Real-Time Detection of Student Engagement: Deep Learning-Based System

Chapter *in* Advances in Intelligent Systems and Computing · January 2022

DOI: 10.1007/978-981-16-2594-7_26

6 authors, including:

Zeyad Abdulhameed Ahmed
Dr. Babasaheb Ambedkar Marathwada University
27 PUBLICATIONS **295** CITATIONS
SEE PROFILE

Mukti Jadhav
Marathwada Institute of Technology
75 PUBLICATIONS **1,012** CITATIONS
SEE PROFILE

Ali Mansour Al-madani
Dr. Babasaheb Ambedkar Marathwada University
29 PUBLICATIONS **288** CITATIONS
SEE PROFILE

Mohammed Tawfik
Dr. Babasaheb Ambedkar Marathwada University
24 PUBLICATIONS **66** CITATIONS
SEE PROFILE

Real-Time Detection of Student Engagement: Deep Learning-Based System

Zeyad A. T. Ahmed, Mukti E. Jadhav, Ali Mansour Al-madani, Mohammed Tawfik, Saleh Nagi Alsubari, and Ahmed Abdullah A. Shareef

Abstract Due to the spread of COVID-19, E-learning has become the only option for the teachers and students. However, it is difficult for a teacher to monitor each student's engagement while teaching online. This paper aims to develop an automated real-time video-based system to detect student engagement during online classes effortlessly and efficiently. The MobileNet model is trained on an eye images dataset from Kaggle and achieved 99% accuracy on data validation. The result obtained from training the model is used with the Viola–Jones algorithm and OpenCV. By using the built-in camera of the laptop, the system can detect whether the student is engaged or disengaged from his/her eye gaze. In the disengagement state detection, a buzzer sound starts.

Keywords Engagement · Deep learning · Eye gaze · Transfer learning · E-learning · Real-time system

1 Introduction

One of the significant problems with online learning is how to improve the quality of learners' engagement in their educational activities. Students have several online educational activities such as online lessons, online exams, and watching video tutorials. Students can be distracted during such activities. The level of engagement/disengagement could be detected through facial expressions, eye movements, eye tracking, gaze patterns, and body movements [1]. The evaluation of engagement can be measured by an external observer. A real-time system is the best candidate for doing so in an E-learning environment. Engagement detection can be applied

Z. A. T. Ahmed (✉) · M. E. Jadhav · A. M. Al-madani · M. Tawfik · S. N. Alsubari · A. A. A. Shareef
Department of Computer Science, Dr. Babasaheb Ambedkar Marathwada University Aurangabad, Aurangabad, India

Shri Shivaji Science & arts College, Chikhli Dist. Buldana, India

A. Khanna et al. (eds.), *International Conference on Innovative Computing and Communications*, Advances in Intelligent Systems and Computing 1387,
https://doi.org/10.1007/978-981-16-2594-7_26

in various areas like monitoring driver behavior and analyzing the eye tracking of autistic people [2].

Engagement refers to the connection between a person and a resource that comprises behavioral, emotional, and cognitive engagement, at any point of time [3]. The quality of the engagement can be measured by eye gaze, eye contact, and facial expressions. The teachers need to recognize their students' states, whether focused or disturbed in education. This paper provides a real-time video-based system for detecting student engagement status. The eye engagement recognition system is preferable because it does not require expensive hardware or highly technical expertise to operate. Simply, the built-in camera of the laptop serves well to capture a video of the student's face and eyes. The captured videos can be transferred to input frames to be input to the CNN model to decide whether the student is *engaged* or *disengaged*. When the student's eyes are open and focused on the screen, the state is *engaged* and a green notification, *engaged*, is shown above the face boundary box. The *disengaged* state occurs when the student looks out of the screen or closes his/her eyes. In a such case, a red notification word *disengaged* is shown and a buzzer starts. This system improves the quality of E-learning by monitoring the student's engagement in the E-learning environment.

This paper starts with reviewing the most recent and relevant works in the area of student engagement detection through deep learning. Then it moves on to describe the proposed methodology and algorithm highlighting some deep learning models. After that, it presents the experimental work and results. Finally, there is a discussion of the study and conclusion.

2 Related Work

Videos, audios, and learner log data are used in the level of engagement detection. This research paper is based on computer vision that uses the facial features. Here is a review of some previous related works.

Sharma et al. [3] proposed a system to detect students' engagement using the eyes and emotion analysis in which they applied two models. The first CNN model was trained on eye images to detect the two classes of eye images; "Focused" and "Distracted". The second pretrained model which was trained on the FER-2013 dataset consists of grayscale images of facial expressions such as "happy" and "angry". The system was meant to detect the focus of students and then analyze their emotions. Nezami et al. [4] developed a CNN model for engagement recognition and facial expression of the students. They collected their own dataset, called *engagement recognition* (ER) dataset, that consisted of 4627 annotated images; 2290 *engaged* and 2337 *disengaged*. For the facial expression, they used the FER-2013 dataset. Student behaviors such as a student looking to the screen or looking down to the keyboard were detected. Simultaneously, they recognized the facial expression such as happy or excited, then compared the two models to get the resultof whether the student was engaged or not. The student's detection was described as *disengaged* when the student was looking anywhere else other than the screen or closed his eyes. Then his/her emotion was classified to be either *bored* or *confused*. CNN and VGG models were trained on their datasets. The classification accuracy of the engagement model was 72.38%. Núñez et al. [5] proposed a real-time system to detect eye gaze direction in videos. The data were collected from 40 non-autistic adults and 31 children: 23 children not diagnosed as autistic and 8 conformed diagnosed as autistic. The CNN model was trained on three classes of eye directions: *left*, *right*, and *unknown*. The CNN model achieved 95.1% accuracy.

Rodríguez et al. [6] applied decision tree algorithm for student engagement detection using nonverbal behavior. The data were collected from 5 students with 60 instances that were classified into three classes: 17 *neutral*, 14 *yes*, and 27 *no*. They used five attributes, i.e., face, eyes, shoulders, mouth, and interest. They classified the output into *interested* or *uninterested* and *neutral*. The highest accuracy achieved was 87% by using the random decision tree. Hashemi et al. [7] built a CNN model to detect drowsy drivers. They used the ZJU dataset of eye images "*close*, *open*" and extended ZJU by their own dataset, consisting of 4841 images, 2458 open eye images and 2383 closed-eye images. Their CNN model achieved 98% accuracy. Kaur [8] developed a DNN and LSTM for localization of student engagement. The data were collected from 78 subjects, 25 females and 53 males. 195 videos were collected. The duration of each video was 5 minutes. Eye gaze and head pose features were extracted by using OpenFace. For training, they used DNN and LSTM and Random Forest. The mean squared error of both RF and LSTM was 0.09%. Thomas et al. [9] predicted student engagement by using the face. The data were collected from 10 students during watching some YouTube videos that were used to motivate people. They analyzed the engagement level of the students while watching the videos. Data were labeled as engaged and distracted. For feature extraction, they used an open-source toolbox for the analysis of the face. The SVM (RBF) classifier was applied to gaze and pose features, and the classification accuracy result was 86%.

3 Methodology

The process of learning will become more effective when the teacher or the educator could keep track of the student's engagement level. This study provides a real-time system to monitor student engagement.

3.1 The Architecture of the Student Engagement System

This section gives details on the algorithm and the architecture of the system. The algorithm is a sequence of steps to solve the problem. The flowchart of the student engagement detection system is shown in Fig. 1. Here, we provide the steps of the algorithm.

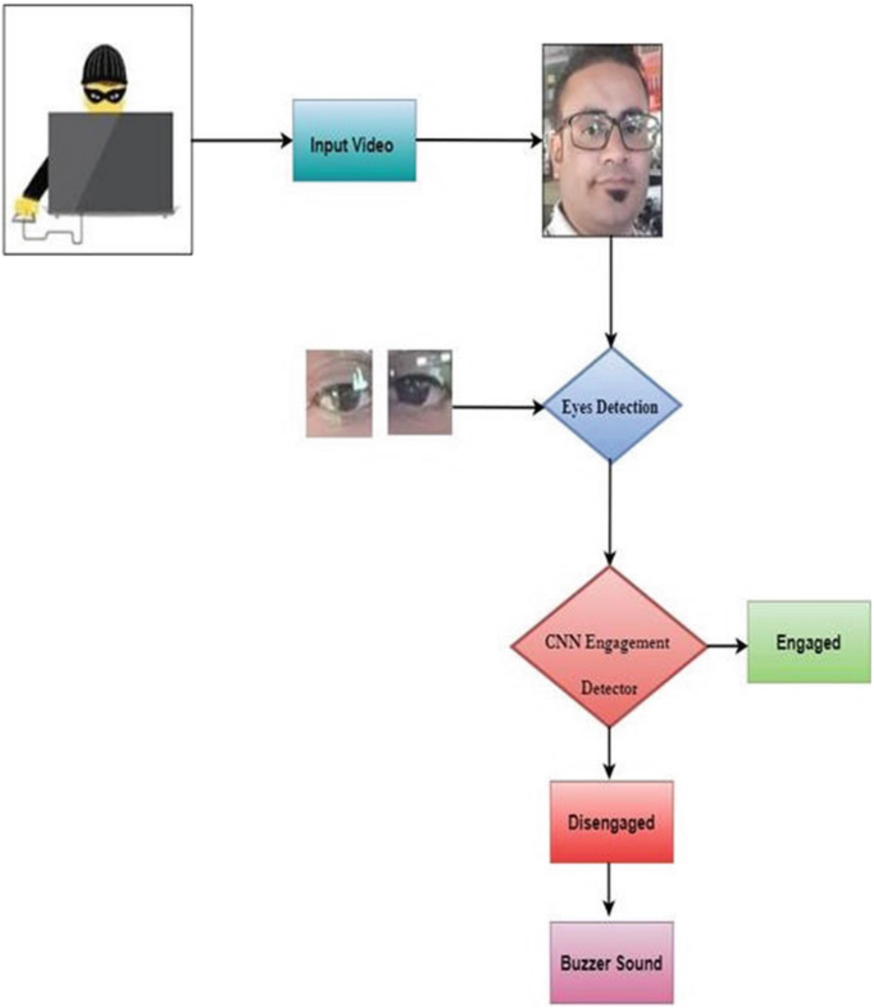


Fig. 1 Architecture of the student engagement system

Step.1: The student should sit in front of the laptop with a built-in camera or desktop computer with a webcam.

Step.2: when the student starts his online meeting, the system captures the video of the meeting.

Step.3: Haar Cascade object detection is used for face and eyes detection. Step.4: The eyes region is cropped by Haar Cascade object detection.

Step.3: The eyes region will be recognized using the CNN model as to check whether the student is in an engaged or disengaged state.

Step.4: If the system recognizes the state of the student as disengaged, the buzzer sound starts.

3.2 Dataset

To develop a high-performance model for student engagement detection, the MobileNet model was trained on a publicly available dataset from Kaggle [10]. This dataset consists of 14360 eye images: 3,828 *close look*, 3,457 *forward look*, 3498 *left look*, and 3,577 *right look*. This study used 8044 eye images as a subset of the dataset: 1,156 *close look*, 1,706 *left look*, 1,725 *right look*, and 3,457 *forward look*.

3.3 Preprocessing

Dataset Preprocessing was done by selecting 4,587 eye images from three classes (*close, left, right look*) to be merged into one class named "*disengaged*" while the other class, named "*engaged*", contains 3,457 *forward eye look* images. The eyes region was cropped from the face images by the database creator [10]. The dataset was then separated into three categories; 6842 for training, 805 for validation, and 397 for testing. The CNN model needed the data to be normalized, such as rescaling all image's parameters from [0, 255] pixel values into binary [0,1] by using a preprocessing class from Keras.

3.4 The Architecture of Pre-training Model

This study implemented CNN pre-training models for student engagement detection based on eye gaze. Deep learning provided a transform learning strategy to use the architecture and weight of the model which was trained on a large image database for the image classification task. MobileNet [11] is one of the most popular image classification and recognition models, which is trained on the ImageNet database. ImageNet database [12] was collected by Google, consisting of 14 million object images [11]. The fine-tuning is used to custom the model for the specific image classification task. Before deciding to use the Pre-training Models, the researchers considered using a trained model on an image dataset similar to eye image dataset. Then they used the model network architecture to extract the features from the image dataset. After that, they randomly initialized all the weights and trained the model according to the *engaged* and *disengaged* dataset. The input eye image size was (64*64*3), and two dense layers were added to the top of the model architecture, as shown in Fig. 2. The first layer consisted of 1024 neurons with a Rule activation function. A dense layer was followed by the dropout layer (0.4). The softmax function was the second layer used for the output prediction of two classes, namely, *Engaged/Disengaged*. Figure 2 shows the MobileNet architecture with SSD layers. For reducing the error during the training, the RMSprop optimizer was used. In

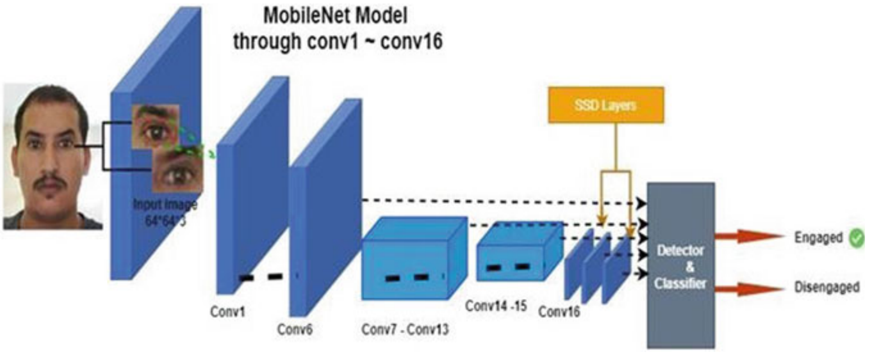


Fig. 2 Proposed model architecture

machine learning, some strategies were used to avoid overfitting such as dropout and batch-normalization [13]. Dropout randomly ignored some neurons during training. While training the model, the validation loss was monitored by the early stop strategy. If the loss does not improve, the model's training will be stopped and save the best performance.

4 Experimental Results

4.1 *Training and Testing*

The MobileNet model was trained on the cloud using the Google Colab environment by implementing Tensorflow and Keras. The model was trained with 100 Epochs, and the Batch size was 128. The early-stopping strategy monitored the validation loss and stopped on 27 epochs while the model's best performance was saved. The MobileNet model achieved 99% accuracy on training data and 99% on the validation data. Figure 3 shows plot, performance, training, and validation accuracy while Fig. 4. shows the training loss and validation loss.

4.2 *Classifier Evaluation*

The model's performance was measured using some metrics such as validation and testing accuracy, sensitivity, and specificity. *Sensitivity* measured a positive prediction in which it was defined in our model as *disengaged*. *Sensitivity* was calculated by the following formula:

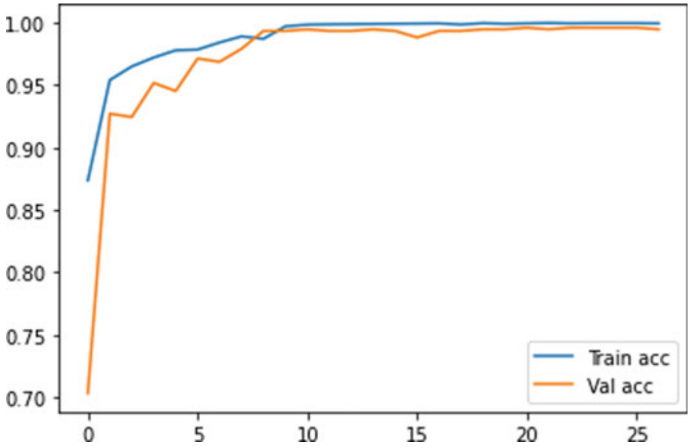


Fig. 3 Plot of model accuracy

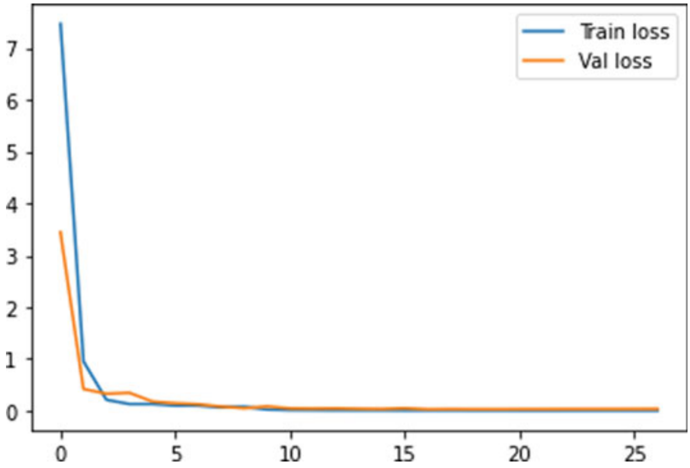


Fig. 4 Plot of model loss

$$Sensitivity = \frac{True\ positives}{True\ positives + False\ positives} \tag{1}$$

The model identified eye images as “Disengaged Negative” with 0.99% accuracy.

$$Sensitivity = \frac{True\,negatives}{True\,negatives + False\,negatives} \tag{2}$$

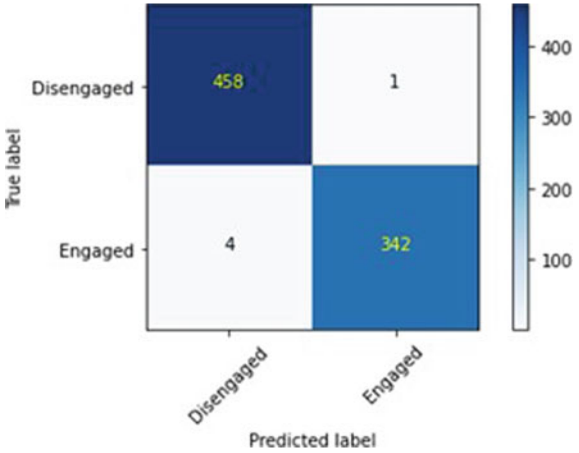


Fig. 5 Plot of confusion matrix

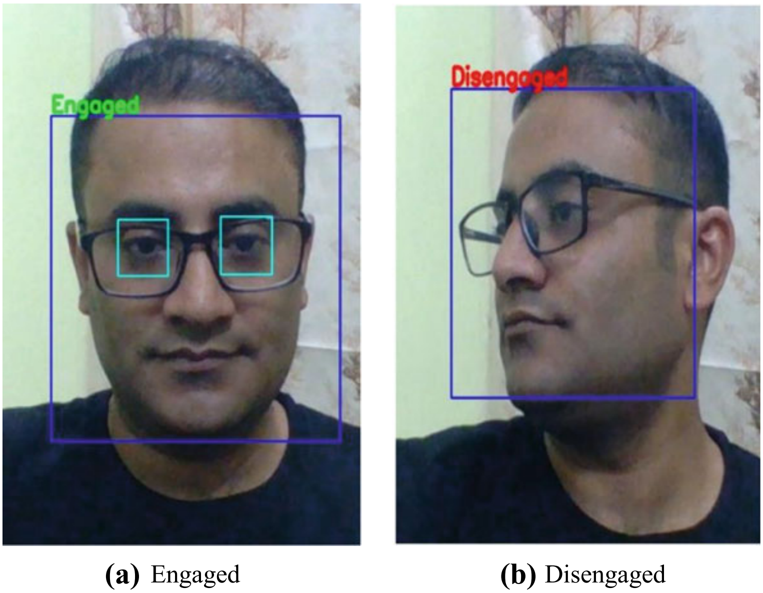
4.3 Confusion Matrix

True positives (TP): the model could correctly predict 458 eye images as Disengaged. True Negatives (TN): the model could accurately predict 342 eye images as Engaged. False Positives (FP): the model could falsely predict one eye image as Disengaged, but the eye image was Engaged. False Negatives (FN): the model falsely predicted four eye images as Engaged, but eye images were Disengaged. The details of the Confusion Matrix are shown in Fig. 5.

4.4 Real-Time engaged detector

The Viola–Jones method is an object detection framework that is a fast and high accurate method for face detection in real time. This method analyzes the pixels in a frame of full-frontal faces. The advantage of the Viola–Jones method [14] is that it is applied in real time with low false positives. The Viola–Jones method was applied to detect the face and eyes in each frame of the video. In this system, the student's face was detected, and the eyes' region was located for CNN model to predict the student engagement state. The MobileNet model has two binary classes to predict the student state, *Engaged* and *Disengaged*.

In the experimental work, the result of training MobileNet classification was saved in Hierarchical Data Format (hdf5) to be used with Haar Cascade of the eye and face OpenCV. The simulation of the system: (i) student sits in front of the laptop, (ii) the video is captured by webcam, (iii) Haar Cascade object detection, which works in real time, detects the face and locates the eyes, and (iv) the CNN model predicts the eyes states of the student. When the student eyes are open and focused on the



(a) Engaged **(b)** Disengaged

Fig. 6 The Real-Time Engaged Detector

screen, the state is engaged and a green notification "*Engaged*" appears above the boundary box of the face, as shown in Fig. 6a. A disengaged state happens when the student looks out of the screen or closes his/her eyes, which means the student is disengaged. In such a case, a red notification word "Disengaged" appears above the face boundary box as shown in Fig. 6b. and a buzz sound starts.

5 Discussion

Online learning gives people the opportunity to learn from their favorite places despite place constraints. However, recognizing the attention of the student is a challenge in online teaching. This study provides a high accuracy classification of student *engagement* and *disengagement*. The MobileNet model achieved 99% accuracy on validation data, 98% sensitivity, and 99% specificity. These results encouraged the authors to develop a system named Real-time Detection of Student Engagement: Deep Learning-based System, Haar Cascade object detection, and OpenCV. Compared to existing systems [3, 4] who used eye gaze with emotions to detect the engaged level of the student, this study is based only on the eye gaze which gives a high performance on real-time testing which proves that the engagement of the student can be detected using the eye gaze more efficiently with high speed.

6 Conclusion

In education, engagement is essential to get the maximum benefit of learning. This paper developed a real-time system based on nonverbal communication using deep learning to detect and improve student engagement. This system can detect the states of students' attention during online classes which also enables the teachers to monitor the class activities. The eye image classification is based on the MobileNet model, which achieved 99% accuracy on validation data. The system's advantage is that it makes use of the basic E-learning Hardware, i.e., a laptop with a built-in camera or desktop computer with a webcam. This system can be adopted in the future to analyze the behavior of participants in online video meeting applications. The future work of this application can be used in the cloud, and it should be secure by using blockchain innovation to preserve all data and give a decentralized information base [15].

References

1. Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: a review. *Smart Learning Environments, 6*(1), 1.
2. Z. A. Taha Ahmed and M. E. Jadhav, "A Review of Early Detection of Autism Based on Eye-Tracking and Sensing Technology," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 160-166, https://doi.org/10.1109/icict48043.2020.9112493.
3. Sharma, P., Joshi, S., Gautam, S., Filipe, V., & Reis, M. J. (2019). Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. arXiv preprint arXiv:1909.12913.
4. Nezami, O. M., Dras, M., Hamey, L., Richards, D., Wan, S., & Paris, C. (2019, September). Automatic Recognition of Student Engagement using Deep Learning and Facial Expression. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 273-289). Springer, Cham.
5. Núñez Fernández, D., Barrientos Porras, F., Gilman, R. H., Vittet Mondonedo, M., Sheen, P., & Zimic, M. (2020). A Convolutional Neural Network for gaze preference detection: A potential tool for diagnostics of autism spectrum disorder in children. arXiv e-prints, arXiv- 2007.
6. Rodríguez, C. V., Lavalle, M. M., & Elías, R. P. (2015, November). Modeling student engagement by means of nonverbal behavior and Decision trees. In 2015 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE) (pp. 81 -85). IEEE.
7. Hashemi, M., Mirrashid, A., & Beheshti Shirazi, A. (2020). CNN-based Driver Drowsiness Detection. arXiv, arXiv-2001.
8. Kaur, A., Mustafa, A., Mehta, L., & Dhall, A. (2018, December). Prediction and localization of student engagement in the wild. In 2018 Digital Image Computing: Techniques and Applications (DICTA) (pp. 1-8). IEEE.
9. Thomas, C., & Jayagopi, D. B. (2017, November). Predicting student engagement in classrooms using facial behavioral cues. In Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education (pp. 33-40).
10. https://www.kaggle.com/abhibasavapattana/eyegaze-classification-using-cnn/data
11. https://keras.io/api/applications/
12. http://www.image-net.org/

13. Chakraborty N., Dan A., Chakraborty A., Neogy S. (2020) Effect of Dropout and Batch Normalization in Siamese Network for Face Recognition. In: Khanna A., Gupta D., Bhattacharyya S., Snasel V., Platos J., Hassanien A. (eds) International Conference on Inno- vative Computing and Communications. Advances in Intelligent Systems and Computing, vol 1059. Springer, Singapore. https://doi.org/10.1007/978-981-15-0324-5_3.
14. g, Y. Q. (2014). An analysis of the Viola-Jones face detection algorithm. Image Processing On Line, 4, 128-148
15. A. M. Al-madani and A. T. Gaikwad, "IoT Data Security Via Blockchain Technology and Service-Centric Networking," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 17-21, https://doi.org/10.1109/icict48043.2020.9112521.