Focus and Engagement Level Detection Using Computer Vision and Machine Learning in a Classroom Environment

Author

Hasnain Ali Poonja

Regn Number

317898

Supervisor

Dr. Muhammad Jawad Khan

Robotics and Intelligent Machine Engineering

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

MAY 2023

Focus and Engagement Level Detection Using Computer Vision and Machine Learning in a Classroom Environment.

Author

Hasnain Ali Poonja

Regn Number

317898

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

MS Robotics and Intelligent Machine Engineering

Thesis Supervisor:

Dr. Muhammad Jawad Khan

Thesis Supervisor's Signature: ____

Robotics and Intelligent Machine Engineering

SCHOOL OF MECHANICAL & MANUFACTURING ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

MAY 2023

Thesis Acceptance Certificate

It is certified that the final copy of MS Thesis written by Hasnain Ali Poonja (Registration No. 317898), of Department of Robotics & AI (SMME) has been vetted by undersigned, found complete in all respects as per NUST statutes / regulations, is free from plagiarism, errors and mistakes and is accepted as a partial fulfilment for award of MS Degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in this dissertation.

Signature: ________________________

Date: ____________________________

Dr. Muhammad Jawad Khan (Supervisor)

Signature HOD: ___________________

Date: ____________________________

Signature Principal: ________________

Date: ____________________________

FORM TH-4

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **Hasnain Ali Poonja** having **Regn. No. 317898**, titled "**Focus and Engagement Level Detection Using Computer Vision and Machine Learning in a Classroom Environment**", be accepted in partial fulfillment of the requirements for the award of MS Robotics & Intelligent Machine Engineering degree.

Examination Committee Members

1. Dr. Hassan Sajid Signature: _______________

2. Dr. Kashif Javed Signature: _______________

3. Dr. Usman Bhutta Signature: _______________

Supervisor: Dr. Muhammad Jawad Khan Signature: _______________

______________________	______________________
Date	Head of Department

COUNTERSINGED

______________________	______________________
Date	Dean/Principal

Declaration

I certify that this research work titled "*Focus and Engagement Level Detection Using Computer Vision and Machine Learning in a Classroom Environment"* is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student
Hasnain Ali Poonja
317898
2023-NUST-MS-RIME-00031798

Plagiarism Certificate (Turnitin Report)

This thesis has been checked for Plagiarism. The Turnitin report endorsed by Supervisor is attached.

Signature of Student
Hasnain Ali Poonja
317898

Signature of Supervisor
Dr. Muhammad Jawad Khan

Copyright Statement

- Copyright in text of this thesis rests with the student author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the author and lodged in the Library of NUST School of Mechanical & Manufacturing Engineering (SMME). Details may be obtained by the Librarian. This page must form part of any such copies made. Further copies (by any process) may not be made without the permission (in writing) of the author.
- The ownership of any intellectual property rights which may be described in this thesis is vested in NUST School of Mechanical & Manufacturing Engineering, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the SMME, which will prescribe the terms and conditions of any such agreement.
- Further information on the conditions under which disclosures and exploitation may take place is available from the Library of NUST School of Mechanical & Manufacturing Engineering, Islamabad.

Acknowledgements

I am grateful to my Creator Allah Subhana-Watala for guiding me through each step of this project and for every new idea that You implanted in my mind to improve it. Indeed, I could have accomplished nothing without Your invaluable assistance and direction. Whoever assisted me during the course of my thesis, whether my parents or anyone else, was Your will; therefore, no one deserves praise but You.

I am exceedingly grateful to my cherished parents, who raised me and continued to support me in every aspect of my life. I would also like to thank my advisor, Dr. Muhammad Jawad Khan, for his assistance throughout my thesis, as well as for the Computer Vision and Deep Learning courses he has taught me. I can confidently state that I have not studied any other engineering subject in such depth as the ones he has taught.

Dr. Riaz Uddin's tremendous support and cooperation at the National Center for Robotics and Automation deserve special recognition. Every time I was stuck in something, he provided a solution. Without his assistance and cooperation, I could not have finished my thesis. I value his patience and guidance throughout the entirety of the thesis. I would also like to thank Dr. Hassan Sajid, Dr. Usman Bhutta, and Dr. Kashif Javaid for serving on the advisory and evaluation committee for my thesis.

Lastly, I would like to express my appreciation to everyone who has contributed to my academic success.

Dedication

Dedicated to my exceptional parents and cherished siblings, whose unwavering support and cooperation enabled me to achieve this great success.

Abstract

Due to Covid 19, the global education system has changed toward online learning, which has a high dropout rate. Therefore, it is vital that students maintain their level of interest. Therefore, detection of engagement level alone is insufficient for analyzing and improving learning and teaching techniques. To promote student engagement in STEM and online learning environments, technologies such as AR/VR and Haptics should be implemented. Utilizing facial emotion, body pose, and head rotation, a web-based computer vision system is developed and implemented to identify student involvement levels using webcams during tasks such as online classrooms, haptic interaction, and augmented reality. In addition, an AR and Haptics-based World Map is being designed and developed. To evaluate and compare three types of learning scenarios, namely (1) Traditional, (2) Augmented Reality-based, and (3) Haptics-based, two methods are employed: (1) Trained Computer Vision models are tested for 3 scenarios, and (2) A user study is conducted using the Positive and Negative Affect Schedule (PANAS) Questionnaire and NASA-Task Load Index, from which conclusions are drawn.

The results of a comparison of Traditional, Augmented reality, and Haptics-based learning indicate that Haptics and Augmented Reality-based learning are the most immersive and increase levels of engagement during online learning and STEM training, whereas Traditional learning methods are the least effective during online classes. User studies and computer vision models are utilized to validate the results.

Key Words: Engagement Detection, Engagement Enhancement, Computer Vision, Augmented Reality, Haptics

Table of Contents

Declaration iii

Plagiarism Certificate (Turnitin Report) iv

Copyright Statement v

Acknowledgements vi

Dedication vii

Abstract viii

Table of Contents ix

CHAPTER 1: INTRODUCTION 1

1.1 Problem Statement 2

1.2 Proposed Solution 2

1.3 Expected Outcomes 3

1.4 Methodology 3

1.5 Thesis Overview 4

CHAPTER 2: LITERATURE REVIEW 5

2.1 Introduction 5

2.2 Engagement Detection in Online Learning 5

2.3 Engagement Enhancement using Augmented reality and Haptics. 7

CHAPTER 3: ENGAGEMENT DETECTION 9

3.1 Emotion Detection 10

3.1.1 Dataset 10

3.2 Code Details 12

3.2.1 Importing Libraries and Data Preprocessing 12

3.2.2 ResNet 50 Architecture for Model Training 14

3.2.3 Training Model 16

3.3 Behavioral /Pose Detection 16

3.3.1 Mediapipe 18

3.3.2 Mediapipe Pipeline 19

3.3.3 Subjects: 20

3.4 Head Pose Estimation 21

3.5 Engagement Model Deployment 22

CHAPTER 4: ENGAGEMENT ENHANCEMENT 25

4.1 Design 25

4.1.1 Hardware Design 26

4.1.2 Augmented Reality Mobile App for World Map 27

4.2 Application Details 28

4.2.1 3D Models 28

4.2.2 Application Layout 30

CHAPTER 5: USER STUDY **37**

5.1 Details of User Study 37

5.2 PANAS Questionnaire 38

5.3 NASA Task Load Index 39

CHAPTER 6: Results and Conclusion **41**

6.1 Results 41

6.2 Conclusion 44

REFERENCES **46**

List of Figures

Figure 3.1: Engagement detection block diagram 9
Figure 3.2: Training and validation accuracy on FER 2013 Emotion Recognition Dataset for engagement and disengagement. 12
Figure 3.3: Skip connections 15
Figure 3.4: Typical ResNet-50 architecture.. 16
Figure 3.5: Landmark model in Media Pipe Pose predicting the location of 33 pose landmarks. 17
Figure 3.6: Realtime disengagement example of pose estimation via googles media pipe......... 18
Figure 3.7: Pose landmarks.. 20
Figure 3.8: Projection of point P in world coordinates UVW onto the point p in the image. 21
Figure 3.9: Engagement detection online demo using a computer vision video as an example. 23
Figure 3.10: Emotion, body pose and head rotation in real-time. 23
Figure 3.11: Real time results for engagement detection models deployed on web. 24
Figure 4.1: Raw World Map design.. 25
Figure 4.2: Final World map design using Adobe Illustrator 26
Figure 4.3: Hardware design for World Map 27
Figure 4.4: Capturing key points using Vuforia 28
Figure 4.5: Adding 3D AR object in Unity. 28
Figure 4.6: Design of 3D globe inside Unity.. 29
Figure 4.7: Design of 3D flag inside Unity 29
Figure 4.8: 3D country information block.. 29
Figure 4.9: 30 Countries for which the above scenes are created.. 30
Figure 4.10: AR Application layout. 31
Figure 4.11: Working of AR application. 31
Figure 4.12: Augmented Reality scene.. 32
Figure 4.13: Complete flow of AR App 33
Figure 4.14: 3D Virtual world integrated with haptics using Novint Falcon. 34
Figure 4.15: 3D Virtual world using Touch Haptics 35
Figure 4.16: Immersive STEM using AR and Haptics.. 36
Figure 5.1: PANAS questionnaire 38
Figure 5.2: NASA Task Load Index... 40
Figure 6.1. Mean positive and negative scores for Traditional, Haptics and AR teaching methodologies.. 41
Figure 6.2. Graph displaying a very strong correlation between haptic and augmented reality instructional strategies. 41
Figure 6.3. Task load index for Traditional teaching is highest among the other two methodologies.. 42
Figure 6.4. Individual factors contributing to task load for Traditional, AR and Haptics based learning methods.. 43
Figure 6.5. Temporal load can be seen to dominate and contribute to major factor of task load in AR learning method.. 43
Figure 6.6. Effort to complete the task is the major contributing factor in Haptics based learning. 44

Figure 6.7. Mental workload and frustration are the key contributors for task load in Traditional learning methods. 44

List of Tables

No table of figures entries found.

List of Acronyms

1.	CNN	Convolutional Neural Network
2.	ReLU	Rectified Linear Unit
3.	FC Layer	Fully Connected Layer
4.	SGD	Stochastic Gradient Descent
5.	ANN	Artificial Neural Network
6.	ML	Machine Learning
7.	DL	Deep Learning
8.	AI	Artificial Intelligence
9.	CV	Computer Vision

CHAPTER 1: INTRODUCTION

In global education, there has been a shift toward online learning because of COVID-19. There is a growing need to leverage educational resources and offer online learning possibilities [1]. However, online learning is reported to have a significant dropout rate. There are numerous possible causes. For example, the material or topic may be too lengthy, there may be technical issues, the instructor's delivery style may be insufficient, or the lecture material may be too boring. In order to provide personalized pedagogical support through interaction with online students, it has become crucial in online education to detect student engagement [2-4]. There are typically three methods for engagement detection (i) manual (offline detecting by giving post lecture survey form), semi-automatic (detection by post lecture rapid fire questions from the students) and fully automatic (detecting runtime features via streaming and/or physiological sensors using computer vision (CV), brain computer interface (BCI) etc.). Compared manual and semi-automatic methods., computer vision-based methods for detecting online learning engagement are more promising, non-intrusive, and cost-effective (add references). Hence the data from CV can be used to accurately assess the efficacy of teachers, students, and most importantly, instructional materials during and after the online classes [5]. These data can be used to accurately assess the efficacy of teachers, students, and most importantly, instructional materials.

Nonetheless, it has been demonstrated that conventional teaching techniques are not suitable for online courses, as they require more engaging and immersive content to maintain and enhance student participation. In light of this, STEM education, an instructional approach that combines science, technology, engineering, and mathematics to provide students with significant learning experiences through hands-on design and research activities, is becoming increasingly popular. To facilitate exceptional STEM education, it is crucial to employ a comprehensive curriculum, teaching methods, and evaluation processes, incorporate technology and engineering into science and math lessons, and promote both scientific inquiry and the engineering design process [2]. As a result, employing immersive technologies such as Virtual and Augmented Reality (VR/AR) and touch-based experiences (haptics) is crucial for fostering engagement in STEM education [4].

Therefore, the concepts of engagement detection and enhancement should be addressed concurrently to increase the interactivity of online learning. In this approach, extensive research has been conducted in both streams, namely engagement detection by computer vision and engagement enhancement via AR/VR and Haptics integration.

1.1 Problem Statement

The COVID-19 pandemic has led to a substantial rise in the prevalence of online learning. Nevertheless, it has been noted that students' engagement tends to be relatively low in this learning format. Conventional teaching methods also appear insufficient in maintaining long-term student interest. This diminished engagement may adversely impact students' motivation and scholastic achievements. It is imperative for educators and policymakers to devise creative strategies to enhance student engagement in both online and in-person learning environments, ultimately leading to improved learning results and overall academic accomplishment.

1.2 Proposed Solution

To tackle the problem of low student engagement in online learning, a computer vision system has been developed. This system can measure the engagement level of a student during online learning using facial cues, head rotation, and body pose. This can provide teachers with real-time feedback on the level of engagement of each student, allowing them to tailor their teaching approach to keep students more engaged and focused.

To enhance the engagement level, two modes of learning are used: Haptics-based learning and Augmented Reality-based learning. Haptics-based learning utilizes touch-based feedback to enhance the learning experience, in contrast, learning through Augmented Reality employs computer-generated elements like audio, visuals, or graphics to establish an engaging and immersive educational setting.

To demonstrate the effectiveness of these two approaches, an AR-Haptics-based world map is designed. This interactive map allows students to explore the world and learn about different countries and cultures through touch and sensory inputs, providing an engaging and immersive learning experience. The computer vision system measures the engagement level of each student during this learning activity, showing a significant increase in engagement compared to traditional online learning methods.

Overall, this solution addresses the problem of low student engagement in learning by using innovative approaches that incorporate technology and sensory inputs, creating a more engaging and immersive learning experience for students.

1.3 Expected Outcomes

Improved Student Engagement: The computer vision system, along with the use of Haptics-based and Augmented Reality-based learning, is expected to improve the engagement level of students during online learning. The real-time feedback provided by the computer vision system can help teachers adjust their teaching approach to keep students engaged and focused. The AR-Haptics-based world map is also expected to provide an engaging and immersive learning experience, leading to better engagement levels.

Better Learning Outcomes: With improved engagement levels, it is expected that students will have better learning outcomes. They are more likely to retain the information and skills they have learned, leading to better academic performance and success.

Increased Motivation: Engaged students are more likely to be motivated to learn and participate actively in online learning. This increased motivation can lead to a positive attitude towards learning, promoting lifelong learning habits.

Enhanced Teaching and Learning Experience: The use of innovative technology-based approaches such as Haptics-based and Augmented Reality-based learning, along with the computer vision system, can enhance the teaching and learning experience. Teachers can use these tools to create interactive and engaging learning experiences for students, leading to a more rewarding experience for both teachers and students.

Overall, the expected outcomes from this solution are better engagement levels, improved learning outcomes, increased motivation, and an enhanced teaching and learning experience.

1.4 Methodology

The methodology for this solution can be divided into two main parts: computer vision-based engagement detection and the design of an AR-Haptics based World Map.

Computer vision-based engagement detection using facial cues, head rotation, and body pose. Relevant literature and datasets were explored to identify the most effective cues for engagement detection. Then the most relevant cues for engagement detection were selected, including facial cues, head rotation, and body pose. The selected cues were used to train the engagement detection system using appropriate algorithms, such as machine learning or computer vision techniques.

Design of AR-Haptics based World Map. The world map was designed using Adobe Illustrator software and cut on wood using a CNC machine to create a physical representation of the map. The AR application was designed using Vuforia Engine and Unity to provide an immersive and interactive learning experience for students. A virtual 3D environment was created using Open Haptics and Unity, which provided touch-based feedback to students for a more engaging and interactive learning experience.

The overall methodology involved a combination of data-driven approaches, software design, and physical prototyping to develop an effective solution. By incorporating computer vision-based engagement detection and innovative AR-Haptics based learning techniques, the solution aimed to improve engagement levels and learning outcomes for students in online learning environments.

1.5 Thesis Overview

This thesis aims to address the problem of low student engagement in online learning by proposing a computer vision-based engagement detection system and the use of innovative learning techniques such as Haptics-based and Augmented Reality-based learning. The thesis includes a comprehensive literature review, a detailed methodology for engagement detection and enhancement, a user study to evaluate the effectiveness of the proposed solution, and a results and conclusion chapter summarizing the key findings and recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In global education, there has been a sudden huge shift towards online learning as a result of the COVID-19 pandemic [1]. There is a growing need to leverage educational resources and offer online learning possibilities. However, issues such as less focus and engagement of students in online learning are being reported to have a significant dropout rate that may have numerous possible causes e.g., either the material/topic is too lengthy (making the lecture boring) or there may be issues with the instructor's delivery style. In order to provide personalized pedagogical support through interaction with online students, it has become crucial in online education to detect students' engagement [2-4].

It is interesting to note that engagement is typically conceived of as a three-dimensional construct that consists of an emotional component, a cognitive component, and a behavioral component [6-8]. There are typically three methods for engagement detection: manual (offline detection by providing a post-lecture survey form), semi-automatic (detection by post-lecture rapid-fire questions from students), and fully automatic (detecting runtime features via streaming using computer vision (CV) or physiological sensors as in brain-computer interface (BCI), etc.). Among these methods computer vision-based methods for detecting online learning engagement are more promising, non-intrusive, and cost-effective than manual or semi-automatic methods [5].

2.2 Engagement Detection in Online Learning

The use of computer vision presents the opportunity to estimate a student's level of engagement in an unobtrusive manner by assessing cues from the face, body posture, and hand gestures [8]. Engagement can be measured in two settings: the first, a traditional classroom, and the second, an online environment. Edu sense is an all-encompassing sensing system that was designed, built, and installed in a classroom environment so that it can function in real-time. In a similar manner, a Kinect One sensor and various computer vision techniques were utilized to recognize the learner's facial and body features, such as their gaze point and body posture and common gestures (such as sitting, raising a hand, standing, sleeping, and whispering) [9-12]. Conversely, initiatives have been undertaken to assess the engagement levels of online students. One study created a deep learning-based technique to identify learners' engagement using facial expressions. This method utilized Local Directional Pattern (LDP) to extract person-independent characteristics and Kernel Principal Component Analysis (KPCA) to identify the nonlinear relationships among the extracted features. By using the DAISEE dataset, it achieved a high classification accuracy (90.89 percent) for a binary classification problem (engaged/not engaged). Another study presented an approach to gauge student engagement levels. The researchers devised a concentration metric consisting of three participation levels, based on eye and head movements as well as facial expressions (highly engaged, moderately engaged, and not engaged at all). The results showed a correlation between the highest concentration indices and the top test scores [13-15].

Various facial feature detection methods have been investigated to detect engagement. A method was developed to categorize face images as engaged or disengaged using an engagement model. Transfer learning to train an engagement model with 4627 engaged and disengaged images outperforms deep learning architectures, histogram of directed gradients, and SVMs. Another approach, using Deep Facial Spatiotemporal Network (DFSTN) predicts engagement. SE-ResNet-50 and LSTM Network with Global Attention (GALN) were used to extract facial spatial data. On DAiSEE, the model predicted engagement better than current research. A multimodal method was also presented. That incorporated three modalities that characterize student behavior: facial expressions, keyboard keystrokes, and mouse movement. Another proposed model combined ResNet and TCN neural network architectures in an end-to-end design. ResNet extracts spatial information from sequential video frames, while TCN examines temporal changes. This model surpasses other approaches when applied to the DAISEE dataset [6, 16-18]. Work in [19] utilized facial expression detection and subjective evaluations to assess participants' emotional engagement while performing game or non-game-based numerical tasks, revealing that the emotional appeal of games aids in learning. A unique manual rating method was developed, providing a proof of concept for a machine vision-based technique. The manual ratings were accurately predicted using a machine vision approach that considered gaze, head posture, and facial expressions [20].

According to [2], STEM education is an instructional approach that involves students participating in engineering design or research to gain valuable learning experiences by integrating science, technology, and mathematics. This method is becoming increasingly popular. The study also demonstrated that traditional learning techniques are inadequate for online courses, recommending the use of interactive and immersive content to maintain and enhance student engagement. To involve students in high-quality STEM education, it is essential to establish a thorough curriculum, instruction, and assessment, incorporate technology and engineering into science and math curricula, and promote scientific inquiry and the engineering design process.

2.3 Engagement Enhancement using Augmented reality and Haptics.

Augmented Reality is a type of technology that allows users to see the real environment while also seeing virtual items superimposed on or composited with it. As a result, rather than entirely replacing reality, AR augments it [21, 22]. The interaction between the computer and the human being in connection to or based on the sensation of touch, as defined by the phrase computer haptics. In other words, haptics refers to technology that uses force feedback to allow a user to physically engage with a virtual world as and therefore it has a great significance for STEM education [23].

In STEM Education, various prototypes by using Augmented Reality or Haptics independently have been built in recent years to explain the challenging task of demonstrating immersive lesson learning [4]. AR-based mobile application to enhance the learning experience for students was developed. These applications were deployed for various topics, like teaching vocabulary and character education for kindergarten students [22, 24, 25]. For kids with neuromotor disorders, a virtual reality (VR) training system that includes wearable haptics has been made. It gives consistent multi-sensory afferent feedback during motor exercises and uses the flexibility of VR to tailor workouts to the needs of the patient [26]. A virtual reality classroom where students use their hands to build hydrocarbon molecules and get haptic feedback from gloves with built-in sensors and hand tracking is developed [27]. The work in [28] developed a haptics-based teaching interface that helps students understand gyroscopic precision forces is described. The study in [29] investigated augmented reality (AR) studies that supported STEM education. In this framework, the general status of augmented reality in STEM education was shown, along with its pros and cons. The work in [30] studied haptic design by comparing how novice and expert haptician design. They identified three successful strategies for how teams created useful and engaging interactions and multimodal experiences. [31, 32] proposed methodological guidelines that show how popular entertainment games can be used to create a fun educational game with learning outcomes. Preschoolers' STEM comprehension and transfer outcomes were also studied after playing a STEM game with haptics.

When mixed reality experiences are included in the learning process, students' knowledge improves. Experiments in classrooms can be carried out without causing any harm. AR technology allows students to explore, practice, and interact with STEM subjects without having to worry about costs or ethical considerations [33].

Consequently, immersive learning utilizing Virtual and Augmented Reality (VR/AR) and the sensation of touch (haptics) is essential for STEM engagement learning [4]. Therefore, the concepts of engagement detection and enhancement should be addressed concurrently to increase the interactivity of online learning. In this approach, extensive research has been conducted in both streams, namely engagement detection by computer vision and engagement enhancement via AR/VR and Haptics integration.

In this thesis, the fact that due to Covid 19 the world education system has shifted towards online learning and online learning has a large dropout rate [3] so there is a need that the engagement level of the students should retain. Therefore, only engagement level detection is not enough to enhance learning experience in STEM and online learning environment but the technologies like AR/VR and Haptics should be used to enhance engagement levels. Therefore, a computer vision system is designed and deployed on web to detect engagement levels of students using webcam during task like online classes, haptic interaction and AR, using facial emotion, body pose and head rotation. Furthermore, AR and Haptics based World Map is designed and developed. To evaluate and compare three types of learning scenarios (1) Traditional (2) Augmented Reality based (3) Haptics based, two methods are used (1) Trained Computer Vision models are tested for 3 scenarios and (2) User study is performed using Positive and Negative Affect Schedule (PANAS) Questionnaire and NASA-Task Load Index and results and conclusion are drawn.

CHAPTER 3: ENGAGEMENT DETECTION

According to Newmann (1992) and Ninaus (2019) [19, 34], engagement is the active and focused involvement in a task or learning activity. This contrasts with pretending to be in class despite having a lack of interest, apathy, or superficial participation in the activity. It is interesting to note that engagement is often viewed of as a three-dimensional construct that comprises of an emotional component, a cognitive component, and a behavioral component [6-8]. This notion of engagement having these three components is what makes the concept so interesting. Computer vision techniques is used to detect emotional and behavioral engagement using facial cues and body pose, respectively. This detection is accomplished by analyzing the subject's body pose. While cognitive involvement is something that can be incorporated into the work that will be done in the future.

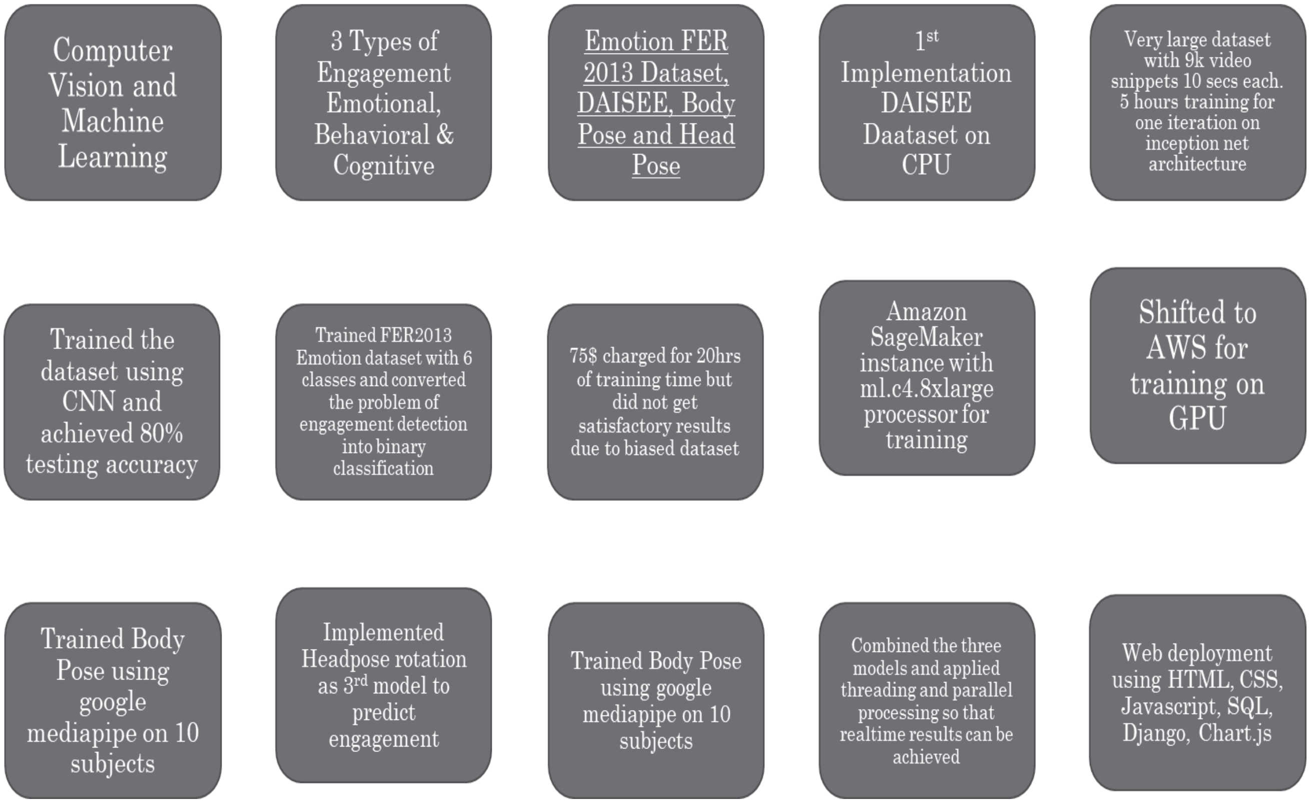

Computer Vision and Machine Learning
3 Types of Engagement Emotional, Behavioral & Cognitive
Emotion FER 2013 Dataset, DAISEE, Body Pose and Head Pose
1st Implementation DAISEE Daataset on CPU
Very large dataset with 9k video snippets 10 secs each. 5 hours training for one iteration on inception net architecture
Trained the dataset using CNN and achieved 80% testing accuracy
Trained FER2013 Emotion dataset with 6 classes and converted the problem of engagement detection into binary classification
75$ charged for 20hrs of training time but did not get satisfactory results due to biased dataset
Amazon SageMaker instance with ml.c4.8xlarge processor for training
Shifted to AWS for training on GPU
Trained Body Pose using google mediapipe on 10 subjects
Implemented Headpose rotation as 3rd model to predict engagement
Trained Body Pose using google mediapipe on 10 subjects
Combined the three models and applied threading and parallel processing so that realtime results can be achieved
Web deployment using HTML, CSS, Javascript, SQL, Django, Chart.js

Figure 3.1: Engagement detection block diagram

3.1 Emotion Detection

Emotional engagement refers to learner's emotional reactions/interest in learning experiences, content, and social connection with teachers and classmates at school [19].

3.1.1 Dataset

FER-2013 [35] dataset, also known as the Facial Expression Recognition 2013 dataset, is a widely-used benchmark dataset in the computer vision and emotion recognition fields. It was introduced by Pierre-Luc Carrier and Aaron Courville as part of the ICML 2013 Challenges in Representation Learning (WREPL) workshop. The dataset consists of 35,887 grayscale facial images, each with a resolution of 48x48 pixels, belonging to seven distinct facial expression categories.

1. The seven facial expression categories in FER2013 are:
2. Angry
3. Disgust
4. Fear
5. Happy
6. Sad
7. Surprise
8. Neutral

The dataset is divided into three subsets:

Training set: 28,709 images

- Public test set (also called the validation set): 3,589 images
- Private test set: 3,589 images

The training set is used for model training, while the public and private test sets are used for model evaluation. The public test set allows researchers to gauge their model's performance during development, while the private test set is used for final evaluation and comparison against other methods.

The FER2013 dataset was collected from the internet, with images representing a diverse range of individuals in terms of age, gender, and ethnicity. These images were preprocessed to ensure consistency in the size and format of the dataset. To provide ground-truth labels, the images were manually annotated by several taggers who were asked to categorize the facial expressions based on the provided categories. Some images in the dataset have been labeled as ambiguous due to the difficulty in discerning the correct expression.

The FER2013 dataset has been extensively used in the development and evaluation of facial expression recognition algorithms, including traditional machine learning approaches, as well as deep learning techniques, such as convolutional neural networks (CNNs). It serves as a valuable resource for researchers aiming to improve the state-of-the-art in emotion recognition and develop models that can effectively recognize and interpret human emotions from facial expressions.

Despite its popularity, the FER2013 dataset has some limitations, such as the relatively low resolution of the images, which may not fully represent the complexity of facial expressions in real-world scenarios. Additionally, the dataset lacks annotations for demographic information, such as age, gender, and ethnicity, which could be valuable for developing more diverse and robust emotion recognition models.

In the proposed research, the seven emotion categories are further grouped into two classes: engagement, representing the positive class (happy, surprise, and neutral) and disengagement, representing the negative class (angry, disgust, fear, and sad). To equalize the number of images in each class, KERAS data generator [36] was employed for data augmentation. The modified FER-2013 dataset was then trained using a deep learning method (RESNET 50 architecture – [37]), achieving training accuracy (92%) and validation accuracy (85%), as illustrated in Figure 2.1: Training and validation accuracy on FER 2013 Emotion Recognition Dataset for engagement and disengagement.

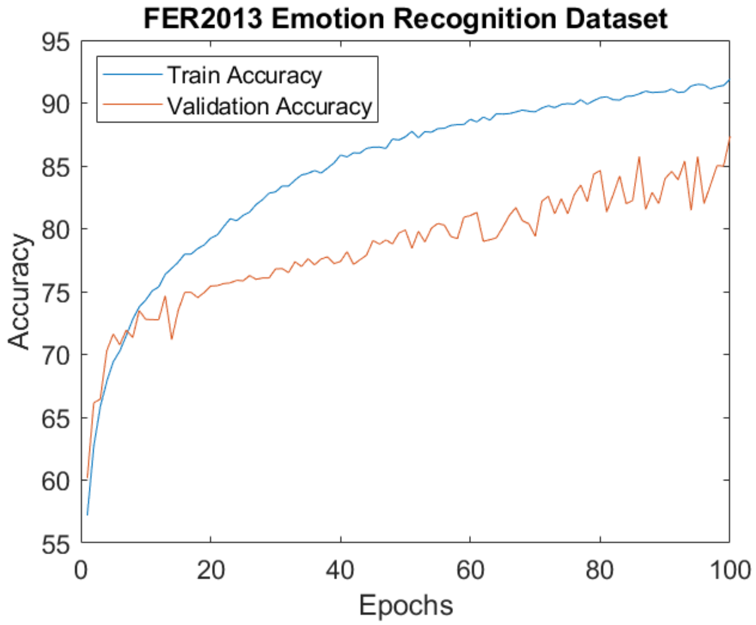


Figure 3.2: Training and validation accuracy on FER 2013 Emotion Recognition Dataset for engagement and disengagement.

3.2 Code Details

3.2.1 Importing Libraries and Data Preprocessing

To code and train FER-2013 dataset most important libraries used were

3.2.1.1 TensorFlow

TensorFlow and Keras. TensorFlow is a free and open-source library that was developed by Google. It has gained a lot of traction in the field of machine learning. TensorFlow provides application programming interfaces (APIs) that make machine learning easier. In addition, the compilation time for TensorFlow is significantly less than that of other Deep Learning libraries, such as Keras and Touch. Both central processing units (CPUs) and graphics processing units (GPUs) can use TensorFlow. Because it was designed from the beginning to process large amounts of numerical data, it is an excellent instrument for use in deep learning. Tensors, which are multidimensional arrays, are the only acceptable form of data for this system. This application is well-suited to the management of significant amounts of data. The fact that the mechanism for execution is in the form of graphs makes it much simpler to put the code into action.

Once data is accessed in TensorFlow, it undergoes computation through a Data Flow Graph, which represents each computation that can be carried out. Nodes are created to form a graph, and these nodes are executed in a session with data from tensors. Each node is assigned to a different mathematical operation, and edges represent tensors, which are multidimensional arrays. After the graph is created, data is processed through the graph. By using TensorFlow, coding becomes easier to write and execute in a distributed way. This is important in deep learning, where models require a significant amount of time to train due to the amount of data involved.

3.2.1.2 Image Data Generator

Keras is a Python-based software library that provides a user interface for creating artificial neural networks and is available under an open-source license. It can be used with the TensorFlow library to interact with external systems. One useful feature of Keras is the ImageDataGenerator class, which can be utilized to enhance images.

Image Augmentation refers to the process of expanding the image training data by applying transformations to the currently available image data. These transformations can include random rotations, shear transforms, shifts, zooms, and flips When we don't have enough training data to properly train our model, we augment images with additional data. In these kinds of circumstances, we can generate new images from the ones that already exist by applying transformations to them. CNN considers these to be entirely new images even though they have a very similar appearance (Convolutional Neural Network). Because of this, we will be able to create a larger training dataset, which will, in turn, allow our model to converge in a more time and effort efficient manner.

The ImageDataGenerator class that is included in Keras makes it possible for us to accomplish the same thing. The ImageDataGenerator is responsible for the generation of batches of tensor image data with real-time enhancement. The data will be processed in batches throughout the loop. We make use of the flow from directory method, which is a generator. This method takes the path to the parent directory that contains various types of image data as an input and generates batches of images that are then fed into the ImageDataGenerator. This is the directory structure that the flow from directory method anticipates having in place.

You should save the images for each class in their own separate folders, and then provide the flow from directory method with the path to the parent directory. Only the edits that we intend to make to your images will be specified within the ImageDataGenerator class.

After this piece of code has been run, the number of augmented images that will be saved in the output folder will be equal to the batch size multiplied by the number of times the code has been iterated. Using the ImageDataGenerator class and the flow from directory method of the Keras library in this manner will result in the enhancement of the images.

3.2.2 ResNet 50 Architecture for Model Training

ResNet50 is a deep residual neural network architecture, which is a part of the ResNet (Residual Network) family. It was introduced by Kaiming He et al. in their 2015 paper "Deep Residual Learning for Image Recognition." ResNet50 has 50 layers, including both convolutional and fully connected layers. The key innovation in ResNet architectures is the residual block, which addresses the vanishing gradient problem in deep networks and enables efficient training of very deep networks.

A residual block consists of several convolutional layers followed by batch normalization and a ReLU activation function. The input of the block is added to its output to create a "shortcut" connection, which is the residual connection. This is the main idea behind residual learning.

Mathematically, let's consider an input x, and the output of the block is F(x). The residual block computes the following function:

$$y = F(x) + x$$

Here, y is the output after adding the input x to the output of the residual block F(x).

The ResNet50 model consists of the following layers:

1. A 7x7 convolutional layer with 64 filters, followed by batch normalization and ReLU activation.
2. A 3x3 max-pooling layer.
3. A series of four stages, each containing multiple residual blocks with different numbers of filters.
4. An average pooling layer.
5. A fully connected layer with softmax activation to produce class probabilities.
6. Each stage doubles the number of filters and reduces the spatial dimensions by half.

Stage 1:

3 residual blocks with 64 filters in each block.

Stage 2:

4 residual blocks with 128 filters in each block.

Stage 3:

6 residual blocks with 256 filters in each block.

Stage 4:

3 residual blocks with 512 filters in each block.

ResNet50 utilizes a bottleneck structure in its residual blocks, which reduces the number of parameters and computational complexity. Each bottleneck residual block consists of three convolutional layers:

1. A 1x1 convolutional layer to reduce the number of channels.
2. A 3x3 convolutional layer that applies the main convolution operation.
3. A 1x1 convolutional layer to restore the number of channels to the original value.

These three layers are followed by batch normalization and ReLU activation, with the exception of the last layer, where the activation function is applied after the residual connection.

ResNet50 has proven to be effective in various computer vision tasks, such as image classification, object detection, and semantic segmentation. Its success can be attributed to the residual connections that enable efficient training of very deep networks and the bottleneck structure that reduces the number of parameters and computational complexity.

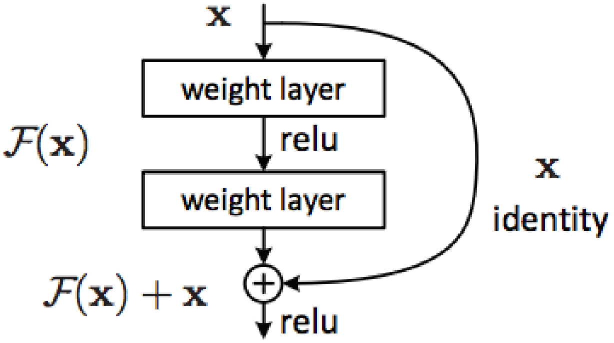

x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 3.3: Skip connections

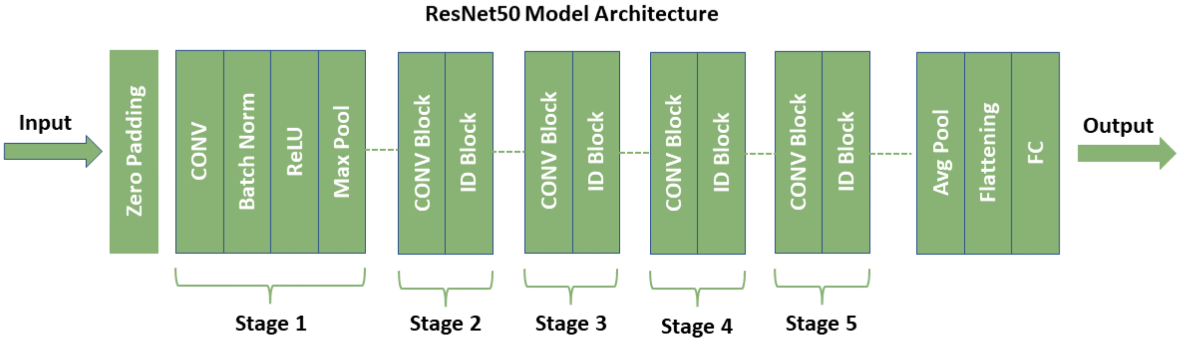


Figure 3.4: Typical ResNet-50 architecture.

3.2.3 Training Model

Optimizer: Adam

Loss: Binary Cross Entropy

Metrics: Accuracy

Epochs: 100

Model Name: Correctedv1

3.3 Behavioral /Pose Detection

Behavioral engagement can be defined in terms of participation, effort, attention persistence, positive conduct, and absence of disruptive behavior. For the implementation of body pose, a customized data set for body pose was created used by a deep learning method for pose detection using media pipe from google to train and test engagement model of body pose. The engaged body poses can be such as Looking towards the camera, Student making notes, Raising hands in class, and Arms closed. Similarly, the Disengaged Poses are Yawning, Hands-on mouth, Fully distracted (Not looking towards camera), Not sitting upright (Dull Posture), and Faces on hand.

For the training of customized dataset, ten (10) subjects (students) of ages 10 to 15 years were chosen for training comprising of 05 males and 05 females. Each subject made 20 secs video of 4 engaged and 4 disengaged poses at 10 fps (frame per secs) i.e. (10 subjects*4(engaged/disengaged poses) *20(secs)*10(fps = 8000 frames for engaged and similarly 8000 frames for disengaged). When the image is given to media pipe library, it gives 33 landmarks for pose detection as shown in. Each landmarks have its own x, y, z coordinates along with a numerically defined visibility factor (0 for missing landmark and 1 for visible landmark) were chosen as features and the data frame with 8000 rows (frames) with 2004 feature columns was made. The created dataset is divided into 3 sets naming Training (60%), Validation (20%) and Testing (20%) sets. Machine learning algorithm like SVM and Random Forest classifier yield accuracy above 95% on training and test set. The real time sample result on a subject is shown in

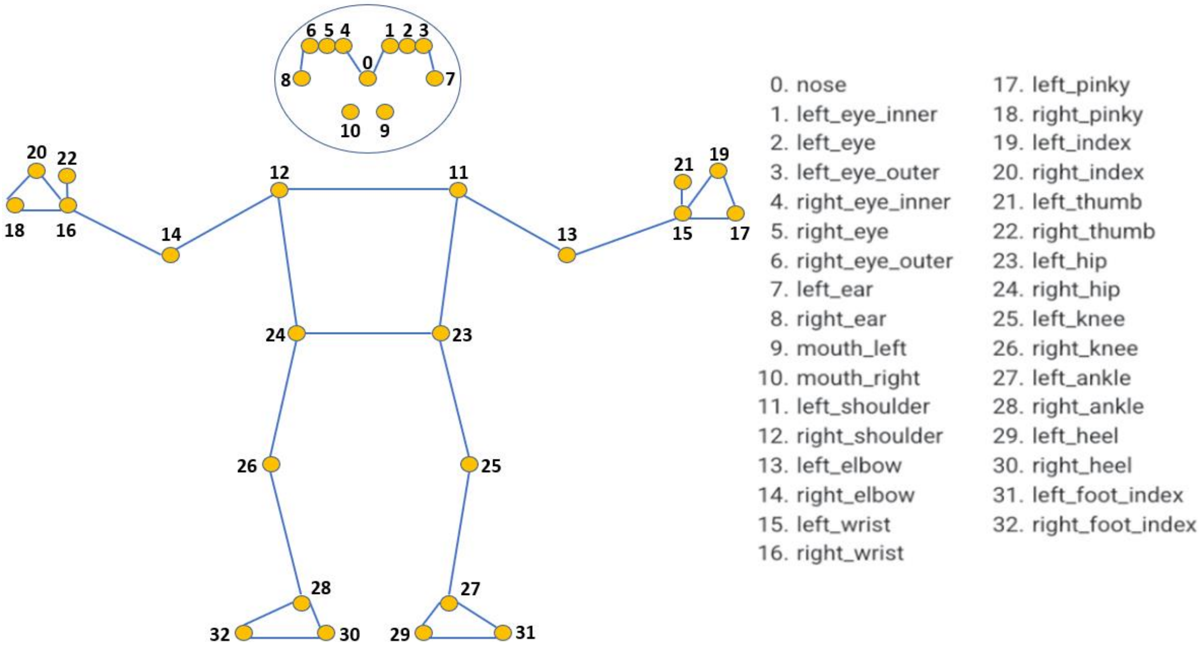


Figure 3.5: Landmark model in Media Pipe Pose predicting the location of 33 pose landmarks.

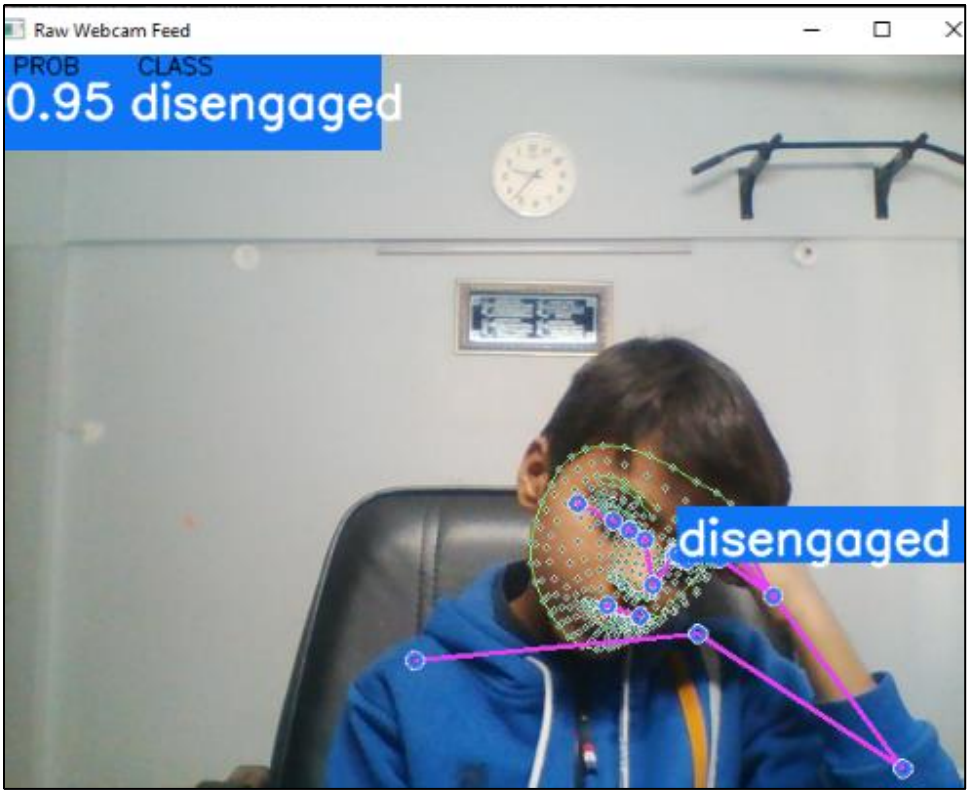


Figure 3.6: Realtime disengagement example of pose estimation via googles media pipe.

3.3.1 Mediapipe

The MediaPipe is a highly efficient framework that enables the development of pipelines capable of making inferences from a wide range of sensory data. It is designed to facilitate the assembly of a perception pipeline by combining a variety of modular components, including model inference, media processing algorithms, data transformations, and other related functions. These components are arranged in a graph format, which can process input streams of sensory data, such as audio and video streams, and output processed information such as object localization and face landmark data.

This framework is primarily targeted at machine learning (ML) practitioners, including researchers, students, and software developers, who are involved in developing production-ready ML applications, sharing code alongside research results, and creating technology prototypes. MediaPipe provides the necessary tools and resources for rapid prototyping of perception pipelines, utilizing reusable components and inference models.

Furthermore, MediaPipe simplifies the incorporation of perception technology into demonstrations and applications that can run on a broad range of hardware platforms. The framework's robust configuration language and various evaluation tools support the iterative improvement of perception pipelines. Our research utilized Google's MediaPipe to implement body pose for training and testing our engagement model, demonstrating the platform's effectiveness in handling complex datasets and its potential for advancing the field of machine learning.

3.3.2 Mediapipe Pipeline

The MediaPipe Holistic pipeline incorporates three individual models for pose, face, and hand components, each tailored for their specific domains. However, using the same input for all three models could lead to inaccurate results, as they have different resolution requirements. For instance, the pose estimation model demands a lower and consistent resolution (256x256), whereas cropping the hand and face regions and passing them to their corresponding models may not provide enough resolution for accurate articulation.

To overcome this challenge, MediaPipe Holistic employs a multi-stage pipeline that processes distinct regions with appropriate image resolutions. Initially, BlazePose's pose detector and subsequent landmark model estimate human pose. This data is then utilized to derive three regions of interest (ROI) crops for the hands (two regions) and face, with a re-crop model employed to refine the ROIs. Next, the input frame is scaled down to these ROIs, and task-specific face and hand models estimate the landmarks for those regions. Finally, all the landmarks from each model, including those from the pose model, are combined to produce the complete set of 540+ landmarks.

The multi-stage approach of MediaPipe Holistic ensures that each region receives the appropriate resolution and specialized treatment, leading to accurate articulation and improved performance. This pipeline's effectiveness is evidenced by its ability to handle complex datasets, enabling the advancement of the field of computer vision and machine learning.

Engaged Body Poses:

1. Looking towards the camera.
2. Student making notes.

3. Raising hands in class.
4. Arms closed.

Disengaged Poses:

1. Yawning hands on mouth.
2. Fully distracted (Not looking towards camera).
3. Not sitting upright (Dull Posture).
4. Faces on hand.

3.3.3 Subjects:

10 students of ages 10 to 15 were chosen for training out of which 5 males and 5 females were there. Each subject made 20 seconds video of 4 engaged and 4 disengaged poses at 10 fps i.e. (10*4*20*10 = 8000 frames for engaged and 8000 frames for disengaged).

Face landmarks (468 landmarks)

Right hand landmark (21 landmarks)

Left hand landmarks (21 landmarks)

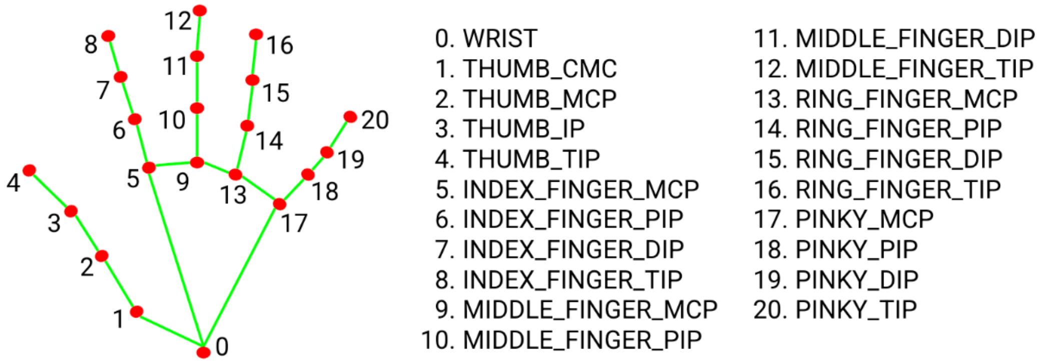


Figure 3.7: Pose landmarks.

Each landmarks has its own x y z coordinate along with visibility factor were chosen as features and the data frame with 8000 rows (Frames) with 2004 feature columns was made. The dataset is divided into 3 sets of Train, Validation and Test set with training set 60%, 20% Validation and 20% test. Simple machine algorithm like SVM and Random Forest classifier yield accuracy above 95%.

3.4 Head Pose Estimation

Head pose estimation plays a crucial role in engagement detection. It can be determined using the Perspective-n-Point (PnP) problem, which estimates the pose of a calibrated camera from n-correspondences between 3D reference points and their 2D projections [41]. As illustrated in Fig. 6, three coordinate systems are employed for head pose estimation: (1) the 3D coordinates of various facial features in world coordinates (UVW), (2) the 3D points transformed from world coordinates to camera coordinates (XYZ) if the rotation and translation (i.e., pose) are known, and (3) the 3D points in camera coordinates projected onto the image plane (i.e., image coordinate system) (XY) using the camera's intrinsic parameters (focal length, optical center, etc.).

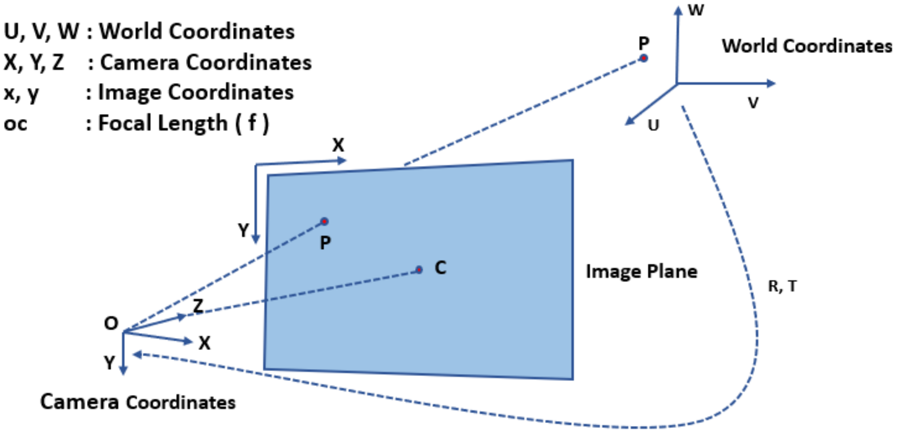


Figure 3.8: Projection of point P in world coordinates UVW onto the point p in the image.

The mathematical representations of transformation from world coordinates to camera coordinates are shown in equation (1).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} r_{00} & r_{01} & r_{02} & t_x \\ r_{10} & r_{11} & r_{12} & t_y \\ r_{20} & r_{21} & r_{22} & t_z \end{bmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix}$$

where the rij is the 3x3 rotation matrix and tx, ty, tz are the translation in x, y, and z axis. In the absence of radial distortion, the coordinates (x, y) of point ‘p’ in the image coordinates is given by equation (2).

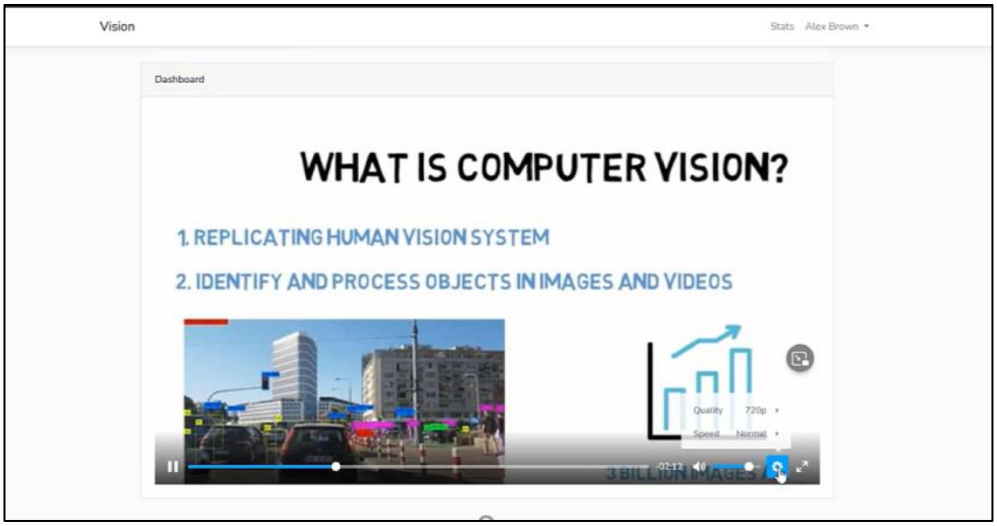
$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \begin{bmatrix} f_x & 0 & C_x \\ 0 & f_y & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

where, f_x and f_y are the focal lengths in the x and y directions, and (c_x, c_y) is the optical center. The OpenCV library (open source computer vision library) [38] is used to solve the PnP problem. The ‘SolvePnP function’ of OpenCV takes the camera’s intrinsic parameters and image-object point pairs as input and returns the transformation matrix **T**. This describes the estimated pose of the object with respect to the stationary camera. Rotations in x, y and z axis is extracted for the moving head. When the face is looking towards camera the angles are zero and hence the label given to the frame is engaged. By hit and trial angles values for x and y axis has been set after which the frame is labeled disengaged.

3.5 Engagement Model Deployment

As the engagement model has been developed in the previous section. Now, it is required to create a complete portal for online courses using, HTML, CSS, bootstrap, JavaScript for frontend, Django and Laravel for the backend as shown in **Figure 3.9:** Engagement detection online demo using a computer vision video as an example. In this regard, trained models are deployed on the portal. Online course videos are prepared and scripts to automatically take a screenshot of the user after every 3 seconds using Camera.js.

At portal, initially the user must register on the portal and signup for the course. As sign-up user enrolls himself for the course, all the videos for the course will be available to the user. A dialog box for the required permission to take the screenshot will be prompted and user will have to allow for that. During the process of taking image snapshots of the user via webcam, they will simultaneously be stored in a database and the trained model will start predicting output labels, that will be finally classified as engaged or disengaged at the admin panel (or teacher panel). Threading is applied to run number of models in parallel and the resulting statistical graphs will be generated using chart.js as can be seen from **Figure 3.11:** Real time results for engagement detection models deployed on web.

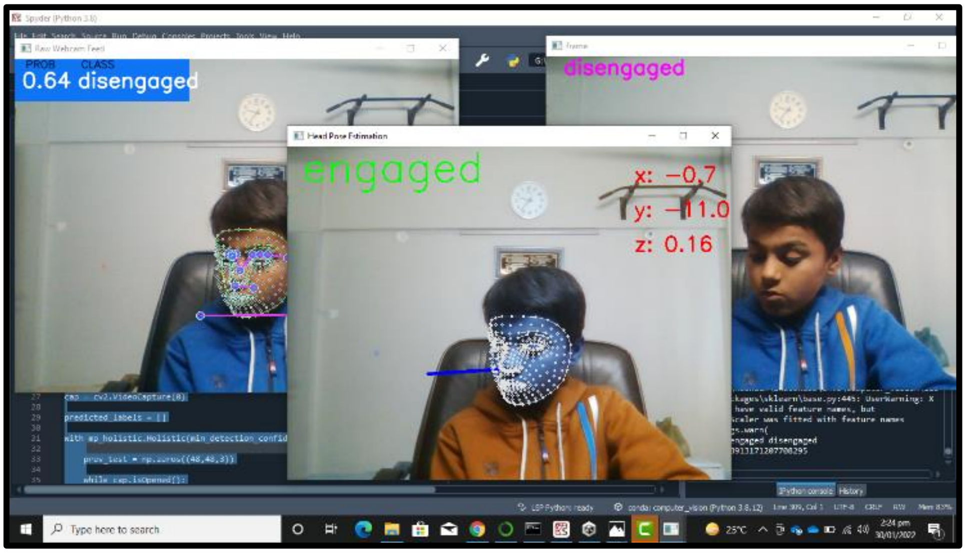


WHAT IS COMPUTER VISION?

1. REPLICATING HUMAN VISION SYSTEM

2. IDENTIFY AND PROCESS OBJECTS IN IMAGES AND VIDEOS

Figure 3.9: Engagement detection online demo using a computer vision video as an example.



```
cap = cv2.VideoCapture(0)

predicted_labels = []

with mp_holistic.Holistic(min_detection_confid

    prev_test = np.zeros((48,48,3))

    while cap.isOpened():
```

```
ckages\sklearn\base.py:445: UserWarning: X
 have valid feature names, but
Scaler was fitted with feature names
gs.warn(
engaged disengaged
0913171287708295
```

Figure 3.10: Emotion, body pose and head rotation in real-time.

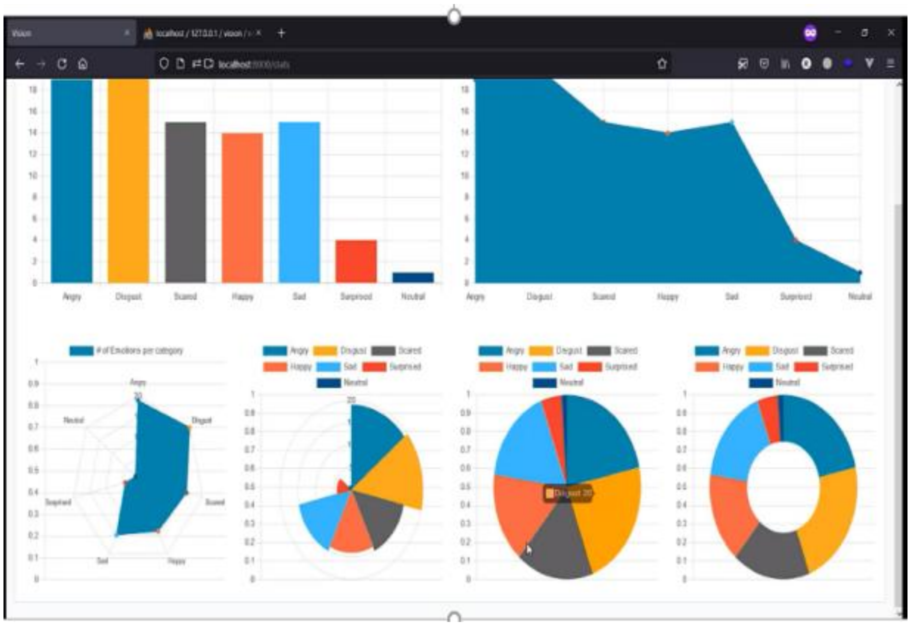


Figure 3.11: Real time results for engagement detection models deployed on web.

CHAPTER 4: ENGAGEMENT ENHANCEMENT

The theme of the world map globe was chosen because it is a generic issue that is studied at various levels in schools and colleges, and there are no difficult concepts that require special effort on the part of teachers.

4.1 Design

Open access image of world map was downloaded from google.

Figure 4.1: Raw World Map design

Using Adobe Illustrator, the raw image was redesigned by adding relevant details like

1. Country name
2. Capital of the country
3. Famous monuments
4. National Animals
5. Continents
6. Oceans

The designed map after adding these details look like this as in the picture below

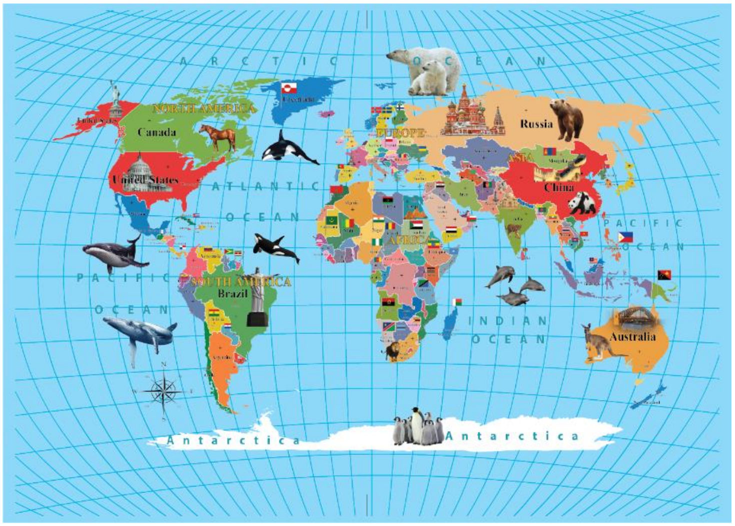


Figure 4.2: Final World map design using Adobe Illustrator

4.1.1 Hardware Design

Final design of the world map was printed. A wooden frame was designed using CNC and the printed map in the sticker form was pasted on the wooden frame. Wooden frame was designed like a book having hinges in between that is easy to carry and light weight. The dimension of the map is 35cm height by 50 cm width.

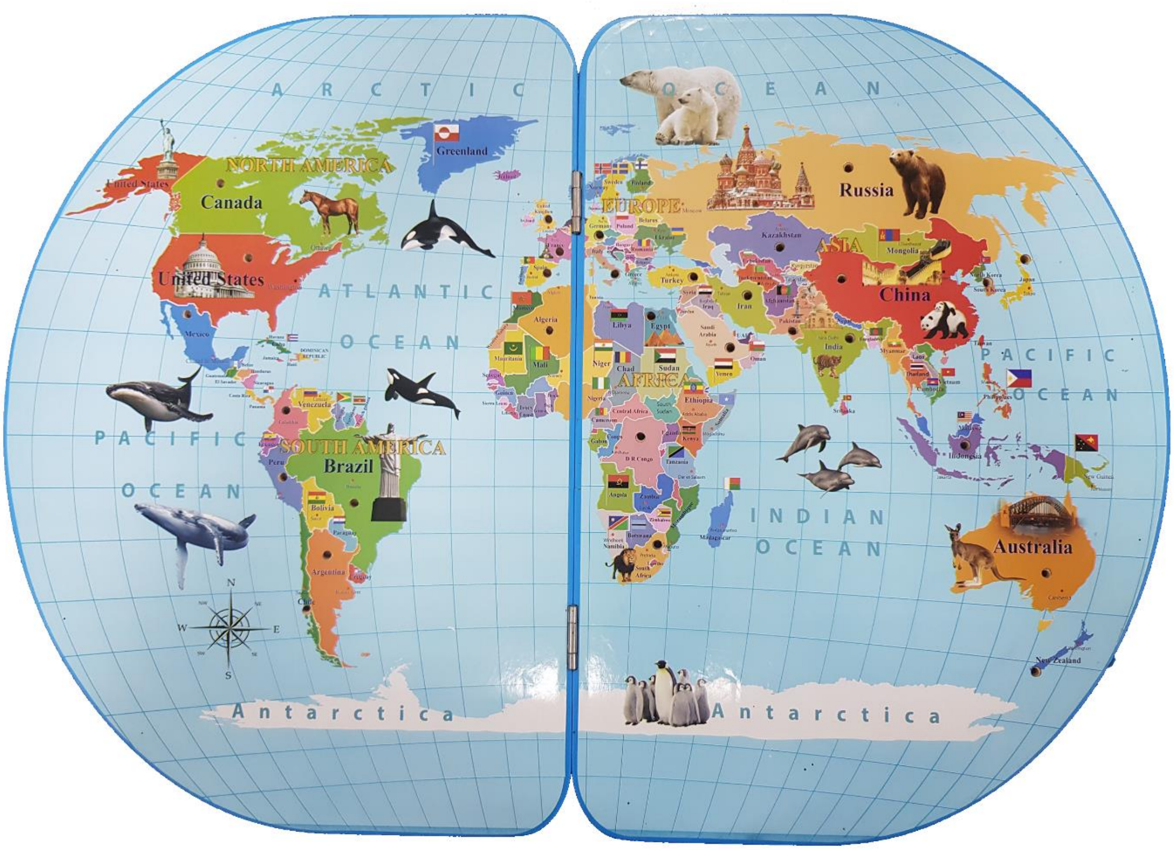
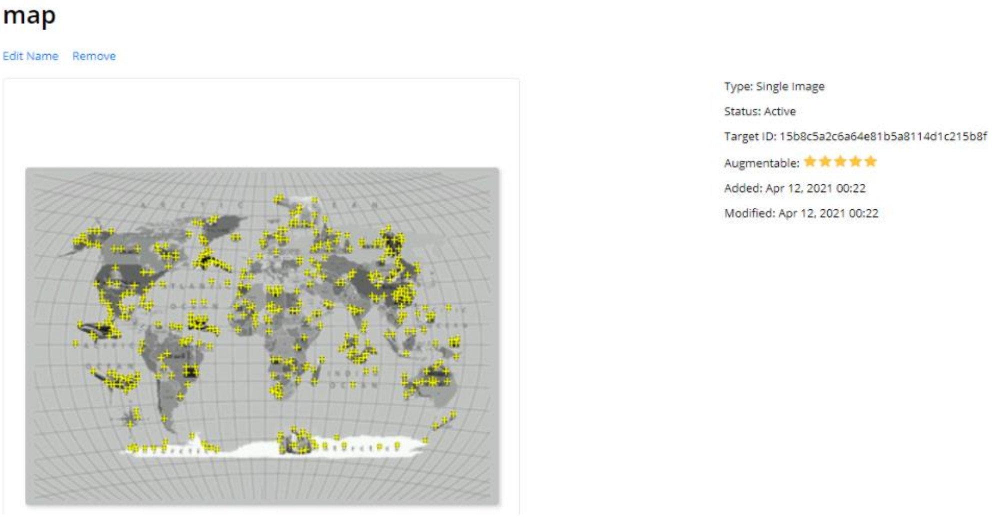


Figure 4.3: Hardware design for World Map

4.1.2 Augmented Reality Mobile App for World Map

Augmented reality based mobile application was designed using Unity3D game engine along with Vuforia SDK that is basically an augmented reality plugin for Unity3D. Image of the world map was uploaded on Vuforia database to check if the image has number of features and is highly augmentable or not. Our image of world map showed five-star rating which indicated that our image is highly augmentable.



map

Type: Single Image

Status: Active

Target ID: 15b8c5a2c6a64e81b5a8114d1c215b8f

Augmentable: ★★★★★

Added: Apr 12, 2021 00:22

Modified: Apr 12, 2021 00:22

Figure 4.4: Capturing key points using Vuforia

4.2 Application Details

4.2.1 3D Models

3D models for globe, flags and information block were designed using Unity3D objects like sphere, cylinders, planes etc. These models were made as prefab so that it can be used for multiple country scenes in Unity. Flags were given cloth effects using Unity Physics Engine.

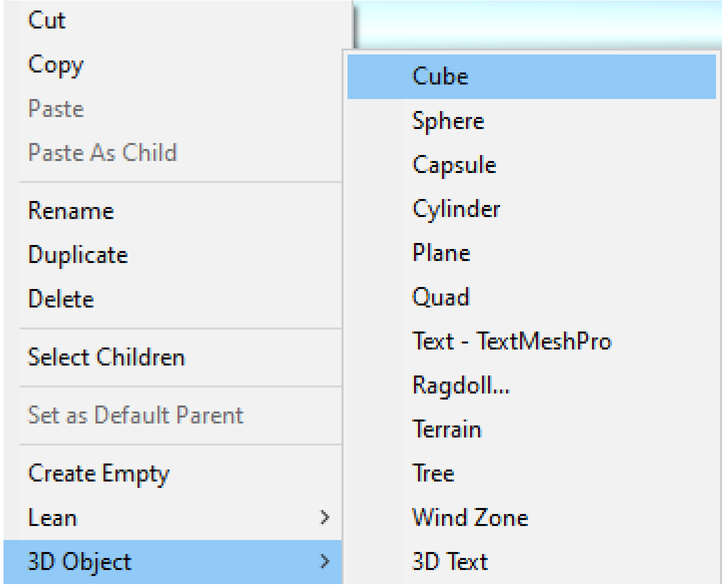


Figure 4.5: Adding 3D AR object in Unity.

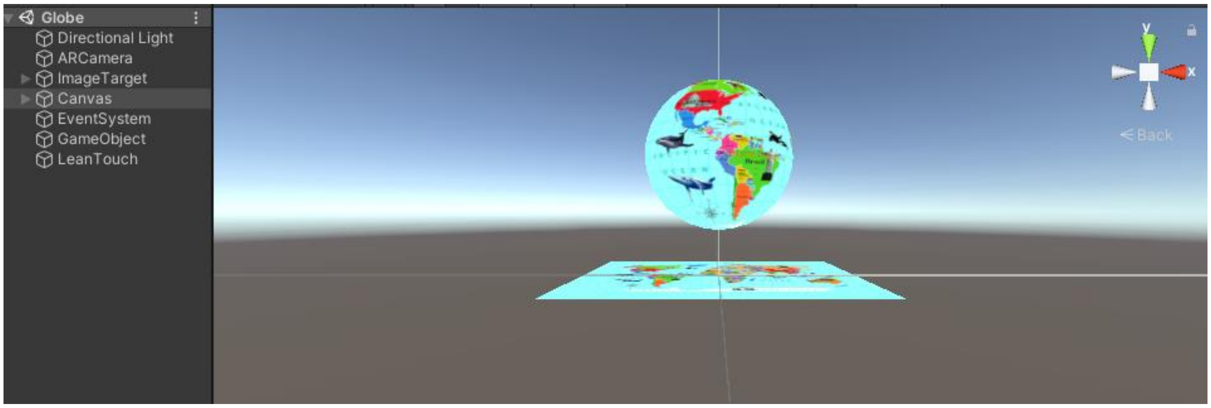


Figure 4.6: Design of 3D globe inside Unity

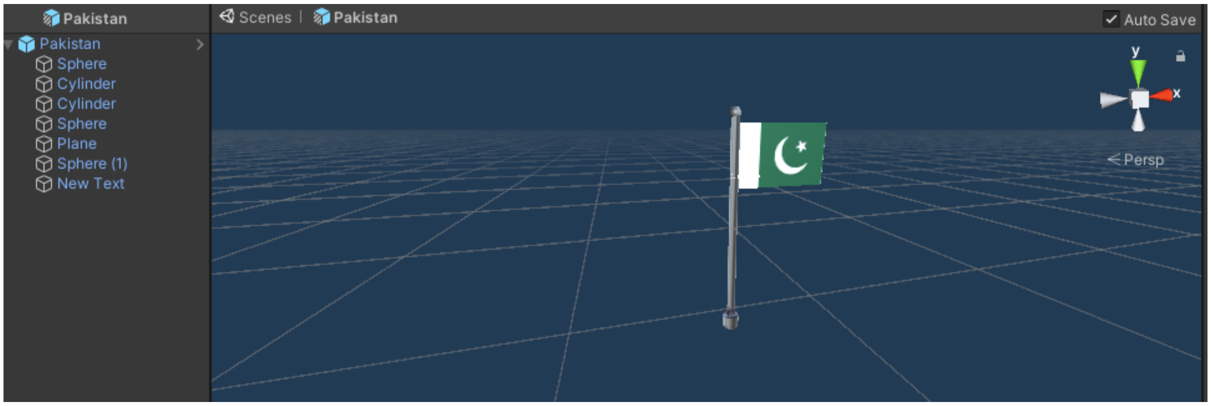
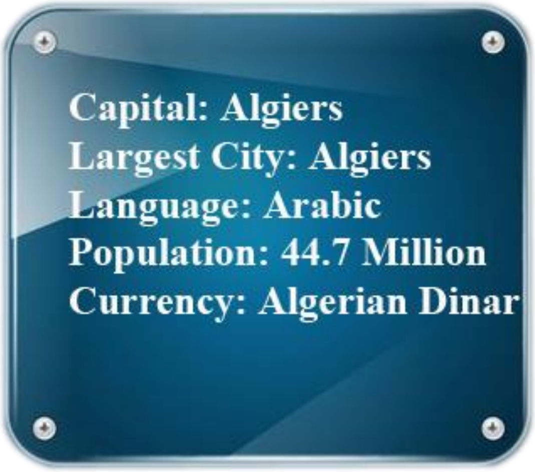


Figure 4.7: Design of 3D flag inside Unity



Capital: Algiers
Largest City: Algiers
Language: Arabic
Population: 44.7 Million
Currency: Algerian Dinar

Figure 4.8: 3D country information block

Same models were created for 30 different countries.

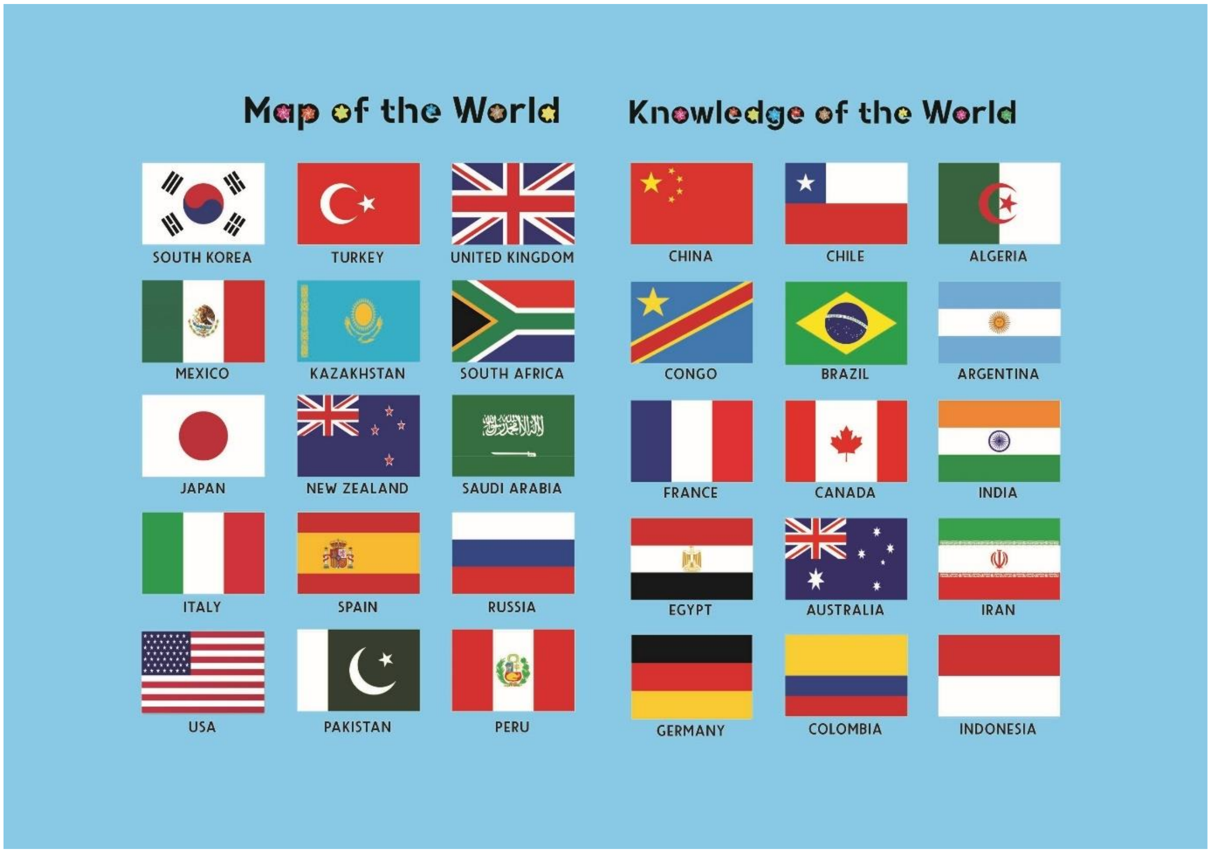


Figure 4.9: 30 Countries for which the above scenes are created.

4.2.2 Application Layout

1. **Main Menu Scene**
   a. AR-Globe
   b. Country Info
   c. Quiz
   d. About

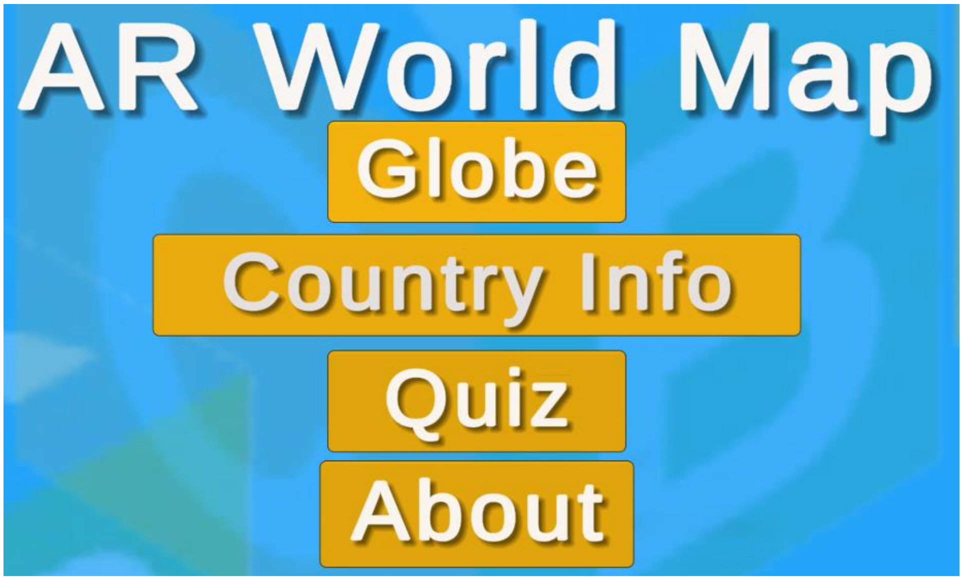


Figure 4.10: AR Application layout.

Functionalities:

1. Hover over buttons
2. Color changes on button pressed and released.
3. Background music

Globe Scene

Functionalities:

1. Touch zoom in and out (Lean Touch Script).
2. Touch rotation (Rotation object Script).
3. Close and back button (Scene Management script).

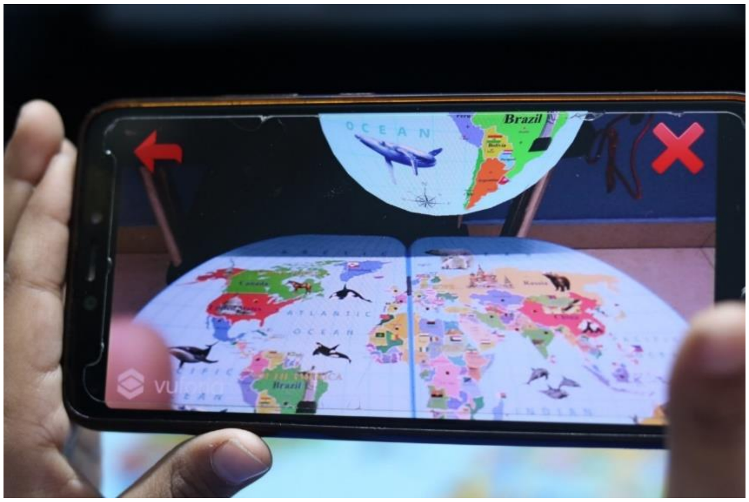
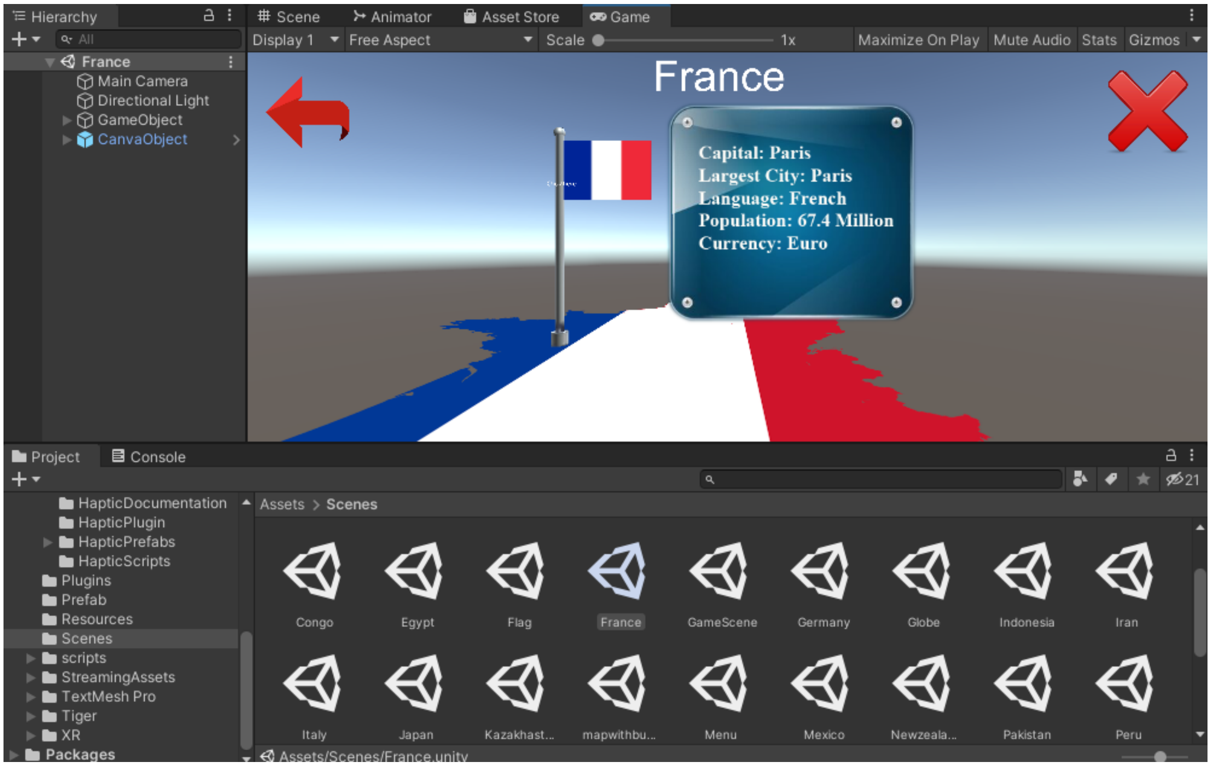


Figure 4.11: Working of AR application.

Country Info Scene (30 Scenes for 30 Countries)

Functionalities:

1. Clothing effect for flag (Cloth component, Capsule collider).
2. Wind direction using Unity Physics Engine.
3. Back and close buttons (Scene management script).
4. Same for 30 scenes.



Capital: Paris
Largest City: Paris
Language: French
Population: 67.4 Million
Currency: Euro

Assets/Scenes/France.unity

Figure 4.12: Augmented Reality scene

Quiz Scene

Functionalities:

1. Prepare your own quiz using inspector panel.
2. You can input image type, sentence type, multiple choice, video type question for students to test their knowledge.

Working of AR- Application

After designing 3D models target images are imported using Vuforia database having a good augmentable rating. AR camera is placed in the Unity scene which enables mobile devices to search for target image. 3D models are placed onto the target object and when the mode is switch to game scene using play button AR camera is switched on and when it detects the target image it displays 3D models in the real world which can be viewed using mobile devices.

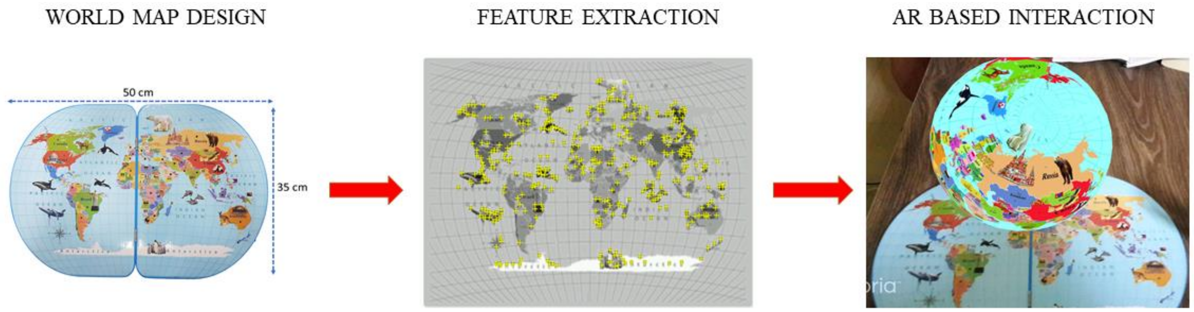


Figure 4.13: Complete flow of AR App

Integrating Haptics on World Map

Haptic technology, also known as kinaesthetic communication or 3D touch, refers to any technology that can create an experience of touch by applying forces, vibrations, or motions to the user. These technologies can be used to create virtual objects in a computer simulation, to control virtual objects, and to enhance remote control of machines and devices (telerobotics).

Some country scenes were modified to include famous monuments and animals like Pyramid of Giza in Egypt and haptic materials were attached that gives exact feel of the monument material. For this purpose we tested our country scenes using two haptic devices.

1. Novint Falcon
2. 3D Systems Touch Omni

Note: For haptics we have created a desktop application of the country scene as the hardware (Haptic Devices) is not supported for mobile whereas For Augmented Reality we have developed mobile application as Vuforia AR plugin only creates mobile application.

Novint Falcon Integration

To integrate haptics in the the country scene using Novint Falcon we purchased a plugin asset name “Touchable Universe” from Unity Asset Store. This asset allow us to add haptic functionalities to our game object using two scripts Haptic Camera (It allows haptic cursor to be controlled by the Novint Falcon) and Haptic Mesh (It apply haptic material to be set on the game object).

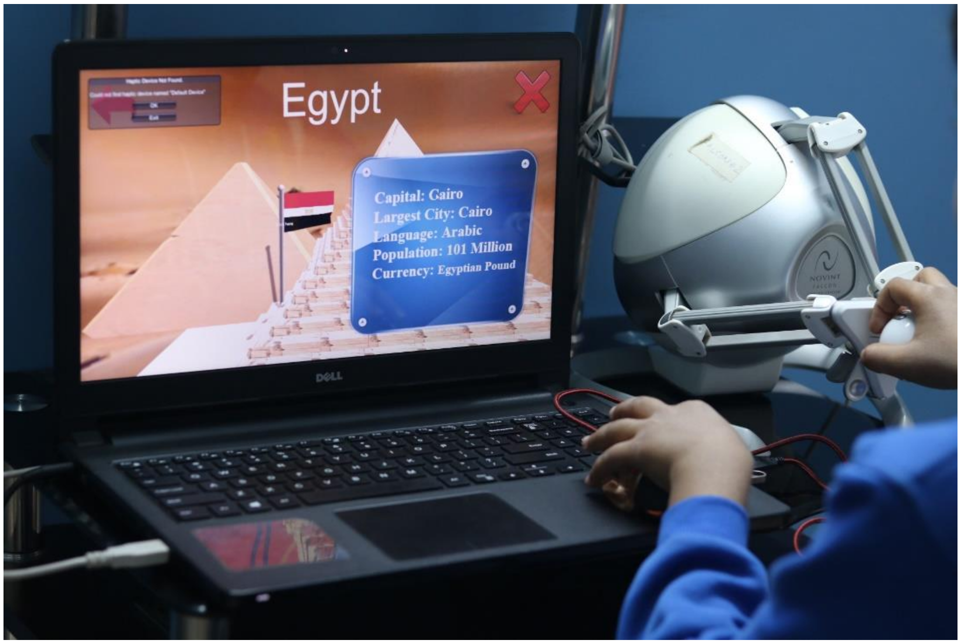


Figure 4.14: 3D Virtual world integrated with haptics using Novint Falcon.

Integration With 3D Systems Touch Omni

The integration of Omni device with country scene was achieved using Open Haptics asset for 3dsystem device available in Unity Asset Store for free. To apply haptic properties to a game object 4 properties of material are required.

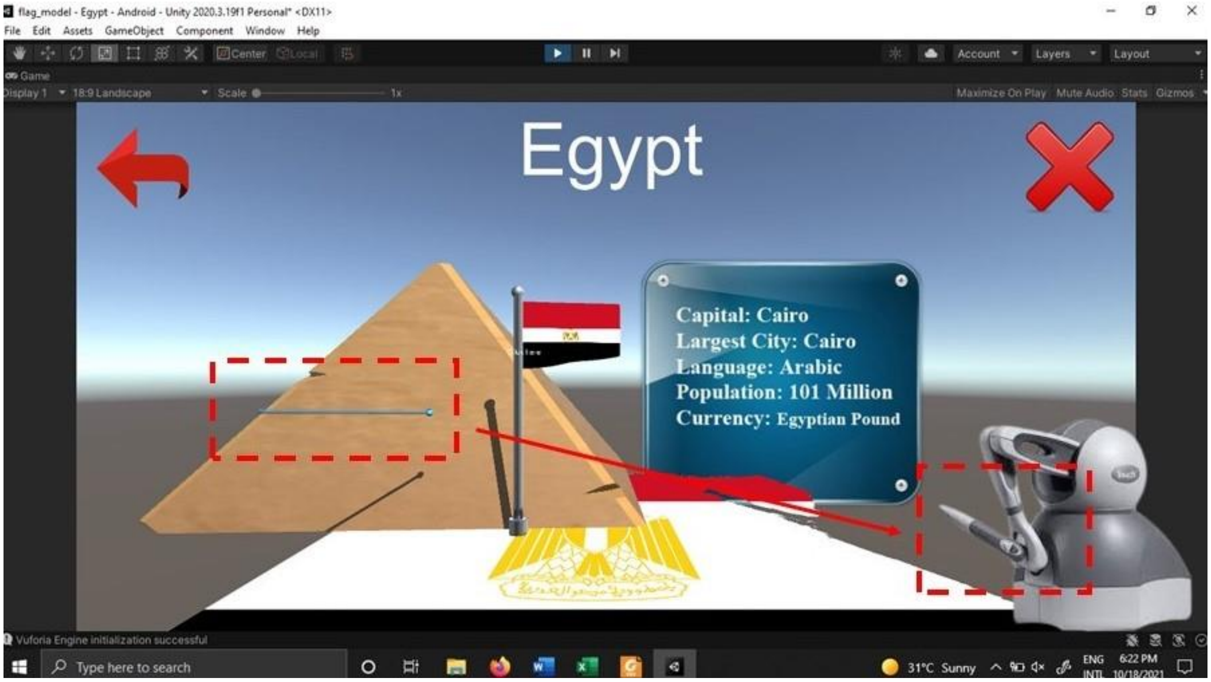
(Range between 0 and 1) definition from Open Haptics Documentation.

1. Stiffness
2. Damping
3. Static Friction
4. Dynamic Friction

Haptic grabber was designed which will act like a cursor when controlled from a Omni device. Haptic grabber script is attached in the scene. Haptic surface is made and set using above parameters for any game object in our case Pyramid of Giza. For complicated design like the Pyramid in the Egypt scene was designed using Blender.

Functionalities:

1. Haptic feedback when cursor is moved and placed over Pyramid.
2. Scene can be rotated using Cinemachine and Rotation object Script when grabber is rotated so that user can feel full 3D view.
3. Cloth effect on flag using Unity Physics Engine


flag_model - Egypt - Android - Unity 2020.3.19f1 Personal* <DX11>
File Edit Assets GameObject Component Window Help
Center Local
Account Layers Layout
Game
Display 1 18:9 Landscape Scale 1x
Maximize On Play Mute Audio Stats Gizmos
Egypt

Capital: Cairo
Largest City: Cairo
Language: Arabic
Population: 101 Million
Currency: Egyptian Pound

Vuforia Engine initialization successful
Type here to search
31°C Sunny
ENG INTL 6:22 PM 10/18/2021

Figure 4.15: 3D Virtual world using Touch Haptics

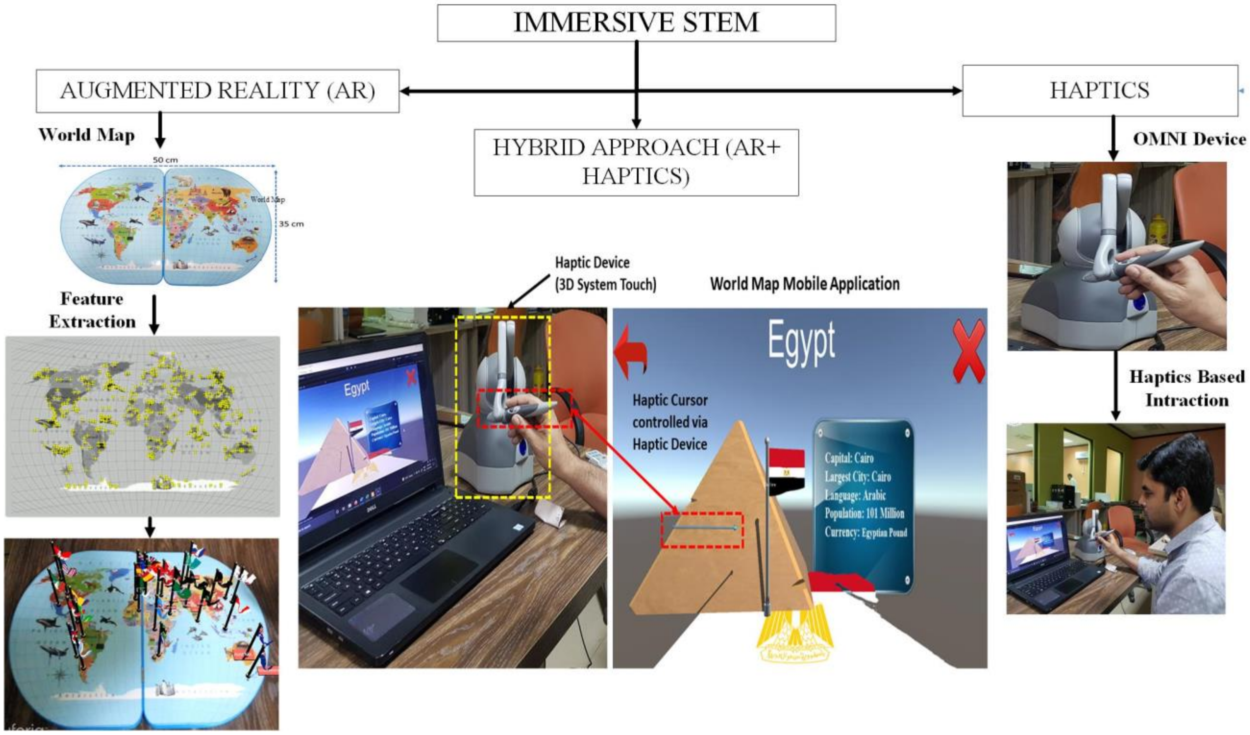


Figure 4.16: Immersive STEM using AR and Haptics

CHAPTER 5: USER STUDY

5.1 Details of User Study

(Augmented Reality, Haptics and Traditional Learning Methods)

Experiment Details

15 girls aged 7 to 12 and 12 boys aged 7 to 12. (Grade 1 to 5) participated in the experiment.

Theme: World Map as a geography topic was chosen for this activity as it is a generic topic and easy to understand and taught at different levels in primary schools.

Three types of learning methods.

Traditional: Students were taught world map idea using traditional learning methods.

Augmented Reality: Students Were given tablets and smart phone and were taught using Augmented Reality.

Haptics: Students were provided with Novint Falcon device by which they can feel and learn monumental landmarks of country like Pyramid of Giza in Egypt.

Duration of Activity: 20 minutes each with a rest time of 10 minutes between activities.

Task: 20 minutes of teaching through 3 methods and after teaching students should complete 3 quizzes and fill PANAS Questionnaire and Nasa Task Load Index.

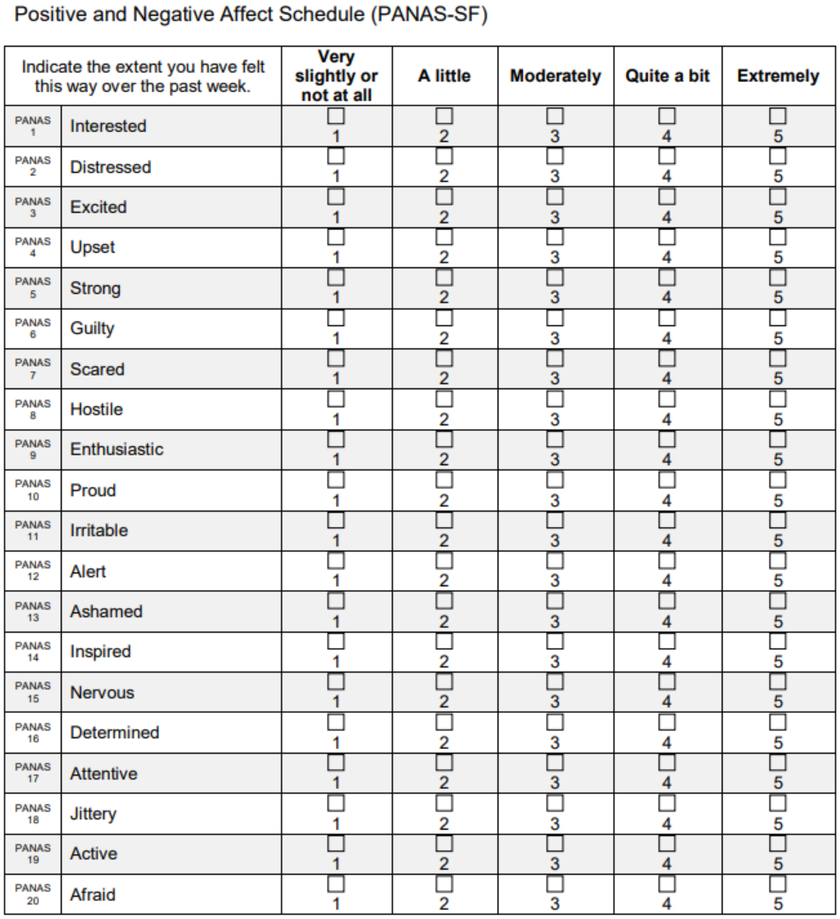
Note: This is not a memory quiz therefore all resources like Wooden World Map, AR Mobile App and haptic devices were provided during the quiz. Time to complete the quiz was 10 minutes with 10 multiple choice questions of basic level.

Two types of methods were used to evaluate the most efficient method of learning (Traditional, Augmented Reality and Haptics). The task load index (NASA Task Load Index) and emotions of students using (PANAS Questionnaire) were measured.

Hypothesis: Haptics and Augmented Reality learning are more immersive way than the traditional methods of learning. Therefore, the task load index for Haptics teaching and AR should be minimum while that of traditional teaching should be maximum. And the positive emotion score for haptics and AR should be higher than the traditional methods while negative emotion should be minimum in case of Haptics and AR as compared to traditional teachings.

5.2 PANAS Questionnaire

The Positive and Negative Affect Schedule (PANAS) is a widely utilized scale for assessing emotions or moods (Watson, Clark, & Tellegen, 1988). This concise 20-item scale comprises 10 components that gauge positive affect (such as excitement and inspiration) and 10 components that evaluate negative affect (for instance, distress and fear). Participants rate each item on a five-point Likert Scale, with 1 signifying "Very Slightly or Not at all" and 5 denoting "Extremely," to determine the degree to which they have experienced the emotion within a specified period. The PANAS was designed to measure affect in diverse contexts, including the present moment, the past day, week, or year, as well as on average. Consequently, the scale can be employed to assess current emotions, trait or dispositional affect, emotional fluctuations over time, or emotional responses to particular events.



Positive and Negative Affect Schedule (PANAS-SF)

Indicate the extent you have felt this way over the past week.		Very slightly or not at all	A little	Moderately	Quite a bit	Extremely
PANAS 1	Interested	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 2	Distressed	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 3	Excited	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 4	Upset	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 5	Strong	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 6	Guilty	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 7	Scared	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 8	Hostile	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 9	Enthusiastic	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 10	Proud	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 11	Irritable	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 12	Alert	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 13	Ashamed	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 14	Inspired	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 15	Nervous	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 16	Determined	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 17	Attentive	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 18	Jittery	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 19	Active	□ 1	□ 2	□ 3	□ 4	□ 5
PANAS 20	Afraid	□ 1	□ 2	□ 3	□ 4	□ 5

Figure 5.1: PANAS questionnaire

5.3 NASA Task Load Index

The NASA-TLX is a tool designed to assess subjective workload, enabling users to evaluate the workloads experienced by operators interacting with diverse human-machine systems. Over time, the NASA-TLX has emerged as the premier benchmark for gauging subjective workloads in an extensive array of applications.

By incorporating a multi-dimensional rating procedure, NASA TLX derives an overall workload score based on a weighted average of ratings on six subscales:

1. Mental Demand
2. Physical Demand
3. Temporal Demand
4. Performance
5. Effort
6. Frustration

Three dimensions deal with the demands placed on the subject (Mental, Physical, and Temporal Demands), while the other deals with the subject's involvement with the work (Effort, Frustration and Performance).

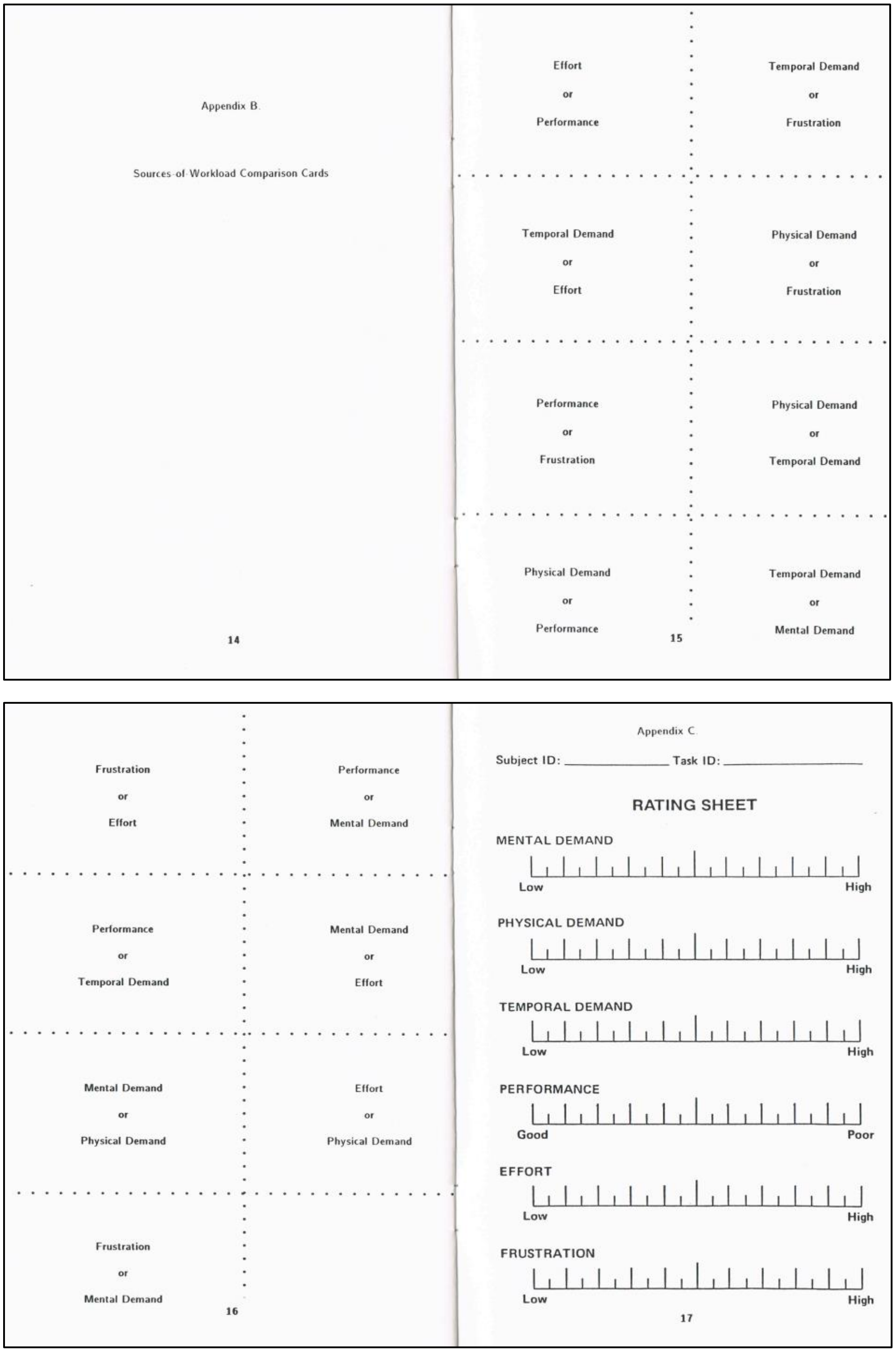
From the perspective of the raters, the degree to which each of the six criteria contributes to the burden of the specific activity to be evaluated is established by their responses to pairwise comparisons among the six factors.

Sources of Load

The NASA TLX assessment consists of two components: weights and ratings. Initially, raters evaluate the contribution (weight) of each factor to the workload of a specific task. This involves 15 potential pair-wise comparisons across the six scales. Participants indicate the factor within each pair that had the greatest impact on the task's workload by encircling it. The frequency with which each factor is selected is tallied, with possible totals ranging from 0 (not relevant) to 5 (extremely relevant, more crucial than any other factor).

Magnitude of Load

The second criterion is to acquire numerical ratings for each scale that indicate the significance of that component in a particular task.



Appendix B.

Sources-of-Workload Comparison Cards

14

Effort

or

Performance

Temporal Demand

or

Frustration

Temporal Demand

or

Effort

Physical Demand

or

Frustration

Performance

or

Frustration

Physical Demand

or

Temporal Demand

Physical Demand

or

Performance

Temporal Demand

or

Mental Demand

15

Frustration

or

Effort

Performance

or

Mental Demand

Performance

or

Temporal Demand

Mental Demand

or

Effort

Mental Demand

or

Physical Demand

Effort

or

Physical Demand

Frustration

or

Mental Demand

16

Appendix C.

Subject ID: ______________ Task ID: __________________

RATING SHEET

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

17

Figure 5.2: NASA Task Load Index

CHAPTER 6: Results and Conclusion

6.1 Results

The result of PANAS Questionnaire is as follow.

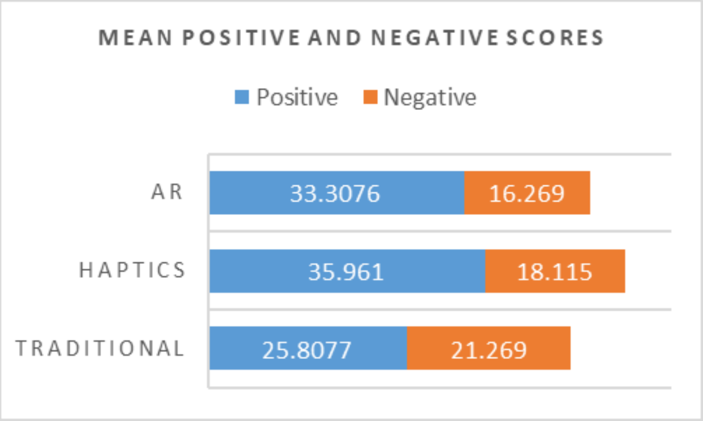


Figure 6.1. Mean positive and negative scores for Traditional, Haptics and AR teaching methodologies.

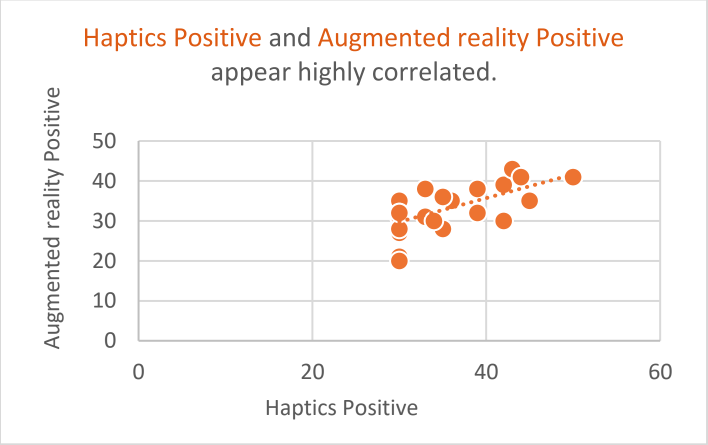


Figure 6.2. Graph displaying a very strong correlation between haptic and augmented reality instructional strategies.

NASA-TLX [39] serves as a tool for gauging subjective workload. It allows users to examine the workloads of operators interacting with diverse human-machine systems. NASA TLX sets the standard for evaluating subjective workload in numerous applications. Utilizing a multi-dimensional rating process, the NASA TLX calculates a comprehensive workload score based on the weighted average of ratings across six subscales:

- Mental Demand
- Physical Demand
- Temporal Demand
- Performance
- Effort
- Frustration

The results for NASA Task Load Index for three types of learning methods are as follows: Traditional learning has the greatest NASA-TLX group mean score of 83 which suggest a higher task load as compared to Augmented Reality and Haptics-based learning as shown in Figure 6.3. It can be seen in Figure 6.4 that the major contributing factor in traditional learning is mental and frustration workload. It can be due to the fact that in traditional learning mental effort is required the most. The major workload in case of Augmented Reality based learning is temporal workload. It is because in Augmented Reality based learning some amount of time is required to perform and experience Augmented Reality via mobile or tablet. Some type of camera calibration is also required due to which user experiences temporal workload (reference) as shown in Figure 6.5. Certain hardware like (3D System Touch Omni) haptic device is required to experience haptic force feedback. The user has to apply certain efforts in order to perform the task correctly therefore in case of Haptics based learning effort contribute majorly to the Task Load Index as shown in Figure 6.6. When it comes to traditional learning methods user faces a great amount of mental load and as the amount of time increases the user feels frustration while completing the task as shown in Figure 6.7.

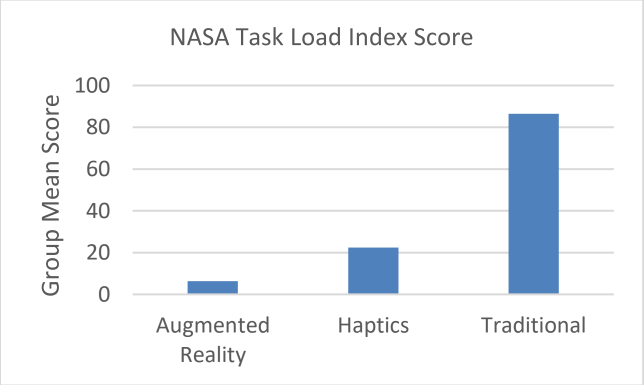


Figure 6.3. Task load index for Traditional teaching is highest among the other two methodologies.

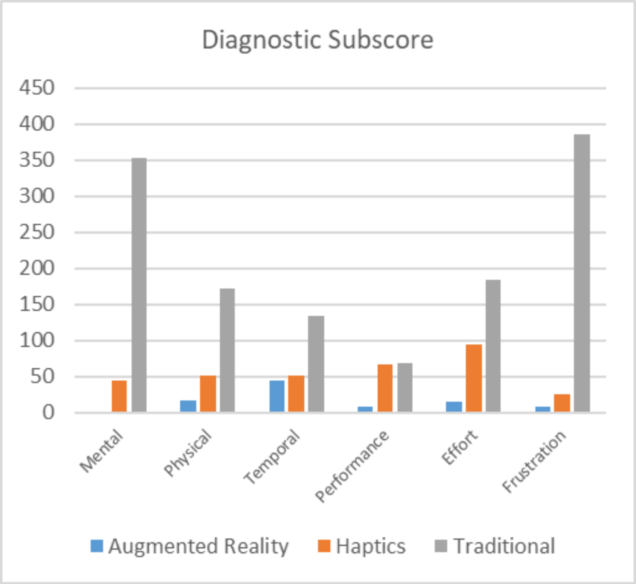


Figure 6.4. Individual factors contributing to task load for Traditional, AR and Haptics based learning methods.

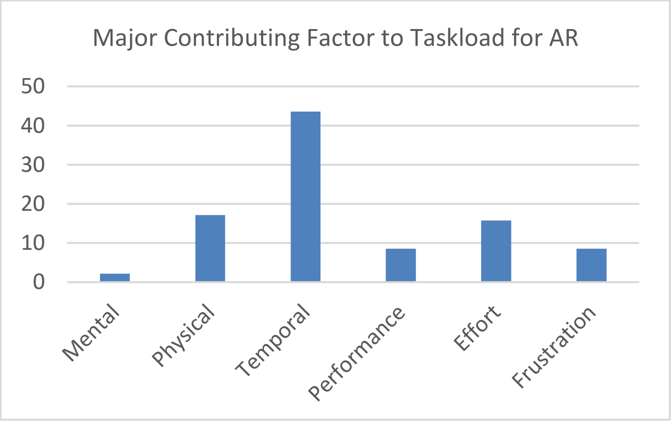


Figure 6.5. Temporal load can be seen to dominate and contribute to major factor of task load in AR learning method.

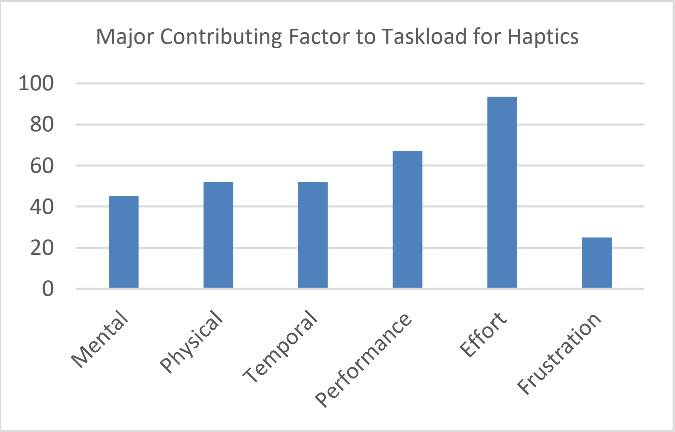


Figure 6.6. Effort to complete the task is the major contributing factor in Haptics based learning.

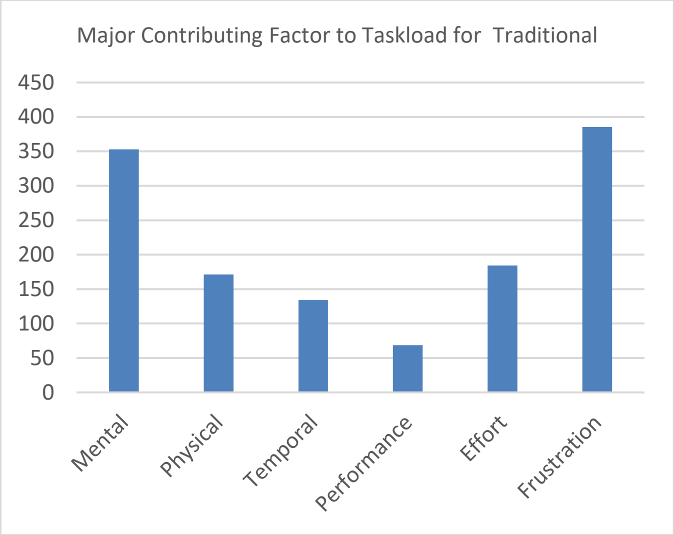


Figure 6.7. Mental workload and frustration are the key contributors for task load in Traditional learning methods.

6.2 Conclusion

This This thesis presents the design and creation of an innovative system aimed at detecting student engagement or disengagement during online learning and STEM education. The system utilizes computer vision-based models to analyze facial emotions, body posture, and head rotation, which helps determine if a learner is engaged or disengaged in the learning process. To enhance engagement levels, three distinct learning approaches have been employed: (i) traditional/conventional methods, (ii) haptic technology, and (iii) augmented reality.

A World Map incorporating Augmented Reality and Haptic technology was developed to examine the effectiveness of the latter two learning methods, AR and Haptics. To evaluate these approaches, a user study was conducted with participants exposed to the three different types of learning techniques. The study aimed to identify the most successful and engaging instructional strategies by analyzing the participants' experiences.

The NASA-TLX and PANAS Questionnaire evaluations were used to gather insights into the participants' perceived workload and emotional responses to each learning method. By comparing their responses, the study aimed to determine which approach resulted in the highest levels of engagement and satisfaction among the learners.

The findings of the study revealed that students exhibited higher engagement levels when using AR and Haptic technology compared to traditional learning methods. These results suggest that incorporating modern technology, such as Augmented Reality and Haptic feedback, can significantly improve the learning experience, leading to better student engagement and overall educational outcomes in online and STEM education environments.

REFERENCES

1. Kennedy, T.J. and M.R.J.S.E.I. Odell, Engaging students in STEM education. 2014. 25(3): p. 246-258.
2. Moore, T.J., K.A.J.J.o.S.E.I. Smith, and Research, Advancing the state of the art of STEM integration. 2014. 15(1): p. 5.
3. Rothkrantz, L. Dropout rates of regular courses and MOOCs. in International Conference on Computer Supported Education. 2016. Springer.
4. Sanfilippo, F., et al., A Perspective Review on Integrating VR/AR with Haptics into STEM Education for Multi-Sensory Learning. 2022. 11(2): p. 41.
5. Dewan, M., M. Murshed, and F.J.S.L.E. Lin, Engagement detection in online learning: a review. 2019. 6(1): p. 1-20.
6. Altuwairqi, K., et al., Student behavior analysis to measure engagement levels in online learning environments. 2021. 15(7): p. 1387-1395.
7. Vanneste, P., et al., Computer vision and human behaviour, emotion and cognition detection: a use case on student engagement. 2021. 9(3): p. 287.
8. Whitehill, J., et al., The faces of engagement: Automatic recognition of student engagementfrom facial expressions. Transactions on Affective Computing, 2014. 5(1): p. 86-98.
9. Ahuja, K., et al., EduSense: Practical classroom sensing at Scale. 2019. 3(3): p. 1-26.
10. Monkaresi, H., et al., Automated detection of engagement using video-based estimation of facial expressions and heart rate. Transactions on Affective Computing, 2016. 8(1): p. 15-28.
11. Yu, M., et al. Behavior detection and analysis for learning process in classroom environment. in 2017 IEEE Frontiers in Education Conference (FIE). 2017. IEEE.
12. Zaletelj, J., A.J.E.j.o.i. Košir, and v. processing, Predicting students’ attention in the classroom from Kinect facial and body features. 2017. 2017(1): p. 1-12.
13. Dewan, M.A.A., et al. A deep learning approach to detecting engagement of online learners. in 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data

Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2018. IEEE.

14. Gupta, A., et al., Daisee: Towards user engagement recognition in the wild. 2016.

15. Sharma, P., et al., Student engagement detection using emotion analysis, eye tracking and head movement with machine learning. 2019.

16. Abedi, A. and S.S. Khan. Improving state-of-the-art in Detecting Student Engagement with Resnet and TCN Hybrid Network. in 2021 18th Conference on Robots and Vision (CRV). 2021. IEEE.

17. Liao, J., Y. Liang, and J.J.A.I. Pan, Deep facial spatiotemporal network for engagement prediction in online learning. 2021. 51(10): p. 6609-6621.

18. Mohamad Nezami, O., et al. Automatic recognition of student engagement using deep learning and facial expression. in Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2019. Springer.

19. Ninaus, M., et al., Increased emotional engagement in game-based learning–A machine learning approach on facial emotion detection data. 2019. 142: p. 103641.

20. Goldberg, P., et al., Attentive or not? Toward a machine learning approach to assessing students' visible engagement in classroom instruction. 2021. 33(1): p. 27-49.

21. Azuma, R.T.J.P.t. and v. environments, A survey of augmented reality. 1997. 6(4): p. 355-385.

22. Sunil, S. and S.S.K. Nair. An educational augmented reality app to facilitate learning experience. in 2017 International Conference on Computer and Applications (ICCA). 2017. IEEE.

23. Koul, M.H. and I. Shahdad, Towards an open source haptic kit to teach basic STEM concepts, in Proceedings of the Advances in Robotics. 2017. p. 1-6.

24. Lee, L.-K., et al. Using augmented reality to teach kindergarten students english vocabulary. in 2017 International symposium on educational technology (ISET). 2017. IEEE.

25. Sarosa, M., et al. Developing augmented reality based application for character education using unity with Vuforia SDK. in Journal of Physics: Conference Series. 2019. IOP Publishing.

26. Bortone, I., et al., Wearable haptics and immersive virtual reality rehabilitation training in children with neuromotor impairments. 2018. 26(7): p. 1469-1478.

27. Edwards, B.I., et al., Haptic virtual reality and immersive learning for enhanced organic chemistry instruction. 2019. 23(4): p. 363-373.

28. Hamza-Lup, F.J.a.p.a., Kinesthetic Learning--Haptic User Interfaces for Gyroscopic Precession Simulation. 2019.

29. Sırakaya, M. and D.J.I.L.E. Alsancak Sırakaya, Augmented reality in STEM education: A systematic review. 2020: p. 1-14.

30. Seifi, H., et al., How do novice hapticians design? A case study in creating haptic learning environments. 2020. 13(4): p. 791-805.

31. Pila, S., et al., Preschoolers' STEM Learning on a Haptic Enabled Tablet. 2020. 4(4): p. 87.

32. Videnovik, M., et al., Increasing quality of learning experience using augmented reality educational games. 2020. 79(33): p. 23861-23885.

33. Petrov, P.D. and T.V.J.I. Atanasova, The Effect of augmented reality on students' learning performance in stem education. 2020. 11(4): p. 209.

34. Newmann, F.M., Student engagement and achievement in American secondary schools. 1992: ERIC.

35. Goodfellow, I.J., et al. Challenges in representation learning: A report on three machine learning contests. in International conference on neural information processing. 2013. Springer.

36. Ketkar, N., Introduction to keras, in Deep learning with Python. 2017, Springer. p. 97-111.

37. Yamaguchi, K., et al., End-to-end learning potentials for structured attribute prediction. 2017.

38. Wang, Y., Y. Li, and J. Zheng. A camera calibration technique based on OpenCV. in The 3rd International Conference on Information Sciences and Interaction Sciences. 2010. IEEE.

39. Hart, S.G. and L.E. Staveland, Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, in Advances in psychology. 1988, Elsevier. p. 139-183.