

Research Article

Estimation of Learning Affects Experienced by Learners: An Approach Using Relational Reasoning and Adaptive Mapping

Anil Audumbar Pise , Hima Vadapalli , and Ian Sanders 

School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

Correspondence should be addressed to Anil Audumbar Pise; anil.pise7@gmail.com

Received 17 December 2021; Accepted 8 January 2022; Published 7 February 2022

Academic Editor: Kalidoss Rajakani

Copyright © 2022 Anil Audumbar Pise et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Various studies have measured and analyzed learners' emotions in both traditional classroom and e-learning settings. Learners' emotions can be estimated using their text input, speech, body language, or facial expressions. The presence of certain facial expressions has shown to indicate a learner's levels of concentration in both traditional and e-learning environments. Many studies have focused on the use of facial expressions in estimating the emotions experienced by learners. However, little research has been conducted on the use of analyzed emotions in estimating the learning affect experienced. Previous studies have shown that online learning can enhance students' motivation, interest, attention, and performance as well as counteract negative emotions, such as boredom and anxiety, that students may experience. Thus, it is crucial to integrate modules into an existing e-learning platform to effectively estimate learners' learning affect (LLA), provide appropriate feedback to both learner and lecturers, and potentially change the overall online learning experience. This paper proposes a learning affect estimation framework that employs relational reasoning for facial expression recognition and adaptive mapping between recognized emotions and learning affect. Relational reasoning and deep learning, when used for autoanalysis of facial expressions, have shown promising results. The proposed methodology includes estimating a learner's facial expressions using relational reasoning; mapping the estimated expressions to the learner's learning affect using the adaptive LLA transfer model; and analyzing the effectiveness of LLA within an online learning environment. The proposed research thus contributes to the field of facial expression recognition enhancing online learning experience and adaptive learning.

1. Introduction

E-learning is flexible and therefore can meet various challenges posed within the sphere of information technology (IT), and most notably it has the potential to widen people's access to knowledge. The joint impact of communication and IT on learning offers various routes of exploration, such as how best to capture and keep learners' attention, as well as developing active and flexible learning environments that motivate students to learn continuously through the use of a variety of IT tools [1].

With the COVID-19 situation, e-learning has been at the forefront in providing quality education to learners. E-learning refers to using computers and online learning tools in a blended learning environment, focusing on collaborative online learning [2]. E-learning can be

advantageous as it allows learners to progress at their own pace and access information at the time and place that suits them best [3, 4]. However, it also presents some challenges. One such challenge has to do with the congruence between the e-learning environments' and students' characteristics, as well as the effect of all users', students', and teachers' cultural-educational background on their personal preferences within the e-learning environment. Currently, rapid social and technological changes emphasize the need for lifelong learning. Due to the enormity of this need, it cannot be met in a traditional classroom setting. Thus, e-learning provides an alternative way to pursue lifelong learning via the internet. This mode of knowledge transmission is rapidly gaining momentum due to recent advances in computer technologies, as well as research into the pedagogical methodologies linked thereto. Online learning has become

part of the “normal” landscape in training and education because it widens access to knowledge. Furthermore, it also offers learners and instructors flexibility and convenience, since they can upload/access information when it suits them best. One disadvantage is that some online learning sites provide learners with a wealth of information but do not offer them any support with using the information to construct meaning. In such situations, students are passive receivers of information, which does not promote the principles of lifelong learning [5].

In a physical classroom, the instructor and students can send and receive non-verbal communication signals. Instructors who are aware of nonverbal signals sent in the classroom can modify their instruction in response to negative signals (confused looks) from students, as well as send positive nonverbal signals that may aid in student learning. These signals are unfortunately absent in a virtual classroom, which makes it very difficult to gauge students’ engagement and boredom levels. Thus, a very informative lecture can completely miss the mark and become ineffective due to the inability of the platform to provide lecturers with students’ non-verbal communication signals. For successful online teaching and learning, it is imperative to overcome this drawback. Thus, there is a need for a system that can operate within an e-learning environment and provide real-time feedback to lecturers on their students’ levels of engagement in a virtual classroom.

Within the world of educational technology, empathetic interactions between the user and the system are very important. This can be achieved through the use of multiple channels such as text, audio, and visual modalities. This paper proposes and discusses the development of a non-intrusive model that can assess the engagement level and emotional state of the learner and then generate appropriate feedback. In the proposed study, a deep learning-based facial emotion recognition (FER) system, in particular a temporal relational network (TRN) FER system, is used to predict the expressions exhibited by a learner as recorded via a webcam. Furthermore, the predicted expressions from the system are used to estimate the emotional state and engagement level of the learner which allows for generation and distribution of appropriate feedback. For that reason, it is believed that the integration of such a system into the online learning environment can improve the overall platform. Figure 1 illustrates an autonomous e-learning evaluation system based on emotion recognition that can offer lecturers real-time feedback on their students’ levels of involvement in a virtual classroom.

The basic framework to identify learners’ affective state depends on the detection of basic emotions, such as happy, sad, sadness, happiness, surprise, fear, anger, and disgust, from facial cues. For this purpose, DISFA + dataset is used which includes both posed and non-posed emotions. Posed emotions include “sadness” and “happiness” labels. Figure 2 provides sample images from DISFA + dataset with posed face expressions. Non-posed emotions are spontaneous, with notable muscle movements of the face. Both posed and non-posed emotions are further classified into two broad categories, namely, positive and negative emotions.

The rest of this paper makes use of the following structure. The relevant literature in the field is provided in Section 2 and methods for estimating learning affects are given in Section 3. Mapping emotions to learning affect are introduced in Section 4 and types of learning affects are introduced in Section 5. The proposed FER system is reviewed in Section 6 along with the methodology employed being discussed briefly and proposed mapping emotions to learning affect are discussed in Section 7. Live testing of proposed models using GUI and experimental results are discussed in Section 8. Conclusion and possible future work are discussed in Section 9. Section 10 concludes the article with a discussion.

2. Related Research

This section highlights the work done so far in the area of emotion recognition, e-learning, and use of facial expressions.

2.1. Emotion and Expression Recognition. Recently, studies have suggested that brain function is affected when basic emotional mechanisms are missing. Thus, emotions are needed for knowledge production. However, it is also indicated that extremely high emotional reactions have an adverse effect on rational thinking [7]. It reasonably follows that positive emotions such as joy, acceptance, satisfaction, and trust can lead to higher levels of creativity and accurate decision-making and problem-solving skills that can enhance learning [8]. Negative emotions such as fear, anger, and sadness have a negative influence on the brain and prolonged emotional distress can adversely affect the learning process and can be demotivating to a learner. For example, a learners’ ability to memorize and remember can be affected by depression and anxiety which can lead to frustration and despair which is ultimately expressed in emotional forms such as anger, fear, and/or sadness [9]. In these scenarios, good feedback practices can help get learners back on track as well as motivate them, which can ultimately lead to enhanced learning. Thus, in an e-learning situation, an effective system should be able to read the learners’ emotions and measure their attention in order to provide intelligent feedback that will enhance the learning experience of the learners. Effective and intelligent feedback in e-learning can be given to learners using embodied conversational agents (ECAs) [10]. These ECAs can be given a digital persona and can communicate with the user verbally or non-verbally. They can also express emotions that are more effective than a “faceless” computer providing feedback. Unfortunately, building these systems can be quite challenging.

When working with emotion recognition, it can be challenging to map emotional states according to facial expressions. To counteract this, Paul Ekman set out to map universal facial expressions [11] for emotions such as disgust, fear, anger, sadness, surprise, and happiness. Furthermore, this field gained more interest in the late 1990s when affect was successfully recognized by a machine from

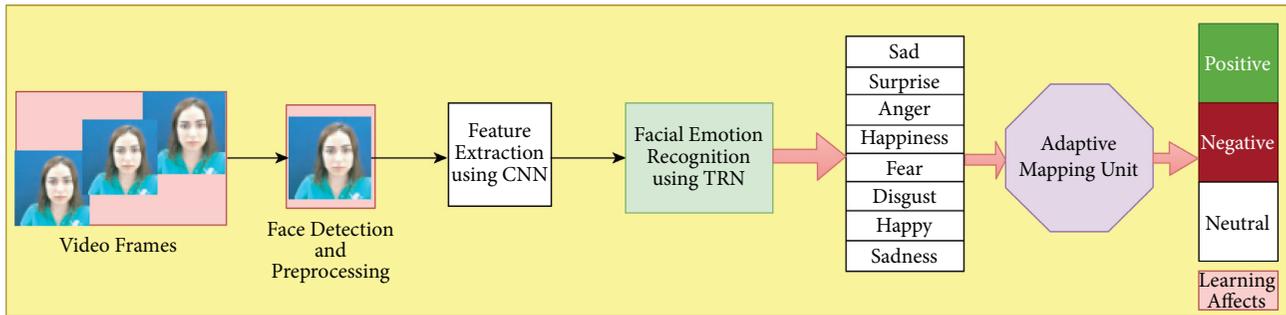


FIGURE 1: Automatic evaluation system for e-learning based on emotion recognition.



FIGURE 2: Sample images of posed facial expressions from DISFA+ dataset [6].

static images as well as from audio-visual signals. Literature suggests that emotion data should be drawn from observing the entire face and specifically noting the use of certain facial muscles. This is called the sign-judgment approach [12]. Additionally, the Facial Action Coding System (FACS) can be used to classify the action units (AUs) of facial expressions, which can then be labeled according to emotion [13]. Bartlett et al. [14] made use of AUs in their studies and found it to be a robust detection system to accurately label emotions. Other models, such as the geometric feature-based model, trace the variations in the shape and size of facial components (mouth, eyes, and eyebrows) to identify emotions. On the other hand, holistic approaches make use of a variety of machine learning approaches to mine facial features to ascertain emotions. Thus, making use of already available applications, it seems possible to identify emotions based on facial expressions. Face Reader [15] employs FACS to differentiate between six basic emotional states. The accuracy of this application is 89%. Happy et al. [16] also reported a substantial accuracy in their work on facial expression recognition by employing local features on a person-specific dataset. It has also been indicated that facial expressions can hint at whether a person is feeling bored or tired; however, this has not been researched in depth as of yet.

Furthermore, posture and gestures can also provide insight into a person's affective state, as well as a level of attention and interest. This means that data based on gestures and body posture can be used to measure the cognitive state of the learner. However, these aspects have not been researched in depth. These aspects can be combined with an affective tutoring system that predominantly relies on

identifying emotional states for appropriate feedback during the learning process. In [17], the authors proposed a preliminary architecture that makes use of both emotional responsiveness and personality for the virtual tutor. Claims have been made by Woolf et al. [18] that a computer-based tutor should be able to detect and analyze levels of motivation, confidence, boredom, frustration, and fatigue in order to provide relevant feedback for each of these states. Other researchers [3, 4] have looked into detecting basic emotions in an e-learning situation. They made use of ECAs to accomplish parallel empathy and thereafter reactive empathy that expresses emotion through voice and articulation.

Another interesting field is automatic engagement recognition. This makes use of computer vision systems that have the ability to discreetly measure learner engagement by analyzing body posture, hand gestures, and facial cues [19]. An engagement recognition system that works in real time could be applied widely to the following scenarios: (i) teachers who work in distance learning could receive immediate feedback based on the engagement levels of their learners; (ii) participants' reactions could be used to identify sections of a video where people are disengaged, which can then be addressed by the maker of the video; (iii) to gather data on the causes and variables that affect learner engagement; and (iv) institutions could use this technology to monitor online engagement.

2.2. Facial Expressions and Their Significance in E-Learning.

E-learning does not yet address the issue of lower levels of student engagement, which can easily be addressed to a great extent in face-to-face learning situations where the teacher can observe when students stop engaging with content. To overcome this problem, adaptive e-learning systems (AESs) have been implemented in online environments. In essence, these systems automatically adapt to a student based on the student's actions and a set of independent conclusions about that student [20]. Thus, these systems help to tailor learning material to individual students based on each student's goals, preferences, level of knowledge, and preferred learning style [21]. Adaptive e-learning systems adapt content and way it should be presented to each student. Over the last decade, the development and use of AES has increased and become an important cornerstone of online learning platforms [22].

Online learning can be adapted based on the following: (i) user preferences—the learning experience and materials are adapted to the needs and preferences of each individual user; (ii) user behavior—the user’s online learning behavior is tracked and informs which adaptations are necessary; and (iii) user performance—the user’s performance in online learning provides input on how the learning experience and material should be adapted per user based on his preference. When data from these three metrics are combined into a user model (UM), it has the potential to enhance the learning experience. According to Stoyanov and Kirchner [23], “An adaptive e-learning system is an interactive system that personalizes and adapts e-learning content, pedagogical models, and interactions between participants in the environment to meet the individual needs and preferences of users if and when they arise.”

In any classroom, the most important interaction is between the teacher and the learners [24]. Communication via facial expressions plays a very important role during this interaction as faces have an ability to impart information about a learners’ mood or present mental and emotional state of being and hence, to an extent, their internal feelings. According to Dragon et al. [25], the facial expressions of the lecturer play an important role in keeping the students motivated during the entire period of the lecture. A lecturer also uses the facial expressions exhibited by students as a source of motivation for future classes. It can also be an indicator to the lecturer about the general information the students would like to impart to the lecturer such as whether the pace is in line with the students’ learning and whether the students are confused about a topic. In a nutshell, a lecturer should be able to analyze the facial expressions and hence the comprehension levels to some extent and use them as an indicator to modify his/her delivery. Failure to understand the facial expressions exhibited by the students can hamper the understanding of the impact of a non-conductive style or pace of delivery and hence the levels of student learning.

2.3. Facial Emotions and Learning Affects. Russell’s (1980) two-dimensional “circumplex model of affect” is widely employed in the field of user emotion modeling, in which emotions are viewed as combinations of arousal and valence [26–31]. It has also been demonstrated that the OCC [32] psychological constructivist approach model should be utilized as the standard cognitive appraisal model for emotions. This model includes 22 emotion categories based on emotional reactions to scenarios such as (i) the goals of key events, (ii) the conduct of a responsible person, or (iii) attitudes towards appealing or repulsive products, among other things. Conati and Zhou [33] used the OCC model to recognize the emotions of users when they built the instructive game Prime Climb in 2002. It was the first time Katsionis and Virvou [34] used the OCC model to map students’ emotions while playing an educational game. Emotions are often utilized in the design and development of instructional materials. Picard et al. [7] embarked on a project that they defined as “creating a math they will adore”

rather than “trying to make youngsters appreciate the arithmetic they loathe.” Participants were included in the design of things to learn in order to elicit emotions that will aid in the learning process.

In both traditional classroom settings and e-learning environments, researchers have sought to evaluate and quantify a student’s comprehension level and learning affects. To assess learning affect, biophysical sensors, eye-gaze, body gesture, facial expressions, questionnaires, and other techniques are employed. However, there has been minimal investigation investigating the relationship between facial expressions and learning consequences [35]. In the next section, the methods for estimating learning affects are examined and elaborated in depth.

3. Methods for Estimating Learning Affects

Researchers attempted to evaluate students’ learning affect in both traditional classroom and e-learning environments in order to estimate and quantify a student’s understanding level as well as their learning affect in order to improve learning outcomes. A variety of methodologies, including questionnaires, biophysical sensors, eye-gaze, body gesture, and facial expressions, are used to evaluate learning affect. Although some research has been conducted on the association between face emotions and learning outcomes, it has been limited.

3.1. Biophysical Signal. Shen et al. [36] investigated biophysical signals to assess how emotions evolve during teaching and learning, as well as whether or not it was possible to apply the findings to increase learning in their study. Their research examined the emotions of students who were taking part in a learning process using Russell’s circumplex model of affect, as well as machine learning techniques such as support vector machine (SVM) and K-nearest neighbor (KNN), among others. They reported that SVMs surpass other methods with an accuracy of 86.3%. The researchers discovered that incorporating emotion data into e-learning will improve the overall performance of learners. Their findings conclude that emotional data are necessary for a learning system to function properly. However, the system was assessed in a laboratory setting, and the equipment that was used limited the movement of learners and made it uncomfortable for them to use. Incorporating such a system into an e-learning environment could prove to be a difficult task. The findings reported from the AutoTutor project suggest that engagement and confusion were the most prominent and relevant emotions in learning.

3.2. Body Gestures. Using an automated automated gaze, Bidwell and Fuchs [37] were able to measure student participation. It was determined whether or not a student was engaged by watching records of a learning phase filmed in the classroom. In a similar manner, Klein and Celik [38] employed body movements to determine student concentration in a classroom context. They trained a CNN on image

data collected during learning sessions to determine whether or not the students were “engaged.” They were able to achieve an accuracy of 89.7%.

3.3. Facial Emotions. Sathik and Jonathan [39] identified four non-verbal communication channels that were most frequently utilized for learning and then examined which of these channels were most frequently used for learning. According to the findings of their study, people utilize facial expressions the most to interact with one another in non-verbal ways. Ayvaz et al. [40] devised a method for determining how motivated people are in an online learning environment by employing an online video conference learning environment. Their system was able to read a student’s emotions and provide information about them to the teacher. The feedback was critical in creating a more immersive learning environment. Hammoumi et al. [41] devised a method for determining how a learner feels in an e-learning environment. The framework was created using instructive game software. The method was evaluated with the help of young learners. They recorded and analyzed their emotions in real time using a webcam. During the learning process, the learners’ emotions were observed even when they did not look into the camera. When pupils encounter an issue, their findings demonstrate that they are remorseful. Pan et al. [42] devised a method for using facial recognition technology to keep track of how many students show up for class. Later, they devised a method for determining how effectively students were engaged in a lesson by observing their facial expressions. To classify the learner’s facial expression and level of attention, the stimulus-response approach was utilized. Six emotions, it turns out, have an impact on how well you learn (concerned, curious, pondering, comprehension, disrespect, and disgust). A model of how each antenna type’s learning effects transfer was created to figure out how the teaching tactics were performed. This strategy enabled a quantitative examination of classroom instruction practices. They provided a theoretical basis for assessing learning affect by observing a learner’s facial expressions. However, the evaluation was done by manually rather than automatically.

3.4. Questionnaires. Self-reporting refers to the procedure of administering questionnaires to external observers or to participants in a learning environment. The administration of questionnaires is straightforward, and they are widely used to assess learners’ involvement or learning affect. Using daily diaries as questionnaires, Zembylas [43] studied the emotions experienced by online distance learners over a six-month period. They next employed open-coding approaches to categorize the emotions of the trainees based on their responses to the questionnaire. In the study, the researchers discovered that when students interact with e-learning platforms, they experience both pleasant and negative emotions. It was revealed that these feelings changed as the learning process continued. As a result of this investigation, it is obvious that emotions and learning are interconnected. Sathik and Jonathan [39] used questionnaires in a similar

way to determine the most often used non-verbal communication strategies in the classroom, as well as the relationship between facial expression and the level of comprehension of the students in the classroom. According to their research, the most frequently used non-verbal cues in learning are facial expressions of happiness and sadness. Additional research demonstrates that when a student comprehends the course content or is satisfied with the course’s delivery, a beneficial affect is demonstrated. When pupils do not comprehend the course topic, on the other hand, a negative reaction is demonstrated. The use of sensors such as cameras, microphones, pressure gages, particular monitoring devices, and other instruments to analyze the learner’s emotional changes is a more complex and analytical approach of analyzing the learner’s affective state changes [7].

This study proposes the use of facial expressions to ascertain the learning affect experienced by learners. This is performed by the use of TRN to predict facial expressions. The extended DISFA dataset is used to train the network. The TRN architecture is used because it is capable of capturing and extracting meaningful and useable attributes from images, as well as having state-of-the-art performance on image classification problems [44, 45].

4. Mapping Emotions to Learning Affect

There has been little research into the relationship between facial expressions and learning affect. During their experiment, Kapoor et al. [46] revealed that learners’ learning experiences were connected with facial action units (AUs). The facial action units, which are determined by the facial muscle motions, define the deformation of the face characteristics. A grin (AU 6 + AU 12), tightening of the eyelids (AU 7), widening of the eyes (AU 5), and rising of the brows (AU 1 + AU 2), according to their research, are all powerful markers that the learner is engaged and passionate about the learning phase or activity. Lower brows (AU 1 + AU 4), nose creases (AU9), and lower lip corner rounding all indicate a lack of motivation (AU 15). Lifting of the lower lip corners (AU 1 + 4) does not imply a lack of motivation.

In response to classroom observations and responses to surveys, Sathik and Jonathan [39] proposed a mapping approach based on their findings. Their work related facial movements to the level of comprehension achieved by the students. Their findings suggested that raised eyebrows and wide-open eyes were indicators of positive learning, whereas narrowing eyes, wrinkles on the forehead, narrowed eyes, and curling lips indicated a negative learning experience, and a smile or a neutral facial expression indicated an undecided learning state. Happy et al. [12] proposed an emotion detection and automated alertness system for empathetic feedback generation during e-learning. According to their findings, empathetic feedback generated during e-learning will be able to estimate the user’s affective state and assess the user’s attention level. They classified detected emotions as positive (e.g., happiness and surprise) or negative (e.g., anger, sadness, fear, anger, and disgust). Based on data collected from learning videos in classrooms,

Pan et al. [42] presented a mapping of seven emotional states (worry, interest, thinking, comprehension, disregard, and disgust) to three learning outcomes (positive, negative, and neutral) in a related study. Despite the fact that their findings did not include a clear mapping of facial emotions to learning affects, a psychologically informed description of the AU's and facial movements served as indications that face emotion may be mapped to either positive or negative learning affects. In [35], Zakka and Vadapalli investigated the use of facial expressions expressed by learners for the interpretation of their learning affect in an e-learning session, which they found to be effective. A standardized mapping mechanism between facial emotions displayed and their related learning effects was also investigated by the researchers. In conclusion, the researchers found that the system detected the participants' facial expressions of emotion and mapped each recognized facial emotion to its corresponding learning affect in accordance with the system's initial mapping. Table 1 provides a summary of the mappings that were discussed previously.

5. Types of Learning Affects

Studies from the field of behavioral science and neuroscience reveal that emotion recognition and emotional processes form the basis of behaviors by not only being interactive but also by integrating with the function of neural mechanisms [47]. The behavioral performance of an individual is influenced by the emotions and learners' cognitive drive and motivation [48]. The levels of creativity and problem-solving ability are developed by positive emotions as they have the potential to influence the learners' cognitive thinking and learning intelligence, whereas negative emotions have negative effects on their thinking [49].

To measure learning affect in an online learning environment, we propose the use of facial expressions and adaptive mapping between estimated facial expressions and learning affect. Participants' learning affect can be typically categorized and placed into the following five groups [50].

5.1. Positive Learning Affect. When a student has a "happy" and "fear" facial expression or frowns while engaging with academic material, it potentially indicates that the student is engaging with what is happening on the screen and is thinking indicating a positive learning affect. A frown also indicates that the material presented is challenging the student. In such feedback scenarios, an instructor might opt to reduce the difficulty level of the content to ensure a very positive learning affect.

5.2. Very Positive Learning Affect. When students have a look of "surprise" and "happiness" on their faces while engaging with the online academic material, it indicates that students are curious about the content they are currently interacting with. A very positive learning affect means that the content is well designed and that students find it interesting and stimulating.

5.3. Neutral Learning Affect. Neutral learning affect occurs when students' expressions are friendly while engaging with online learning material. This indicates that students are able to comprehend the content.

5.4. Negative Learning Affect. A negative learning affect is characterized by students' indifference towards the online learning of the material they are viewing or disregard for the instructions from the lecturer. This maps to the presence of negative expressions such as "sadness" and "disgust."

5.5. Very Negative Learning Affect. Very negative learning affect refers to a situation where students who do not look at the material on the screen and who also exhibit negative expressions that show "anger." This may be a result of the learning material being pitched at too high a level for the students to comprehend, or there might be personal reasons as to why a particular student shows a very negative learning affect.

These categories of learning affect can be employed in the current analysis to evaluate students' levels of engagement and to provide feedback to instructors on their online teaching practices, which in turn can be used to enhance the online learning environment. Note that, in the proposed experiment, "very positive" and "positive" learning affect will be considered "positive" while "very negative" and "negative" will be considered "negative" learning affect.

6. Methodology and Proposed FER-Learning Affect-Feedback System

The objective of our FER system was to predict the engagement level of learners using relational reasoning and a deep learning approach. Figure 3 illustrates the system architecture of the TRN-based FER system with adaptive mapping module. The main stages of the system include (i) face detection and extraction of representative frames, (ii) emotion recognition using a pretrained TRN model, (iii) adaptive mapping to estimate learning affect, and (iv) feedback generation.

6.1. Proposed Methodology. In the framework, a TRN is used for identifying the changes in facial emotions exhibited by students. A TRN model takes in a video V which is composed of n frames $V = \{f_1, f_2, \dots, f_n\}$, where f_i represents the i^{th} frame. TRN can be applied on distinct multiple sets of frames or on a single set of frames. A multi-scale TRN (MS-TRN) [42] is a composite function and depicts frame relations at distinct scales and is given as

$$MT_N(V) = T_2(V) + T_3(V) + T_4(V) + T_5(V) + \dots + T_d(V), \quad (1)$$

where T_d is a single-scale TRN (SS-TRN) that establishes a temporal relation between a single set of " d " sorted frames with respective $h_\phi^{(d)}$ and $g_\theta^{(d)}$ values. T_d for a given value of " d " is given as

TABLE 1: Summary of facial expressions to learning affect by Pan et al. [42], Sathik and Jonathan [39], Kapoor et al. [46], Zakka and Vadapalli [35], and Happy et al. [12].

Description	Possible emotion type [12, 33, 39, 42, 46]	Learning affect				
		Pan et al. [42]	Sathik and Jonathan [39]	Kapoor et al. [44]	Zakka and Vadapalli [33]	Happy et al. [12]
Lowering eyebrows	Sadness	Negative	Negative	Negative	Not stated	Negative
Nose wrinkling	Anger	Negative	Negative	Negative	Negative	Negative
Depressing lower lip	Sadness	Negative	Negative	Negative	Not stated	Negative
Smile/pleasant expression	Happy	Positive	Not stated	Positive	Positive	Not stated
Eye lightening	Anger	Negative	Negative	Positive	Negative	Negative
Upper lid raiser	Fear, surprise	Not stated	Positive	Positive	Positive	Negative
Raising eyebrow	Fear, surprise	Not stated	Positive	Positive	Positive	Negative
Wrinkle forehead	Anger	Positive	Negative	Not stated	Negative	Negative
Not focused	Disgust	Negative	Not stated	Negative	Negative	Negative
Display surprise	Surprise	Positive	Not stated	Not stated	Positive	Positive
No expression	Neutral	Positive	Not stated	Not stated	Neutral	Not stated

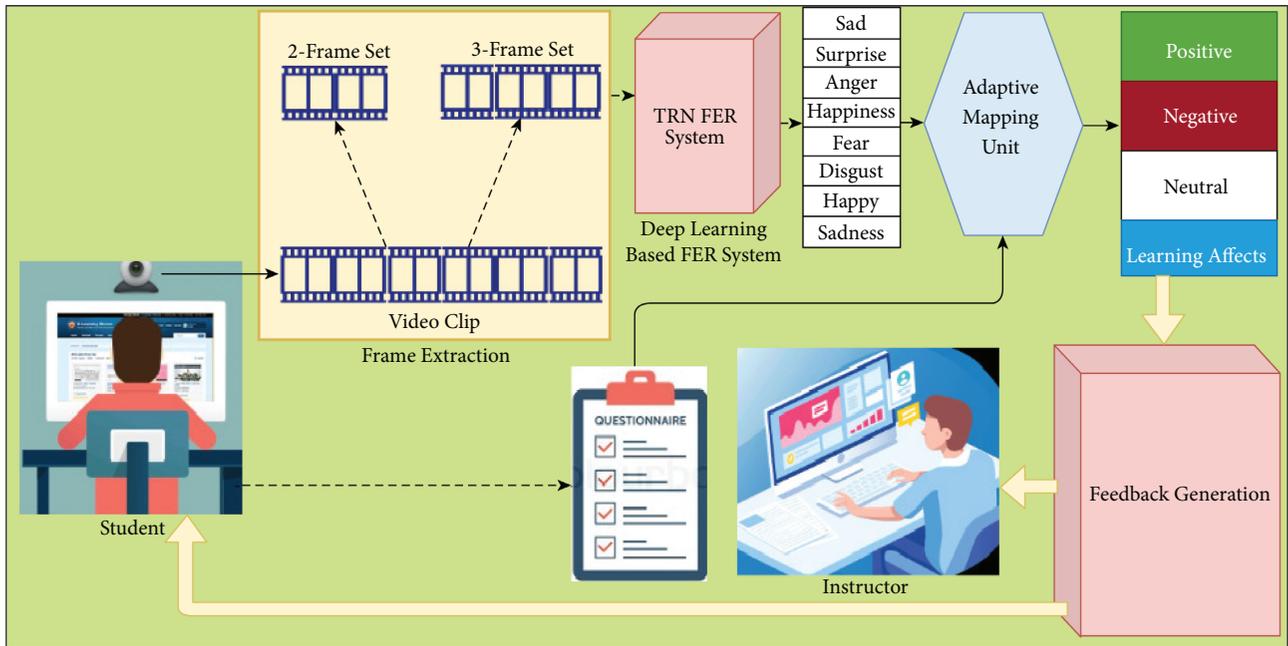


FIGURE 3: Proposed system architecture.

$$T_2(V) = h_\phi \left(\sum_{i < j} g_\theta(f_i, f_j) \right), \quad (2)$$

where $d = 2$, f_i and f_j are frames at times i and j such that $i < j$ and is known as two frame relation TRN. Similarly three frame relations TRN is given as

$$T_3(V) = h'_\phi \left(\sum_{i < j < k} g'_\theta(f_i, f_j, f_k) \right). \quad (3)$$

6.2. Face Detection and Extraction of Representative Frames.

Face detection and extraction of representative frames steps are required to align and normalize the input samples, allowing the deep neural network to learn meaningful facial features. Aspects that are unrelated to facial expressions,

such as the background or the angle and pose of the head, are relatively similar and should be effectively managed for efficient and accurate estimation of labels.

After retrieving all frames from the 251 video segments in the DISFA + dataset, representative video frames are acquired and meta-files are generated. By expanding the amount of input attributes, underfitting and bias are avoided [51]. Due to minute differences in facial expression, it was determined that the transition between frames is difficult to discern. When multiple filters are used, little changes in the frame are lost, making it impossible to train the dataset on certain emotions [52]. To remedy this, the frames are decreased to emphasize the face exclusively. However, it demonstrates that increasing the number of input features can reduce the high bias. To tackle the model's underfitting issue, we used several training and validation samples. The video frame scale is also trimmed to 256×256 . Due to the fall in pixel value, it is difficult to notice subtle changes in the frame for long-term videos.

During a scheduled live session, the user interface for the e-learning model in the proposed work automatically initiates video capture whenever a learner plays a video from a given e-learning site. This GUI was designed for live testing and was created using PyQT5 and Python3. The OpenCV 4.2 function, in conjunction with PyQT5, is used to capture a video of the learner using a computer webcam. Figure 3 depicts a snapshot of video footage and frame extraction during live testing. Faces detected were scaled to 256×256 and normalized. These images are then fed into the trained TRN FER system, which is used to anticipate a learner's emotions.

6.3. Emotion Recognition Using a Pretrained TRN Model.

The generated set of frames is passed as an input to the deep learning-based FER system that uses a TRN model. The TRN system has the ability to link meaningful transformations between different frames and recognize the existence of different facial emotions (sad, surprise, anger, happiness, fear, disgust, happy, and sadness) [45, 53]. As previously stated, because of the small dataset, there was an issue with underfitting. 6024 samples were developed from the first 251 samples to remedy this issue. Following sample enlargement, it was separated into training and validation sets. Additionally, the normalization technique has been designed specifically for the Squeezenet1_1 design from the BN-Inception architecture. The adjustment is driven by the fact that SqueezeNet's deep compression method is model reduction-friendly. The SqueezeNet models were trained for 180 epochs on the DISFA + dataset using the training set.

After training a proposed model on the SqueezeNet base model using 6024 repeated samples from the DISFA + dataset, it is observed that loss is reducing and that the precision values for the Top 1 and Top 5 are approaching 100%. SqueezeNet's deep compression approach has been used to reduce the size of the model. Downsampling using pooling layers is applied later in the network to increase accuracy without being greedy in the number of parameters used. Formally, the fire module is a squeeze convolutional layer followed by an expanded convolutional layer. According to Iandola et al. [54], SqueezeNet (Left) begins with an isolated convolution layer (conv1), followed by eight separate fire modules (fire2–9) (conv10). The model steadily increases the number of kernels per fire module. Following the conv1, fire4, fire8, and conv10 layers, a stride of two is used to perform the max-pool procedure. In addition to simple bypass (Middle) and advanced bypass, SqueezeNet is used in this work (Right). Bypass is utilized in accordance with ResNet's findings [55].

In Table 2, testing of the multi-scale TRN model with the DISFA + test dataset samples is represented using a confusion matrix table. In this test, proposed architecture test samples of the dataset are tested with multi-scale TRN. The same hyperparameters are used as a learning configuration, with learning rate set to 0.001, dropout set to 0.5, and momentum set to 0.6. It is observed that after completing testing, the best Top 1 and Top 5 precision values achieve 100% with a loss of 0.00%. For the multi-scale TRN, overall

test accuracy achieved is 91.30%. It is observed that the emotion that has the highest emotion recognition accuracies is surprise.

Towards achieving the goal of emotion recognition, the TRN model detects a face from the frames which are extracted from a real-time recording of a video clip of a learner. The extracted facial features are then used to identify the correct emotion the participant is experiencing and classify it as one of eight basic emotions. This is achieved in real time.

6.4. Representative Frame Detection. In this research, we offer a process able to recognize representative frames in videos utilizing a relational reasoning approach. Once a representative frame is found, the TRN-based facial emotion recognition technique described in Pise et al. [45] will be executed on the observed representative frames. The proposed scheme offers two advantages. First, the CPU time corresponding to the recognition component will be greatly decreased since only a few representative frames are considered. Second, representative frames are specified by the frames where the facial expressions change significantly. Thus, one can determine a representative frame by scanning the sequence for representative frames that have been modified by the expression frame. Figure 4 illustrates the whole block architecture for representative frame detection utilizing multi-scale TRN by probability score.

Users are unaware of the technical aspects of proposed architectures (multi-layer perceptron (MLP), SS-TRN, and MS-TRN), which should be enabled only by researchers. Following model activation and selection, we must press the START button to initiate the live video stream processing. When the user initiates a live test, the window displays the predicted emotion label. Each anticipated label is saved in a text file at the backend, together with the system timestamp and label. When the same emotion occurs more than once, the label automatically keeps it once until it disappears or the label category changes. The label category assigned to the frame that is considered representative changes. This indicates that among the series of these frames, the frame with the highest intensity of features capable of changing the expression label was identified as the representative frame. Also, the frames that represent the changes that took place in the sequence were considered as representative frames.

7. Proposed Mapping Emotions to Learning Affects

The next phase of the proposed model involves mapping a detected emotion to a relevant learning affect. This research incorporates the mappings of Happy et al., Zakka and Vadapalli, Sathik and Jonathan, Pan et al., and Kapoor et al. [12, 35, 39, 42, 46]. The integration was carried out in order to compile the mapping from the literature and arrive at a direct mapping between exhibited facial emotions and learning affect. Table 3 shows an initial mapping based on mappings identified in the literature. In the initial mapping, "happy" and "fear" were mapped onto

TABLE 2: Confusion matrix of FER results with multi-scale TRN on test set until 120 epochs.

Actual	Predicted							
	Sad	Surprise	Anger	Happiness	Fear	Disgust	Happy	Sadness
Sad	4					1		1
Surprise		6						
Anger			6			1		
Happiness				5				
Fear					7			
Disgust					1	5		
Happy							5	
Sadness								4
Overall Accuracy=91.30%								
TN+FN	TP	FN	FP	Precision	Recall	F-Measure		
6	4	2	0	100.00	66.67	80.00		
6	6	0	0	100.00	100.00	100.00		
7	6	1	0	100.00	85.71	92.30		
5	5	0	0	100.00	100.00	100.00		
7	7	0	0	100.00	100.00	100.00		
6	5	1	2	71.43	83.33	76.92		
5	5	0	0	100.00	100.00	100.00		
4	4	0	1	80.00	100.00	88.88		
46	42							

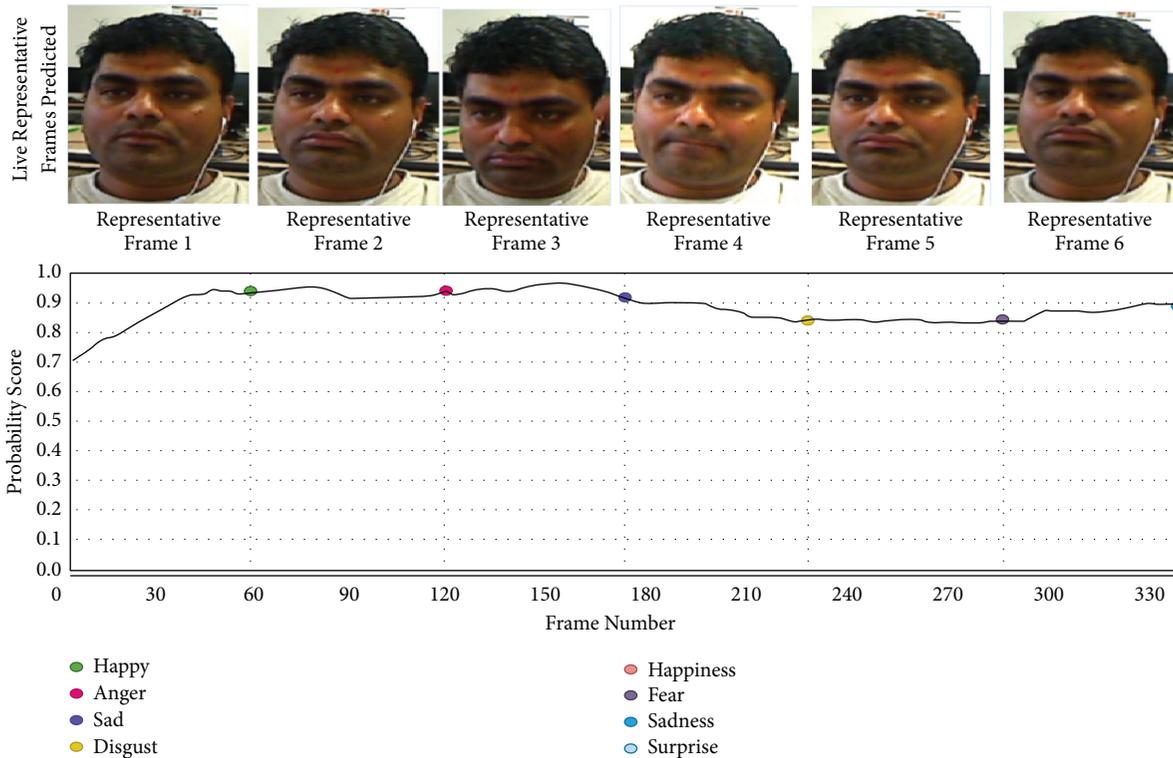


FIGURE 4: Probability score given by multi-scale TRN for representative frames.

positive learning affect, “surprise” and “happiness” were mapped onto very positive learning affect, “sad,” “sadness,” and “disgust” were mapped onto negative learning affects, while “anger” was mapped onto a very negative learning affect.

As seen in Table 3, the “fear” emotion is associated with a positive learning affect. As a result of this, the following question arises: is “fear” a good indicator of learning? To that purpose, this paper also suggests an adaptive mapping approach in which the initial mapping in Table 3 is compared

TABLE 3: Proposed facial emotion mapping to learning affect.

Emotion	Pan etal. [42]	Sathik etal. [39]	Kapoor etal. [46]	Zakkaand Vadapalli [35]	Happy etal. [12]	Initial Mapping
Happy	Positive	NotStated	Positive	Positive	NotStated	Positive
Fear	NotStated	Positive	Positive	Positive	Negative	Positive
Surprise	NotStated	Positive	Positive	Positive	Positive	VeryPositive
Happiness	NotStated	NotStated	NotStated	NotStated	Positive	VeryPositive
Sad	NotStated	NotStated	NotStated	Negative	NotStated	Negative
Sadness	Negative	Negative	Negative	NotStated	Negative	Negative
Disgust	Negative	NotStated	Negative	Negative	Negative	Negative
Anger	Negative	Negative	Negative	Negative	Negative	VeryNegative

to the responses provided by learners during a live testing session: that is, the learning experience they had while using the proposed system. These responses will aid in the adaptation of the mapping and the establishment of a dependable relationship between face emotions and learning effect.

8. Live Testing and Experimental Results

Twenty volunteers agreed to take part in a test run of the proposed framework. The live testing took place in a laboratory setting, with each participant seeing a video through a computer system equipped with a webcam. As soon as the video began collecting images of the learner’s face, the camera was activated. Before being fed into the trained TRN FER model, these images are preprocessed (resized and reshaped). Using the first mapping in Table 3, classified emotions are mapped to their relevant learning outcomes. The program estimates face emotions at the end of each learning session and correlates the top three recognized facial emotions with associated learning affects.

8.1. Learning Affect Estimation. The emotion recognition model is used to study the effect of learning on the recognition of face emotions. As indicated in Table 3, each facial emotion identified by the algorithm is mapped onto a learning affect. The learning affects were projected based on the system’s estimation of the top three facial emotions. According to the proposed mapping in Table 3, each facial emotion is mapped to its related learning affect. As previously stated, twenty (20) volunteers participated in live system testing. The system was evaluated in a laboratory setting, with each participant watching a twenty-minute video about the “Future Importance of Data Science.” At the completion of class, the participants enthusiastically answered survey questions on their educational experience (learning affect). Figure 5 depicts each participant’s emotions during the live test.

During the live testing sessions, the emotions “fear” and “disgust” were never observed. As a result, it is difficult to validate the initial mapping between these emotions and their accompanying learning affect. Therefore, one of the first and most concerning mappings, from “fear” to positive learning affect, could not be validated. The proposed model

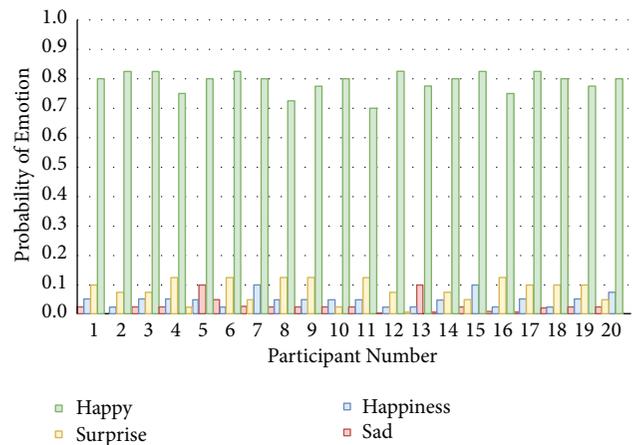


FIGURE 5: Emotion detected per participant during live testing.

needs be submitted for more rigorous testing in order to effectively adapt the initial mapping.

The “happy” facial expression was anticipated to be the most prevalent emotion regardless of the learner’s learning experience, i.e., both types of learners who expressed positive and negative learning affect displayed a “happy” feeling for the majority of the learning session (Figure 5). Additionally, a close examination of the video frames revealed that the majority of learners maintained a “happy” expression during the majority of the learning session, with only a few exceptions where learners lost interest, were distracted, were unable to understand, or found a portion of the course content particularly engaging.

At the end of the learning session, the learners replied to the survey questions. Their responses were compared to the estimated learning affect from the proposed approach.

The results suggest that the first estimated learning affect corresponded with 81% of the participants’ responses, 17% did not correspond, and 2% did not have a predicted first learning affect. The second learning affect prediction corresponded to 32% of the participants’ responses, 48% were identified as incorrectly predicted, and 20% of the participants’ had no second predictions. It was discovered that the third predicted learning affect did not match the participants’ experience. For example, the system predicted that all participants would have a “positive” learning experience, whereas participants’ responses indicated that they had a “negative” learning experience. Figure 6 depicts a summary

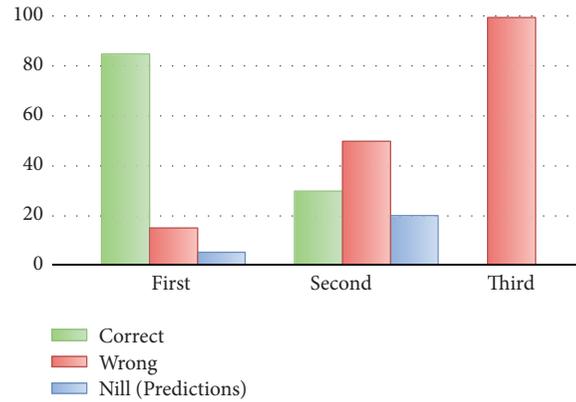


FIGURE 6: Correlation between the first, second, and third estimated learning affects with participants' responses.

TABLE 4: Emotion distribution per learner during learning session (%).

Learner	Emotions recognized by			Predictions			Learners' responses
	MLP	SS-TRN	MS-TRN	First	Second	Third	
1	Surprise	Happiness	Happy	Positive	Positive	Negative	Positive
2	Happiness	Surprise	Happy	Positive	Positive	Negative	Positive
3	Sad	Sad	Surprise	Negative	Negative	Negative	Negative
4	Happy	Surprise	Happy	Positive	Positive	Negative	Positive
5	Sad	Sad	Sad	Negative	Negative	Negative	Negative
6	Sad	Happy	Happiness	Positive	Negative	Negative	Positive
7	Happiness	Happy	Happy	Positive	Positive	Negative	Positive
8	Sad	Sad	Sad	Positive	Negative	Negative	Negative
9	Happy	Happiness	Surprise	Positive	Negative	Negative	Positive
10	Happy	Happy	Happy	Positive	Positive	Negative	Positive
11	Happiness	Happiness	Happy	Positive	Positive	Negative	Positive
12	Happy	Surprise	Sad	Negative	Negative	Negative	Positive
13	Sad	Sad	Sad	Negative	Negative	Negative	Negative
14	Happy	Happiness	Surprise	Positive	Negative	Negative	Positive
15	Surprise	Happy	Happiness	Positive	Negative	Negative	Positive
16	Sad	Happy	Happy	Positive	Negative	Negative	Positive
17	Sad	Happy	Happy	Positive	Negative	Negative	Positive
18	Happy	Happy	Happiness	Positive	Positive	Negative	Positive
19	Happiness	Surprise	Happy	Positive	Negative	Negative	Positive
20	Happy	Happy	Happiness	Positive	Negative	Negative	Positive

of the three predictions and their relationship to the learners' responses.

According to an analysis of the questionnaire responses, sixteen of the twenty participants in the evaluation had a "positive" learning experience, while the remaining four participants had a "negative" learning experience. According to the estimated emotion during the learning session, 60% of the 16 participants who experienced "positive" learning had a "happy" expression throughout the learning session, while 20% of the participants had other facial expressions recognized. Participants who responded that they had a "negative" learning experience, on the other hand, expressed "sad" (75%) or a "surprise" and "sad" (25%) expression as the top most expressions (see Table 4).

It is worth noting that, when compared to the first and second predictions, the first prediction was the most closely related to the learner's responses. However, due to the frequency with which the "happy" emotion occurs, the first prediction cannot be ignored. For example, the algorithm

identified "sad" and "surprise" face emotions as the second most expressed emotion among learners who reported a "negative" learning experience. Learners who had a "positive" learning experience expressed "surprise" or "happiness" throughout the learning session, with "happiness" being the second most prevalent emotion. As a result, the mapping mechanism's top-two emotion predictions will be more accurate, which is similar to the findings of Mollahosseini et al. [56] and Saravanan et al. [57]. It is also worth noting that none of the participants claimed to have experienced "anger" and "sadness" learning experience. The projections that the system made as to which basic emotion a participant is experiencing, based on their facial features, are stored in a customized dataset for each person.

8.2. Learners' Reactions to System-Generated Feedback. In addition to being an effective tool, providing fast feedback to learners has been demonstrated to increase overall learning

outcomes [58]. In this study, feedback was generated for each learner individually at the end of each learning session. Following viewing the feedback, participants' comments revealed that the most of them were encouraged to learn and gain knowledge. A few learners who experienced a positive learning affect were eager to continue with the video content, whereas others who experienced a negative learning affect were eager to rewatch the video in order to better understand the content. In general, students were pleased with the feedback they received.

The questionnaire was developed using the literature that is available in the field of computer science education and on student understanding in general. It is vital that students' feedback is collected making sure that a high response rate is registered with a low level of researcher bias [39]. The questionnaires obtained from the candidates along with estimated facial emotions by the TRN model are fed into an adaptive mapping unit to predict the learning affects such as positive, negative, or neutral. In this way, the use of facial emotions for estimation of learning affect involves recognition of facial emotions exhibited by a student during an e-learning session and estimation of the learning affect using an adaptive mapping mechanism between facial expressions and learning affects. An overview of the proposed system is given in Figure 3.

9. Conclusion and Future Work

The proposed TRN model was evaluated in this paper using both the DISFA + dataset and live images from a computer web camera. The system achieved an accuracy of 91.3%, 86.95%, and 80.43% for multi-scale TRN, single-scale TRN, and MLP on the DISFA + test dataset, respectively, and was also capable of making effective real-time predictions. As a result, the model was adopted as a trained emotion recognition model for the learning affect-based feedback system. The learning affect-based feedback system successfully mapped emotions to learning affect; study of the mapping and responses from learners suggested that the first prediction had a higher correlation with the participants' responses. The feedback generated by the system satisfied and inspired learners (participants).

The current work proposed an autonomous prediction model to measure learning affect and emotion depiction in an online learning environment that has the potential to be used in putting together tailored feedback to both learners and teachers. This was achieved through the use of a non-intrusive deep learning-based system that used visual cues obtained from participants. Based on participants' facial expressions, the system was capable of accurately predicting their cognitive and affective states. The proposed system can be integrated into an existing LMS. We propose that this stand-alone system should work with various LMSs, and we suspect that it will increase participant productivity.

This research shows that it is possible to estimate the affective state of participants as well as their engagement levels. The system performed admirably in recognizing participants' learning affect during a online learning session and subsequently categorized it as negative or positive

learning affect. The learning impacts were determined by categorizing participants' degrees of attention and discomfort as negative or positive. These classifications were used together as input for the feedback system. In future work, we would like to build a comprehensive assessment, and to achieve this, we propose to fuse the subunits. Furthermore, tailored feedback based on participants' emotional state and level of attention is also still underway. For instance, when a participant displays negative emotion, affective feedback will be provided to help the participant overcome the issue faced. Lastly, we suggest that a database should be created to help recognize different cognitive and emotional states for validation purposes.

10. Discussion

This work presents an e-learning model (system) that can measure learners' learning affect and generate real-time feedback. This was accomplished by utilizing TRN on the DISFA + dataset to construct a facial emotion recognition model, which was then used to forecast the learners' facial emotions in real time. Based on the mapping proposed in this study, the predicted face emotions were mapped to learning impacts.

Twenty (20) individuals participated in live testing of the proposed technology. The program recognized the participants' facial emotion expressions and, using the initial mapping, mapped each recognized facial emotion to its associated learning affect. The top three most estimated learning affects were examined, and the results were compared to learner survey replies. The findings of the analysis indicate that there is a correlation between facial emotions and learning affect, and that using the top two predictions for facial emotions is superior to using one or three predictions for learning affect detection.

Additional testing on larger samples of the suggested mapping mechanism may result in a more weighted mapping mechanism for learning influence predictions as a result of the results of the testing. It is recommended that future research in the e-learning system provide feedback to learners based on the estimated learning affect and also incorporate multi-model pattern analysis techniques such as body expressions, eye gazing, and head motions to produce an even more accurate result in identifying learning effects. Survey questions and feedback regarding the proposed system are given in the supplementary information file (available here).

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Supplementary Materials

In the Appendix Section, we have included questions that were presented to participants following the conclusion of

the live testing lecture. The questionnaires are divided into three sections. The first subsection contains general questions about the framework, the second subsection contains questions about how to use the system, and the final subsection contains general feedback about the system. Furthermore, under Supplementary Material, we have provided two distinct forms: a Consent Form for declarations of voluntary activity handed to participants and a Participant Information Sheet for future reference, as well as contact information for the University Human Research Ethics Committee. (*Supplementary Materials*)

References

- [1] R. M. Bottino, "The evolution of ICT-based learning environments: which perspectives for the school of the future?" *British Journal of Educational Technology*, vol. 35, no. 5, pp. 553–567, 2004.
- [2] S. Tsai and P. Machado, "E-learning basics: essay," *eLearn*, vol. 2002, no. 7, p. 3, 2002.
- [3] C. N. Moridis and A. A. Economides, "Affective learning: empathetic agents with emotional facial and tone of voice expressions," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 260–272, 2012.
- [4] S. Deniz, D. Lee, G. Kurian et al., "Computer vision for attendance and emotion analysis in school settings," in *Proceedings of the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0134–0139, Las Vegas, NV, USA, January 2019.
- [5] W. Horton, *E-Learning by Design*, John Wiley & Sons, Hoboken, NJ, USA, 2011.
- [6] S. M. Mavadati, P. Sanger, and M. H. Mahoor, "Extended DISFA dataset: investigating posed and spontaneous facial expressions," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2016*, pp. 1452–1459, IEEE Computer Society, Las Vegas, NV, USA, June 2016.
- [7] R. W. Picard, S. Papert, W. Bender et al., "Affective learning—a manifesto," *BT Technology Journal*, vol. 22, no. 4, pp. 253–269, 2004.
- [8] M. Lewis, J. M. Haviland-Jones, and L. F. Barrett, *Handbook of Emotions*, Guilford Press, New York, NY, USA, 2010.
- [9] C. N. Moridis and A. A. Economides, "Toward computer-aided affective learning systems: a literature review," *Journal of Educational Computing Research*, vol. 39, no. 4, pp. 313–337, 2008.
- [10] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and H. Hermens, "Design features of embodied conversational agents in ehealth: a literature review," *International Journal of Human-Computer Studies*, vol. 138, Article ID 102409, 2020.
- [11] J. B. Engelmann and M. Pogosyan, "Emotion perception across cultures: the role of cognitive mechanisms," *Frontiers in Psychology*, vol. 118, no. 4, 2013.
- [12] S. L. Happy, A. Dasgupta, P. Patnaik, and A. Routray, "Automated alertness and emotion detection for empathic feedback during e-learning," in *Proceedings of the 2013 IEEE Fifth International Conference on Technology for Education (t4e 2013)*, pp. 47–50, IEEE, Kharagpur, India, December 2013.
- [13] T. Skiendziel, A. G. Rösch, and O. C. Schultheiss, "Assessing the convergent validity between the automated emotion recognition software noldus facereader 7 and facial action coding system scoring," *PLoS One*, vol. 14, no. 10, Article ID e0223905, 2019.
- [14] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," vol. 2, pp. 568–573, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 568–573, IEEE, San Diego, CA, USA, June 2005.
- [15] M. D. Uyl and H. Van Kuilenburg, "The facereader: online facial expression recognition," *Proceedings of Measuring Behavior*, vol. 30, pp. 589–590, 2005.
- [16] S. L. Happy, A. George, and A. Routray, "A real time facial expression classification system using local binary patterns," in *Proceedings of the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*, pp. 1–5, IEEE, Kharagpur, India, December 2012.
- [17] C. Elliott, J. Rickel, and J. Lester, "Integrating affective computing into animated tutoring agents," in *Proceedings of the IJCAI Workshop on Animated Interface Age: Making Them Intelligent*, vol. 113121, Nagoya, Japan, August 1997.
- [18] B. Woolf, W. Bureson, and I. Arroyo, "Emotional intelligence for computer tutors," in *Proceedings of the Workshop on Modeling and Scaffolding Affective Experiences to Impact Learning at 13th International Conference on Artificial Intelligence in Education*, Los Angeles, CA, USA, July 2007.
- [19] D. Prakash, J. Van Haneghan, W. Blackwell, S. Jackson, G. Murugesan, and K. S. Tamilselvan, "Classroom engagement evaluation using computer vision techniques," in *Pattern Recognition and Tracking XXX*, M. S. Alam, Ed., vol. 10995, pp. 192–199, International Society for Optics and Photonics, Bellingham, DC, USA, 2019.
- [20] F. Karel and J. Klema, "Adaptivity in e-learning," *Current Developments in Technology-assisted Education*, pp. 260–264, 2006, http://karlovi.cz/filip/adaptivity_in_e-learning.p.
- [21] E. Popescu, "Diagnosing students' learning style in an educational hypermedia system," in *Cognitive and Emotional Processes in Web-Based Education: Integrating Human Factors and Personalization*, pp. 187–208, IGI Global, Hershey, PA, USA, 2009.
- [22] R. Oppermann and R. Rasher, "Adaptability and adaptivity in learning systems," *Knowledge Transfer*, vol. 2, pp. 173–179, 1997.
- [23] S. Stoyanov and P. Kirchner, "Expert concept mapping method for defining the characteristics of adaptive e-learning: alfanet project case," *Educational Technology Research and Development*, vol. 52, no. 2, pp. 41–54, 2004.
- [24] M. B. Mohamed Sathik and G. G. Sofia, "Identification of student comprehension using forehead wrinkles," in *Proceedings of the 2011 International Conference on Computer, Communication and Electrical Technology (ICCCET)*, pp. 66–70, Tirunelveli, India, March 2011.
- [25] T. Dragon, I. Arroyo, B. P. Woolf, W. Bureson, R. Kaliouby, and H. Eydgahi, "Viewing student affect and learning through classroom observation and physical sensors," in *Proceedings of the 9th International Conference On Intelligent Tutoring Systems, ITS '08*, pp. 29–39, Springer-Verlag, Berlin, Heidelberg, June 2008.
- [26] R. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1161, 1980.
- [27] P. Fagerberg, S. Anna, and H. Kristina, "eMoto: emotionally engaging interaction," *Personal and Ubiquitous Computing*, vol. 8, no. 5, pp. 377–381, 2004.
- [28] R. Kleine Deters, J. Naaijen, M. Rosa et al., "Executive functioning and emotion recognition in youth with oppositional defiant disorder and/or conduct disorder," *World*

- Journal of Biological Psychiatry*, vol. 21, no. 7, pp. 539–551, 2020.
- [29] B. Kort and R. Reilly, “Analytical models of emotions, learning and relationships: towards an affect-sensitive cognitive machine,” in *Proceedings of the Conference on Virtual Worlds and Simulation (VWSim 2002)*, pp. 1–15, 2002, <https://web.media.mit.edu/~reilly/its2002.pdf>.
- [30] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, “A user-independent real-time emotion recognition system for software agents in domestic environments,” *Engineering Applications of Artificial Intelligence*, vol. 20, no. 3, pp. 337–345, 2007.
- [31] R. W. Picard, E. Vyzas, and J. Healey, “Toward machine emotional intelligence: analysis of affective physiological state,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.
- [32] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, Cambridge, UK, 1990.
- [33] C. Conati and X. Zhou, “Modeling students’ emotions from cognitive appraisal in educational games,” in *Proceedings of the International Conference on Intelligent Tutoring Systems*, pp. 944–954, Springer, Biarritz, France, June 2002.
- [34] G. Katsionis and M. Virvou, “Adapting OCC theory for affect perception in educational software,” in *Proceedings of the International Conference on Human-Computer Interaction*, Berlin, Heidelberg, January 2005.
- [35] B. E. Zakka and H. Vadapalli, “Estimating student learning affect using facial emotions,” in *Proceedings of the 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pp. 1–6, IEEE, Kimberley, South Africa, November 2020.
- [36] L. Shen, M. Wang, and R. Shen, “Affective e-learning: using “emotional” data to improve learning in pervasive learning environment,” *Journal of Educational Technology & Society*, vol. 12, no. 2, pp. 176–189, 2009.
- [37] J. Bidwell and H. Fuchs, “Classroom analytics: measuring student engagement with automated gaze tracking,” *Behavior Research Methods*, vol. 49, no. 113, 2011.
- [38] R. Klein and T. Celik, “The wits intelligent teaching system: detecting student engagement during lectures using convolutional neural networks,” in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2856–2860, IEEE, Beijing, China, September 2017.
- [39] M. Sathik and S. G. Jonathan, “Effect of facial expressions on student’s comprehension recognition in virtual educational environments,” *SpringerPlus*, vol. 2, no. 1, p. 455, 2013.
- [40] U. Ayvaz, H. Gürüler, and M. Osman Devrim, “Use of facial emotion recognition in e-learning systems,” *Information Technologies and Teaching Aids*, vol. 60, pp. 95–104, 2017.
- [41] O. E. Hammoumi, F. Benmarrakchi, N. Ouherrou, J. E. Kafi, and A. E. Hore, “Emotion recognition in e-learning systems,” in *Proceedings of the 2018 6th international conference on multimedia computing and systems (ICMCS)*, pp. 1–6, IEEE, Rabat, Morocco, May 2018.
- [42] M. Pan, J. Wang, and Z. Luo, “Modelling study on learning affects for classroom teaching/learning auto-evaluation,” *Science Journal of Education*, vol. 6, no. 3, pp. 81–86, 2018.
- [43] M. Zembylas, “Adult learners’ emotions in online learning,” *Distance Education*, vol. 29, no. 1, pp. 71–87, 2008.
- [44] B. Zhou, A. Andonian, A. Oliva, and A. Torralba, “Temporal relational reasoning in videos,” 2018, <https://arxiv.org/abs/1711.08496>.
- [45] A. Pise, H. Vadapalli, and I. Sanders, “Facial emotion recognition using temporal relational network: an application to e-learning,” *Multimedia Tools and Applications*, pp. 1–21, 2020.
- [46] A. Kapoor, S. Mota, and R. W. Picard, “Towards a learning companion that recognizes affect,” in *Proceedings of the AAAI Fall Symposium*, vol. 543, pp. 2–4, Barnstable, MA, USA, November 2001.
- [47] J. Yao, “Multilayer model for on-line learning resources based on cognitive load theory,” *World Transactions on Engineering and Technology Education*, vol. 13, no. 3, pp. 245–250, 2015.
- [48] L. Pessoa, “How do emotion and motivation direct executive control?” *Trends in Cognitive Sciences*, vol. 13, no. 4, pp. 160–166, 2009.
- [49] C. J. H. Fowler, *The role of arousal in memory and attention*, PhD Thesis, University of London, London, UK, 1977.
- [50] Y.-C. Kuo, A. E. Walker, B. R. Belland, and K. E. E. Schroder, “A predictive study of student satisfaction in online education programs,” *International Review of Research in Open and Distance Learning*, vol. 14, no. 1, pp. 16–39, 2013.
- [51] H. Jabbar and D. R. Z. Khan, “Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study),” *Computer Science, Communication and Instrumentation Devices*, 2015.
- [52] A. Geron, *Hands-on Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O’Reilly Media, Inc., Newton, MA, USA, 2017.
- [53] A. A. Pise, H. Vadapalli, and I. Sanders, “Relational reasoning using neural networks: a survey,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 29, 2018.
- [54] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and 10.5 MB model size,” 2016, <https://arxiv.org/abs/1602.07360>.
- [55] B. Li, W. Wei, A. Ferreira, and S. Tan, “ReST-net: diverse activation modules and parallel subnets-based CNN for spatial image steganalysis,” *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 650–654, 2018.
- [56] A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” 2016, <https://arxiv.org/abs/1511.04110>.
- [57] A. Saravanan, G. Perichetla, and K. S. Gayathri, “Facial emotion recognition using convolutional neural networks,” 2019, <https://arxiv.org/abs/1910.05602>.
- [58] M. Ornella Treglia, “Feedback on feedback: exploring student responses to teachers’ written commentary,” *Journal of Basic Writing*, vol. 27, no. 1, pp. 105–137, 2008.