

Determine the Level of Concentration of Students in Real Time from their Facial Expressions

Bouhlal Meriem*, Habib Benlahmar, Mohamed Amine Naji, Elfilali Sanaa, Kaiss Wijdane

Laboratory of Information Technology and Modeling
Hassan II University of Casablanca, Faculty of Sciences, Casablanca, Morocco

Abstract—In teaching environments, student facial expressions are a clue to the traditional classroom teacher in gauging students' level of concentration in the course. With the rapid development of information technology, e-learning will take off because students can learn anytime, anywhere and anytime they feel comfortable. And this gives the possibility of self-learning. Analyzing student concentration can help improve the learning process. When the student is working alone on a computer in an e-learning environment, this task is particularly challenging to accomplish. Due to the distance between the teacher and the students, face-to-face communication is not possible in an e-learning environment. It is proposed in this article to use transfer learning and data augmentation techniques to determine the concentration level of learners from their facial expressions in real time. We found that expressed emotions correlate with students' concentration, and we designed three distinct levels of concentration (highly concentrated, nominally concentrated, and not at all concentrated).

Keywords—*Emotion recognition; level of concentration; transfer learning; data augmentation*

I. INTRODUCTION

In recent years, E-learning is very popular because this type of learning system uses modern educational technologies to implement an ideal learning environment by integrating information technology into the program.

E-learning is gaining prominence in universities, colleges, and industries by examining its advantages over traditional approaches, as students can access all the data they need for their research. Through webinars, they can access information they might not otherwise be able to access in person due to finances, distances, or time constraints. Depending on the level of understanding of the student, they can study at their own pace, which may increase their satisfaction with the course and reduce their stress levels. It also helps students with special needs or having difficulty getting to school for one reason or another, such as illness or unforeseen accidents.

In traditional classroom teaching, the experienced teacher can know the level of concentration by the students from their behaviors which can be studied by heart rate, pose, gesture, gaze, height of the voice and facial expressions and, accordingly, it can adjust the pace, improve the educational system and provide content. While the online environment separates teachers from students and students from students, there is a lack of face-to-face communication to understand students' emotions and cognitive state.

An effective individualized learning system must therefore be not only intelligent but also emotional. Researchers in neuroscience and psychology have found that emotions are largely related to cognition / concentration. These eLearning system models place particular emphasis on assessing learner emotional states and adjusting teaching strategies.

Among the challenges teachers face is examining how learners acquire course content, as noted in [1]. In order to improve the educational system, it is of paramount importance to ensure student concentration through participation in the learning environment.

Virtual classrooms were introduced as early as the mid-1990s [2]. Deconcentration of students is an issue that is addressed daily.

The way that students are taught is also a factor behind student concentration. Bradbury [3] reported that between 25% and 60% of students became bored in the classroom and lost concentration for a long time [4]. According to Ekman, Friesen, and Ellsworth, [5], facial expression can be the fastest way to understand an individual's emotions. You can use a student's emotional state during their learning period (in the classroom or another setting) to determine if they are paying attention to the content.

To identify students' concentration, other authors suggest using pupil dilation, which occurs when students view images of emotional arousal [6], or the length of time the eyes are closed [1]. Students' facial expressions can be captured by using the embedded webcam in their laptop computers in a typical e-learning environment. The teacher can use this information to determine the level of concentration of the students, and measure how focused they are (or aren't). The teacher can use this information to make the learning environment cost-effective.

The concentration of learners is influenced by several factors. The emotional life of learners has a profound impact on academic success, learning techniques, and academic achievement [7]. If emotions could be recognized as impacts on motivation and concentration, educational outcomes could be improved [8]. Emotion in this context corresponds to the psychological and physiological characteristics of a being, which are individual, efficient, and personal in nature, which relates to habits, manners, thoughts, and sensations [9]. It has been shown that facial muscles move in relation to different emotions, such as happiness, sadness, anger, fear, surprise, and disgust [10].

*Corresponding Author.

Most facial expression recognition approaches are based on a posed expression database, such as the Japanese Female Facial Expression Database (JAFFE), Cohn-Kanade Database (CK) which are built on six Universal emotions such as happiness, sadness, surprise, anger, disgust and fear due to the lack of database of facial expressions in educational environments.

The proposed system automatically identifies the emotional state of the learner based on facial expressions to know the level of concentration. This helps the teacher to identify slow learners who have difficulty understanding the lesson, therefore the lesson may be changed.

Students' engagement and concentration during learning is a prerequisite for successful learning effects and is positively correlated with students' academic success and development of higher-level abilities (Pascarella, Seifert, & Blauch, 2010). Effective detection of students' learning situations can provide information to instructors so that they can identify struggling students in real time. And this is so that a student's level of concentration and state of learning engagement can help intelligent tutoring systems provide students with individualized learning resources.

The purpose of this article is to examine the concentration level of students in a typical e-learning scenario, by analyzing their facial emotions in real time. In our analysis, we attempt to define how emotions affect concentration level and to devise a concentration index based on this. A facial emotion index is generated in real time by Python and the Keras algorithm, which is derived from the Haar-cascade algorithm and we will compare the transfer learning (a pre-trained convolutional neural network (CNN): VGG16, VGG19, XCEPTION, ALEXNET) models to determine the best performing model for our proposal.

This paper is structured as follows: Section 2 presents a review of related work; Section 3 discusses the methodology adopted for facial emotion recognition. The implementation environment, datasets, as well as the algorithms used, the experiment performed and the results are presented in Section 4. We concluded the paper by summarizing the research performed and providing some remarks in Section 5.

II. RELATED WORK

Currently available e-learning systems have a flaw: they do not monitor student concentration. Recent years have seen increased interest in finding clues to determine student concentration. Due to the absence of a teacher and the inability to grasp emotions, emotions, etc., are particularly important for students using standalone e-learning systems. According to the study conducted by Du, Tao, and Martinez [11], there are 22 different types of emotions: seven basic emotions and twelve compound emotions. There are forty-six fundamental muscles that form the UA of the face, including those that produce facial expressions. Based on individual UA, the system classifies facial categories by combining each individual UA (Fig. 1). In the case of AU12 and 25, if the system recognizes that the image indicates a "happy" emotion (Table I), the system will classify the image accordingly (Table I).

Bidwell and Fuchs [13] used an automated gaze system to measure student engagement. Classifiers were created based on video recordings of classrooms. A face tracking system was used to track student gaze. After the automated gaze model and observations of a panel of experts were combined, a Hidden Markov Model (HMM) was constructed. HMM incorrectly categorized the data, suggesting eight discrete categories of behaviors, but they were only able to determine whether students were "engaged" or "not engaged".

TABLE I. THE UNITED STATES OBSERVED UAS IN BOTH BASIC AND COMPOUND EMOTION CATEGORIES [12]

Category	AU	Category	AU
Satisfied	12,25	Sadly, disgusted	4,10
Unhappy	4,15	Fearfully angry	4,20,25
Frightful	1,4,20,25	Fearfully surprised	1,2,5,20,25
Annoyed	4,7,24	Fearfully disgusted	1,4,10,20,25
Amazed	1,2,25,26	Angrily disgusted	4,25,26
Disgusted	9,10,17	Disgusted surprised	1,2,5,10
Happily sad	4,6,12,25	Happily fearful	1,2,12,25,26
Happily surprised	1,2,12,25	Angrily disgusted	4,10,17
Happily disgusted	10,12,25	Awed	1,2,5,25
Sadly fearful	1,4,15,25	Appalled	4,9, 10
Sadly angry	4,7,15	Hatred	4,7,10
Sadly surprised	1,4,25,26	-	-

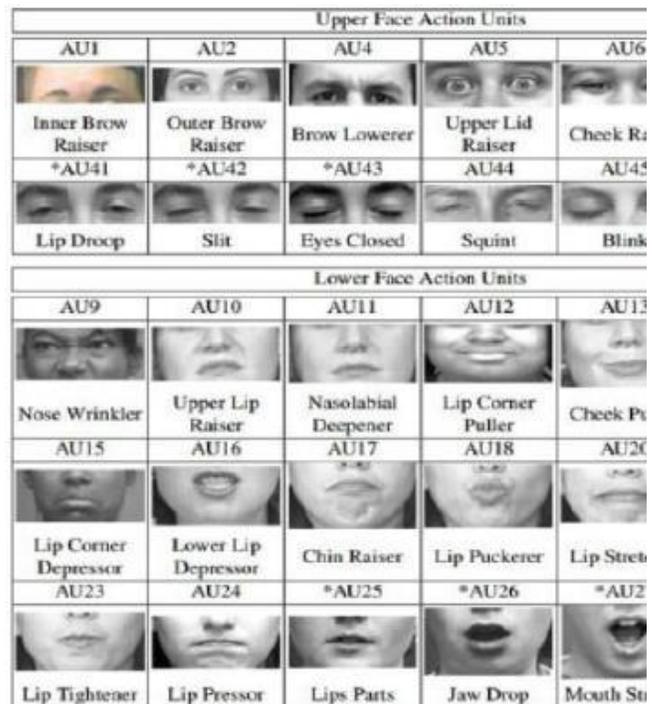


Fig. 1. Different Units of Action for the Upper and Lower Part of the Face [11].

In their paper [14,15], Cha and Kim proposed the use of webcams to measure the duration of attention and a learner's movements. The changes of these facial features were analyzed

in [14]: face up, face down, face turned, eyes closed, eyes open. It is claimed that they "determine the focus and unfocus states for the face and the eye from the coordinates of the extracted characteristic points". The authors fail to explain or define how they achieved this "focused state" or "unfocused state". When the learner's face is in a "front face" position, they conclude they are concentrated. Additionally, they claim that whether a student is "focused" or "non-focused" depends on the data obtained from "the learner's face tilts or turns sideways or their eyes open or close". However, they do not demonstrate a degree of concentration or explain how this analysis is performed. Adding a set of new features to [14], [15], the authors added smiles, surprise, sadness, anger, and closed and open mouths. In the study, "concentration" was defined as "the state of an open eye, open mouth, depressed expression, face turned, and facial expressions of emotion," while "non-concentration" was defined as "the state of a closed eye, open mouth, or dejected expression".

If both the criterion value and the value are above 0.9 milliseconds ", the learner's eyes are closed, indicating that he is not focused. The blinking eye occurs if the value is under 0.9 ms", indicating concentration. Those who do not concentrate will have their eyes closed, their mouth open, their faces turned, or facial expressions of emotion such as smile, surprise, sadness, or anger.

Students' gaze tracking behavior in front of a computer screen is studied by Yi et al. [16]. The resulting learning status is calculated. By assessing this learning status, the quality of the teaching can be evaluated. During the cognitive learning process, the eyes perform three basic actions: scanning, searching, and seclusion. According to them, pupils do not use the information captured by their eyes in a situation of inactivity on a cognitive level. This method, however, only utilizes eye movement information.

Yale and JAFFE [17], Xia Mao and Zheng Li designed an intelligent online learning system to learn about a student's emotional states through facial expression, speech, and text.

Lan Li, Li Cheng, and Kun-xi Qian explored the use of affective computing in facial expression-based e-learning system, and classify learners' emotional states into four categories such as surprise, confusion, frustration and confidence [19].

An analysis and representation of facial dynamics is described in [20]. Using facial expressions, the algorithm calculates the optical flow to determine the direction of movement.

By automatically detecting subtle changes in expressions, [21] is aimed at developing an optical flow-based approach to capturing facial expressions.

Fig. 2 illustrates the comparison of Sanchez et al.'s [22] two FER methods, which are based, respectively, on feature point tracking and dense flow tracking.

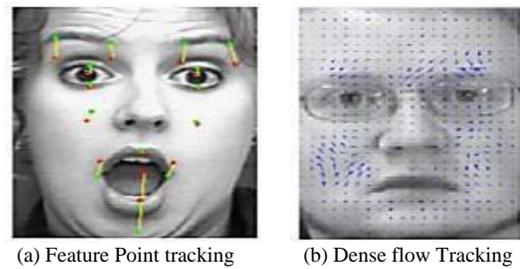


Fig. 2. Applications of Optical Flow based Methods on Facial Images in [22].

When using deep learning for FER, CNN is well suited to detect DU. FACS-based CNN-based FER methods have shown the capacity to generalize both cross-task and cross-data networks associated with FER [23]. Microexpressions are detected by the model in a well-executed manner. By using the CNN of Kim et al. [24], the LSTM is trained to learn the temporal characteristics of a spatial representation through facial expressions. We determine the most representative expressions in facial sequences regardless of the intensity or duration of their expressions as part of network learning. Kritika [25] monitors the position of pupils' eyes and heads, and generates an alert if concentration is low. Videos were analysed and dissected. MATLAB was used to implement and detect faces and Violas-Jones features using different functions. Students can find out whether they are feeling negatively in e-learning environments by using the system. Even though considerable progress has been made in this field of research, emotion and focus still need further exploration. This article attempts to establish an index of student concentration by analyzing facial expressions.

III. PROPOSED METHOD

In this article, a system is proposed that automatically discovers the learner's state of concentration in real time from facial expressions using a web camera.

From the learner's camera, facial expressions are analyzed to determine the student's state of concentration, but there is no standard database for the teaching environment and most studies are based on databases of posed expressions based on six universal expressions such as happiness, sadness, anger, surprise, neutral and fear. These are not suitable for an online learning environment.

The datasets used in this research were collected from different datasets. Therefore, The Concentration State Ranking System consists of three modules, as shown in Fig. 3:

A. Face Detector

This algorithm detects and extracts the student's features quickly and efficiently by using the Haar Cascade algorithm developed by Viola & Jones. In recent years, this method has become one of the most popular methods to accomplish this purpose [26].

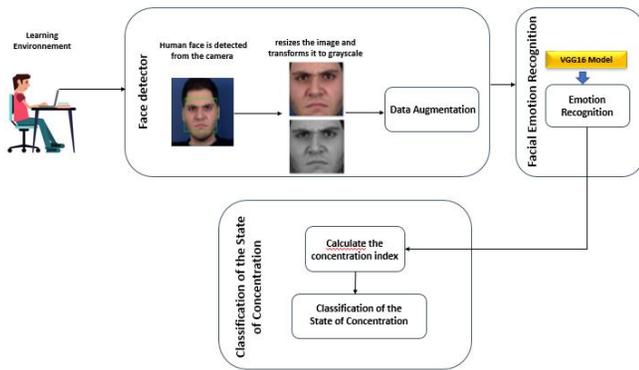


Fig. 3. System Proposed for Determining the Concentration Level of Students in Real Time from their Facial Expressions.

B. Recognition of Facial Emotions

An alternative model (that is useful for detecting facial expressions) uses transfer learning to identify the dominant emotion represented by a student's face at any given moment. There are six categories of facial expressions based on the emotion expressed: angry, scared, happy, sad, surprised, or neutral. Transfer learning is a machine learning approach that focuses on the ability to apply relevant knowledge from previous learning experiences to a different but related problem. We used a transfer learning approach to create a phrase recognition framework using databases (CK+, fer2013, and JAFEE). We used 4 pre-trained models (VGG16, VGG19, Resnet50, AlexNet, Xception) [27], which are deep convolutional networks designed for object recognition [28] and have shown good results on the ImageNet dataset [29] for object recognition. We replaced the last fully connected layers of our models with a dense layer with six outputs. The number of outputs in the last dense layer corresponds to the number of classes to be recognized. We trained the last dense layer with images from the database using the softmax activation function and the ADAM optimizer [30].

C. Classification of Concentration

The concentration index (CI) is calculated by multiplying the probability of dominant emotions (DEP) by the corresponding emotion weights (EW) see Table II. Another way to put it is that emotional weight is the measure of how well one's mental state reflects one's concentration at a given time. A value between 0 and 1 is assigned to it.

According to the results, a student's concentration level is classified into one of three categories: highly concentrated, nominally concentrated, and not concentrated.

TABLE II. CLASSIFICATION OF CONCENTRATION

Emotion	EW
neutral	0.9
happy	0.6
surprised	0.5
sad	0.3
fear	0.3
angry	0.25

- Very concentrated: a student falls into this category when the value of his concentration index based on facial emotion is between 50% and 100%.
- Nominally concentrated: a student falls into this category when the value of the Facial emotion concentration index is between 50% and 20%.
- Not at all concentrated: the student's concentration is in this category when the value of the facial emotion concentration index is less than 20%.

Fig. 3 shows an example of a real-time system that provides information data to teachers in real-time. Teachers and e-learning systems use these data to monitor learners in real time as they stream content, so that the teacher can adjust the teaching accordingly, that is, if the concentration level gets too low score, then the teaching material is too difficult for the learners and the difficulty level of the material can be adjusted.

IV. COLLECTING AND PREPROCESSING DATA

To ensure that the output was not biased in favor of a particular dataset, multiple datasets were gathered.

The following standard facial databases are available online: CK & CK+ [31], FER2013 [32], and JAFEE [33]. There were 29,207 images in the training dataset prior to the increase in data.

- Class 'anger' contains 4160 images,
- Class 'fear' contains 4204 images,
- Class 'happy' contains 7453 images,
- Class 'neutral' contains 4995 images,
- Class 'sadness' contains 4945 images, and the class 'surprise' contains 3450 images.

The validation dataset contained a total of 3,533 images.

- Class 'anger' contains 467 images,
- Class 'fear' contains 496 images,
- Class 'happy' contains 895 images,
- Class 'neutral' contains 607 images,
- Class 'sadness' contains 653 images, and the class 'surprise' contains 415 images.

Examples of model datasets are shown in the following Fig. 5.

A. Face Detection and Cropping

Detecting the location of faces from images is known as the face detection process, or face registration. The faces from the images were detected using OpenCV Cascade [34]. The face was detected and then cropped to avoid the complexity of the background, thereby improving the efficiency of the model.

B. Converting to Grayscale

Red, green, and blue channels were added to the images to make them 224 x 224 pixels. By converting the images to

Grayscale with only one channel, we were able to reduce the pixel complexity in the dataset [18] [35]. The training process was streamlined as a result.

C. Image Augmentation

The amount of data could be increased to improve the model's performance. An image augmentation process produces additional images by performing certain operations on existing image data sets, such as random rotation, zoom, shear, flip, etc. (see Fig. 4).



Fig. 4. Dataset Samples.



Fig. 5. Data Augmentation.

V. EXPERIMENTAL RESULTS

A. Comparison between the Models

Using Keras and TensorFlow (<https://keras.io/>) [15], we created a neural network and image processing system using a Python-based neural network API. The program offers many functions and models, which are important for improving the quality of images.

It provides easy-to-use tools for creating custom neural networks, which facilitate rapid experimentation. Google Colaboratory, a free cloud service that supports GPUs, was used to reduce training time.

To train our dataset, we used 100 epochs of 871 steps each. We analyzed the pre-trained models, VGG-16 and VGG-19, Xception, Alexnet.

The VGG 19 model performed excellently with a training accuracy of 90% (show Fig. 8 and Fig. 9) achieved in the 70th epoch, while the VGG16 and Xception (show Fig. 10, Fig 11 and Fig. 12 and Fig 13) model achieved its full 90% accuracy in the 50th epoch. However, the AlexNet (show Fig. 6 and Fig. 7) model was a bit long with a 91% accuracy in the 90th epoch. The accuracy of Alexnet training was 99.8%, but it could have been higher.

From Table III, we can observe that Vgg16 obtains the highest accuracy 100%. Vgg16 and Vgg19 get the most optimal scores in terms of Accuracy and error rate - 100%, 19% and 99.8%, 18%, respectively.

As a conclusion, VGG16 is a powerful deep learning model for facial emotion classification, specifically for CK+, JAFEE, and FER13 images classification. VGG16 produced the highest training and testing accuracy compared to other models. Vgg19, outperforming AlexNet, obtained Xception the second highest accuracy.

B. Determine the Level of Concentration of Students

The concentration level can be detected by analyzing the result of the seven emotions. The level of concentration can be categorized into three levels: high, medium and low, respectively.

TABLE III. COMPARISON BETWEEN THE MODELS

Model name	Maximum training accuracy	Maximum Validation accuracy
Xception	100%	85%
Vgg16	100%	86%
Vgg19	99.8%	87%
AlexNet	99.8%	64.5%

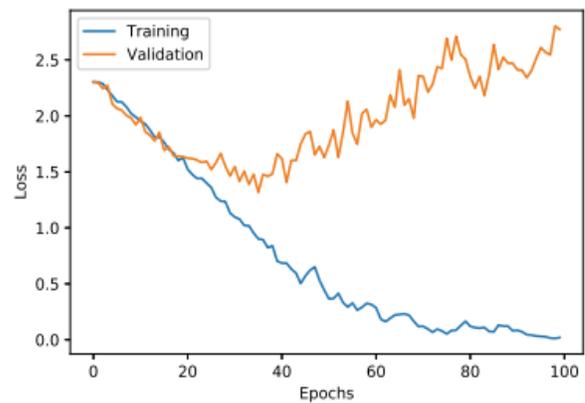


Fig. 6. AlexNet Training and Validation Accuracy Loss.

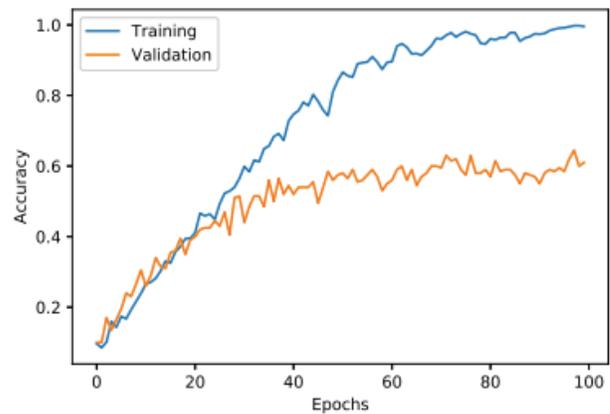


Fig. 7. AlexNet Training and Validation Accuracy.

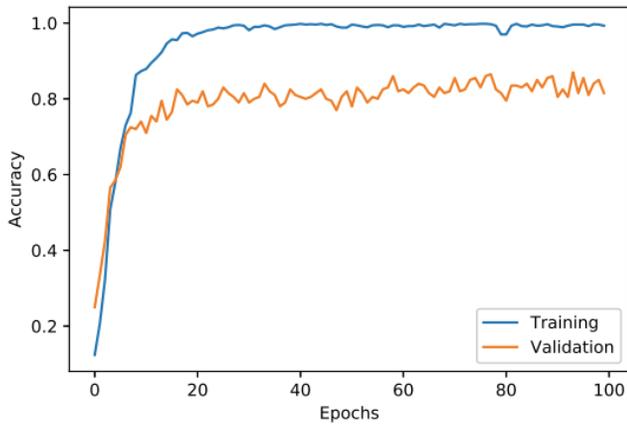


Fig. 8. VGG19 Training and Validation Accuracy.

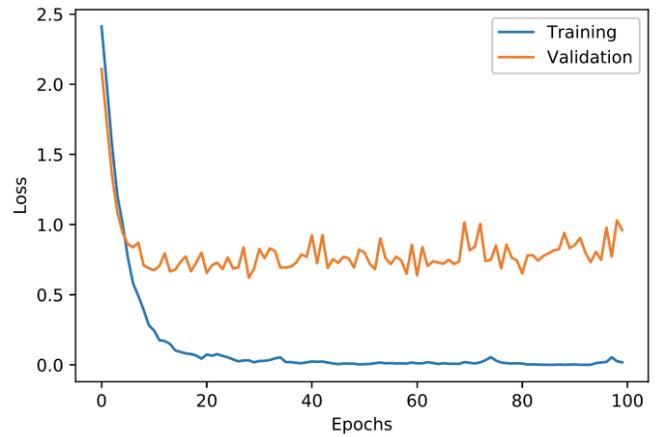


Fig. 11. VGG16 Training and Validation Loss.

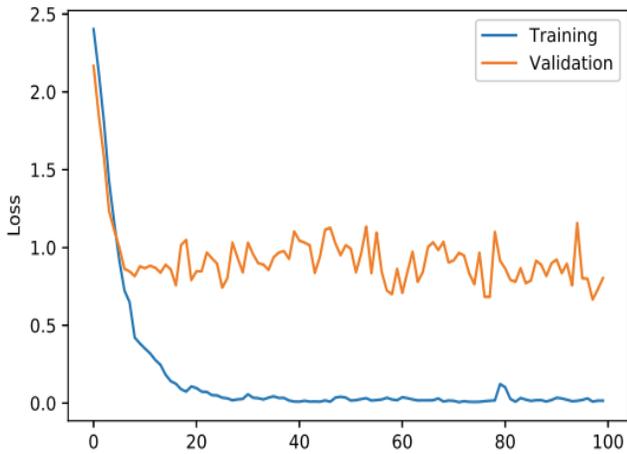


Fig. 9. VGG19 Training and Validation Loss.

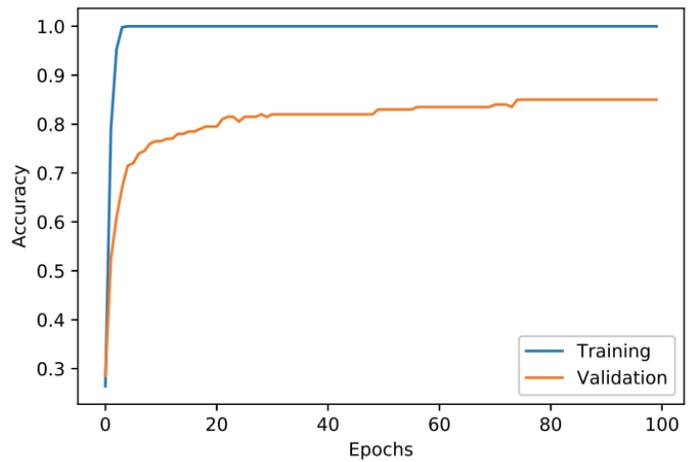


Fig. 12. Xception Training and Validation Loss.

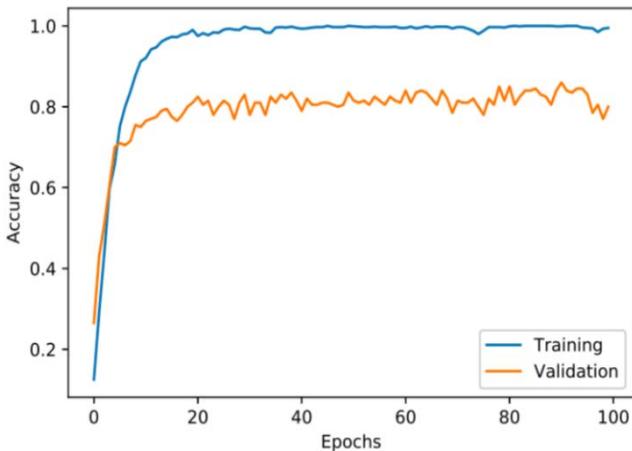


Fig. 10. VGG16 Training and Validation ACCURACY.

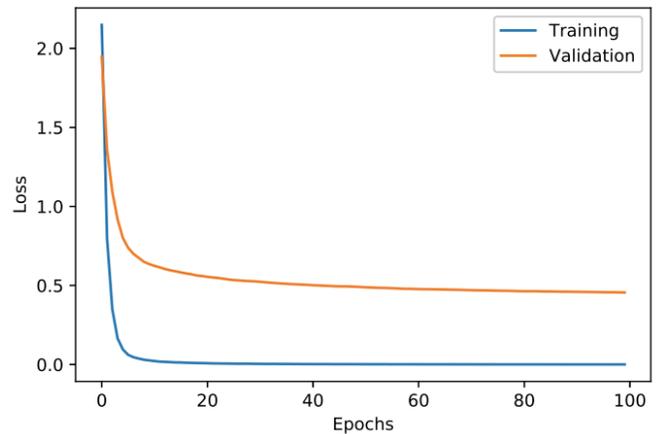


Fig. 13. Xception Training and Validation Accuracy.

An evolution of the student's concentration based only on facial expressions. For the index of concentration to be calculated, each emotion adds its own value. Based only on facial emotion detection, we calculated the percentage concentration using the following rules shown in Table IV:

Fig. 14 shows an overview of the system operating in real time and presenting information data to the teacher. This data, which includes the learner's facial emotions, and the level of concentration detected (Very concentrated, Nominally concentrated, Not at all concentrated).

TABLE IV. CALCULATE THE LEVEL OF CONCENTRATION

Emotion	CI
neutral	$(DEP * 0.9) * 100$
happy	$(DEP * 0.6) * 100$
suprised	$(DEP * 0.5) * 100$
sad	$(DEP * 0.3) * 100$
fear	$(DEP * 0.3) * 100$
angry	$(DEP * 0.25) * 100$
Not Emotion	0 (i.e., Not at all concentrated)

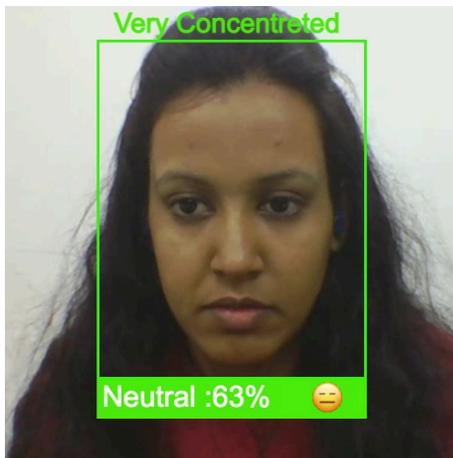


Fig. 14. General view of the System Operating in Real Time.

VI. CONCLUSION

Recognizing human emotions is one of the most challenging tasks in e-learning. In this paper, we explored emotion detection to predict students' concentration level. One of the biggest challenges and benefits of distance learning, and especially online learning, is having a system that can determine the concentration levels of students. We experimented with transfer learning models on the dataset we combined (JAFEE + FER2013+CK+) for facial expression recognition to determine the concentration level of students during educational tasks and then estimated their effectiveness. The results show that VGG16 performs better than other proposed models, while the results show that the VGG19 model could also achieve decent accuracy in facial expression recognition. In our research we present an approach to a system for detecting students' concentration level from facial expressions. Only the web camera provides information to the system. We developed a system that produces a concentration index based on the facial expressions captured by the camera. It was designed to work in real time. Three different concentration levels are presented: "highly concentrated", "nominally concentrated", and "no concentration whatsoever". This system can also help teachers to know the learning state of their students, so that the teacher can adapt the teaching material accordingly, i.e. if the concentration level is too low, the teaching material is too difficult for the learners and the difficulty level of the material can be adjusted. Furthermore, the research can be improved by including an analysis of the history of emotions detected over a given period of time in

order to establish a predictive model of when students are likely to drop out or fail in a subject.

REFERENCES

- [1] Jacob Whitehill, Zewelanj Serpell, Yi-Ching Lin, Aysha Foster, and Javier R Movellan. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, 5(1):86–98, 2014.
- [2] Michael K Barbour and Thomas C Reeves. The reality of virtual schools: A review of the literature. *Computers & Education*, 52(2):402–416, 2009.
- [3] Neil A Bradbury. Attention span during lectures: 8 seconds, 10 minutes, or more?, 2016.
- [4] Reed W Larson and Maryse H Richards. Boredom in the middle school years: Blaming schools versus blaming students. *American journal of education*, 99(4):418–443, 1991.
- [5] Paul Ekman, Wallace V Friesen, and Phoebe Ellsworth. *Emotion in the human face: Guidelines for research and an integration of findings*, volume 11. Elsevier, 2013.
- [6] Margaret M Bradley, Laura Miccoli, Miguel A Escrig, and Peter J Lang. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4):602–607, 2008.
- [7] Hayes, D.: ICT and learning: lessons from Australian classrooms. *Comput. Educ.*2(49),385–395 (2007).
- [8] Arkorful, V., Abaidoo, N.: The role of e-learning, the advantages and disadvantages of its adoption in higher education. *Int. J. Educ. Res.*2(12), 397–410 (2014).
- [9] Yewale, P., Zure, S., Awat, A., Kale, R.: Emotion recognition using image processing. *Imperial J. Interdisciplinary Res.*3(5) (2017).
- [10] Ekman, P.: Universal facial expressions of emotions. *California Mental Health Res. Digest*8(4), 151–158 (1970).
- [11] Du, S., Tao, Y., Martinez, A.: Compound facial expressions of emotion. *Proc. Natl. Acad. Sci. U.S.A.*111(15), 1454–1462 (2014).
- [12] Benitez, Q.C., Srinivasan, R., Martinez, A.: EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, pp. 5562–5570 (2016).
- [13] Jonathan Bidwell and Henry Fuchs. Classroom analytics: Measuring student engagement with automated gaze tracking. *Behav Res Methods*, 49:113, 2011.
- [14] Cha, S., Kim, W.: Analyze the learner's concentration using detection of facial featurepoints. *Adv. Sci. Technol. Lett.*92,72–76 (2015).
- [15] Cha, S., Kim, W.: The analysis of learner's concentration by facial expression changes & movements. *Int. J. Appl. Eng. Res.*11(23), 11344–11349 (2016).
- [16] Yi, J., Sheng, B., Shen, R., Lin, W., Wu, E.: Real time learning evaluation based on gazetracking. In: 14th International Conference on Computer-Aided Design and Computer Graphics, Shanghai, pp. 157–164 (2015).
- [17] Sathik, M.M. and Sofia, G. Identification of student comprehension using forehead wrinkles. *IEEE International Conference in Computer, Communication and Electrical Technology (ICCCET)*, 2011, 66–70.
- [18] Mao, X. and Li, Z. Implementing emotion-based user-aware e-learning. *ACM in CHI'09 Extended Abstracts on Human Factors in Computing Systems*, 2009, 3787–3792.
- [19] Li, L., Cheng, L. and Qian, K.X. An e-learning system model based on affective computing. *IEEE In International Conference on Cyber worlds*, 2008, 45–50.
- [20] Yacoub, Y.; Davis, L.S. Recognizing human facial expressions from long image sequences using optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.*1996,18, 636–642.
- [21] Cohn, J.F.; Zlochower, A.J.; Lien, J.J.; Kanade, T. Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 14–16 April 1998; p. 396.

- [22] Sánchez, A.; Ruiz, J.V.; Moreno, A.B.; Montemayor, A.S.; Hernández, J.; Pantrigo, J.J. Differential optical flow applied to automatic facial expression recognition. *Neurocomputing* 2011, 74, 1272–1282.
- [23] Breuer, R.; Kimmel, R. A deep learning perspective on the origin of facial expressions. *arXiv* 2017, arXiv:1705.01842.
- [24] Kim, D.H.; Baddar, W.; Jang, J.; Ro, Y.M. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition. *IEEE Trans. Affect. Comput.* 2017.
- [25] LB Krithika. Student emotion recognition system (SERS) for e-learning improvement based on learner concentration metric. *Procedia Computer Science*, 85:767–776, 2016.
- [26] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.
- [27] K. Simonyan, A. Zisserman, Very deep convolutional networks for largescale image recognition, in: *Arxiv*, 2014.
- [28] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (6) (2018) 1452–1464.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: *Computer Vision and Pattern Recognition*, 2009.
- [30] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, *CoRR* arXiv:1412.6980.
- [31] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 94–101.
- [32] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, “Deep learning approaches for facial emotion recognition: A case study on fer-2013,” in *Advances in Hybridization of Intelligent Methods*. Springer, 2018, pp. 1–16.
- [33] Lyons, M.; Akamatsu, S.; Kamachi, M.; Gyoba, J. Coding facial expressions with gabor wavelets. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 14–16 April 1998; pp. 200–205.
- [34] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *IEEE*, 2001, p. 511.
- [35] M. Grundland and N. A. Dodgson, “Decolorize: Fast, contrast enhancing, color to grayscale conversion,” *Pattern Recognition*, vol. 40, no. 11, pp. 2891–2896, 2007.