Multimedia Tools and Applications (2023) 82:33835–33863
https://doi.org/10.1007/s11042-023-14543-6

Impact of lockdown on Generation-Z: a fuzzy based multimodal emotion recognition approach using CNN

Sirshendu Hore[1] · Tanmay Bhattacharya[2]

Received: 6 April 2021 / Revised: 15 November 2021 / Accepted: 31 January 2023 /
Published online: 8 March 2023

Abstract

The primary direction of most of the research done so far on the effects of Lockdown due to pandemic have been limited to areas such as clinical studies, possible impact on the global economy, or issues related to migrant workers. However, during this period, little attempt has been made to understand the emotions of Generation Z, one of the prime victims of this pandemic. Members of this generation were born after 1996. So, most of them are studying in various schools, colleges, or universities. In the proposed work, the emotions of some students of an engineering college in West Bengal, India, have been analyzed. A multimodal approach has been applied to obtain vivid pictures of 74 students' minds. The valence-arousal inspired Organize-Split-Fuse (OSF) model has been proposed to achieve this objective. Two conventional Convolutional Neural Network (CNN) models have been employed separately, to classify human emotions using Acoustic Information (AcI) and Facial Expressions (FE) from the generated dataset. The employed models have achieved satisfactory performance (*91% and 72.7% accuracy respectively)* on the benchmark dataset. Afterward, *classified* emotions have been organized and split successfully. Finally, a fuzzy rule-based classification system has been used to fuse both emotions at the decision level. The results show that junior students have higher positivity and less Neutral emotions than the senior. In-depth analysis shows that boys are more apprehensive than girls while girls have a more optimistic outlook for the future. The year-wise observations show the chaotic state of students' minds.

Keywords Acoustic information · CNN · Emotion · Facial expressions · Fuzzy · Generation-Z

✉ Sirshendu Hore
shirshendu.hore@gmail.com

Tanmay Bhattacharya
dr.tb1029@gmail.com

[1] Department of CSE, HETC, Hooghly, India

[2] Department of IT, Techno Main Salt Lake, Kolkata, India

1 Introduction

A lot of research work has been carried out to measure the impact of Lockdown on society as it has put the whole world into a turmoil state. So, it is at the center of everyone's mind. Although it is at the center of everyone's mind, the mentality and philosophy of each of our generations are not the same. This is why the effects of Lockdown have impacted each generation individually, changing them in various means [6, 16, 38, 54]. According to some experts, this lockdown crisis will have far-reaching and profound effects on Generation-Z. Experts are apprehensive that it will have a deep impact on Generation-Z minds and will last a long time [9, 65]. People are prone to emotions. Our conversation is full of different kinds of emotions. That is why in various studies, emotion has been given a lot of importance [12]. Philosophical studies advocate that Facial based expression (non-verbal) plays a big role in determining human emotions though it has some limitations [2, 5, 8, 13–15, 21, 24, 26, 28, 31–33, 36, 40, 42–44, 51–53, 55, 57, 58, 60, 69, 70]. Those studies also show that we can get an in-depth picture of human emotions from a spontaneous voice or Acoustic Information (verbal) [4, 10, 20, 22, 29, 35, 45, 47, 64, 67].

Earlier several studies have been conducted to determine the impact of Lockdown. The primary focus of those studies was to find some clinical solutions, measure the impact on the global economy, issues related to the migrant workers, etc. [6, 9, 16, 17, 38, 54, 65]. However, very little research has been done to analyze the emotions of Generation-Z. Therefore there is a clear research opportunity in the literature. It inspires us to determine Generation-Z emotions based on a multimodal approach, which combines facial expressions with acoustic information at the decision level using the Fuzzy rule-based techniques.

The advent of multimodal methods for determining human emotions has given a big thrust in this research area. In a multimodal approach, emotions are obtained from multiple sources such as facial expression, acoustic Information, body gesture, etc. [25, 30]. In this approach, decision-level fusion is a well-known practice that integrates emotions obtained from different sources [59]. The difficulty in decision-level fusion is that number of emotion labels defined in some widely used datasets is different and the types of labels are different as well. For example, FER2013 [19] is a well-known dataset used to study facial expressions. The number of labels defined in this dataset is seven. Another frequently used dataset for facial expression analysis is CK+ [36]. In the CK+ dataset, the number of labels is eight also the type of labels are not identical to what is defined in FER2013. RAVDESS, a popularly used dataset to study AcI-based emotions [34]. The number of emotion labels defined in RAVDESS is eight but the types of labels are not identical as defined in FER2013 or CK+.

To overcome the problem in the current study Organize-Split-Fuse (OSF) model has been proposed which is largely influenced by the valence-arousal model. In general, human emotions are bi-polar i.e. positive or negative, Watson [62]. Research works of Russell [50] suggest that all emotions can be represented by combining two things, one related to valence (negative/positive) and the other to arousal (intensity of valence). Similarly, Whissell [63] has represented emotions in terms of evaluation and activation. The objective of the work is to analyze the emotions of a section of Generation-Z during the Lockdown period, based on the proposed OSF model.

Problem Formulation: In the current study, the following steps have been adopted to satisfy the stated objective:

At the outset, the FE and AcI of each student have been separated from the generated dataset and stored separately.

Two separate pre-trained CNN-based models have been employed to classify FE and AcI based emotions for each student from the generated dataset

Classified emotions have been organized and stored separately based on their valence (negative or positive), using the first phase of the proposed OSF model i.e. Organize.

Afterward stored emotions are further sub-divided into six sub-classes of emotion namely High Negative (HN), Moderately Negative (MN), Low Negative (LN), Low Positive (LP), Moderately Positive (MP), and High Positive (HP) using the second phase of proposed OSF model i.e. Split.

Finally, these six classes of the emotion of two modes (FE and AcI) have been combined at the decision level i.e. Fuse, the last phase of the proposed OSF model. To achieve this fusion, the Fuzzy rule-based classification systems have been engaged [7].

Novelty: The proposed OSF model is effectively able to analyze students' emotions using the multimodal approach. The proposed model successfully fused obtained emotions at the decision level using a Fuzzy rule-based classification system. The samples used are spontaneous and natural in the generated dataset. Facial expressions have been extracted from Colour and B/W videos. The rest of the sections are organized as follows:

Some of the works which are relevant for this study have been discussed in the literature review part. The next section describes the methodologies adopted which are followed by the proposed methodology. The outcome of the study has been given in the form of results and discussion. Finally, the observation of the current study has been mentioned in the conclusion section.

2 Literature review

In this section, some contributions of the researchers related to FER, SER, Fuzzy, Multimodal, and Human emotions have been discussed. People use facial expressions as a non-verbal medium to express their emotions and stimuli [12, 57]. With the advancement of human-computer interaction, automatic facial expression analysis has gained a lot of popularity. Facial expression recognition (FER) is a very popular area of research. In a FER-based system, facial signs are converted into facial expressions. Ekman and Friesen [13] through their work showed that human beings express six basic emotions, anger, disgust, fear, happiness, sadness, and surprise, irrespective of their culture. This is known as the categorical model of FER. There are two major types of FER systems: Static image [33, 44] and Dynamic sequence [26, 69]. In the Static image FER system, features are extracted from a single image but in Dynamic sequence FER, the temporal relation among adjoining frames of facial expression sequence are considered. In the traditional FER methods, Shallow learning and handmade features [5, 31] were mostly used. All traditional dynamic sequence-based FER system follows three basic steps. In the first step, facial parts are being detected, cropped and other necessary things are being done [58, 60] to make the system robust. In the second phase, required features either texture [32, 40] or appearance [36, 51] based are being extracted, and finally, using these extracted features conventional machine learning classifiers like SVM, kNN, etc. [8, 28, 70] are used to classify these emotions. However, achieving the desired performance in real-life situations remains elusive. From 2013 onwards due to the availability of cheap with improved processing capabilities and with the introduction of the deep neural network, the accuracy level of the FER system has been improved significantly [21, 53]. In recent years, a large number of research papers have been published on automatic facial expression analysis [14]. For the stated purpose researchers haves employed various deep learning mechanism [2, 24, 52, 55]. The advantage of DNN is that unlike other traditional systems it can extract useful features more accurately [15]. Lately, Mohan et al. [42] have applied local gravitational force descriptors to identify human emotions in various challenging situations using an improved deep neural network model. In their work, geometric features were fused with the holistic feature using score level fusion mechanism. In the following year, Mohan et al. have introduced 'FER-net' based DNN to identify the FE [43]. In both of these works, authors have employed five widely used databases to evaluate the performance of their proposed work which achieved significant accuracy.

Research suggests that apart from the FER-based system we can also get people's emotions from acoustic information (SER). It is currently one of the hot topics in research and its presence is being felt in various sectors of our life. Researchers applied acoustic information to solve various real-life problems; Psychological Assessment [35] Human-Computer Interaction [10], Call Center [20], etc. Iqbal et al. show that acoustic information can be generated from speech in real-time. In their study, the authors have implemented the gradient boosting method to achieve their objective [22]. Pinto et al. employed Deep CNN to develop one emotion model that can understand people's emotions based on spoken language [47]. In their work author(s) has engaged RAVDESS as a dataset. To evaluate the performance of their model F1 score has been used as a metric. The weighted score achieved on the test data is 0.91. Zhang et al. [67] developed three shared models using RAVDESS as a dataset. The objective is to find the relationship between speech and songs. Their work suggests that although the process of recognition of speech and song are dissimilar, they have some relation and can be treated as the same. Mel Frequency Cepstral Coefficients (MFCCs) is a well-known non-parametric method used to extract features from acoustic information [4, 29, 45, 64]. Muda et al. [45] has successfully built one model to recognize acoustic information using MFCC techniques. In their study author(s) has used Dynamic Time Warping (DTW) to measure the testing patterns. Kuang et al. [29] used MFCC, STFT, and SIFT features to classify human emotions. In their proposed study, they employed RAVDESS as a dataset and Alexnet as a classifier that achieved significant accuracy (95.88%). Earlier in the year 1995, Lecun and Bengio [30] have suggested that CNN can be used to extract information from different sources such as images, speeches, etc. Of late researchers have combined facial expressions with acoustic information to get a complete spectrum of people's minds. In the year 2018, Jannat et al. successfully fused audio data with video data to get people's emotions [25]. In the very next year, Tzirakis et al. proposed an automatic affect recognition system that works in a real-world environment based on audiovisual signals [59]. In the case of a multimodal way of emotion recognition, the basic challenge is how to merge the emotions obtained from different modes. D.Zhang et al. [68] proposed that this can be achieved at the decision level where the emotions obtained from different modes are combined at the last level. Mohan et al. [42] have applied score level fusion at the final stage of the classification process to merge geometric and holistic facial features in a FER-based system using DCNN.

The Fuzzy rule-based classification system (FRBCs) is a well-known classification mechanism that can be applied at the decision level to, combine decisions obtained from different modes [7]. The FRBCs is based on the concept of Zadeh's Fuzzy principles [66]. Mohammadpour et al. [41] successfully classify coronary artery disease based on the FRBCs approach. In their work author(s) has developed 144 fuzzy rules to classify the disease into four classes. In this process author(s) has achieved 92.8% accuracy. The presence of Fuzzy principles and their implements are found during Lockdown [17]. The hybridization of Fuzzy is also felt in the image processing domain [61]. For the design of a multimodal recognition system involving facial, acoustic information, gesture, etc., a sufficient number of labeled training data with all possible variations of the populations and environments are required. Many publicly available and widely used few of those datasets are FER2013 [19], CK+ [36], RAVDESS [34], JAFFE [37], EmotioNet [3], SAVEE [56], and TESS [46].

Human emotions are complex therefore classifying human emotion is a stimulating job. Watson et al. suggested that human emotions can be broadly classified into two categories i.e. positive or negative [62]. Russell [50] proposed one 2D model to represent the emotional state in terms of valence and arousal where valence represents a positive or negative state and the intensity of valence is represented in terms of arousal. Whissell [63] developed a dimensional model 'The Dictionary of affect in language' to represent the emotion in terms of evaluation and activation. The work of Gasper [18] suggests that there is little consensus among the researchers regarding the 'neutral affect' and therefore leaves the decision on the experiment's needs and its type. The research work of Damasio [11] and Izard [23] strongly questioned the presence of 'neutral affect'. According to them, we are living in a world that is full of emotions and there is nothing called a 'neutral world' because we are always feeling something and our expressions are full of emotions. They further suggest that there can be nothing called an 'affectless mind' and all our emotions are tuned with emotions be it positive or negative. Summary of a few works related to FER and Acoustic have been shown in Table 1.

3 Methodology

In this section, a few baseline techniques relevant to this study have been discussed. The Convolutional Neural Network model has been applied by several researchers to solve the problem related to real-life [1]. It's a type of DNN algorithm that takes the image as input and recognizes an image from the rest of the images. The recognition is being made based on weight and bias values obtained during the learning process. The main principle behind the CNN architecture is the 'Convolutional Layer'. In CNN models each input image goes through a sequence of Convolutional layers. The connecting layers are filtered (kernel), pooling, fully connected layer (FC), and at the end of the layers, the "softmax" function is used to classify an image with prospective values between 0 and 1. In the case of CNN based model ReLU act as an activation function. It accumulates the weighted inputs. If the value is greater than the threshold value (0), it passed the signal into the next convolutional layer otherwise, inputs are rejected.

$$y = max\left[\left(\sum_i w_i x_i + a\right), 0\right] \tag{1}$$

In the CNN model to prevent exploding gradient problems and vanishing gradient problems, Batch normalization is used. Let ×1; ×2; xm, be a small batch then the mean value m and deviation σ can be obtained respectively using

$$u = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{2}$$

Table 1 Summary of work (FER and Acoustics)

Research work[Ref no]	Year	Dataset used	Feature Extraction Techniques	Learning Models	Performance
Mohan et al. [42]	2020	FER2013, JAFE,CK+,KDEF, RAF	Static Expression	CNN	78.9%, 96.7%, 97.8%, 82.5% and 81.68%, respectively
Li et al. [31]	2019	RAFDB, Affect Net	ACNN	CNN	85.07% and 58.78%
Breuer & Kimmel [5]	2017	CK+, FER2013	FACS and Action Units (AU)	CNN, CNN-LSTM	Shown state-of-the-art performance. 98.62 for CK+ and 72.1% on FER13
Kim et al. [26]	2017	MMI, CASME+MMI	Spatio-temporal feature	CNN	78.61% and 72.83%
Mollahosseini et al. [44]	2016	MultiPIE, MMI, CK+, DISFA, FERA, SFEW, and FER2013	HOG, LBPH, and Gabor	DNN	Offer high performance (66.41% for Top-1 and 81.7% Top-2
Zhao et al. [69]	2016	Oulu-CASIA and CK+	peak expression	PPDN	Shown state of art performance
Liu et al. [33]	2014	Ck+, JAFEE	Std. Facial features	BDBN	96% and 68%
Pinto et al. [47]	2020	RAVDESS	MFCC	CNN	91%
Kuang et al. [29]	2020	RAVDESS	MFCC,SIFT, STFT	Alex Net	95.88%
Iqbal et al. [22]	2019	RAVDESS SAVEE	MFCC, energy, spectral entropy, etc.	Gradient Boosting, SVM, KNN	Offer higher performance On test data
Yang et al. [64]	2019	EMODB	MFCC	BPNN, ELM, PNN, SVM	92.4%
Boulmaiz et al. [4]	2017	Generated	TRD-MFCC-SS	SVM	Shown satisfactory performance (Avg accuracy >90% for three types of Test cases)
Muda et al. [45]	2010	Features matching techniques	MFCC	DTW	Shows reasonable performance

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu)^2 \tag{3}$$

Then normalize ×1; ×2; xm by using a small number € in case σ = 0 by

$$\hat{x}_i = \frac{x_i - \mu}{\sqrt{\sigma^2 + €}} \tag{4}$$

The purpose of training is to reduce the loss as much as possible. Cross-Entropy is a popularly used Loss function expressed as

$$Loss = -\left[\sum_{i=1}^{n} P_i ln y_i\right] / n \tag{5}$$

Vgg16 is a kind of CNN model proposed by Simonyan and Zisserman [53]. The Network gains its popularity due to its simplicity. In this model 3 × 3 convolutional layers have been placed on top of each other in growing depth. Max pooling has been used to reduce volume size. It has two fully-connected layers, each having 4096 nodes. In the end, there is a 'Softmax classifier'. In the current study, VGG16 has been used to build FER based system as well as to classify facial expression-based emotions. Fuzzy Logic is the brainchild of Zadeh [66]. In a Fuzzy based system at first, we fit the linguistic variables in the form of crisp value to a Fuzzy system. Then these crisp values have been Fuzzified (converted to Fuzzy sets), thereafter the degree of the set memberships are determined using some membership function such as Triangle, Gaussian, etc. In the next level a set of if/then rules which are also known as a Fuzzy rule-based classifier has been applied. Finally, the fuzzy values are reconverted into the crisp value. In the case of a Fuzzy system, the membership function is used to represent/assign the degree of membership

$$A = \{(x, \mu A(x)) | x \in X\} \tag{6}$$

Here, μ*A* (*x*) represents the membership function or degree of membership function, of x in A and X, is the Universal set. The triangle Fuzzy membership can be expressed using the given equation.

$$trimf(x : a, b, m) = \begin{cases} 0 & x \leq a \\ \dfrac{x-a}{m-a} & a \leq x \leq m \\ \dfrac{b-x}{b-m} & m \leq x \leq b \\ 0 & c \leq x \end{cases} \tag{7}$$

Fuzzy rule-based classification mechanism is a well-known method in the machine learning domain because of its simplicity. It is extensively used to solve various real-life problems such as image process, sentiment analysis, etc. [17, 41, 61]. The if/then fuzzy rule is based on two parts i.e. IF (antecedent) THEN (consequence). The first one is specifying the membership function of antecedent Fuzzy sets and the second one is determining consequent class C_j and certainty grade CF_j of the fuzzy if/then rule (R_j). Given a *n*-dimensional, *c*-class problem, we apply the Fuzzy if-then rule in the following form

$$\text{Rule } R_j : \text{If } x_1 \text{ is } A_{j1} \text{ and}\ldots\text{and } x_n\ is\ A_{jn} \text{ Then class } C_i \text{ with } CF_i\ j = 1\ldots\ N. \tag{8}$$

Where R_j is the i^{th} Fuzzy if-then rule, *N* is the total number of Fuzzy if-then rules, $X = [x_1,. .., x_n]$ is *n*-dimensional pattern vector, A_{j1} presents antecedent Fuzzy sets for the i^{th} attribute, C_j represents a consequent class i.e. one of the *c* classes, and CF_j is a probable grade of the fuzzy if-then rule (R_j). The De-Fuzzification of the Fuzzy value into a crisp value can be done using Eq. (9)

$$Y = {\int_{\min}^{\max}\mu(y)ydy} \Big/ {\int_{\min}^{\max}\mu(y)dy} \tag{9}$$

Where Y is the result of Defuzzification, μ (y) is the membership function, y is the output variable, min is the lower limit, and max is the maximum limit for defuzzification.

4 Proposed method

This section describes the methods adopted for the current study. Both FE and AcI-based emotions have been considered together to analyze the minds of a section of Generation-Z.74 students who are pursuing engineering were involved in this process. The overall process has been depicted using a block diagram, see Fig. 1(a-c). The proposed work has three major phases; build two separate conventional CNN models [47, 53] using benchmark dataset, classification of emotions (FE and AcI) using these two models, employ the 'OSF' model to understand Gen-z emotions.

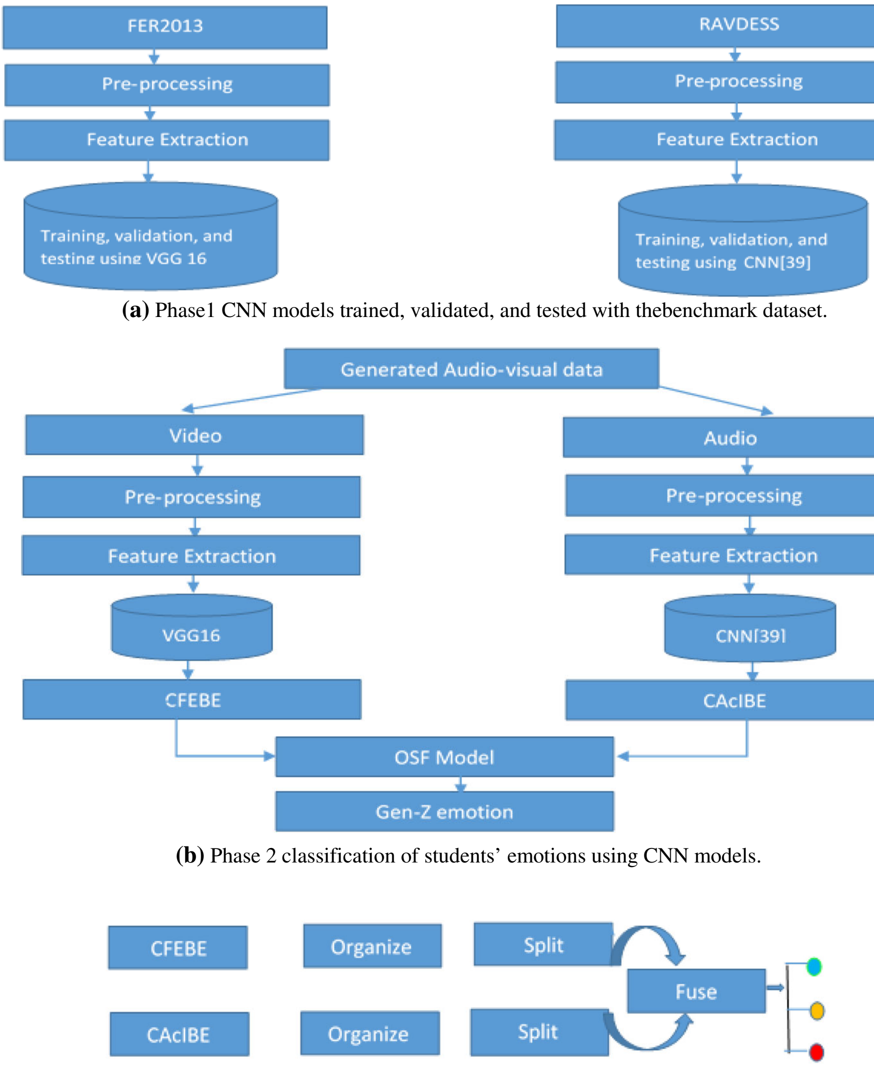
About the dataset In the proposed work 'FER-2013' dataset has been employed for training, validation, and testing the Facial expression-based CNN. In the case of acoustic information, RAVDESS datasets have been engaged. Finally, student datasets have been employed to classify emotions based on two modes. In the following section, a brief description of the datasets used in this study has been given.

FER2013 This is an image dataset comprising 35,889 48* 48-pixel gray-scale facial expression images. The images are labeled with the seven universal emotions (Table 2).

RAVDESS The Dataset is a collection of audio and video clips of 24 actors. They were expressing the same two lines with eight different emotions. Here in this study, only the audio clips of the RAVDESS dataset have been considered. Table 3 shows the distribution of eight such emotions.

Students dataset The dataset contains 74 numbers of videos. The students who appeared in these 74 videos are from an engineering college in West Bengal, India. Students of all years i.e. first to the fourth year were participated. Most of them are belong to the Computer Science and Engineering disciplines. Out of those 74 students, 29 of them are girls while the rest 35 are boys. The average age group of these students is 20. The average length of each video is 10.98 seconds. The total duration of all the videos is 813 seconds. Out of these 813 seconds, the contribution of girls is 386 seconds. The contribution of boys is 427 seconds. The average contribution of the girls in those videos is 13.31 seconds while the average contribution of boys is 9.48 seconds. Each facial and audio file has a unique name in the dataset. The videos were recorded during the period of lockdown i.e. in between the 1st week of April to the 3rd week of April 2020. The message, communicated in those video communications is based on the following dos and don'ts to be followed during the lockdown:

(a) Maintain social distance in public space (b) Wash your hand with soap/sanitizer regularly (c) Wear a mask and make use of hand gloves whenever goes out (d) Do not spread fake news (e) Respect and extend help to the corona worriers (f) Do not show any apathy towards the infected person or persons. (g)Stay home to stay safe (h) Practice yoga (i) Read Books (j) Watch movies with the family.



(a) Phase1 CNN models trained, validated, and tested with thebenchmark dataset.

(b) Phase 2 classification of students' emotions using CNN models.

(c) Phase 3 proposed OSF model.

Fig. 1 (**a-c**) The block diagram of the proposed work. Here 'FE' =Facial Expression, 'AcI' = Acoustic Information, 'CFEBE' =Classified Facial expression Based Emotions, 'CAIBE'= Classified Acoustic Information Based Emotions, 'OSF' =Organize-Split-Fuse, Negative—🔴, Neutral———🟡, Positive…🔵

About the role of CNN models In the current study CNN is used to serve two purposes. First, it is employed to build the FER and SER system following the traditional process [43]. In the case of FER based system, a VGG16 model [53] has been adopted (Table 4). The model was

Table 2 Emotions labels and numbers of images in the FER2013 dataset

Emotion Label	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
No of Images	4952	546	5120	8988	4829	6076	6197

Table 3 Emotions labels and number of audio samples in RAVDESS dataset

Emotion Label	Neutral	Clam	Happy	Sad	Angry	Fearful	Disgust	Surprised
No of Audio	96	192	192	192	192	192	192	192

trained, validated, and tested using the FER2013 dataset. In the case of SER, a CNN model proposed by [47] has been engaged. The RAVDESS datasets have been used to train, validate, and test the CNN model (SER). Subsequently using these two models emotions of 74 students have been classified. Figure 3(a-b) shows the emotion classification process for both modes and subsequent computation of classified emotions using these pre-build models.

Separation of facial and audio part At the facial and the acoustic information of each such student have been separated from their respective videos and stored separately. The process has been depicted in Fig. 2

Classification of gen-Z emotions using the pre-built CNN models Applying the pre-build VGG16 model, section 4.2, emotions of Gen-Z have been classified and several occurrences of the classified emotion labels in these videos have been computed for further processing. To achieve the stated purpose traditional FER-based process has been adopted [43], see Fig. 3(a). In the case of audio, acoustic information has been divided into 'n' numbers of chunks. Necessary padding has been done for the last chunks to make the length size equal. Then employing the CNN model [47] emotions are classified and several occurrences of the classified emotion labels in this audio were also computed, see Fig. 3 (b).

Valence based Organization of Emotions Surprise is the thematic term for describing a standing response. It starts with abrupt attention and then progresses into astonishment and finally converted into befuddled amazement. According to Whissell, [63] and Robinson [49] surprise is a positive emotion. Moreover, some popularly available public datasets have considered the surprise as a pleasant state [46]. Therefore in the current study, the 'surprise' has been considered as a positive state of emotions. Neutral does not Figure in this approach of emotion organization (valenced based) since some literature says that Neutral' does not represent any valence state [62]. It does not Influence Cognition or Behaviour [18]. Moreover, some research suggested that "It is not Possible to Feel Neutral Because People are Always Feeling Something" [11, 23]. Table 5 shows the emotion labels and their corresponding valence states.

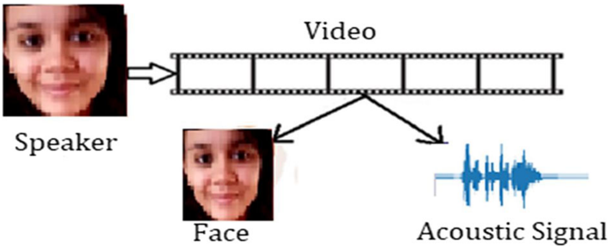
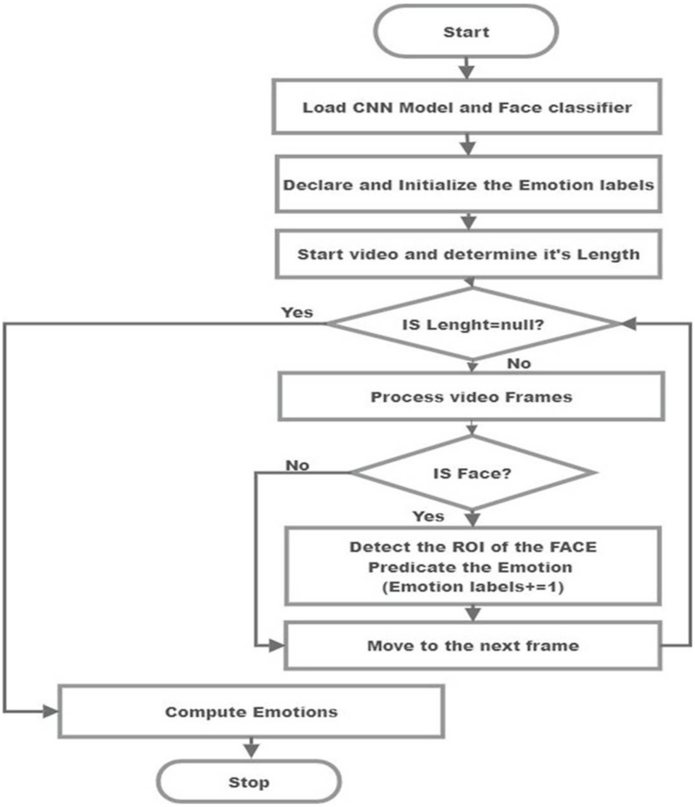
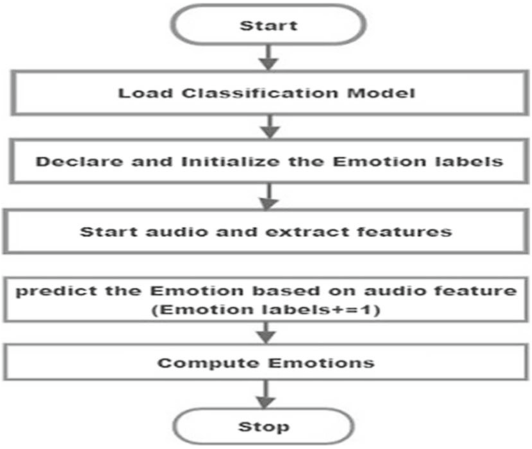


Fig. 2 The separations of FE and AcI from the captured Video



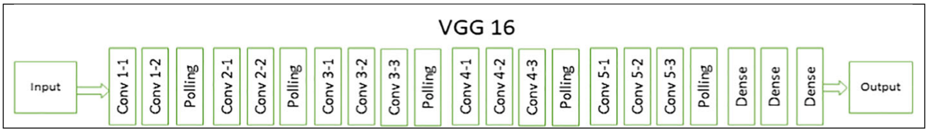
(a)


Start
Load Classification Model
Declare and Initialize the Emotion labels
Start audio and extract features
predict the Emotion based on audio feature
(Emotion labels+=1)
Compute Emotions
Stop

(b)

Fig. 3 **a** Classification of Facial expression and computation of classified emotion labels using pre-build CNN model **b** Classification of Acoustic information and computation of classified emotion labels using pre-build CNN model

Table 4 VGG16 Model architecture



Based on Table 5 the emotions of each student have been categorized and subsequently organized using Eq. (10). The process of valence-inspired emotion organization for both modes has been presented using Algorithm 1. Figure 4(a-b) depicted the process of valence-inspired emotion organization.

$$T_v = \begin{cases} P_v & Dv > 0 \\ Nu_v & Dv = 0 \\ N_v & Dv < 0 \end{cases} \tag{10}$$

Where T_v= Types of valence; D_v=Difference between the cumulative value of all positive and negative valence; Nu_v=Neutral or absent of valence; P_v is the positive valence and N_v is the negative valence.

The Following algorithm has been used separately to organize the classified emotions (FE and AcI).

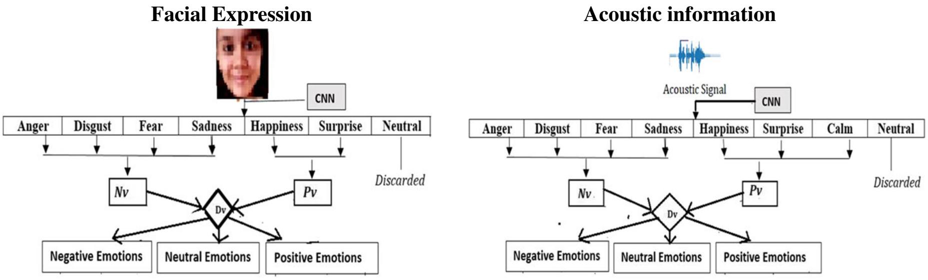
Algorithm 1: Valence based organization of classified emotions

```
Input:   The classified emotions labels
Output: Classes of emotions
Begin
      $P_V$= summations of all positive emotion labels
      $N_V$=summations of all negative emotion labels
      $D_V=P_V- N_V$
      If  $D_V>0$
          Valence: Positive
      Else if  $D_V <0$
          Valence: Negative
      Else
          Valence: Neutral
      End if
End
```

Splitting of Valenced emotions Once emotions of both types have been obtained and organized based on their valences of emotion, see Fig. 4(a-b), it has been further divided into six sub-classes of emotions using the degree of valence (D_v) i.e. arousal. Algorithms 2 and 3 presented the process of splitting while Fig. 5(a-b) depicted the outcome of the process.

Table 5 Emotion labels and their corresponding valence

Valence	Negative Emotions/Valence	Anger, Fear, Disgust, Sadness
	Positive Emotions/Valence	Happiness, Calm, Surprise



(a) Valence based organization of classified

(b) Valence based organization of classified

Fig. 4 **(a)** Valence based organization of classified emotions, FE-based. **(b)** Valence based organization of classified emotions AcI-based

Algorithm 2: Sub-classification of facial expression based on the degree of valence emotions.

```
Input:    $D_v$s from algorithm 1
Output: Sub-classified emotions based on the degree
          of $D_v$ i.e. arousal
Begin
    If $D_v$ >=50
        Emotions: High Positive
    Else if 50< $D_v$ >25
        Emotions: Moderate Positive
    Else if 25<=$D_v$ >=0
        Emotions: Low Positive
    Else if <0 $D_v$ >=-25
        Emotions: Low Negative
    Else if -25 <$D_v$ >-50
        Emotions: Moderate Negative
    Else
        Emotions: High Negative
    End if
End
```

Algorithm 3: Sub-classification of acoustic information based degree of valenced emotions.

```
Input:   $D_vs$ from algorithm 1
Output: Sub-classified emotions based on the degree
          of $D_v$i.e. arousal
Begin
    If $D_v$ >=5
        Emotions: High Positive
    Else if 5< $D_v$>2.5
        Emotions: Moderate Positive
    Else if 2.5 <=$D_v$ >=0
        Emotions: Low Positive
    Else-if 0< $D_v$ >=-2
        Emotions: Low Negative
    Else if -2 <$D_v$>-5
        Emotions: Moderate Negative
    Else
        Emotions: High Negative
    End if
End
```

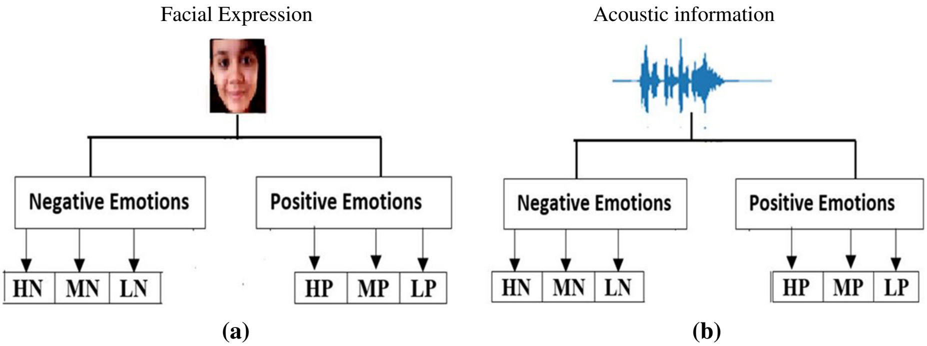


Fig. 5 **(a)** Sub-classifications of valenced emotions based on Algorithm-2 (**b**) Sub-classifications of valence emotions based on Algorithm-3

Decision Level Fusion using Fuzzy rule-based classification system Here in this section sub-classified six facial expression-based emotions have been amalgamated with six sub-classified acoustic information-based emotions. To achieve these objectives Fuzzy rule-based classification system has been employed. The process of amalgamation has been explained using Algorithm 4. In Fig. 6, the combination process has been depicted. The inputs and outputs of the proposed fuzzy system have been shown in Table 6. Table 7 shows the proposed Fuzzy rules to be applied in the current study to fuse the emotions obtained from both modes.

Algorithm 4: Fusion of split emotions using Fuzzy rule-based classification system.

```
---------------------------------------------------------------------------------------------------------
Input: sub-classes from algorithms 2 and 3
Output: Classified Emotions
Begin
     Create a 'Mamdani' based Fuzzy Inferences System 'EMO-FIS'
     Read emotions of voice and face into 'EMO'
     For each emotion 'e' in 'EMO'
         Voice=e (1)
         Face=e (2)
         Read Emo-FIS in 'ES'
         Classification (e) = evaluate ([Voice, Face] in ES)
      End
End
```

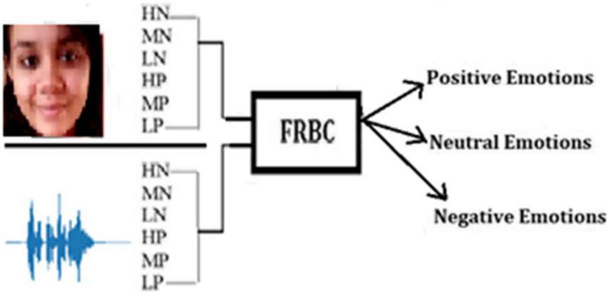


Fig. 6 The decision level fusion of facial and acoustic information based emotions using Fuzzy rule-based classification system

Table 6 The input of the Linguistic variables in the Fuzzy system and decision based on De-Fuzzification

Type	Linguistic Variables
Input	The degree of valence
Output	Fuzzy based Classification(Positive, Neutral, and Negative)

Table 7 Fuzzy rules to fuse both modes of emotions based on the degree of emotion i.e. D_v

SL No	Intensity of emotions		Combined emotional state based on Algorithm 4	SL No	Intensity of emotions		Combined emotional state based on Algorithm 4
	Facial	Acoustic			Facial	Acoustic	
01	HN	HN	Negative	31	LP	HN	Negative
02	HN	MN		32	LP	MN	
03	HN	LN		33	LP	LN	Neutral
04	HN	HP	Neutral	34	LP	HP	Positive
05	HN	MP	Positive	35	LP	MP	
06	HN	LP		36	LP	LP	

5 Results and discussion

During the initial observation, it was observed that the quality of facial expressions in a few recorded videos is very poor but the audio quality is good. The same is observed in the case of acoustic data. The possible reasons are mentioned in the challenges and limitations of this study. In this case, the decision has been made based on either one of the two approaches. The number of samples considered for the facial expressive emotions is 68 and for acoustics, the number of samples considered is 71.

The performance of the employed CNN model (VGG16) has been depicted in Tables 8 and 9. While in Table 10 performance has been compared with some previous work. Figures 7 and 8 show the emotions obtained from facial expression and acoustic information, respectively. Based on the valence state, identified emotions have been organized into two classes (Eq. 10 and Algorithm 1). The results obtained have been shown in Figs. 9 and 10.

Once the emotions of both modes have been organized successfully then organized facial and acoustic information-based emotions have been further sub-classified into six sub-classes based on the degree of intensity of their valence (D_v). To achieve this objective Algorithm 3

Table 8 Confusion matrix for the employed CNN model

	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	**356**	0	24	30	17	17	23
Disgust	19	**12**	2	6	5	2	10
Fear	30	1	**263**	18	70	44	70
Happy	15	0	16	**756**	22	23	63
Sad	45	0	70	42	**387**	5	104
Surprise	9	0	28	12	7	**348**	11
Neutral	22	0	12	29	42	11	**489**

The correct classification is indicated in bold font type

Table 9 Performance of the employed CNN model

Labels	Precision	Recall	F-Measure	Accuracy: 0.727906
Angry	0.718	0.762	0.739	
Disgust	0.923	0.214	0.348	
Fear	0.634	0.53	0.577	
Happy	0.847	0.845	0.846	
Sad	0.704	0.593	0.643	
Surprise	0.773	0.839	0.805	
Neutral	0.635	0.808	0.711	

Table 10 Comparative analysis of various FER based works

Sl No.	Method	Dataset	Accuracy (%)
01	Ensemble DCNNs [48]	FER2013	53.00
02	MCNN [2]		64.00
03	Deep-emotion [39]		70.00
04	STF+LSTM [27]		71.00
05	FER-net [43]		79.00
06	VGG16 (Employed)		72.70

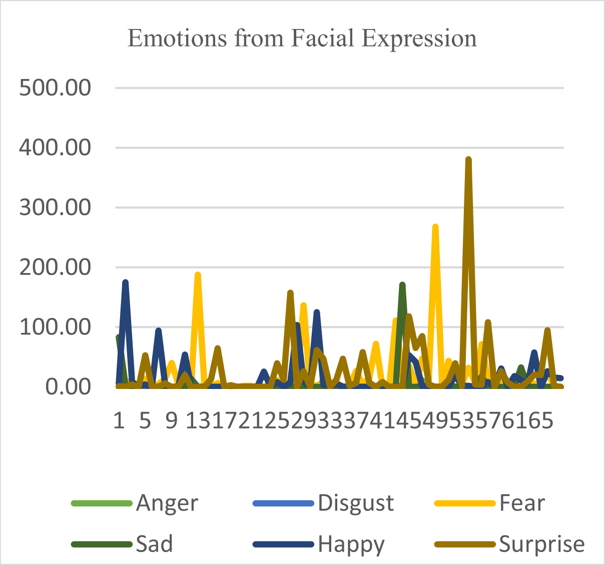


Fig. 7 Emotions of the students based on facial expression using VGG16

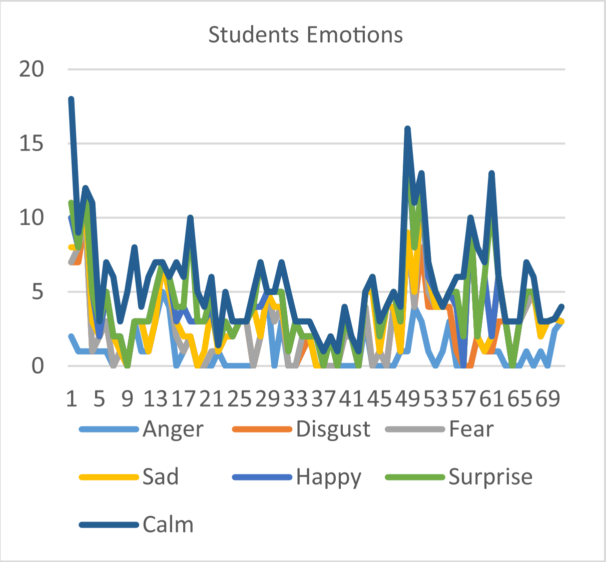


Fig. 8 Emotions of students based on acoustic information using CNN [47]

and 4 have been applied. The output of this sub-classification i.e. the second part of the proposed OSF model has been depicted in Figs. 11 and 12.

Finally, six sub-classified emotions, obtained from both of these modes have been combined. To achieve this goal a Fuzzy rule-based classification mechanism has been engaged.

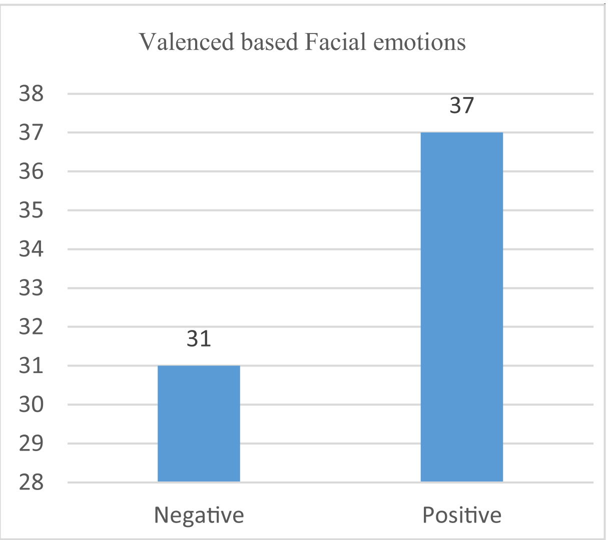


Fig. 9 Classifications of facial expression based emotions using Algorithm 1

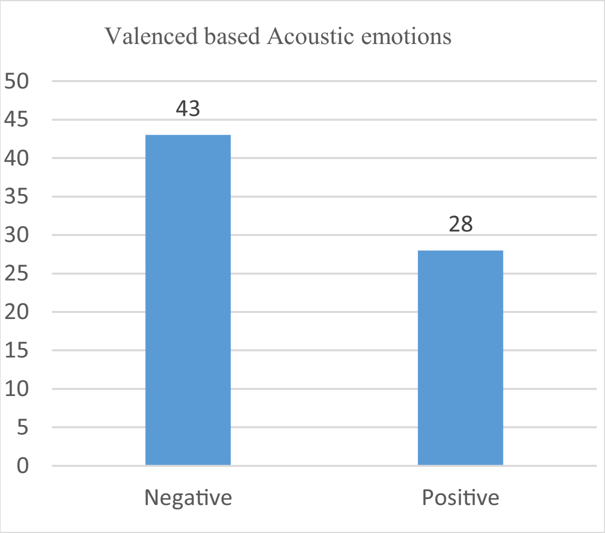


Fig. 10 Classifications of acoustic information-based emotions using Algorithm 1

Figure 13(a) shows the proposed Fuzzy system while Fig. 13(b-c) shows the degree of membership of each mode. The value of each mode lies between −10 to +10. In Figs. 14, 15, 16, the findings of fusions of two modes, based on Fuzzy rule-based classification systems have been displayed. The year-based emotions of the students have been illustrated in Fig. 17. The comparative emotional states of boys and girls during this lockdown period have been shown using Figs. 18, 19, 20, 21.

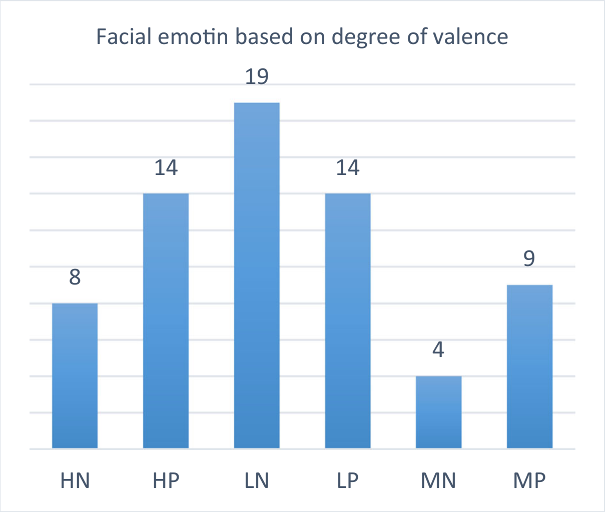


Fig. 11 Sub-classifications of facial emotions based on the intensity of valence

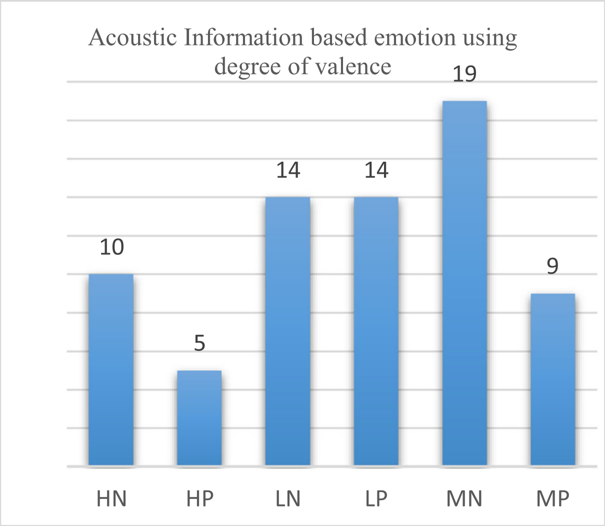


Fig. 12 Sub-classifications of acoustic information based emotions based on the intensity of valence

6 Discussion

History reminds us that there is some acceleration of negative emotions in society after every great catastrophe like World War I and World War II. The impact of global Lockdown is one such big catastrophe. In the present study, the results/emotions obtained from the student's facial and acoustic information also support this claim. The emotions obtained from their facial expression show the presence of more positive emotions in comparison to negative emotions, Fig. 7. On the other hand, the results obtained from acoustic pieces of information show the dominance of negative emotions over positive emotions, Fig. 8. Thus to conclude our findings, the OSF model has been employed. At the outset, emotion labels have been identified based on their valence (positive/negative) and then numbers of such emotions have been added and compared. The result obtained from Fig. 9 shows that like Fig. 7, the supremacy of positive emotions (37) over its counterpart negative (31). The results obtained from acoustic pieces of information show the dominance of negative emotions (43) over positive emotions (28), Fig. 10. At the end of the first step of the proposed OSF model, we got two class labels, namely, positive and negative, for both modes.

Fig. 11 shows the sub-classified emotions, obtained from the facial expression based on the degree of valence. The results show the values of HP (14) and MP (9) are more compared to HN (9) and MN (4). It also shows the dominance of LN (19) over LP (14). The sub-classified emotions, obtained from acoustic information, based on the degree of valence show that HN (10), and MN (19) prevailed over HP (5), MP (9), Fig. 12. It also shows the presence of LP (14) and LN (14) are on the same scale. After the successful completion of splitting i.e. the second part of the OSF model, six sub-classes of emotions have been obtained based on the degree of valence for both modes (facial and acoustic).

Fig. 13(a) shows the process employed to fuse the emotions of both modes using the 'Mamdani' based Fuzzy Inference System (FIS). Fig. 13(b) and (c) show the degree of memberships of emotions for both modes. Here in this study based on Algorithms 2 and 3,

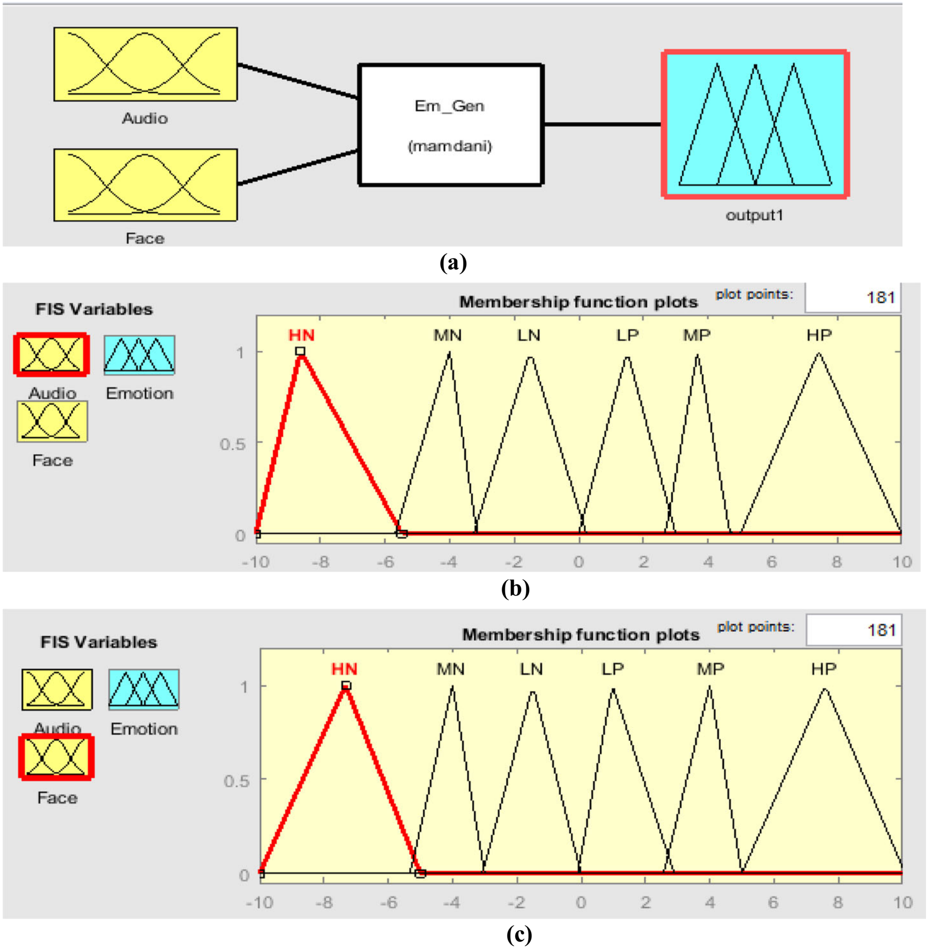


Fig. 13 **(a)** Proposed Fuzzy Interface System (b) Degree of membership for Facial emotions **(c)** Degree of membership for acoustic information based emotions

the membership values of the proposed Fuzzy system have been determined. The value lies between −10 to +10. Where −10 represents the Highest Negative (HN) value while +10 represents the Highest Positive value (HP). The rest of the memberships has been defined between these two values based on Algorithm 2 and 3.

Fig. 14 shows the student's emotions in terms of three classes. It shows that the negative emotions (34) are comparatively more than positive ones (31), while there are 9 neutral emotions. Figure 15 shows the student's emotions in terms of the degree of valence. It shows that HP (15) prevails over HN (15), indicating that students are more positively oriented when the degree of emotions is at its highest level. It also exhibits the supremacy of MN (13) over MP (7) and marginal dominance of LN (6) over LP (5). Figure 16 shows that the presence of

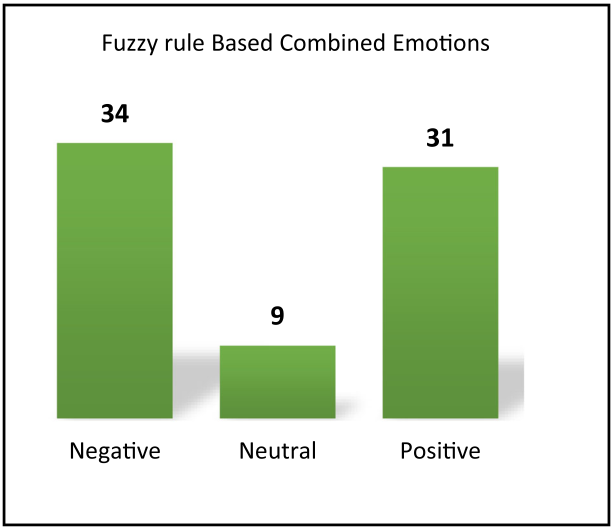


Fig. 14 Fuzzy rule-based combined emotions in terms of their valence

positive, negative, and neutral emotions is 42%, 46%, and 12% respectively. Figure 17 unveils a very important scenario. It shows that as students progress toward senior classes, negative emotions increase. (0.42, 0.44, 0.50, and 0.56 respectively). It also shows a gradual decrease in positive emotions (0.50, 0.46, 0.33, and 0.22 respectively). It further shows the presence of more neutral emotions in the senior students compared to their juniors (0.08, 0.10, 0.17, and 0.22 respectively).

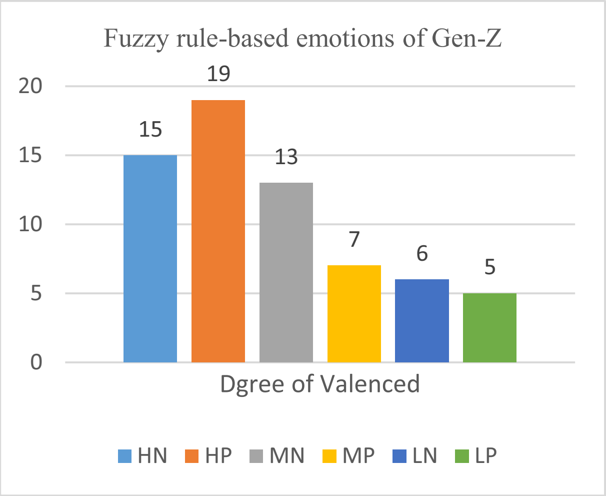


Fig. 15 Fuzzy rule-based classification of combined emotions in terms of their degree of valence

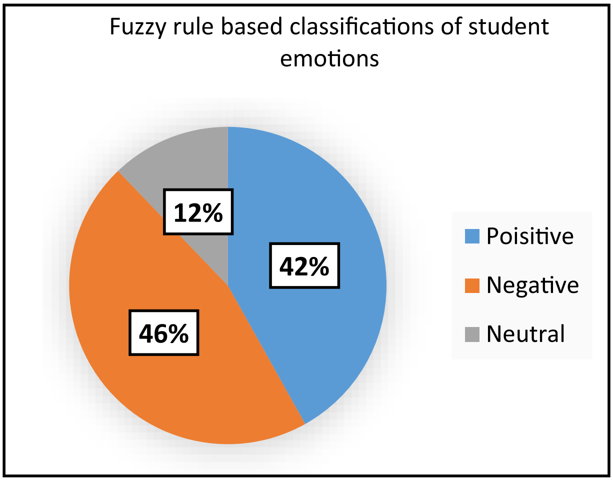


Fig. 16 Fuzzy rule-based polarized emotions of students

To get the entire spectrum of students' emotions further analysis shows that boys are more apprehensive than girls while girls have more optimism. Figure 18 narrates, that boys (26) have more presence of negative emotions compared to girls (9). It also shows that the presence of positivity in girls is more (19) compare to boys (11) although the presence of neutral emotions in the boys (8) is more than in girls (1). The year-wise comparison of emotions also shows very interesting statistics, Fig. 19. It shows the presence of less negativity and more positivity in the girls compared to boys except for the final year students where the presence of positivity in boys is more compared to girls. It also tells that apart from 1st-year, the emotions of girls are completely bi-polar while for other years there are few presences of neutral emotions. The year-wise analysis shows the chaotic state of students' emotions, Figs. 20 and 21. It does not follow any patterns or defined directions.

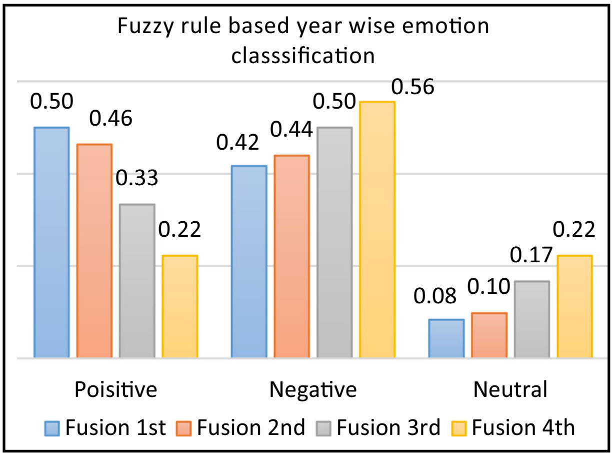


Fig. 17 Year-wise students emotions using a fuzzy rule-based classification system

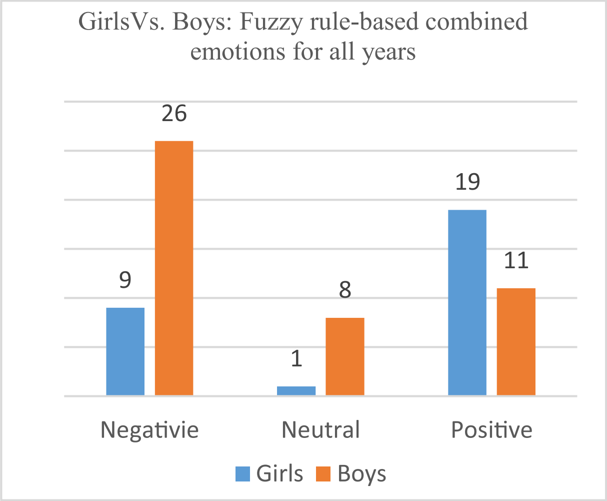


Fig. 18 Boys vs. Girls, Fuzzy rule-based combined emotion of all the years

7 Conclusion

There is no denying that the effects of Lockdown will completely change the whole world, as this kind of catastrophe is extremely unusual. However, experts believe that this will have a major impact on Generation-Z. It will completely change their view of the world. In the submitted work, emotions of a section of Generation-Z have been classified by combining the FE-based emotions with the AcI. A fuzzy rule-based classification system has been employed at the decision level to fuse six sub-classified information obtain from two modes (video and audio) into three class labels (positive, neutral, and negative). This has been done based on the

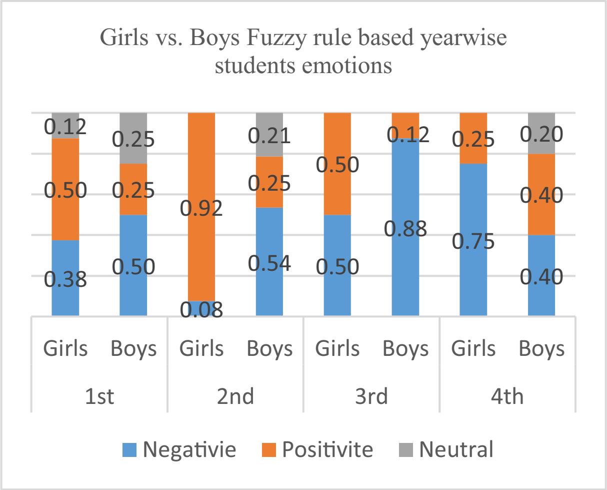


Fig. 19 Boys vs. Girls: year-wise classes of emotions

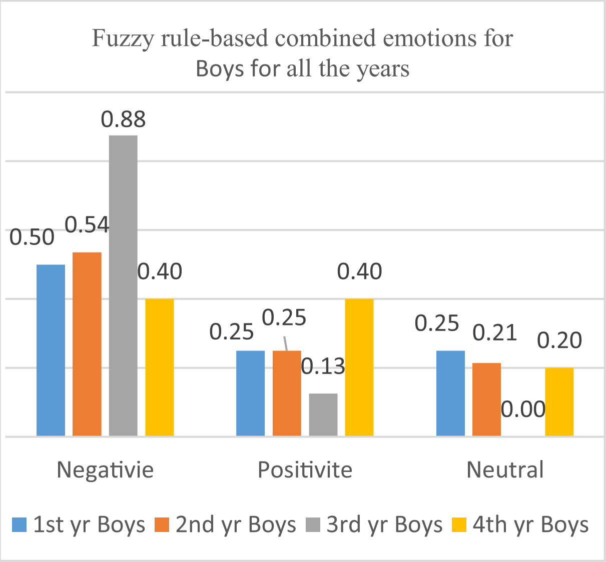


Fig. 20 Fuzzy rule-based combined emotions of Boys for all the years

degree of their valence(D_v). The test dataset used is comprised of 74 small videos. The overall study reveals that

- The junior students have fewer Negative, emotions compared to their seniors. The probable reason is that they are highly optimistic, more energetic, and confident to handle any odds that may come their way. Again it could be that they failed to comprehend the threats that lie in the feature due to their lack of maturity.

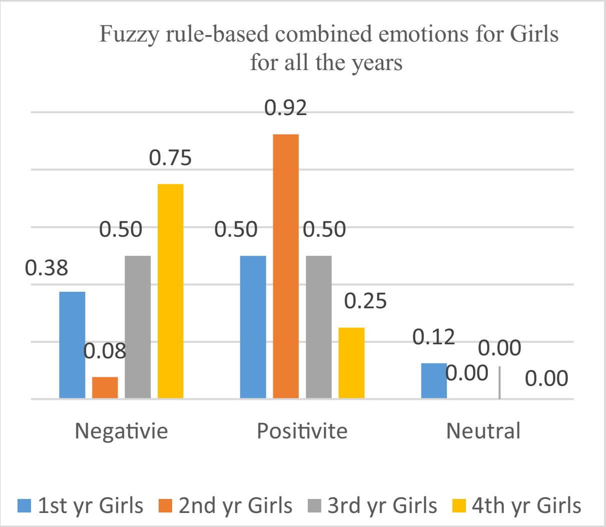


Fig. 21 Fuzzy rule-based combined emotions of Girls for all the years

- The result analysis also shows that first-year students are more happy compare to other years while the fourth-year students are more apprehensive about the fallout of this lockdown and corona crisis. They are emotionally more neutral compare to their juniors, suggesting that they are more confused about the fallout.
- It also shows that the sign of fear or surprise is very less in all the cases though signs of sadness have been observed in some time.
- The study also reveals that compared to boys, girls are more optimistic about their future. The outcome of Figs. 18 and 19 also supported this claim to some extent.
- The year-wise comparison between boys and girls in Figs. 20 and 21 shows students' chaotic state of mind.

Challenges: Since we have considerable control over a large part of facial muscles, we can mask our facial emotions from the rest of the world to some extent. It has been recorded that sometimes instead of revealing many negative feelings like annoyance or aggravation, emotions such as happiness, joy, etc. have been expressed. Therefore in such a circumstance, the face makes a mockery of the mind and acts as a mask. Thus, making the facial expression-based emotion classification process more stimulating or sometimes inducing imprecision. In the case of acoustic signals addition of background or ambiance noise makes the quality of the signal poor and noisy. Thus, sometimes suppressed the necessary acoustic features, making the detection process hard or inducing imprecision.

Limitations: During this study, we face some problems, firstly, it can detect only the front part of the face with some restrictions. Second, in addition to the different ambiance, the resolution and quality of the capture device used to record student expressions were different. Third, the study does not represent the entire Generation-Z. In the future, to extract facial and acoustic-based emotions, different CNN models and datasets can be engaged. To extract emotions based on acoustic information apart from MFCC, other features can also be included.

Acknowledgments We appreciate the efforts of the Hooghly Engineering and Technology College students in expressing their views during the lockdown time. We are also expressing our gratitude towards the College authority and all the stakeholders for their support and encouragement.

Declarations

Conflict of Interest/Competing Interest The authors certify that there is no conflict of interest regarding the material discussed in the manuscript. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

Appendix Figs. 22 and 23.

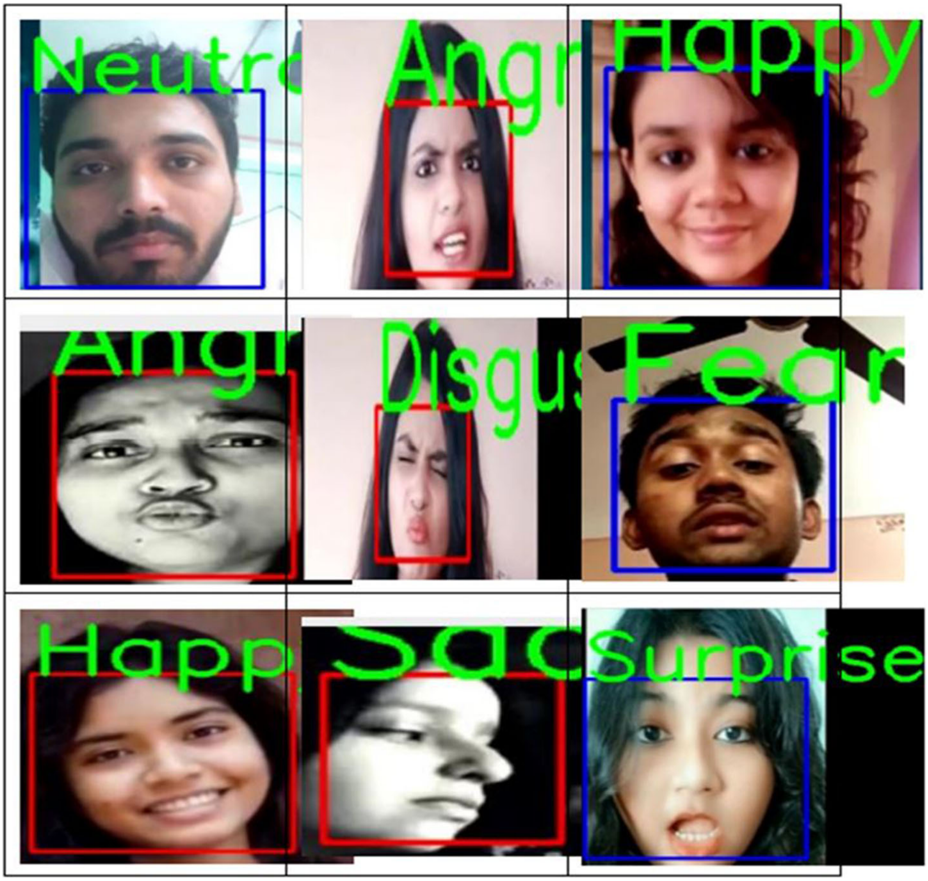


Fig. 22 Facial Expression of Engineering student

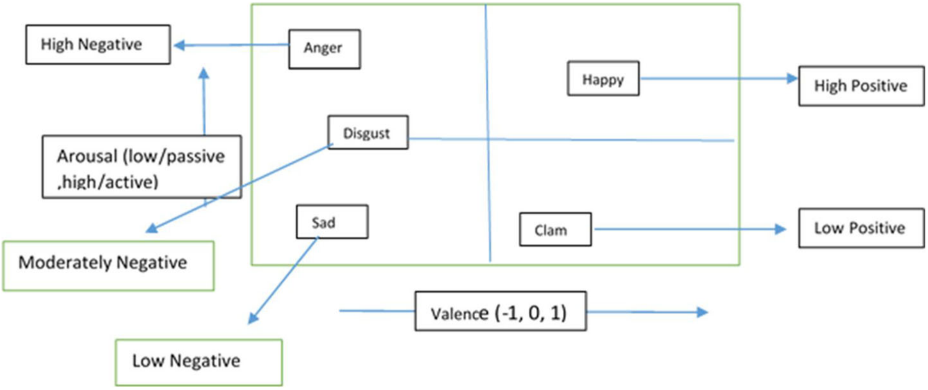


Fig. 23 Valence-Arousal based 2D emotion model

References

1. Ali MNY, Sarowar MG, Rahman ML, Chaki J, Dey N, Tavares JMR (2019) Adam deep learning with SOM for human sentiment classification. Int J Ambient Comput Intell (IJACI) 10(3):92–116
2. Alizadeh S, Fazel A (2017) Convolutional neural networks for facial expression recognition arXiv:1704:06756. https://doi.org/10.48550/arXiv.1704.06756
3. Benitez-Quiroz CF, Srinivasan R, Martinez AM (2016) Emotional: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: 2016 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, pp 5562–5570. https://doi.org/10.1109/CVPR.2016.600
4. Boulmaiz A, Messadeg D, Doghmane N, Taleb-Ahmed A (2017) Design and implementation of a robust acoustic recognition system for waterbird species using TMS320C6713 DSK. Int J Ambient Comput Intell (IJACI) 8(1):98–118
5. Breuer R, Kimmel RA (2014) deep learning perspective on the origin of facial expressions. arXiv 2017, arXiv:1705.01842
6. Chakraborty I, Maity P (2020) COVID-19 outbreak: migration, effects on society, global environment, and prevention. Sci Total Environ 728:138882. https://doi.org/10.1016/j.scitotenv.2020.138882
7. Chandrasekar R, Khare N (2016) Review of Fuzzy Rule-Based Classification systems. Res J Pharm Tech 9(8):1299–1302. https://doi.org/10.5958/0974-360X.2016.00247.X
8. Chen C-R, Wong W-S, Chiu C-T (2010) A 0.64 mm 2 real-time cascade face detection design based on reduced two-field extraction. IEEE Trans Very Large Scale Integr (VLSI) Syst 19(11):1937–1948 20
9. Covid-19 impact on young people and the youth sector (2020) Knowledge HUB: COVID-19 impact on the youth sector Council of Europe European Union. https://pjp-eu.coe.int/en/web/youth-partnership/covid-19
10. Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion recognition in human-computer interaction. IEEE Signal Process Mag 18(1):32–80. https://doi.org/10.1109/79.911197
11. Damasio A (2003) Virtue in mind. New Sci 180(49–51):2003
12. Darwin C, Prodger P (1998) The expression of the emotions in man and animals. Oxford University Press, Oxford
13. Ekman P, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17(2):124–129
14. Fasel B, Luettin J (2003) Automatic facial expression analysis: a survey. Pattern Recogn 36(1):259–275
15. Fathallah A, Abdi L, Douik A (2017) Facial expression recognition via deep learning. In: 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA). IEEE, pp 745–750. https://doi.org/10.1109/AICCSA.2017.124
16. Fong SJ, Dey N, Chaki J (2020) Artificial intelligence for coronavirus outbreak, pp 23–45. https://doi.org/10.1007/978-981-15-5936-5_2
17. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E (2020) Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl Soft Comput 93:106282
18. Gasper K (2018) Utilizing neutral affective states in research: theory, assessment, and recommendations. Emot Rev 10:255–266. https://doi.org/10.1177/1754073918765660
19. Goodfellow IJ, Erhan D, Carrier PL et al (2013) Challenges in representation learning: a report on three machine learning contests. Neural Networks : the Official Journal of the International Neural Network Society 64:59-63. https://doi.org/10.1016/j.neunet.2014.09.005
20. Gupta P, Rajput N (2007) Two-stream emotion recognition for call center monitoring. Proc Interspeech 2007:2241–2244. https://doi.org/10.21437/Interspeech.2007-609
21. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition, Las Vegas, pp 770–778. https://doi.org/10.1109/CVPR.2016.90
22. Iqbal A, Barua K (2019) A real-time emotion recognition from speech using gradient boosting. In: 2019 IEEE international conference on electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, pp 1–5. https://doi.org/10.1109/ECACE.2019.8679271
23. Izard CE (2007) Basic emotions, natural kinds, emotion schemas, and a new paradigm. Perspect Psychol Sci 2:260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x
24. Jain DK, Shamsolmoali P, Sehdev P (2019) Extended deep neural network for facial emotion recognition. Pattern Recogn Lett 120:69–74
25. Jannat R, Tynes I, Lime LL, Adorno J, Canavan S (2018) Ubiquitous emotion recognition using audio and video data. In: 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, ACM, pp 956–959. https://doi.org/10.1145/3267305.3267689

26. Kim DH, Baddar W, Jang J, Ro, YM (2017) Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Trans Affect Comput 10:223–236. https://doi.org/10.1109/TAFFC.2017.2695999
27. Kim DH, Baddar WJ, Jang J, Ro YM (2017) Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. IEEE Trans Affect Comput 10(2):223–236
28. Kotsia I, Pitas I (2006) Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans Image Process 16(1):172–187
29. Kuang Y, Wu Q, Wang Y, Dey N, Shi F, Crespo RG, Sherratt RS (2020) Simplified inverse filter tracked affective acoustic signals classification incorporating deep convolutional neural networks. Appl Soft Comput 97(Part A):106775
30. Lecun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. Handb Brain Theory Neural Netw 3361:10
31. Li Y, Zeng J, Shan S, Chen X (2019) Occlusion aware facial expression recognition using CNN with attention mechanism. IEEE Trans Image Process 28:2439–2450
32. Liu C, Wechsler H (2002) Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. IEEE Trans Image Process 11(4):467–476
33. Liu P, Han S, Meng Z, Tong Y (2014) Facial expression recognition via a boosted deep belief network. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Columbus, pp 1805–1812. https://doi.org/10.1109/CVPR.2014.233
34. Livingstone SR, Russo FA (2018) The Ryerson audio-visual database of emotional speech and Song (RAVDESS): a dynamic, multimodal set of facial and vocal expressions in north American English. PLoS One 13(5):e0196391
35. Low LA, Maddage NC, Lech M, Sheeber LB, Allen NB (2011) Detection of clinical depression in adolescents' speech during family interactions. IEEE Trans Biomed Eng 58(3):574–586. https://doi.org/10.1109/TBME.2010.2091640
36. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: 2010 IEEE computer society conference on computer vision and pattern recognition-workshops. IEEE, pp 94–101. https://doi.org/10.1109/CVPRW.2010.5543262
37. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with gabor wavelets. In: 1998 IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp 200–205. https://doi.org/10.1109/AFGR.1998.670949
38. Mahalle P, Kalamkar AB, Dey N, Chaki J, Shinde GR (2020) Forecasting models for coronavirus (covid-19): a survey of the state-of-the-art. SN Comput Sci 1(4):197. https://doi.org/10.1007/s42979-020-00209-9
39. Minaee S, Abdolrashidi A (2019) Deep-emotion: facial expression recognition using attentional convolutional network. Computer Vision and Pattern Recognition. arXiv:1902.0101. https://doi.org/10.48550/arxiv.1902.01019
40. Mohammadi MR, Fatemizadeh E, Mahoor MH (2014) Pca-based dictionary building for accurate facial expression recognition via sparse representation. J Vis Commun Image Represent 25(5):1082–1092 13
41. Mohammadpour RA, Seyed M, Abedi M, Bagheri S, Ghaemian A (2015) Fuzzy rule-based classification system for assessing coronary artery disease. Comput Math Methods Med 2015(564867):8. https://doi.org/10.1155/2015/564867
42. Mohan K, Seal A, Krejcar O, Yazidi A (2020) Facial expression recognition using local gravitational force descriptor based deep convolution neural networks. IEEE Trans Instrum Meas 70:1–12
43. Mohan K, Seal A, Krejcar O, Yazidi A (2021) FER-net: facial expression recognition using deep neural net. Neural Comput Applic 33:9125–9136. https://doi.org/10.1007/s00521-020-05676-y
44. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter Conference on Applications of Computer Vision(WACV), pp 1–10. https://doi.org/10.1109/WACV.2016.7477450
45. Muda L, Begam M, Elamvazuthi I (2010) Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (dtw) techniques, ArXiv, abs/1003.4083. https://doi.org/10.48550/arXiv.1003.4083
46. Pichora F, Kathleen M, Kate D (2020) Toronto emotional speech set (TESS), Borealis, V1. https://doi.org/10.5683/SP2/E8H2MF
47. Pinto MG, Polignano M, Lops P, Semeraro G (2020) Emotions understanding model from spoken language using deep neural networks and Mel-frequency cepstral coefficients. In: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), pp 1–5. https://doi.org/10.1109/EAIS48028.2020.9122698
48. Pons G, Masip D (2017) Supervised committee of convolutional neural networks in automated facial expression analysis. IEEE Trans Affect Comput 9(3):343–350

49. Robinson DL (2008) Brain function, emotional experience and personality. Neth J Psychol 64:152–167
50. Russell J (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178. https://doi.org/10.1037/h0077714
51. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis Comput 27(6):803–816
52. Shao J, Qian Y (2019) Three convolutional neural network models for facial expression recognition in the wild. Neurocomputing 355:82–92
53. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556. https://doi.org/10.48550/arXiv.1409.1556
54. Socio-economic impact of COVID-19 (2020) Briefs and Report. https://www.undp.org/content/undp/en/home/coronavirus/socio-economic-impact-of-covid-19.html
55. Sun N, Li Q, Huan R, Liu J, Han G (2017) Deep spatial-temporal feature fusion for facial expression recognition in static images. Pattern Recogn Lett 119(49–61):31
56. Surrey Audio-Visual Expressed Emotion (SAVEE). (n.d.), http://kahlan.eps.surrey.ac.uk/savee/
57. Tian YI, Kanade T, Cohn JF (2001) Recognizing action units for facial expression analysis. IEEE Trans Pattern Anal Mach Intell 23(2):97–115
58. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: 1991 IEEE Conference on computer society computer vision and pattern recognition, Maui, pp 586–591
59. Tzirakis P, Zafeiriou S, Schuller B (2019) Real-world automatic continuous affect recognition from audiovisual signals. In: Pineda A, Sebe R (eds) Multimodal Behavioral Analysis in the Wild: Advances and Challenges. Academic Press Ltd-Elsevier Science Ltd, pp 387–406. https://doi.org/10.1016/B978-0-12-814601-9.00028-6
60. Viola P, Jones P (2001) Rapid object detection using a boosted cascade of simple features. In: 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, pp 511–518. https://doi.org/10.1109/CVPR.2001.990517
61. Wang D, He T, Li Z, Cao L, Dey N, Ashour AS, Shi F (2018) Image feature-based affective retrieval employing improved parameter and structure identification of adaptive neuro-fuzzy inference system. Neural Comput Applic 29(4):1087–1102
62. Watson D, Wiese D, Vaidya J, Tellegen A (1999) The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. J Pers Soc Psychol 76:820–838. https://doi.org/10.1037/0022-3514.76.5.820
63. Whissell CM (1989) The dictionary of affect in language. In: Plutchik R, Kellerman H (eds) The measurement of emotion. Academic Press, pp 113–131. https://doi.org/10.1016/B978-0-12-558704-4.50011-6
64. Yang N, Dey N, Sherratt RS, Shi F (2020) Recognize basic emotional statesin speech by machine learning techniques using mel-frequency cepstral coefficient features. J Intell Fuzzy Syst 39(2):1925–1936 ISSN 1875-8967
65. Youth and COVID-19: Response, Recovery and Resilience (2020) OECD Survey on COVID-19 and Youth. http://www.oecd.org/coronavirus/policy-responses/youth-and-covid-19-response-recovery-and-resilience-c40e61c6/
66. Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353. https://doi.org/10.1016/S00199958(65)90241-X
67. Zhang B, Essl G, Provost EM (2015) Recognizing emotion from singing and speaking using shared models. In: IEEE 2015 International Conference on Affective Computing and Intelligent Interaction (ACII) IEEE, pp 139–145. https://doi.org/10.1109/ACII.2015.7344563
68. Zhang D, Song F, Xu Y, Liang Z (2009) Decision level fusion, advanced pattern recognition technologies with applications to biometrics. IGI Global, pp 328–348. https://doi.org/10.4018/978-1-60566-200-8.ch015
69. Zhao X, Liang X, Liu L, Li T, Han Y, Vasconcelos N, Yan S (2016) Peak-piloted deep network for facial expression recognition. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part II 14 425–442. Springer International Publishing
70. Zhong L, Liu Q, Yang P, Huang J, Metaxas DN (2014) Learning multiscale active facial patches for expression analysis. IEEE Trans Cybern 45(8):1499–1510

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.