Vol. 05, No. 4 (2023) 793-804, doi: 10.24874/PES05.04.021

Proceedings on Engineering Sciences

www.pesjournal.net

BI-MODEL EMOTION RECOGNITION SYSTEM FOR DIFFERENT AGE GROUPS STUDENTS ON ZOOM PLATFORM USING FUZZY AND DEEP LEARNING

Ati Jain[1]
Hare Ram Sah
Harsha Atre

Received 26.02.2023.
Accepted 04.09.2023.
UDC – 159.925

Keywords:

Convolutional, Emotions, Facial Expressions, Fuzzy Logic, e-learning, Speech Recognition

A B S T R A C T

Modality is the word that refers to different modes of recognition. In early researches it has been seen that single modality more over taken into considerations for emotion recognition which always not give better results due to insufficient information. Multi-model functionality helps in overall recognition process for increasing reliability. Fusing multiple feature sets and classifiers into one system will produce a comparably more accurate system. This research works includes combination of speech and facial expressions, as this hybrid mode will help in evaluating results perfectly and as per desire. The proposed methodology used here consists of two main parts; first is Facial Expression Recognizer (FER) and Speech Emotion Recognition (SER) is second part and for the final result hybrid mode will work that consider output of both FER and SER. This is multi-sensory and multi-model emotion recognition system. The research work used MLP classifier for classification of emotions from speech/audio for speech emotion recognition with the use of MFCC for feature extraction that gives accuracy of about 81% and Convolutional Neural Network classifier is used for video/images that gives accuracy of 98.68%. As a hybrid system, outputs found better than that of individual systems. Age groups like kindergarten, primary, adults, senior citizens are compared for finding emotions based on the models. Such applications may be used at schools, colleges, universities, training centers etc.

1. INTRODUCTION

Emotions are the responses that happens when any situation is faced by human being. Importance of emotions are very much needed in our daily lives, so understanding emotions is matter of work with effortlessness and steadiness. Emotions can be sometimes physiological, behavioral or subjective. Response from the human body, action in the brain and thoughts, mindset and other activities relates to different

[1] Corresponding author: Ati Jain
Email: jain.ati@gmail.com

types of emotions (Panksepp 2011).

As a motivation, $COVID_{19}$ started with late 2019 from the month of November-December and it had become dangerous by the month of February- March of early 2020. By this time, it has been found compulsory that entire education system must be shifted to online mode in the whole world.

By online learning, student's opportunities and learnings continued but this has become a challenging task for teachers to recognize student's engagement during online class (Song et al 2004). For any educational systems, it has become an important task to understand the emotions of students learning in online class and knows their mood whether students are attentive in a class or not in terms of understandings and grasping the delivered information (Inkeles & Sirowy 1984).

This type of system can obviously overcome traditional pen and paper method, as they are not real time feedbacks and may be false written (Halder et al., 2016).

As earlier discussed, there are several metrices on which emotion detection is based of facial expression as well as speech. Artificial Intelligence's Deep Learning methods and algorithms are proved better than machine learning algorithms (Latha & Priya 2016) when there is matter for detection of emotions figure 1.

Emotion recognition can be done by detecting Facial Expression Recognition (FER) techniques in which face and its parts expressions are used as an input, do show a vital role (Huang, et al. 2019). With the notion of Image processing like techniques (capturing face image) face sentiments can be analyzed to detect emotions. Apart from this, nowadays a speech or an audio is also used to understands the emotions of a human by judging prosodic and temporal features. Speech Emotion Recognition (SER) uses a voice file (preferably .wav format) into the system that is preprocessed (Aparna, 2023) and based on certain metrics to detect emotions from speech (Spyrou et al. 2019).

To make more better and accurate system than existing systems hybrid approach technology using bimodal fusion technology is truly important. But instead of using single datasets for both models different must be tried. RAVDESS datasets are best for speech emotion recognition and FER2013 is appropriate for using for facial expressions to train the model. For face detection Haar Cascade method will be best suited and for classification CNN is applied using famous deep convolutional neural networks (concept of machine learning) for facial expression recognition and MLP classifier for speech recognition. As an individual as well as hybrid model gives better accuracy using mentioned approaches (Yang, Jiang & Guo 2019).

Apart from this, live face will be captured using online Zoom platform as well live audio will be recorded for emotion recognition and hence at last results will be combined to calculate attentiveness of students learning in e-class. Proposed model of this thesis research work fits best for achieving higher accuracy. One more newness added into this research work that this model will be studied for learners of different age groups like kindergarten (2-5 Years), primary (5-12 Years), juniors (12-18 Years), adult (18-24 Years) and company employees (25-60 Years).

Figure 1. Basic Emotions

- **Happy:** It is a positive emotion that shows joy from satisfaction. It is related to a smile that appears on the face of human. It is subject of human well-being and satisfaction that can be either just a smile to laughter's. Expression shown when narrowed eyes i.e. corner of the eyes shown like crow feet wrinkles.
- **Sad:** Failure or any loss many leads to Unhappiness i.e. sadness. It is a negative emotion that is happens usually when human is in lack of some degree. This emotion is associated when there is disappointment, helplessness or sorrow. Being cry is an indication for sad emotion. This is expressed on face with dropping upper eyelids, unfocused eyes and pull down of lips corner slightly.
- **Fear:** It is a negative emotion felt when human is under some risk, harm related to emotional or physical, or any disturbing conditions. One could also face fear in case of social interactions. For e.g. fast speed of COVID19 made human with full of fear. On face it can be shown as stretched lips, eyebrows slighted raised and towards each other and eyes may be open full.
- **Anger:** It is a negative and an intense emotional reaction that can harm to self as well as others as violence. Any non-cooperative response, unexpected event and behavior is also the cause for being angry. Staring in one direction, wide eyes, lowered eyebrows and drawn with each other are main expressions of showing anger.
- **Surprise:** When anything happens sudden that is if unexpected moment or sentence that causes Surprise i.e. seems like wonder. It is positive emotion that comes for a while and without any

questioning. Surprise may be positive or negative sometimes depends upon the situation of a person. It is shown with eyebrows raised, eyes widened and dropping of jaws.

- **Disgust:** It is a negative emotion normally evoked with factors like smell, vision, sound, taste, touch and related senses. They may also occur due to unexpected of some ideas and appearance of person or not liking of served meal. It can be judged when in the rise of the upper lip, eyebrows when lower and wrinkled nose.
- **Neutral:** It is neutral emotion it means than there may be balance in positive and negative emotion. Some predict that there is no neutral emotion exist as there may be some kind of feeling that is present in human.

The paper further organized as follows: Section 2 describes the literature survey i.e. work done by other authors in the related field. Next, Section 3 describes about Mathematical formula that is created for calculation emotions mathematically. Section 4 describes about proposed methodology of the proposed problem and at last Section 5 notifies screenshots and discussions about the built model and last, section 6 presents conclusion and future scope of present work.

2. LITERATURE SURVEY

A vast of work has been already started in the field of emotion detection using either facial expressions or speech emotion recognition systems (table 1). But there are some systems that used hybrid system for the better results. Hybrid here means final results of student's emotions on the basis of both face emotions and speech. Most of the systems used Deep Convolutional Neural Networks for emotion detection using various different datasets and good accuracy is achieved from such systems.

No authors worked on live data of students learning with zoom platform and neither emphasized on different age group differences for emotion detection considering both facial expressions and speech. Some of the researcher's work are as:

- **Khan et al 2002**, in paper explained that face and speech is the best tool to identify emotions and better way is always tried by scientists to find the solution. Instead of unimodal using face, speech must also be fusion. Proposed method only communicated idea to identify emotions without dealing with depth details. Numerous methods There are several methods used for the detection of face, one out of them named as Viola Jones algorithm which is widely used and Relative Sub Image Based Features method is proposed while for speech RBFC method was proposed. SVM used as classifier to identify emotions. LibSVM tool is prescribed for implementation but no accuracy of the system was discussed.
- **Yutai Wang et al. in 2013** (Yutai, 2013), in paper, proposed bimodal fusion method for emotion recognition. Gaussian Mixture model was proposed for further findings and Chinese datasets used for both facial and speech recognition. It has been mentioned by an author that accuracy achieved is 6% higher than that of single model for bimodal fusion proposal.
- **Humaid Alshamsi et al. in 2018** (Alshamsi, 2018), in paper, identified emotions for real time in a mobile application built for Android systems. SAVEE and RAVDESS datasets were used for the same that results up to 97% appx. Cloud technology is used here and for feature extraction MFCC techniques applied and SVM is used as a classifier and here technique used is multi model for feature extraction from the audio input for emotions. For the implementation of this work, MATLAB with Android features are combined for better accuracy.

2.1 Emotions in Different Age Groups

As per researches, it's a lot easier to hold the attention of higher-level students/attendees during online classes rather than those in kindergarten and primary levels (Uçar, 2017). Laissez-faire approaches must see carefully to much more physical classroom rules in compare with online classes that range widely from student behavior expectations. This students' learning attitude put learning effect that makes a difference to with this approach. Facial lexis and emotions by speech may be the easiest for anyone to recognize in learners.

- **Gao and Maurer et al (2010)**, proposed that different age groups show emotions at different level and time. At the age of five it is seen that most of the children are able to signify happy emotions on their faces with adult like accuracy and other emotions like sadness, anger, disgust, and surprise take much longer timer to grow up that creates an emotion level difference in age groups.
- **Chronaki et al (2015),** stated that until child is 11 years old, it is difficult to find out the complete emotions out of them as compare them with adult like competency. Regular development of emotion insight is a critical criterion for social capability and communication skills in children. It has now been well-known that the ability to notice and judge emotions in faces and voices advances and increases with age throughout childhood (Dhirender, 2021).

Table 1. Comparative Study of related papers

Reference	Year	Purpose	Method	Datasets	Accuracy
Malyala Divya et al.	2019	automated live facial emotion recognition	CNN model	CSV file	66%
Rituparna Halder et al.	2016	human emotions from facial expressions	neural network	JAFFE database	90%
Harshala Chaudhari et al.	2015	comparative techniques for recognizing emotions	SVM	JAFFE database and Cohn Kanade database	85%
Leh Luoh et al.	2010	Recognize facial expression	Gaussian mixture model	JAFFE database	90%
M. Ali Akber Dewan et al.	2019	engagement detection	LBP-TOP and Deep Multi-Instance Learning (DMIL)	DAiSEE dataset	75.77%
Archana Sharma et al.	2019	facial emotion recognition	CNN model	Cohn-Kanade (CK+)	85%
Hemanth Singh et al.	2017	facial expressions using deep learning	convolutional neural network	JAFFE	86%
Sahla K. S. et al.	2016	deep learning method for emotion analysis	deep convolutional neural networks	JJAFEE	90%
Reakaa S D et al.	2021	Comparison of Speech emotion Recognition	SVM	RAVDESS	60%
Nahla Nour et al.	2020	Facial Expression Recognition	CNN model with SVM classifier	CK+	88%
P. K. Sidhu et al.	2022	Face Emotion Recognition	CNN	Jonathan Oheix	67%
Supreet U Sugur et al.	2019	Facial Expression recognition	CNN	FER2013	87%
Renju Renjith et al.	2021	Facial Expression recognition	VGG16 Net (CNN)	CK+	97%
Dr. B. S. Daga et al.	2021	Speech Emotion recognition	CNN	RAVDESS	92%
Nikhil Rai et al.	2021	Facial Expression recognition	CNN with transfer learning	FERC2013	97%
Lucy Nwosu et al.	2017	Facial Expression recognition	CNN	JAFEE	90%
Vaibhav K.P. et al.	2021	Gender and emotion through Speech	CNN	TESS	78%
Dr. N. Herald Anantha Rufus et al.	2021	Speech Emotion Recognition	CNN	RAVDESS	90%

3. MATHEMATICAL IMPLEMENTATION BY FUZZY EXPERT SYSTEM REPRESENTATION

Fuzzy means steps from being precision to fuzziness (figure 2). It is defined for approximation of human reasoning that uses inputs and outputs.

There are many systems in which it is difficult to predict ‘yes’ or ‘no’ in some situations. In this case, fuzzy logic helps to reach some level of reasoning for the uncertainty. Decision making application uses fuzzy logic to solve need of daily lives.

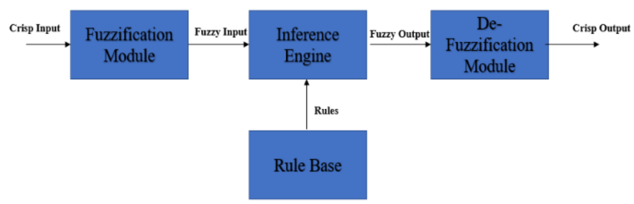

Figure 2. Fuzzy System overview

- **Rule Base:** Decision making is an essential part of fuzzy system in which framing of rule is important with membership function. This helps in building a decision and controlling any system by some conditions like IF-THEN.
- **Fuzzifier:** In order to get fuzzy sets, inputs are fed to fuzzifier that do conversion of input to fuzzy sets. Fuzzy sets are useful for system for various processing for finding uncertainty.

- **Inference Engine:** Every input is specified with some particular rule. This is the tool that performs the said operation. Fuzzy output is always generated by input used with rule based.
- **Defuzzifier:** This is last stage that gives output in the form of crisp values.

Fuzzification is always based on rules and it is necessary to create rules to produce output.

3.1 Facial Expression Recognition:

- **If** (upper eyelids pulled up, lower eyelids pulled up, Lowered and knit together, staring intensely) and (margins of lips rolled in lips may be tightened) and (One side of the mouth raised) **then** (Stage 1 is **Anger**).
- **If** (upper eyelids pulled up, lower eyelids pulled up, Lowered and knit together) and (Eyebrows pulled down) and (margins of lips rolled in lips may be tightened) **then** (Stage 2 is **Disgust**).
- **If** (Eyebrows pulled down) and (lip stretcher) and (mouth stretched, Open mouth) **then** (Stage 3 is **Fear).**
- **If** (Muscle around the eyes tightened, “crows feet” wrinkles around the eyes) and (lip corners raised diagonally) and (Raised corners) **then** (Stage 4 is **Happy**).
- **If** (Raised and arched eyes) and (lip corners raised diagonally) and (Raised corners) **then** (Stage 4 is **Happy**).
- **If** (Inner corners of eyebrows raised, eyelids loose, Lowered and knit together, looking away) and (lip corners pulled down) and (Corners that are drawn down) **then** (Stage 5 is **Sad**).
- **If** (Raised and arched eyes, eyelids pulled up) and (lips part) and (mouth hangs open, A dropped jaw) **then** (Stage 6 is **surprise**).
- **If** (Inner corners of eyebrows raised) and (lips part) and (mouth hangs open, A dropped jaw) **then** (Stage 6 is **surprise**).
- • **If** (eyes looking away) and (lip corner puller, lip together) and (mouth hangs open, A dropped jaw) **then** (Stage 6 is **Neutral**).
- **If** (eyes looking away) and (eyes brows looking away) and (mouth hangs open, A dropped jaw) **then** (Stage 6 is **Neutral**).

Rule A: If (Muscle around the eyes tightened, “crow’s feet” wrinkles around the eyes) and (lip corners raised diagonally) and (Raised corners).

$$\text{Min}\left\{\frac{0}{20}+\frac{0.7}{30}+\frac{0.71}{35}+\frac{1}{40}+\frac{0}{50}\right\}=0$$

Rule B: If (Raised and arched eyes) and (lip corners raised diagonally) and (Raised corners).

$$\text{Min}\left\{\frac{0.5}{55}+\frac{0.4}{66}+\frac{0.71}{35}+\frac{0.5}{40}+\frac{0.7}{43}\right\}=0.4$$

Now, obtain the degree of membership for the system as max (Rule A, Rule B) = max (0,0.4) = 0.4, that means output is Stage 1: Happy (according to linguistic variable).

3.1.1 De-Fuzzification:

The membership function for Stage 4(Happy) (figure 3):

$$\mu_{\text{Stage 4}}(\text{Happy})=\left\{\frac{0}{2.0}+\frac{0.6}{3.3}+\frac{0.6}{3.7}+\cdots+\frac{0}{4}\right\}$$

$$X^{*}=\frac{0.6*3.3+0.6*3.7}{0.6+0.6}=\mathbf{3.5}$$

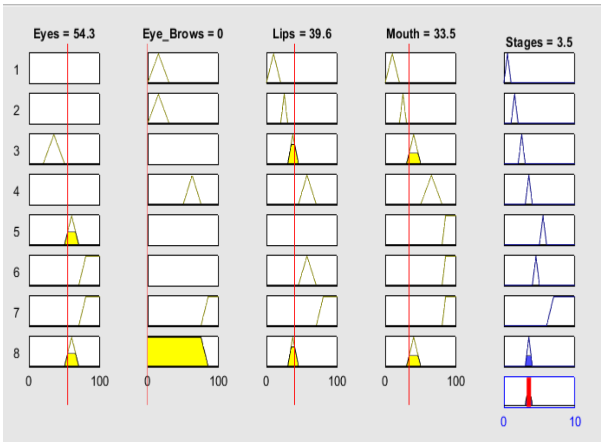

Figure 3. Face Rule Viewer

It can also be seen that the value of output variable stage is 4 i.e., student considered Happy during online class which is in stage 4 according to linguistic variable.

3.2 Speech Emotion Recognition:

- **If** (Pitch level is high) and (Voice quality/ Rate is very High) and (Volume /Intensity is Very high) and (Pronunciation Rhythm) **then** (Stage 1 is **Anger**).
- **If** (Pitch level is Low) and (Voice quality/ Rate is Low) and (Volume /Intensity is Very Low) and (Pronunciation Fumble) **then** (Stage 1 is **Fear**).
- **If** (Pitch level is high) and (Voice quality/ Rate is high) and (Volume /Intensity is high) and (Pronunciation Correct) **then** (Stage 1 is **Happy**).
- **If** (Pitch level is Normal) and (Voice quality/ Rate is Normal) and (Volume /Intensity is high) and (Pronunciation Correct) **then** (Stage 1 is **Happy**).
- **If** (Pitch level is Low and (Voice quality/ Rate is Normal) and (Volume /Intensity is Normal) and (Pronunciation incorrect **then** (Stage 1 is **sad**).
- •**If** (Pitch level is thin voice) and (Voice quality/ Rate is Normal) and (Volume /Intensity is normal) and (Pronunciation correct) **then** (Stage 1 is **Neutral**) (figure 4).

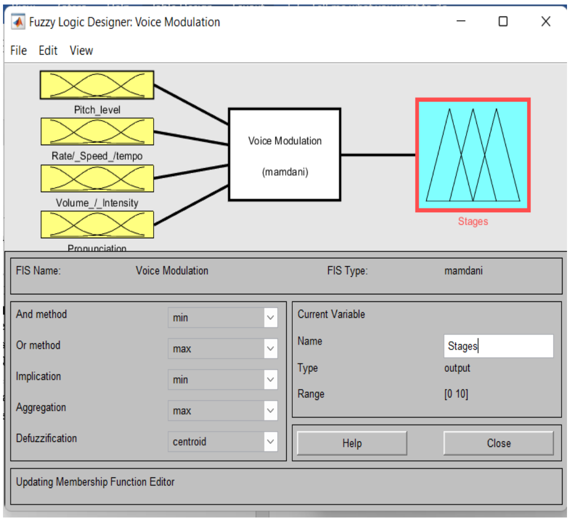

Figure 4. Voice Modulation Factors

Rule A: If (Pitch level is high) and (Voice quality/ Rate is high) and (Volume /Intensity is high) and (Pronunciation Correct).

$$\text{Min}\left\{\frac{0.5}{60}+\frac{0.6}{65}+\frac{0.8}{80}+\frac{1}{75}\right\}=0.6$$

Rule B: If (Pitch level is Normal) and (Voice quality/ Rate is Normal) and (Volume /Intensity is high) and (Pronunciation Correct).

$$\text{Min}\left\{\frac{0.7}{30}+\frac{0.2}{38}+\frac{0.24}{66}+\frac{1}{75}\right\}=0.2$$

Now, obtain the degree of membership of the system as max (Rule A, Rule B) = max (0.6,0.2) = 0.6, that means output is Stage 4: Happy (according to linguistic variable).

3.2.1 De-Fuzzification:

The membership function for Stage 4(Happy):

$$\mu_{\text{Stage 4}}(\text{Happy})=\left\{\frac{0}{5}+\frac{0.67}{6}+\frac{0.3}{7.5}+\frac{0.07}{7.9}+\frac{0}{8}\right\}$$

$$X^{*}=\frac{0.67*6+0.3*7.5+0.07*7.9}{0.67+0.3+0.07}=6.5$$

It can also be seen that the value of output variable stage is 4 i.e., student considered Happy during online class which is in stage 4 according to linguistic variable of voice (figure 5).

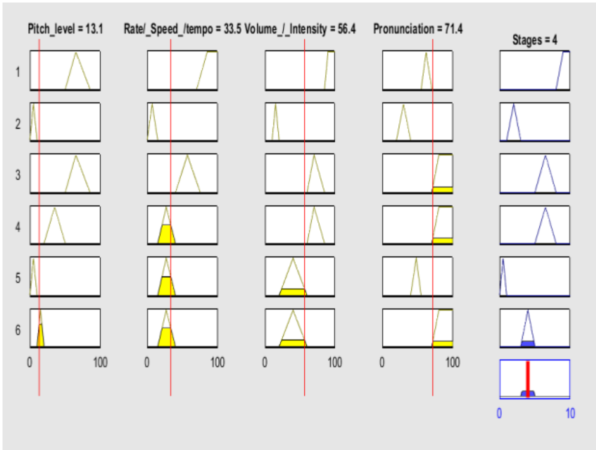

Figure 5. Speech Rule Viewer

4. BI-MODEL HYBRID SYSTEM

The way to find out human's expression is by face. Face has many parts like mouth, lips, nose, head and mainly eyes (Laraib et al. 2023). From all this, it is easy to find and understand what they are feeling as these parts responds appropriately. According to some studies, it has been cleared that other than face, speech is one of the most important aspects to connect with emotions of the human (figure 6).

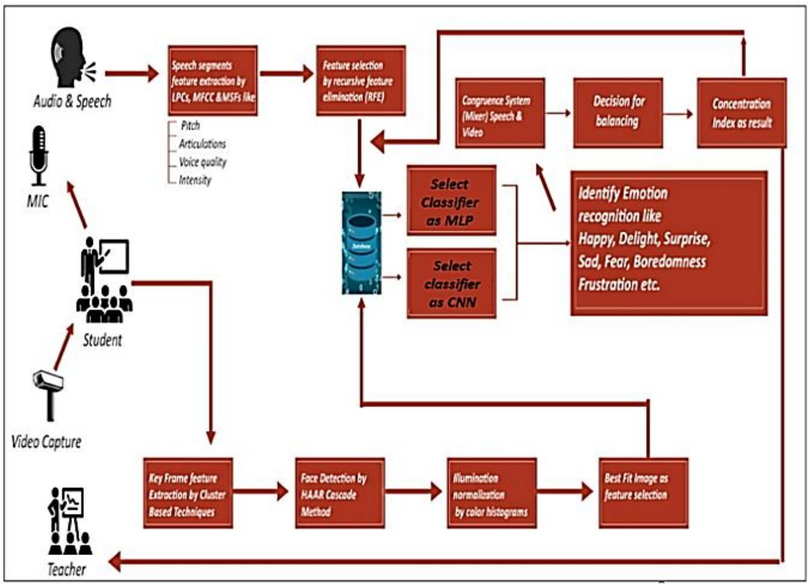

Figure 6. Hybrid System for Facial Expression and Speech

Emotional state can be better understood by speech of a person that contains many speech factors. While speaking to someone, speech can be judged as either monotone, sounds down and sometimes tired. Nervousness, excitement and enthusiasm are some of the factors that is easily judged by a voice (table 2).

In order to judge attentiveness and emotional state of a person, the results would be better by the combinations of emotions by speech as well as facial expressions. Due to this hybrid mode, both positive and negative emotions can be judged better as human psychological signals gets integrated with facial expressions (figure 7).

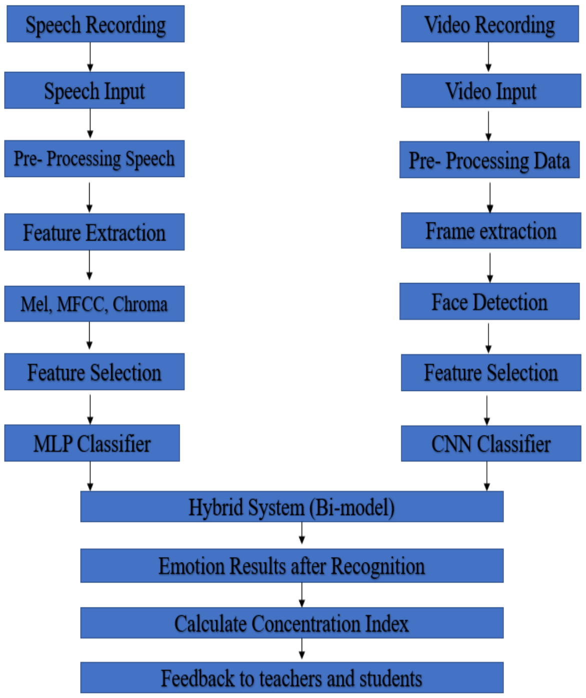

Figure 7. Steps for Hybrid System for Facial Expression and Speech

Table 2: Emotion Combinations

Primary Emotion	Primary Emotion	Combined Emotions
Neutral	Surprise	Confusion
Neutral	Angry	Frustration
Neutral	Happy	Satisfaction
Sad	Surprise	Dissatisfaction
Fear	Surprise	Alarm
Disgust	Angry	Remorse
Fear	Sad	Despair
Disgust	Surprise	Disbelief
Sad	Angry	Envy
Fear	Disgust	Shame
Surprise	Angry	Indignation

4.1 Face Emotions Recognition in Children

This is big question of anyone’s mind that is there is possibility of children’s emotion recognition in early childhood. By some study, it has been proved that infants with 4-6 months old can easily discriminate between anger and fear, starting from negative emotions. 4-5 years old children able to understand basic emotions and upto 7-8 years children hidden emotions, their desires and some communication beliefs that helps in understanding the other person mental status (Du, Crespo & Martínez 2023). Up to 11-12 years little more increment is being seen with some more moral values with all emotion types. Mixed emotions when seen in children that is considered as an adolescent state, this means that proper well emotion recognition can be done in this stage only. With the age, emotion recognition becomes better and results also improved. These clearly means that from kindergarten to primary, primary to secondary, secondary to adolescent and young age to middle and senior age emotions can be categorized according to age and many developments had been seen with growing age (table 3).

Table 3. Emotions at Different Age Groups

S.N.	Category	Age Groups	Emotions
1	Kindergarten	2-5years	Happy, Neutral, Anger
2	Primary	5-12 Years	Happy, Neutral, Anger, Disinterested, Confusion
3	Juniors	12-18 Years	Happy, Neutral, Anger, Disinterested, Confusion, Sadness, Disgust, Surprise
4	Adult (College)	18-24Years	Happy, Neutral, Anger, Disinterested, Confusion, Sadness, Disgust, Surprise, Delight, Fear, Learning
5	Experienced (Company Employee)	25-60Years	Happy, Neutral, Anger, Disinterested, Confusion, Sadness, Disgust, Surprise, Delight, Fear, Frustration, Learning

4.2 Speech Emotions Recognition in Children

Vocal emotions are best recognized from the prosody of spoken sentences, that helps in recognizing emotions with development trajectories. Infants recognize mother's voice just after their birth only and also open eyes with higher propensity in case of happy emotion rather than sad, angry or neutral emotions (Pervaiz & Ahmed, 2016). With aging, like face expressions recognition, speech emotion recognition also improves; as speech goes more clearer with growing age. Stability in emotions can be normal only with when children enter into secondary stage from primary. More number of emotions can be judged after the age of 12 and even happiness and sadness improve a lot after this age. The time course of children emotional development is totally set with ages and clearly shows that number of emotions in infant is less than in comparison with adolescent as best development seems at this time.

By some study, it has been cleared that in comparison with face and speech emotions recognition, face emotions are clear with early childhood than speech and speech are better to recognize in late childhood rather than early stage. Child's voice quality and consistency in speech is quite low as compare with adolescent.

"These are informational signals that can be useful; they're not necessarily decisive" Josh Dulberger, Head of Product (Data and AI) at Zoom, quoted.

Exhibition of expressions and emotions of a human is well given by face and speech. Multimodal is used when advantages of both the speech and facial expressions are taken to recognize right emotions. Though, it a job of challenge to detect right emotions from speech and face input. In this design framework, microphone will be used to collect speech for emotion recognition and facial expression is recognized with the help of webcam, closed circuit, video surveillance or any other available camera.

Finding a right way to extract features can obviously give best classification technique for best accuracy of the system. Feature vectors are used to add extracted features from both audio and video, there are certain methods proposed for the same. For the facial expression recognition CNN classifier is used, while for speech emotion recognition MLP classifier is applied. Finally result will show that how hybrid mode using both face and speech results better than using single model either speech or face individually. Recorded audio files and video files converted into images are taken as in input for hybrid system.

4.3 Concentration Index

Concentration Index is calculated by multiplication of emotions that has been recognized with the emotion's weight which is corresponded (as shown in table below as well) (Haritha, 2021) (table 4). This value is assigned between 0 to 1, considering positive emotions near to 1 and negative emotions near 0, that can be categorized as high concentration level, low concentration level and medium concentration level. Mental status of a person is also showing different emotion weights.

Table 4. Concentration Classification Chart

S. N	Name of Emotion	Concentration Index	Type of level
1	Happy	10	High Concentration
2	Neutral	9	High Concentration
3	Surprise	8	High Concentration
4	Delight	7	High Concentration
5	Sad	6	Medium Concentration
6	Boredomness	5	Medium Concentration
7	Angry	4	Low Concentration
8	Fear	3	Low Concentration
9	Disgust	2	Low Concentration
10	Frustration	1	Low Concentration

5. IMPLEMENTATION & RESULTS

Implementations and results are presented on figures 8-12.

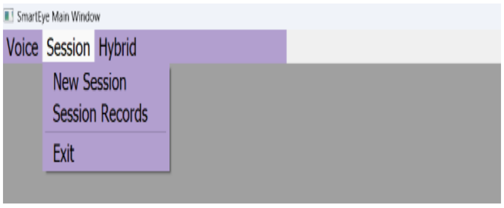

Figure 8. Screenshot of second step to either select recorded session or start new session

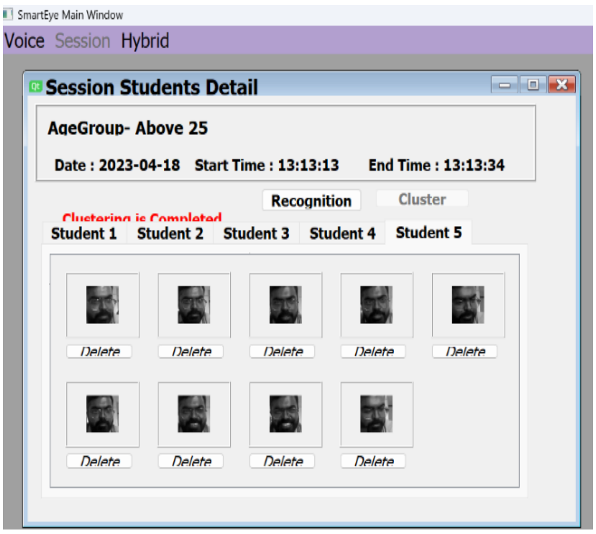

Figure 9. Screenshot of next step for recognition of emotions from clustered images for Age Groups (Senior)

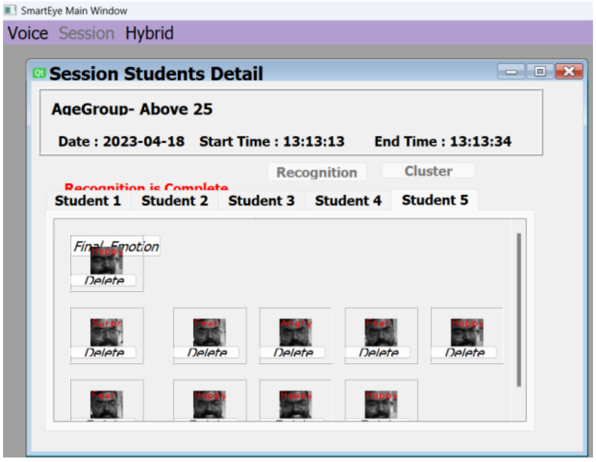

Figure 10, Screenshot of next step with images of students learning online with different identified images for Age Groups (Senior)

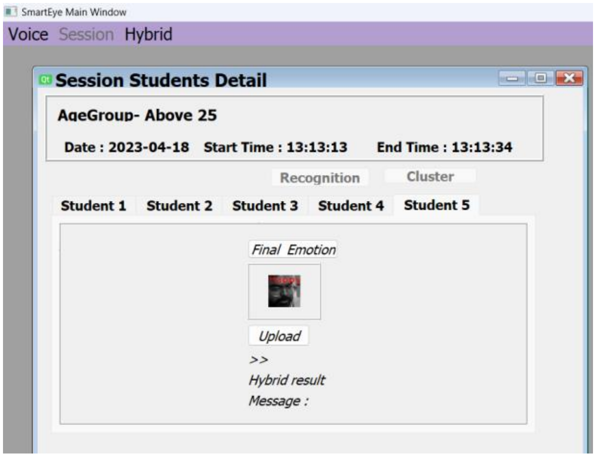

Figure 11. Screenshot showing final result of facial expression for Age Groups (Senior)

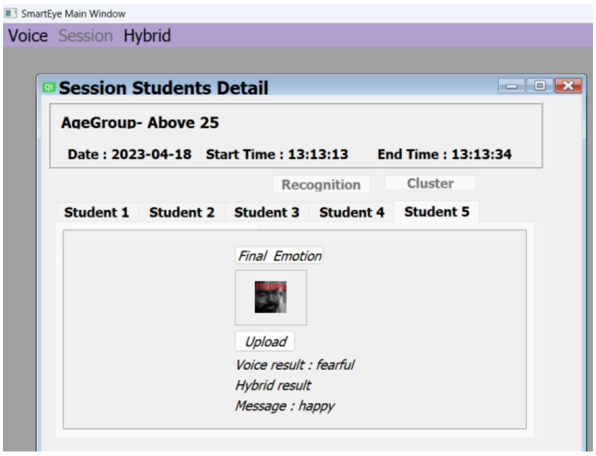

Figure 12. Screenshot showing audio as well as hybrid model final emotion for Age Groups (Senior)

6. CONCLUSION & FUTURE SCOPES

One great line is written in Shakespeare's play that it is difficult for anyone to speak about any another individual's concentration or trustfulness from only the face, though voice is also useful for finding emotions.

The human–computer communication will be further more natural if computers are gifted to see and respond to human nonverbal communication such as emotions. It has been observed that though numerous approaches have been projected to identify human emotions either from the speech or grounded on facial expressions, comparatively very few works have been thru to fuse these two, in command to improve the accuracy and robustness of the emotion recognition system. In this thesis work, this feature is addressed by examining the assets and which also adds to its boundaries of different systems based only on facial terms or acoustic information, and suggest a fusion of numerous information channels such as audio and video and finds how this ability develops with age. The system represents different emotions by facial and speech recognition individually. And hence, concentration index of students is calculated and by the same feedback to the faculty is provided for the improvement of lecture delivery.

When there is matter of Speech emotion recognition there are plenty of data sets available. RAVDESS datasets are used widely which is perfect datasets for speech emotions and so model which is built up consider RAVDESS data set only. These apprehensions the audio file and hence extract features using MFCC and categorize the emotions using MLP classifier. The accuracy that is recognized for the considered model finds to be 81%. Two approaches were applied here for recognition: first one is used to recognize emotions from live audio fed by user from an audio device to the system and second one by uploading zoom recording of participants attending online class.

For the facial expression recognition, FER2013 datasets are used on which facial expression recognition model is built up. This captures the images of the face by frame extraction using clustering, extract features and classify the emotions using CNN classifier by recognition. The recognition accuracy of chosen base model finds to be 98.68%. Two approaches were applied here for recognition: first one is used to recognize emotions from live video (converted to images) captured from the zoom participants attending online class (via webcam) and second one by inputting image by own. The FER2013 datasets is applied to test the accuracy of the approach and hence it achieves up to 98.68%.

At last, results of facial and speech emotion recognition finds as an individual are combined and produced as a fusion, that helps in better recognition. Further, two frequently used approaches of the fusion named as fusion with feature and score fusion are also observed and anticipated a new multi-level fusion method for enhancing the person-dependent and person-independent classification performance for different emotions by comparing it with different age groups of attending participants.

Acknowledgement: I would like to thanks my guide Dr. Hare Ram Sah, my friend Harsha Atre and Dr. Abhay Kothari sir for guiding all the time.

References:

Alshamsi, H., Kepuska, V., Alshamsi, H., & Meng, H. (2018). Automated facial expression and speech emotion recognition app development on smart phones using cloud computing. *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 730–738. https://doi.org/10.1109/IEMCON.2018.8614831

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, *6*(1), 1. https://doi.org/10.1186/s40561-018-0080-z

Divya, M., Reddy, Dr. R. O. K., Raghavendra, C., & Dundigal, H. (2019). Effective facial emotion recognition using convolutional neural network algorithm. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(4), 4351–4354. https://doi.org/10.35940/ijrte.D8275.118419

Halder, R., Sengupta, S., Pal, A., Ghosh, S., & Kundu, D. (2016). Real time facial emotion recognition based on image processing and machine learning. *International Journal of Computer Applications*, *139*(11), 16–19. https://doi.org/10.5120/ijca2016908707

Nour, N., Elhebir, M., & Viriri, S. (2020). Face expression recognition using convolution neural network (Cnn) models. *International Journal of Grid Computing & Applications*, *11*(4), 1–11. https://doi.org/10.5121/ijgca.2020.11401

Pervaiz, M., & Ahmed, T. (2016). Emotion recognition from speech using prosodic and linguistic features. *International Journal of Advanced Computer Science and Applications*, *7*(8). https://doi.org/10.14569/IJACSA.2016.070813

Reakaa, S. D., & Haritha, J. (2021). Comparison study on speech emotion prediction using machine learning. *Journal of Physics: Conference Series*, *1921*(1), 012017. https://doi.org/10.1088/1742-6596/1921/1/012017

Sahla, K. S., & Senthil Kumar, T. (2016). Classroom teaching assessment based on student emotions. In J. M. Corchado Rodriguez, S. Mitra, S. M. Thampi, & E.-S. El-Alfy (Eds.), *Intelligent Systems Technologies and Applications 2016* (Vol. 530, pp. 475–486). Springer International Publishing. https://doi.org/10.1007/978-3-319-47952-1_37

Sharma, A., Mansotra, V. (2019). Deep learning-based student emotion recognition from facial expressions in classrooms. *International Journal of Engineering and Advanced Technology*, *8*(6), 4691–4699. https://doi.org/10.35940/ijeat.F9170.088619

Sidhu, P. K., Kapoor, A., Solanki, Y., Singh, P., & Sehgal, D. (2022). Deep learning based emotion detection in an online class. *2022 IEEE Delhi Section Conference (DELCON)*, 1–6. https://doi.org/10.1109/DELCON54057.2022.9752940

Ucar, A. (2017). Deep Convolutional Neural Networks for facial expression recognition. *2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 371–375. https://doi.org/10.1109/INISTA.2017.8001188

Wang, Y., Yang, X., & Zou, J. (2013). Research of emotion recognition based on speech and facial expression. *TELKOMNIKA Indonesian Journal of Electrical Engineering*, *11*(1), 83–90. https://doi.org/10.11591/telkomnika.v11i1.1873

Chronaki, G., Garner, M., Hadwin, J. A., Thompson, M. J., Chin, C. Y., & Sonuga-Barke, E. J. (2015). Emotion-recognition abilities and behavior problem dimensions in preschoolers: evidence for a specific role for childhood hyperactivity. Child Neuropsychology, 21(1), 25-40.

Gao, X., & Maurer, D. (2010). A happy story: Developmental changes in children’s sensitivity to facial expressions of varying intensities. Journal of experimental child psychology, 107(2), 67-86.

Khan, M. U., Choudry, Z. A., Aziz, S., Naqvi, S. Z. H., Aymin, A., & Imtiaz, M. A. (2020, June). Biometric authentication based on EMG signals of speech. In 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE) (pp. 1-5). IEEE.

Laraib, U., Shaukat, A., Khan, R. A., Mustansar, Z., Akram, M. U., & Asgher, U. (2023). Recognition of Children’s Facial Expressions Using Deep Learned Features. Electronics, 12(11), 2416.

Du, Y., Crespo, R. G., & Martínez, O. S. (2023). Human emotion recognition for enhanced performance evaluation in e-learning. *Progress in Artificial Intelligence*, *12*(2), 199-211.

Panksepp, J. (2011). Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloS one*, *6*(9), e21236.

Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2004). Improving online learning: Student perceptions of useful and challenging characteristics. *The internet and higher education*, *7*(1), 59-70.

Inkeles, A., & Sirowy, L. (1984). Convergent and divergent trends in national educational systems. In Current issues and research in macrosociology (pp. 137-165). Brill.

Latha, C. P., & Priya, M. (2016). A review on deep learning algorithms for speech and facial emotion recognition. *APTIKOM Journal on Computer Science and Information Technologies*, *1*(3), 92-108.

Huang, Y., Chen, F., Lv, S., & Wang, X. (2019). Facial expression recognition: A survey. *Symmetry*, *11*(10), 1189.

Spyrou, E., Nikopoulou, R., Vernikos, I., & Mylonas, P. (2019). Emotion recognition from speech using the bag-of-visual words on audio segment spectrograms. *Technologies*, *7*(1), 20.

Yang, C., Jiang, W., & Guo, Z. (2019). Time series data classification based on dual path CNN-RNN cascade network. *IEEE Access*, *7*, 155304-155312.

Ati Jain
Institute of Advance Computing,
Sage University,
Indore,
India
jain.ati@gmail.com
ORCID 0000-0002-6567-2591

Hare Ram Sah
Professor, Sage University,
Indore,
India
ramaayu@gmail.com
ORCID 0000-0003-3640-0869

Harsha Atre
Research Scholar, Department of
Applied Mathematics,
Jabalpur Engineering College,
India
atre.harsha.p22@gmail.com
ORCID 0000-0001-7753-1625