Engagement Detection in E-Learning

Anish Kumar
Information Technology *Techno International New Town* Kolkata, West Bengal, India
anishkumaro8082001@gmail.com

Vikram Kumar
Information Technology *Techno International New Town* Kolkata, West Bengal, India
thevsingh1@gmail.com

Abhishek Kumar
Information Technology *Techno International New Town* Kolkata, West Bengal, India
kumarabhishek962628@gmail.com

Dr Khondekar Lutful Hassan
Assistant Professor
Aliah University West Bengal,India
klh.cse@gmail.com

Deep Kumar Dutta
Information Technology *Techno International New Town* Kolkata, West Bengal, India
deepduttaofficial@gmail.com

Arindam Ghosh
Information Technology *Techno International New Town* Kolkata, West Bengal, India
arindamghosh.org@gmail.com

***Abstract*— E-learning is transforming the teaching and learning process from traditional classrooms to online platforms, focusing on personalized learning. Emotion recognition in e-learning platforms allows students to express their feelings in text, image, and audio modes. Online students engage in various activities, exhibiting varying engagement levels. To assess engagement, educators use machine learning and computer vision methods. Data from eye trackers and camera sensors is used to measure learner engagement and categorize it. Perceptual user features are extracted from the collected data, and feature selection and classification techniques are used to create classifiers. The difficulties in detecting engagement are investigated, and suggestions for future technology improvements are offered. This approach aims to provide individualized pedagogical support through interventions.**

Keywords- Online Learning, Engagement, Students, Audio & Video, etc.

I. Introduction

The usage of computer-based technology is growing in numerous directions due to its easy availability, effectiveness, etc. Technological advancements like smartphones, laptops, and other intelligent devices help us use online learning facilities termed E-learning. E-learning platforms have become a significant tool for knowledge sharing and understanding for almost every student, especially after the pandemic. E-learning has numerous advantages like eco-friendliness as it saves paper, reduces the cost of traveling and time, etc. The students can attend the classes from their places even if they are not feeling well. During a pandemic situation like COVID-19, all parents and students were worried about the futures of their children in lockdown situations, but E-learning has solved this problem, and every student is capable of learning through their mobile phones and laptops.

On the other hand, the physical education system has benefits like interaction between teacher and students, assessment of student's understandability, and hands-on sessions. In traditional classroom teaching, teachers evaluate their students' learning effect, the level of understanding and comprehension, by mainly observing students' behavior. The behavior aspects may include body language, eye gaze, facial expressions, and emotions exhibited through vocal feedback. Analyzing all these attributes a tutor in a physical classroom could give personalized feedback to a particular student. But in online learning analyzing these attributes is a challenge in the absence of a physical tutor. Detection of a student's emotional state is crucial to personalize their learning in an automated learning platform. Multiple researchers have proposed the use of natural language processing, hand gesture recognition, eye gaze estimation, facial emotion recognition, and body language detection to estimate learners' learning effects and provide a measure that will provide a more effective learning experience. It is becoming essential for e-learning platforms to be able to educate their learners according to their personalized features.

Due to the enormous development in the Machine Learning sphere, the LMSs can now classify the learners according to their different personalized features like learning style, attentiveness, cognitive ability, specialized requirements, etc. The learners of various classes are recommended different learning objects, learning contents, tutors, etc. after discovering the most appropriate one for them by applying different Machine Learning algorithms. Recognizing a learner's emotional state (fear, anger, depression, joy, confusion, confidence, etc.) is also very important for personalizing their learning experience. Recommending appropriate learning content according to their emotional state could enhance their learning quality as a result improving the overall quality of the learning platform. Researchers have developed several frameworks and models predominantly based on Artificial Intelligence to recognize the emotional state of a person from textual, audio, image, and video content. In this work, we have reviewed different research works for recognizing the emotion of an e-learner through analyzing their video or image.

The purpose of this survey is to encourage further study into ways to improve the learning drive of an e-learner. Since a person's motivation is closely related to their emotion we have surveyed recognizing emotions of a person from their video mostly in the context of e-learning. This will direct new avenues to enhance personalized learning in Intelligent Tutoring Systems to attract and sustain a greater number of students in online learning. Using the above exposition, we justify the novelty of this work. This paper is organized in the following sections. In section 2 we have conducted a brief literature review of the topic. Section 3 represents a comparative study of the research works discussed in section 2. In Section 4 we have discussed the future direction of this research topic in brief.

II. BRIEF SURVEY

We observed as we explored through the previous works that frameworks developed for the emotional recognition of a person are majorly based on two categories. Either deep learning-based or Conventional Machine learning-based. So as per our findings we have divided our review in two categories. In category 2.1 we have reviewed those works which are developed around deep learning-based algorithms and in 2.2 we have reviewed those works in which deep learning algorithms have hardly intervened.

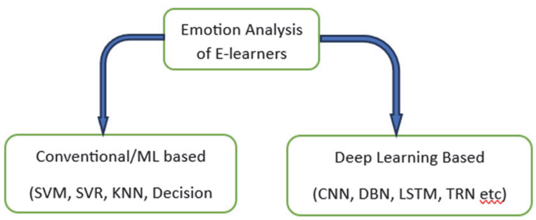

Fig.1 Category for the Emotion Analysis

A. Deep Learning Based

All the deep learning-based works are reviewed in this section.

Nonverbal behavior of forty-four undergraduate students was observed using a USB front-facing camera while participants completed an on-screen multiple-choice question-and-answer test [1]. An ANN is fed the information on their question-answer scores and behavioral patterns to classify their comprehension states almost instantly. Future work by the authors is planned to increase the classifier's accuracy. Furthermore, this investigation might be extended to examine the relationship between behavior and the type of question asked. In the future, this technique might also be used to analyze various kinds of behaviors and mental states. In [2] a concept of similarity is introduced generally to preserve the actual data with features and group them

ISBN : 978-81-978522-2-0

creating a degree of similarity among the pairs which is achieved through fuzzification. Authors have proposed to develop theoretical methods to generate rules and select features based on the similarity relation in the future.Zakkaetal. conducted a study between traditional classrooms and e-learning to improve the e-learning environment to match the traditional one [3]. A framework to detect the motivation level of learners is incorporated which senses emotion and sends feedback to the teacher and further develops a response mechanism to distinguish expressions and attention.The authors created a smart computer system (EAC Net) that can better recognize facial expressions (AUs) [4]. Unlike other methods, it doesn't need perfectly aligned faces and understands facial features more effectively. It showed big improvements in accuracy on a dataset. This system is useful because it wo rks well even if faces are at different angles or partly covered; making it more versatile for real-world situation.

In [5] authors emphasized more on multi-modal emotion recognition than the single modal ones. They used Affectnet as their FER model. The use of predicted emotions extends beyond understanding student behavior to include visual summarization of classroom films and classification of the group-level emotions on videos.

Based on their behavior and biological data, several attempts have been made to ascertain the e-learners; level of focus through a neuro-fuzzy inference system that tracks the position of the eye's iris to find out the concentration level. In it, SVM is also used to determine concentration level [6]. This proposed model has an overhead of timing for the preprocessing of data which in the future could be eliminated by fully automating the process so that it could be implemented in real time also.

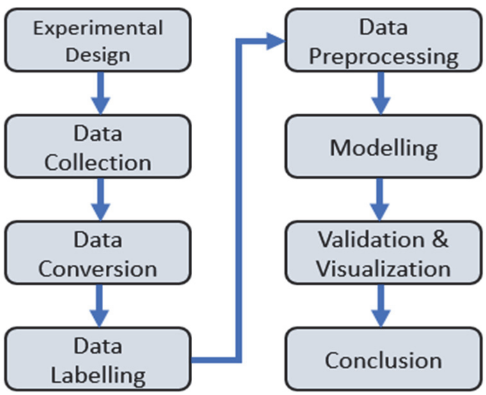

Fig 2. Overview of Procedures

Pise et al. employed a two-phase method to categorize 3D face expression images extracted from the video to quantify the optical flow intensity with the help of the Hidden Markov model and Naïve Bayes [7]. Moreover, video can be used to measure even minute changes in contrast to still images.Using temporal appearance and facial landmark points facial gestures are extracted and then integrated to best recognize the expression [8]. CNN is used for object detection, and feature extraction, in this model. In the future, the framework could be extended to work in GPU-supported machines for a better training time. Researchers utilize multimodal learning analytics in online education to comprehend the feelings and engagement of students [9]. They use information from posture, gestures, and emotions to estimate students' level of engagement in the classroom. Computer vision techniques examine lecture footage and detect feelings such as joy or indifference. Some systems even use head and eye motions to categorize different levels of engagement. By taking into account the feelings and participation of the students, these methods seek to improve online instruction. Thiruthuvanathan et al. discussed the challenges in detecting e-learners' engagement level from their facial expression recognition. They have proposed to extend their model for detecting group-level

ISBN : 978-81-978522-2-0

engagement in the future [10]. In [11] the author is trying to remove the challenges and weaknesses of the mostly used blended learning model. Additionally, the gamification concept has been applied. The fer and gamification system were developed using an object-oriented approach using unified modeling language (UML).The methods used are ANN, CNN, and JavaScript library with open-source code TensorFlowJS. It has two stages of testing the facial expression recognition system and gamification application.

In [12] the authors have used a video dataset called Children's Spontaneous Facial Expressions (LIRIS-CSE) and proposed a system that uses Convolutional-Neural-Network (CNN)-based models, such as VGG19, VGG16, and Resnet50, for feature extractions and Support Vector Machine (SVM) and Decision Tree (DT) for classification. This system will automatically recognize children's expressions. Several experimental configurations, such as 80–20% split, K-Fold Cross-Validation (K-Fold CV), and leave-one-out cross-validation (LOOCV), are used to assess the system for both image and video-based categorization.

Keerthana et al. proposed a hybrid model for identifying student facial emotions by monitoring eye gaze and head movements to analyze student engagement levels [13]. Haar cascade model and binary patterns are used to detect the head movements. CNN models are used for FER. Using the OpenCV object detection framework authors determined the status of the student, whether he is Distracted or engaged.

[14] In this paper, the proposed framework calculates the concentration index of students. CNN model using Keras is used with fer2013 datasets for emotion recognition. Additionally, Mamdani MATLAB software is used to create fuzzy rule sets and implement membership functions utilizing the principles of Fuzzy Logic. It mainly deals with three major steps: face detection, feature extraction, and feature classification.

Dewan et al. explore various methods for learners' engagement detection and classify them into three main categories —automatic, semi-automatic, and manual [15]. Each category is further divided into audio video and text depending on the type of data used for detection. The authors reviewed the automatic methods for engagement detection in more detail as they found it more effective than other methods in the case of online learning platforms. They further discussed the challenges of all of these methods and future directions about how they could be used in a more advanced way.

[16] In this paper, a framework using computer vision to detect student engagement is proposed. Facial expressions, body language, and other cues are used to determine whether a student is paying attention and engaged in the material. The proposed model uses an SVM (Support Vector Machines), Random Forest, Neural Networks, CNN (Convolutional Neural Networks), LSTM (Long Short-Term Memory), InceptionV3, and VGG16 for object recognition in video scenes to analyze students' facial expressions, body postures, and hand gestures. The data set considered for analyzing the model is too small consisting of only 45 students. The number of features considered to detect engagement is also very small which could be enhanced in the future for a more accurate detection with an increased number of labels like neutral, low engaged, highly engaged, etc. apart from only engaged or not engaged labels.

Ozdamli et al.explore various algorithms and models for facial recognition in education [17]. For face detection and recognition, it mentions software like MATLAB and Python utilizing techniques like PCA and 3WPCA-MD. Classifications are done with diverse algorithms like SVM, Bayesian, or neural networks. When it comes to recognizing emotions from facial expressions, the paper discusses static models, Action Unit (AU) based models, and the Facial Action Coding System (FACS). Deep learning frameworks like CNNs are also increasingly used. To detect cheating in online exams, models like Multi-Class Markov Chain Latent Dirichlet Allocation and supervised dynamic Bayesian models are employed. Various datasets like GI4E for gaze tracking and FEI for general face recognition tasks are used in this model.

The authors investigate methods for evaluating teaching quality and student engagement in classrooms [18]. Traditional approaches, including tests and observations, are criticized for their limitations in providing comprehensive and real-time data. Attention shifts towards assessing student engagement, defined across behavior, cognition, emotion, and social interaction dimensions. Technological advancements, particularly in computer vision, offer non-invasive means to detect engagement through facial expression analysis and gaze tracking. Studies demonstrate high accuracy in predicting engagement levels using deep learning models. Overall, the review highlights a growing interest in leveraging technology to enhance classroom evaluation and improve teaching practices.

Proposed models define structures for representing learning behaviors in classrooms, enabling effective feature extraction [19]. Evaluation metrics demonstrate the efficacy of machine learning models in accurately detecting and categorizing student behaviors. More data sources could be added to the model in the future. Advances in the computer vision field, specifically in Human Pose and Graph Neural Networks could be incorporated also to increase the efficacy.

The Authors have used Multimodal Emotion Recognition in Multiparty Conversations (MERMC) which focuses more on audio and text while ignoring visual information in [20]. It incorporates two-stage frameworks. Facial expression-aware Multimodal Multi-Task learning and Multimodal facial expression-aware emotion recognition model which helps in extraction of face and help improve emotion recognition. They have planned to leverage multimodal fusion mechanisms to improve the performance of this task in the future.

B. Conventional ML-Based

Communication requires the use of facial expressions, which differ between individuals and civilizations. Effective teaching requires a knowledge of students' emotions, especially with the development of online learning brought on by COVID-19. Understanding facial expressions allows educators to modify their approaches and add interest to their lessons. While negative emotions might cause disengagement, positive emotions support academic progress.

The authors conducted a thorough review of emotion classification on facial emotion recognition [21]. It elaborates analysis on emotion classifiers and datasets used in FER. Different approaches considered by the researchers for preprocessing and feature extractions are discussed. The authors highlighted the strengths and limitations of this approach. Their study revealed that deep learning is the most used approach for FER in the academic arena. Whereas the most used dataset and emotion classifier are DAiSEE and SVM respectively.

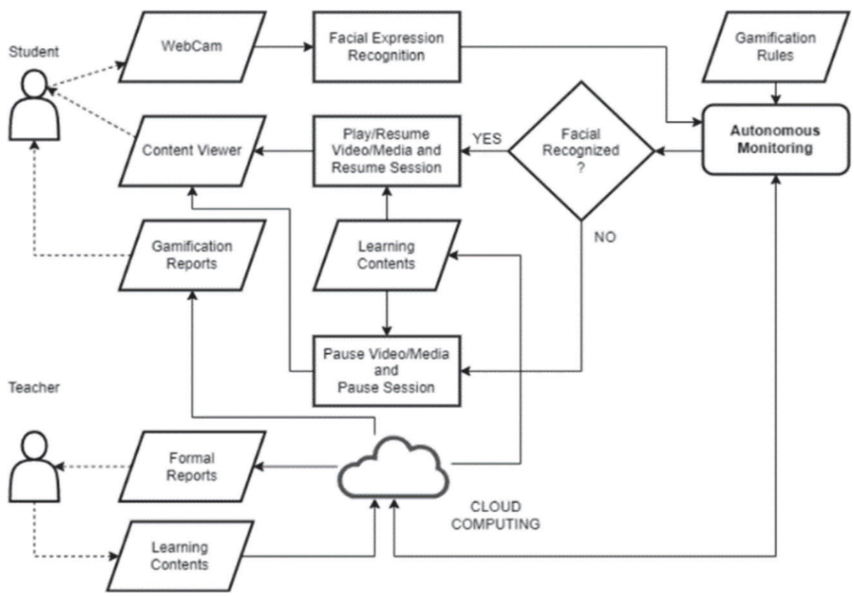

Fig. 3. Research Design Block Diagram

Zhanget. Al. developed an algorithm that can accurately detect students' engagement in online learning environments [22]. Supervised learning is used to recognize emotional gestures through which it distinguishes emotions like frustration, happiness, boredom, confusion, etc. Sensors are being used in this research. Measures of the degree of engagement are divided into 3 categories namely single-sensor, multiple-sensor, and sensor-free methods.

[23] The principal aim of this study is to explore reliable facial information models that can describe how people interact in a learning environment. Different approaches for automated recognition of student engagement levels are studied. Authors stated that engagement recognition will be more effective and variant in long-term learning situations rather than short-term studies of current scenarios.

In [24] authors wanted to highlight the growing interest in facial expression recognition and eye tracking, to assess and enhance student engagement in digital learning environments. Various methods, including deep learning models and facial action coding systems, have been explored to measure concentration levels and emotional states. These approaches aim to provide real-time feedback to instructors, allowing for personalized adjustments to content delivery. Despite challenges such as dropout rates in virtual classrooms, ongoing research underscores the importance of understanding and addressing student engagement for effective digital education.

Dukicet al. discover connections between emotions, activities, and gender in online learning [25]. Authors have analyzed two different perspectives (1) classroom experiment-related and (2) FER data-related. To gather feedback on active teaching strategies, students' emotions are tracked as they complete programming assignments. The focus of the methodology in this research was primarily on the activity portion. However, the variance due to age difference is not taken into consideration in this study. The authors intend to use additional data to test this model in the future with improved sticker conditions and camera positioning on participant behavior.

The authors have designed a framework to recognize emotions and categorize them into 3 parts. First is the face tracker. The second one is the facial motion tracking optical flow algorithm and the third is the recognition engine. This method proposed using a channel attention network with depth separable convolution to enhance the linear bottleneck structure. When tested on the FER2013 dataset, it outperforms other methods significantly. Mainly because it pays more attention to extracting features, resulting in better accuracy in recognizing emotions [26].

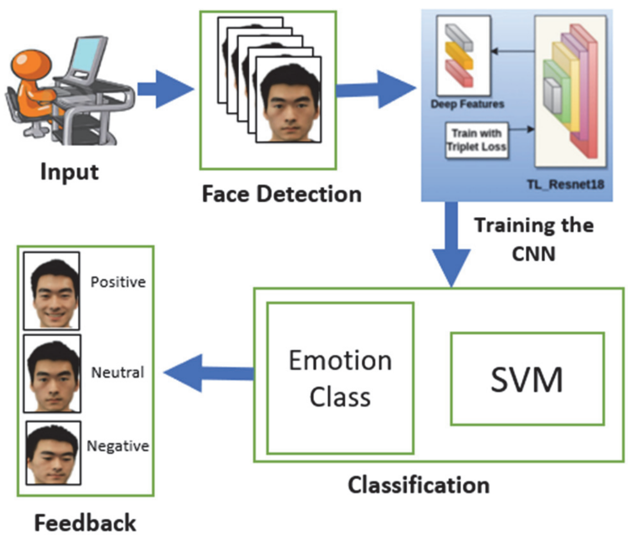

Fig.4. The Proposed TLF-ResNet Pipeline for facial expression recognition with a face detector to detect

ISBN : 978-81-978522-2-0

In [27] Authors conducted a systematic review of existing frameworks for Facial Emotion Recognition (FER)and how they are used in classifying academic emotions mainly in the context of online learning. Authors observed that low illumination, Lack of frontal pose, and small size datasets are some of the major hindrances in FER in e-learning. They proposed that continuous video recording, wearable sensor monitoring of facial emotions over an extended period of time, and the elimination of any human biases could all lead to improved accuracy when it comes to FER in online learning.

Alkabany et al. proposed a methodology that assesses students' degree of engagement in both traditional classroom settings and online learning environments [28]. The proposed framework captures the user's footage and tracks their facial movements across the frames. It can be used to monitor the progress of e-learners with varying degrees of learning disabilities and to examine how nerve palsy affects facial expressions and social interactions. The Facial Action Coding System (FACS) decomposes facial expressions into the basic actions of individual muscles or groups of muscles (i.e., action units) by extracting various features from the user's face video, such as facial fiducial points, head pose, eye gaze, and learning features. These decoded action units (AUs) are then used to measure the student's behavioural engagement, or willingness to participate in the learning process, and emotional engagement, or attitude towards learning

Emotional changes of 67 students during a lecture on Information technology are studied in [29]. The Microsoft Emotion Recognition API and the C# programming language are used in the software's development to classify students' emotions into categories such as disgust, sadness, happiness, fear, contempt, anger, and surprise. Study is done on the importance of the relationship between students' feelings and their departments, gender, lecture hours, computer location in the classroom, lecture type, and session information. Finally, the association between student’s emotional change and their achievements is analyzed to examine how the emotional recognition of students could contribute to increasing the overall quality of education.

In the research work done by Gong et al. [30], the high-definition video of classroom teaching is recorded using the camera in front of the classroom. The faces of every student in the classroom are located and intercepted using the AdaBoost algorithm from the sampled frame images, and the images are pre-processed to produce an expression area of 64 by 64 pixels. After PCA+ dimensionality reduction, Gabor and ULBPHS feature fusion is integrated with the KNN classification algorithm for expression classification. At last, the assessment and results of the emotional learning of the students are achieved.

III. FUTURE WORKS

Computer vision-based methods for engagement detection in learning contexts have some limitations, such as analyzing head motion and facial occlusions, which can lead to data loss and segmentation errors. Researchers are increasingly focusing on creating datasets for engagement detection from facial expressions, but three major challenges remain: defining the link between specific facial expressions and learning activities, determining the number of affective states or engagement levels effective for recognizing online learners, and determining the frequency of affective states reported in input videos. Additionally, it is unclear how frequently engagement detection should be made, the length of video clips suitable for assigning a single level, and the standard for determining emotions a learner is experiencing. When capturing videos for engagement detection in online learning, environmental constraints must be taken into account. It is also unclear how reliable trained judges are at labelling training data. By addressing these issues, computerized learning environments can improve their automatic engagement detection capabilities, which will benefit students by improving learning outcomes and their overall experience. To fully comprehend the relationship between engagement detection results and task performance, more research is required.

A. Comparative Study

In this section through Table.1 we have expounded upon our findings of this work. It is a concise tabular depiction of the Brief Survey in section 2. Table 1 has four columnar data. The work's goal is succinctly stated in the first column. The second column of the document indicates which algorithm or method is used. The third column displays the limitations of the work. Column four contains a synopsis of the future directions for the works. In a single glance, this table captures the importance of the entire work.

TABLE I. COMPARATIVE TABLE

Sl. No.	OBJECTIVE	ALGORITHM	LIMITATION	SCOPE
1	Non-verbal behavioral pattern classification of E-learners	Feed-forward MLP & MSE	Question type, gender, and demographic variables are not considered for classification	Deep learning techniques can enhance NVB modelling and behaviour labelling.
2	Identifying Learning Affect in E-Learning Platforms through the Expression of Facial Emotions	OpenCV, HAAR, CNN2		We will develop fuzzy rough theoretical techniques for feature selection and rule generation.
3	Estimating Student Learning Affect Using Facial Emotions	CNN, FER2013 dataset	It is better to use two rather than one or three facial emotion predictions when learning affect detection.	To obtain a more accurate result, incorporate the use of multi-model pattern analysis, such as body expression, eye gaze, and head movements.
4	EAC-Net: Deep Nets for Facial Action Unit Detection with Enhancing and Cropping	CNN, BP4D, and DISFA AU datasets	Automatic generation of attention maps	Finding more responsive areas to improve and crop nets instead of currently finding the positions by hand
5	Encoding Feelings and Involvement in Virtual Education Using a Singular Face Expression Identification Neural Network	CNN, Affectnet	Improve the quality of the engagement prediction engine	1) Identify not only the facial expressions but also the arousal and valence. 2) Face clustering
6	A methodology to predict e-learners' concentration	enduring short-term memory and recurrent neural networks	1)applied only to well-structured process models. 2) applied in a controlled environment	1) Automation of the process. 2) Should be tested in a real environment. 3) Expanding the model to CNN and TCNs
7	application of a facial image analysis model based on deep learning to gauge the impact of learning and assess student involvement.	temporal relational network, MLP, FER	Deep neural networks struggle to discover the long-term temporal dependencies between frames.	Combine other modalities, like the audio modality, with additional datasets and the TRN technique.
8	Proposes a method for accurate facial expression recognition using a lightweight deep learning model.	Stacked sparse auto-encoder (SSAE)	It took longer to train because it could only be done on CPU-based computers.	Employing a framework capable of supporting GPU will shorten the training period.

ISBN : 978-81-978522-2-0

9	A comprehensive overview of the research on FER in online learning	SVM, CNN, KNN, DNN & LSTM	1) High memory and sophisticated computing requirements. 2) It is not in frontal pose and has poor illumination.	Examine methods for tracking facial emotions over an extended period of time, such as wearable sensors or ongoing video recording.
10	An algorithm that can accurately detect students' engagement in online learning environments	Adaptive Weighted Local Gray Code Patterns (LGCP)	Numerous single-screen learning sites where students don't have to scroll to see updated content	Managing a brief educational content page
11	Provide methods for automatically detecting student engagement based on their facial expressions.	CNNs	Limitation of short-term laboratory studies.	Focusing on long-term learning situations
12	To investigate student engagement levels context of online learning through the analysis of facial behaviour.	LSTM	Limited data set and computational power.	They wish to combine the data that our system currently provides with the data that
13	to create deep learning models for active teaching in the context of real-time facial expression recognition (FER).	CNNs	This restricts the options solely to the age range.	Reconduct the classroom experiment with improved camera positioning and gather additional data to help with model learning.
14	Analyze online lecture videos, detect students' engagement levels and emotions	CNNs	Camera Orientation, Obstructions and Lighting Conditions, Engagement Detection from Profile Pictures	In the future, they hope to address the issues and carry out a user study.
15	The development of a robust and accurate system for classifying human facial emotions	ResNet18, Triplet Loss Function	Overfitting, network complexity, linear bottleneck structure, complex facial structures	Multimodal strategies that integrate image- and video-based techniques
16	An extensive review of the literature on the application of facial expression recognition (FER) systems to the categorisation of emotions in the classroom.	SVM	Low Illumination, Lack of Frontal Pose, sample number of Dataset	1) Long-term monitoring of facial emotions.2) privacy, consent, and potential biases
17	Detecting engagement levels and emotions of online learners using facial expression recognition	CNNs	Emotion Recognition Accuracy, Ethical and Privacy Concerns	Group engagement detections and evaluate the valence and arousal of the group

18	An autonomous monitoring system that supports learning with a blended learning model by employing gamification techniques and facial expression recognition.	UML, ANN, CNN	Constructing an independent surveillance system. cloud computing-based implementation. Examining gamification and facial expression recognition	Facial expression recognition accuracy. Generalization to different contexts
19	using human face and head movements as a model to depict participation in a learning environment.	PnP, FACS, BP4D, CNN	Collection and annotation of a substantial engagement dataset.	Provide a system that can identify anxiety and depression in students early on.
20	To examine how the emotions of the students change during a lecture	Facial Movements Coding System (FACS)	the quality of the camera's image, the height of the students, and their seating arrangement.	Different age groups, departments, and lectures can all analyse emotions. .
21	Judge the emotions of students in a classroom through their videos.	FACS, Adaboost, KNN	A learner's high degree of facial movement intensity may indicate a higher level of internal emotional arousal.	Learning tolerance was negatively correlated with the intensity of eyebrow lowering.
22	Identifying expressions on children's faces (6–12 years old)	CNN, (K-Fold CV)	Owing to the lack of additional child FE datasets, the suggested system has only been tested on one dataset.	For a comparative analysis, the methodology can be tested on additional datasets.
23	Real-time monitoring and analysis of learner engagement in e-learning environments.	CNN, Haar Cascade algorithm	The level of engagement recognized using eye tracking and head movement detection	Adding facial area recognition to uniquely identify the individual student and provide personal feedback
24	Deep learning techniques to recognize student emotions during online learning sessions	AdaBoost Algorithms, CNN, Haar Cascade algorithm	Because age groups differ, learners' responses change along with the concentration index.	By making a hybrid system result can be made better
25	Engagement detection in online learning	FACS, CERT	Engagement may not always be correlated with observational metrics.	establishing standards for annotation to mark the benchmarking datasets
26	Measure and understand student engagement in a real classroom environment	CNN, RCNN, SVM	A comparatively small dataset is used to analyse the effects of gender and timestamps.	More students in an offline classroom setting could be taken into consideration while improving the dataset.
27	To develop a new system that monitors and manages students during online learning sessions and exams.	CNN, FACS, OpenCV	Institutions use a variety of techniques to administer exams.	Speech recognition to aid the exam follow-up system in detecting cheating

28	recognising gaze and facial expressions in authentic collaborative learning situations to forecast students' participation in class	DAiSEE, CNN, KPCA, LDP	recorded how often students looked at particular targets.	Improve facial expressions by combining more potent AI computer vision recognition algorithms.
29	Acknowledgement of Student Conduct via Actions in the Classroom	R-CNN, YOLO, SSD	a number of variables, including how different camera angles affect how people stand, how occlusion obscures important details, miscommunications, and the unpredictability of human posture.	incorporating new developments in computer vision, particularly in the areas of Human Pose and Graph Neural Networks, and expanding the number of data sources
30	The multimodal emotion recognition (text, audio, vision)	InfoMap, KNN,	The proposed FacialMMT approach is a two-stage framework that is not fully end-to-end.	Explore better cross-modal alignment and multimodal fusion mechanisms

REFERENCES

[1] Mike Holmes, Annabel Latham, Keeley Crockett, and James D. O'Shea "Near real-time comprehension classification with artificial neural networks: decoding e-Learner non-verbal behaviour"

[2] Sukrit Bhattacharya, Vaibhav Shaw, Pawan Kumar Singh, Ram Sarkar, and Debotosh Bhattacharjee "SV-NET: A Deep Learning Approach to Video Based Human Activity Recognition"

[3] Benisemeni Esther Zakka and Hina Vadapalli "Detecting Learning Affect in E-Learning Platform Using Facial Emotion Expression"

[4] Wei Li, Farnaz Abtahi, Zhigang Zhu, and Lijun Yin "EAC-Net: Deep Nets with Enhancing and Cropping for Facial Action Unit Detection"

[5] Andrey V. Savchenko, Lyudmila V. Savchenko, Ilya Makarov "Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network"

[6] Young-Sang Jeong, Nam-Wook Cho "Evaluation of e-learners' concentration using recurrent neural networks"

[7] Anil Pise, Hima Vadapalli, Ian Sanders "Facial emotion recognition using temporal relational network: an application to E-Learning"

[8] Mubashir Ahmad, Saira, Omar Alfandi, Asad Masood Khattak, Syed Furqan Qadri, Iftikhar Ahmed Saeed, Salabat Khan, Bashir Hayat and Arshad Ahmad "Facial expression recognition using lightweight deep learning modeling"

[9] Jennifer Xin-Ying Lek and Jason Teo "Academic Emotion Classification Using FER: A Systematic Review"

[10] Zhaoli Zhang, Zhenhua Li, Hai Liu, Taihe Cao, and sannyuya Liu "Data-driven Online Learning Engagement and Mouse Behavior Recognition Technology"

[11] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan "The Faces of Engagement: Automatic Recognition of Student Engagement from Facial Expressions"

[12] Unqua Laraib, Arslan Shaukat, Rizwan Ahmed Khan, Zartasha Mustansar, Muhammad Usman Akram, and Umer Asgher "Recognition of Children's Facial Expressions Using Deep Learned Features"

[13] David Dukic, Ana sovic Krzic "Real-Time Facial Expression Recognition Usinig Deep Learnin with Application in the Active Classroom Environment"

[14] Mohammad Nehal Hasnine, Huyen T.T.Bui, Thuy Thi Thu Tran, Ho Tran Nguyen, Gokhan Akcapinar, Hiroshi Ueda "Students emotion extraction and visualization for engagement detection in online learning"

[15] Irfan Haider, Hyung-Jeong Yang, Guee-Sang Lee, Soo-Hyung Kim, Wataru Sato "Robust Human Face Emotion Classification Using Triplet-Loss-Based Deep CNN Features and SVM"

[16] Jeniffer Xin-Ying Lek and Jason Teo, Sarah E.Domoff "Academic Emotion Classification Using FER: A Systematic Review"

[17] Michael Moses Thiruthuvanathan, Balachandran Krishnan, Madhavi Rangaswamy "Engagement Detection Through Facial Emotional Recognition Using a Shallow Residual Convolutional Neural Networks"

[18] Indra Kurniawan, Yeffry Handoko Putra "Autonomous Monitoring with Facial Expression Recognition and Gamification to Support Blended Learning Model"

[19] Islam Alkabany, Asem Ali, Amal Farag, Ian Bennett, Mohamad Ghanoum, Aly Farag "Measuring Student Engagement Level Using Facial Information"

[20] Guary Tonguc, Betul Ozaydin Ozkara "Automatic Recognition of Student Emotions from Facial Expressions during a Lecture"

[21] Bing Gong and Jing Wei "Quantitative Analysis of Facial Expression Recognition in Classroom Teaching Based on FACS and KNN Classification Algorithm"

[22] Unqua Laraib, Arslan Shaukat, Rizwan Ahmed Khan, Zartasha Mustansar, Muhammad Usman Akram and Umer Asgher "Recognition of Children's Facial Expressions Using Deep Learned Features"

[23] K. Keerthana, D. Pradeep Dr. B. Vanathi "Learner's Engagement Analysis for E-Learning Platform"

[24] Ati Jain, Dr. Hare Ram Sah and Harsha Atre "Student's Emotion Recognition through Facial Expressions during E-Learning using Fuzzy Logic and CNN classification"

[25] M. Ali Akber Dewan, Mahbub Murshed and Fuhua Lin "Engagement detection in online learning: a review"

[26] Sana Ikram, Haseeb Ahmad, Nasir Mahmood, C.M. Nadeem Faisal, Qaisar Abbas, Imran Qureshi and Ayyaz Hussain "Recognition of Student Engagement State in a Classroom Environment Using Deep and Efficient Transfer Learning Algorithm"

[27] Fezile Ozdamli, Aayat Alijarrah, Damla Karagozlu, and Mustafa Ababneh "Facial Recognition System to Detect Student Emotions and Cheating in Distance Learning"

[28] Yi Chen, Jin Zhou, Qiating Gao, Jing Gao, and Wei Zhang "MDNN: Predicting Student Enagegement via Gaze Direction and Facial Expression in Collaborative Learning"

[29] Nha Tran, Hung Nguyen, Hien Luong, Minh Nguyen, Khiet Luong, Huy Tran "Recognition of Student Behaviour through Actions in the Classroom"

[30] Benyoussef Abdellaoui, Aniss MOUMEN, Younes ELBOUZEKRI EL IDRISSI, Ahmed Remaida "Face Detection to Recognize Students Emotion and Their Engagement: A Systematic Review"