Proceedings of the Third International Conference on Smart Electronics and Communication (ICOSEC 2022)
IEEE Xplore Part Number: CFP22V90-ART; ISBN: 978-1-6654-9764-0

Exam Proctoring Classification Using Eye Gaze Detection

Jagendra Singh
School of Computer Science Engineering and Technology, Bennett University,
Greater Noida, India
jagendrasngh@gmail.com

Reet Aggarwal
School of Computer Science Engineering and Technology, Bennett University,
Greater Noida, India
e19cse322@bennett.edu.in

Shubhi Tiwari
School of Computer Science Engineering and Technology, Bennett University,
Greater Noida, India
e19cse346@bennett.edu.in

Vinayak Joshi
School of Computer Science Engineering and Technology, Bennett University,
Greater Noida, India
e19cse024@bennett.edu.in4

2022 3rd International Conference on Smart Electronics and Communication (ICOSEC) | 978-1-6654-9764-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICOSEC54921.2022.9951987

***Abstract*—Currently, as there is a spike in the number of examinations being conducted online, there is also an increasing need for proctors, a resource that soon got overwhelmed. Proctors are supposed to monitor each and every examinee individually to ensure that the code of conduct is being followed by the examinee, contrary to the offline examination system in which a large group of students are overlooked by a handful of invigilators. But even then, the rate of unfair means being opted by the examinees in online examinations is quite higher in comparison to offline examinations. Thus, making the current process extensive, tiresome and wasteful. To make the process more coherent we came up with Eagle Eye. This system would be implemented and made efficient with the use of the technique of Eye gaze detection using Artificial Intelligence and Machine Learning. At the beginning, of the examination, the examinee would be required to undergo a calibration test to determine the situations around the examinee and to set a designated border-box area into which the examinee would be required to get their eyeball movements tested. This data gathered from the detection would be used to examine whether or not the examinee is looking outside or inside the box, which in turn would be used to classify the data as fraudulent and/or fair respectively. If detected to be fraudulent by the system, an alert will be sent to the examiner and the examinee both so that further actions can be taken according to the mandate. To achieve the best prediction results, we have made our own dataset with the help of volunteers so as to obtain unfiltered and more authentic samples to work with.**

Keywords—Calibration, Cellular neural networks, Classification Algorithms, Data Preprocessing, Face detection, Feature Extraction, Image Classification

***Impact Statement* —** Whilst the shift of the education sector into the online mode has been rather swift than not. Nevertheless, there are a few areas in which some modifications or improvements could be made to improve efficiency and decrease the need for enormous human capital. One of such areas is Online Examinations, to be specific, the proctoring method. Even with the presence of individual proctors with every single examinee, the rate of unfair means being opted by the examinees in online examinations is quite higher in comparison to offline examinations wherein a single proctor keeps a track of a group of examinees. Thus, making the current process extensive, tiresome and wasteful. To make the process more coherent we came up with Eagle Eye. A system which would use less financial and human resources and capital investment.

I. Introduction

In today's time, wherein everything is dependent on technology to operate successfully be it any sector, education, work, or retail, it makes sense to develop something that helps us to move ahead of the situation at hand and make these times easier and work and education, more effective. Keeping this ideology in mind, the team has come up with *Eagle Eye.* As the name suggests, this system or method will act as a third eye, a surveillance tool that can be used in various sectors of the economy. But we limit our focus to the Education Sector for the scope of this paper. We believe that this paper when completed successfully can have many more applications in the fields of Crime, Marketing, Automobile among many others.

The motive behind the paper was simple yet impactful, we intended to make a paper with the purpose of enhancing the online examination and the education system. The Core Problem that we could point our finger at in this system was the lack of an inviolable *proctoring method.* Due to the lack of an absolute method, there is a huge gap left for the students to indulge in unfair means without any hassle. Even though unlike the traditional offline examination system wherein a single invigilator monitors an entire group of students, in the Online mode, a singe proctor or examiner is assigned to every individual student, which requires a huge financial investment from the organisation. Even after the presence of a proctor it becomes difficult to detect if the examinee is par-taking in any unfair means like having a notepad or using their mobile phones to search for the answers, or any other practice that doesn't cause any noise or changes in the video making it harder to monitor.

As a consequence of the poorly achieved results, the current process, burdens the organizations that are conducting these examinations with the responsibility of a result that is unfair and non-equitable while also using a large amount of the human resource and the capital needed to employ them. As a result of the above-mentioned factors, there is a huge void left in the system of Online Examinations and the proctoring method currently in use. Consequently, to provide a more rigid method for the examinations we came up with *Eagle Eye.*

We believe, *Eagle Eye* would be an effective solution to this problem and could be used to fill the void. Using Eagle Eye, would help the proctors or the examiners to take the unnecessary load off of themselves during examinations. This could also help the organisations conducting the examination to execute the said examinations in a more fair and equitable manner all the while using less human capital, making this this system less wasteful than the one currently being used in the industry.

To provide a basic and understanding into how the system would work, in layman terms we can say that it could help the examiner by easily detecting if an examinee is time and again looking outside of the designated screen area of the device which can either be a computer or laptop screen or a mobile phone screen. Following the detection, it will be evaluated if the examinee has taken part in any dishonest practices by looking outside the mandated screen area, which will be designated at the start of the examination using the process of *calibration.*

This designated are will be the area that will be used to track and evaluate the examinee's eye ball movement to gather the required data for the evaluation. If a breach of code of conduct has been monitored, an alert would be immediately sent to the examiner or administrator in charge and thus he/she could take the required actions. As examinees can easily use the Internet to search their questions by switching their tabs all the while looking inside the designated area, thus their screens would be fixed before the start of the examination and any examinee switching their tabs could result in disbarment from the examination.

As Eagle Eye's proctoring process would be digitalised, it would mean that there would be less requirement of actual proctors and a single examiner or proctor could handle multiple examinations simultaneously with this system in a much easier and efficient manner.

II. Related Work

Gaze Detection and Eye Tracking are not relatively new domains of research for scientists and researchers in the field, they have been the topic of ample number of research papers over the past years. In 1849, Du Bois-Reymond discovered a relationship between electrode potential on the surface of the human skin and the movements of the eye [1] and since then, As long ago as 1935, there have been papers that suggest galvanometric recording and detection of eye movement, suggesting that using electrodes placed on temples and further connected to a galvanometer, movement of the eye can be detected using potential differences [2].

Technology has evolved a lot since 1935, and since the relatively recent emergence of computer vision technology, better and safer eye tracking and gaze detection methods have been developed. These modern methods usually rely on the images of eyes or faces captured by a digital camera. in quite a few cases, the camera can be the webcam of the device that the user is using as seen in this paper from 2010 [3], or Webgazer, an eye-tracking library developed by Brown University [4]. These modern methods can usually be applied through two categories of devices, firstly, through a head-mounted tracker, that is often mounted on glass frames, and secondly, through a remote device that is placed at a certain distance from the user [5], an example of this are the webcams that we discussed earlier.

The modern, computer vision-powered methods are further divided into these two kinds of approaches, the first type of approaches are the ones that use specific sensors and devices (in addition to the camera) in order to process geometric data and information about the image, such as eye corners, depth, and contours to detect the gaze of the user, these kinds of approaches are usually slightly different for 2D dataset and 3D dataset. Technically, it is much easier to process 2D data into geometric data in comparison to 3D data, consequently, 3D data is often corrected for its depth, shadows, etc. and by using highly specific hyperparameters such as the viewing angle of the user and the radius of the cornea, it is further according to for better processing of geometrical features.

The second kind of approach includes appearance-based methods that directly uses mapping functions to understand the image, unlike geometric approaches, they don't need specific sensors and devices to detect several geometric features, instead, they completely rely on the information preprocessed from the image such as deep features or image pixels [6], in order to detect the gaze of the user. Because of their less reliability on multiple devices, Appearance Based methods are going to be the center point of our study because they can be effectively used for generating data for detection of eye gaze using Machine Learning and Deep Learning.

Ever since the mid-90s, there have been various attempts at using ML and DL for detecting gaze, in this research from 1994 [7] by S. Baluja and D. Pomerleau, they have discarded the traditional gaze tracking method of Recording Specular Reflection, and have used Artificial Neural Networks, and in order to do so, they collected 1000 vertical and 1000 horizontal images as a training dataset, and in order to estimate the exact direction of the user's gaze, the pixels of the image are the inputs to the artificial neural network, and they have used only one hidden layer that if further split according to data on the x-axis and the y-axis [7].

This research spearheaded the research on the use of Neural Networks for gaze detection and further became a model and an inspiration for various other researches in the domain. But as research in this field increased, a problem that was often encountered was the problem of the movement of a head during gaze detection from a video, but a research paper from 2006 [8] Bayesian differential pose tracking, which uses past keyframes to estimate the current frame by using increments between the neighboring images, and using this estimates the position of the head. This research, and others like it, inspired multiple research papers, such as this one from 2008 [9,10] that used the movement detection methods of Bayesian differential pose tracking, and created an incremental learning method that worked to detect gaze.

Discussing the dataset, as discussed earlier, the dataset can be divided into images taken by a head-mounted camera that takes an image of the face, or in some cases of just the eyes, in other cases, the dataset is generated from an image or a video taken from a distance. For example, in this dataset from research at Columbia [11], the users, or the subjected, have been requested to place their chin on a surface and rest their head, and further, they are asked to focus at a point on the wall in front of them, and images are collected to record their gaze.

While this dataset does give quite high-quality images that can be used for processing, these images are not very diverse, as there is little to no head movement from the subjects. An improvement of this has been done lately where subjects have been asked to look at several targets on the screen of a monitor, or a mobile phone and their information was recorded by webcams or front cameras, for example, in this research paper [12] twenty unique target aims were shown on a laptop screen in a session for a period of time and videos of the subject was recorded. For data collection in our research, this [13,14] is our major inspiration, we plan to give students a sample test and record their videos while they're doing it.

In 2015 a study by Zhang, Sugano, Fritz, and Bulling [10], proposed the first Convolutional Neural Network based approach towards gaze detection, and the performance of this simple gaze estimation method surpassed most research done using conventional appearance-based methods, and even though its more economically straining its effectiveness makes us overlook that. In a nutshell, even though the research on gaze detection has been going on for a long time, there is still a lot of scope, especially due to the emergence and popularity of Computer Vision, and Deep and Machine Learning, in recent years, a lot of research is yet to be done.

III. METHODOLOGY

The Flow of our work "Eye Gaze Detection" is reflected in the fig. 1. The flow is divided into two major parts- Part 1 being, Model Training and the other being, Realtime Testing. The first part that is the Model Training is the core part in where we will be building a model and deploying it for the real time application usage using various techniques [15,16]. The second part is the Application use, that is, the real time use of the deployed and fully optimized model, This the part with which the user will be interacting with and using.

While building the model, the first thing we will needing is a dataset to work with. To optimize the model a well featured dataset will be required that will further go under processing using various pre-processing techniques of machine learning. To obtain a more *authentic and reliable* dataset, a dataset will be made from scratch and to do so, we will be needing a few volunteers, who will be required to take part in a mock examination, with the solid aim to create a comparatively similar environment to that of an actual examination. We will be recording the videos of the volunteers so as to obtain the data required.

From the obtained videos we will be generating the image frames and use them to process the data. The data obtained after processing, will be used to train and test the model. Following this step, this processed data will be further cleaned and organized, in this step a lot of time will be invested as the data is self-created, hence organization of the data will require efforts. After the procurement of the processed data, the next step will be to build the model.

The first step to approach the model would be, to extract the frames from the given video up, which will be used for feature extraction in the next step. We will be extract frames from the live video, but the number of frames depends on the FPS of a video. For example, the video We took was of 30 fps this means that in a second 30 frames are there, but we just need one frame after every second. To extract the frames, we use cv2. But to take just a single frame in each second we used a timer, through which we were able to extract it.

Coming to the next step, that is feature extraction. According to the requirement of the paper, our goal is to detect the eyeballs of the user and then in turn, detect the coordinates of eyeball in relation to the screen and face both. These coordinates are essential to initiate the model and set up the calibration procedure. To understand and detect the gaze of the user it is very important to first detect the location of the eyeball, and in essence, accuracy of the image based eye gaze detection system is proportional to the accuracy of the localization of the center of the user's pupil.

The process of localization of the center of the user's pupil can be explained in three distinctive parts, the first part involves locating the face in the image, the second part involves locating the position of the eye and extracting the details of the area around it, and the third part involves the detection of the pupil from the eye and the calculation of it's center, we are using these three steps as it takes much lesser computational space, and is also a faster process, as it bypasses the requirement of the classifiers to parse through the entire image for processing and classification, this also makes it less prone to errors, making the entire process more streamlined.

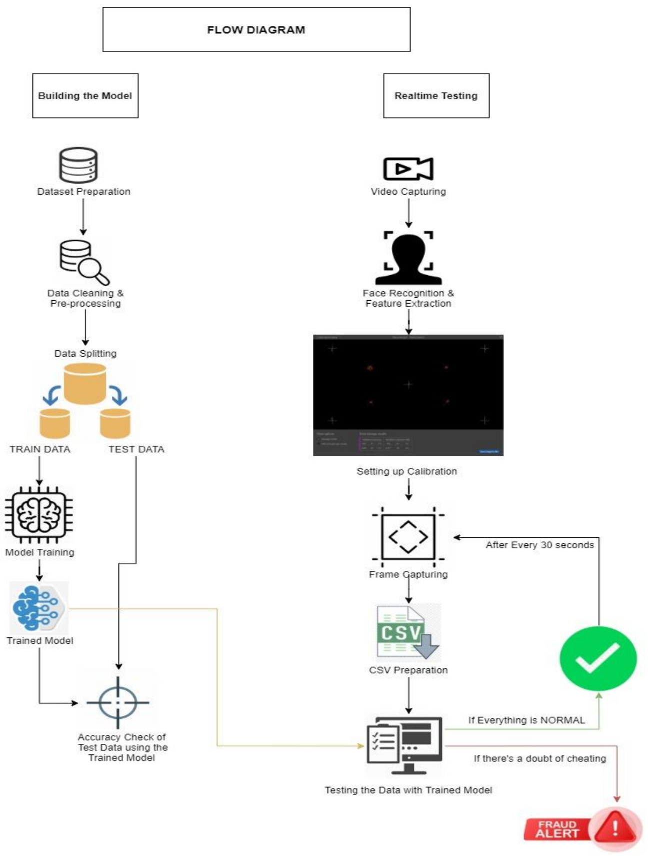

Fig. 1. The diagram explaining the elaborated flow of the following method implemented

For locating the face and the eyes in the image, our research uses Haar Cascade Classifiers, the Haar Based Classification Model has been introduced by Paul Viola and Michael Jones in their 2001 paper[13], It is an approach in image processing and machine learning to detect the objects in an image or a video by using a collection of negative and positive images, these are terms that represent the set of images on the basis of their importance or value to the classification, it also uses the concept of integral image, where the classifier extracts smaller rectangular images from the main image, these smaller images essentially are the important part of the image, that helps the machine learning classifier in classification and object detection, and this process of creating integral images also increases the efficiency and the speed of the model as the data that needs to be processed reduces.

Haar cascade classifiers rely heavily on using xml files for processing, this processing is based on the huge amount of data collected in such xml files that can be used for faster detection and processing of objects. We have used two Haar Cascade XML files, namely, 'haarcascade_\rontalface_ default.xml', and 'haarcascade_eye.xml'.

For face detection, the first step involves converting the normal image to a black and white (or Gray) image, and after that, using the haar cascade classifier, the location of the image is detected, and after that another haar cascade classifier is similarly used to detect the eyes.

After this, for detecting the center of the eyeball, the program first cuts a smaller rectangular image that contains just the eye, and after that the eye is converted to a grey image and gaussian blur is applied to reduce the noise of the image, gaussian blur uses a threshold value as an input, but the threshold value changes from image to image, so our program calculates the value of the threshold by quickly parsing through the range of values, that lie in between 0 and 255, in real time, and then assigns the best suited value to the program(which it calculates by finding the threshold value for which the coordinate results can are obtained, it keeps on checking in real time, so that any errors cause due to lighting are solved in real time.

After this the eyebrow is removed from the image, and after the image is processed through gaussian blur, blob detection machine learning algorithm from open cv is applied to the image and the center of the blob is calculated, this center is the center of the pupil of the eye, and its coordinates are calculated with respect to the initial image.

To start with the model, we would require the data that is pre-processed, cleaned and organized and we will split the data into two parts - TRAINING DATA and TESTING DATA. The splitting will be done in such a manner that the ratio of train data to test data will be 80:20.

After the splitting, the train data will be used to train the model by the CNN technique. As mentioned earlier, in 2015 a study by Zhang, Sugano, Fritz, and Bulling [10], proposed the first Convolutional Neural Network (CNN) based approach towards gaze detection. The performance of this simple gaze estimation method surpassed most research done using conventional appearance-based methods, and even though its more economically straining its effectiveness makes us overlook that. Hence, we are using the CNN approach for the same, but we will try to test the model using a few more techniques so as to compare the accuracy with the goal of providing a fully optimized training model by finalizing the technique that is most accurate.

Moving to the next stage of the paper, we will be using this trained model on our Test data, to check its accuracy score. If the accuracy is good to go, we will continue to the next part of the paper that is Real Time Testing, but if the score is not up to the desired level, we will try to increase the accuracy of the model using different techniques before moving on the next part.

Now, coming to the second part, that is the Real time testing. Real Time testing is the main part or application of our paper. In here, we'll be accessing the real-time videos of the users or the examinees performing the task that in the given case will be giving an examination. On the obtained videos, we will be performing Image processing and face recognition methodologies. After the recognition of the face, we'll be performing feature extraction, using this we are able to locate the detailed parts in the video like eyes and specifically the eyeballs whose movement is to be tracked by the model. Following this, the model will be able to access the eyes and the movement of the eyeballs of the user from the videos and the image frames.

Further, we will be setting up the calibration procedure, this will be used to set a fixed boundary range into which the user has to keep his eyeball movement restricted, this boundary will be used to check whether the user is taking part in or indulging in any sort of fraudulent activities or not.

Consequently, from these videos, as mentioned above, frames will be captured. Frame is in the image format and it will be captured in the time gap of every 5-10 secs. Now, these captured frames from the video will be converted to the data into an CSV file. This data will be used for testing, but before going to testing phase, this data will be cleaned, and pre-processing of the data will be done to make it a bit organized. After refining of the data, testing will be performed. To test the data, we will be using the Trained model which we prepared in part 1 of this section, using aforementioned and explained, CNN technique.

The metric used in the paper here is CLASSIFICATION. The model will result into 0 and 1. If the user is looking into the boundary range this means everything is okay, and the user is not taking part in or indulging in any sort of fraudulent activities, so the model will continue the same procedure after 30 seconds again, until the exam ends or any fraudulent activity is detected.

If in case the user is looking out of fixed calibrated boundary in the any of the obtained frames, this means that the user is not paying attention to the task or examination and is indulging into unfair means. So, the model will declare FRAUD DETECTED for the particular candidate. Following which, an alert will be sent to notify the proctor or the examiner in charge to pay attention to the particular user. And take necessary actions according to the given mandate or the code of conduct.

IV. RESULT

The most critical part of the eye gaze detection task is to track the eyeball movement and it's location coordinates, and we have successfully done this. The fig. 2 shown below reflects the tracked eye of one of our team members (Shubhi Tiwari).

For better accuracy, we first converted the extracted eye image into grayscale image shown in fig. 3, and then into gaussian blurred inverted image, whose threshold value was calculated in real-time live video. The image below explains this statement better.

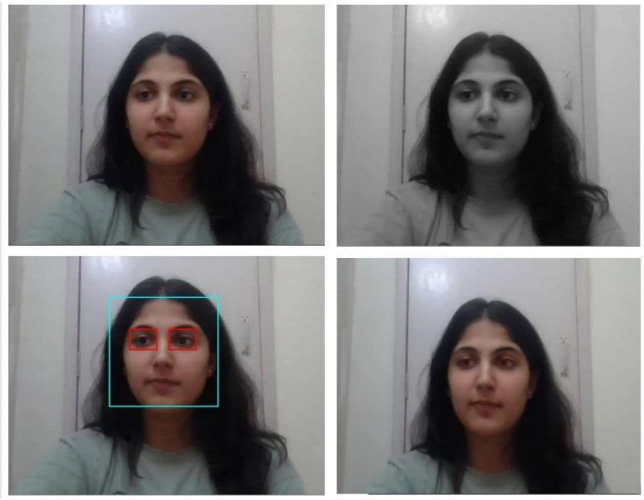

Fig. 2. The diagram explaining the capturing of face picture

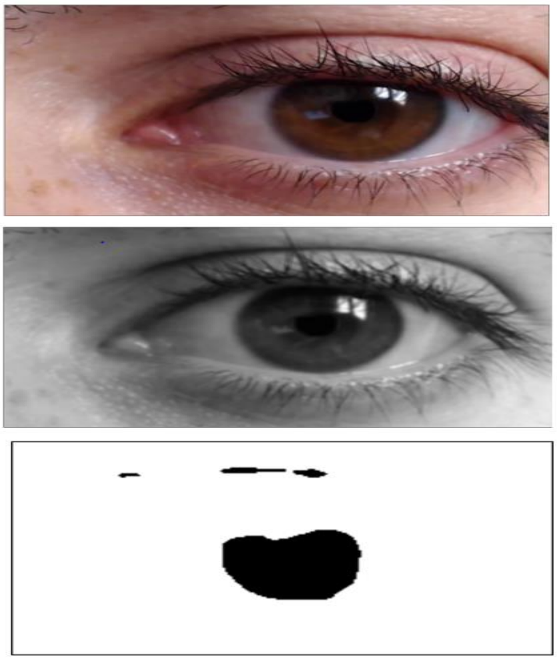

Fig. 3. The eye gaussian blurred inverted image

We calculated Derived Distance (d2) of the coordinates by using the distance formula from the origin, we assumed the origin to be the bottom left corner of the screen, then we measured the actual coordinates of the eye manually and through it calculated the Actual Distance(d1).

Then we calculated the Error(E) as the modulus of the magnitudes of the Actual and Derived distances, and finally calculated the Accuracy Percentage according to the below mentioned formula in fig. 4. For our paper we calculated the Accuracy Percentage for three different frames, and we got the following accuracies: 98.23%, 97.41% and 98.15%.

For,
Actual Distance (d1)
Derived Distance (d2)
Error (E) = |d2-d1|

Accuracy % (A) = (d1-E)/d1 * 100

Fig. 4. The calculation the Accuracy

The fig. 5 shows the classification of the built model while one of the team member (Reet Aggarwal) is looking into the frame and everything is normal and while looking outside the frame and a fraud warning is reflected.

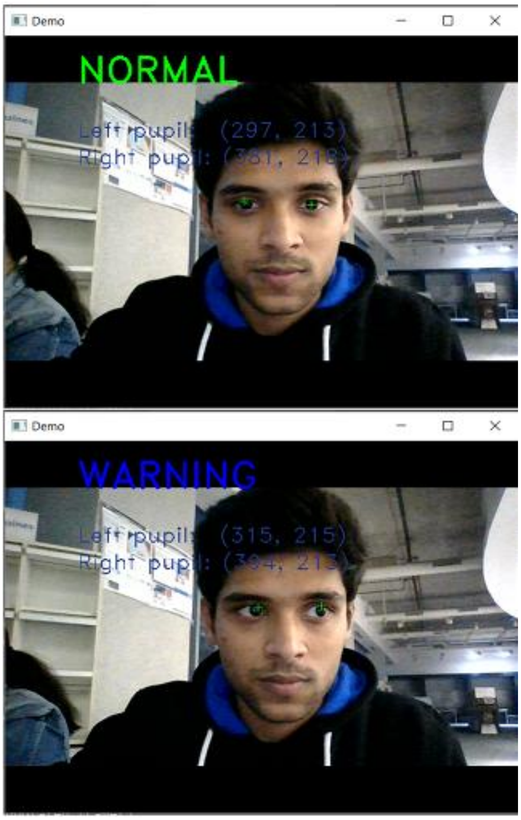

Fig. 5. The classification of the built model

V. Conclusion

The aim for this paper was to create a system which could classify an examinee's activity during the examination as fraudulent or fair. This system could come in handy for the proctors as well as for the organizations' conducting the examination. In this paper, we have used the techniques of face detection, eyeball detection and eyeball tracking and combined them to obtain a model of *eye gaze detection* which is further used to use in a classifier which classifies the data in 0 and 1 format. This result will be then used by the administrators further to send out a warning if needed.

References

[1] Bordt, S., Finck, M., Raidl, E., & von Luxburg, U. Post-hoc explanations fail to achieve their purpose in adversarial contexts. arXiv preprint arXiv:2201.10295, 2022.

[2] J. Singh, "Collaborative Filtering based Hybrid Music Recommendation System," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 186-190.

[3] N. H. Cuong and H. T. Hoang, "Eye-gaze detection with a single WebCAM based on geometry features extraction," 2010 11th International Conference on Control Automation Robotics & Vision, 2010, pp. 2507-2512, doi: 10.1109/ICARCV.2010.5707319.

[4] J. Singh and A. Sharan. A new fuzzy logic-based query expansion model for efficient information retrieval using relevance feedback approach," Neural Computing and Applications, vol. 28, 2017, p. 2557–2580.

[5] Yihua Cheng, Haofei Wang, Yiwei Bao and Feng Lu, "Appearance-based Gaze Estimation With Deep Learning: A Review and Benchmark" arXiv:2104.12668v1 [cs.CV] 26 Apr 2021.

[6] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 10, pp. 2033–2046, 2014

[7] J. Singh, et. al., "Change Detection in Remote Sensing Image Data Comparing Algebraic and Machine Learning Methods", Electronics, Article id: 1505208, 2022.

[8] Wang, Q., Zhang, W., Tang, X., Shum, H.Y.: Real-time bayesian 3-d pose tracking. IEEE Transactions on Circuits and Systems for Video Technology 16(12), 1533–1541 (2006).

[9] Y. Sugano, Y. Matsushita, Y. Sato, and H. Koike, "An incremental learning method for unconstrained gaze estimation," in European Conference on Computer Vision, ser. ECCV '08. Berlin, Heidelberg: Springer-Verlag, 2008, p. 656–667.

[10] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2015.

[11] Smith, B.A., Yin, Q., Feiner, S.K., Nayar, S.K.: Gaze Locking: Passive Eye Contact Detection for Human-Object Interaction. In: ACM Symposium on User Interface Software and Technology. pp. 271–280 (2013).

[12] Zhang, X., Sugano, Y., Fritz, M., Bulling, A.: Appearance-Based Gaze Estimation in the Wild. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 4511–4520 (2015).

[13] J. Singh, et. al., "Ensemble Deep Learning and Internet of Things Based Automated COVID-19 Diagnosis Framework", Contrast Media & Molecular Imaging, Article id: 7377502, 2022.

[14] Sangeetha, S. K. B. "A survey on Deep Learning Based Eye Gaze Estimation Methods." Journal of Innovative Image Processing (JIIP) 3, no. 03 (2021): 190-207.

[15] Sathesh, A. "TYPING EYES: A HUMAN COMPUTER INTERFACE TECHNOLOGY." Journal of Electronics and Informatics 1, no. 2 (2019): 80-88.

[16] C. T. Lin et al., "IoT-Based Wireless Polysomno- graphy Intelligent System for Sleep Monitoring," in IE EE Access, vol. 6, pp. 405-414, 2018, doi: 10.1109/ACCESS.2017.2765702.