Automatic Facial Expression Recognition using CNN and RNN Algorithm's

N. LOKESWARI, M.Tech, (Ph.D.) [#1], K. AMARAVATHI, M.Tech, (Ph.D) [#2]

[#1] Assistant Professor, [#2] Assistant Professor,

Department of Computer Science & Engineering,

Anil Neerukonda Institute of Technology & Sciences,

Sanghivalasa, Visakhapatnam, AP 531162.

ABSTRACT

In current days for identifying a person emotion takes a lot of time because of varying expressions and features present in human faces. There was a lot of study going on to detect the several types of expressions from human's in different environments such as work,play,relax,break and so on. Home room correspondence includes instructor's conduct and understudy's reactions. There was a lot of research work done on the investigation of below average students and average students by the instructor to identify those students and take some special care.In the primitive day's almost human emotions are identified manually based on the current expression and feelings of the human's in day to day environment. But there was no accurate method designed in primitive days for identifying the emotions or expressions automatically. This emotion recognition is becoming an interesting aspect which is used mainly for diagnosis of human brain and psychological disorders. In a recent survey conducted by a team of experts, we came to know that deep learning has gained a lot of user's attention in the field of image classification. These emotions is used for not only diagnosis of human brain but also used as a recommended systems to assist users in finding items that match their needs and preferences. Hence this motivated me to develop a system which can effectively and efficiently recognizes emotions from the facial expressions of the user. In this proposed article we try to develop an application which can be used for prediction of expressions of both still images. Here we try to develop a system by using two well-known models such as CNN and RNN and then check which model suits best for facial expression recognition. The highest probability value to the corresponding expression will be the predicted expression for the image. We have conducted experiments on FER_2013 Data set which we collected from KAGGLE website and try to train the system to detect the emotions accurately. Experimentation results show critical execution acquire on boundaries like exactness, F1-score, and review.

Keywords:

Convolution Neural Networks (CNN); Recurrent Neural Networks (RNN), Psychological Disorders, Expressions, Kaggle, Emotion Recognition

1. INTRODUCTION

Now a days it is very interesting and challenging task to detect expressions from pictures and recordings and has become extremely huge on account of its various applications in the PC vision area, for example, in human computer collaboration [1], expanded and computer generated reality [2, 3], advance driver help framework (ADASs) [4], and video recovery and security frameworks [5–7]. Human expressions and emotions have been analyzed in examinations with the assistance of acoustic and semantic highlights [8], looks [9–12], body pose, hand development, bearing of look [13], and use of electroencephalograms (EEGs) and electrocardiograms (ECGs) [14]. Educators show various feelings which are brought about by different reasons [15], e.g., a teacher may encounter delight when an instructive goal is being satisfied or when understudies follow given headings. At

the point when understudies show an absence of interest and reluctance to get a handle on an idea, it causes dissatisfaction. Also, outrage is reflected when understudies need discipline. As indicated by educators, these looks regularly emerge from disciplinary study hall collaborations, and dealing with these looks oftentimes helps them in accomplishing their objectives [16]. The impression of data preparing has totally changed by these profound learning draws near. Because of its amazing capacity of self-learning, profound learning is viewed as a superior choice for vision and characterization issues. Different methodologies for arrangement incorporate pretrained networks which decrease the cycle of long preparing by presenting the utilization of pretrained loads. Nonetheless, learning here includes tuning of millions of organization boundaries and colossal marked information for preparing. Since FER is fundamentally pertinent to various fields, we trust FER utilizing profound highlights can be appropriate in understanding the semantics of educator's looks in a study hall climate.

Now a days it is becoming a challenging task to identify or detect the facial features in the fields of video surveillance camera or image cameras, biometrics and a lot more. There are nearly seven basic emotions in the universe to detect the current emotion of human beings, namely neutral, angry, disgust, fear, happy, sad, and surprise and these basic emotions can be recognized from human's facial expression[2]. In general to solve the problem of recognizing facial features is very complex job because each and every individual has several facial features when compared with one person to other person.

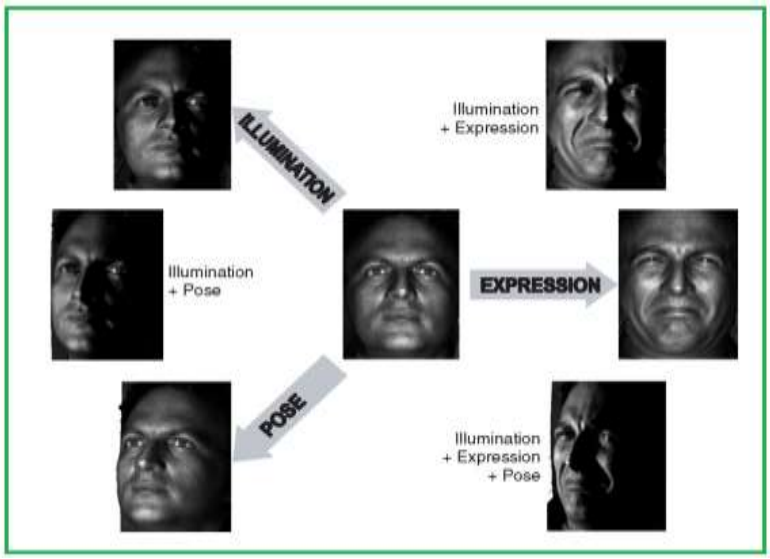


Figure 1. Denote the Several Factors which are considered for Face Emotion Recognition

In general there are a lot of factors which influence the features like physical characteristics, sex, genes, and age. In the real world environment there are many factors that have to be taken emotion recognition system which is shown in figure 1.The main step for any face processing system is the ability to classify the face accurately and then try to detect the emotion [3] from that facial expression. A general emotion detection system has four stages of processflows:

1) Face Detection
2) Preprocessing
3) Feature Extraction, and

4) Emotion Recognition

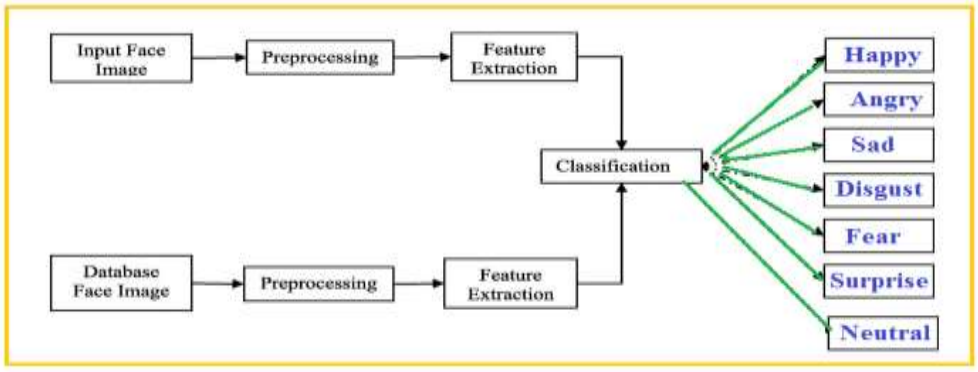


Figure 2. Denotes the General Flow of Emotion Detection from Facial Expressions

From the above figure 2, we can clearly identify the system is going to identify the emotions collected from input image and initially we try to load the input dataset which contains a set of images. All the images are given for the system for training purpose and once the images are loaded into the application they are pre-processed and features are extracted[8]. All the features are extracted and kept ready for classification of input image. If any user who wish to find out the emotion he need to load the image either through web cam or he can browse for an image and then try to train with already pre-loaded images and now after classification is done, emotion is detected. During the process of emotion detection the accuracy of emotion depend on the quality of image and factors which are influenced for emotion detection. If the image is clear and having bright appearance then emotion will be accurately detected and if the input contains any noise or factors which affect the image recognition then emotion may not be accurate[9].

2. LITERATURE SURVEY

Literature survey is that the most vital step in the software development process. Before developing the new application or model, it's necessary to work out the time factor, economy, and company strength. Once all these factors are confirmed and got approval then we can start building the application. The literature survey is one that mainly deals with all the previous work which is done by several users and what are the advantages and limitations of those previous models. This literature survey is mainly used for identifying the list of resources to construct this proposed application.

MOTIVATION

1) Automated Facial Expression Recognition System Using Neural Network Classifiers

Creators: Jyh-Yeong et al.

In this proposed paper the author proposes automated facial expression recognition system using neural network classifiers. The author used Rough Contour Estimation Routine (RCER) technique for extracting the features from a human face like eyebrows, eyes, mouth with the help of Point Contour Detection Method (PCDM)

[16] in order to improve and detect the precision of eye and mouth. In this proposed paper the author try to find out a novel method like Action Units (AU) [17] in which we can able to see the basic movements of face muscles.

2) A real time face emotion classification and recognition using deep learning model

Creators: Dr. Shaik Asif Hussain, Ahlam Salim Abdallah Al Balushi

The proposed authors try to discuss about face emotion classification and recognition under real time manner by using deep learning model. In this paper the authors try to extract the main features with deep learning, Haar cascade and VGG 16 model to recognize face and try to build the classification and recognition[18]. From the experimental results the authors clearly prove that the network architecture which was designed for this current paper has better advancements than compared with existing algorithms. Here the proposed deep learning models are comparatively having more improvement than compared with several other models which were used in the literature of facial expression detection.

3) A Real-Time Recognition System for User Characteristics Based on Deep Learning.

Creators: Dan Duncan

This author proposes a real time recognition system for user characteristics based on deep learning models. In this paper the author try to design a VGG is an innovative object-recognition model that supports up to 19 layers. This is mainly built on the top of CNN and based on this CNN we can able to outperform baseline on many tasks and datasets outside of ImageNet. This proposed system is designed by overcoming the pre-processing difficulties [19]which is present in the CNN and this proposed system we can able to prove much larger dataset in order to improve the model's generality. The proposed system can able to provide accuracy in perfect manner and prove straight angle.

3. EXISTING SYSTEM AND ITS LIMITATIONS

In the existing system, there was no concept like facial expression recognition using CNN & RNN models. All the prediction is done using manual approach or by using primitive Machine Learning models. In the ML we can able to classify only few emotions accurately depend on the human expression at that appropriate situation, but the manual approach or primitive methods are not accurate to classify the emotions accurately in all situation.

LIMITATIONS OF THE EXISTING SYSTEM

1. All the existing schemes are limited to the manual classification of expression.
2. All the existing systems are failed to detect all emotions from the human expressions.
3. All the existing methods try to classify the emotions either 3 or 4 emotions but not all the seven emotions.
4. All the existing ML approaches fail to classify the emotions accurately if the image is captured in low lighting conditions or poor appearance.

5. There is no accurate model to classify the human emotions automatically from the given input image with prediction values.

4. PROPOSED SYSTEM AND ITS ADVANTAGES

In this proposed article we try to develop an application which can be used for prediction of expressions of both still images. Here we try to develop a system by using two well-known models such as CNN and RNN and then check which model suits best for facial expression recognition. The highest probability value to the corresponding expression will be the predicted expression for the image. We have conducted experiments on FER_2013 Data set which we collected from KAGGLE website and try to train the system to detect the emotions accurately. Experimentation results show critical execution acquire on boundaries like exactness, F1-score, and review.

ADVANTAGES OF THE PROPOSED SYSTEM

1. The proposed scheme is very accurate in emotion detection from several expressions.
2. The proposed system gives accurate results for the end users in order to identify the current mental state of that human from his expressions.
3. The proposed system is capable of classifying a large set of images which are captured either short distances or long distances.
4. The proposed system is designed based on CNN and RNN model to improve the accuracy of human emotion detection.
5. The proposed system can accurately identify the human emotions based on expressions in very low lighting conditions or poor appearance.
6. We can get comparative analysis of two models for automatically facial expression recognition for the given input images.

5. PROPOSED MODEL FOR AUTOMATIC FACIAL EXPRESSION RECOGNITION

In this section we try to discuss about proposed RNN & CNN models, which is used to recognize the facial expression automatically from the static images or from moving video file.

MOTIVATION

The overall structure of the proposed look acknowledgment framework is introduced in Figure 3. Its initial step includes face discovery of an educator from the talk recordings. I.e. separated teacher's face outlines are then exposed to scratch outline extraction where repetitive edges are disposed of and just mainframes are saved for every appearance. These key casings are additionally prepared to separate profound highlights from various layers of a CNN, these separated profound highlights are utilized to prepare a RELM classifier for perceiving teacher's looks inside five classes. RELM is one of the variations of the outrageous learning machine (ELM), which

depends on a primary danger minimization standard. .e underlying danger minimization rule is utilized to improve the construction of ELM, and regularization is used for exact expectation.

FACE DETECTION AND KEY FRAME SELECTION

Perceiving a teacher's looks is testing since it is unique in relation to the regular look acknowledgment framework. Basically, the information are obtained in a homeroom climate. .is includes difficulties like face imperceptibility, e.g., when the educator is composing on the board, impediment, e.g., when the teacher is perusing the slide from the PC and half of the face is taken cover behind the screen, and changing easing up conditions, e.g., when the teachers stroll under the projector's light. .e proposed calculation is planned in such a manner in order to defeat such difficulties. Keeping in see the indoor and single item climate, countenances of educators are distinguished utilizing the Viola–Jones face location approach [11]. .e location of appearances in a picture by Viola–Jones calculation looked for full upstanding front facing faces that likewise diminish the non-facial demeanor outlines [12]. For strong commonsense identification, the face should be noticeable to the camera; henceforth, just front facing faces are thought of. When the face is recognized, the bouncing box around the substance of the educator is edited, to frame a district of interest. As per the writing, the fundamental advance in handling recordings is to portion the video into worldly shots. A shot comprises of an arrangement of casings. Among every one of the casings, a key edge gives notable data of the shot. It sums up the substance of the video by eliminating repetitive data and conveying just huge dense data.

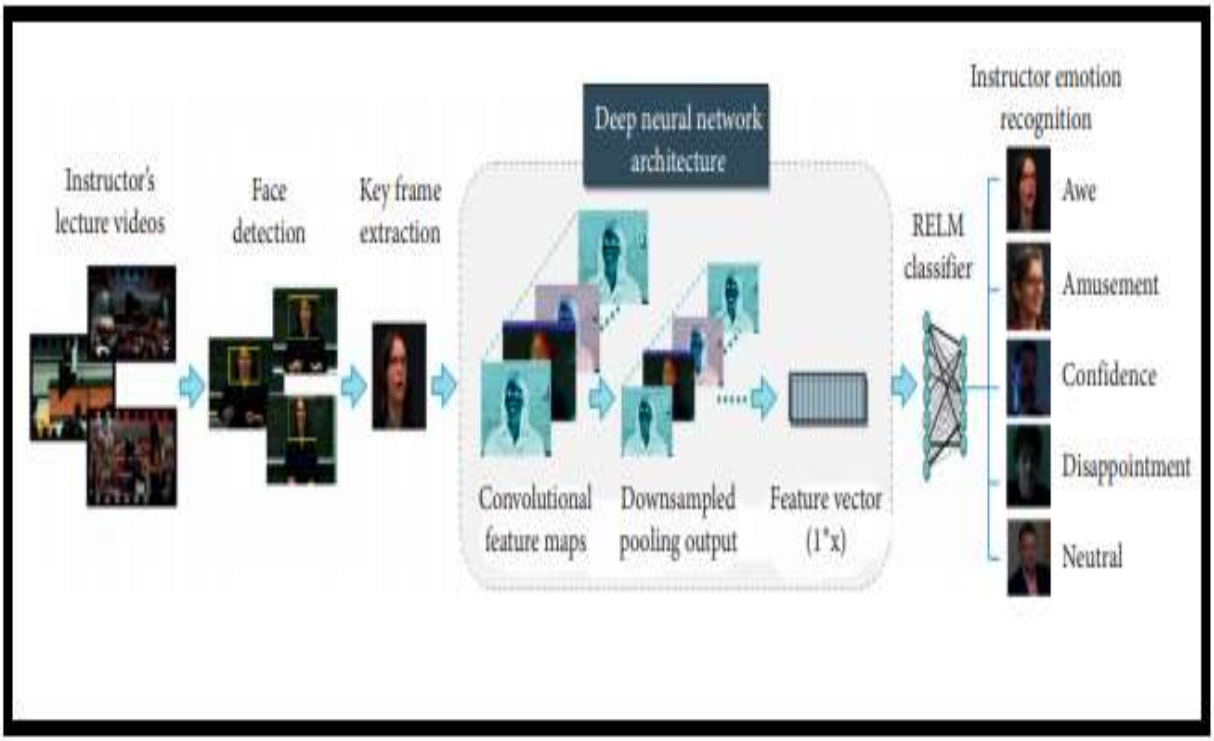


Figure 3. Denotes the Proposed Framework for Identifying the Expression of Education Instructor

Algorithmic steps

Given a set of deep neural features x, target output t, an activation function g (x) with N number of hidden nodes. the weight vector, w connects the hidden node with the output nodes.

Step 1: assign the input weights and the bias of hidden layer nodes B randomly (wi, bi), i = 1, 2, ..., L.

Step 2: compute output matrix H (w1, ..., wL, x1, ..., xN, b1, ..., bL)

Step 3: compute output weights: $\beta = H^T T$.

Step 4: for regularization the output term is computed by using $\tilde{\beta} = (\lambda + H^T H)^{-1} H^T T$

Return: parameters β, W, and b.

FIGURE 4: The RELM algorithm for classification.

TABLE 1: Tabular view of the new Instructor Expression Video (IEV) dataset along with sample images for each expression class.

Sample images from lecture videos	Expression	No. of samples	Gender
	Amusement	425	
	Awe	425	
	Confidence	425	Male: 23 Female: 7
	Disappointment	425	
	Neutral	425	

FIGURE 5: Sample sequence of seven expressions from the CK dataset.

FIGURE 6: Sample images taken for each expression from the JAFFE dataset.

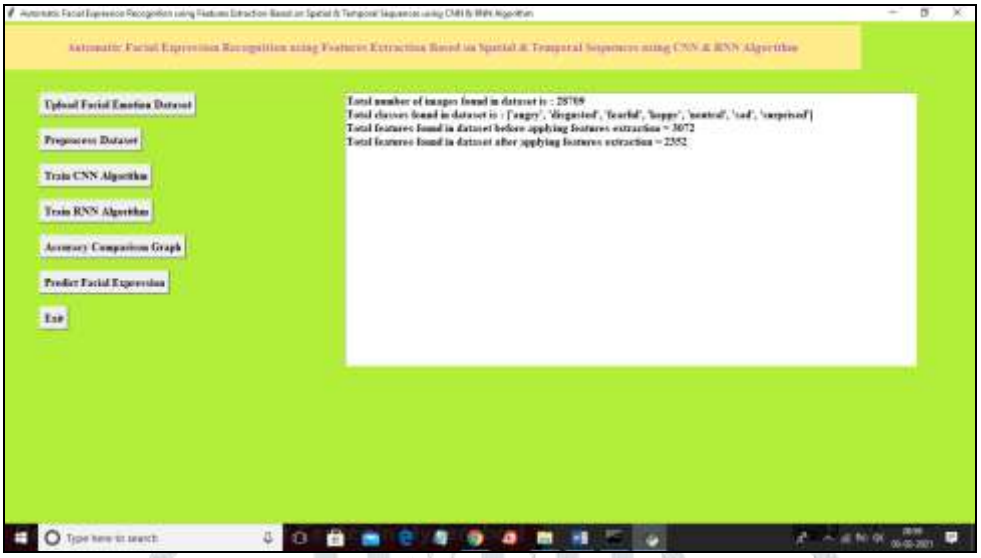
6. EXPERIMENTAL RESULTS

Implementation is a stage where the theoretical design is converted into a programmatic manner. In this proposed application we try to use PYTHON as a programming language in which Google Collaboratory or Jupiter Notebook as a working platform to process the current application.

STEP 1: IMPORTING ALL NECCESARY LIBRARIES

pip install sklearn
pip install scikit-image

STEP 2: UPLOAD FACIAL EMOTION DATASET & PRE-PROCESS

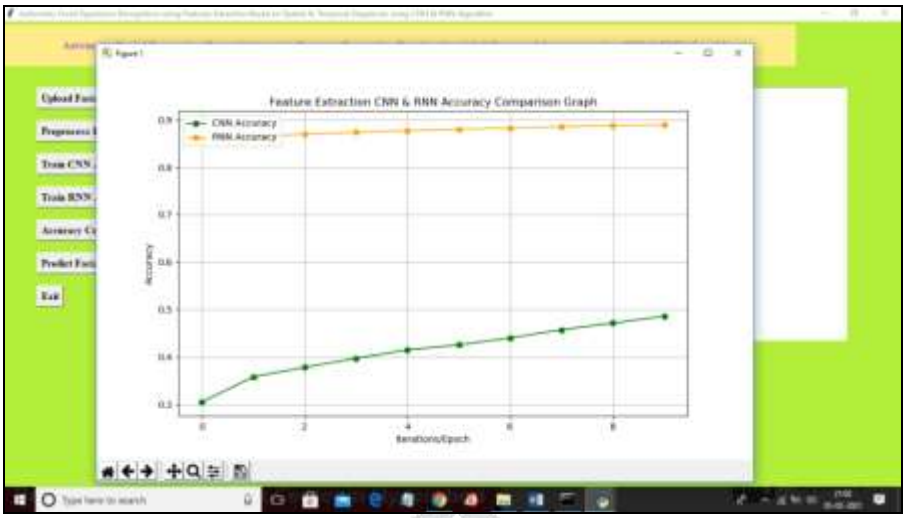


In above screen we can see dataset contains total 28709 images and before applying feature extraction algorithm total images features/pixels are 3072 and then after applying features reduces to 2352 as PCA remove unimportant pixels and used only important pixels/features. Now image data is ready and now click on 'Train CNN Algorithm' button to train CNN with process image features. In above screen CNN accuracy is 48 and now click on 'Train RNN Accuracy' button to train dataset with RNN.

STEP 3: NOW TRY TO RUN RNN ALGORITHM

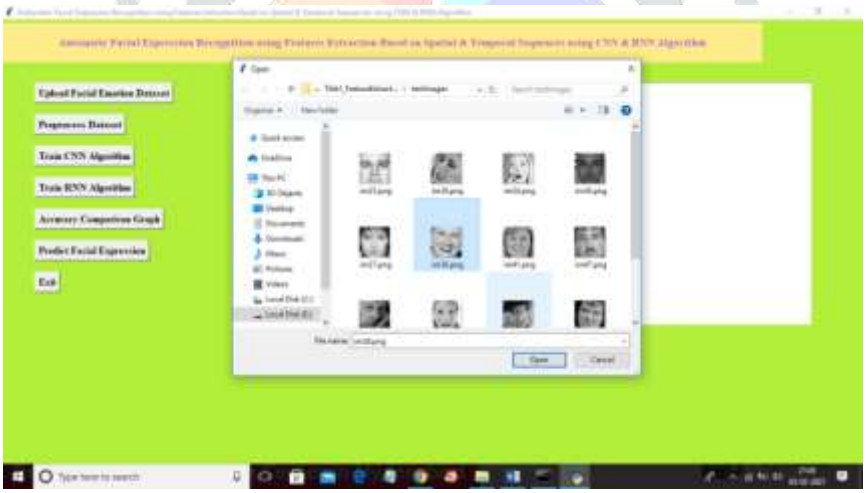
In above screen RNN accuracy is 88% and now click on 'Accuracy Comparison Graph' to get below graph of both algorithms.

STEP 4: RUN ACCURACY GRAPH



In above screen x-axis represents epoch/iteration ad y-axis represents accuracy and in above graph orange line represents RNN accuracy and green line represents CNN accuracy and from above graph we can see with further epoch/iteration both algorithm accuracy get better and better and from above graph we can conclude that RNN is giving better result. Now click on 'Predict Facial Expression' button to upload new test image and the application predict emotion from it

STEP 5: TEST MODEL WITH SAMPLE IMAGES



In above screen selecting and uploading im38.png image and then click on 'Open' button to get below result.

STEP 6: TEST RESULT

In above screen we got detected emotion as 'happy' and similarly you can upload any image and then predict emotion. So this is the output of TITLE 1.

7. CONCLUSION

In this paper, a novel approach has been proposed for facial expression recognition of instructors in a classroom environment by incorporating a feed forward learning model with deep features. In contrast to back propagation approaches, the proposed model works in a feed forward fashion. It extracts the deep features from a neural model for high-level representation gain, without updating the weights iteratively, causing a reduction in computational time complexity. Extensive experimentations are performed with state-of-the-art techniques, traditional classifiers, and other deep neural models. .e proposed method has proven to be successful in evaluating five instructor's expressions in a classroom environment. For future research, we will investigate the performance of the model with more features such as instructor's speech and activity recognition approaches in order to improve the effectiveness of classroom teaching methods.

8. REFERENCES

[1] D. C. B. Silva, P. P. Cruz, A. M. Gutierrez, and ´ L. A. S. Avendaño, Applications of Human-Computer Interaction and Robotics Based on Artificial Intelligence, Editorial Digital del Tecnol´ogico de Monterrey, Monterrey, M´exico, 2020.

[2] C.-H. Chen, I.-J. Lee, and L.-Y. Lin, "Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders," Research in Developmental Disabilities, vol. 36, pp. 396–403, 2015.

[3] S. Hickson, N. Dufour, A. Sud, V. Kwatra, and I. Essa, "Eyemotion: classifying facial expressions in VR using eyetracking cameras," in Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa Village, HI, USA, January 2019.

[4] M. A. Assari and M. Rahmati, "Driver drowsiness detection using face expression recognition," in Proceedings of the 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), November 2011.

[5] T. Bai, Y. F. Li, and X. Zhou, "Learning local appearances with sparse representation for robust and fast visual tracking," IEEE Transactions on Cybernetics, vol. 45, no. 4, pp. 663–675, 2015.

[6] T. Chen and K. H. Yap, "Discriminative BoW framework for mobile landmark recognition," IEEE Transactions on Cybernetics, vol. 44, no. 5, pp. 695–706, 2014.

[7] X. Zhu, X. Li, and S. Zhang, "Block-row sparse multiview multilabel learning for image classification," IEEE Transactions on Cybernetics, vol. 46, no. 2, pp. 450–461, 2016.

[8] K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in Proceedings of the Fifteenth Annual Conference of the International Speech Communication Association, Singapore, September 2014.

[9] R. J. J. Huang, Detection Strategies for Face Recognition Using Learning and Evolution, Citeseer, Princeton, NJ, USA, 1998.

[10] B. Gokberk, A. A. Salah, and L. Akarun, "Rank-based decision ¨ fusion for 3D shape-based face recognition," Lecture Notes in Computer Science, Springer, Berlin, Germany, pp. 1019–1028, 2005.

[11] W. Y. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, Lille, France, May 2019.

[12] T. Abegaz, G. Dozier, K. Bryant et al., "Hybrid GAs for Eigenbased facial recognition," in Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM), Paris, France, July 2011.

[13] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 677–695, 1997.

[14] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, "ASCERTAIN: emotion and personality recognition using commercial sensors," IEEE Transactions on Affective Computing, vol. 9, no. 2, pp. 147– 160, 2018.

[15] F. Cubukcu, ".e significance of teachers' academic emotions," Procedia-Social and Behavioral Sciences, vol. 70, pp. 649–653, 2013.

[16] S. Prosen and H. S. Vitulic, "Emotion regulation and coping ´ strategies in pedagogical students with different attachment styles," Japanese Psychological Research, vol. 58, no. 4, pp. 355–366, 2016.

[17] E. Jensen, M. Dale, P. J. Donnelly et al., "Toward automated feedback on teacher discourse to enhance teacher learning," in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13, Florence, Italy, April 2020.