets. Early studies in FER focused on basic techniques such as feature extraction from images using handcrafted methods. However, the advent of deep learning has revolutionized the field, introducing methods that can automatically learn and extract features from raw image data. Deep Convolutional Neural Networks (DCNNs) have become a cornerstone in FER due to their ability to model complex patterns in image data. These networks consist of multiple layers that can learn hierarchical representations of facial features, making them highly effective for emotion detection. Studies like those by Chen *et al*. (2020); Mohan *et al*. (2021) have demonstrated the superior performance of DCNNs in FER tasks, showing significant improvements over traditional methods. The integration of DCNNs with Extreme Learning Machines (ELMs) has further advanced FER. ELMs are known for their fast training times and ability to handle large datasets efficiently. When combined with DCNNs, ELMs can enhance classification performance by providing robust and quick learning capabilities. This hybrid approach has been shown to outperform standalone deep learning models in several studies. Facial Emotion Recognition (FER) plays a crucial role in fields such as education, healthcare, and human-computer interaction, where understanding emotional states enhances communication. However, traditional models like DCNN and ELM face limitations in real-time accuracy and efficiency. To address these challenges, this study introduces a hybrid approach that combines the strengths of DCNN's deep feature extraction with ELM's rapid learning capabilities, significantly improving FER performance.

Proposed System

The goal of recent studies in human-computer interaction is to create an intuitive interface by taking the user's emotional state into account. As a result, humans would have a fighting chance and be useful in many areas, such as healthcare and education. Expressions, facial pictures, physiological signals, and neuroimaging approaches are only a few examples of how human emotions can define a variety of methods. The methodology involves four key stages: Preprocessing, feature extraction, model training, and performance evaluation. Preprocessing is conducted using a kernel filter to enhance image clarity by reducing noise. For feature extraction, Step Wise Linear Discriminant Analysis (SWLDA) is employed to select the most significant features, optimizing the data fed into the model. The proposed model, a DCNN-ELM hybrid, integrates the convolutional layers of DCNN for deep feature abstraction with ELM's fast learning capabilities, ensuring real-time emotion detection. Data consistency, the number of emotions seen, characteristics extracted, and the categorization method all play a role in accuracy. The following are some of the main points covered by many ideas on the methodology of emotional detection and modern research on emotions. As a result, researchers would be more motivated to learn about the physiological signs of the scientific community's present-day emotional awareness issues.

Here we will go over the method, which involves preprocessing, feature extraction, testing, and training; Fig. (1) shows the graphical representation of this framework. For preprocessing the proposed approach uses a kernel filter to remove noise and greyscale conversion. SWLDA is used for feature extraction. For training the model uses DCNN-ELM Model.

Dataset

The proposed approach makes use of publicly available datasets. This is one of the largest facial expression databases, with over a million photos sourced from the internet. These images were manually annotated to indicate the existence of seven different facial expressions: Joyful, sad, surprise, fear, disgust, rage, and contempt.

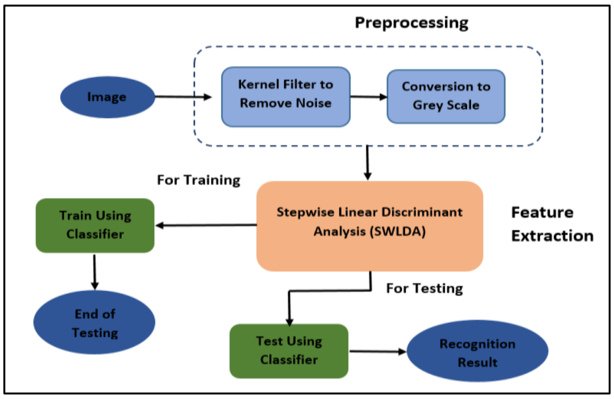


Fig. 1: A Model for recognizing facial emotions as proposed

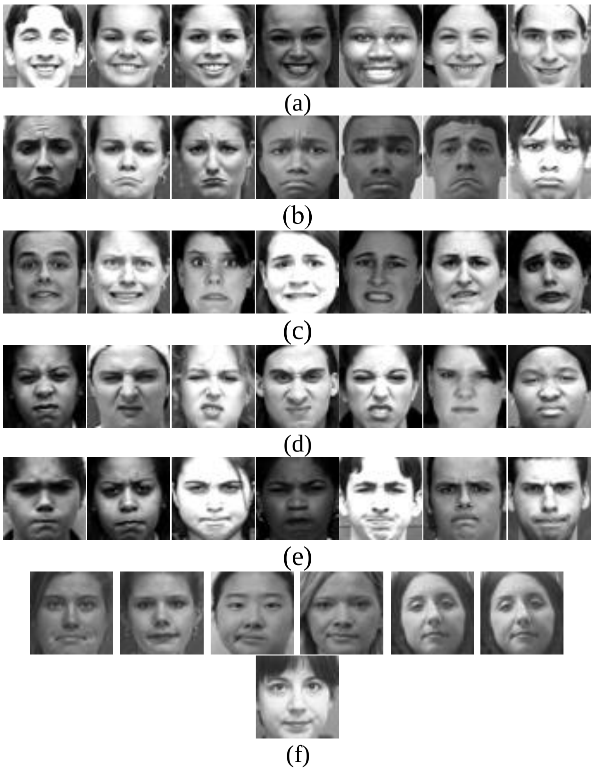

(a)
(b)
(c)
(d)
(e)
(f)

Fig. 2: Examples images from CK+ dataset (a) Happy (b) Sad (c) Surprise (d) Anger (e) Fear (f) Contempt

Table 1: Dataset description (%)

NAME	Size (no of images)	Image color	Size of image (in pixels)	Face variations types
CK+	593	Mostly gr11/18/2024ay	640×480	Happiness, sadness, surprise, Anger, Fear, disgust and contempt

The CK+ dataset, an extension of the CK dataset (Zhang *et al.,* 2019), is commonly used for facial expression recognition. Table (1) provides a summary of the CK+ dataset, whereas Fig. (2) displays representative images.

This dataset is frequently used to train and validate facial emotion recognition models. Some examples of datasets are depicted in Fig. (2).

FACS-coded emotion labels were applied to the peak frames of 327 image sequences, ranging from neutral to peak emotions for both posed and spontaneous expressions. The database has 123 individuals ranging in age from 18 to 50. 79% are female, 81% are European-American, 13% are Afro-American and 6% are of other races. This dataset is the most comprehensive yet, as it includes contempt among its eight facial expression classifications (Alok, 2024). To focus on only the 7 basic emotions in the CK+ dataset, we translated contempt labels to disgust since they are similar.

Preprocessing

One of the most important filters for face recognition in images is the kernel filter, which helps to flatten out edges, decrease dimensionality, and lessen blurriness in the image. Islam *et al.* (2018) This project uses the kernel filter to preprocess images. Edge detection, which predicts the ideal edge and decreases the dimensional space, is one of the most important kernel filter techniques. The three filter evaluations that make up the smoothening kernel are the median, average, and box filters. Adding a noise-removal filter and converting an RGB image to greyscale are the two main goals of the preprocessing procedure.

One of the most important kernel filter techniques is edge detection, which predicts the ideal edge and lowers the dimensional space. The three kinds of filter evaluations-box, average, Gaussian, and median-make up the smoothening kernel:

$$(o * n)(t, s) = \sum_{l=-\infty}^{l=\infty} \sum_{k=-\infty}^{\infty} o(l, k) n(t - l, s - k) \qquad (1)$$

An edge detection kernel will have three different operators: The Prewitt, Sobel, and Laplacian operators. The primary method for implementing the kernel filter on the input data is convolution. After that, it takes the mistake level into account and applies the appropriate filter to eliminate specific errors. Eliminating noise from the input image is a key function of the smoothening kernel. A low-pass filter is the simplest description of the Gaussian filter. As a rule, the input picture has some degree of Gaussian noise. To identify this kind of noise, we will use the white Gaussian noise filter. Images that are blurry on a large scale can also be helped by using the Gaussian noise filter. It accurately predicts the fixation of the image:

$$N(z, y) = {\frac{1}{2*3.14*\phi^{2^{p}}}}^{\frac{-(z^2+y^2)}{\beta^2}} \qquad (2)$$

The kernel filter's Gaussian mathematical expressions are shown in (1). One kind of noise-removal filter, the median filter, adds a little salt and pepper to the input picture to bring the noise down.

Feature Extraction

SW-LDA

FLD is the gold standard when it comes to finding the best hyper-plane to separate two classes. When the two classes have equal covariance and are Gaussian, FLD gives robust classification and is easy to compute. Improving the within-class variance to the between-class variance ratio yields the best projection or transformation in traditional LDA. The predicted feature weights are provided as follows and Fisher's linear discriminant and the ordinary least-squares regression solution are comparable for binary classification tasks like these:

$$\hat{x} = (W^{a}W)^{-1}W \tag{3}$$

Within which $\hat{x}$ is the vector of class labels and W is the matrix of observed feature vectors.

To build a multiple regression model with important features as the classifier, Stepwise Linear Discriminant Analysis (SWLDA) combines LDA with forward and backward regression for feature selection. Using SWLDA, one can choose appropriate predictor variables to incorporate into a multiple regression model. The suggested method for extracting localized features from data on facial expressions was found to be effective in this research. Stepwise regression, both forward and backward, was used in this method. The model was supplemented with the highest statistically significant predictor variable, which had a p-value less than 0.05 after it began with no starting model terms. Each time a new variable was added to the model, the least significant ones with p-values greater than 0.1 were removed using a backward stepwise regression. This procedure was carried out until either the number of terms in the model reached a certain threshold or until no more terms met the criteria for inclusion or exclusion. There was a cap of 100 features for the final discriminant function in this instance.

Materials and Methods

In order to run the experiment, a Windows 11 PC with a 32-core, 3.20 GHz Intel Core i7-8700 processor, 128 GB of RAM, and two NVIDIA GeForce GTX 3080 Ti graphics processing units (GPUs) was used.

SW Regression Methodology

If the conditions are beneficial, the stepwise regression approach will try to construct more W variables one at a time after selecting an equation with the best W variable. If you want to know which variable goes in first or second, you can use the partial F-test values to determine the order of addition and selection. Next, we compare the highest partial F-value to the F-to-enter value, which can be either selected or left as the default. The deletion process, also known as backward deletion, begins after the adding process, which is also known as forward entry. During this procedure, we computed the partial test values for every predictor variable that was already in the Queue. Afterward, the preselected or default significance level, F0, is compared to the lowest partial test value, OI. To begin the F-Test calculation process again, remove the variable UI if $\mathrm{OI} < \mathrm{O0}$ .Rely on the regression equation if $\mathrm{OI} > \mathrm{O0}$.

Classification of the Model

Extreme Learning Machine

SLFNs, or single-hidden-layer feedforward neural networks, were the original target of ELM's proposal. Compared to traditional techniques for learning neural networks, it's light years ahead. It finds the output weights analytically and chooses the hidden node parameters at random. In this way, the training is finished quickly and effectively, without the need for tedious iterations.

Using thirteen inputs to categorize the degree of heart stroke, our proposed classifier architecture is based on the standard database and five goal values of the disease category. Figure (3) shows the structure of the ELM classifier that is created. An ELM random feature space with L dimensions is used to map the input data and the network output is:

$$o_I(w) = \sum_{l=1}^{I} \varphi_l n_l\,(w) = m(w)\varphi \tag{4}$$

where, φ is equal to$[\varphi_1, \varphi_2, \ldots, \varphi_I]$. The output weights matrix is denoted by A and the function $h(x)$ = [b 1 (x),..., b L (x)] $m(w)$ = $[n_1(w), \ldots, n_I(w)]$ is defined. A represents the outputs of the hidden nodes for the input w. A hidden node's output is represented by $n_l(w)$. The ELM is able to obtain an approximation of the G training samples $(w_l,\ a_l)$ with zero error, indicating that:

$$M\varphi = A \tag{5}$$

The matrix of intended output is represented by $A = [a_1, a_2, \ldots a_G]^A$ and $M = [m^A(w_1), \ldots, m^A(w_G)]^A$. The regularised least squares approach can be used to determine the output weights in the following way:

$$\varphi = \left(\frac{L}{R} + M^A M\right)^{-1} M^A A \tag{6}$$

To improve generalization performance, the regularisation parameter, denoted as R, is utilized.

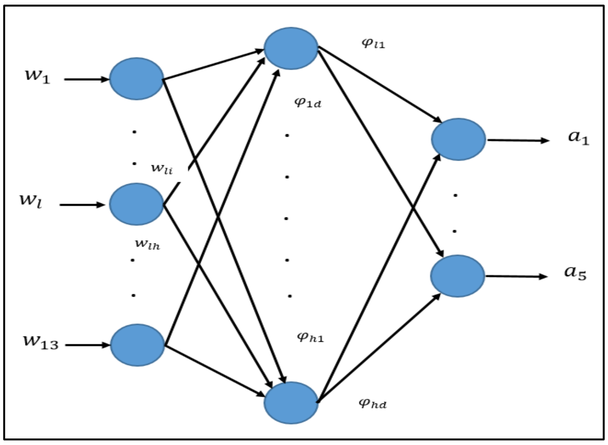


Fig. 3: The proposed ELM network

DCNN-ELM

DCNN-ELM integrates the speed of extreme learning with the feature abstraction capabilities of convolutional neural networks. A DCNN-ELM's structure is as follows: An input layer, an output layer, and many hidden layers, with each hidden layer alternating between a convolution layer and a pooling layer. Grouped together by convolution nodes are multiple feature maps that make up the convolution layer. While multiple feature maps use different input weights, they all share the same feature map. The network is made translationally invariant by using the square root pooling layer. Its feature map size and quantity are identical to those of the prior convolution layer. While each node in a convolutional layer's feature map is linked to every feature map in its preceding pooling layer, each node in a pooling layer's feature map is linked to just one equivalent feature map in its preceding convolutional layer. By using a stochastic pooling method, the feature maps of the final pooling layer are shrunk. Completely linked to the output layer it operates. Three factors were taken into account when designing this network. One important aspect of image classification jobs is the ability of the numerous hidden convolution and pooling layers to efficiently extract high-level features from input images. Secondly, our technique can learn to picture local correlations and handle image rotation invariance properly because of the shared local receptive weights. Thirdly, our method outperforms deep learning methods in terms of runtime thanks to ELM's batch training.

Algorithm 1: Proposed DCNN-ELM algorithm

Algorithm 1: DCNN-ELM CLASSIFIER
Input: $x \in (D1...n, D2.n)$
Output: Classification of the input image

```
ending epochs=100
initialize the neural network model with the parameter
while (training epochs < ending epochs) do
    for $D1$, $D2X$ in dataset do
        preprocessing image
        Feature Extraction
        Extracting feature from CNN by using $x = (W^a W)^{-1} W$
        Extracting feature from pooling layer by using $e_l = r_l \div \Sigma g \in C_k r_g$
        Extracting feature from fully connected layer by using $r_{l,k,j}(w) = \Sigma\Sigma\ w_{l+h-1,k+g-1}.\ t^j_{n,g}$
        ELM iamge classification by using $o_I(w) = \Sigma\varphi_l n_l(w) = m(w)\varphi$, $M\varphi = A$, $\varphi = (L \div R + MAM) -1MAM$
        if(testing data==classification result) then
            Classification image intro corresponding class
            Calculate the accuracy, computation time
        Else// incorrectly classified labels
            Updating gradient descent with backpropagation algorithm
        end if
    end for
end while
```

The fuzzy logic method adjusts the hidden node parameters of ELM t_l and s_l (input weights, biases, centers, and impact factors) during training. These parameters may also be allocated random values. The DCNN-ELM procedure's pseudo code is summarised as in Algorithm 1. In the first step of DCNN-ELM training, known as "convolution feature abstracting," the model builds high-level features layer by layer. The second step, "ELM classifier calculation," involves combining the features produced by the convolution and pooling layers into a vector. With the help of the regularised least squares approach, the output weights are computed analytically.

The Process of Creating and Aligning Local Weights

DC-ELM uses a continuous probability distribution to randomly create the input weights between the input layer and the first convolution layer, as well as the local weights between the pooling layer and the subsequent convolution layer. As a sampling distribution for input weights, we pick the Gaussian probability function in our paper.

Convolution

Following the feature maps generated by the prior pooling layer or the input image, the convolution layer applies the convolution process to extract features. For the first convolutional layer's convolutional node located at (l, s)on the jth feature map, the calculation is as:

$$r_{l,k,j}(w) = \sum_{h=1}^{c}\sum_{g=1}^{c} w_{l+h-1,k+g-1}.\, t_{n,g}^{j} \quad l, k = 1, \ldots (q - c + 1) \qquad (7)$$

The input image is represented by w. Nonlinear functions are not subject to the feature maps, in contrast to CNN. The map in the higher-level convolution layer is linked to all the feature maps in the previous pooling layer. To acquire the pooled feature map, we first calculate the convolution with each feature map using its local weights as (5). Then, we sum them all up.

Pooling

Two common methods, mean pooling and max pooling, are available for reducing the dimensionality of the last hidden layer, given that the size of the square root pooling map remains constant with the previous convolution layer. While max pooling chooses the biggest node, the former takes the mathematical mean of all nodes in each pooling region. When training deep convolutional networks, however, each of these pooling methods has its own set of limitations. Max pooling has a tendency to overfit the training set and impact generalization performance, whereas average pooling downweights strong activations and sometimes results in modest pooled responses. A stochastic pooling system was suggested as a solution to these shortcomings. Its stochastic character aids in preventing the overfitting problem while still providing the benefits of max pooling. When compared to mean and max pooling, stochastic pooling performs better in picture classification applications, according to the experimental data. The final pooling layer in this study is implemented using the stochastic pooling strategy. Using a multinomial distribution, the pooled maps are determined in stochastic pooling by sampling from each pooling zone.

To start, we normalize the values of the nodes in the preceding convolution layer so that we can calculate the probabilities for each pooling region (t, s):

$$e_l = \frac{r_l}{\sum_{g \in c_k} r_g} \tag{8}$$

Next, we select a location i inside the region by sampling from the multinomial distribution based on j. It follows that the value of the node at position $i(c_i)$ is the same as the pooled value r_i:

$$b_k = r_i \; where \; i \sim E\left(e_1, \ldots, e_{|C_k|}\right) \tag{9}$$

The pooled value is the element in the location i that we select from the multinomial distribution, which is $[1, 2, \ldots., 9]$. To illustrate, the value in the first grid of the pooled region would be 2.0, or the pooled value if $i = 1$. Choosing the biggest node for the pooling area is not always the best option. Therefore, stochastic pooling can depict regional multimodal distributions of convolution nodes. Another benefit of stochastic pooling over square root pooling is the significantly reduced size of the pooled feature map it produces. A hierarchical feature extractor, made up of all the alternate convolution and pooling layers, converts the original input images into high-level features, allowing for more efficient categorization. Also, DCNN-ELM is able to drastically cut down on computing load thanks to stochastic pooling, which means training time is significantly reduced.

Output Weights Calculation of ELM

The final step in feature generation is to create a row vector by joining the values of all the nodes in the stochastic pooling layer. The hidden layers output matrix $M \in C^{G.J.c_b^2}$ is obtained by combining the rows of Ninput samples, where c_b is the size of the feature map in the stochastic pooling layer. Next, we use the regularised least squares approach to analytically compute the output weights as:

$$\varphi = \begin{cases} M^A \left(\frac{L}{R} + MM^A\right)^{-1} A \; if \; G \leq J.c_b^2 \\ \left(\frac{L}{R} + M^A M\right)^{-1} M^A A \; if \; G > J.c_b^2 \end{cases} \tag{10}$$

The regularisation parameter F is used to manage the trade-off between the norm of output weights and the training error term, whereas A represents the labels of the input sample images. Improving the algorithm's generalization performance is possible with the right value for this parameter.

Results and Discussion

Among the most difficult and time-consuming aspects of interacting with people is emotion detection. Facial movements are a common and obvious way for people to communicate with one another. Facial expressions are the defining characteristic of non-verbal communication. Preprocessing, feature extraction, and classification approaches are the three main components of face emotion recognition research. In this study, we suggest comparing several Keras-based deep learning architectures for emotion identification with the use of deep facial features in images and neural networks trained on well-known pre-trained models, such as DCNN-ELM. The dataset is used to assess the models' performance. The combination of DCNN with ELM represents a novel technique to improve the speed and accuracy of emotion detection. The emphasis on online learning environments fills a current need in FER research by offering a realistic alternative for real-time emotion monitoring. The findings clearly demonstrate how the suggested model surpasses existing methodologies, giving value to the area.

Figure (4) shows the strengths and shortcomings of the technique used to detect facial emotions. With an accuracy of over 76% when taking diagonal factors into account and a classifier performance of 82%, it is possible to categorize all emotions. If we dissect the data, we find that out of 14 pictures depicting rage, the system correctly identified 10 as sad and 4 as angry. People think this is because tightly pursed lips are a universal sign of emotion, whether it's rage or sadness. Twelve out of fourteen pictures accurately depicted disgust, while the other two were misidentified as depicting melancholy. The similarity in mouth shape between the two emotions led to the misunderstanding. Out of 14 fear photos fed into the algorithm, 11 were correctly identified as a surprise. This is because the closeness of the eyebrows in both cases causes misunderstanding. While 12 of the 14 photos depicting happiness were accurately identified, the other 2 were mistakenly identified as anger due to the presence of teeth in both expressions. In a similar vein, the algorithm correctly identified 10 out of 14 photos depicting expressions of melancholy, while incorrectly identifying 4 as disgust. Both neutral and surprised expressions were correctly identified by the system.

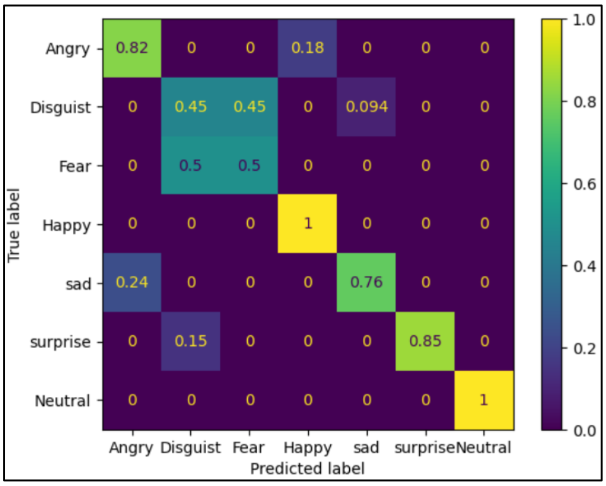
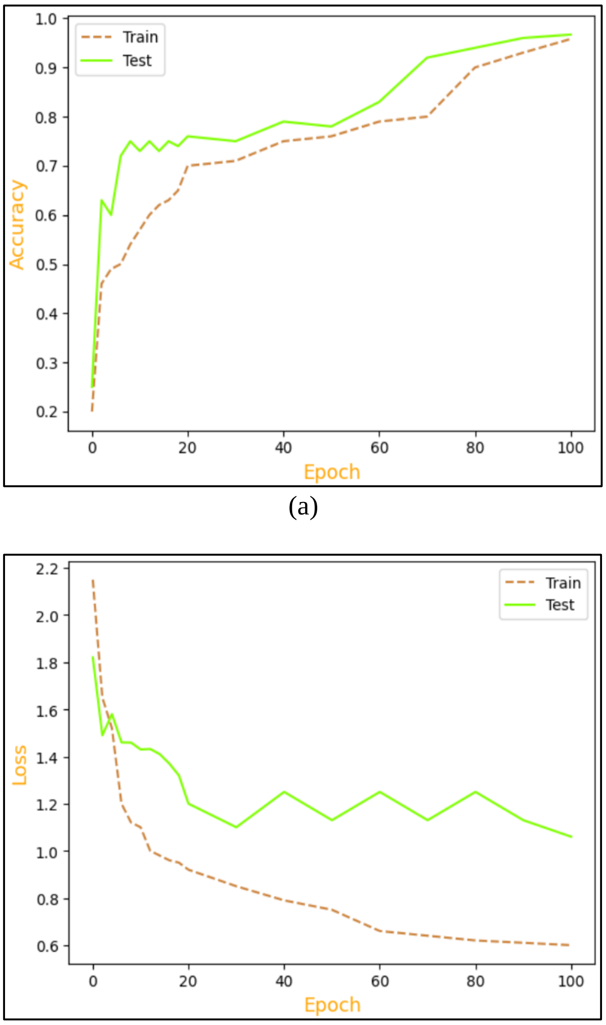


Fig. 4: Confusion matrix for the proposed model



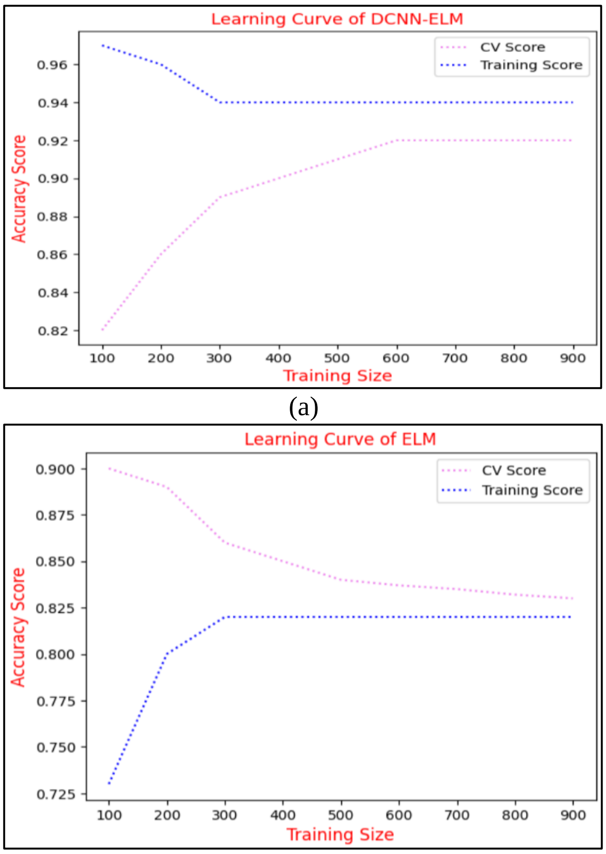
(a)

(b)

Fig. 5: Model accuracy and model loss for the classification using DCNN-ELM

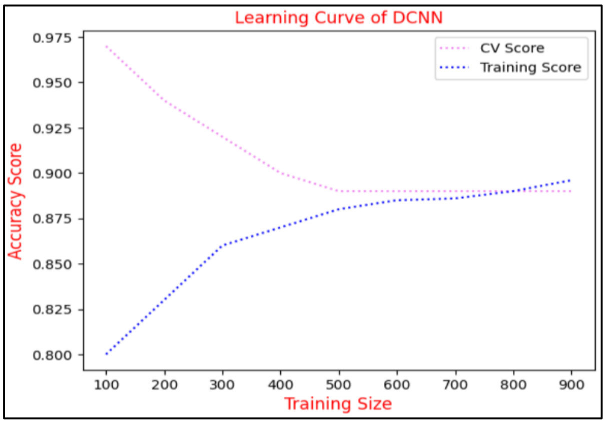
Figure (5) shows the main metrics that were monitored during the 100 epochs of model training: Training accuracy, testing accuracy, testing loss, and training loss. In each iteration, the training loss shrinks from 2.15 to 0.6 as the model refines its fit to the training data through stochastic gradient descent. The validation loss shows signs of generalization as well, although it's still significantly greater than expected, which could be due to overfitting. Both the training and testing accuracy levels reach above 95.72-95.64%, respectively, after 60 epochs. This suggests that further improvement in performance could be achieved with the addition of capacity or regularization. Overfitting starts to happen on the training set when the difference between the train and testing accuracy curves starts to widen around epoch 40. In sum, the results show that the model is learning to distinguish between different emotions based on facial expressions and other distinguishing characteristics.

Using eight classifiers to categorize facial expressions across datasets, we may assess how well the suggested model performs. Here are the hyperparameters used by each classifier. Using ten features retrieved from the dataset of photos, the classification is performed. Figure (6) shows the learning curves for the adopted classifiers, which demonstrate the behaviors and cross-validation scores for different training sizes.



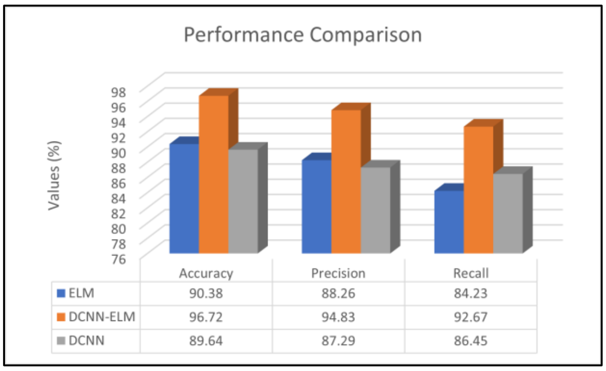
(a)

(b)



(c)

Fig. 6: The accuracy vs training size for different classifiers (ELM, DCNN-ELM, and DCNN) for facial expression recognition using the proposed model



	Accuracy	Precision	Recall
■ ELM	90.38	88.26	84.23
■ DCNN-ELM	96.72	94.83	92.67
■ DCNN	89.64	87.29	86.45

Fig. 7: Performance conparison of the models

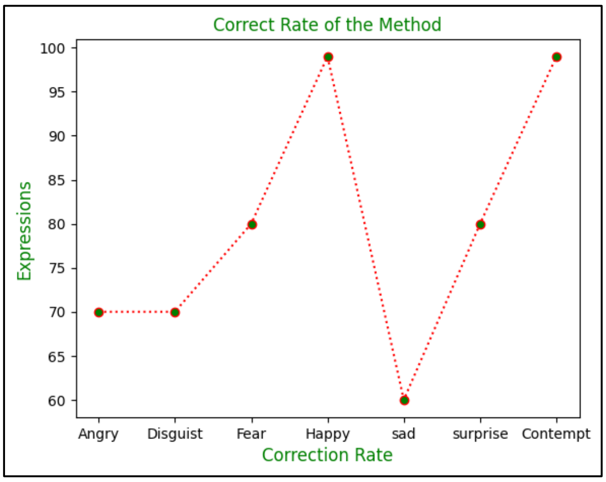


Fig. 8: The graph depicting the accuracy rate of the seven fundamental facial expressions

Figure (7) depicts the performance results for each of the three models presented. When compared to two other models, DCNN and ELM, our proposed model, DCNN-ELM, has the highest accuracy.

Figure (8) is a graph that shows the total rate of correction for the test dataset. A model's or system's accuracy in producing predictions or adjustments is represented by its correction rate. Here, it's probable that the test dataset includes some examples or samples that were omitted from the model's training. To evaluate the model's generalizability, the graph represents its performance on unseen data graphically. A higher correction rate usually means better performance and trends or oscillations in the curve could show how the model acts in different situations. You may learn a lot about the system's performance and where to find room for improvement by analyzing the graph. The results demonstrated that the proposed DCNN-ELM model significantly outperformed the conventional methods in recognizing facial emotions in an online learning environment. The real-time emotion recognition capability of the model was validated through live classroom sessions, where it successfully monitored and reported student engagement and emotional states. The DCNN-ELM hybrid model achieved a significant performance improvement, with an accuracy of 96.72% on the CK+ dataset. In comparison, traditional DCNN and ELM methods recorded lower accuracy, demonstrating the superiority of the hybrid approach. The model also excelled in real-time emotion detection, offering robust results across different emotional categories such as happiness, sadness, and anger.

Conclusion

It has always been easy for humans to read facial expressions and determine an individual's emotional state, but computer systems have a far harder time with this task. Emotion recognition from facial expressions is an area of social signal processing that finds use in several domains, especially in the realm of human-computer interaction. There have been a lot of studies on automatic emotion recognition and the majority of them have used a machine learning approach. However, basic emotion recognition remains a challenging area of computer vision research. This includes feelings like surprise, rage, joy, contempt, fear, and sadness. Emotion identification is only one of several practical problems that have recently attracted more attention to neural networksThe proposed DCNN-ELM model marks a significant advancement in facial emotion recognition, offering a robust solution for real-time emotion detection in online learning environments. By combining the deep feature extraction capabilities of DCNN with the efficiency of ELM, the model outperforms traditional approaches in both accuracy and speed. Future work could explore the integration of additional data modalities, such as voice and text, to further enhance the emotion detection system's robustness and applicability across various fields. The proposed approach outperforms well when compared with the other two traditional methods and produces an accuracy of about 96.72%.

Acknowledgment

The authors would like to thank anonymous reviewers for their constructive comments and suggestions to update the manuscript.

Funding Information

This research received no external funding.

Author's Contributions

All authors equally contributed to this study.

Ethics

This manuscript is an original work. The authors declare that there are no ethical concerns associated with this submission.

Conflict of Interest

The authors have no competing interests to declare relevant to this article's content.

References

Allen, I. E., & Seaman, J. (2017). Digital Compass Learning: Distance Education Enrollment Report 2017. *Babson Survey Research Group.*

Alok, S. (2024). *CK+ Dataset*. Kaggle. https://www.kaggle.com/datasets/shuvoalok/ck-dataset

Cason, C. L., & Stiller, J. (2011). Performance Outcomes of an Online First Aid and CPR Course for Laypersons. *Health Education Journal*, *70*(4), 458–467. https://doi.org/10.1177/0017896910379696

Chen, L., Zhou, C., & Shen, L. (2012). Facial Expression Recognition Based on SVM in E-learning. *IERI Procedia*, *2*, 781–787. https://doi.org/10.1016/j.ieri.2012.06.171

Chen, X., Li, D., Wang, P., & Yang, X. (2020). A Deep Convolutional Neural Network with Fuzzy Rough Sets for FER. *IEEE Access*, *8*, 2772–2779. https://doi.org/10.1109/access.2019.2960769

Dachapally, P. R. (2017). Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units. *ArXiv*, arXiv.1706.01509. https://doi.org/10.48550/arXiv.1706.01509

Dolan, E., Hancock, E., & Wareing, A. (2015). An Evaluation of Online Learning to Teach Practical Competencies in Undergraduate Health Science Students. *The Internet and Higher Education*, *24*, 21–25. https://doi.org/10.1016/j.iheduc.2014.09.003

Donahue, J., Hendricks, L. A., Guadarrama, S., Rohrbach, M., Venugopalan, S., Darrell, T., & Saenko, K. (2015). Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2625–2634. https://doi.org/10.1109/cvpr.2015.7298878

Ebrahimi K., S., Michalski, V., Konda, K., Memisevic, R., & Pal, C. (2015). Recurrent Neural Networks for Emotion Recognition in Video. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 467–474. https://doi.org/10.1145/2818346.2830596

Fatima, S. A., Kumar, A., & Raoof, S. S. (2021). Real Time Emotion Detection of Humans Using Mini-Xception Algorithm. *IOP Conference Series: Materials Science and Engineering*, *1042*(1), 012027. https://doi.org/10.1088/1757-899x/1042/1/012027

Fu, Z.-P., Zhang, Y.-N., & Hou, H.-Y. (2014). Survey of Deep Learning in Face Recognition. *2014 International Conference on Orange Technologies*, 5–8. https://doi.org/10.1109/ICOT.2014.6954663

Islam, B., Mahmud, F., & Hossain, A. (2018). Facial Region Segmentation Based Emotion Recognition Using Extreme Learning Machine. *2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, 1–4. https://doi.org/10.1109/icaeee.2018.8642990

Kim, H.-R., Kim, Y.-S., Kim, S. J., & Lee, I.-K. (2018). Building Emotional Machines: Recognizing Image Emotions Through Deep Neural Networks. *IEEE Transactions on Multimedia*, *20*(11), 2980–2992. https://doi.org/10.1109/tmm.2018.2827782

Lee, J., Kim, S., Kim, S., & Sohn, K. (2020). Multi-Modal Recurrent Attention Networks for Facial Expression Recognition. *IEEE Transactions on Image Processing*, *29*, 6977–6991. https://doi.org/10.1109/tip.2020.2996086

Li, J., Qiu, T., Wen, C., Xie, K., & Wen, F.-Q. (2018). Robust Face Recognition Using the Deep C2D-CNN Model Based on Decision-Level Fusion. *Sensors*, *18*(7), 2080. https://doi.org/10.3390/s18072080

Lopes, A. T., de Aguiar, E., & Oliveira-Santos, T. (2015). A Facial Expression Recognition System Using Convolutional Networks. *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, 273–280. https://doi.org/10.1109/sibgrapi.2015.14

Mohan, K., Seal, A., Krejcar, O., & Yazidi, A. (2021). FER-Net: Facial Expression Recognition Using Deep Neural Net. *Neural Computing and Applications*, *33*(15), 9125–9136. https://doi.org/10.1007/s00521-020-05676-y

Riyantoko, P. A., Sugiarto, & Hindrayani, K. M. (2021). Facial Emotion Detection Using Haar-Cascade Classifier and Convolutional Neural Networks. *Journal of Physics: Conference Series*, *1844*(1), 012004. https://doi.org/10.1088/1742-6596/1844/1/012004

Shea, P., Bidjerano, T., & Vickers, J. (2016). Faculty Attitudes Toward Online Learning: Failures and Successes. *SUNY Research Network*, 1–17

Sowmya, B. J., Meeradevi, Sini, A. A., Anita, K., Supreeth, S., Shruthi, G., & Rohith, S. (2024). Machine Learning Model for Emotion Detection and Recognition Using an Enhanced Convolutional Neural Network. *Journal of Integrated Science and Technology*, *12*(4), 1–10. https://doi.org/10.62110/sciencein.jist.2024.v12.786

Yang, D., Alsadoon, A., Prasad, P. W. C., Singh, A. K., & Elchouemi, A. (2018a). An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment. *Procedia Computer Science*, *125*, 2–10. https://doi.org/10.1016/j.procs.2017.12.003

Yang, G., Ortoneda, J. S. Y., & Saniie, J. (2018b). Emotion Recognition Using Deep Neural Network with Vectorized Facial Features. *2018 IEEE International Conference on Electro/Information Technology (EIT)*, 0318–0322. https://doi.org/10.1109/eit.2018.8500080

Yu, Z., & Zhang, C. (2015). Image Based Static Facial Expression Recognition with Multiple Deep Network Learning. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 435–442. https://doi.org/10.1145/2818346.2830595

Zhang, S., Pan, X., Cui, Y., Zhao, X., & Liu, L. (2019). Learning Affective Video Features for Facial Expression Recognition Via Hybrid Deep Learning. *IEEE Access*, *7*, 32297–32304. https://doi.org/10.1109/access.2019.2901521