MJSAT 4 (1) 32-43

MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

Emotion Detection Using Machine Learning: An Analytical Review

Ashadu Jaman Shawon*[1], Anika Tabassum[1], and Rifath Mahmud[1]

[1] *Computer Science Department, American International University-Bangladesh. Dhaka, Bangladesh.*

KEYWORDS

Emotions
Detection
Machine Learning

ABSTRACT

This paper provides an analytical review of emotion detection using machine learning techniques. The study focuses on the attempts made in the past and present to develop emotion detecting systems. Numerous methods for identifying emotions have been introduced by students, and several outstanding works have already been found. In order to identify emotions, machine learning has been the primary focus. This paper also includes a summary of the frameworks, algorithms, and datasets that have been employed thus far to address this issue.

ARTICLE HISTORY

Received 23 September 2023
Received in revised form 22 December 2023
Accepted 26 December 2023
Available online 8 January 2024

1. INTRODUCTION

Facial expression is a nonverbal technique of expressing one's emotions [1]. And emotions are a fundamental and significant component of human behaviour that influence how people communicate [2]. Nobody can hide their feelings. So, in recent decades many automated intelligence systems have paid close attention to emotion detection. It might enhance how people and computers communicate. Emotion detection in educational contexts can inform students' engagement and comprehension. However, it can be challenging for a teacher to keep track of every student's eye. Other students may divert students' attention. Different learning styles may exist among students. Probably, the teacher won't have enough time to focus on each student individually. More distractions can occur in a larger class, which can affect both the teacher and the students. Emotion detection in the classroom is the process of identifying and understanding the emotional states of students in a learning environment. This can be done through a variety of methods, including facial expression recognition, body language analysis, and speech analysis. Emotion detection is a powerful tool that can be used to improve student learning and create a more positive and productive learning environment. However, there is no proper system available which can detect student's emotions in a classroom and give the report to the class faculty or teacher. To resolve this problem an automated emotion detection system can established.

2. BACKGROUND STUDY

Humans communicate emotion mostly through facial expressions. There are numerous efforts. As it features an automatic facial expression analysis function, to use in numerous fields [3]. Some can gauge the emotions of drivers to avoid accidents [4]. The accuracy of the predictions of 7 facial expressions is being tested in several research papers [5],[6]. Whereas some papers only predicted 5 facial expressions [7]. Based on multimodal data, researchers have been able to identify student engagement in several studies. Their research's primary goal is to aid online learning [8]. They mostly concentrate on six facial expressions [9],[10]. A system able to automatically detect human emotions was developed by Lim et al. in 2020. Based on input from several sensors, including electroencephalography (EEG) data, face expression, and eye tracking, the emotion detection system will assist in identifying human emotions [11]. Except this several research papers Investigating human emotion

*Corresponding author:
E-mail address: Ashadu Jaman Shawon <shawon573253@gmail.com>.
https://doi.org/10.56532/mjsat.v4i1.195

recognition through electroencephalogram (EEG) signals [12]. Besides this, voice signals are also used to determine emotions [13]. Regarding the purpose of creating reviews of students depending on the mental health of the students of a class, there are numerous manual software methods available. In 2021, Siam et al. introduced this kind of computer vision and deep learning-based student emotion detection system [14]. Furthermore, in 2020 Nishchal J. et al. will be able to automatically detect suspicious or unethical activity during a test. In their approach, the procedure is carried out by identifying the student's body position and seven different kinds of simple expressions throughout the exam using CCTV footage from the classroom [15]. on the other hand, build a bullying detection algorithm based on speech emotion recognition and motion recognition to accurately identify bullying incidents in real-time is the major goal of an article [16]. A study paper suggests an automated attendance system that makes use of face recognition technology, which is an innovative concept in the field [17]. An article applying deep learning techniques to detect emotions from facial expressions also uses an AI system, with the overall goal of improving the effectiveness of current models and making a contribution to the field [18]. In comparing classifiers for real-time facial emotion recognition, A separate research employs Gabor filters, Histogram of Oriented Gradients (HOG), and Local Binary Pattern (LBP) for feature extraction [19]. Physiological signals such as photoplethysmography (PPG) and galvanic skin response (GSR) are used to detect emotions in video clips in another article [20]. GoogLeNet uses an Inception module to collect both local and global information in the field of emotion recognition [21]. Face detection, facial feature extraction, and expression categorization are all involved in emotion detection. The technology recognizes facial traits, classifies emotions, and extracts features like the lips and eyes. A different approach makes use of SVM for classification and Oriented Fast and Rotated (ORB) for feature extraction [22]. While all the articles use a single model, one paper uses a hybrid model to identify emotions [23]. Emotions are categorized into positive and negative groups in an article and also use Speech Emotion Recognition (SER) [24]. In an article, facial expressions are evaluated utilizing image processing and machine learning for precise emotion identification in autistic children [25]. Physiological cues are utilized in another article to create a model for recognizing human emotions [26]. With less computing complexity, real-time emotion identification is accomplished with the virtual marker technique [27]. A different application for research Computer vision methods makes use of technology to identify and evaluate a range of human emotions [28]. Further research builds on this, investigating the development of an automated facial expression identification system specifically designed for those experiencing stress. The ultimate goal of this research is to improve human-machine interaction by integrating the processes of image acquisition, preprocessing, and classification [29]. Simultaneously, another study uses Support Vector Machines (SVM) and facial features to focus on the robustness of face identification and emotion detection. The goal of this research is to contribute to a number of fields, such as psychological research and identification [30]. A separate research project aims to build a reliable CNN-based real-time emotion detection system concurrently. For personal robots to understand and communicate with humans more individually, this system is considered essential [31]. The final research in this series focuses on using machine learning techniques to develop and deploy a real-time employee emotion detection system (RtEED). The system is designed to track employee satisfaction and send out automatic messages depending on identified feelings. Its goal is to provide people the knowledge they need to make wise decisions and promote happier, healthier lifestyles [32]. The comprehension and utilization of emotion detection technologies are improved by the cumulative efforts of this linked line of study. Moving on to the educational space, research is being done to create a real-time attention monitoring system for the classroom. It uses deep learning techniques to identify student behaviours and compares the effectiveness of several YOLOv5 models [33]. The primary objective of all these papers is to deploy the face expression system as accurately as possible by utilizing various technologies or systems.

3. REVIEW BASED ON THE MODEL

3.1 Convolution Neural Network

Convolutional Neural Network (CNN) is a member of the deep neural network or "deep learning" family of algorithms. It plays an important role in the field of emotion detection [34]. Depending on the input format, the array dimension changes. Convolutional filters, pooling/subsampling, and classification layers make up the three basic components of CNN [35]. A CNN may include hundreds of layers, each of which picks up the ability to identify different types of images [36]. This explains why CNNs has had such great success in identifying objects from images, facial identification, clinical image analysis, and many other areas. It can be trained using both B&W and RGB images [37]. Although some research indicates that convolutional neural networks are best suited for simple facial emotion recognition [38], they show better performance when working with larger datasets [39].

3.2 K- Nearest Neighbors

For classification and regression issues, k-nearest-neighbors (KNN) is a machine learning technique. It is regarded as a straightforward and understandable algorithm that fits into the supervised learning subset. Although KNN may produce the greatest results for classification, it has drawbacks of its own. Despite being computationally slow and requiring that all instances be kept in memory, it provides the highest accuracy [40].

3.3 Artificial Neural Network

Artificial neural networks (ANNs) mimic the functioning of brain cells and share similarities, with brain networks. Like the neurons in our brains, ANNs consist of interconnected layers that process data gradually, just as humans learn from their experiences. Similar to how people improve by making mistakes and learning from them, these networks adapt to minimize errors. They excel in tasks like language comprehension and object recognition in images. ANNs have proven beneficial across domains, including aiding autonomous driving systems and providing medical guidance. For instance, eye-pupil analysis systems have been developed using networks to detect emotional states by considering factors like cognitive load and stress levels acknowledging the complex relationship, between inputs and emotions [41]. In ways, they resemble problem solvers who learn from past mistakes and assist individuals effectively

3.4 Deep Neural Network

Deep neural networks (DNN) can be applied to extract complex patterns from unprocessed data to forecast emotions. Multi-layer deep connections, a novel idea, are introduced with deep neural networks (DNN). Since many connections are using this idea, the network's computation time and accuracy rise even if there are several layers. In deep neural networks, zero padding is a helpful method for preserving spatial dimensions, managing output sizes, and preserving edge information during convolution and pooling processes.

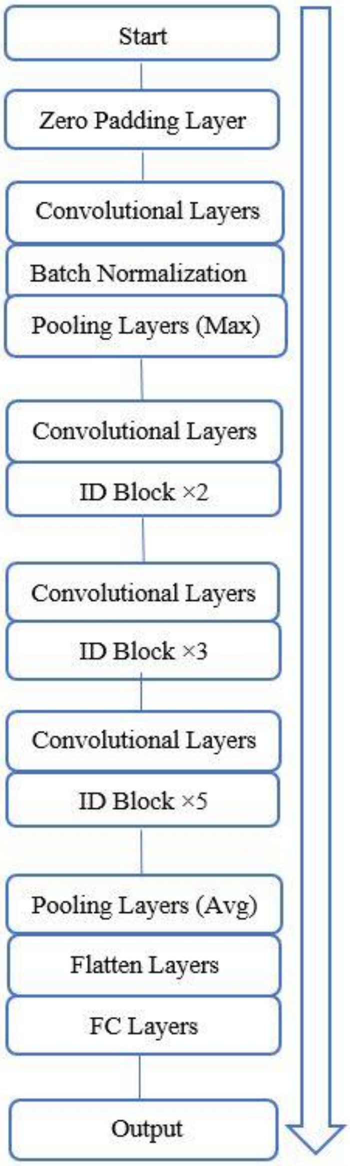

Fig. 1. A Proposed deep neural network [4]

The DNN processes the input image through a convolutional layer that employs filters to extract features from various regions of the image. To safeguard details about edges and maintain spatial dimensions, zero padding is applied. To reduce feature dimensions and information abstraction, a pooling layer is used after convolution. It facilitates the identification of key traits and lessens processing complexity. To gather crucial data and lower parameter dimensions, two kinds of pooling - average and max pooling -are used. Use of batch normalization to maximize instruction. Each layer's input data is normalized, and this aids the network in learning how the data is distributed. This increases training efficiency, avoids overfitting, and ensures the network's generalizability.

3.5 Other Models

A particular object detection model that expands on CNNs is called Faster R-CNN [42]. It is made to precisely locate and detect things in photos. It can detect faces at different resolutions, enhancing performance and computing efficiency. Faster R-CNN has recently achieved tremendous success in real-time facial expression detection in various scenarios [43].

The categorization of expressions, into emotional categories, relies on the use of a Multi-Layer Perceptron Neural Network (MLPNN). This neural network offers improved training flexibility. Effectively tackles the challenges associated with classification [44]. MLPNNs are specifically designed as feed-forward networks making them well-suited for classifying problems involving multiple classes. By combining layers of interconnected nodes MLPNNs can effectively understand patterns and reliably classify emotions based on facial expressions. Due to their design and ability to interpret emotional signals MLPNNs serve as an effective tool, for understanding human emotions through visual cues.

Additionally, there are also two types of neural network Excitatory Post Potential (EPP) and Partially Positive Recurrent (PPR) neural networks. However, the hybrid Excitatory Synaptic Potential/PPR neural network (EPPNN) brings together the features of both. By integrating the functionality of EPP with the memory capacity of PPR, EPPNN excels in processing data, identifying patterns, and comprehending temporal changes. It's akin to having a neural network at one's disposal [45]. This EPPNN is specifically designed to handle situations where both excitement and memories are crucial. The hybrid EPPNN, which combines EPP and PPR stands out in the field of networks because it shows promise, for tasks that require deep thinking and managing complex information.

Deep learning model are becoming increasingly popular in the emotion detection space because of their capacity to automatically extract complex patterns and representations from data. These models can efficiently extract complicated information from a variety of modalities, including speech signals, pictures, and physiological data. They are typically neural networks with many layers [46].

Hidden Markov Models (HMMs) are statistical models that have been used in various number of applications, including emotion detection. In speech, facial expressions, or physiological signals, temporal sequences of characteristics are retrieved and analysed using HMMs in the context of emotion detection [47].

An extensive and popular instrument for accurately characterizing facial motions is the Facial Action Coding System (FACS). Each face muscle contraction or relaxation is represented by a unique numerical code that FACS assigns to each AU. Experts can interpret someone's emotional state and read their facial expressions by combining these Aus [48]

The VGG 16 CNN model is widely known for its effectiveness, in image recognition tasks [49]. It is praised for its architecture consisting of 16 layers with 138 million learning parameters. While training this model from scratch can be time-consuming and data-intensive it can be used as a trained model where the CNN layers act as skilled feature extractors. The key advantage of VGG 16 lies in its layout, which utilizes convolutional filters across all layers. This allows it to extract hierarchies of information and accurately detect objects. However, due to its complexity using VGG 16 requires power. As a result, VGG 16 is commonly used as a benchmark, in computer vision applications representing a cornerstone of learning.

A subclass of artificial neural networks known as DCNNs is built using specialized convolutional layers that effectively capture spatial patterns, allowing for better performance in image analysis tasks such as facial expression identification [50].

Table 1. Summary of different Models for detecting emotions and layers.

Papers	**Models**	**Layers**
[1]	CNN	
[2]	AlexNet, GoogleNet	
[3]	FERC	2,4-layers
[4]	DNN	
[5]	XCEPTION-CNN	
[6]	CNN	
[8]	CNN	
[9]	VGG-19, FASTER R-CNN ResNet-50	50
[11]	ANNs	
[13]	CNN.LSTM	2
[12]	CNN, DCNN, FCN	
[14]	VGG-16, CNN	
[15]	3DCNN, XGBoost	
[18]	CNN	
[21]	CNN, SNA	8
[22]	ANN	
[24]	CONV1D	
[26]	DBN	
[29]	MLP	
[31]	CNN	20
[33]	YOLOv5	
[34]	CNN	
[35]	CNN	4
[39]	CNN	
[40]	NBd, NB, ID3, KNN	
[41]	ANN	3
[42]	Faster R-CNN	
[43]	Faster R-CNN	
[45]	EPPNN	
[46]	Deep learning	
[47]	HMM	
[48]	FACS	
[49]	VGG-16, CNN	
[50]	CNN, DCNN	

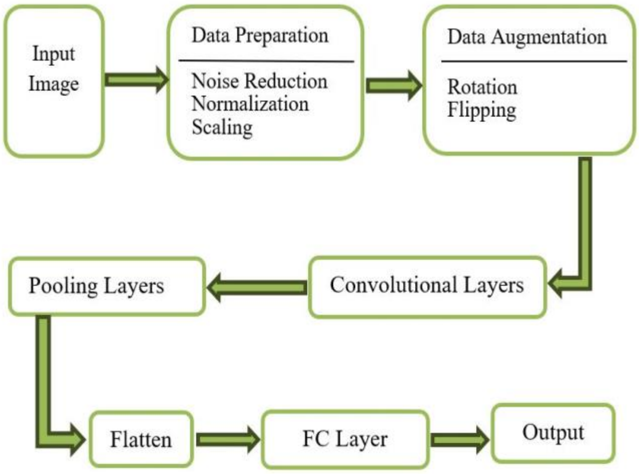

Fig. 2. A workflow chart of Emotion Detection System

Here it is found that CNN and DNN models prove to be the most effective. CNNs and DNNs represent superior choices among neural network architectures. It is best because DNNs are more flexible and can be used for a wide range of data types, including structured data (like numerical features in a table), sequential data (like text or time series), and even unstructured data (like raw audio). On the other hand, CNNs are primarily used for processing grid-like data, such as images or any data with a grid-like structure. They exploit the spatial relationships between neighbouring pixels in images.

4. REVIEW BASED ON DATASET

According to their needs, different kinds of datasets are used in all work in the field of emotion detection. Some of them applied speech, while others applied image, and so on. However, the primary goal of their dataset was to help train the machine to recognize various types of emotions. The datasets cited by the writers in their published works are provided below:

4.1 CMU Dataset

Videos of people interacting in a real-world setting can be found in this CMU dataset. It has audio data, a 3D body pose, and expressions on the faces. It is frequently used in linguistics research, synthesis of text into speech and detection of speech tasks in natural language processing. CMU has released several face datasets for research in facial recognition and analysis. These datasets often include images of faces in different poses, lighting conditions, and expressions.

4.2 Fer-2013 Dataset

The 2013 Facial Expression Recognition dataset (FER2013), a dataset made available through Kaggle, was introduced by Pierre-Luc Carrier and Aaron Courville at the 2013 International Conference on Machine Learning. Each image in the FER-2013 dataset is a grayscale image measuring 48 × 48 pixels. Each face in this dataset has been classified according to numerous emotion categories. The total number of micro expressions in the FER-2013 dataset is 35,887 and there are seven different types of emotions are there [1].

Fig. 3. Some Images of the Fer-2013 [51].

This is a widely used dataset for emotion detection systems which has seven facial emotion categories. Many researchers used this fer-2013 dataset in their system for emotion detection and some of them gained a good accuracy, on the other hand, most researchers get an average accuracy.

Table 2. FER-2013 dataset expressions and image numbers [51]

Expressions	**Test Data**	**Train Data**	**Total Data**
Angry	958	3995	4953
Happy	177	7215	8989
Sad	1247	4830	6077
Surprise	831	3171	4002
Disgust	111	436	547
Fear	1024	4097	5121
Neutra	1233	4965	6198
Total	7178	28709	358887

4.3 RaFD Dataset

RaFD (Radboud Faces Database) datasets vary mostly in the quality of images, quantity of images, and cleanness of images. RaFD displays clear facial expressions [15]. The Netherlands' Radboud University is where the RaFD dataset was created. RaFD includes a wide range of emotional expressions, including fundamental feelings like joy, sorrow, anger, fear, surprise, and disgust. Multiple subjects portray each expression, capturing various expressions of the same emotion. The RaFD dataset is used by researchers and programmers to test and improve facial expression analysis models, facial emotion recognition algorithms, and related computer vision applications. The dataset is a useful tool for advancing research in these fields because of its diversity and well-documented annotations.

4.4 AMIGOS Dataset

A large collection developed especially for researching human-robot interactions is called AMIGOS (A Multimodal Dataset for Affect, Personality, Image, and Language in Human-Robot Interactions). AMIGOS is an artificial intelligence and affective computing research platform that includes text transcripts, physiological signals, audio-visual recordings, and annotations of participants' emotional states during robot interactions. The development of emotionally intelligent robots is aided by the investigation of intricate socio-emotional dynamics made possible by this multimodal dataset. By using AMIGOS, researchers can better understand how machines can perceive and react to human emotions, which will help develop AI systems that are more responsive and empathic.

4.5 CK+ Dataset

Image sequences recording emotional face expressions make up the CK+ dataset. There are 327 labelled image sequences in all. These sequences are categorized according to various emotions, such as surprise, disgust, fear, happiness, neutrality, contempt, sadness, and anger. It includes 920 photos from the original 920 CK+ dataset that have previously been resized to 48x48 pixels.

Table 3. CK+ dataset expressions and image numbers [52].

Emotions	Testing and Training Image
Happiness	69
Neutral	593
Angey	45
Sadness	28
Fear	25
Disgust	59
Surprise	83
Contempt	18
Total	920

4.6 Wider Face Dataset

A well-liked standard dataset for testing and refining face identification systems is the WIDER FACE dataset. It is intended to test face detection models' abilities to deal with a wide range of real-world situations, such as changes in scale, position, occlusion, illumination, and more. With 32,203 coloured photos, it is one of the largest datasets. Based on several categories for both the training and testing images, starting with the index label range of 0 to 61.

Fig. 4. Some Images of the wider face dataset [53].

4.7 DISFA Dataset

The Denver Intensity of Spontaneous Facial Action (DISFA) dataset is an essential collection of facial movement data that has improved the science of facial expression detection. It is made up of 27 people performing a range of natural facial expressions in high-quality video recordings. The intensity of 12 facial action units (AUs), which are the fundamental components of facial expressions, are annotated on each frame in the videos. Several studies have made use of DISFA, such as those on autism spectrum disorder, emotion recognition, and automatic facial expression recognition. Additionally, commercial facial expression recognition software has been developed using it. Scholars can use the DISFA dataset for free for scholarly purposes.

4.8 Others Dataset

The Caltech Faces dataset is designed to aid researchers in developing and testing facial recognition algorithms. The dataset contains a collection of over 450 grayscale images of 27 individuals, with each person depicted in various poses, lighting conditions, and expressions. These images are relatively small, typically measuring 48x40 pixels, making the dataset particularly challenging due to its low resolution. The National Institute of Standards and Technology (NIST) has played a crucial part in fostering advancements in technology, including datasets used for benchmarking and evaluating various technologies, particularly in biometrics and cybersecurity. For testing recognition of handwriting, speaker recognition, and other pattern recognition tasks, NIST has made datasets available. These datasets are crucial for comparing algorithms, encouraging creativity, and fostering teamwork among researchers.

Japanese researchers at the National Institute of Information and Communications Technology (NICT) in Japan worked together to create the JAFFE dataset. To guarantee uniform lighting, backgrounds, and image quality, the images are taken under carefully monitored circumstances. The dataset contains images that depict the seven different emotional states of happiness, anger, contempt, fear, surprise, sadness and neutrality. To guarantee uniform lighting, backgrounds, and image quality, the images are taken under carefully monitored circumstances. Due to the controlled environment, researchers can only concentrate on the subject's facial expressions. The JAFFE dataset has been used by developers and researchers to test and train facial emotion recognition algorithms. Due to this, deep learning and machine learning models have been created that can precisely identify emotions from facial images.

A renowned set of facial expression images created for use in research in psychological science, neuroscience, and computer vision is the Karolinska directed emotional faces (KDEF) dataset. The KDEF dataset is dedicated to capturing a wide variety of emotional expressions and offering a standardized set of facial stimuli for various experiments and studies. The dataset contains pictures of different people displaying a variety of emotions, such as happiness, sadness, surprise, anger, fear and neutral faces. To ensure accuracy, these expressions are carefully posed and labelled by humans. Researchers at Kyungpook National University in South Korea developed the facial expression recognition dataset recognized as the KMU-FED. The images were captured using an NIR camera. The dataset includes 55 image sequences of 12 different persons representing the six basic emotions of surprise, fear, anger, happiness, sadness, and disgust. An NIR camera, which is less sensitive to changes in illumination than a visible light camera, was used to take the pictures. As a result, the KMU-FED dataset is ideal for use in driver assistance systems, where the capacity to detect facial expressions in low-light situations is crucial.

A significant 3D facial expression dataset called BU-3DFE was gathered by Binghamton University researchers. It comprises 2500 3D facial scans of 100 participants displaying the six basic facial expressions of anger, disgust, fear, happiness, sadness, and surprise at four different brightness levels (low, medium, high, and extreme). The scans, which were made with a structured light scanner, give a precise representation of the facial surface geometry of the subjects, including the location and configuration of the eyes, nose, mouth, and jaw. The BU-3DFE dataset was used by researchers to examine how pose and illumination affect 3D facial expression recognition. They observed that pose and illumination can significantly affect how well 3D facial expression recognition algorithms perform. However, they also discovered that by utilizing normalization techniques, these effects could be reduced.

4.9 Process with Image

Numerous studies have investigated various datasets and approaches in the area of facial expression identification to increase accuracy. In their study, Zahara et al. made use of the fer-2013 database to create a system that could identify seven different facial expressions in test images shot at a 0-degree angle. The system's capacity to recognize several expressions from these various angles was demonstrated by testing using images rotated from -15 degrees to +15 degrees, as well as 15 degrees and 345 degrees [1]. Moving forward, Giannopoulos et al. used the FER-2013 dataset, which has 48x48 pixel values for each image, as their starting point. They created an emotion recognition system by employing Python programs to recreate dataset photos [2]. Similarly, another study by Cohn-Kanade utilized the Cohn-Kanade expression dataset, albeit with its initial restrictions of 486 sequences and 97 posers. The researchers added more web datasets and their photographs of varied facial expressions to their dataset to get around these constraints. The accuracy increased as the dataset size increased [3]. Sukhavasi et al. chose well-known benchmark datasets for emotion recognition of drivers, such as the expanded Cohn Kanade database (CK+), FER 2013, Karolinska directed emotional face (KDEF), and the KMU-FED database. Notably, the KMU-FED database recorded drivers' facial expressions while they were driving in real time [4]. In the context of emotion recognition training, the FER 2013 dataset played a pivotal role. Images of seven different facial expressions were included in it: Angry, Disgust, Scared, Happy, Sad, Surprised, and Neutral. Two methods were used to evaluate the suggested system's accuracy: first, training and testing on a GPU, and second, evaluating the trained model on an embedded system (Raspberry Pi 4) [5]. Additionally, D Y Liliana's research in 2019 involved the CK dataset, a facial expression database with expert-validated ground truth annotations. A critical step in getting the dataset ready for training was scaling the photos to 100x100 pixels as part of the preprocessing phase [35].Mainly these datasets are based on different kind of images which used to train models.

4.10 Process with Video

Basically, some authors propose or create systems based on video data. Nishchal et al. suggested using a CCTV continuous video capture to observe the students take the test and look for signs of cheating. When a signal is received, the key point detection algorithm uses the Radboud Faces Database (RaFD) to evaluate CCTV footage frame by frame to extract the output or the nature of the student's activity. Facial Expression Recognition Challenge (FERC-2013) and Extended Cohn-Kanade (CK+) [15]. Wei et al. took a different approach and recruited 15 students of varying ages to wear movement sensors while engaging in routine activities. They gathered the motion data in this manner. The entire process was filmed in trials for data synchronization. The actor's or actress's waist was fixed with the movement sensor, which could collect 3D accelerations and 3D gyroscope data at 50 Hz [16]. Chowdhury et al. proposed a technique in which they evaluated the system by recording video with a laptop's webcam. Various amounts of photos of 17 separate individuals are included in their dataset. Each user trained a specific number of images, and the precision was evaluated by using the system to separate the trained images from five different video source frames. In the training phase, they experimented four times while increasing the number of photos per person [17]. Next, only the video-based emotion recognition sub-challenge was completed by Fan et al. Each film in this collection is categorized into one of seven emotions: disgust, fear, happiness, sadness, surprise, and neutral. Giving each test set video an emotional label is the main challenge. The Emotion 2013–15 was able to tackle the problem, and the test sets freshly added reality TV videos underwent major alteration. The 1750 video clips in the AFEW 6.0 database have been divided into three categories: training (774), validation (383), and testing (593) [49].

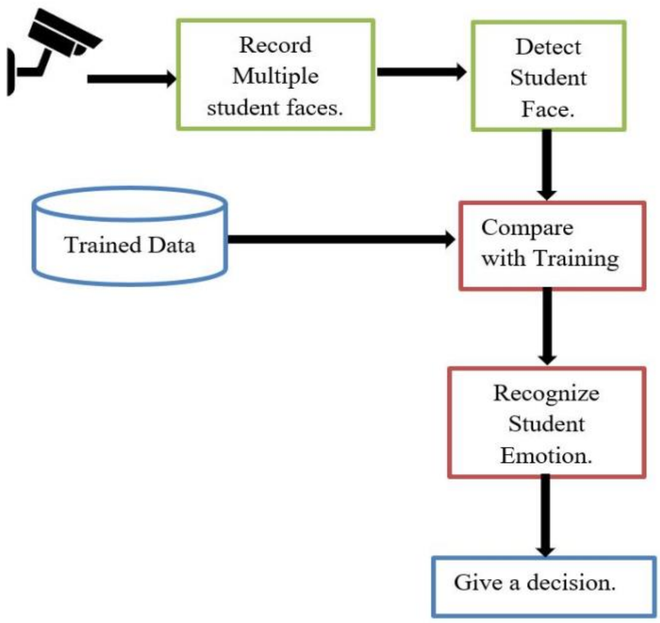

Fig. 5. A proposed Emotion Detections system with Real time video data.

Table 4. Summary of different datasets and working field images/videos.

Papers	Dataset	Image	Video
[1]	FER-2013	✓	✓
[2]	FER-2013	✓	✓
[3]	NIST, CK+, CMU, CALTECH FACE	✓	✓

[4]	Ck+, KEDF, Fer-2013, KMU-FED	✓		
[5]	FER-2013			
[6]	JAFFE	✓		
[7]	Own (collected using Mobile Camera)	✓		
[8]	CK+, KEDF, JAFFE			✓
[9]	WIDE FACE, CK+, FER-2013	✓		✓
[10]	RAF-DB	✓		
[11]	IAPS	✓		✓
[12]	AMIGOS			
[13]	ANAD, BAVED, SAVEE, EMO-DB			✓
[14]	FER-2013	✓		
[15]	FERC-2013, CK+, RAFD			✓
[17]	Own			✓
[18]	FER-2013, JAFFE	✓		
[19]	DISFA	✓		
[20]	DEAP			✓
[21]	FER-2013	✓		
[22]	Own (Created by capturing 13 participates picture)	✓		
[24]	SAVEE, RAVDESS, TESS, CREMA-D			
[25]	Google Images	✓		
[26]	DEAP	✓		
[27]	FER-2013	✓		
[29]	CK+, JAFFE, RAF-DB	✓		
[30]	CK, CK+, IMM	✓		✓
[31]	FER-2013, CK+, JAFFE, FEI, IMFDB, TFEID	✓		
[32]	CMU Multi-PIE Face Data			✓
[33]	Affect-NET	✓		✓
[34]	IEMOCAP, MELD, CK+, RECOLA			
[35]	CK+	✓		
[37]	CK+. JAFFE, Google Images	✓		✓
[38]	FER- 2013, JAFFFE, KDEF	✓		
[39]	CK+, JAFFE, BU-3DFE			✓
[40]	Own (created)			✓
[41]	Own (Created using 600 thousand of samples)	✓		
[42]	NVID			✓
[43]	WIDERD Face, FDDB, IJB-A	✓		
[44]	JAFFE	✓		
[45]	MIT Media Lab	✓		
[46]	DEAP	✓		
[48]	CMU	✓		
[49]	AFEW 6.0		✓	
[50]	FER	✓	✓	

5. Review Based ON Result

The paper introduces a system designed to predict facial emotions in real time using Convolutional Neural Networks (CNN) through TensorFlow and Keras. The system attained a predictive precision of 65.97% [1]. It is presented how two models were trained using the FER2013 dataset. After 7 epochs of testing, the first model, VGG-16, had a 54% accuracy rate, whereas the second model had a 69% accuracy rate after about 40 epochs [14].An outstanding one-dimensional convolutional neural network with an accuracy of 96.60% shone out in study on the identification of negative emotions in Thai language studies. Utilizing a ten-fold cross-validation technique, the study assessed deep learning methods with a variety of open emotive speech datasets [24]. A remarkable 97% accuracy was shown by the proposed technique for detecting human emotions using machine learning algorithms, including SVM, KNN, NB, LR, RF, and MLP [29]. A computer vision system is unveiled for detecting subtle facial expressions utilizing different methods. Averaging 88% utilising feature point tracking and 81% using high gradient component analysis, the lower face's expressions could be recognised [38]. When evaluated using the CK+ database, a recommended method for identifying facial expressions has a competitive precision of 96.76%. Furthermore, it was shown that regulation techniques improved the method's accuracy [39]. A range of machine learning algorithms underwent evaluation for detecting emotions using a novel database of facial expressions. Bayesian Networks, SVMs, and Decision Trees were applied alongside cross-validation and voting techniques to enhance classification outcomes. The findings are presented with statistical significance [40]. Experiments illustrated the potential for emotion detection accuracy reaching up to 90%, with the peak precision (90.27%) noted for "humorous" emotions. However, recognition issues were noted, particularly for neutral emotions, accompanied by a 2-second delay and an average deviation of roughly 10% [41]. An advanced deep learning algorithm aimed at recognizing emotions from EEG signals is suggested, accomplishing an average precision of 82.01% through 10-fold cross-validation within a four-class emotion classification task [46].

A couple of paper achieve more than 90% of accuracy by using different kinds of algorithms where two common algorithms are SVM, ADAM.

Table 5. Summary of Algorithm and Accuracy.

Papers	Algorithm	Accuracy
[2]	AlexNet, GoogleNet	
[3]	FERC	96%
[4]	MLM, DL	96.15%,84.58%,99.18%, 99.09%
[5]	Deep Learning	94% ,89%
[6]	SVM, LDA, PPA	83.33%
[7]	Adam	
[8]	SVM	97.53%
[9]	SGD	92.58
[13]	LSTM	96.72%,88.39%,97.13%,96 .72%
[14]	Adam	65%
[15]	LBPH	63%
[16]	PPCA, SCM, Relief-F	75.76%
[18]	Adam	70.14%,98.65%
[19]	SVM, Random Forest, NNA	
[20]	SVM	
[21]	Viola-Jones	69%
[22]	SVM	
[24]	CONV1D	96.60
[25]	SVM	
[26]	Fine Gaussian Support Vector Machine (FGSVM)	
[27]	Virtual Market Algorithm	73%
[28]	Viola-Jones	
[29]	Naïve Bayes, Logistic Regression, Random Forest	97%
[30]	SVM	
[31]	Cnn	74%
[32]	AdaBoost	
[33]	DeepShort	76%
[35]	RMB	92.81%
[37]	Viola-Jones	80.90%
[38]	HMMs	88%
[39]	MTJSR	96.76%
[42]	MIL	64.07%,70.59%
[43]	SGD	
[44]	MLP	76.70%
[46]	Deep Learning	82.01%
[47]	FACS, HMMs	93%
[49]	Hybrid	59.02%
[50]	Adam	86.05%

6. LIMITATIONS

Though some papers achieved a high accuracy, they have some limitations also. Couple of papers are selected for this analysis, as they are the most relevant in the entire study. However, this paper has some limitations as well. In this paper the accuracy of the system in predicting facial expressions is 65.97%, which means that there is still room for improvement. The testing distance of 500 cm allows the system to detect facial objects 22 times and detect and recognize facial expressions, and it only succeeded ten times out of 35 experiments, which indicates that the system may not work well at longer distances [1]. The current work focused on designing an algorithm that uses a trained model to extract facial expressions and generate reviews. However, there are limitations that can be addressed to improve the system. The dataset used in this work could be improved by using larger image sizes like 224 by 224 and training the data on pre-trained architectures such as inception, ResNet, VGG, etc. This could potentially enhance the accuracy of the model. However, due to limitations in data collection, these options could not be tested in the current work [14]. The accuracy of facial recognition systems is generally lower than fingerprint and iris recognition systems. However, facial recognition is preferred due to its non-contact, time-saving, and non-interfering process. The camera used for face recognition should be positioned in a way that it has a clear view of all the students to ensure accurate detection and recognition. The system may label a student as 'Unknown' if there is not a match of high enough accuracy during face recognition [17]. The constraints of the deep learning technique are not explicitly discussed in the research on negative emotion identification in Thai. The quantity and variety of the Thai language dataset utilized in the research are not disclosed, nor are any difficulties encountered throughout the testing process mentioned. Furthermore, there is no discussion of the possible biases associated with categorizing negative emotions in Thai speech patterns alone. The study does not address generalizability to different languages or cultural contexts, and it does not look at ethical issues in real-world applications, like telecommunications. These omissions cast doubt on how thorough and reliable the research findings are [24]. Limitations of the study analysed frontal views of subjects under constant illumination, which may not represent real-world conditions. The study included subjects of different ages and ethnicities but did not report on potential variations in recognition accuracy among these groups [38]. Limitations of the presented method requires the locations of each eye for image pre-processing steps, which can be obtained using eye detection methods. The accuracy of some expressions, such as sadness, was lower (around 84%) compared to the overall accuracy of the method (around 96%), indicating that there is not enough variation between these classes to separate them effectively. The presented method is limited to controlled environments with frontal face images of subjects. To address these limitations, one approach is to create a specialized classifier for expressions with lower accuracy, which can be used as a second classifier. Another limitation is the need for a large set of training data to handle these constraints and allow for a deeper network [39]. The limitations of this paper are the sample size of the study is relatively small, with only 30 participants, which may not be representative of the general population. The system determines the emotion with a 2-second delay with approximately 10% deviation on average, which may not be suitable for real-time emotion recognition applications [41]. One limitation could be that the newly created authentic facial expression database may not be representative of all possible emotional states and may not generalize well to other populations. Additionally, the paper only focuses on facial expressions as a channel for machine perception of emotions, while other channels such as speech and physiological signals could also be informative [40].

7. DISCUSSIONS

In recent years, advances in technology have paved the way for innovative solutions in education. One of these innovations is the introduction of automated emotion monitoring systems in classrooms, aimed at giving teachers insight into student behaviour. emotional states during educational activities. One of the main challenges is that they are not always accurate. The accuracy can vary depending on the type of emotion being measured, the environmental conditions, and the individual's facial expressions. Here are some additional points that could be included in the discussion

- The implementation of automatic emotion monitoring systems can also affect teacher-student dynamics. Introducing technology that continually monitors and assesses students. Emotions can unintentionally disrupt the natural teacher-student relationship. Students may feel uncomfortable or being monitored, resulting in altered behavioural patterns that don't truly reflect their emotional experiences. This change underscores the need to create a safe and supportive learning environment that encourages open communication while embracing technology.
- One of the biggest challenges in integrating automated emotion-tracking systems into classrooms is the privacy and ethics issues surrounding student data. Collecting and analysing emotional data without explicit consent raises questions about students' rights to their emotions and the possibility of data misuse. The need for strict data protection measures becomes crucial to prevent unauthorized access, disclosure, or manipulation of sensitive emotional information.
- Given the challenges of automated emotion monitoring systems, finding a balance between leveraging technological advances and preserving the human aspects of education seems to be a key issue. Integrate these systems as additional tools to help teachers understand and support student skills. Emotions can offer a more comprehensive approach, rather than relying solely on automated judgments.
- The introduction of automated emotion monitoring systems in classrooms is both promising and concerning. While the potential benefits of in-service teacher education are clear to students, emotions are evident, and challenges related to confidentiality, accuracy, psychological implications, and teacher-student dynamics require careful evaluation and ethical considerations.

8. CONCLUSION

The paper addresses emotion detection through machine learning methods, especially facial emotion recognition algorithms, and emphasizes the effectiveness of Convolutional Neural Networks (CNNs) in this domain. In an attempt to foster a supportive learning environment and offer focused interventions for students who require them, the authors stress the importance of conducting thorough evaluations and taking ethical issues into account when putting emotion detection systems into use, especially in educational settings.

However, when compared against other facial expression recognition methods, a different approach generates competitive findings, with an accuracy rate of 96.76% in the CK+ database with the CNN model. The detection of driver mood in real-time circumstances while driving, however, is done using two new deep network techniques that take into account things like position versions, illumination, and occlusions. Deep neural network techniques provided high accuracy rates for driver emotion identification, ranging from 83.68% to 99%.18% using CK+, FER 2013, KDEF and KMU-FED datasets.

In the entire study, one can identify four significant datasets and two primary models, namely CK+, FER 2013, KDEF, KMU-FED dataset, and CNN and DNN. Employing these datasets and models can facilitate the detection of student emotions with high accuracy in real time.

There are plenty of advantages to using student emotion detection in the classroom for both teachers and students. Teachers can identify students who are having difficulties comprehending what they're learning or who have become disengaged by using emotion detection. To help these students succeed, this can assist teachers in providing targeted interventions. Identifying and resolving any problems that are making students feel stressed or anxious can assist teachers in creating a positive learning environment. Giving them the impression that their teachers are more conscious of their needs and difficulties can also help students feel more connected to their teachers. This might result in a more conducive and productive learning environment.

It's essential to keep in mind that detecting emotions is not an accurate method. Accurately identifying emotions from facial expressions can be challenging, and the outcomes of emotion detection are affected by a variety of elements, including the classroom's lighting and the student's socioeconomic situation.

ACKNOWLEDGEMENT

The American International University-Bangladesh's Bachelor of Computer Science and Engineering program requires students to complete a thesis, and this article served that purpose. The writers thank all the advisers who helped with this thesis.

REFERENCES

[1] Zahara L, Musa P, Prasetyo Wibowo E, Karim I, Bahri Musa S. The Facial Emotion Recognition (FER-2013) Dataset for Prediction System of Micro-Expressions Face Using the Convolutional Neural Network (CNN) Algorithm based Raspberry Pi. 2020 5th International Conference on Informatics and Computing, ICIC 2020, Institute of Electrical and Electronics Engineers Inc. 2020.

[2] Giannopoulos P, Perikos I, Hatzilygeroudis I. Deep learning approaches for facial emotion recognition: A case study on FER-2013. Smart Innovation, Systems and Technologies, Springer Science and Business Media Deutschland GmbH 2018, 1–16.

[3] Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). SN Appl Sci 2020; 2.

[4] Sukhavasi SB, Elleithy K, El-Sayed A, Elleithy A. Deep Neural Network Approach for Pose, Illumination, and Occlusion Invariant Driver Emotion Detection. Int J Environ Res Public Health 2022; 19.

[5] Helaly R, Hajjaji MA, M'Sahli F, Mtibaa A. Deep Convolution Neural Network Implementation for Emotion Recognition System. Proceedings - STA 2020: 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering, Institute of Electrical and Electronics Engineers Inc. 2020, 261–265.

[6] Putra WB, Arifin F. Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom. Journal of Physics: Conference Series, Institute of Physics Publishing 2019.

[7] Pranav E., Kamal S, Chandran C. S, Supriya M.H. Facial Emotion Recognition Using Deep Convolutional Neural Network. International Conference on Advanced Computing & Communication Systems (ICACCS) 2020;

[8] Hammoumi O El, Benmarrakchi F, Ouherrou N, Kafi J El, Hore A El. Emotion Recognition in E-learning Systems.

[9] Gupta S, Kumar P, Tekchandani R. A multimodal facial cues-based engagement detection system in e-learning context using deep learning approach. Multimed Tools Appl 2023; 82: 28589–28615.

[10] Deshmukh RS, Jagtap V, Paygude S. Facial Emotion Recognition System through Machine Learning approach. International Conference on Intelligent Computing and Control Systems ICICCS 2017;

[11] Lim JZ, Mountstephens J, Teo J. Emotion recognition using eye-tracking: Taxonomy, review and current challenges. Sensors (Switzerland) 20 2020.

[12] Santamaria-Granados L, Munoz-Organero M, Ramirez-Gonzalez G, Abdulhay E, Arunkumar N. Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS). IEEE Access 2019; 7: 57–67.

[13] Alsabhan W. Human–Computer Interaction with a Real-Time Speech Emotion Recognition with Ensembling Techniques 1D Convolution Neural Network and Attention. Sensors 2023; 23.

[14] Siam SC, Faisal A, Mahrab N, Haque AB, Suvon MNI. Automated student review system with computer vision and convolutional neural network. Proceedings - IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2021, Institute of Electrical and Electronics Engineers Inc. 2021, 493–497.

[15] Nishchal J, Reddy S, N NP. Automated Cheating Detection in Exams using Posture and Emotion Analysis.

[16] Wei C, Zhang H, Ye L, Meng F. A school bullying detecting algorithm based on motion recognition and speech emotion recognition. Proceedings - 2020 International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2020, Institute of Electrical and Electronics Engineers Inc. 2020, 276–279.

[17] Chowdhury S, Nath S, Dey A, Das A. Development of an Automatic Class Attendance System using CNN-based Face Recognition; Development of an Automatic Class Attendance System using CNN-based Face Recognition. 2020;

[18] Jaiswal A, Raju AK, Deb S. Facial Emotion Detection Using Deep Learning. International Conference for Emerging Technology (INCET) 2020;

[19] Mehta D, Siddiqui MFH, Javaid AY. Recognition of emotion intensities using machine learning algorithms: A comparative study. Sensors (Switzerland) 2019; 19.

[20] Domínguez-Jiménez JA, Campo-Landines KC, Martínez-Santos JC, Delahoz EJ, Contreras-Ortiz SH. A machine learning model for emotion recognition from physiological signals. Biomed Signal Process Control 2020; 55.

[21] Ivanova E, Borzunov G. Optimization of machine learning algorithm of emotion recognition in terms of human facial expressions. Procedia Computer Science, Elsevier B.V. 2020, 244–248.

[22] Kundu T, Saravanan C. Advancements and recent trends in Emotion Recognition using facial image analysis and machine learning models.

[23] Kahou SE, Michalski V, Konda K, Memisevic R, Pal C. Recurrent neural networks for emotion recognition in video. ICMI 2015 - Proceedings of the 2015 ACM International Conference on Multimodal Interaction, Association for Computing Machinery, Inc 2015, 467–474.

[24] Mekruksavanich S, Jitpattanakul A, Hnoohom N. Negative Emotion Recognition using Deep Learning for Thai Language.

[25] Rani P. Emotion Detection of Autistic Children Using Image Processing. 2019;

[26] Hassan MM, Alam MGR, Uddin MZ, Huda S, Almogren A, Fortino G. Human emotion recognition using deep belief network architecture. Information Fusion 2019; 51: 10–18.

[27] Gowri SM, Rafeeq A, Devipriya S. Detection of real-time facial emotions via deep convolution neural network. Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021, Institute of Electrical and Electronics Engineers Inc. 2021, 1033–1037.

[28] Karilingappa K, Jayadevappa D, Ganganna S. Human emotion detection and classification using modified Viola-Jones and convolution neural network. IAES International Journal of Artificial Intelligence 2023; 12: 79–86.

[29] Siam AI, Soliman NF, Algarni AD, Abd El-Samie FE, Sedik A. Deploying Machine Learning Techniques for Human Emotion Detection. Comput Intell Neurosci 2022; 2022.

[30] Rajesh K M, Naveenkumar M. A Robust Method for Face Recognition and Face Emotion Detection System using Support Vector Machines.

[31] Jaiswal S, Nandi GC. Robust real-time emotion detection system using CNN architecture. Neural Comput Appl 2020; 32: 11253–11262.

[32] Anwar S, Anjaneyulu M. Machine learning based Real Time-Employee Emotion Detection.

[33] Trabelsi Z, Alnajjar F, Parambil MMA, Gochoo M, Ali L. Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. Big Data and Cognitive Computing 2023; 7.

[34] Cai Y, Li X, Li J. Emotion Recognition Using Different Sensors, Emotion Models, Methods and Datasets: A Comprehensive Review. Sensors 23 2023.

[35] Liliana DY. Emotion recognition from facial expression using deep convolutional neural network. Journal of Physics: Conference Series, Institute of Physics Publishing 2019.

[36] Ding W, Xu M, Huang D et al. Audio and face video emotion recognition in the wild using deep neural networks and small datasets. ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction, Association for Computing Machinery, Inc 2016, 506–513.

[37] Sahla KS, Senthil Kumar T. Classroom teaching assessment based on student emotions. Advances in Intelligent Systems and Computing, Springer Verlag 2016, 475–486.

[38] Hung JC, Lin K-C, Lai N-X. Recognizing learning emotion based on convolutional neural networks and transfer learning. The Official Journal of the World Federation on Soft Computing (WFSC) 2019;

[39] Lopes AT, de Aguiar E, De Souza AF, Oliveira-Santos T. Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order. Pattern Recognit 2017; 61: 610–628.

[40] Sun Y, Sebe N, Lew MS, Gevers T. Authentic Emotion Detection in Real-Time Video.

[41] Raudonis V, Dervinis G, Vilkauskas A, Paulauskaite -Tarasevičiene A, Kersulyte -Raudone G. Evaluation of Human Emotion from Eye Motions. 2013.

[42] Nayak S, Nagesh B, Routray A, Sarma M. A Human–Computer Interaction framework for emotion recognition through time-series thermal video sequences. Computers and Electrical Engineering 2021; 93.

[43] Jiang, Huaizu, and Erik Learned-Miller. "Face detection with the faster R-CNN." 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017). IEEE, 2017.

[44] Perikos I, Ziakopoulos E, Hatzilygeroudis I. Recognizing Emotions from Facial Expressions Using Neural Network. IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI) 2014; 10th: 236–245.

[45] Intrator N, Reisfeld D, Yeshurun Y. Face recognition using a hybrid supervised/unsupervised neural network. Pattern Recognit Lett 1996; 17: 67–76.

[46] Sharma R, Pachori RB, Sircar P. Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomed Signal Process Control 2020; 58.

[47] Jenn-Jier Lien J, Cohn JF, Li C-C. Subtly Different Facial Expression Recognition and Expression Intensity Estimation. 1998;

[48] Cowie, Roddy, et al. "Emotion recognition in human-computer interaction." IEEE Signal processing magazine 18.1 (2001): 32-80.

[49] Fan Y, Lu X, Li D, Liu Y. Video-Based emotion recognition using CNN-RNN and C3D hybrid networks. ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction, Association for Computing Machinery, Inc 2016, 445–450.

[50] Depuru S, Nandam A, Ramesh PA, Saktivel M, Amala K, Sivanantham. Human Emotion Recognition System Using Deep Learning Technique. J Pharm Negat Results 2022; 13: 1031–1035.

[51] M. Sambare, "Fer-2013. Kaggle," [Online]. Available: https://www.kaggle.com/datasets/ msambare/fer2013 (accessed Aug. 20, 2023).

[52] D. Sena, "CK+ dataset. Kaggle," [Online]. Available: https://www.kaggle.com/datasets/davilsena/ckdataset (accessed Aug. 18, 2023).

[53] Saad, M., "Wider face. Kaggle," [Online]. Available: https://www.kaggle.com/datasets/mksaad/wider-face-a-face-detection-benchmark (accessed Aug. 20, 2023).