**RESEARCH ARTICLE**

# Video Traffic Analysis for Real-Time Emotion Recognition and Visualization in Online Learning

**AYOUB SASSI**[1], **WAEL JAAFAR**[2], **(Senior Member, IEEE), SAFA CHERIF**[1],
**JIHENE BEN ABDERRAZAK**[1], **AND HALIM YANIKOMEROGLU**[3], **(Fellow, IEEE)**

[1]ESPRIT School of Engineering, El Ghazala, Ariana 2083, Tunisia
[2]Department of Software and IT Engineering, École de Technologie Supérieure (ÉTS), Montreal, QC H3C 1K3, Canada
[3]Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada

Corresponding author: Wael Jaafar (wael.jaafar@etsmtl.ca)

**ABSTRACT** Since the outbreak of the COVID-19 crisis, the transition to remote education presented several challenges to educational institutions. Unlike face-to-face classes where educators can modify and keep track of the lessons and content according to the students' observed emotions and participation, such activities are difficult to complete in online learning environments. To address this issue, we propose here a novel and comprehensive framework that leverages advanced computer vision and analysis techniques to detect students' emotions during online learning and assess their state of mind regarding the taught content. Our framework is composed of three modules. The first module uses a novel lightweight machine learning method, called convolutional neural network-random forest (CNN-RF), to efficiently detect the students' basic emotions, e.g., sad, happy, etc., during the online course. Our approach surpasses existing benchmarks in terms of accuracy (over 71%) on the FER-2013 dataset, while being less complex (i.e., using a smaller number of parameters). The second module consists of mapping the basic emotions to an education-aware state of mind, e.g., interest, boredom, distraction, etc. Unlike the few works that proposed simplistic mapping, we propose here a Plutchik wheel's inspired mapping system, which is more precise and reflects better the relationship between combinations of basic emotions and the resulting education-aware state of mind. Thus, our understanding of the students' cognitive and affective experiences during online learning can be enhanced. The third module is a visualization dashboard that offers clear and intuitive real-time representations of basic emotions and states of mind. This tool provides educators with invaluable insights into students' emotional dynamics, enabling them to identify learning difficulties with high precision and make informed recommendations for improvements in course content and online teaching methods. In summary, the proposed framework presents a novel and powerful tool that addresses the challenges related to online learning. By accurately detecting the students' emotions, assessing their states of mind, and providing real-time visualization, our approach represents a significant advancement toward the optimization of online education, which is critically needed in rural and remote areas of the globe.

**INDEX TERMS** CNN, computer vision, emotion recognition, face recognition, online learning, state of mind.

## I. INTRODUCTION

During the last few years, the development and large-scale deployment of remote education around the globe has been accelerated [1]. Notably, there has been an increase in the number of virtual schools [2], and the use of virtual tutoring,

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

online learning software, and video conferencing tools [3]. This type of learning is now part of the education system, used by several schools and universities, either to reach students in remote areas or reduce large gatherings at schools for healthcare reasons.

Indeed, online learning improves time management and offers location-wise accessibility, as well as self-paced and cost-effective learning. However, it might result in social

isolation, health problems (from spending too much time in front of screens), and topic specialization restrictions [4]. Aside from this, it has been noted that comprehension during online learning is not always successful, regardless of how hard teachers work to create their remote courses' content and materials. In the past, the majority of online learning-related efforts were on pedagogy, resources, and efficiency [5]. As a result, little to no focus has been made to pinpoint the emotional stress associated with these teaching methods. In fact, it is now impossible to ignore the emotional and psychological conditions of students while teaching them remotely. Hence, there is an obvious necessity to tailor the learning process to the student's state of mind during remote classes.

In face-to-face or human tutoring, the teacher may quickly spot emotional and psychological changes in the student's facial expression or behavior thanks to human awareness and experience. Hence, the teacher can quickly react by opting for the proper educational strategy to recapture the student's attention. However, this is more complex in online learning. To detect students' emotional states, several methods have been developed. Dewan et al. classified them in [6] into three main categories, namely automatic, semi-automatic, and manual while taking into account the methods' dependencies on students' participation, as shown in Table 1. The manual engagement detection category involves the direct participation of learners in the process. A common technique in this category is self-reporting, which involves learners completing a questionnaire to indicate their level of attention, distraction, excitement, or boredom [7]. In the semi-automatic engagement detection category, learners are involved indirectly in the process. Engagement tracing is a popular method that assesses engagement based on the timing and accuracy of learner responses to practice problems and test questions [8]. The automatic engagement detection category utilizes various traits captured through sensors, such as eye movements, facial expressions, and posture, as well as physiological and neurological data, such as heart rate and electroencephalograph (EEG). It also includes tracking learners' activities in their learning environments, such as the amount of time spent on studying, forum posts, problem-solving time, and accuracy of submissions. Indeed, the human face is the perceptual evaluation that can clearly express one's internal mood and psychological condition more than gesture and posture. For this reason, facial expression is the most potent and useful visual evaluation tool for determining a student's emotional involvement [9].

On the other hand, emotions are frequently described in psychology as complicated states of feeling that cause physiological and psychological changes affecting cognition and behavior. Psychologists have acknowledged the presence of fundamental emotions, however, there is disagreement about the actual number of basic emotions. In a colored wheel, Robert Plutchik ordered the eight fundamental emotions he proposed [10]: Anger, fear, sadness, disgust, surprise, anticipation, trust, and joy, while Ekman proposed seven basic emotions: Fear, anger, joy, sadness, contempt, disgust, and surprise; but then he revoked the contempt emotion [11].

To assess the student's emotional state, several innovative methods were designed. Most of the existing systems used a screen and a learning management system (LMS) platform to monitor each student [12], which is woefully insufficient to control the dynamics of students' emotions on online learning platforms. As a result, novel approaches are required. For instance, recent emotion recognition methods relied on automation based on physiological factors recorded by wearable devices [13], spoken words from audio recordings [14], or written expressions from text [15], or facial expressions from photos and videos [16].

In the context of online learning, the use of videos during COVID-19 was found to be promising. This technique is non-intrusive, and the hardware and software it makes use of are affordable and easily achievable. This resulted in the provision of a huge amount of data that can be used for emotion recognition. For instance, Yan et al. proposed in [12] a framework that provides an intelligent 3D visualization of the classroom's atmosphere based on the valance-arousal values detected from the students' facial expressions. Authors of [9] developed a classification model that focused different types of emotions on the selected activity. For instance, emotion classes for online learning might fall under the categories of confusion, satisfaction, disappointment, or frustration. Also, Megahed et al. suggested in [17] to incorporate students' responses to questions and their basic emotional states into a methodology for modeling an intelligent adaptive online learning environment.

These works attempted to relate emotion recognition to the learning environment. However, they may lack targeted emotions design to the online learning context where emotions evolve over time, and lack explicit representation of emotions and education-aware states of mind. Motivated by the aforementioned issues, we propose, for the first time, a complete framework for real-time video traffic analysis to detect students' emotions and assess their states of mind during online learning. The framework is composed of four main parts: 1) Face recognition, 2) facial expression, a.k.a., basic emotions, recognition, 3) a mapping system between basic emotions and states of mind, and 4) a dynamic visualization dashboard. Subsequently, the contributions of the paper can be summarized as follows:

1) Unlike state-of-the-art studies that captured, with simplified approaches based on physiological or text data, the basic emotions of students at a given time, we propose here a more advanced study where video-based analysis of facial expressions is realized to identify emotions and mix them over a period of time, which is more effective in assessing the students' states of mind during online learning. Specifically, we introduce a novel lightweight convolutional neural

**TABLE 1.** Students' engagement detection methods.

| Type | Methods | Requirements | Measures |
|---|---|---|---|
| Automatic | Computer Vision | Camera | Facial Expression<br>Gesture and Posture<br>Eye Movement |
| | Sensor Data | Sensors | Heart Beat<br>Blood Pressure |
| | Log File Analysis | Documents | Time Tracking |
| Semi-automatic | Engagement Tracking | Practice Test | Test Timing and Accuracy |
| Manual | Observational Check-list | Questionnaires (Observers) | – |
| | Self Reporting | Questionnaires (Students) | – |

network-random forest (CNN-RF) method that allows emotion recognition with high accuracy.

2) From the Plutchik wheel's basic emotions classification, we propose a novel mapping system from combinations of basic emotions to deduce the education-aware states of mind, i.e., tailored to the online learning environment. This mapping system allows for a more nuanced and precise assessment of the student's emotional state of mind, e.g., boredom, interest, etc.

3) Finally, we develop a visualization dashboard that illustrates real-time statistics about every student's basic emotions and state of mind during an online class, thus allowing educators to make informed decisions about their teaching methods and course content.

The remainder of the paper is organized as follows. Section II reviews the related works. Section III describes the proposed framework. In section IV, experiments' setup and results are presented. Finally, section V concludes the paper.

## II. RELATED WORK

The task of emotion recognition in online learning scenarios has raised challenges regarding the used methodology to predict the learning states from raw data, as well as the models' suitability in the context of the learning experience. Below, we provide an overview of related works that focused on those issues, which are also summarized in Table 2.

In [12], Yan et al. provided an intelligent 3D visualization of the classroom environment based on the valence-arousal values. They used a spatial transformer network (STN) in order to enhance the spatial invariance to non-rigid deformations and other spatial transformations including translation, scaling, rotation, and cropping. In addition, at any point in the lesson, a student's status can be monitored by visualizing his/her detected emotion curve in the valence-arousal space. Shen et al. used in [18] the physiological data captured using heart rate (HR), skin conductance response (SCR), blood volume pulse (BVP), and EEG sensors, for emotion detection in online learning environments. The authors proposed a support vector machine (SVM) prediction model for the valence-arousal space, and demonstrated how emotion-aware technologies can increase the student's connection and engagement. Also, Rodríguez et al. introduced in [19] a

framework that monitors students' learning states as they watch a piece of content with a knowledge-based focus. Four learning states were specifically extracted, namely interested, bored, confused, and distracted. Their method was based on facial expression analysis and CNN. They identified a link between abrupt changes in the course and the changes in the audience's learning states, demonstrating a strong relationship between well-structured and bounded information and the students' learning behaviors. Using multimodal natural sensing, Luo et al. proposed in [20] an intelligent approach to examine students' interest in a learning environment. In their work, they provided a 3D learning interest model that takes into account cognitive attention, learning emotion, and thinking activity to fully represent the students' interests. Based on this model, multimodal data are compiled through head pose estimation, face expression detection, and interactive data gathering. Finally, multimodal data fusion is used to fully assess the students' interests. Alternatively, Gupta et al. analyzed in [21] the emotive content of a student's writing using the latter's facial expressions. Using the convolutional neural network architecture, they considered four student moods, namely high positive affect, low positive affect, high negative affect, and low negative affect, which were used to calculate course engagement scores.

Recently, data-driven techniques have introduced sophisticated models to analyze emotions in online learning environments. For instance, the use of CNN for facial emotion recognition has become a very popular approach. The authors of [17] combined two systems, namely CNN and fuzzy logic, to adapt the learning process to the students' levels. Specifically, based on the facial expression states retrieved from their CNN model and various students' response parameters, the fuzzy system selects the next learning level. Experimental findings show that the suggested approach offers adaptive learning that corresponds to the learning capacities of each student within the class. Nezami et al. proposed in [22] a CNN-based model to improve engagement recognition from photos and overcome the data sparsity barrier. First, deep learning is used to train a facial expression recognition model. Then, the trained model is used to recognize students' engagement, named the engagement model. The latter recognizes engagement and disengagement from their built dataset.

**TABLE 2.** Summary of related works.

| Reference | Online Learning Approach | Basic Emotions | State of Mind | Real-time Visualization | Dataset | Proposed Algorithm |
|---|---|---|---|---|---|---|
| [9] | Emotion Recognition State of Mind Identification | ✓ | ✓ | – | FER-2013 | CNN |
| [12] | Concentration and Engagement Level Estimation | – | ✓ | ✓ | FER-2013 | STN |
| [17] | Behavioral and Emotional Expressions Recognition | ✓ | – | ✓ | ES Dataset LA Dataset FER-2013 | CNN |
| [18] | Emotion Recognition | – | ✓ | – | Private Dataset | SVM KNN |
| [19] | Learning State Detection | – | ✓ | ✓ | LS Dataset Affectnet CK+ DAiSEE | CNN |
| [20] | Interest Analysis | ✓ | – | ✓ | CCNU Dataset LFW Dataset Pointing'04 Dataset | Multimodal Data Fusion |
| [21] | Affective Content Analysis | ✓ | – | ✓ | FDDB YALE | CNN |
| [22] | Engagement Recognition | ✓ | ✓ | – | FER-2013 ER Dataset | CNN |
| [23] | Facial Expression Recognition | ✓ | – | – | FER-2013 | CNN-LSVM |
| [24] | Facial Expression Recognition | ✓ | – | – | FER-2013 | CNN |
| [25] | Facial Expression Recognition | ✓ | ✓ | – | CK+ DAiSEE JAFFE | CNN-SVM |
| [26] | Emotion Recognition | ✓ | – | – | FER-2013 | CNN |
| [27] | Facial Expression Recognition | ✓ | – | – | FER-2013 | VGGNet |
| **Our Work** | **Basic Emotion Recognition State of Mind Identification** | ✓ | ✓ | ✓ | **FER-2013** | **CNN-RF** |

Rao et al. introduced in [25] a hybrid-CNN model that uses both manually created and convolutional neural network-extracted information to identify a student's cognitive state from its facial expressions. Their approach was tested on different datasets, namely the Japanese female facial expression (JAFFE) dataset, the extended Cohn-Kanade dataset (CK+) dataset, and the spontaneous dataset (DAiSEE). Results prove the superior accuracy of their method compared to the CNN-based and manual feature extraction methods. Similarly, the authors of [26] proposed a field programmable gate array (FPGA) architecture that uses a trained CNN on the facial emotion recognition (FER)-2013 dataset [28]. Their model achieved an accuracy of 60.4%. Tang [23] introduced a CNN model where they tried to replace the softmax layer with a linear support vector machine. They demonstrated that their approach gives significant gains as they achieved an accuracy of 69.3% on FER-2013. Also, Liu et al. [27] proposed a model consisting of several structured subnets where each subnet is a CNN model trained apart. Their greatest single subnet achieved 62.44% and their whole model achieved 65.03% of accuracy. Finally, Agrawal et al. improved in [24] the results of [26] by fine-tuning the CNN parameters, such as kernel size and number of filters, thus achieving a higher accuracy of 65%.

## III. PROPOSED FRAMEWORK
Our framework consists of four modules: *Face recognition*, *basic emotion recognition*, *basic emotions-state of mind*

*mapping system*, and *visualization dashboard*, explained below and illustrated in Fig. 1.

### A. FACE RECOGNITION MODULE
The *face recognition* module is responsible for detecting faces on the screen and recognizing the students' identities. The identification process flow consists of a number of automatic steps:
1) Activate cameras in video streaming mode. Then, using the open-source computer vision (OpenCV) library, do the following for each captured video image:
2) Detect the student's facial characteristics
3) Create a grayscale image using the face traits
4) Scale down the final image to the proper dimensions
5) Identify the student by comparing the final image to those stored in a class database.

Subsequently, the *face recognition* module outputs the final image, to be used as input into the *basic emotion recognition* module.

### B. BASIC EMOTION RECOGNITION MODULE
The module's main function is to classify the student's facial expressions as basic emotions. It consists of a CNN-based model that outputs, for the evaluated image, the probability of occurrence of seven basic emotions classes namely, anger, disgust, fear, happy, neutral, sad, and surprise, abbreviated respectively by (A), (D), (F), (H), (N), (S), and (R). Our basic emotion recognition model is constructed using a
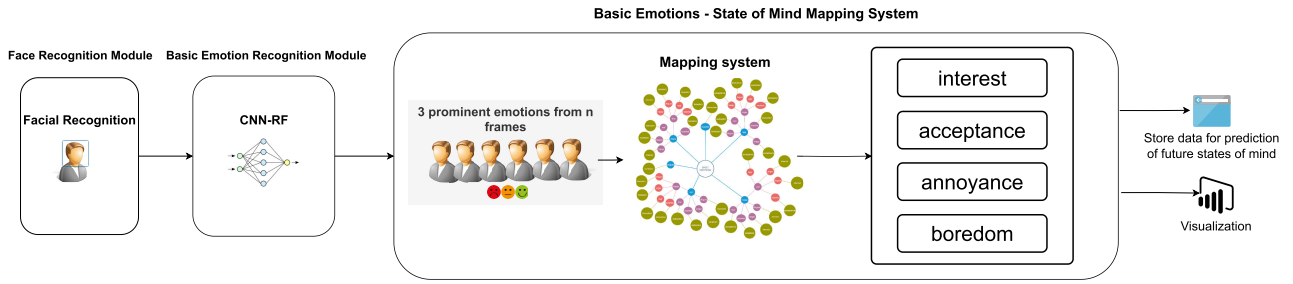
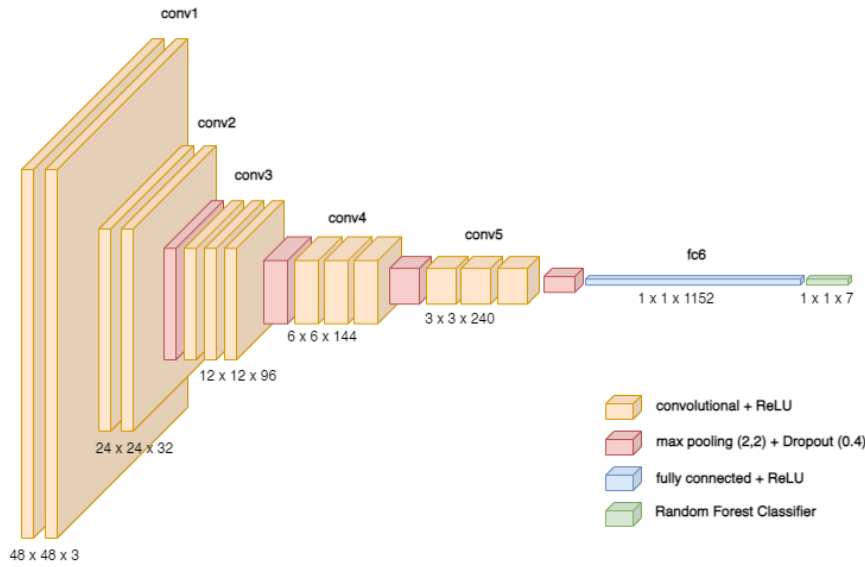**FIGURE 1.** Proposed emotion recognition framework.
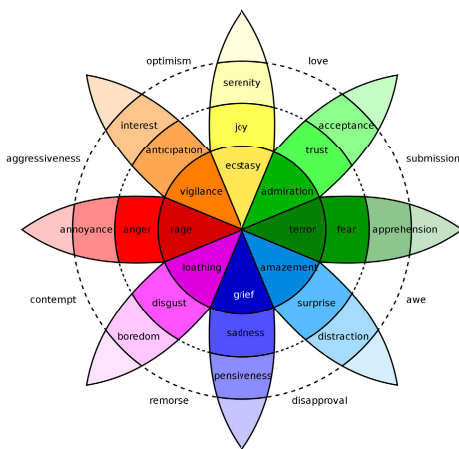


**FIGURE 2.** Proposed CNN-RF model architecture.



**FIGURE 3.** Three-level based Plutchik wheel of emotions.

It has been noted that during the learning process, a student's mood or emotion does not alter immediately but rather gradually [9]. Moreover, the reflection of the emotion on the facial expression is not abrupt and may take some time to be emphasized. Hence, it is very difficult and inaccurate to determine a student's emotion by analyzing a single image of his/her facial expression. For a more rigorous outcome, a series of images taken over time should be analyzed. Specifically, let $N$ be the number of images sequentially captured in time to analyze the student's emotional state. Since authors of [9] have determined that a human emotion transition may last about 6 seconds, we assume that $N = 6$, where images are captured each second. For image $i \in \{1, \ldots, N\}$, let $P_{i,j}$ be the probability of occurrence of emotion $j \in \{A, D, F, H, N, S, R\}$ in image $i$. Thus, the average probability of occurrence of emotion $j$ for the analyzed $N$ images can be written as

$$P_j = \frac{1}{N} \sum_{i=1}^{N} P_{i,j}, \quad \forall j \in \{A, D, F, H, N, S, R\}. \quad (1)$$

combination of a CNN and an RF classifier. Its architecture is presented in Fig. 2, where it consists of a series of convolutional and pooling layers and an RF classifier.
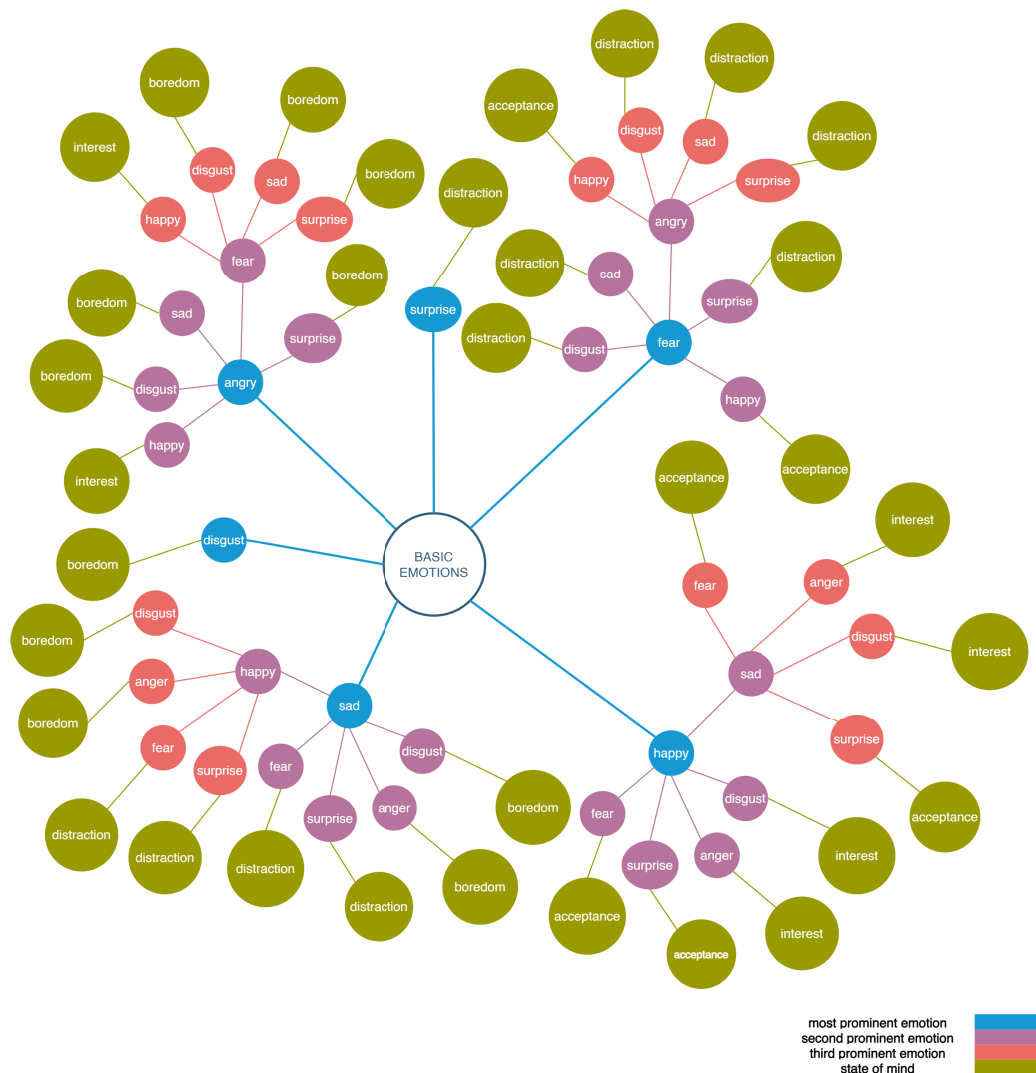
**FIGURE 4.** Designed basic emotions - state of mind mapping system.

Let $P_E = \{P_A, P_D, P_F, P_H, P_N, P_S, P_R\}$ be the set of calculated averaged probabilities for the considered basic emotions, in descending order. Then, the first three emotions are considered the prominent ones and thus will constitute the input for the *mapping system*.

### C. BASIC EMOTIONS-STATE OF MIND MAPPING SYSTEM

Humans often orchestrate a variety of emotions, either at once or in sequence, rather than expressing their feelings openly through a single distinct emotion. According to Plutchik's theory [10], these emotions work together to create a state of mind that is frequently decided by combining the basic emotions. Knowing how a student would feel about an online course may be determined by evaluating these states of mind. In order to determine the student's state of mind, we map here the three prominent emotions to the third level of Plutchik's wheel of emotions (i.e., third outer circle), illustrated in Fig. 3. Based on the general relevance of the e-learning

environments, we have chosen four types of states of mind, namely, interest, acceptance, distraction, and boredom. These states of mind are equidistant emotions located on the third level (i.e., separated by the same number of other emotions on the third outer circle) of Plutchick's wheel.

Considered basic emotions are located on the second level of Plutchick's wheel, however, Neutral is not seen as an emotion,[1] hence it is ignored in the mapping to states of mind. On the second level of the wheel, prominent basic emotions disgust (D) and surprise (R) lead inevitably to complex emotions boredom, and distraction, respectively, regardless of the second and third prominent basic emotions. This is explained by the fact that the positions of D and R on the second level of the wheel are surrounded by other negative emotions that won't affect the outcome. However, the positions of the other considered basic emotions, i.e., $\{A, F, H, S\}$, suggest further investigation into the second

---

[1]Neutral is more likely to be a facial expression.

**TABLE 3.** FER-2013 dataset basic emotions class distribution [28].

| Class | Nbr. images |
|---|---|
| Anger (A) | 4953 |
| Disgust (D) | 547 |
| Fear (F) | 5121 |
| Happy (H) | 8989 |
| Neutral (N) | 6198 |
| Sad (S) | 6077 |
| Surprise (R) | 4002 |



**FIGURE 5.** Basic emotions recognition.



**FIGURE 6.** State of minds recognition.

and even third prominent emotion, to define the related state of mind. For instance, if the most prominent emotion is H, having the second prominent emotion as R leads to acceptance while having the second prominent emotion as S, suggests checking the third prominent emotion, as shown in Fig. 4. In total, there are thirty-six combinations to map the selected basic emotions to the state of mind.

### D. VISUALIZATION MODULE

The role of this module is to provide a real-time reference for teachers, to monitor the level of students' interest, acceptance, annoyance, and boredom during the online course, and to evaluate their interaction with the course's content. To provide an efficient emotions visualization and understand the classroom's atmosphere intuitively, both basic and complex emotions should be visualized in real-time.

## IV. EXPERIMENTS AND RESULTS
### A. EMOTION RECOGNITION SETUP

For the basic emotion recognition model training, we select the FER-2013 dataset, which is an open dataset in Kaggle containing images of faces expressing different emotions, as described in Table 3 [28]. This dataset is not related to

online learning, but it is suitable for the aimed environment and has been used in several other relevant works [12], [17], [22], [23], [24]. FER-2013 has 35,887 grayscale $48 \times 48$-pixel photos stored in a spreadsheet with the pixel values of each image in row cells. Faces in photos have been automatically saved such that they are centered and occupy approximately the same space in each frame. Photos are labeled with the basic emotions $\{A, D, F, H, N, S, R\}$, and the corresponding class distribution is presented in Table 3. For the needs of our experiments, we respect the predefined division of the FER-2013 dataset into 80% for training, 10% for validation, and 10% for testing. Finally, to enrich our training dataset, we opted for data augmentation by means of horizontal mirroring, rotation by $\pm10°$, image zoom by $\pm10\%$, and horizontal/vertical shifting by $\pm10\%$.

For the designed CNN-RF emotion recognition model, we get inspired by the hyperparameters setup of Vulpe-Grigoraşi et al. [29]. Thus, the used hyperparameters are set as follows: The number of epochs is 310, the batch size is 50, and the learning rate is set to 0.001. These values have been selected after extensive experimentation and evaluation to optimize the model's performance for our specific task.

**TABLE 4.** Performances of emotion recognition models.

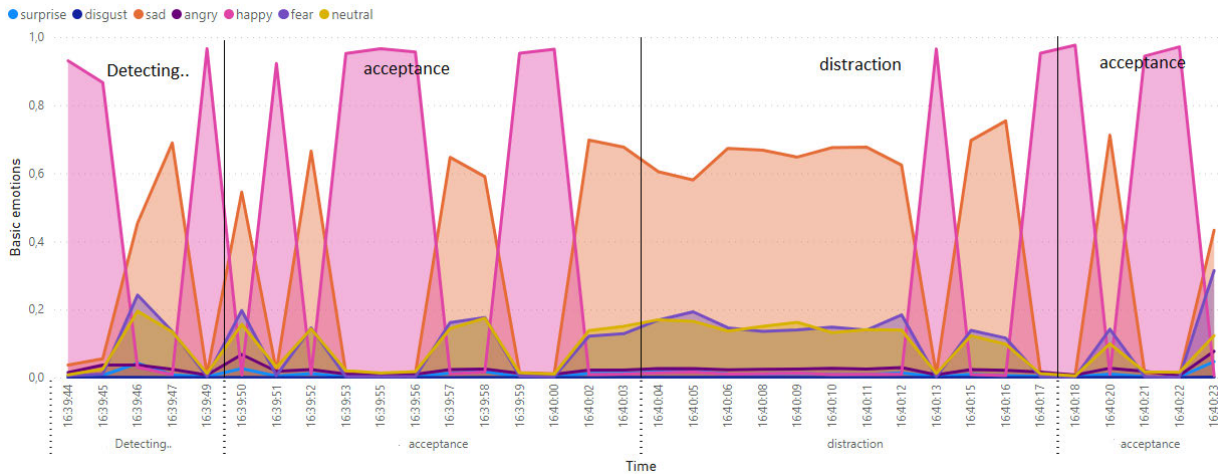| Reference | Accuracy $\alpha$ | Precision $\phi$ | Recall $\rho$ | F1-score $\delta$ | # Parameters | Online learning | Baseline architecture |
|---|---|---|---|---|---|---|---|
| [9] | 65% | 69% | 43% | – | – | ✓ | CNN |
| [17] | 67.15% | 69.13% | 63.92% | 65.74% | – | ✓ | CNN |
| [23] | 69.3% | – | – | – | 7.17M | – | CNN-LSVM |
| [24] | 65.23% | 64% | 61% | 62.46% | 0.46M | – | CNN |
| [26] | 60.4% | 60.71% | 59.86% | 59.8% | – | – | CNN |
| [27] | 61.74% | – | – | – | 84M | – | VGGNet |
| **Our Work** | **71.86%** | **70.56%** | **74.35%** | **72.09%** | **5.17M** | ✓ | **CNN-RF** |



**FIGURE 7.** Evolution of basic emotions through time (Omar).

## B. EXPERIMENTAL RESULTS

### 1) BASIC EMOTION RECOGNITION

Following the training of our proposed CNN-RF model, we evaluated its performances in terms of accuracy $\alpha$, precision $\phi$, recall $\rho$, and F1-score $\delta$, defined respectively by

$$\alpha = \frac{TP + TN}{TP + TN + FP + FN}, \qquad (2)$$

$$\phi = \frac{TP}{TP + TN}, \qquad (3)$$

$$\rho = \frac{TP}{TP + FN}, \qquad (4)$$

and

$$\delta = \frac{2\ TP}{2\ TP + FP + FN}, \qquad (5)$$

where TP, TN, FP, and FN refer to true-positive, true-negative, false-positive, and false-negative outcomes.

As shown in Table 4, our model achieved $\alpha = 71.86\%$, $\phi = 70.56\%$, $\rho = 74.35\%$ and $\delta = 72.09\%$, which are respectively higher performances than those of benchmarks. The high accuracy rate means that the model correctly classified the samples, while the high precision indicates that the model has a low rate of false positive prediction, i.e., less likely to classify a negative sample as positive. The high recall ensures that the model is less likely to miss true positives, and the F1-score summarizes the previous metrics into a single metric, which reflects the overall performance of the model. As evident from the results, our proposed approach outperforms existing models in terms of F1-score. This superior performance is achieved through a streamlined model architecture that not only enhances accuracy but also ensures lightweight operation, with a mere 5.17 million parameters. By efficiently managing the number of model parameters, our approach significantly reduces complexity. This attribute positions it as an attractive choice for real-world applications, particularly in resource-constrained environments.

For the sake of illustration, we visualize in Fig. 5 the occurrence probabilities of the basic emotions $\{P_A, \ldots, P_R\}$ every $N = 6$ seconds (frames from top to bottom) within a video sequence [30]. As shown, the list of basic emotions is displayed below the student's name with its probability values and is updated over time. This dynamic representation aids in deciphering the evolving basic emotions of students as they progress through the course. For instance, the right subject shifts from being predominantly happy to a fearful state.

### 2) BASIC EMOTIONS-STATE OF MIND MAPPING AND VISUALIZATION

From the output of the CNN-RF module, basic emotions are mapped to states of mind according to the mapping system explained in Section III. The result can be displayed in real-time using our visualization dashboard. For instance, Fig. 6 displays the detected state of minds in a video stream.
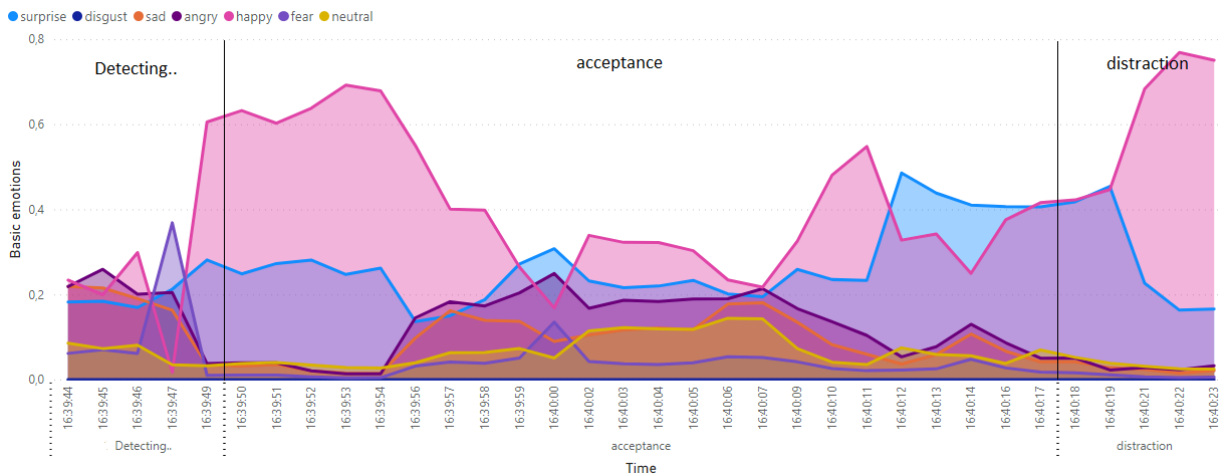
**FIGURE 8.** Evolution of basic emotions through time (Nolan).

The state of mind, i.e., interest, acceptance, annoyance, or boredom, is displayed within the face recognition red rectangle and is updated every $N = 6$ seconds (frames from top to bottom).

*Remark 1: Due to the difficulty encountered to obtain adequate videos of online classrooms for testing, we opted for a video of 13 minutes found on Youtube [30]. The latter features two online streamers, who are interacting and showing emotions in a manner that resembles an online classroom environment. We acknowledge that this video source is not ideal and may not fully represent the dynamics of a real online classroom. However, we believe that it provides a reasonable approximation of the emotional interactions that take place in such an environment, and thus serves as a useful tool to illustrate the results of our basic emotions - state of mind mapping system.*

In addition to the real-time illustration of basic emotions and states of mind, the collected data can be stored and leveraged to predict the future states of mind of students using machine learning algorithms [31], and visualized in an evolving graph that takes time in the X-axis and basic emotions' occurrence probabilities in the Y-axis, as presented in Figs. 7-8. This type of visualization allows superimposing the reactions of different students during online learning, identifying emotional trends, and establishing correlations between students' emotions (and states of mind) during particular times of the learning process.

## V. CONCLUSION

In this paper, we proposed a complete framework to analyze in real-time students' emotions and states of mind during online learning. The proposed framework enhances the online learning experience in different ways. First, the utilization of the novel CNN-RF model for students' emotion recognition demonstrates higher performances and lower complexity than existing methods. Second, the proposed mapping system of basic emotions to states of mind, inspired

by the Plutchik wheel, adds depth to the assessment of students' cognitive and affective reactions during online lessons. Third, the inclusion of the visualization dashboard simplifies the interaction and analysis of the collected emotional data. This novel real-time monitoring tool would allow educators to clearly associate students' emotions and states of mind with phases of the online learning process, thus identifying lessons' difficulties, which would suggest adaptation of their teaching strategies and provision of timely interventions to create a supportive and engaging learning environment. Nevertheless, our framework has some limitations. For instance, it would be inefficient in detecting emotions in subtle or discreet facial expressions. Moreover, external factors such as camera setup, lighting conditions, and the quality of video capture equipment, may influence the accuracy of emotion recognition in our system. Also, it is difficult to evaluate the accuracy of our basic emotions-state of mind mapping system since the final results have not been validated by a psychologist or through real experiments using different questionnaires for students. The latter will be tackled in future work. Specifically, we will target expanding the scope of our framework to make it efficient for diverse student populations, cultural backgrounds, and learning contexts. Finally, the use of lightweight methods for emotion recognition in masked faces and optimization algorithms [32], [33] will be investigated to further improve the emotion recognition and mapping performances of our framework.

## REFERENCES

[1] R. L. Quezada, C. Talbot, and K. B. Quezada-Parker, "From bricks and mortar to remote teaching: A teacher education program's response to COVID-19," *J. Educ. Teach.*, vol. 46, no. 4, pp. 472–483, Aug. 2020.

[2] G. Theonas, D. Hobbs, and D. Rigas, "The effect of facial expressions on students in virtual educational environments," *Int. J. Educ. Pedagogical Sci.*, vol. 1, no. 11, pp. 625–632, Nov. 2007.

[3] A. R. Artino, "Emotions in online learning environments: Introduction to the special issue," *Internet Higher Educ.*, vol. 15, no. 3, pp. 137–140, Jun. 2012.

[4] M. Sadeghi, "A shift from classroom to distance learning: Advantages and limitations," *Int. J. Res. English Educ.*, vol. 4, no. 1, pp. 80–88, Mar. 2019.

[5] G. Salmon, *E-Moderating: The Key to Online Teaching and Learning*, 3rd ed. Evanston, IL, USA: Routledge, Jul. 2011.

[6] M. A. A. Dewan, M. Murshed, and F. Lin, "Engagement detection in online learning: A review," *Smart Learn. Environ.*, vol. 6, no. 1, pp. 1–20, Jan. 2019.

[7] H. L. O'Brien and E. G. Toms, "The development and evaluation of a survey to measure user engagement," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 61, no. 1, pp. 50–69, Jan. 2010.

[8] J. E. Beck, "Engagement tracing: Using response times to model student disengagement," in *Proc. Conf. AI Educ.*, 2005, pp. 88–95.

[9] M. Mukhopadhyay, S. Pal, A. Nayyar, P. K. D. Pramanik, N. Dasgupta, and P. Choudhury, "Facial emotion detection to assess learner's state of mind in an online learning system," in *Proc. 5th Int. Conf. Intell. Inf. Technol. (ICIIT)*, 2020, pp. 107–115.

[10] R. Plutchik and H. Kellerman, *EMOTION: Theory, Research, and Experience*. New York, NY, USA: Academic, 1980.

[11] P. Ekman, "An argument for basic emotions," *Cognition Emotion*, vol. 6, nos. 3–4, pp. 169–200, May 1992.

[12] F. Yan, N. Wu, A. Iliyasu, K. Kawamoto, and K. Hirota, "Framework for identifying and visualising emotional atmosphere in online learning environments in the COVID-19 era," *Appl. Intell.*, vol. 52, pp. 1–17, Jun. 2022.

[13] S. Saganowski, A. Dutkowiak, A. Dziadek, M. Dziezyc, J. Komoszynska, W. Michalska, A. Polak, M. Ujma, and P. Kazienko, "Emotion recognition using wearables: A systematic literature review–work-in-progress," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–6.

[14] C. S. Ooi, K. P. Seng, L.-M. Ang, and L. W. Chew, "A new approach of audio emotion recognition," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 5858–5869, Oct. 2014.

[15] Z. Erenel, O. R. Adegboye, and H. Kusetogullari, "A new feature selection scheme for emotion recognition from text," *Appl. Sci.*, vol. 10, no. 15, p. 5351, Aug. 2020.

[16] A. R. Dores, F. Barbosa, C. Queirós, I. P. Carvalho, and M. D. Griffiths, "Recognizing emotions through facial expressions: A largescale experimental study," *Int. J. Environ. Res. Public Health*, vol. 17, no. 20, p. 7420, Oct. 2020.

[17] M. Megahed and A. Mohammed, "Modeling adaptive E-learning environment using facial expressions and fuzzy logic," *Expert Syst. Appl.*, vol. 157, Nov. 2020, Art. no. 113460.

[18] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using 'emotional' data to improve learning in pervasive learning environment," *Educ. Technol. Soc.*, vol. 12, pp. 176–189, Apr. 2009.

[19] J. Rodríguez, J. Comas, and X. Binefa, "Affective state-based framework for e-learning systems," in *Proc. 23rd Int. Conf. Catalan Assoc. Artif. Intell.*, Oct. 2021, p. 357.

[20] Z. Luo, C. Jingying, W. Guangshuai, and L. Mengyi, "A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology," *Interact. Learn. Environ.*, vol. 30, no. 6, pp. 1–14, Jan. 2020.

[21] S. K. Gupta, T. S. Ashwin, and R. M. R. Guddeti, "Students' affective content analysis in smart classroom environment using deep learning techniques," *Multimedia Tools Appl.*, vol. 78, no. 18, pp. 25321–25348, Sep. 2019.

[22] O. M. Nezami, M. Dras, L. Hamey, D. Richards, S. Wan, and C. Paris, "Automatic recognition of student engagement using deep learning and facial expression," in *Machine Learning and Knowledge Discovery in Databases* (Lecture Notes in Computer Science), vol. 11908, U. Brefeld, E. Fromont, A. Hotho, A. Knobbe, M. Maathuis, and C. Robardet, Eds. Cham, Switzerland: Springer, Apr. 2020, pp. 273–289.

[23] Y. Tang, "Deep learning using linear support vector machines," Jun. 2013, *arXiv:1306.0239*.

[24] A. Agrawal and N. Mittal, "Using CNN for facial expression recognition: A study of the effects of kernel size and number of filters on accuracy," *Vis. Comput.*, vol. 36, no. 2, pp. 405–412, Jan. 2019.

[25] K. P. Rao, M. Chandra, and S. Rao, "Recognition of learners' cognitive states using facial expressions in e-learning environments," *J. Univ. Shanghai Sci. Technol.*, vol. 22, no. 12, pp. 93–103, Dec. 2020.

[26] H. Phan-Xuan, T. Le-Tien, and S. Nguyen-Tan, "FPGA platform applied for facial expression recognition system using convolutional neural networks," *Proc. Comput. Sci.*, vol. 151, pp. 651–658, Jan. 2019.

[27] K. Liu, M. Zhang, and Z. Pan, "Facial expression recognition with CNN ensemble," in *Proc. Int. Conf. Cyberworlds (CW)*, Sep. 2016, pp. 163–166.

[28] I. J. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," in *Neural Information Processing*. Berlin, Germany: Springer, 2013, pp. 117–124.

[29] A. Vulpe-Grigorasi and O. Grigore, "Convolutional neural network hyperparameters optimization for facial emotion recognition," in *Proc. 12th Int. Symp. Adv. Topics Electr. Eng. (ATEE)*, Mar. 2021, pp. 1–5.

[30] I. Lütkebohle. *How to Record Long-Distance Podcast Interviews*, Accessed: Jun. 14, 2023. [Online]. Available: https://www.youtube.com/watch?v=poPNtONw_Kk&ab_channel=Buzzsprout%E2%80%94LearnHowtoPodcast

[31] M. Frank, D. Drikakis, and V. Charissis, "Machine-learning methods for computational science and engineering," *Computation*, vol. 8, no. 1, p. 15, Mar. 2020.

[32] Y. Yuan, J. Ren, S. Wang, Z. Wang, X. Mu, and W. Zhao, "Alpine skiing optimization: A new bio-inspired optimization algorithm," *Adv. Eng. Softw.*, vol. 170, Aug. 2022, Art. no. 103158.

[33] Y. Yuan, Q. Shen, S. Wang, J. Ren, D. Yang, Q. Yang, J. Fan, and X. Mu, "Coronavirus mask protection algorithm: A new bio-inspired optimization algorithm and its applications," *J. Bionic Eng.*, vol. 20, no. 4, pp. 1747–1765, Mar. 2023.

**AYOUB SASSI** received the master's degree from the ESPRIT Engineering School. He was an AR/VR Research Intern with the Internet Technology and Data Science Laboratory (IDLab), IMEC, Ghent University, Ghent, Belgium, from July 2021 to September 2021. He was also a Research Intern with the Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada, from January 2022 to October 2022. He is currently an AI Engineer with the ESPRIT Engineering School. His current research interests include AI, ML, computer vision, and AR/VR technologies.

**WAEL JAAFAR** (Senior Member, IEEE) received the B.Eng. degree from the Higher School of Communication (SUPCOM), Tunis, Tunisia, in 2007, and the M.A.Sc. and Ph.D. degrees in electrical engineering from Polytechnique Montreal, QC, Canada, in 2009 and 2014, respectively. He is currently an Associate Professor with the Software and IT Engineering Department, École de Technologie Supérieure (ÉTS), Montreal, QC. He was a Visiting Research Intern with the Department of Computer Science, UQAM, Montreal, Canada, from February 2007 to April 2007. He was also a Visiting Researcher with Keio University, Tokyo, Japan, from March 2013 to June 2013; and Khalifa University, Abu Dhabi, United Arab Emirates, from November 2019 to December 2019. From 2014 to 2017, he has pursued a career in telecommunications industry. In 2018, he was a Research Fellow and a Lecturer with the Computer Science Department, UQAM. From December 2018 to August 2022, he occupied an NSERC Postdoctoral Fellowship position with the Systems and Computer Engineering Department, Carleton University. His current research interests include wireless communications, resource allocation, non-terrestrial networks, edge caching and computing, and machine learning.

**SAFA CHERIF** received the B.Eng. degree in telecommunications from the Higher School of Communication of Tunis (SUP'COM), in 2012, and the Ph.D. degree in information and communication technologies in science from IMT Atlantique, France, in 2018. She is a Teaching Assistant with the ESPRIT Engineering School, Tunisia. Her main research interests include signal and image processing, with a special interest in artificial intelligence applications for e-health.

**JIHENE BEN ABDERRAZAK** received the B.Eng. degree from the Higher School of Communication (SUPCOM), Tunis, Tunisia, in 2006, the M.A.Sc. degree in computer science from the University of Tsukuba, Japan, in 2010, and the Ph.D. degree in information and communication technology from SUPCOM, in 2018. She is a Research Assistant with the ESPRIT Engineering School, Tunis, where she is leading the Wireless Communications Research Team. Her current research interests include future wireless networks, smart agriculture, autonomous vehicles, and machine learning.

**HALIM YANIKOMEROGLU** (Fellow, IEEE) is currently a Chancellor's Professor with the Department of Systems and Computer Engineering, Carleton University, Canada. He has extensive collaboration with industry, which resulted in 39 granted patents. He has given more than 150 keynotes, tutorials, and invited seminars in the last ten years. His research group has made substantial contributions to 4G/5G wireless technologies. His group's current focus is the wireless infrastructure for the 6G and B6G era with terrestrial, aerial (HAPS and UAV), and satellite network elements. His research interests include wireless communications and networks.

He is a member of the IEEE ComSoc Conference Council and the IEEE PIMRC Steering Committee and a fellow of the Engineering Institute of Canada (EIC) and the Canadian Academy of Engineering (CAE). He received several awards for his research, teaching, and service, including the IEEE ComSoc Fblue W. Ellersick Prize, in 2021; the IEEE VTS Stuart Meyer Memorial Award, in 2020; and the IEEE ComSoc Wireless Communications TC Recognition Award, in 2018. He also received the Best Paper Awards from IEEE ICC 2021 and IEEE WISEE 2021. He is serving as the Steering Committee Chair for the IEEE Wireless Communications and Networking Conference (WCNC). He has served as the general chair and the technical program chair for several IEEE conferences. He has also served in the editorial boards for various IEEE periodicals. He is an IEEE Distinguished Speaker of ComSoc and VTS.

• • •