STUDENT DISTRACTION DETECTION USING COMPUTER VISION AND MACHINE LEARNING

SUBMITTED BY

ZEESHAN 35342

A thesis submitted in partial fulfillment of the requirements for the degree of

BSCS

IQRA UNIVERSITY

2020

COORDINATOR

DR. MANSOOR Ebrahim

SUPERVISOR

MUHAMMAD FAIZAN TAHIR
A thesis presented on the Student destruction detection using Computer vision and Machine learning, automated learning analytics is becoming an essential topic in the educational area, which needs effective systems to monitor the learning process and provides feedback to the Teacher. Recent advances in visual sensors and computer vision methods enable automated monitoring of behavior and affective states of learners at different levels, from University to pre-school. The objective of this research was to build an automatic system that allowed the faculties to capture and make a summary of student behaviors Distracted or not in the Virtual Class as a part of data acquisition for the decision making process. The system records the session and identifies when the students distracted In the Online class, and then reports to the facilities.
Table of Contents

List of Figures .................................................................................................................i
Abstract .............................................................................................................................. ii
Chapter I: Introduction .....................................................................................................1
  Objective and background of project .............................................................................1.1
      problem, importance, and your motivation ...............................................................1.2
Chapter II: Technology Background ............................................................................2
  Computer Vision ..............................................................................................................2.1
  Machine Learning ..........................................................................................................2.2
Chapter III: Methodology ..............................................................................................40
  Selection of Celestial Bodies .......................................................................................41
  Selection of Subjects .....................................................................................................43
  Collection of Data .........................................................................................................50
  Analysis of Data ............................................................................................................57
Chapter IV: Findings and Discussion .............................................................................60
Acknowledgments

It is our name that appears on the front page but behind every effort there are two persons who have contributed openhandedly of their time, talent, support and have enriched the report you are about to read. My sincere thanks to Sir Faizan Tahir & Dr. Mansoor Ebrahim who have provided us with their dedication and cooperation. We thank them for giving us access to the jewels of knowledge and priceless advice. Their generosity and selflessness are gifts to their profession.
1.1 Introduction

Final Year Projects (FYPs) are an integral part of a Computer Science (CS) degree and the most critical achievement for a CS graduate. Unlike a taught course, the projects are mainly student-driven and the students have to perform the practices like project and time management, which are major issues for students. Dividing the projects into smaller tasks, close mentoring by supervisors and then tracking the students’ progress individually can improve the involvement and hence the learning of the students.

In a conventional university, the students are asked to follow the traditional software lifecycle models for projects, such as Waterfall. The serial nature of the Waterfall model allows activities to be performed in a sequential manner only. Further, it does not require frequent interaction between the developers and the client. However, inexperienced students fail to completely follow the sophisticated model correctly. The limited meetings with the supervisor leave many gaps in learning as well as the project. It has been observed that following this model the students work mostly towards the end of the semester only as compared to the Waterfall model, the Agile model enforces regular team meetings to focus on the project tasks, plan for implementation, proper monitoring of the project progress, and team collaboration.

A backlog is maintained throughout the project to track team progress and current working status. At the Virtual University (VU), an e-learning institute in Pakistan, FYPs are supervised by a team of project supervisors who coordinate with students for the entire lifecycle of the project. Most of the students at VU are either from remote areas with restricted educational facilities or are employed with time and capital constraints which force the students to opt for distance education rather than conventional study at a university.
These factors cause limited student-supervisor interactions and constrained team coordination leading to low student involvement and hence resulting in low-quality projects. Consequently, the skill set of the students fails to meet the industry expectations that are an increasing concern for the university. Therefore, this study has been carried out to identify the shortcomings of the existing process. Based on the study a hybrid approach of Agile-Scrum and in-practice Waterfall model has been proposed.

**Objective**

Many factors affect a student’s academic performance. Student achievement depends on teachers, education programs, learning environment, study hours, academic infrastructure, institutional climate, and financial issues. Another extremely important factor is the learner’s behavior. H.K. Ning and K. Downing believe that major constructs of study behavior, including study skills, study attitude, and motivation, to have strong interaction with students’ learning results. Students’ perceptions of the teaching and learning environments influence their study behavior. This means if teachers can grasp the bad attitudes of students, they can make more reasonable adjustments to change the learning environment for the students. To conclude whether good or bad behavior for a particular student is not an easy problem to solve, it must be identified by the teacher who has worked directly in the real environment. The teacher can track student behavior by observing and questioning them in the classroom. This process is not difficult in a classroom that has few students, but it is a big challenge for a classroom with a large number of students. It is valuable to develop an effective tool that can help teachers and other roles to collect data of student behavior accurately without spending too much human effort, which could assist them in developing strategies to support the learners. In this way, the students’ performances could be increased. The online courses can take advantage of this technological advancement for personalized intervention design, and reduce learners’ frustration and dropout rates.

**Background**

Many researchers have commented on the behaviors that influence students’ performances. Arnold L. Glass and Mengxue Kang have pointed out that students who are distracted by watching videos, playing games, or texting while taking lecture notes on digital
devices are far more likely to have their long-term memory affected. In this manner, the students perform more poorly in exams, even if short-term memory is not impacted. People like to think they can multitask. But this is a myth. What people are doing when they say they are multitasking is constant task switching. Although switching costs may be relatively small, sometimes just a few tenths of a second per switch, they can add up to significant amounts when people repeatedly switch back and forth between tasks. When we switch from one task to another task, the brain cannot continue and keep up with everything that it has just done. Therefore, there will be a delay as one’s attention moves from one task to another.

When the students pay more attention in class, there is a higher probability of better achievement, as stated in the book of Dorothy Piontkowski, Robert Calfee “Shannon (1942) reported positive correlations between the degree of attentiveness as measured by Morrison’s cues and student achievement.” The evidence which shows that digital devices influence the attention of students in the classroom is shown in a study by Bernard McCoy. It showed “a belief among teachers that constant use of digital technology hampered their student’s attention spans and ability to persevere in the face of challenging tasks.” Additionally, a survey written in the study showed that 71% of teachers thought technology damaged students’ attention. And 64% people who took another survey said that technology did “more to distract students than to help them academically.” Bernard’s study also pointed out that students have also identified learning distractions caused by digital technology. Wei, Wang, and Klausner found out that texting during class partially affected a student’s ability to self-regulate his/her sustained attention to classroom learning. In an earlier study, Wei and Wang noted college a student’s ability to text and perform other tasks simultaneously during class might become a habit over time. Such habits may be defined as automatic behaviors triggered by minimum consciousness.

To keep track of the actions/behaviors of students, two potential approaches can be taken: surveys and quizzes. However, these two approaches are inconvenient, and lack objectiveness, since the people might not remember what they did exactly. With the development of the computer vision field, the work of recording and analyzing students’ behaviors in the classroom in real-time is not an impossible thing at present. Il-Hyun Jo et
al. believe that a systematic understanding of each learner’s educational needs are required, and they prepared customized instructional strategies and customized content by collecting, analyzing, and systematizing learners’ data. Today, academic analytics is one of the actions that can be captured with real-time data-reporting and predictive modeling, which helps suggest likely outcomes from familiar patterns of behavior. The faculty might soon be able to use these data on behaviors as guides for course redesign and as evidence for implementing new assessments and lines of communication between instructors and students.

In particular, one of the possible reasons that make students do things other than pay attention to the lessons is poor lesson content during lecture time. Since then, from the data observed, the department of students might communicate to the lecturers to modify the content to be as suitable as possible for students. On the other hand, lecturers themselves redesign that content to let students interested in the lessons instead of neglecting them. Another action the faculty might intervene in is to directly communicate with students who have had negative attitudes during lecture time in recent days to detect the reasons why they have had those them. Our study aims to develop a software system based on computer vision to recognize students’ behavior in the classroom environment.

**Literature Review**

**Building Video-level Representations**

Modeling video-level temporal structure is crucial for better understanding of actions in videos. Recently, plenty of methods have been proposed to build video-level representations based on CNN features. Some of them [Girdhar et al., 2017; Duta et al., 2017; Lan et al., 2017] integrates BoW based encoding techniques, such as Fisher Vector and VLAD, into CNN architectures to learn video-level representations. For example, Girdhar et al. [Girdhar et al., 2017] aggregated local convolutional features across the entire extent of the video using VLAD encoding method. The other methods build video-level representations explicitly considering the temporal structure of videos. Most of them [Srivastava et al., 2015; Richard et al., 2017] feed the CNN features of multiple video segments into a RNN with LSTM [Donahue et al., 2015] or GRU [Cho et al., 2014] units.
For example, Srivastava et al. [Srivastava et al., 2015] used the multi-layer LSTM network to learn action representations. Richard et al. [Richard et al., 2017] divided all videos into small blocks and eventually modeled long and complex video information using GRU. In addition, Bilen et al. [Bilen et al., 2016] proposed a temporal pooling method, which captures the temporal evolution of a video by looking for a function that is capable of ordering the frames of the video temporally.

**Action Recognition with CNNs:** Recently, plenty of works focus on using CNNs for action recognition. Karpathy et al. [Karpathy et al., 2014] trained a deep network on the Sports-1M activities dataset. It operated on individual frames and turned out to be less accurate than the state-of-the-art hand-craft representation. To capture the motion information in videos, Simonyan et al. [Simonyan and Zisserman, 2014] proposed a two-stream network, which takes RGB frames and optical flows as input respectively and fused the predictions from the two streams as the final output. Based on the two-stream architecture, several improved methods [Feichtenhofer et al., 2016b; 2016a; Wang et al., 2016; Carreira and Zisserman, 2017; Zhang et al., 2016] have been proposed. For example, Feichtenhofer et al. [Feichtenhofer et al., 2016b] developed an architecture that can fuse spatial and temporal cues at several levels of granularity in feature abstraction and achieved great improvements. Except for the two-stream based methods, there are also other aspects of explorations.

**Existing Systems:** There are several solutions for the proposed system such that the monitoring of students' behaviors can be achieved to evaluate their studying performances. Assessing the progress of learners has been explored in an environment without digitally quantified inputs and their uninformative possibilities were calculated for the implementation. They developed a system that can monitor attention in the classroom during the lecture which can lead to two possible outcomes: a real-time reporting system or summary report.

The main focus aspects are quantifying body motion and the estimation of eye-gaze direction. With eye-gaze direction, there can be three distinct directions: the teacher/slide, notebook/bench, and other directions. The motion metrics were tested by annotating the regions in which each student resides and measuring the amount of movement inside of it.
The data was collected and ready for supervised machine-learning. However, in this paper, only assumptions and theories about the way to address students’ behaviors are deeply investigated. A doctoral thesis for the program in computing and communication showed the approaches to evaluate attention by metrics: motion, gaze estimation, and body-pose detection. For motion, differences in attention are manifested on the level of audience movement synchronization with the idea that attentive students would have a common behavior pattern. The relationship between head orientation and gaze direction was also studied. The combination of head detection and pose estimation was used to extract measures of audience head and gaze behavior. Meanwhile, the synchronization of student’s head orientation and teacher’s motion serves as a reliable indicator of the attentiveness of students. They showed that the behavior which can be used for the project is moving, but they needed to work around their assumption about the experiment. A data analysis module with the integration of computer vision technologies and machine learning algorithms to perform attendance taking was investigated to understand the students’ behaviors and students’ motion with minimum human intervention.

The computer vision system uses cameras placed in a suitable location in the classroom as its data collector module; facial recognition and body-motion detection are applied to take attendance and behavior analysis. Haar cascade face detection is applied to detect faces, and Eigenface and Fisher face approaches. These approaches are used to train and recognize students’ faces. For body detection, the cascade classifier and histogram of oriented gradients (HOG) are used. There are four rules of body detection which are based on “face is detected,” “upper body is detected,” “full-body is detected.” Furthermore, they lead to performances: sitting and concentrating in classroom, sitting but not concentrating in classroom, and standing and ready to leave the classroom. Some specialized digital devices, such as Kinect from Microsoft, have been employed to utilize the capabilities of collecting behavioral data of multiple students.

The students’ attention was evaluated by five human observers, who noted types of behavior from each student: writing, yawning, supporting head, leaning back, or gazing, and then found the attention level for each of the behavior; each behavior had a different range to evaluate the level of attention, and that was calculated by taking the mean of them.
But there were some limitations: the ground truth data on attention, computed from human observer estimates, was not entirely reliable (need better evaluation of attention level); the training data was not large enough; and the Kinect sometimes detected incorrectly and produced erroneous results.

In addition, the seven features computed from low-level Kinect data were not comprehensive enough to be able to describe all observed behavioral differences of the test persons (e.g., cannot detect writing). Recently, a school in Hangzhou, China, is using facial recognition to monitor the behavior of their students. The technology that classifies the students is generally based on their range of emotions—from antipathy to happiness (and a whole host of others). The system also cross-checks the faces of all students against the school database to mark the attendance and has the ability to predict if a student is feeling sick. Unfortunately, the results of most actions have not yet been published. However, this showed the possibility to use facial recognition technology to help and monitor students.

**Contribution of This Paper**

The major contribution of this paper was to develop a complete algorithmic process that addresses appropriate processing methods for an automatic system of monitoring student behavior. The system acts as a data collection and aggregation tool for decision making. There are many different types of behaviors of students, we focused on determining where the learner was observing across time. Our system was designed to surpass all the existing student behavior monitoring systems by evaluating and applying several computer vision techniques, such as face detection, facial landmark detection, face embedding, face classification, and gaze estimation. We implemented the algorithm for 3D position estimation. The combination of estimated locations and eye-gaze is a reliable assessment of the estimated user attention. It allows us to be able to respond in the real environment, where the layout of classrooms is different (special layout, large area, etc.) by using a combination of several cameras. Together with the combination algorithm, we used the inference ability based on the statistics of previous observations to enhance the accuracy of computer vision techniques. As an additional contribution, a web application that supports the lecturers and academic staff has been developed. The web application can
take part in the academic portal as part of the business intelligence module. Videos recorded during a student session will be processed, and then the system performs several computer vision techniques automatically. We visualized the analyzed data in the form of charts and slideshows on the web. This application, faculties not only assess the general situation of all students in the class but also grasp the details of the situation of each specific student to propose strategies to improve the quality of learning.
CHAPTER-02
Technology Background and objective of project

2.1 Artificial intelligence Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving. The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal.

Understanding Artificial Intelligence: When most people hear the term artificial intelligence, the first thing they usually think of is robots. That's because big-budget films and novels weave stories about human-like machines that wreak havoc on Earth. But nothing could be further from the truth. Artificial intelligence is based on the principle that human intelligence can be defined in a way that a machine can easily mimic it and execute tasks, from the most simple to those that are even more complex. The goals of artificial intelligence include learning, reasoning, and perception.

As technology advances, previous benchmarks that defined artificial intelligence become outdated. For example, machines that calculate basic functions or recognize text through optimal character recognition are no longer considered to embody artificial intelligence, since this function is now taken for granted as an inherent computer function.

AI is continuously evolving to benefit many different industries. Machines are wired using a cross-disciplinary approach based in mathematics, computer science, linguistics, psychology, and more.

“Algorithms often play a very important part in the structure of artificial intelligence, where simple algorithms are used in simple applications, while more complex ones help frame strong artificial intelligence.”
**Machine Learning:** Machine learning is the concept that a computer program can learn and adapt to new data without human intervention. Machine learning is a field of artificial intelligence (AI) that keeps a computer’s built-in algorithms current regardless of changes in the worldwide economy.

**KEY TAKEAWAYS**

Machine learning is an area of artificial intelligence (AI) with a concept that a computer program can learn and adapt to new data without human intervention.

A complex algorithm or source code is built into a computer that allows for the machine to identify data and build predictions around the data that it identifies.

Machine learning is useful in parsing the immense amount of information that is consistently and readily available in the world to assist in decision making.

Machine learning can be applied in a variety of areas, such as in investing, advertising, lending, organizing news, fraud detection, and more.

**Understanding Machine Learning:** Various sectors of the economy are dealing with huge amounts of data available in different formats from disparate sources. The enormous amount of data, known as big data, is becoming easily available and accessible due to the progressive use of technology, specifically advanced computing capabilities and cloud storage. Companies and governments realize the huge insights that can be gained from tapping into big data but lack the resources and time required to comb through its wealth of information. As such, artificial intelligence measures are being employed by different industries to gather, process, communicate, and share useful information from data sets. One method of AI that is increasingly utilized for big data processing is machine learning.
The various data applications of machine learning are formed through a complex algorithm or source code built into the machine or computer. This programming code creates a model that identifies the data and builds predictions around the data it identifies. The model uses parameters built in the algorithm to form patterns for its decision-making process. When new or additional data becomes available, the algorithm automatically adjusts the parameters to check for a pattern change, if any. However, the model shouldn’t change.

**Uses of Machine Learning**: Machine learning is used in different sectors for various reasons. Trading systems can be calibrated to identify new investment opportunities. Marketing and e-commerce platforms can be tuned to provide accurate and personalized recommendations to their users based on the users’ internet search history or previous transactions. Lending institutions can incorporate machine learning to predict bad loans and build a credit risk model. Information hubs can use machine learning to cover huge amounts of news stories from all corners of the world. Banks can create fraud detection tools from machine learning techniques. The incorporation of machine learning in the digital-savvy era is endless as businesses and governments become more aware of the opportunities that big data presents.

**Application of Machine Learning**: How machine learning works can be better explained by an illustration in the financial world. Traditionally, investment players in the securities market like financial researchers, analysts, asset managers, and individual investors scour through a lot of information from different companies around the world to make profitable investment decisions. However, some pertinent information may not be widely publicized by the media and may be privy to only a select few who have the advantage of being employees of the company or residents of the country where the information stems from. In addition, there’s only so much information humans can collect and process within a given time frame. This is where machine learning comes in.
An asset management firm may employ machine learning in its investment analysis and research area. Say the asset manager only invests in mining stocks. The model built into the system scans the web and collects all types of news events from businesses, industries, cities, and countries, and this information gathered makes up the data set. The asset managers and researchers of the firm would not have been able to get the information in the data set using their human powers and intellects. The parameters built alongside the model extracts only data about mining companies, regulatory policies on the exploration sector, and political events in select countries from the data set.

Example of Machine Learning: Say mining company XYZ just discovered a diamond mine in a small town in South Africa. A machine learning tool in the hands of an asset manager that focuses on mining companies would highlight this as relevant data. The model in the machine learning tool would then use an analytics tool called predictive analytics to make predictions on whether the mining industry will be profitable for a time period, or which mining stocks are likely to increase in value at a certain time, based on the recent information discovered, without any input from the asset manager. This information is relayed to the asset manager to analyze and make a decision for their portfolio. The asset manager may then make a decision to invest millions of dollars into XYZ stock.

In the wake of an unfavorable event, such as South African miners going on strike, the computer algorithm adjusts its parameters automatically to create a new pattern. This way, the computational model built into the machine stays current even with changes in world events and without needing a human to tweak its code to reflect the changes. Because the asset manager received this new data on time, they are able to limit their losses by exiting the stock.

Computer Vision: Computer vision is a field of study focused on the problem of helping computers to see.

"At an abstract level, the goal of computer vision problems is to use the observed image data to infer something about the world"
It is a multidisciplinary field that could broadly be called a subfield of artificial intelligence and machine learning, which may involve the use of specialized methods and make use of general learning algorithms.

Overview of the Relationship of Artificial Intelligence and Computer Vision

As a multidisciplinary area of study, it can look messy, with techniques borrowed and reused from a range of disparate engineering and computer science fields.

One particular problem in vision may be easily addressed with a hand-crafted statistical method, whereas another may require a large and complex ensemble of generalized machine learning algorithms.

"Computer vision as a field is an intellectual frontier. Like any frontier, it is exciting and disorganized, and there is often no reliable authority to appeal to. Many useful ideas have no theoretical grounding, and some theories are useless in practice; developed areas are widely scattered, and often one looks completely inaccessible from the other."

The goal of computer vision is to understand the content of digital images. Typically, this involves developing methods that attempt to reproduce the capability of human vision.
Understanding the content of digital images may involve extracting a description from the image, which may be an object, a text description, a three-dimensional model, and so on.

“Computer vision is the automated extraction of information from images. Information can mean anything from 3D models, camera position, object detection and recognition to grouping and searching image content.”

Computer Vision and Image Processing

Computer vision is distinct from image processing.

Image processing is the process of creating a new image from an existing image, typically simplifying or enhancing the content in some way. It is a type of digital signal processing and is not concerned with understanding the content of an image.

A given computer vision system may require image processing to be applied to raw input, e.g. pre-processing images.

Examples of image processing include:

- Normalizing photometric properties of the image, such as brightness or color.
- Cropping the bounds of the image, such as centering an object in a photograph.
- Removing digital noise from an image, such as digital artifacts from low light levels.

Challenge of Computer Vision

- Helping computers to see turns out to be very hard.

“The goal of computer vision is to extract useful information from images. This has proved a surprisingly challenging task; it has occupied thousands of intelligent and creative minds over the last four decades, and despite this we are still far from being able to build a general-purpose “seeing machine.”
Computer vision seems easy, perhaps because it is so effortless for humans.

Initially, it was believed to be a trivially simple problem that could be solved by a student connecting a camera to a computer. After decades of research, “computer vision” remains unsolved, at least in terms of meeting the capabilities of human vision.

*Making a computer see was something that leading experts in the field of Artificial Intelligence thought to be at the level of difficulty of a summer student’s project back in the sixties. Forty years later the task is still unsolved and seems formidable.*

One reason is that we don’t have a strong grasp of how human vision works.

Studying biological vision requires an understanding of the perception organs like the eyes, as well as the interpretation of the perception within the brain. Much progress has been made, both in charting the process and in terms of discovering the tricks and shortcuts used by the system, although like any study that involves the brain, there is a long way to go.

“*Perceptual psychologists have spent decades trying to understand how the visual system works and, even though they can devise optical illusions to tease apart some of its principles, a complete solution to this puzzle remains elusive*”

Another reason why it is such a challenging problem is because of the complexity inherent in the visual world.

A given object may be seen from any orientation, in any lighting conditions, with any type of occlusion from other objects, and so on. A true vision system must be able to “see” in any of an infinite number of scenes and still extract something meaningful.

Computers work well for tightly constrained problems, not open unbounded problems like visual perception.

**Tasks in Computer Vision:** Nevertheless, there has been progress in the field, especially in recent years with commodity systems for optical character recognition and face detection in cameras and smartphones.
“Computer vision is at an extraordinary point in its development. The subject itself has been around since the 1960s, but only recently has it been possible to build useful computer systems using ideas from computer vision.”

The 2010 textbook on computer vision titled “Computer Vision: Algorithms and Applications” provides a list of some high-level problems where we have seen success with computer vision.

- Optical character recognition (OCR)
- Machine inspection
- Retail (e.g. automated checkouts)
- 3D model building (photogrammetry)
- Medical imaging
- Automotive safety
- Match move (e.g. merging CGI with live actors in movies)
- Motion capture (mocap)
- Surveillance
- Fingerprint recognition and biometrics

It is a broad area of study with many specialized tasks and techniques, as well as specializations to target application domains.

Computer vision has a wide variety of applications, both old (e.g., mobile robot navigation, industrial inspection, and military intelligence) and new (e.g., human computer interaction, image retrieval in digital libraries, medical image analysis, and the realistic rendering of synthetic scenes in computer graphics).

It may be helpful to zoom in on some of the more simpler computer vision tasks that you are likely to encounter or be interested in solving given the vast number of publicly available digital photographs and videos available.
Many popular computer vision applications involve trying to recognize things in photographs; for example:

**Object Classification**: What broad category of object is in this photograph?
**Object Identification**: Which type of a given object is in this photograph?
**Object Verification**: Is the object in the photograph?
**Object Detection**: Where are the objects in the photograph?
**Object Landmark Detection**: What are the key points for the object in the photograph?
**Object Segmentation**: What pixels belong to the object in the image?

**Object Recognition**: What objects are in this photograph and where are they?

Other common examples are related to information retrieval; for example: finding images like an image or images that contain an object.

**Image Processing**: Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Nowadays, image processing is among rapidly growing technologies. It forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools;
- Analysing and manipulating the image;
- Output in which result can be altered image or report that is based on image analysis.
There are two types of methods used for image processing namely, analogue and digital image processing. Analogue image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. Digital image processing techniques help in manipulation of the digital images by using computers. The three general phases that all types of data have to undergo while using digital technique are pre-processing, enhancement, and display, information extraction.

**Deep learning:** is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural networks.

Deep Learning is Large Neural Networks

Andrew Ng from Coursera and Chief Scientist at Baidu Research formally founded Google Brain that eventually resulted in the productization of deep learning technologies across a large number of Google services.

He has spoken and written a lot about what deep learning is and is a good place to start.

In early talks on deep learning, Andrew described deep learning in the context of traditional artificial neural networks. In the 2013 talk titled “Deep Learning, Self-Taught Learning and Unsupervised Feature Learning” he described the idea of deep learning as:

*Using brain simulations, hope to:*

– *Make learning algorithms much better and easier to use.*

– *Make revolutionary advances in machine learning and AI.*
Detailed Background of the technology

**Anaconda:** Anaconda being the World’s most popular Data Science platform will be used in our final year project. Anaconda is a conditional free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python. There is also a small, bootstrap version of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.

**Anaconda Navigator:**

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository, install them in an environment, run the packages and update them. It is available for Windows, macOS and Linux.

**The following applications are available by default in Navigator:**

- JupyterLab
- Jupyter Notebook
- QtConsole
- Spyder
- Glue
- Orange
- RStudio
- Visual Studio Code
**Python (programming language):** Python is an interpreted, high-level and general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured (particularly, procedural), object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

**TensorFlow:** TensorFlow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. A Tensor Processing Unit (TPU) is a programmable AI accelerator designed to provide high throughput of low-precision arithmetic and oriented toward using or running models rather than training them.

**OpenCV (Open Source Computer Vision Library):**

OpenCV (Open Source Computer Vision Library) is a library of programming functions mainly aimed at real-time computer vision. OpenCV features GPU acceleration for real-time operations.

**Applications:**

**OpenCV's application areas include:**

- 2D and 3D feature toolkits
- Egomotion estimation
- Facial recognition system
- Gesture recognition
- Human–computer interaction (HCI)
- Motion understanding
- Object detection
- Segmentation and recognition
- Stereopsis stereo vision: depth perception from 2 cameras
- Structure from motion (SFM)
- Motion tracking
- Augmented reality

**To support some of the above areas, OpenCV includes a statistical machine learning library that contains:**

- Boosting
- Decision tree learning
- Gradient boosting trees
- Expectation-maximization algorithm
- k-nearest neighbor algorithm
- Naive Bayes classifier
- Artificial neural networks
- Random forest
- Support vector machine (SVM)
- Deep neural networks (DNN)

**Keras:** Keras is an open-source library that provides a Python interface for artificial neural networks. Keras acts as an interface for the TensorFlow library.

**Features:** Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier to simplify the coding necessary for writing deep neural network code. In addition to standard neural networks, Keras has support for convolutional and recurrent neural networks. It supports other common utility layers like dropout, batch normalization, and pooling. Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of distributed training of deep-learning models on clusters of Graphics processing units (GPU) and tensor processing units (TPU).
**DLib**: DLib is an open source C++ library implementing a variety of machine learning algorithms, including classification, regression, clustering, data transformation, and structured prediction.

DLib is an open source modern C++ library implementing many machine learning algorithms and supporting functionality like threading and networking.

**DLib-ml implements numerous machine learning algorithms:**

- SVMs,
- K-Means clustering,
- Bayesian Networks,
- and many others.

**DLib also features utility functionality including:**

- Networking,
- Threading,
- Numerical Algorithms,
- Image Processing,
- and Data Compression and Integrity algorithms.

DLib includes extensive unit testing coverage and examples using the library. Every class and function in the library is documented. This documentation can be found on the library's home page. DLib provides a good framework for developing machine learning applications in C++.

DLib is much like DMTL in that it provides a generic high-performance machine learning toolkit with many different algorithms, but DLib is more recently updated and has more examples. DLib also contains much more supporting functionality.
Support Vector: Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot).

Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

Jupyter Notebook: The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter. Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.
Convolutional Neural Networks (CNN):

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision.

The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics.
The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

**Problem domain with historical background**

Online Education System was introduced in the world before Covid-19, but it came into demand when Covid-19 affected the whole world. Teachers have given their best to educate the students by online means of learning, but how can we ensure that students are reciprocating with their teachers by attending the classes with the same focus and enthusiasm as they show in physical class. This has been quite a problem for the teachers to maintain the records for the students who are interested in lectures and who are not. To ensure this whether a student is focusing on the studies by all means, we have proposed this system.

Automated learning analytics is becoming an essential topic in the educational area, which needs effective systems to monitor the learning process and provides feedback to the Teacher. Recent advances in visual sensors and computer vision methods enable automated monitoring of behavior and affective states of learners at different levels, from University to pre-school. The objective of this research is to build an automatic system that allowed the faculties to capture and make a summary of student behaviors Distracted or not in the Virtual Class as a part of data acquisition for the decision-making Process. The system records the session and identifies when the students distracted In the Online class, and then reports to the facilities.
Discuss all similar solutions

Solution 1:

Human Distraction Detection from Real Time Video Data using Artificial Intelligence and Deep Learning Techniques by Raffesia Khan (Khulna University)

Abstraction: In case of learning activity, both physical and virtual classroom, behavior detection system can detect the learner’s attention level, difficulties in interaction, collaboration among them. Based on these information teachers can change their course material and also take care of learners those need additional help. The goal of this research is to find out the symptom of distraction among human in different situation, for example students in classroom, drivers in vehicle or any other people. A real-time algorithm to detect attention distraction in a video sequence from a standard camera is proposed in this work. This artificial emotional intelligent or emotional AI algorithm is proposed to detect change in visual attention for learners in a classroom. The primary hypothesis of this dissertation is to detect changes in human facial movement and calculate their attention level using the deviations.

Conclusion: In this paper, we propose a distraction detection model for human. Our model continuously monitors a video and recognizes attentive as well as distracted behavior of a person. We believe this distraction detection model can be helpful for so many sectors such as students learning, where human attention as well as distraction detection is necessary. The performance evaluation for the proposed system shows that the algorithm can effectively detect the change in individual learner’s attention level. This algorithm can also detect any abnormal facial expression during learning activity based on the facial attribute’s alignment deviation from standard value over time. This proposed work can detect distraction for real time video without any manual annotation. Due to our defined attribute model with selected landmarks, unnecessary computations are avoided and the system configuration is faster than state of art. Also, because of the novel behavior model, no additional machine learning or deep learning training is required for classifying distracted and attentive behaviors.
Solution 2:


Abstraction: Monitoring drivers’ visual behavior using machine learning techniques has been identified as an effective approach to detect and mitigate driver distraction to enhance road safety. In our previous work, detection system based on supervised Extreme Learning Machine (ELM) was developed and tested with satisfactory performance. However, supervised ELM requires all training data to be labeled, which can be costly and time-consuming. This paper proposed and evaluated a semi-supervised distraction detection system based on Semi-Supervised Extreme Learning Machine (SS-ELM). The experimental results show that SS-ELM outperformed supervised ELM in both accuracy (95.5% for SS-ELM vs. 93.0% for ELM) and model sensitivity (97.6% for SS-ELM and 95.5% for ELM), suggesting that the proposed semi-supervised detection system can extract information from unlabeled data effectively to improve the performance. SS-ELM based detection system has the potential of improving accuracy and alleviating the cost of adapting distraction detection systems to new drivers, and thus is more promising for real world applications.

Conclusion: all the distraction detection systems proposed so far are based on supervised learning, meaning that the training of such systems need to be ‘supervised’ by human experts by providing a target set for the training data containing distraction status. The supervised learning paradigm is only suitable for early stage research and may not be suitable for implementation in real driving cases, because of the huge cost and difficulty of creating target distraction status set, which requires additional subjective ratings by the driver [12], post-processing by the experimenters [16,7], or additional computation based on data from other sources [3]. On the other hand, data without labels of distraction status (unlabeled data) are in fact easy to collect without additional costs, e.g., collecting from driver’s daily driving records.
Solution 3:

Driver distraction detection using machine vision techniques by Rominson

**Abstraction:** This article presents a system for detecting states of distraction in drivers during daylight hours using machine vision techniques, which is based on the image segmentation of the eyes and mouth of a person, with a front-face-view camera. From said segmentation states of motion of the mouth and head are established, thus allowing to infer the corresponding state of distraction. Images are extracted from short videos with a resolution of 640x480 pixels and image processing techniques such as color space transformation and histogram analysis are applied. A decision concerning the state of the driver is the result from a multilayer perceptron-type neural network with all extracted features as inputs. Achieved performance is 90% for a controlled environment screening test and 86% in real environment, with an average response time of 30 ms.

**Conclusion:** The proposed system is able to identify driver distraction states using movement and orientation of eyes, mouth and head as identification parameters. Global detection achieved a precision of 99% under the following controlled conditions as: maximum rotation of the head of 120 °, soft movements and a distance between driver and camera of no more than 70cm. Relying on several factors to perform the discrimination of the distraction state allows the system to increase the robustness against similar work found in the literature under controlled conditions. Under non-controlled conditions, system performance drops in a lower amount compared to other work. Furthermore, global performance achieves a precision of 86% using a solution which had not been previously considered. Mouth banalization through histogram analysis did not fully resolve problems caused by variations in light; as future work, robustness through machine vision algorithms against changes in light condition is required.

Solution 4:

Using Video to Automatically Detect Learner Affect in Computer-Enabled Classrooms by Nigel Bosch (University of Norte dame)
Abstraction: Affect detection is a key component in intelligent educational interfaces that respond to students’ affective states. We use computer vision and machine-learning techniques to detect students’ affect from facial expressions (primary channel) and gross body movements (secondary channel) during interactions with an educational physics game. We collected data in the real-world environment of a school computer lab with up to 30 students simultaneously playing the game while moving around, gesturing, and talking to each other. The results were cross-validated at the student level to ensure generalization to new students. Classification accuracies, quantified as area under the receiver operating characteristic curve (AUC), were above chance (AUC of 0.5) for all the affective states observed, namely, boredom (AUC = .610), confusion (AUC = .649), delight (AUC = .867), engagement (AUC = .679), frustration (AUC = .631), and for off-task behavior (AUC = .816). Furthermore, the detectors showed temporal generalizability in that there was less than a 2% decrease in accuracy when tested on data collected from different times of the day and from different days. There was also some evidence of generalizability across ethnicity (as perceived by human coders) and gender, although with a higher degree of variability attributable to differences in affect base rates across subpopulations. In summary, our results demonstrate the feasibility of generalizable video-based detectors of naturalistic affect in a real-world setting, suggesting that the time is ripe for affect-sensitive interventions in educational games and other intelligent interfaces.

Conclusion: Our long-term vision is for next-generation learning environments to improve the process and products of learning by considering affect in addition to cognition. We hope that the present research that detects students’ affect in a noisy real-world environment and with evidence of multiple levels of generalizability helps us achieve this goal. Next critical step is to use the detectors to trigger affect-sensitive interventions in order to provide a more enjoyable, efficient, and effective learning experience for all students.
Comparison:

<table>
<thead>
<tr>
<th>Other Solution</th>
<th>Our Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solution 01</td>
<td>Our project is focused on student behavior only but their solution is more generalized they are detecting driver’s behavior with same algorithm.</td>
</tr>
<tr>
<td>Solution 02</td>
<td>This solution is same as our solution, the only difference is that they are detecting driver behavior to enhanced road safety</td>
</tr>
<tr>
<td>Solution 03</td>
<td>This solution is also detecting driver behavior but only during daylight but our system can detect student behavior all the time.</td>
</tr>
<tr>
<td>Solution 04</td>
<td>Solution 4 is way more advanced system it depends on body movement also, our system only detects facial features.</td>
</tr>
</tbody>
</table>
Summary of research papers related to the technology

1. Emotion AI, Real-Time Emotion Detection using CNN

The ability to confidently detect human emotions can have a wide array of impactful applications, and therefore emotion recognition has been a core area of research in computer vision. We wanted to focus on the issue of emotion recognition, and build a real-time emotion detection system. When we began to work on the area of emotion detection, we quickly realized that there is an innate problem which is that all data sets are based on "acted" emotions instead of "real" emotions. Many of these data sets such as CK+ ([Lucey et al., 2010]) and JAFFE ([Lyons et al., 1998]) are collections of actors who demonstrated core emotions in front of a camera. Therefore the field isn't detecting real emotions, but rather detecting the emotion that the subject is acting or the observer is perceiving. This problem was also very obvious while testing our model, as we saw confidence scores increase as the subject portray very exaggerated facial expressions that would be defined as "fake" by a human. When we discussed possible applications of a successful emotion recognition tool, one application we thought of is to use emotion labels and prediction scores combined with social science on emotion research led by Paul Ekman [Ekman, 1992] to predict emotion intensities. As indicated in Frijda et al. [1992] emotion intensity prediction is a really hard problem and a very valuable insight for the field of psychology. The main reason we didn’t pursue emotion intensity prediction is that there were no existing data sets or research that can serve as the ground truth. Therefore we concentrated on building a successful emotion recognition model that can work in real-time. In this project we built a model that uses Convolution Neural Networks to successfully classify faces as one of the core seven emotions: anger, contempt, disgust, fear, sadness, happiness, surprise, and neutral [Darwin et al., 1998]

2. Face Recognition System With Face Detection

The computational models, which were implemented in this project, were chosen after extensive research, and the successful testing results confirm that the choices made by the researcher were reliable. The system with manual face detection and automatic face
recognition did not have a recognition accuracy over 90%, due to the limited number of eigenfaces that were used for the PCA transform. This system was tested under very robust conditions in this experimental study and it is envisaged that real-world performance will be far more accurate. The fully automated frontal view face detection system displayed virtually perfect accuracy and in the researcher's opinion further work need not be conducted in this area. The fully automated face detection and recognition system was not robust enough to achieve a high recognition accuracy. The only reason for this was the face recognition subsystem did not display even a slight degree of invariance to scale, rotation or shift errors of the segmented face image. This was one of the system requirements identified in section 2.3. However, if some sort of further processing, such as an eye detection technique, was implemented to further normalise the segmented face image, performance will increase to levels comparable to the manual face detection and recognition system. Implementing an eye detection technique would be a minor extension to the implemented system and would not require a great deal of additional research. All other implemented systems displayed commendable results and reflect well on the deformable template and Principal Component Analysis strategies. The most suitable real-world applications for face detection and recognition systems are for mugshot matching and surveillance. There are better techniques such as iris or retina recognition and face recognition using the thermal spectrum for user access and user verification applications since these need a very high degree of accuracy. The real-time automated pose invariant face detection and recognition system proposed in chapter seven would be ideal for crowd surveillance applications. If such a system were widely implemented it’s potential for locating and tracking suspects for law enforcement agencies is immense.

The implemented fully automated face detection and recognition system (with an eye detection system) could be used for simple surveillance applications such as ATM user security, while the implemented manual face detection and automated recognition system is ideal of mugshot matching. Since controlled conditions are present when mugshots are gathered, the frontal view face recognition scheme should display a recognition accuracy far better than the results, which were obtained in this study, which was conducted under adverse conditions.
3. **Student monitoring/counseling in e-learning platforms**

This PhD thesis presents a way to detect some context information regarding e-learning students, namely stress, that will allow to try to adapt and personalize the way content is presented to that student. The main goal is to detect individualized characteristics of students that are needed for better understand the way he/she learns, thus having precious information for improving learning. Some state of the art investigation is done, and e-learning evolution and importance is remarked. Some issues regarding e-learning platforms are identified, namely and more important the lack of expressivity of such systems.

As a way to achieve some "expressivity" in such systems, artificial intelligence is considered, and its evolution in e-learning usage is reviewed. Intelligent tutoring systems try to mitigate such issues, but also lack some important characteristics. A framework to develop it is proposed. Latter in the next chapter, a way to detect stress in e-learning students is presented, through the use of a dynamic stress recognition module. It uses keyboard and mouse as sensors to monitor e-learning students. The results achieved through the use of the log tool developed under the scope of this work in a case study with 74 students, are encouraging. With this case study, two main types of students are identified, regarding stress influence in their behavior. This possibly will allow that, identified the student behavior under stress some adequate measures will then be taken, to avoid that the students go beyond their stress level comfort zone, thus personalizing learning and being aware of the specific student characteristics, in a given moment.

4. **Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.**

With the increase of distance learning, and e-learning environments in particular, having a system capable of determining students' engagement is of primordial importance and one of the biggest challenges both for teachers, researchers and policy makers.
Here, we presented a new approach of a system to detect the engagement level of the students. The system uses only the information provided by the built-in web-camera present in typical laptop computer. Our system uses the images grabbed by the camera to extract information about the movements of the eyes and head, and combines this information with the facial emotions, also retrieved from these images, to produce a concentration index. The presented system produces three classes of engagement: "very engaged", "nominally engaged" and "not engaged at all". The system proposed here was designed to work in real time.

By including the facial emotions information reflected by students about the learning topic, which includes the seven typical emotions, a teacher/instructor/learning management system will have live feedback, hence helping the system/teacher/instructor to automatically adapt the learning contents to the needs of the students. This will definitively contribute to dynamically enrich the learning environment and hence, improve the performance of the students.

We have tested our system with fifteen students in a typical e-learning scenario, and the results show that the system correctly identifies each period of time where students were "very engaged", "nominally engaged" and "not engaged at all". Additionally, the results also show that the students with best scores also have higher resultant concentration index. In the future, we want to merge the information currently provided by our system with the information retrieved with the help of other sensors, such as heart rate, EEG signals, and oxygen level, among other. We are also working to change to 3D facial expressions detection, because they better facilitate an examination of the fine structural changes inherent to spontaneous expressions. However, this will pose additional difficulties and the need to use other capturing Cameras, besides the laptop typical built in web-camera.
5. Real-Time Facial Affective Computing on Mobile Devices

We have proposed a light-weight CNN architecture for real-time facial affective computing on mobile devices. The key design principle of proposed network architecture is to minimize the number of parameters and computational complexity. This network uses facial images as input and outputs the categories of facial expression and values of valance and arousal. Compared to conventional CNNS, the proposed method well balances the high performance and low computation complexity. Moreover, the performance of the proposed method outperforms a series of existing methods in our experiment. We have also implemented a real-time facial affective computing mobile application that has a low consumption of memory and storage on actual mobile devices to demonstrate the feasibility of the proposed method for mobile development.

The performance of our proposed method still has limitations and possibilities for further improvement. In terms of valence prediction, the proposed light-weight CNN architecture still has a gap compared with the method proposed by Mollahosseini et al. [33]. We understand that there is a correlation between facial expression classification and valence/arousal prediction which is the possibility of further improving the performance. The alternative to two separated networks, implementing Multi-Task Learning (MTL) in one network structure is an alternative solution. Future work will address these limitations of the proposed methods and further improve the performance of facial affective computin
CHAPTER-03
Requirements and Methodology

How we are going to solve the problem that we have described in chapter 2

Computer Vision and Machine learning

The construction of explicit, meaningful descriptions of physical objects from images. Here we use Classification techniques such as Image Classification, Object detection, face features extraction and machine learning algorithms for better results and accuracy score

Dataset and evaluation techniques

Information about open and online datasets and knowing about metrics and evaluation techniques might be of great interest for researchers in any research field. A review on available datasets, metrics for evaluation and evaluation techniques suitable for users’ engagement detection in the context of learning, and some research results are discussed in this section.

Dataset

The need for large, labeled, publicly available datasets for training, evaluating, and benchmarking has been widely acknowledged, and a number of efforts to address this need have been made in the last few years. In user engagement detection, while many research studies use their in-house datasets, very few of them are made publicly available online and Here we use DAiSEE dataset. Gupta et al. (2018) created the DAiSEE dataset with an intent to capture learners’ engagement in online courses. This dataset includes 112 individuals, where 80 male and 32 females. The videos in the dataset were collected in unconstrained environments, such as at dorm rooms, crowded lab spaces and libraries, with three different illumination settings—light, dark, and neutral. The videos were captured with a webcam mounted on a computer focusing on learners’ watching some video tutorial. The annotation of the video frames was done in four different levels—engaged, bored, confused, and frustrated, by relying on the “wisdom-of-the crowd”.

Final Year Project Thesis Page |
Agile (Scrum)
Scrum is an agile framework for developing, delivering, and sustaining complex products, with an initial emphasis on software development, although it has been used in other fields including research, sales, marketing and advanced technologies.

What logic, constraints, platform and strategy you would use to implement requirements given in this chapter.

The state of the art of engagement detection methods in the context of online learning, and then it identifies the challenges of detecting engagement in online learning. We classify the existing methods into three main categories automatic, semi-automatic and manual considering the methods' dependencies on learners' participation. And, then the methods in each category are divided into subcategories based on the types of data used, e.g., audio, video, learner log data etc. In particular, the computer vision-based methods in the automatic category that use facial expressions are examined because they are promising in an online learning environment, nonintrusive in nature, and cost-effective when considering the hardware and the software needed for capturing and analyzing video data. Finally, we explore available datasets and performance metrics for engagement measurement, and provide recommendations for the future to advance the technology of engagement measurement for online education. Among the different methods, the computer vision-based methods in the automatic category are found to be beneficial so we use computer vision method and machine learning technique

ADOPTED PROJECT LIFE CYCLE MODEL

The proposed FYP lifecycle is a blend of Waterfall and Agile models. Semester 1 is completed using Waterfall model whereas Semester 2 is completed using the Agile model as shown in Figure-1.
Project Management AgileWrap was adopted for student projects management because it can optimize the distributed product development process through its modules.

**SPRINTS AND TASKS**

The supervisor acted as a scrum master for the project. The sprint plans and user stories were defined by supervisor. Depending upon the sprint duration and tasks complexity a project was distributed into three to four weeks release plans.

The students were required to create the product log at the project management website and identify the tasks for each user story of the sprints. They were required to estimate the work hours for each task and justify their estimate. It was mandatory for the students to update daily work. The individual effort of each member of the group was discussed in scrum meetings. The students had to demonstrate the running software to the supervisor for evaluation after each sprint.

**Justifications why we would use these strategies or platform or language**
We use all these techniques, strategies and platform language because these are manly used for Machine learning and computer vision as mentioned above, **Python** is one of the most popular language, the extensive support of libraries makes it outstanding from other progrmaming languages. We use Python because of its

- Ease of coding
- Fast prototyping
- Vast libraries for machine learning
- It is open source
- It can be directly integrated with web frameworks
- And mostly commonly used

We Use **Machine Learning** because it provides smart alternatives to analyzing vast volumes of data. By developing fast and efficient algorithms and data-driven models for real-time processing of data, Machine Learning can produce accurate results and analysis.

We Use **Computer vision** because Computer vision is a field of artificial intelligence that trains computers to interpret and understand the visual world. Using digital images from cameras and videos and deep learning models, machines can accurately identify and classify objects and then react to what they see.

We use **DAiSEE dataset** because the need for large, labeled, publicly available datasets for training, evaluating, and benchmarking has been widely acknowledged, and a number of efforts to address this need have been made in the last few years. In user engagement detection, while many research studies use their in-house datasets, very few of them are made publicly available online.

We use **Power BI Desktop** is a free, self-service data analysis and report authoring tool that we install on a Windows computer. It can connect to more than 70 on-premises and cloud data sources to turn information into interactive visuals. ... Share reports with others using the **Power BI** service.
Proposed model: For Research based project, you should draw your proposed model here. For other projects, draw simple Project Flow Diagram.
New Features (must be few, simple and precise, not detailed like functional requirements)

1) Student Engagement detection
2) Student Performance analysis
3) learners’ perceived engagement
4) Monitoring
5) Reporting
6) Face detection
7) Face landmarks detection
8) Distraction Detector
9) Facial Emotion Recognition
10) Engagement Classification

Type of roles (users)

- Students
- Professionals
- Tutors
- Institutions can adopt this system for their MooCs and integrate on their own LMS

User requirements + Features + Functional requirements (detailed)

The framework is consisted with five different modules that include detection, feature extraction, tracking, classification, and decision, for learner’s perceived engagement detection using the computer vision-based methods

Distraction Detector

In each video frame the student's face is detected using the Viola & Jones algorithm. Next, within the detected face, the eyes region is located. The eyes region feeds a Convolutional Neural Network (CNN), used as a binary classifier, to predict the student's attention state in the two categories "Distracted" or "Focused"
Facial Emotion Recognition

Only when the student is Focused", further facial emotions analysis will take place. For this purpose, another CNN model recognizes the dominant emotion expressed by the student's face at each moment. The classification is based on the emotion shown in the facial expression which can be one of the seven categories: Angry, Disgust, Fear, Happy, Sad, Surprise or Neutral.

Engagement Classification

The concentration index is calculated using the confidence score of dominant emotion and emotion weights. The resultant concentration index, a score between 0% and 100%. is used to classify the student's level of engagement in one of three categories: Very Engaged, No normally Engaged and Not Engaged.

The system operates according to the following main steps:

- Step 1: The student logs into the learning environment and the camera starts image acquisition. Methods.
- Step 2: The face is detected and processed
- Step 3: The eyes region is detected and cropped
- Step 4: The student’s attention state is classified in Distracted or Focused I

User Requirement:

As we are making a project “Student’s Distraction Detection” so the prime goal is to monitor the student behavior and attention level.

This automatic system that allowed the faculties to capture and make a summary of student behaviors and detect student weather attentive or distracted. Here we present a system to detect engagement level of the students. It is uses only information provided by the typical built-in web-camera present in laptop or computer, and was designed to work in real time. We combine information about the movements of the eyes and head, and facial emotions to produce a concentration index with three classes of engagement: “very engaged”, “nominally engaged”, and “not engaged at all”. This information can be viewed by faculty members and
teachers. Faculty can view complete analysis of student’s behavior; a teacher can’t see student’s behavior at a time this system will provide behavior information to faculty for every 5 minutes.

Facial movement detection and human emotion recognition are the two popular research areas. Recently, these research methodologies are being used for distraction detection. Proper attention is a must for students at classroom.

**Anaconda:**

Anaconda is a conditional free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment. The distribution includes data-science packages suitable for Windows, Linux, and macOS. It is developed and maintained by Anaconda, Inc., which was founded by Peter Wang and Travis Oliphant in 2012. As an Anaconda, Inc. product, it is also known as Anaconda Distribution or Anaconda Individual Edition, while other products from the company are Anaconda Team Edition and Anaconda Enterprise Edition, both of which are not free.

Package versions in Anaconda are managed by the package management system conda. This package manager was spun out as a separate open-source package as it ended up being useful on its own and for other things than Python. There is also a small, bootstrap version of Anaconda called Miniconda, which includes only conda, Python, the packages they depend on, and a small number of other packages.

**TensorFlow and Keras:**

Anaconda does also work as a package manager. So, it does provide an option to install required packages directly from the Anaconda Navigator.

Type TensorFlow in the search box, select the TensorFlow and apply to install the compatible TensorFlow package. Now type in Keras and install Keras library to our environment.

**TensorFlow:**
TensorFlow is a tool for machine learning. While it contains a wide range of functionality, TensorFlow is mainly designed for deep neural network models.

**Keras:**

Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano.

**OpenCV:**

OpenCV was started at Intel in 1999 by Gary Bradsky, and the first release came out in 2000. Vadim Pisarevsky joined Gary Bradsky to manage Intel's Russian software OpenCV team. In 2005, OpenCV was used on Stanley, the vehicle that won the 2005 DARPA Grand Challenge. Later, its active development continued under the support of Willow Garage with Gary Bradsky and Vadim Pisarevsky leading the project. OpenCV now supports a multitude of algorithms related to Computer Vision and Machine Learning and is expanding day by day.

OpenCV supports a wide variety of programming languages such as C++, Python, Java, etc., and is available on different platforms including Windows, Linux, OS X, Android, and iOS. Interfaces for high-speed GPU operations based on CUDA and OpenCL are also under active development.

OpenCV-Python is the Python API for OpenCV, combining the best qualities of the OpenCV C++ API and the Python language.

**New Features:**

**Facial Landmark Detection:**

Facial landmark detection component (LV) detects different relevant facial landmarks from face. Not all landmarks on a face are essential for attention level change detection. Hence, after detecting landmark points we consider only those points that are required for our model's Emotional AI attribute model. Then, only the required points are used for one-time computation.
Emotional AI Attribute Model:

Among the facial landmarks, not all of them are relevant for distraction detection, therefore, this work proposes Emotional AI attention model to identify only relevant attributes for attention level change detection. One of the major contributions of this work is that we develop an Emotional AI landmark model that identifies the efficient relevant facial attributes using facial landmarks for tracking individual’s attention level.

Head Posture Alignment:

Head posture has become one of the necessary attributes for attention decrease detection. Turning head posture is directly related to attention level change. If the head is straight or forward of an individual learner, the learner is paying attention. However, if the head is leaning back, rotating or translating the learner is potentially distracted.

Eye Aspect Ratio:

Scientific measuring believes that sleep deprivation leads to lower alertness and concentration. So, we choose sleepiness or drowsiness as another vital attribute for our distraction detection model. To detect drowsiness, we count the eye closure frequency of a person for every minute. Whenever this frequency is larger than a normal threshold, drowsiness is detected.

Eye Direction:

As another attribute of our distraction detection model, we select eye direction. In case of concentration, eye’s pupil supposed to stay at center. A person is assumed distracted if his eye pupil directs to the left or right direction for certain period of time.

Non-Functional requirements

- The system will provide reliability and quality
- Reduce learners’ frustration and dropout rates
- Useful for online learning
- The online courses can take advantage of this technological advancement for personalized intervention design
CHAPTER – 04
Project Plan & Initial Design Total

How we planned to complete this project with in time.

Our Project plan as follows:

1. Design specification
2. Full document
3. Purchase of equipment
4. Equipment in place
5. Development
6. Testing
7. Bug release
8. User manual / training
9. Delivery of Final project

Currently we are on Development phase and we have covered following things:

- GUI
- Facial Detection
- Data set

GUI:
We have developed initial GUI for the project which will user see when he runs the program.

Facial Detection:
We have successfully developed a program that has ability to detect facial expression

Data Set:
Now we are working on dataset that wasn’t proper so we decided to make it perfect, first we extracted frames from the dataset then we labeled them then we have converted those labeled frames in arrays to make proper data set

Our next will be to train model for engagement recognition
Different modules of our project and estimate of how long will each module take, there are also some challenging modules that require more time, than we must mention it and also the give detailed analysis.

**Module 1:**
In week 1 we gather the data for facial expressions detection.

**Module 2:**
In week 2 we setup tools and libraries installation.

**Module 3:**
In week 3 we load dataset and build an initial face landmark detector.

**Module 4:**
In week 4,5 and 6 we Work on Pose Estimation and Face Feature Extraction

**Step #1:** Checking the face of the student by using frames.

**Step #2:** Detect the features and estimate the pose.

**Module 5:**
In week 7 we on face expression detection.

**Module 6:**
In week 8 we build initial GUI for the user.

Working on Face Expression Detection and GUI building

**Step #1:** Detecting facial expression.

**Step #2:** Building GUI.

**Module 7:**
In week 9 we gather the data for engagement recognition. Read the data and understood the logistics.

**Module 8:**
In week 10 we create a new dataset from video clips and frames extraction for pose estimation.

Creating a new dataset for Pose Estimation.

**Step #1:** Downloading video clips and frames from a different dataset.

**Step #2:** Creating a new data set for pose estimation by using video clips and frames.
<table>
<thead>
<tr>
<th>Task</th>
<th>Start Date</th>
<th>Duration</th>
<th>Finish Date</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>In Progress</td>
</tr>
<tr>
<td>Test</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Design Review and Tools</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Necessary Licenses Installed</td>
<td>11/20/20</td>
<td>11/20/20</td>
<td>11/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Final Deliverables</td>
<td>11/20/20</td>
<td>11/20/20</td>
<td>11/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Dedication</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Regression</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Combination of Feed</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Expression and Decision</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Management of Code</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>12/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Front-end Development</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Deployment</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Finding Bugs</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Black Box Testing</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>Gray Box Testing</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
<tr>
<td>TD</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>10/20/20</td>
<td>Complete</td>
</tr>
</tbody>
</table>
All diagrams including use case, sequence, activity and DFD
Prototype (blueprint) or initial design of our UI, which was made before starting the project.
Attentive... 67.00
Eating... 56.90
Smoking... 44.045
Destructive... 60.4
Using Cell Phone... 63.5
Not Present... 34.03
CHAPTER – 05
Chapter Title: Project Design & Development

The actual design and implementation of our project and Front End Design: This includes color theme, logo, template, screenshots of user interface of all screens and web pages etc. with description.

We are using a unique transparent AutoMonitor LOGO for our Student engagement detection system the heading of Logo defines the system name and the tag line describe organization.

We are using the dark Blue multi colored GUI for our system that contains a window for real time video capture and a side window for progress graphs and reports.
Backend Design: this includes physical System Architecture
Different stages of development, code, algorithms and generating results screenshots, screenshots of our project structure, classes and resources.

The computer vision based methods offer a number of ways to measure learners’ engagement by investigating the cues from the gestures and postures, eye movement, and facial expressions.

Facial expressions

It has been hypothesized that a good deal of information used by humans to make engagement judgment is based on human faces, and facial expressions are directly linked to the perceived engagement.

Gestures and postures

Gesture and postures are two important forms of non-verbal communication through our body language.

Eye movement

Users’ gazes and regions of interests from eye trackers have been used to understand the moods of learners while engaging in any educational activity in online learning.

Dataset and evaluation techniques

Information about open and online datasets and knowing about metrics and evaluation techniques might be of great interest for researchers in any research field. We are generating our new engagement detection data set for detecting engagement level.

Evaluation techniques and metrics

Different metrics and techniques are used to evaluate the performance of engagement detection systems. A widely used technique for evaluating the performance of engagement detection is the investigation of correlation between human and automatic perceptions of engagements.
DAiSEE dataset with an intent to capture learners’ engagement in online courses. This dataset includes 112 individuals, where 80 male and 32 female. The videos in the dataset were collected in unconstrained environments, such as at dorm rooms, crowded lab spaces and libraries, with three different illumination settings—light, dark, and neutral.
Screenshots of a sample run of our project

Working SS of our first facial expression detection model on FER2013 DataSet
Libraries we used also the version of language and platform used to develop the project

```
[condaenv]
python = 3.6.9
anaconda = 2019.11
```

- **conda**
  - 2019.11
  - Required for project development.

- **Python 3.6.9**
  - Essential for project implementation.

- **Anaconda**
  - 2019.11
  - Used for managing the project's dependencies.

These tools were crucial in ensuring the project's smooth execution and maintaining its compatibility across different platforms.
CHAPTER-06

Testing

We have separated our data set in to 2 parts one is for training and the 2\textsuperscript{nd} is for testing we can also test the model by applying the model on different faces of the persons.

Machine Learning is changing the way software products and applications think and respond to queries. In the strife to provide our machines with our intelligence, we are empowering our software to think for them and proactively learn based on experiences, just what we do!

Machine Learning results in areas of image, speech, and text recognition, disease detection, human extension, error reduction, and stock prediction. It has a wide array of applications in all major industries in the world like Healthcare, Retail, Life Sciences, Manufacturing, Enterprise, Education, Medicinal sciences, research and development, and Finance and Accounting.

When we talk about testing these Machine Learning projects, we are often confused about how this can be tackled. Software Testing Services, by definition, is a fairly straightforward task. For every input, there is a definite output. We enter values into the software application, choose to make some processing, and then check the outcome. On the basis of the expected outcome and the actual outcome, we ensure rightness or wrongness of a software products feature.

The main point here is that we already have in mind the expected output. But, Machine Learning is one scenario that does not depend on the output in a test case. Machine learning is based on neural networks which are layered algorithms whose variables are adjusted according to a learning process. The learning process effectively comprises of using known data inputs to get the outputs which are then compared to the known results. When the algorithm code reflects the known results with the desired degree of accuracy, we freeze the algebraic coefficients and generate the production code.
CHAPTER-07
Evaluation & Conclusion

### Eye Gaze to Attention Weights

<table>
<thead>
<tr>
<th></th>
<th>Open</th>
<th>Semi-Open</th>
<th>Close</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centre</td>
<td>5</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Upright</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Upleft</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Right</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Left</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Downright</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
<tr>
<td>Downleft</td>
<td>2</td>
<td>1.5</td>
<td>0</td>
</tr>
</tbody>
</table>

### Emotion Weights

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.9</td>
</tr>
<tr>
<td>Happy</td>
<td>0.6</td>
</tr>
<tr>
<td>Surprised</td>
<td>0.6</td>
</tr>
<tr>
<td>Sad</td>
<td>0.3</td>
</tr>
<tr>
<td>Anger</td>
<td>0.25</td>
</tr>
<tr>
<td>Fearful</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Concentration Index

<table>
<thead>
<tr>
<th>Engagement Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Engaged</td>
<td>&gt;65%</td>
</tr>
<tr>
<td>Engaged</td>
<td>25-65%</td>
</tr>
<tr>
<td>UnEngaged</td>
<td>&lt;25%</td>
</tr>
</tbody>
</table>

100% = 4.5 = (5\*0.9)
X, y, eye-size, mapping

<table>
<thead>
<tr>
<th></th>
<th>Open</th>
<th>Semi-Open</th>
<th>Close</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye Size</td>
<td>&gt;0.3</td>
<td>0.2 - 0.3</td>
<td>&lt;0.2</td>
</tr>
<tr>
<td>Center</td>
<td>x = 1~2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up</td>
<td>y &lt;0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Down</td>
<td>y&gt;=0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>x &gt; 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>x &lt; 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x * (1-y) * size * emotion weight</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

https://github.com/balram2697/Face-Emotion-Recognition

**Limitations**

- The performance of our proposed method still has limitations and possibilities for further improvement. In terms of valence prediction, the proposed light-weight CNN architecture still has a gap compared with the method proposed by Mollahosseini et al. We understand that there is a correlation between facial expression classification and valence/arousal prediction which is the possibility of further improving the performance. The alternative to two separated networks, implementing Multi-Task Learning (MTL) in one network structure is an alternative solution. Future work will address these limitations of the proposed methods and further improve the performance of facial affective computing.
• It very well may be inferred that various computer vision and ML strategies can be utilized to perceive predefined stances including hand gestures. The CNN model offered a preferred execution over different methodologies however for certain limits. The calculation needs high computational assets making hard to be applied in certain conditions with basic equipment engineering (e.g., inserted frameworks). Additionally, preparing a CNN needs a lot of information, which is likewise hard to acquire in certain situations.

**Challenges**

• Teachers must be able to monitor students’ behavior and identify valid cues in order to draw conclusions about students’ actual engagement in learning activities.
• Teacher training can support (inexperienced) teachers in developing these skills by using videotaped teaching to highlight which indicators should be considered. However, this supposes that (a) valid indicators of students’ engagement in learning are known and (b) work with videos is designed as effectively as possible to reduce the effort involved in manual coding procedures and in examining videos.
• One avenue for addressing these issues is to utilize the technological advances made in recent years in fields such as machine learning to improve the analysis of e-learning videos. Assessing students’ attention-related processes through visible indicators of (dis)engagement in learning might become more effective if automated analyses can be employed. Thus, in the present study, we validated a new manual rating approach and provided a proof of concept for a machine vision-based approach evaluated on pilot e-learning recordings of three lessons with university students.
• The discussion focuses on challenges and important next steps in bringing the automated analysis of engagement to the e-learning.
• The level of students’ (dis)engagement in learning activities can be considered a major indicator of both cognitive activation and e-learning management because it signals students’ engagement in the deep processing of learning content and reveals the time on task provided by the teachers for students’ learning. To this end, teachers are required to take note of their students’ attentional focus and...
make sure the students are engaging in the desired learning activities. Thus, the ability to monitor students’ attention and to keep it at a high level is part of the competencies that novice teachers need to acquire.

However, research has indicated that teachers might not always be aware of their students’ attentional focus, and this may be particularly true for novice teachers.

- In general, beginning teachers have trouble monitoring all students in the e-learning evenly and noticing events that are relevant for student learning.
- Consequently, providing an improved measurement approach for student attention will be beneficial for research and can potentially contribute to teacher training. Research has already demonstrated that both inexperienced and experienced teachers’ ability to notice relevant cues in the e-learning benefits from observing and reflecting on their own videotaped teaching.
- The required effort further increases when investigating students’ attention across an entire lesson and analyzing attention at the group level instead of among individuals. In this vein, attention- and engagement-related behavior during e-learning instruction has rarely been studied due to the difficulty of data collection and labeling.
- In addition, we performed an automated analysis of the video recordings to extract features of the students’ head pose, gaze direction, and facial expressions using modern computer vision techniques. Using these automatically extracted features, we aimed to estimate manually annotated attention levels for each student. Because we had continuous labeling, this could be done by training a regression between the visible features and the manual labels.
- Student attention is a key construct in research on both teaching and learning. However, definitions vary widely and are discussed from multiple perspectives. Learners’ attention also plays a crucial role in research on teaching. Teachers must determine whether their students are attentive by considering visible cues, continually monitoring the course of events in order to manage the e-learning successfully (Wolff et al. 2016) and providing ambitious learning opportunities. A student’s attention or lack thereof (e.g., when distracted or engaging in mind
wandering) can signal whether she or he is on-task or off-task. This in turn can provide hints about instructional quality and the teacher’s ability to engage his or her students in the required learning activities.

**Discuss Future work**

Computer vision has gotten broadly progressed and various complex issues can be addressed utilizing various parts of Computer vision. One of the angles is facial milestone location and process the change, pivot and different highlights from facial tourist spots over the long haul. Great examination works are continuous regarding facial acknowledgment, milestone identification, feeling location and so forth. In any case, there are not very works in the territory of human association and mental investigation dependent on the outward appearance and body development. For human Computer connection it is essential to decide how conduct of individual changes after some time. Additionally, it is critical to discover the visual parts of conduct change. For instance, if an individual is yawing regularly over the long haul, it very well may be a sign of languor, sleepiness or fatigue. Additionally, if an individual is pivoting head every now and again, that could be a sign of various conduct for various settings. On the off chance that a student is looking left or right, too regularly, rather than taking a gander at the whiteboard in a homeroom and the time span of every revolution is relatively excessively long or more than an edge, it means that absentmindedness. Similarly, if a driver is looking left or appropriate for a more drawn out time span that may cause a genuine street mishap and the driver needs a quick ready. On the off chance that we could recognize these oblivious practices, it is gainful to numerous areas. For example, on the off chance that we can distinguish scatterbrained understudies in a homeroom through insightful reconnaissance framework, we can help them for being dynamic and furthermore tell the instructors who have more negligent understudies in his study hall to be more cautious about understudies where both dynamic and occupied understudies are perceived by interruption recognition framework. Recognizing the time span of animation and interruption for every understudies of a homeroom and giving a factual examination those outcomes will help both the understudies and the guardians and educators of those
understudies. This will be a capable commitment if there should be an occurrence of guaranteeing viable learning just as e-learning.

An effective e-learning stage needs to address the above difficulties. Likewise, study hall adapting should be more intelligent as e-learning. Thus, this work proposes a misleadingly canny calculation dependent on facial arrangement and distinctive facial characteristic investigation. Our primary destinations are: (1) to build up an interruption identification model for human consideration level (i.e., mindful or occupied) location, (2) to locate the best facial ascribes that characterizes consideration and interruption, (3) to ascertain the inferences inside those credits, (4) to screen the deviation in facial arrangement and related facial credits dependent on an edge model over the long run, (5) to utilize a passionate AI model to distinguish any adjustment in human consideration level, (6) to recognize the adjustment in consideration level and consistently checking them, (7) to give a factual aftereffect of human consideration and interruption period after some time.

In our proposed model, we select the best ascribes whose inference characterizes interruption of human. All applicable facial ascribes are extricated utilizing facial milestones from consistent video transfer, along these lines, this calculation can identify changes in consideration level progressively.

**Facial landmark detection (LV)**

Facial milestone identification segment (LV) recognizes distinctive applicable facial tourist spots from face. In this proposed model, rather than utilizing open source devices like: OpenFace and OpenVino and getting enormous video investigation results, just significant credits for interruption recognition are processed utilizing facial tourist spots. A pre-prepared milestone recognition model of dlib library is utilized to assess the area of 68 (x, y) - milestone arranges that guide to facial designs as demonstrated (a). Not all milestones on a face are fundamental for consideration level change identification. Consequently, figuring the deviation for each of the 68 milestones registered by OpenFace isn't frequently required. Henceforth, subsequent to recognizing milestone focuses we consider just those focuses that are needed for our model's Emotional AI quality model. At that point, just the necessary focuses are utilized for one time
calculation. Shows the chose milestone focuses utilized for distinguishing eye, head and lip correspondingly. This work utilizes the troupe of relapse trees based location calculation for facial milestone recognition. 68 tourist spots on an individual's face recognized via milestone location model.

**Emotional AI attribute model (ML)**

Among the facial tourist spots, not every one of them are significant for interruption location, along these lines, this work proposes Emotional AI consideration model to recognize just applicable ascribes for consideration level change discovery. For instance, in a homeroom, it is essential to distinguish the ascribes that impact the consideration of understudies just as to identify the adjustment in their conduct and consideration level continuously. Enthusiastic AI can contribute essentially to creating successful practices for various game characters in gaming stage advancement. Additionally, a viable enthusiastic AI model can recognize the applicable credits that best identifies with the human consideration level. One of the significant commitments of this work is that we build up an Emotional AI milestone model that recognizes the proficient pertinent facial credits utilizing facial tourist spots for following person's consideration level. Portrays the rundown of traits that develop the Emotional AI quality model for consideration change identification. We have finished 4 distinct ascribes i.e., Head Posture arrangement (H), Eye Aspect Ratio (Er), Eye Direction (Ed) and Lip Distance (Ld) as applicable facial credits for interruption identification. Utilizing head (H) characteristic we process whether the individual isn't going straight or heading at left, right, top or base. Utilizing eye proportion (Er) property we figure whether the individual is habitually squinting and feeling languid or sluggish. Utilizing eye heading (Ed) trait we register if the individual is gazing directly. Utilizing lip (Ld) characteristic we register whether the individual is talking or yawning. All these fundamental ascribes add to human interruption recognition and make the model more productive. Accordingly, this Emotional AI trait model can be utilized for such countless areas where interruption discovery is important and reached out for different sorts of conduct location.