International Journal of Research in Engineering, Science and Management
Volume 4, Issue 6, June 2021
https://www.ijresm.com | ISSN (Online): 2581-5792

Virtual Meeting System Architecture with Gesture Control, Attentiveness Tracker and Automatic Attendance Recording

Vatsal Gupta[1], Puru Malhotra[2*], Saurabh Gautam[3], Vinay Kumar Saini[4]
[1,2]*Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, New Delhi, India*
[3,4]*Assistant Professor, Department of Information Technology, Maharaja Agrasen Institute of Technology, New Delhi, India*

***Abstract*: As a result of the COVID-19 global pandemic and its concerned lockdown policies, a number of activities have been forced to shift from traditional physical modes to a complete online distanced virtual mode. This has demanded education, offices, conferences and social meets, among others to abruptly adopt virtual communication tools and online meeting platforms. Here, we present a hybrid system architecture with functionalities to enhance the experience of an online virtual meeting. We focus on improving usability through automation provided by Artificial Intelligence and Computer Vision. Our study not only depicts the important pipelines to support current remote experiences in this pandemic but also provides a foundation towards building these systems in a more adaptive and interactive manner for the future even after the pandemic.**

***Keywords*: attendance recording, COVID-19, emotion detection, facial recognition, gesture recognition, virtual meeting.**

1. Introduction

Virtual Meetings have become a hot trend among organizations recently amid the COVID-19 pandemic and virtual collaborations are expected to grow in the future [1]. It's not just the pandemic but natural calamities such as floods, hurricanes, cyclones and earthquakes often disrupt our day-to-day activities and commuting to school or work becomes a challenge. Sometimes, they lead to closures or long traffic jams which, if happening frequently, can hinder learning and growth in the long run. So, there is a need to find a strong alternative that can help us tackle such situations where we find it difficult to be physically present. The current situation of the pandemic challenged us to completely shift to a virtual mode forcibly and abruptly. This led to a mass adaptation of online virtual platforms for collaborations. Virtual Meetings have shown their use in a variety of sectors including education, industry, fitness and healthcare. As more and more people join and make use of this technology, the more its benefits and weaknesses come forward.

A SWOC (Strengths Weaknesses Opportunities and Challenges) analysis of online learning conducted under a research [2] portrayed the time and location flexibility of this system both as a boon and a bane. These factors are a big strength for obvious reasons, but they are fragile and create problems. The freedom and flexibility have led to the non-serious behaviour of the attendees without someone to keep a check on them. This is a big problem reported among students attending online lectures, where they are frequently found drowsy and inattentive. Another major issue is the use of proxies to register fake attendances by taking the advantage of sitting distant.

In a survey [3] conducted among working professionals and non-working professionals who have shifted their office/schools on an online meeting platform, users felt the need for a facilitator who can make the system easier to understand and use. In some follow-up questions, 90% of them supported the idea of using Artificial Intelligence in a virtual meeting system to support the participants. 70% of them suggested the need for gesture features to make the experience more automated with minimal input.

There is an imbalance created by the inadequate compatibility between the design of technology and component of psychology [2]. By exploring the scope of innovation and digital development [2] our research proposes the architecture of a Virtual meeting system making the use of Artificial intelligence and computer vision. Making adequate customisations to the existing systems, the proposed collaborative system is supportive, reliable and stable.

The interaction system architecture in the paper is further composed of three parts:

- Attendance Recording System
- Gesture Control
- Attentiveness Tracker

The main structure of the interactive system is shown in Figure 1. Since hand gestures are able to express enriched information, we make use of a hand gesture recognition system that can be customised by the user to provide personalization control over one's meeting systems. They can choose a gesture, its functionality and provide samples to train a Neural network in real-time. Making the use of facial detection and

*Corresponding author: purumalhotra99@gmail.com

identification techniques, we construct our Attendance recording system. The system takes in samples of face snapshots for a buffer of 5 seconds during the creation of an ID/profile to study a particular user's face and associates it with other information that can act as a primary key, say roll number, to identify a particular user. During each meeting, the system studies the input video frames and verifies if the particular person associated with the user account is physically present or not and marks the attendance automatically along with the respective time frames of activity. There are a lot of facial features that can indicate the attentiveness of a user or not. There are various researches going on to develop systems for emotion detection by facial traces. We make use of an eye tracker and ensure if the user is attentive or not by studying the pattern of the user's eyes during the meeting. The use of real-time data for training also helps to calibrate the system to the configuration of hardware and the physical condition of the environment around a particular user for better results.

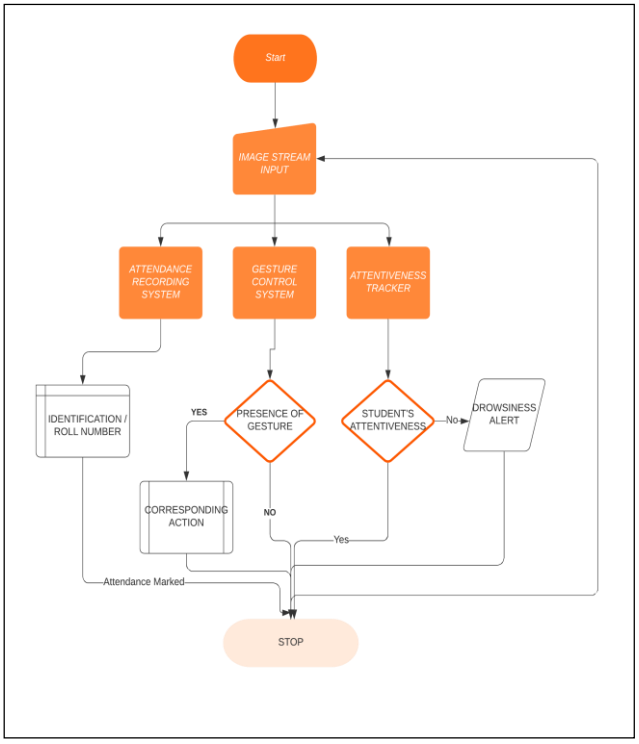

Fig. 1. Architecture of the proposed system

2. Related Work

With the rise in COVID-19, online meeting systems like Skype and Zoom have gained popularity. Researchers and Scholars have also explored ways through research projects to improve these systems which are expected to be a part of our lives for quite some time. Commonly used online meeting systems provide functionality like sharing audio, video and chat to provide a virtual environment. There are efforts made to investigate other features that can be introduced to make virtual meeting systems more effective and useful.

Coventry University developed a virtual agent [4] that assisted human facilitators in a meeting. Introducing personal agents in a meeting software has been a hot research area [5-9]. The use of such assistance, though seeming tempting, has been restricted by the limited vocabulary on which the systems were trained.

Other researchers use methods to provide structure and hence adding layers of automation to support these meeting systems. In an extended effort, Roman F. et al. developed a server-based meeting system that provided immediate feedback based on a user's participation and activity during a meeting. They also provided the extraction of measurable parameters easily translated into CSV or XML formats for statistical analysis [10].

In their work [11], M. Mohanty and W. Yaqub exploited the online meeting systems to sufficiently identify and solve security concerns over them. They used a photo resonance non-uniformity-based camera for digital authentication of a user in order to ensure privacy and security.

Each meeting is unique and can present a different combination of purpose, culture, characteristics and structure. Developing a system with an essence that could fit and support every type of meeting is a huge milestone which is never reached as different kinds of virtual meeting situations and demands continuously emerge with time.

3. Proposed System

A. *Attendance Recording system*

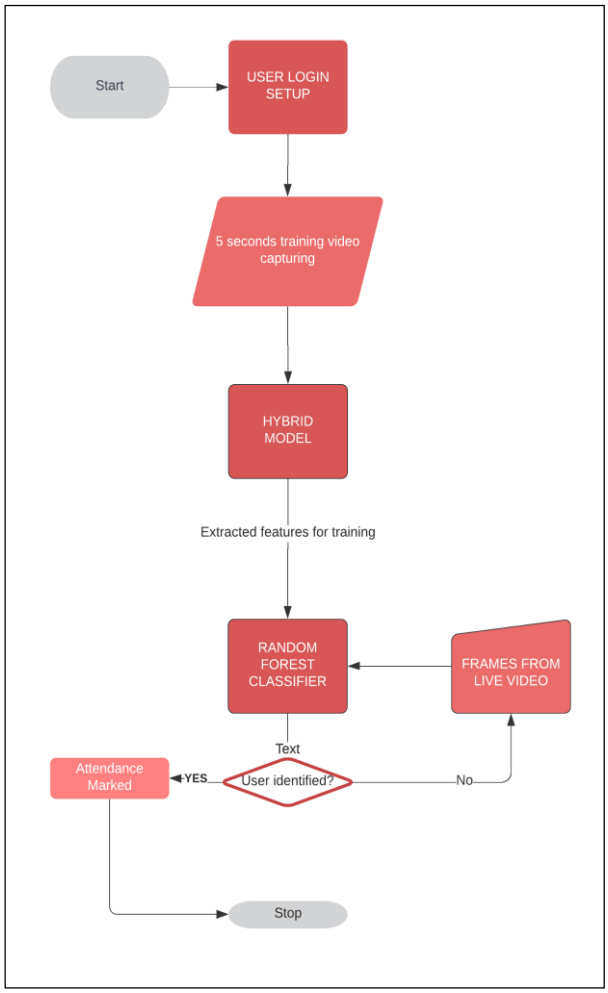

Fig. 2. Working of attendance recording system

The real-time attendance recording system is realized by analyzing the results described in the paper [12] which were further extended and implemented to fine-tune the hyper parameters and ensure repeatability of results under the case of a variable number of output classes. The use of variable number of output classes comes in handy due to the number of members in a meeting varying with time and use. So, we can easily add or subtract new members form our database.

The structure of the attendance recording system is as shown in the figure 2. Each of the components is individually explained below in detail.

1) Feature Extraction

Feature extraction is finding the set of parameters that define an object uniquely and precisely. The features contain the information required to distinguish between the classes [13]. To limit the amount of training data required in real-time, we train a model on a Kaggle dataset [14] that helps in setting the weights of our model for feature extraction. The dataset contains 17534 images belonging to 105 different classes.

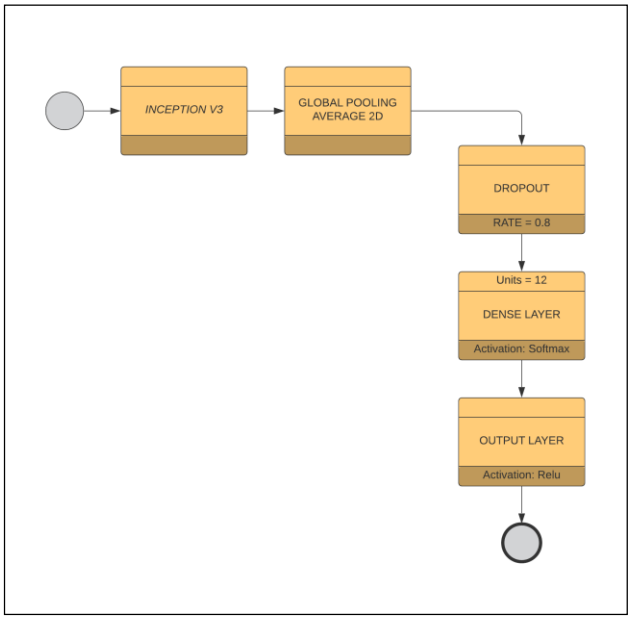

Fig. 3. Hybrid model architecture

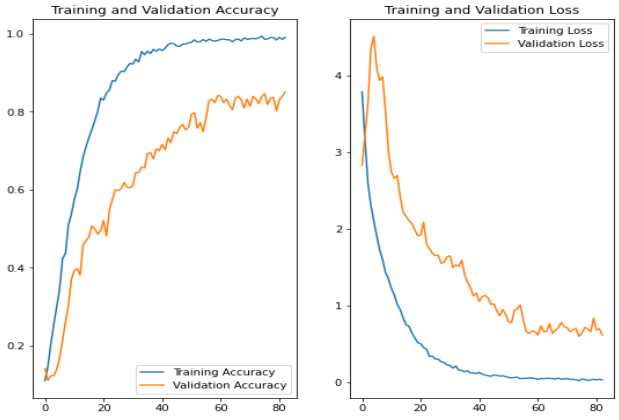

Fig. 4. Accuracy and Loss curves for hybrid model training

The model we propose for feature extraction, which we refer here as hybrid model, is a pre-defined Inception V3 model [15] modified by appending a Global average pooling layer, a dropout layer, a dense layer followed by an output layer with relu activation.

Training of the hybrid model was carried out using a Categorical cross entropy loss, Adam optimizer and Accuracy as metrics. The model compiled with the loss and accuracy functions as shown in figure 1.

2) Random forest classifier

The Random Forest classifier was implemented to process the output features extracted from the trained model. Using a random forest classifier over a Neural network for object identification is more efficient because of simplification of the model to be finally trained and use at the user's end. The adopted approach stands faster and less resource consuming without compromising on accuracy. An experimented approach was used for hyperparameter tuning where we varied the n_estimators, max_features and bootstrap values, keeping all other hyperparameters default.

Table 1
Accuracies for different hyper parameter combinations

Max_features	Bootstrap	Relative Performance Metrics (%)
sqrt	True	49.6051
Log2	True	47.219
sqrt	False	51.6925
Log2	False	49.3733

As it can be seen from table 1, the combination of max_features as 'sqrt' and bootstrap as False performed with an edge of ~2.9% over the other combinations for an average for 2600 test cases. This combination was used to find the optimum value for n_estimators which was concluded to be 198 after averaging over 50 test cases with a varying number of output classes.

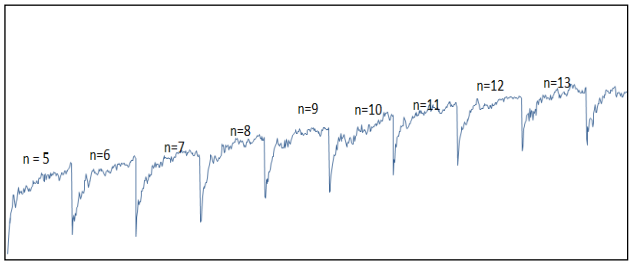

Fig. 5. Trend of accuracies by varying n_estimators (n specifies the number of output classes used in the experiment phase)

Figure 5 shows the plot of accuracies by varying n_estimators after changing number of output classes dynamically and keeping other parameters constant. As it can be seen that the pattern shows a consistent trend for accuracies despite changing the number of classes, our experiment ensures repeatability of results. This comes beneficial in the deployment when a number of participants join and exit during the meeting changing the number of participants at each instance for different meetings. This proves the stability in terms of a single system being efficient in dynamic real-time scenarios.

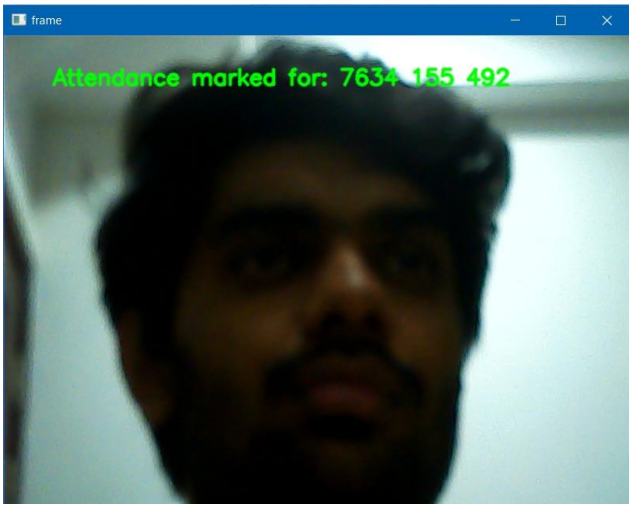

Fig. 6. Demo showing working of attendance system

It is difficult to gather a large dataset to train our models for image identification when they are deployed for use in real-time. This leads us to use the pre-trained model as described in the previous sub subsection for reliable feature extraction. The use of random forest classifier yielded an accuracy of 83.41% with the calculated optimum values for parameters with the use of a small training sample size.

B. Gesture Control

Using the advancement in object identification and Computer vision, we are able to identify and categorize images in real-time. The real-time hand gesture tracking is realised by passing video frames for prediction into a Neural network pipeline designed by us which gets trained on image taken in as an input from the user in real-time. A class identifier is assigned to each category of image which is used to classify samples when the system is in use. We adopt the MobileNet(v1) [16] neural network to recognize hand gestures and modify it to set up the number of classes the image is classified into as variables that can be modified in real-time as per the samples given by the user.

The structure of MobileNet includes depth wise convolutional layers, pooling layer, fully-connected layer and Softmax layer. Each depth wise convolution layer consists of a depthwise convolution and a point wise convolution.

The users are given choice to select as many as 10 different hand gesture categories where they can choose a gesture and its functionality. To improve the accuracy of the system we provide a framework to input sample images and train the model in real-time.

An experimental study conducted by us concluded that an average of 11-13 images per gesture is sufficient enough to train a robust model for prediction and the system was tested giving accuracies above 94% varying the number of output classes from 2 to 25.

C. Attentiveness Tracker

The paper adopts the Dlib Library for python and makes use of its get_frontal_face_detector() and shape_predictor() functions to track the face landmarks with good robustness. The use of these functions helps to visualize a face in the form of 68 landmark points over the face.

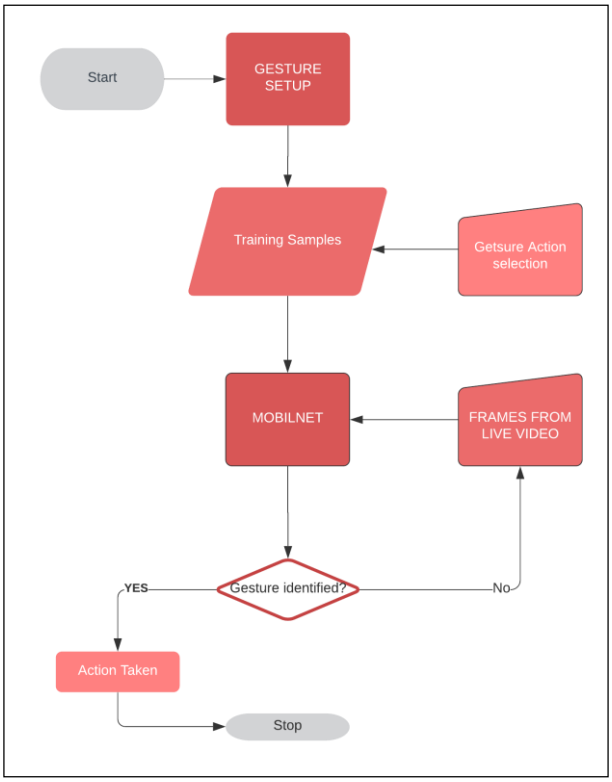

Fig. 7. Working of gesture control system

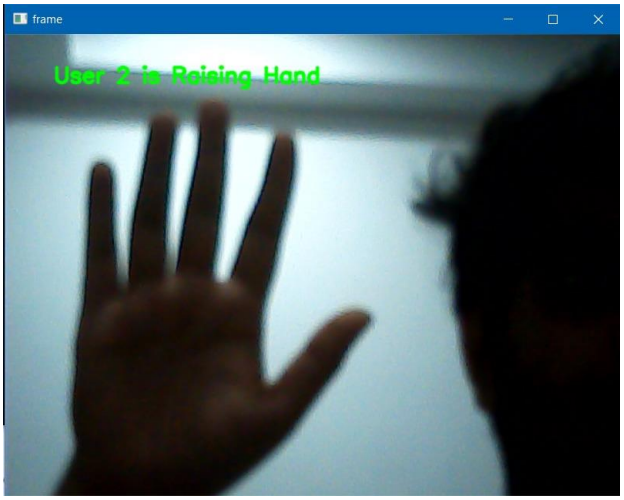

Fig. 8. Demo showing working of gesture control system

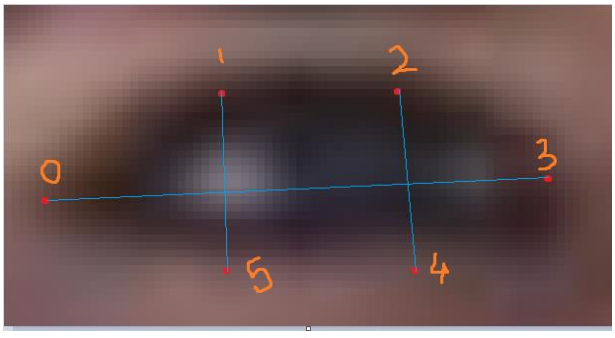

Fig. 9. Eye track points for aspect ratio calculation

Using the Facial landmark points, the region containing the eyes is identified and is analysed on the basis of markings as shown in the figure 9.

The aspect ratio of an eye is calculated as:

$$Aspect\ Ratio = \frac{A + B}{2 * C}$$

where A, B and C represent the Euclidian distance between the points 1 and 5, 2 and 4, 0 and 3 respectively from figure 9.

Fig. 10. Demo showing working of Attentiveness checker (1)

Fig. 11. Demo showing working of Attentiveness checker (2)

The aspect ratio for both the eyes is calculated with each passing frame and is compared with a Threshold value to decide the level of activeness of a user. To declare a drowsiness alert, the aspect ratio for both eyes needs to be less than the threshold of 0.3 continuously for 48 frames. The buffer of 48 frames is taken to skip false alerts due to blinking of eyes based on values of average time taken to blink an eye. This can be seen in figure 5 and 6 where a drowsy alert is thrown when the Aspect ratio goes below the decided threshold.

4. Software Applications and Future Takeaway

The software application of the research is acquired by synthesising an Online meeting system as a base software that allows multiple users to connect in a single virtual room via many-to-many audio videos streams. The discussed modules can be extended to form individual functionalities which allow additional features to make the experience of a virtual meeting convenient, interactive and automated. They can be incorporated in the form of API's or web extensions and mounted onto existing online meeting applications.

5. Conclusion

The paper carries out research on forming an architecture for Online meeting system. Providing additional functionality in the form of attendance recording, attentiveness tracking and gesture control we tend to ensure a 2-way quality interaction and increase the ease of usability for the user.

References

[1] L. L. Martins, L. Gilson, and M. Maynard, "Virtual teams: What do we know and where do we go from here?" Journal of Management, vol. 30 (6), pp. 805–835, 2004.

[2] Dhawan, Shivangi. "Online Learning: A Panacea in the Time of COVID-19 Crisis." Journal of Educational Technology Systems 49, no. 1 (September 2020): 5–22.

[3] A. G. Nanos and A. E. James, "A Virtual Meeting System for the New Age," 2013 IEEE 10th International Conference on e-Business Engineering, 2013, pp. 98-105.

[4] P. Thompson, A. James, and E. Stanciu, "Agent based Ontology Driven Virtual Meeting Assistant", Future Generation Information Technology, LNCS, Vol 6485, pp. 51-62, 2010.

[5] B. Campagnolo, C. A. Tacla, C. A., E. C. Paraiso, G. Y. Sato, and M. P. Ramos, "An architecture for supporting small collocated teams in cooperative software development," in Proc 13th Int'l Conf. on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2009, pp. 264-269.

[6] J. P. A. Barthès, "OMAS - a flexible multi-agent environment for CSCWD," Future Generation Computer Systems, 27(1), pp. 78-87, 2011.

[7] S. A. Rahman, D. Yadav, P., Agerwal, P., and P.S. Bisth, "Multiagent Knowledge Management Architecture," Journal of Software Engineering and Application, vol. 5, 2012.

[8] K. Sugawara, Y. Manabe, Y., and S. Fujita, "Mobile symbiotic interaction between a user and a Personal Assistant agent," in Proc. 11th International Conference on Cognitive Informatics & Cognitive Computing (ICCICC), IEEE, 2012, pp. 341-345.

[9] E. C. Paraiso, and A. Malucelli, A. "Ontologies Supporting Intelligent Agent-Based Assistance", Computing and Informatics, vol. 30 (4), pp. 829-855, 2012. 10] M. N. Huhns and M. P. Singh, "Personal assistants", Internet Computing, IEEE, 2(5), pp 90-92, 1998.

[10] F. Roman, O. Mubin, and P. Dillenbourg, "Reflect World: A Distributed Architecture for Meetings and Groups Evolution Analysis," in Proc. of the 10th Int'l Conf. Collaboration Technologies and Systems (CTS), 2012, pp. 389 – 396.

[11] M. Mohanty and W. Yaqub, "Seamless authentication for online teaching and meeting," 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), 2020, pp. 120-124.

[12] Vatsal Gupta and Saurabh Gautam, "Comparative Analysis of Machine Learning Algorithms with and without Feature Extraction", International Journal for Modern Trends in Science and Technology, 6(12): 235-239, 2020.

[13] G. Kumar and P. K. Bhatia, "A Detailed Review of Feature Extraction in Image Processing Systems," 2014 Fourth International Conference on Advanced Computing & Communication Technologies, 2014, pp. 5-12.

[14] https://www.kaggle.com/hereisburak/pins-face-recognition

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826.

[16] Andrew G. Howard et. al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications".