STUDENT ENGAGEMENT AWARENESS DASHBOARD IN ASYNCHRONOUS E-LEARNING ENVIRONMENT

Abdalganiy Kebede

STUDENT ENGAGEMENT AWARENESS DASHBOARD IN ASYNCHRONOUS E-LEARNING ENVIRONMENT

Thesis submitted to

Indian Institute of Technology, Guwahati

for the award of the degree

DOCTOR OF PHILOSOPHY

by

Abdalganiy Kebede

under the supervision of

Dr. Samit Bhattacharya

Department of Computer Science and Engineering

Indian Institute of Technology, Guwahati

Guwahati, Assam

June, 2021

DECLARATION

I certify that

a. the work presented in this thesis is original and has been done by me under the guidance of my supervisor.
b. the work has not been submitted to any other Institute for any degree or diploma.
c. I have followed the guidelines provided by the Institute in preparing the thesis.
d. I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
e. whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Abdalganiy Kebede

CERTIFICATE

This is to certify that the thesis entitled Student Engagement Awareness Dashboard in Asynchronous E-Learning Environment, submitted by Abdalganiy Kebede to Indian Institute of Technology, Guwahati, is a record of bona fide research work under my supervision and is worthy of consideration for the award of the degree of Doctor of Philosophy of the Institute.

____

Supervisor

Dr. Samit Bhattacharya
Dept. of Computer Science and Engineering
Indian Institute of Technology, Guwahati
Guwahati 781039, INDIA

Dedicated to my family

Acknowledgements

First and foremost, I would like to express my heartfelt appreciation to my supervisor, Dr. Samit Bhattacharya for his guidance, support, and inspiration throughout the progress of my research work. His tireless effort, constant direction shall always be deep in my heart.

I would like to thank all the members of my DC committee for their valuable suggestions during my PhD seminars.

It is my pleasure to extend and express the sacrifices of my dad, my mom, my siblings and my wife –Fetiya Musa, who took all the responsibilities of my family, and helped me to carry out this work smoothly. My son Ejaz and little Daughter, Sumaya, have always encouraged me to focus on my work with their beautiful sound of “Daddy, when will you come home?” that always inspire me to complete this thesis.

I am fortunate to have many good friends: Muhammed, Cherenet, Nayabrasul, Tamam, Abdulwahid, Abdulqadir, Tefera, Tufa, Dr. Abdisa, Dr. Haji, Dr. Harison, Achenif, Mandefro, Mebratu, Masresha, Shiferaw, Tesfay, Dr. Moera, Sibu who have helped and inspired me in many ways. I want to thank Dr. Nasim, Dr. Hakimnias and Prof. Rafi, Sir for inspiring me in all the ways.

Finally, my sincere gratitude goes to all the faculty members, staffs of the Computer Science and Engineering Department and my lab mates– Dr. Subrata, Ujjwal, Shakeel, Nilotpal, and Dr. Hema.

Lastly, but yet very importantly, I would like to acknowledge the continuous support of Dr. Temesgen Kebede Wakjira from the very beginning of my Ph.D. life.

Abstract

The most important activity in life is to acquire and be equipped with knowledge. It is generated and acquainted through learning and trainings. Often there are adult learners who face competing demands, including family and work responsibilities. These competing demands hinder them from attending their lesson from a typical physical classroom. E-learning technology is unique and represents a new era of distance learning, which is categorized as a fourth generation distance learning technology. Asynchronous e-learning has a disadvantage as students feel isolated. The feeling of the isolation leads to drop-out. In the e-learning environment, where the teacher is not physically present, monitoring a student for interest or engagement is a challenge. To solve these problems, we have accomplished the following objectives: (a) We set up an experiment involving 12 participants to collect data from behavioural, collaboration and emotional features for detecting student engagement status in an asynchronous e-learning environment. We identified the most important features affecting student engagement levels out of the total of 13 features from three factors: behavioural, collaboration and emotional factors using Pearson correlation analysis and Pratt's index. (b) We built student engagement prediction model from three factors: behavioural, collaboration and emotional factors across micro level time scale such as 5 minutes. We applied the features that correlated significantly with the levels of engagement from three factors: behavioural, collaboration and emotional factors to build the student engagement prediction model using non-linear regression techniques. We also validated the student engagement prediction model through empirical study and found high accuracy. (c) We built a student engagement visualization dashboard that visualizes the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner. The dashboard is based on a student engagement prediction model, which we also developed. We also performed the validation of these proposed visualizers in controlled experiment. The validation indicated that the users' satisfaction of the visualizers was high. This helps a teacher to gain insight about the engagement levels of all students at a glance. This will also allow the teacher to take immediate action.

Keywords: student engagement, engagement awareness, engagement dashboard, asynchronous e-learning, behavioural states, collaborative states, emotional states, Non-linear regression

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Presently, the most important activity in life is to acquire and be equipped with knowledge. Knowledge is generated and acquainted through learning and trainings. Learning and trainings become massive business opportunities globally. Success of organizations and institutions depend on how competent and talented their employees are who are equipped with the latest information and advanced technical and practical knowledge. Competitive business environment requires employees who are highly skilled, well educated, and can competitively perform in the global work-force[Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.. Often there are adult learners who face competing demands, including family and work responsibilities. These competing demands can hinder them from attending their lesson from a typical classroom. Students are unable to decide whether to study online or face-to-face, [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.. These adult learners have a need of lifelong learning. That need comes with the desire for time and cost savings and with the need of remote learning [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218..

Thus, they cannot use the traditional classroom. The concept of traditional learning does not fit well with the new world of lifelong learning, in which the roles of instructor, students, and curriculum are changing. Teaching and learning are no longer restricted within traditional classrooms [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.. The outcry for higher learning, especially in the developing economies, is getting louder each day even while the cost of providing it is on the increase. The use of distance learning under these circumstances seems more an imperative than an option [Harry, 2002] Harry, K. ed., 2002. Higher education through open and distance learning. Routledge. They can choose to use distance learning which is a cost-effective learning infrastructure that enables anytime, anywhere, self-paced, and interactive learning [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.. Moreover, they choose this mode of learning, because it provides flexibility, enabling them to balance external commitments with their studies [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.. Distance learning is discussed in the next section.

1.1. Distance Learning

Distance learning is ‘a method of studying in which lectures are instructed by correspondence, without the student needing to attend a school or college’ [Harry, 2002] Harry, K. ed., 2002. Higher education through open and distance learning. Routledge. Distance learning is seen as one way of increasing the access to learning. In most cases, print, sometimes supported by broadcasting or by the use of cassettes, dictates open and distance learning [Harry, 2002] Harry, K. ed., 2002. Higher education through open and distance learning. Routledge. Distance learning differs by more access and flexibility than traditional on-campus courses. Distance learning platforms offer a second chance to those bound by time or disabilities, workers who want to upgrading their skill, and older people giving them new chances [Ruhe and Zumbo, 2008] Ruhe, V. and Zumbo, B.D., 2008. Evaluation in distance education and e-learning: The unfolding model. Guilford Press..

With distance learning, "students study at the time and place of their choice (home, work or learning centre) and without face-to-face contact with a teacher". Distance learning delivery methods started with print distance delivery. It also includes video teleconferencing, and CD-ROM, and can serve either on- and off-campus learners [Ruhe and Zumbo, 2008] Ruhe, V. and Zumbo, B.D., 2008. Evaluation in distance education and e-learning: The unfolding model. Guilford Press..

Indeed, institutions of open and distance learning are shifting to online learning applying technologies for helping them to reach new clients [Harry, 2002] Harry, K. ed., 2002. Higher education through open and distance learning. Routledge.. We discussed these technologies in the next section.

1.2. Learning Technologies

The word technology was derived from the Greek word "**tekhnologiá",** meaning a systematic treatment of an art or craft [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice.

Taylor & Francis. This new importance on systematic treatment and an implied attachment to science and especially the scientific method, has inspired the formal field of learning technology to embrace a modernistic, and scientific, view of its activities. Technology directly influences the presentation, the communication, the budget, and the design of the learning results. The definition of learning technology, in this thesis, takes a more widespread view of technology with an emphasis on tools as opposed to techniques [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis.

Learning technologies are defined as: 'those tools used in formal learning practice to broadcast, illustrate, communicate, or immerse learners and teachers in activities purposively designed to induce learning'[Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis. Learning technologies are on the rise in terms of number and variety. CMSs, such as Blackboard, Desire2Learn, and WebCT, have become common at institution of higher learning, and distance learning organizations. Learning management systems can be used as ecommerce sites, online learning and digital libraries. The web, changed from a text-only medium to multimedia, interactive media, and learning objects and into e-learning courses. Learning objects are applets, animation, simulation, maps, and games. Web content management can act like libraries and digital repositories to be used by learners at their residence. Learners are able to access the web using interactive television(iTV), without the need for additional computer. "nomadicity," is "multiple devices of mobile computing and communications", including "multi-function cell phones, voice-over IP (VoIP), peer-to-peer file sharing, digital video capture and wireless data cards” [Ruhe and Zumbo, 2008] Ruhe, V. and Zumbo, B.D., 2008. Evaluation in distance education and e-learning: The unfolding model. Guilford Press.. “Teens use the Internet to multi-task—instant message, reserve books at the library, order online, and participate in an online quiz or games” [Ruhe and Zumbo, 2008] Ruhe, V. and Zumbo, B.D., 2008. Evaluation in distance education and e-learning: The unfolding model. Guilford Press..

Other technologies have grown up alongside video which, by their web-enabled nature make no difference regarding the geographical location of learners. They are: Massively Online Open Courses (MOOC), "Flipped" classroom, Proliferation of video and video-on-demand: YouTube and other "Tubes", Web Real-Time Communications (WebRTC) and HTML5, Mobile devices and the "Bring Your Own Device" (BYOD) phenomenon. A teaching style of these technologies is expressed in the motto, "Learning: Any Time, Any Where." Learning technology is no longer all about overcoming distance obstacles. Now it's Collaborative Learning on distance learning technologies [Economides, 2013] Economides, T., 2013, October. The state of the art in educational technology. In 2013 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 285-287). IEEE. We introduced this concept in the subsequent section.

1.3. Distance Learning Technology

The technologies of distance learning were categorized into generations based on the technological tools that support each generation. These ‘generational’ grouping systems help us to grasp and define the various components chronologically.

First Generation

The technology most related with this generation is the printed textbook and associated course guide. It should not be anticipated that these print course materials are just text or reference books that are naturally found in academic libraries. Rather, the material is wisely designed and created by a carefully identified course team made up of specialized, skilful professionals. The course guide reflects a dialogue approach between the absent teacher and the independent learner. A crucial feature of first-generation technology is the full freedom and independence for learners. Learners can commence their studies at any time of the year. They are not forced to follow an institutionally demarcated timeline. Individual learners may finish learning activities at a speed they define. Learners studying under first-generation distance learning systems are isolated from the guidance of an instructor. Initially, such interaction was continued in asynchronous arrangement through mail [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis.. The feeling of the isolation leads to drop-out [Leo et al., 2009] Leo, T., Manganello, F., Pennacchietti, M., Pistoia, A. and Chen, N.S., 2009, July. Online synchronous instruction: Challenges and solutions. In 2009 Ninth IEEE International Conference on Advanced Learning Technologies (pp. 489-491).IEEE.

Second Generation

The second generation is associated with the newer technologies of broadcast media. This generation also does not put restrictions on time or place of study.

Big and costly media productions (telecourses) were produced. These telecourses allowed learners to visit the laboratory and the work place with the audio and/or video images. Direct communication between learners and instructors was restricted to telephone and mail. ‘Interactive, computer-assisted courses ‘were delivered to learners using networked or standalone computers with courseware carried on CD-ROM or DVD disks. To date, these efforts have been unsuccessful.

Third Generation

The third generation is known by audio, video, and computer mediated conferencing that allowed both asynchronous and synchronous human interaction.

Fourth Generation

A fourth generation has association with the Internet. It encompasses the first three main features of the Net: data retrieval of huge amounts of content; the interactive capacity of computer mediated communications (CMC); and the processing power of locally dispersed processing. These are new tools, with the ability of joining CMC and Web resources (through products such as WebCT, and Blackboard).

Fifth Generation

This fifth generation represents an integrated system of web-based administration of student support services. It involves the use of teacher and learner agents that integrate various types of intelligence. That will result in productive searching and navigation. The fifth generation includes artificial intelligence to the Web for building semantic meaning into the Web, to allow processing by both humans and nonhuman autonomous agents. The technology of the Web is associated with fourth- and fifth-generation distance learning systems.

The fourth generation distance technology is the most popular [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis. We are going to discuss it in detail next.

1.4. E-learning

E-learning technology is unique and represents a new era of distance learning, which is categorized as a fourth generation distance learning technology [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis. What has changed, compared to 1st, 2nd and 3rd generation is the ‘speed and power of communications and the expanded capacity to send, receive, and use information’ and the capacity to bridge time and space for educational purposes[Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis. E-learning can be defined as the use of electronic devices and technology for learning new information and skills [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. E-learning can be classified into three types based on how students can access the content of the course, timing of when students can access the content, and whether students interact with each other or not [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning

& Data analytics. IEEE Access, 6, pp.39117-39138. Figure 1.1 below shows the schematic diagram of the classification of e-learning.

Based on how students can access the content of the course, e-learning can be classified as online and offline [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. When students use internet to access the content, then it is online learning. In online learning, there is no face to face session for the course. When the students access the content offline through CDs and DVDs, then it is offline learning, which is similar with 2nd generation of distance learning. Based on the timing of when students can access the content, e-learning can be either synchronous or asynchronous. In synchronous mode, content is accessed in real-time where students attend classes at the same time via conference calls [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. Synchronous conferencing systems are time dependent systems and have played a supplementary role of socializing, brainstorming, or virtual office hours in online courses especially in higher education settings [Park & Bonk, 2007] Park & Bonk, 2007,Synchronous Learning Experiences: Distance and Residential Learners' Perspectives in a Blended Graduate Course, Journal of Interactive Online Learning. However, the synchronous mode has a limitation where the teacher may be required to manage the pace of his teaching by considering those students with slow connection [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. In asynchronous mode, content is accessed at any time by students. Students use forums to interact to each other or with the instructor [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. Asynchronous online communication occurred in a time-independent environment [Park & Bonk, 2007] Park & Bonk, 2007,Synchronous Learning Experiences: Distance and Residential Learners' Perspectives in a Blended Graduate Course, Journal of Interactive Online Learning. Other classification was based on whether students interact with each other or not as individual learning or group learning. In individual learning, there is no interaction between learners. Each student communicates only with the computer. In the group learning, students can interact with each other via forums [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138. It is also reported that interactions among peers and with instructors and group collaborations influence the online learning experience [Park & Bonk, 2007] Park & Bonk, 2007,Synchronous Learning Experiences: Distance and Residential Learners' Perspectives in a Blended Graduate Course, Journal of Interactive Online Learning. An online learning can be either synchronous or asynchronous. Asynchronous can be achieved through individual learning or group learning as shown on the Figure 1.1.

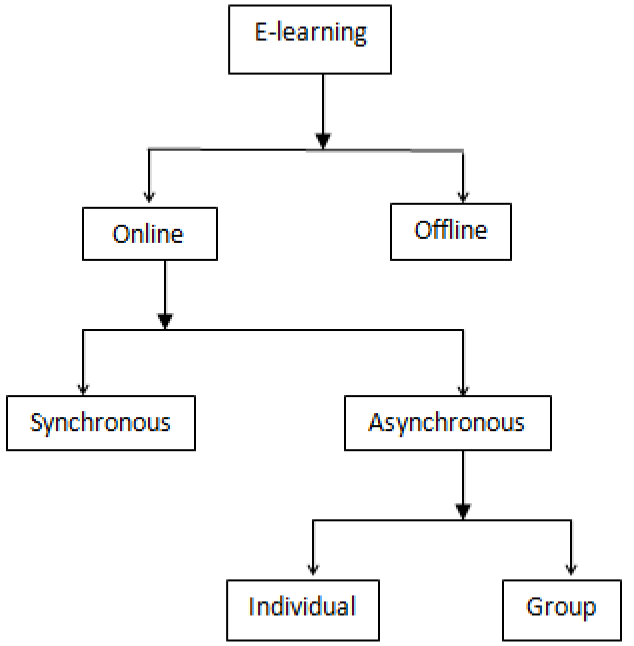


Figure 1.1: Taxonomy of e-learning adopted from *[Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138.*

Learning management systems (LMS) is an online program with a variety of features that support teaching and learning [Edmunds and Hartnett, 2014] Edmunds, B. and Hartnett, M., 2014.Using an online Learning Management System to personalise learning for primary students.Journal of Open, Flexible, and Distance Learning, 18(1), pp.11-29. These asynchronous online learning technologies give students the ability to communicate with their teachers and peers with a fulltime open access to content including course materials, lecture notes, tutorials, messages and recordings [Torun, 2013] Emel Dikbas Torun, 2013, Synchronous Interaction in Online Learning Environment with Adobe Connect Pro,Procedia-Social and Behavioural Science 106(2013)2492-2499,Science Direct The top LMSs that lead the market are the Blackboard and Moodle [Pireva et al., 2015] Pireva, K., Imran, A.S. and Dalipi, F., 2015, August.

User behaviour analysis on LMS and MOOC. In 2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (pp. 21-26). IEEE. In our study, we applied distance learning which is online, asynchronous, in both individual and group learning types using Moodle.

1.4.1. Advantages of E-Learning

The following section highlights several significant benefits of e-Learning with respect to other distance learning technologies particularly the third, second and first generation distance technologies.

- Time and Location Freedom

E-Learning removes the obstacles of time and distance by offering “just-in time” or “on- the-job” learning, and may reach to many such as incapacitated and job-sharing people.

- Self-paced Learning

E-Learning nurtures self-paced learning by organizing learner-centred activities. Each learner can learn at his or her time, especially compared to the third generation distance technologies.

- Collaborative Learning Environment

E-Learning allows communication between physically separated learners to form an online collaborative learning community. Learners can ask questions and share different ideas with each other more easily through online forums.

- Unlimited Use of Learning Materials

E-Learning permits unlimited access of learning materials. Information and knowledge are available to learners 24 hours a day [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

However, the asynchronous e-learning has the following limitations.

- Isolation Leads to Drop-out

Asynchronous e-learning allows learning independent of time, place and pace; however, it has a disadvantage as students feel isolated. The feeling of the isolation leads to drop-out [Leo et al., 2009] Leo, T., Manganello, F., Pennacchietti, M., Pistoia, A. and Chen, N.S., 2009, July. Online synchronous instruction: Challenges and solutions. In 2009 Ninth IEEE International Conference on Advanced Learning Technologies (pp. 489-491).IEEE.

- Students are Unmonitored

In the e-learning environment, where the teacher is not physically present, monitoring a student for interest or engagement is a challenge [Al-Alwani, 2016] Abdulkareem Al-Alwani, 2016,A Combined Approach to Improve Supervised E-Learning using Multi-Sensor Student Engagement Analysis,American Journal of Applied SciencesAccessed on 25 July 2019. In online environments, students are unmonitored [Sarder, 2014]

- Lack of Face to Face Faculty Student Interactions

One of the most common criticisms relates to the quality of educational outcomes of online education is due to the lack of face to face faculty student interactions [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.

- Faculties Lack the Tools

In online environments, faculties lack the tools to ensure complete viewing of the lecture content [Sarder, 2014] .

1.4.2. Information Technologies that are Used to Improve E-Learning Systems

How we learn and what we learn is continuously influenced by technology. The advancement of the Internet and information technologies makes e-Learning more widespread. In this section, we want to explain different information technologies that are used to improve e-Learning systems.

- Internet Technology

The Internet has been giving both information access and a speedy and cheap means of communication to the public, starting from 1994.The growth of e-Learning is constrained by both financial side and effective learning. The Internet is helping us regarding making the learning effective. It moves the notion of “anytime, anywhere” to a higher level as far as learning is concerned. That is done through allowing collaboration and discussion to occur through the Internet [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

- Knowledge Management

Knowledge Management (KM)comprises gathering, management and distributing knowledge, assisting individuals decide what knowledge is needed, and supervising acquisition and distribution of knowledge. There is similar need of knowledge management in e-learning, as there is a need to collect, update, and reuse knowledge that are ultimately delivered to learners. Most of current e-Learning systems are constructed upon a knowledge base that is available over the Internet. The use of database technologies for storing and manipulating e-Learning resources (knowledge) provides support to learner community [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

- Collaboration Technology

Learning is a social practice and becomes more effective through interpersonal communications. Distant learners are learning together through collaborative learning. Groupware sustained collaborative learning leads to better student participation, better performance, and productivity than individual learning [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

- Human-Computer Interaction (HCI)

HCI (Human-Computer Interaction) is "a discipline concerned about the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them". One significant HCI issue is that diverse users have diverse ideas about interactions with computers. Users may be diverse in terms of values. Their interface liking may vary over time. A good e-Learning system should have a Web user interface that provides all possible actions related to learning process [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

- Evaluation of Learning

There are two approaches with respect to evaluation of e-learning: evaluation of learners' progress in learning, and evaluation of system functioning. Keeping track of learning progress allows delivery of the right resources to learners and allowing learning to be effective. Tailored learning model or profile that comprises learner's interests, and problems faced is required to evaluate a learner's progress in an e-learning system. Measuring the effectiveness of an e-learning is essential. The evaluation of effectiveness of distance learning can be achieved through gathering information about pre- and post-course surveys completed by students, records of learners' online activities (records of system usage and access), assessment grades, direct observations, and learner-teacher and learner-learner interactions in the learning process [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

1.5. Roles of Interaction in E-learning

In E-learning technologies, teacher–student activities can be supported either among groups or individually, and in either real time (synchronously) or in non-real time (asynchronously). Below, we briefly review the six forms of interactions supported through e-learning. We present a diagram of the six interactions possible among the three critical players in a formal educational setting – students, teachers, and content as shown in Figure 1.2.

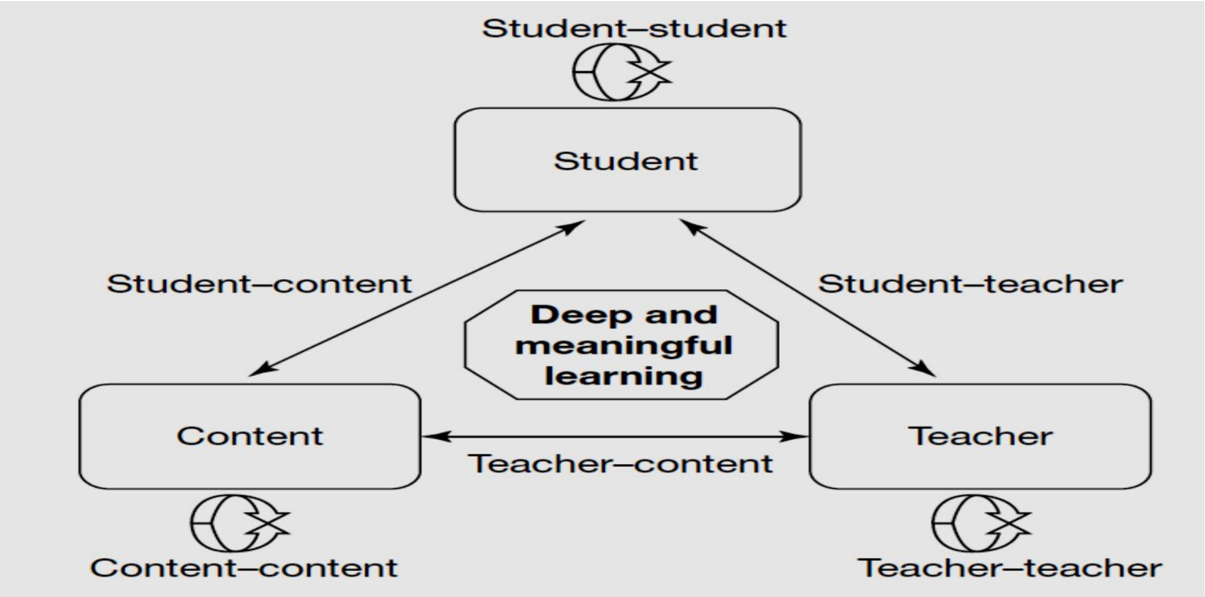


Figure 1.2:Modes of interaction [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis.

1.5.1. Teacher–Student Interaction

Many of the qualities of interaction between students and instructors in e-learning contexts can be both defined and measured and have impact on learning outcomes. We did not consider this in our study to measure engagement.

1.5.2. Student–Student Interaction

In an e-learning context this interaction is supported through a variety of communications technologies, in both synchronous and asynchronous formats. E-learning expands the rich tradition of independent study associated with earlier generations of distance education and provides and often mandates a variety of synchronous and asynchronous learning activities. The design of appropriate amounts of interaction is critical and depends on a variety of factors, many of which are rooted in the expectations and capacity for interaction expressed by the students. This kind of interaction is applied in our study using discussion forum.

1.5.3. Student–Content Interaction

Students spent the majority of their time by interacting with educational content. In e-learning contexts, content can be expressed in text for reading on screen or on paper, but content is often supplemented with a rich variety of computer assisted instruction, simulations, micro worlds, and presentation creation tools. In the past, content was assumed to be static and slow – waiting for consumption by students. Now content can be animated and given agent like properties of autonomy, volition, and rationality and can be programmed to take a more active part in student–content interactions. This kind of interaction is applied in our study. We applied static content.

1.5.4. Teacher–Content Interaction

This form of interaction refers to interaction between teachers and content. The development and application of content objects has become an increasingly important component of the teacher's role in e-learning. We did not apply this type of interaction to measure engagement in our study.

1.5.5. Teacher–Teacher Interaction

The pervasive existence of low-cost, multimedia networks is providing unprecedented opportunities for teacher–teacher interaction. Teacher–teacher interaction is the cornerstone of community within which teachers function. We did not apply this type of interaction to measure engagement in our study.

1.5.6. Content–Content Interaction

Computer scientists and educators are creating 'intelligent' programs or agents that 'differ from conventional software in that they are long-lived, semi-autonomous, proactive, and adaptive'. Agents are currently being developed and deployed that are capable of retrieving information, operating other programs, making decisions, and monitoring other resources on the network. We can imagine an era when content is automated to update itself from various sensory inputs and then to alert students and teachers when these alterations reach a significant level [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis. This type of interaction is out of scope of our study.

1.6. Motivation

Engagement is a "multifactor construct". Previously known are three main factors of engagement: "behavioural, emotional and cognitive" [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. However, according to [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204., within online environments, there are five factors of engagement related to online learning environment: "social engagement, cognitive engagement, behavioural engagement, collaborative engagement, and emotional engagement". The five factors mentioned above are interconnected to each other and revealed to be critical for active learner engagement and impact engagement in online learning [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. We will focus on the three engagement factors namely: behavioural factor, collaborative factor and emotional factor. Some earlier works tried to detect student engagement from behavioural or interaction factors [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.Others detected student engagement from collaboration factors such as discussion on discussion forums [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).in LMS. Some also detected student engagement from emotional factors [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.; [Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences.with the facial emotion recognition tools. All these works applied many features. [Sarsa and Escudero, 2016] Sarsa, J. and Escudero, T., 2016.A Roadmap to Cope with Common Problems in E-Learning Research Designs.*Electronic Journal of E-learning*, *14*(5), pp.336-349. remarked that the high number of features involved in e-learning processes complicates and masks the identification and isolation of the intervening factors. There was no also empirical evidence for identifying the relationship between features of the three factors and student engagement levels, that is whether there is a positive or negative relationship and identifying which ones are the most important features. This motivates us to identify the relationship between features of the three factors which are behavioural, collaboration and emotional factors and student engagement levels, that is whether there is a positive or negative relationship through correlation analysis and identifying which ones are the most important features applying Pratt's index.

The importance of ensuring quality in programs and pedagogy has to be emphasised so that online students receive the same level of support as face-to-face students, cautioning that a tiered system of educational segregation could potentially result if this was not consciously addressed. This is particularly important given the disproportionate number of minority, part-time, and working-class students who elect to study online. Moreover, although asynchronous e-learning allows learning independent of time, place and pace, it has a disadvantage as students feel isolated. The feeling of the isolation leads to drop-out [Leo et al., 2009] Leo, T., Manganello, F., Pennacchietti, M., Pistoia, A. and Chen, N.S., 2009, July. Online synchronous instruction: Challenges and solutions. In 2009 Ninth IEEE International Conference on Advanced Learning Technologies (pp. 489-491).IEEE.. Support for students is therefore vital to ensure student engagement and positive learning outcomes from studying online [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R.

and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. As we explained above, previously known are three main factors of engagement: “behavioural, emotional and cognitive” [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. The existing works did not consider building student engagement prediction models from three factors namely behavioural, collaboration and emotional factors. They neglected adding the element of collaborative engagement factor as explained by [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. which asserts that individuals’ interactions with teachers or other students have been identified as key influencer of engagement. Moreover, [Calvo and D'Mello, 2010] Calvo, R.A. and D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing, 1(1), pp.18-37. remarked that affect detection systems that integrate data from different factors have been widely advocated but rarely implemented. [Kizilcec et al., 2013] Kizilcec, R.F., Piech, C. and Schneider, E., 2013, April. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179). also pointed out that constructing a model of engagement with smallest granule of time has not been implemented widely, but implementing it is important as it allows to uncover subtler patterns. These motivate us to build student engagement prediction model using non-linear regression technique from three factors: behavioural, collaboration and emotional factors across micro level time scale such as 5 minutes in an asynchronous online learning environment to identify at risk students as quickly as possible before they disengage [Falkner and Falkner, 2012] Falkner, N.J. and Falkner, K.E., 2012, September. A fast measure for identifying at-risk students in computer science. In Proceedings of the ninth annual international conference on International computing education research (pp. 55-62)..

One of the most common criticisms related to the quality of educational outcomes of the online learning is due to the lack of face to face faculty student interactions [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.. In online environments, students are unmonitored. Faculty lack the tools to ensure complete viewing of the lecture content [Sarder, 2014] Moreover, in the e-learning environment, where the teacher is not physically present, monitoring a student for interest or engagement is a challenge [Al-Alwani, 2016] Abdulkareem Al-Alwani, 2016,A Combined Approach to Improve Supervised E-Learning using Multi-Sensor Student Engagement Analysis,American Journal of Applied SciencesAccessed on 25 July 2019. Instructors are overwhelmed by data reports provided to them in the online courses [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18.. Such challenges were solved in the literature by classifying the student into different classes of engagement levels using the model they built. [Coffin et al., 2014] Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G., 2014, March. Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 83-92).applied a model and classified students into three categories: auditors, active and qualified and visualized the outputs of the model predictions. However, they did not consider further classifying or filtering students in one of these categories. After classifying a learner as auditors, there may be a large number of students in this particular category. This motivates us to propose a student engagement visualization dashboard that visualizes the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner.

1.7. Objectives

The general objective of this thesis is to present a student engagement awareness system for an asynchronous e-learning platform, which helps in predicting a student engagement level from their interaction with a content, their interaction with another student and their emotion. We have chosen a LMS called Moodle because it is an asynchronous online learning technology that gives students the ability to communicate with their teachers and peers with a fulltime open access to content including course materials, lecture notes, tutorials, messages and recordings. In the LMS, the tasks that were performed were content viewing, quiz, assignment and discussion forum as online tasks using desktop equipped with a webcam for detecting facial expression for emotion recognition in real time while performing the above mentioned tasks. We chose the LMS because its modular design makes it easy to create new courses. It also allows to create interactive course material such as assignments, lesson and quiz. Students can interact with each other through activities such as forum. Moreover, it keeps detailed logs of all activities that students perform. It logs every click that students make for navigational purposes and has a log viewing system built into it. Log files can be filtered by course, participant, day and activity. Content viewing, quiz, and assignment were designed to detect behavioural engagement factor. We also applied posting to a discussion forum to be used as one of the tasks to detect collaborative engagement factor. The facial emotion recognition was performed to get emotional factors.

To achieve the general objective, we have three sub-goals. The sub-goals are as follows.

(a) Detection of behavioural, collaboration and emotional states of the students for the assumed asynchronous e- learning platform.

(b) Building and validating a student engagement level prediction model to predict an engagement level of the students based on their behavioural, collaboration and emotional states in the LMS activities.

(c) Building and validating a student engagement level visualization system to visualize the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner in the asynchronous e- learning platform.

1.8. Contributions of Thesis

In the asynchronous online learning environment, where the teacher is not physically present, monitoring a student for interest or engagement is a challenge. The interactions data can be correlated to four levels of engagement applying Pearson correlation analysis to determine the significant features that affected a given level of engagement. The significant features can be applied in building the student engagement prediction model. We can visualize the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner. The following are the contributions of this thesis.

1.8.1. Empirical Data Collection and Analysis

We required empirical data to determine the significant features and most important features from three factors: behavioural, collaboration and emotional factors. We used the significant features to build the student engagement level prediction model. We also collected the empirical data to validate the student engagement level prediction model. We collected the empirical data in the laboratory setting, i.e., through controlled experiments (CEs). We collected the data because there were no readymade corpora in the literature, which could be used to build and validate the model. Specially, the data related to emotional factors as a log file of rate of emotional features detected through facial emotion recognition tool was lacking. We used the LMS for allowing interaction of student with a content and also interaction of student with another student to detect the behavioural and collaboration features respectively. We implemented facial emotion recognition tool to detect the rate of emotion in real time and download as log file for analysis. We applied Pearson correlation analysis to determine the significant features and Pratt’s index to determine the most important features.

1.8.2. The Most Important Factors Affecting Student Engagement Levels

We performed an empirical study to determine the most important features from three factors: behavioural, collaboration and emotional factors using Pearson correlation analysis and Pratt's index. The contributions of our study was the finding that the most important feature to affect very low level of engagement was the Time of assignment submission (TA) which was a behavioural feature when compared to emotional feature that was surprise (SUR). The most important feature to affect low levels of student engagement was surprise (SUR) which was emotional feature when compared with another collaboration feature that was time between post and reply (TPR) and three behavioural features which were time of assignment submission (TA), time of reading content (TRC) and score of quiz (SC). Happy (HAP) was the most important emotional feature that affected high level of student engagement when compared with other three emotional features such as sadness (SAD), anger (ANG) and surprise (SUR) emotions and other collaboration and behavioural features. The most important feature to affect very high level of student engagement was number of replies (NR) which was collaboration feature when compared with other emotional features that were anger emotion (ANG), surprise (SUR) and happy emotion (HAP) and four behavioural features which were Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV) and score of quiz (SC), and one collaboration feature which was Time between post and reply (TPR).

1.8.3. Student Engagement Level Prediction Model

We have built and validated a student engagement level prediction model to predict the engagement levels of the students in asynchronous e-learning platform. We built the model using 9 features that were significant out of 13 features to affect the levels of student engagement and emerged in the final model. The student engagement prediction model was built using non-linear regression technique from three factors: behavioural, collaboration and emotional factors. The model is able to specify whether a student is at very low engagement level (VL), low engagement level (L), high engagement level (H) or very high engagement level (VH) across micro level time scale such as 5 minutes to identify at risk students as quickly as possible before they disengage.

1.8.4. Implementation of Student Engagement Awareness System

We have implemented a student engagement level visualization system to visualize the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner to the teacher in the asynchronous e- learning platform.

We incorporated the student engagement level prediction model in the system to predict the engagement levels of the students into one of the four levels namely, very low engagement level (VL), low engagement level (L), high engagement level (H) or very high engagement level (VH).

1.9. Thesis Organization

The ultimate goal of the thesis is building and validating a student engagement awareness system for a teacher in an asynchronous e-learning environment. To achieve this goal, we made the above three contributions. Below, we present a thesis flow diagram in Figure 1.3 that helps understand all our contributions more clearly, and a brief description of each chapter of the thesis.

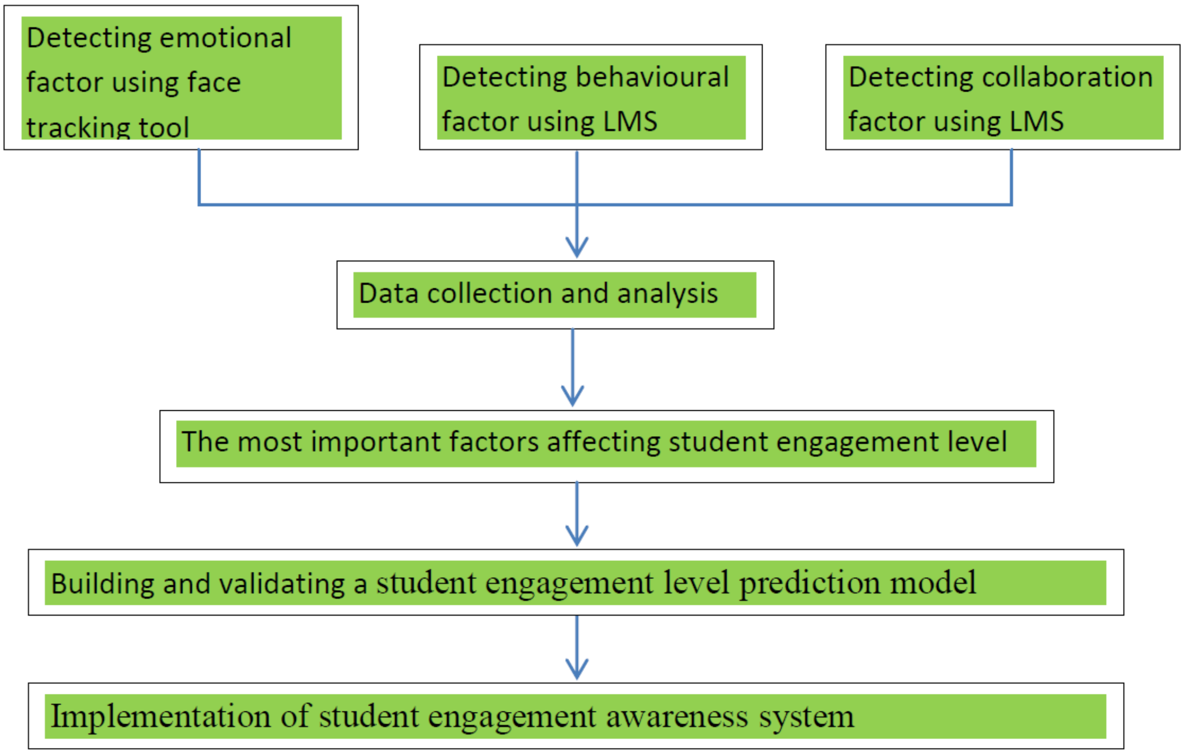


Figure 1.3: Thesis-flow diagram

The thesis consists of six chapters.

Chapter 1, entitled “Introduction", which briefly introduces the problems answered in the thesis, along with a description of the basic terms to understand the general work.

Chapter 2, entitled “Review of Related Work", presents the existing works which are related to the proposed system. This chapter reports the existing systems, and the works related to the components of the same, with their critical analyses.

Chapter 3, entitled “Factors Affecting Student Engagement”, describes an empirical study to determine the significant features and the most important features from three factors: behavioural, collaboration and emotional factors using Pearson correlation analysis and Pratt’s index.

Chapter 4, entitled “Student engagement level prediction model",, reports all the empirical studies conducted to build a student engagement prediction model using non-linear regression technique from three factors: behavioural, collaboration and emotional factors. This chapter contains a detailed description of the validation of the model.

Chapter 5, entitled “Student Engagement Awareness Dashboard ", presents the details of a student engagement level visualization system to visualize the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner to the teacher in the asynchronous e- learning platform.

Chapter 6, entitled “Conclusion and Future Scope", concludes the thesis by presenting a summary of the thesis, along with the discussion on ways for future research based on the current work.

Chapter 2

Review of Related Work

2.1. Introduction

We reviewed works in the literature which are related to the proposed student engagement awareness system for an asynchronous e-learning platform. We focused on research works which dealt with quantitative methods. In relation to this, we surveyed the literature on the following topics: roles of engagement in distance learning, engagement detection methods, factors influencing engagement, engagement prediction models and visualizations of engagement levels. In this chapter, we explained each of the topics in detail.

2.2. Role of Engagement in Distance Learning

Engagement refers to the behavioural intensity and emotional quality of a person's active involvement during a task [Reeve et al., 2004] Reeve, J., Jang, H., Carrell, D., Jeon, S. and Barch, J., 2004. Enhancing students' engagement by increasing teachers' autonomy support. Motivation and emotion, 28(2), pp.147-169. Engagement in the e-learning environment never obtained due consideration in the past. Student engagement is associated with critical results such as scores, perseverance, and graduation [Manwaring et al., 2017] Manwaring, K.C., Larsen, R., Graham, C.R., Henrie, C.R. and Halverson, L.R., 2017. Investigating student engagement in blended learning settings using experience sampling and structural equation modeling.The Internet and Higher Education, 35, pp.21-33.. Student engagement requires advance study as the e-learning existence of universities has improved. The e-learning is sought by those who want to pursue their education while accomplishing the other responsibilities of life such as work and family. These students who shoulder responsibilities face attrition, which makes the area of e-learning student engagement to be an area that needs extra investigation. Student in the e-learning setting can regularly feel isolated and as such it requires attention as it affects the learner's education[Dixson, 2015] Dixson, M.D., 2015. Measuring student engagement in the online course: The Online Student Engagement scale (OSE). *Online Learning*, *19*(4), p.n4.

Moreover, if a student loses interest or is not getting engaged in the e-learning session, the teacher cannot easily monitor as the setting is e-learning [Al-Alwani, 2016] Abdulkareem Al-Alwani, 2016,A Combined Approach to Improve Supervised E-Learning using Multi-Sensor Student Engagement Analysis,American Journal of Applied SciencesAccessed on 25 July 2019. Because engagement represents a direct pathway to learning, disengagement (losing interest or not getting engaged) provides barriers to achieving learning outcomes [Hancock and Zubrick, 2015] Hancock, K.J. and Zubrick, S., 2015.Children and young people at risk of disengagement from school.Commissioner for Children and Young People, Western Australia.

2.3. Methods of Engagement Detection

Engagement is a latent that requires layers of indicators. In this section, engagement measurement was discussed in detail.

2.3.1. Measuring Engagement

Engagement is not readily observed, because it is an inner quality of concentration and effort, so it must be inferred from manifest indicators such as the amount of participation in academic work (attendance, amount of time spent on academic work), and interest and enthusiasm exhibited by students [Rumberger and Rotermund, 2012] Rumberger, R.W. and Rotermund, S., 2012. The relationship between engagement and high school dropout. In Handbook of research on student engagement (pp. 491-513). Springer, Boston, MA.

The existing works for detecting the engagement of the students can broadly be categorized into two groups: quantitative and qualitative methods. These are shown in Figure 2.4 which is a hierarchical diagram and discussed in the following subsections.

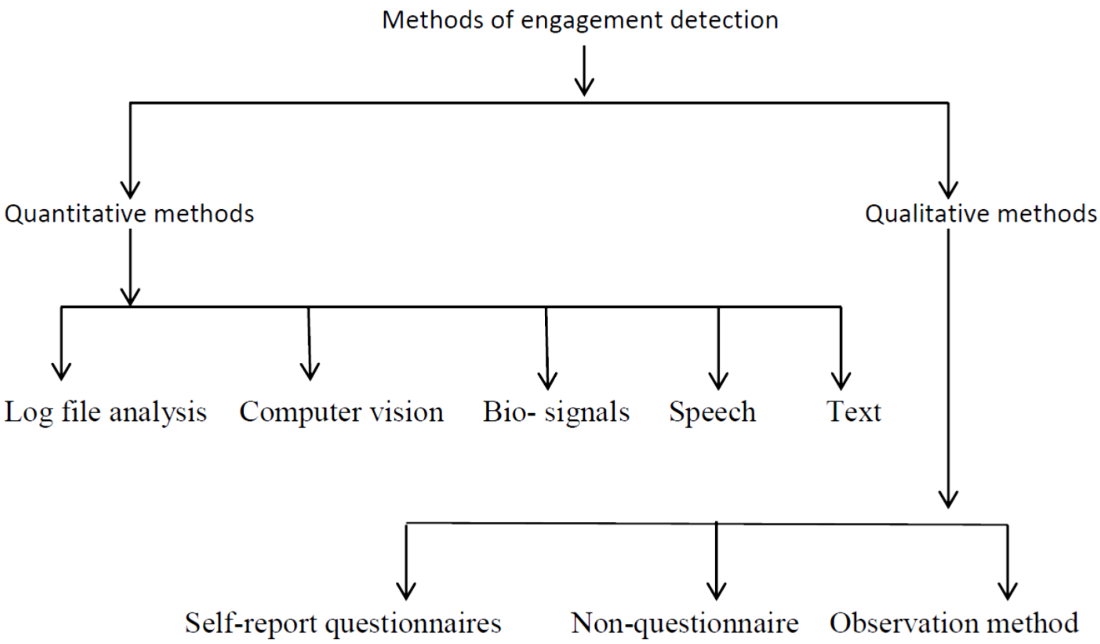


Figure 2.4: A Hierarchical diagram of engagement detection methods

2.3.2. Quantitative Methods

Quantitative methods of student engagement measurements use behaviour data from e-learning activities [Hu and Li, 2017] Hu, M. and Li, H., 2017, June. Student engagement in online learning: A review. In *2017 International Symposium on*

Educational Technology (ISET) (pp. 39-43).IEEE. Quantitative methods involve gaining data with the help of analysis of log data and device (e.g. a webcam).

• ***Log file Analysis***

In the log-file analysis, learners' actions preserved in log files are analysed for the engagement detection. Especially, in an e-learning environment, the learners' actions are stored in log files. This can provide valuable information for the engagement detection. Different data mining and machine learning approaches are used in the log-file analysis [Dewan et al., 2019] Dewan, M.A.A., Murshed, M. and Lin, F., 2019. Engagement detection in online learning: a review. Smart Learning Environments, 6(1), pp.1-20.

[Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124. analysed log-files in a web-based learning environment called HTML-Tutor. A similar work was conducted by [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.who used log-file analysis through mouse click behaviour to predict engagement. [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789 examined the construct validity of activity logs as a measure of student engagement. Specifically, they investigated the relationship between features of student activity derived from LMS web logs, and instructors' ratings of student engagement. They found that estimators derived from LMS web logs are closely related to instructor ratings of engagement. [Badge et al., 2012] Badge, J.L., Saunders, N.F. and Cann, A.J., 2012. Beyond marks: new tools to visualise student engagement via social networks. Research in Learning Technology, 20.used student contributions to a social network to detect student engagement in a peer to peer discussion in an activity stream environment. [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).used health based support forums like DailyStrength as social networking domain to detect engagement.

- ***Computer Vision Methods that Use Facial Expressions Analysis***

A computer based emotion recognition system using facial images in video streams will need to address face identification and tracking, feature extraction, classify and predict emotions. Facial expression recognition is a suitable method to use for emotion recognition in a learning management system due to two reasons. First, extensive research conducted in face expression recognition means that the latest approaches are very mature and have reached an arguably high level of precision. Second, the method only requires a video camera and no other hardware [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore..

[Whitehill et al., 2014] Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A. and Movellan, J.R., 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), pp.86-98.automatically measured behavioural engagement from videos. The face and facial landmark (eyes, nose, and mouth) positions are localized automatically in the image. They collected training data from 34 undergraduate students who interacted with an in-house cognitive skills training software system on an Apple iPad. The researchers recorded video with a commercial webcam aimed directly at students' faces and used computer vision techniques to detect facial expressions and facial textures. Trained coders annotated the videos for behavioural engagement using the following ordinal scale: 1 (not engaged at all), 2 (nominally engaged), 3 (engaged in tasks), and 4 (very engaged). The model's estimate of behavioural engagement (Level 4 vs. Levels 1, 2, and 3) correlated with performance gains assessed before and after training, providing some evidence for its predictive validity.

[Manseras et al., 2018] Manseras, R., Eugenio, F. and Palaoag, T., 2018, March. Millennial Filipino Student Engagement Analyzer Using Facial Feature Classification. In IOP Conference Series: Materials Science and Engineering (Vol. 325, No. 1, p. 012006). IOP Publishing. utilized OpenFace as a tool. They used individual faces for identifying the various facial Action Units, for classification of facial features using Support Vector Machine. They predicted two engagement categories: engaged or disengaged. They were able to visualize percentage of labelled engaged students.

[Soloviev, 2018] Soloviev, V., 2018. Machine learning approach for student engagement automatic recognition from facial expressions. Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics, Informatics and mechanics, 10(2), pp.79-86.applied Faces through images containing a snapshot of students sent to Microsoft Azure Cognitive Services for recognizing the emotion, facial landmark. They predicted two engagement categories: engaged or disengaged. They visualized aggregation of average engagement for groups, courses, faculty on interactive dashboard.

- ***Bio- signals***

Popular methodologies for such systems include blood pressure, oxygen level in the blood, skin conductance, heart rate or pulse rate and electrocardiogram signals. While the use of physiological signals for emotion recognition is arguably more precise and less prone to ambiguities, it requires the use of hardware in the form of bio signal readers. Such devices add to financial cost and may be inconvenient to the user [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore..

- ***Speech Signals***

Typically, emotion in speech can be detected through vocal or acoustic characteristics such as pitch, tone or energy. Learning algorithms in such emotion recognition systems detect the change in pitch and/or energy patterns of streaming audio and a classifier assigns the type of emotion associated to the change. A typical example of a speech based emotion recognition system is the Jerk-O- Meter [Jerk-O-Meter41 - MIT Media], which monitors phone conversation and detects stress levels in speech. We also argue that the use of speech input for emotion detection is not suitable in e-learning considering it is unnatural for e-learning students to talk to a computer system in most learning management systems. The fact that many students access learning management systems in public spaces such as libraries makes it very difficult to implement a speech based e-learning system [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore..

- ***Text***

Emotional context in text is often examined by literary critiques and linguists. Literary critiques scrutinise literary texts to look for cues to interpret characters' direct or indirect emotions. Social and cultural researchers have found that emotion is expressed differently in text and emoticons across different cultures. Secondly, it is rather easy for users to hide their emotions intentionally in their texts by using neutral texts [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore..

2.3.3. Qualitative Methods

Qualitative methods use data from questionnaire or interview [Hu and Li, 2017] Hu, M. and Li, H., 2017, June. Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)* (pp. 39-43).IEEE.

- ***Self-report Questionnaires***

The most widely used measures of engagement are self-report questionnaires. Although relatively inexpensive, easy to administer, and generally reliable, the validity of the self-reporting results depends on a number of factors that are outside of the control of the researchers, such as learners' honesty, their willingness to report their emotion, and the accuracy of learners' perception about their emotions [Dewan et al., 2019] Dewan, M.A.A., Murshed, M. and Lin, F., 2019. Engagement detection in online learning: a review. Smart Learning Environments, 6(1), pp.1-20..

- ***Non-questionnaire Engagement Measures***

Several non-questionnaire engagement measures have also been developed. Examples include experience-sampling methods (ESM) and interviews. However, because they still rely on self- and informer-reports, they are subject to similar biases as questionnaires [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123..

- ***Observational Methods***

Observational methods are attractive alternative to self- and informer-reports because they are arguably more objective. Unfortunately, these methods entail considerable human effort, which might not be a major limitation for small scale studies, but poses a significant challenge for repeated long-term measurement at scale. Further, observations cannot be conducted in some learning contexts, such as students' homes. Finally, engagement can be detected from academic and behaviour records, such as homework completion, absences, achievement test scores, and teacher ratings of classroom conduct but these measures are limited in what they can reveal about engagement at the micro-analytic level [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123.. The usefulness of such behaviour records are limited in terms of making activity-specific interventions that are based on the findings [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18..

2.3.4. Benefits of Quantitative Methods

The qualitative detection methods and three of the quantitative methods of engagement detection namely, bio-signals, speech signals and text have limitations. The validity of self-report questionnaires is not guaranteed, and questionnaires are obtrusive which means they interfere with the learning of the student [Bahraini et al.,2016] Bahreini, K., Nadolski, R. and Westera, W., 2016. Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments, 24(3), pp.590-605. Interviews are biased, observational methods require human effort [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123. and bio-signal devices are costly and will create inconvenience to the learner whose engagement is supposed to be detected [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore..

Logfile analysis through the use of LMS [You, 2016] You, J.W., 2016. Identifying significant indicators using LMS data to predict course achievement in online learning.*The Internet and Higher Education*, *29*, pp.23-30.and computer vision methods through the input devices (such as webcams) with the facial emotion recognition tool [Bahraini et al.,2016] Bahreini, K., Nadolski, R. and Westera, W., 2016. Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments, 24(3), pp.590-605. are two of the quantitative engagement detection methods that have benefits. The LMS enables log file analysis that has the unobtrusiveness advantage [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.. The input devices (such as webcams) are inexpensive and are used with facial emotion recognition tools to give natural interactions with the online learning applications. Moreover, they also offer capability of unobtrusive and continuous data gathering [Bahraini et al.,2016] Bahreini, K., Nadolski, R. and Westera, W., 2016. Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments, 24(3), pp.590-605..

Because of these benefits, we applied the two advantageous quantitative methods: the log file analysis and facial emotion recognition tool to measure an engagement. This engagement is assumed to be an interaction of the individual qualities the learner brings to the learning situation (Emotional factor) and the contextual qualities facilitated by the learning tool (Behavioural and collaboration factors). These factors which influence engagement will be discussed next.

2.4. Factors Influencing Engagement

Engagement is a "multifactor construct". Previously known are three main factors of engagement: "behavioural, emotional and cognitive" [Redmond et al.,2018)]

Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. However, according to [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204., within online environments, there are five factors of engagement related to online learning environment: "social engagement, cognitive engagement, behavioural engagement, collaborative engagement, and emotional engagement". The five factors mentioned above are interconnected to each other and revealed to be critical for active learner engagement and impact engagement in online learning [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. Figure 2.5 below shows online engagement framework overview of these five factors.

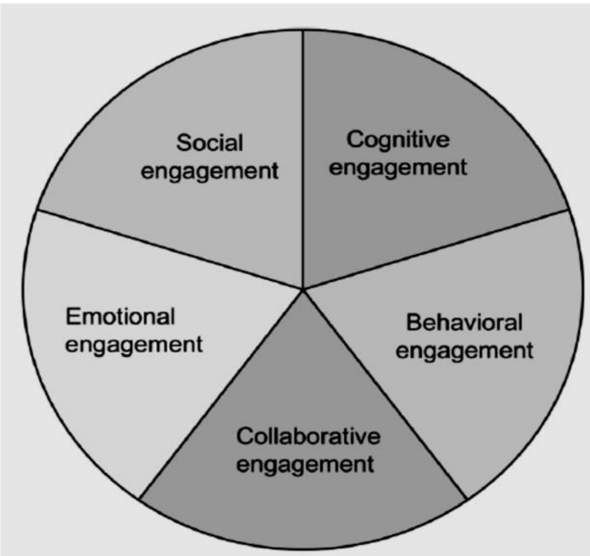


Figure 2.5**:** Online engagement framework overview [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An

online engagement framework for higher education. Online learning, 22(1), pp.183-204.

These factors are defined by [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. as follows: Behavioural engagement is defined as “active participation in academic activities” and it is described as “doing the work and following the rules”. Collaborative engagement is related to “the development of different relationships and networks that support learning, including collaboration with peers and instructors.” Emotional engagement refers to “students’ emotional reaction to learning. It is related to their feelings or attitudes towards learning”. Social engagement refers to “students’ social investment in the collegiate experience”. “It includes participation in academic as well as non-academic activities which occur outside the virtual classroom, such as recreation or social functions, along with discussions of a social nature”. Cognitive engagement is “the active process of learning. It is related to what students do and think to promote learning” [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. [Baker and Rossi, 2013] Baker, R.S. and Rossi, L.M., 2013. Assessing the disengaged behaviors of learners. Design recommendations for intelligent tutoring systems, 1, p.153.remarked that deciding which factor(s) of engagement to model is a challenge. Not all factors (or aspects of each factor) need to be detected in order to support effective intervention. Specific factors impact learning outcomes and longer-term engagement in different ways, and some are more important to identify and adapt to than others, depending on the learning context.

We did not consider measuring social engagement as it is not relevant to our learning context. As we stated above, our learning context is Learning Management System that stores data related to students' activities such as content reading, writing, taking test, and communication with peers. On the other hand, social engagement requires participation in non-academic activities which occurs outside the virtual classroom, such as recreation [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. Similarly, we did not consider measuring the cognitive engagement. In our learning context, students' activities are time stamped and logged by defaults which are used as measures of engagement [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789. They pointed out that what these logs are measuring about mind set remains unclear. Cognitive engagement factor requires, for instance, measuring whether a student justifies an idea in his post or not [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204., which is not readily available in our learning context. Moreover, if we implemented the capability to measure the above instance of cognitive engagement factor, it requires content analysis technique that will use text-based activities [Woodfine et al., 2008] Woodfine, B.P., Nunes, M.B. and

Wright, D.J., 2008. Text-based synchronous e-learning and dyslexia: Not necessarily the perfect match!.Computers& Education, 50(3), pp.703-717.. However, such text-based activities can marginalise, demotivate and disappoint students with dyslexia with difficulties in reading, spelling, word order and argumentation [Woodfine et al., 2008] Woodfine, B.P., Nunes, M.B. and Wright, D.J., 2008. Text-based synchronous e-learning and dyslexia: Not necessarily the perfect match!.Computers& Education, 50(3), pp.703-717.. At times the line between cognitive and behavioural engagement became blurred [Henrie et al., 2015] Henrie, C.R., Halverson, L.R. and Graham, C.R., 2015. Measuring student engagement in technology-mediated learning: A review.*Computers& Education*, *90*, pp.36-53..

We will focus on reviewing the literature regarding the three engagement factors namely: behavioural factor, collaborative factor and emotional factor.

2.4.1. Behavioural Factors

Earlier works in [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018., [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789 and [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning:

Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.detected student engagement from behavioural or interaction factors. [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123.mentioned that interaction features are best suited for behavioural engagement. [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages.

https://doi.org/10.1145/3303772.3303789applied 19 features of the behavioural factor to build a model to predict student engagement. [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.identified relevant features from a web-based interactive environment, HTML-Tutor, to predict whether a learner is disengaged. Their research study analysed 30 attributes, second data set containing 10 features and the third dataset containing 6 features of the online learners' from the log file including a number of pages accessed, average time spent on pages, number of tests attended, number of correctly answered tests, and number of incorrectly answered tests. [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018. reported 10 activities in VLE (Virtual Learning Environment) to be important predictors of student engagement. However, activities and features on activities were not differentiated. For instance, forumng is an activity in discussion forum, and the number of time each student clicks on this activity is a feature. Thus, it is supposed to be features on activities that are predictors of student engagement, not the activities themselves. One disadvantage of taking activity as predictor is sometimes an activity may not be observed but the features may. One such scenario is emotional features from facial expressions analysis. On the other hand, they described features such as the total number of times VLE activities were accessed and final results to be important predictors of student engagement. Nevertheless, the mechanism to determine which feature is the most important is based on which feature occurred most frequently which is not correct method. [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.and [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018. used learner to content interaction. Taking test is an activity used by the two papers and the common features extracted for use in the prediction were: Number of correctly answered tests, and number of incorrectly answered tests. Hussain et.al., (2018) also found out that student clicks on forum discussion, accessing content, subpage, and uRL were moderately correlated with level of engagement in VLE activities. The number of student clicks on accessing resources and collaborate were weakly correlated, the number of clicks on the accessing homepage was highly correlated with the level of engagement and the number of clicks on accessing glossary and data were unrelated to the student level of engagement.

2.4.2. Collaboration Factors

[D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123. explained collaboration engagement factors to be related to the development of different relationships and networks that support learning, including collaboration with peers that is related to engagement for academically worthwhile purposes such as discussion. [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).detected engagement from collaboration factors such as discussion on health forums. They applied 16 features to predict continued participation. They used the coefficients (weights) for the independent variables (features) in logistic regression, to determine the most important features. One basic attribute that can play an important role in predicting future engagement is the amount of activities performed by a user in the observation period. For example, in online forums or bulletin boards the amount of activity can be just the number of pieces of content a user has posted. They applied the number of posts for participation prediction task. Whereas they found replies received from other users had little to no effect on a user's future participation. In their research, [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and

Information Analysis (pp. 12-20).showed that lurking period (time gap between a user's registration and first activity, time gap between a user's last activity and end of an observation period), and average idle time between activities can be useful predictors of a user's future participation in a social media. All three of these features had positive correlation with a user's discontinuation of participation, that is, the higher these numbers were, the more likely that user was going to leave the forum. However, their study was not implemented in the e-learning environment.

[Badge et al., 2012] Badge, J.L., Saunders, N.F. and Cann, A.J., 2012. Beyond marks: new tools to visualise student engagement via social networks. Research in Learning Technology, 20.used student contributions to a social network to detect student engagement in a peer to peer discussion in an activity stream environment. Contributions could be in the form of status updates, comments on others updates, or shared links, but to count for credit, external links must be accompanied by a short commentary explaining how and why it is relevant to academic study. Students were introduced to Friendfeed social network at the start of the module. The Friendfeed data can be used to visualise three types of network :(1) Subscriptions (or ‘‘following’’), (2) Comments (made and received) and (3) Likes (affirmations which refocus attention by moved the liked item to the top of the activity stream). In order to investigate network relationships, they used Gephi, [Bastian et al. ,2011] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media to visualise interactions in student networks. [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.applied also learner to learner interaction model which uses the discussion forum activity.

2.4.3. Emotional Factors

There is a role by the teacher played in the traditional classroom and e-learning needs similar mechanism to recognize emotion and communicate to the teacher. The emotion from facial expression is recommended to get the affective state of students because emotion first reflects on the face [Ray and Chakrabarti, 2012] Ray and Chakrabarti (2012), Design and Implementation of Affective E-Learning Strategy Based on Facial Emotion Recognition,Proceedings of the InConINDIA 2012, AISC 132, pp. 613–622.. The emotion from facial expression contains significant details and plays a communicative role [Razuri et al., 2013] Rázuri, J.G., Sundgren, D., Rahmani, R. and Cardenas, A.M., 2013, November. Automatic emotion recognition through facial expression analysis in merged images based on an artificial neural network. In 2013 12th Mexican International Conference on Artificial Intelligence (pp. 85-96). IEEE..

[Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences.proposed an affective model that measured student engagement based on their emotions. They mapped different emotions to five levels of engagement. These levels are strong engagement, high engagement, medium engagement, low engagement and disengagement. They used observation of facial expression from recorded videos and self-reporting method to detect the emotions. They used self-reporting method to detect the level of engagement of participants. They analysed

22 emotions, listed in detail in Table 2.1, in each level of engagement to detect strong emotions. The emotion that was felt by the largest number of participants indicated that the emotion was strongest. That strongest emotion will be mapped to a strong engagement level. Using self-report to detect emotion and levels of engagement has limitations. The validity of self-report questionnaires is not guaranteed, and questionnaires are obtrusive which means they interfere with the activities of the participant [Bahraini et al.,2016] Bahreini, K., Nadolski, R. and Westera, W., 2016. Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments, 24(3), pp.590-605..

[Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.combined information about the movements of the eyes, head, and facial emotions to produce a concentration index with three classes of engagement: “very engaged”, “nominally engaged” and “not engaged at all”. The model they built recognized a dominant emotion which is an emotion with the highest probability score. The concentration index is calculated by multiplying the dominant emotion probability and emotion weight. They developed a model that detected engagement in real-time. [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913. did not consider calculating engagement from different factors other than emotional factor while engagement is a multifaceted construct [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and

Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123..

[D'Errico et al., 2016] performed a correlation analysis to examine the relationship between positive and negative emotions during learning activities and engagement dimensions. However, their study never revealed what would happen to affective engagement with respect to the particular positive or negative emotion such as joy, anger or surprise. Furthermore, the impact of particular basic emotions felt on the level of student engagement has never been explored. They applied 14 items to measure the levels of intensity of emotions. Moreover, they used self-reporting, which has limitations to detect emotions.

[Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R. and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA. studied the impact of academic emotions on students’ cognitive, motivational, behavioural, cognitive-behavioural, and social-behavioural engagement. They considered positive affect to be comprised of various positive states (e.g., enjoyment, pride, satisfaction) and negative affect consisting of various negative states (e.g., anger, anxiety, and frustration). According to [Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R. and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA., negative emotions such as anger, anxiety, shame, boredom, and hopelessness were associated with task-irrelevant thinking and reduced flow (low engagement), whereas enjoyment related negatively to irrelevant thinking (low engagement) and positively to flow (high engagement). Moreover, positive affect leads to behavioural disengagement. Negative emotions such as sadness and anxiety may signal that there is a threat in the environment, suggesting that they may also contribute to intensified effort (high engagement). Positive emotions such as enjoyment of learning are positively associated with effort, and that negative deactivating emotions such as hopelessness and boredom are negatively associated with effort. In contrast, emotions such as anger, anxiety, and shame often show negative overall correlations with effort, but in some cases, they may support behavioural engagement as they can serve to energize students. However, they studied how emotions affect student engagement in classroom settings, which is different from the asynchronous online learning environment. Moreover, we used sensor (a webcam) and face tracking software while they used self-reporting to detect the emotion which is obtrusive.

Table 2.1 below shows type of factors, total number of features applied, measurement method and limitation of the reviewed works.

2.4.4. Limitations of the Existing Work

Some earlier works tried to detect student engagement from behavioural or interaction factors [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018., [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages.

https://doi.org/10.1145/3303772.3303789 and [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.. Others detected student engagement from collaboration factors such as discussion on discussion forums [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).in LMS. Some also detected student engagement from emotional factors [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.; [Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences. with the facial emotion recognition tools. Many of these works applied many features. [Sarsa and Escudero, 2016] Sarsa, J. and Escudero, T., 2016.A Roadmap to Cope with Common Problems in E-Learning Research Designs.*Electronic Journal of E-learning, 14*(5), pp.336-349. remarked that the high number of features involved in e-learning processes complicates and masks the identification and isolation of the intervening factors. We have seen no previous work that reported determining the most important factors whose features affected student engagement levels.

There are existing works which applied these three factors namely, behavioural factor, collaboration factor and emotional factor for building engagement prediction models. These works will be discussed in the next section 2.5.

S/no	Reference	Type of factors	Total number of features	Measurement method	Limitation
1	[Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789	Behavioural factor	19	Quantitative (log file analysis)	Many features(>4)
2	[Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.		30, 10 and 6 different attributes	Quantitative (log file analysis)	Many features(>4)

3	[Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.		10	Quantitative (log file analysis)	Did not consider multifactor engagement detection
4	[Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).	Collaborative factor	16	Quantitative (log file analysis)	Many features(>4)

5	[Badge et al., 2012] Badge, J.L., Saunders, N.F. and Cann, A.J., 2012. Beyond marks: new tools to visualise student engagement via social networks. Research in Learning Technology, 20.		4 features	Quantitative (log file analysis) and qualitative(self-report)	Did not consider multifactor engagement detection
6	[Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J.,	Emotional factor	7 emotions	Quantitative through data gained with the help of device(webcam)	Did not consider multifactor engagement detection

	2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.				
7	[Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences.		22 emotions	Quantitative through data gained with the help of device(webcam) and qualitative(observation and self-report)	Many features(>4)
8	[D'Errico et al., 2016]		14 items	Qualitative (self-report)	Many features (>4) and Only qualitative method (self-report) was applied

Table 2.1: Type of factors, total number of features applied, measurement method and limitation

2.5. Engagement Prediction Models

Predictive modelling is an activity of creating a model that will predict the values (or class if the prediction does not deal with numeric data) of new data based on observations. Predictive modelling is based on the assumption that a set of known data (referred to as training instances in data mining literature) can be used to predict the value or class of new data based on observed variables (referred to as features in predictive modelling literature) [Brooks and Thompson, 2017] Brooks, C. and Thompson, C., 2017. Predictive modelling in teaching and learning. Handbook of learning analytics, pp.61-68.. In this section, we presented the literature review concerning the engagement prediction models from the three factors namely, behavioural, collaboration and emotional factors.

2.5.1. Engagement Prediction Models from Behavioral Factors

[Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018. applied supervised machine learning algorithms to predict low student engagement from interaction in virtual learning environment. The features they applied in their study include highest education level, final results, score on the assessment and the number of clicks in virtual learning environment (VLE) activities. The machine learning models implemented were Decision Tree, J48, JRIP and gradient boosted algorithm. The output variables were engaged or not engaged. The prediction time they applied was weekly, which was very long time scale.

[Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789investigated the relationship between features of student activity derived from log files of LMS called Canvas and instructors ratings of student engagement. They applied logistic regression model through clustering technique to predict student engagement. They applied 19 features. Some of these features were time related, number of actions on activities and visits to activities. The detail of these features is given in Table 2.2. The prediction time they applied is a semester long, which is very long time scale. The output variables in their study were engaged or not engaged.

[Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124. developed disengagement prediction model on data of an e-learning systems called HTML tutor. They applied 8 data mining methods. These are Bayesian Nets with K2 algorithm and a maximum of three parent nodes (BN), Logistic regression (LR), Simple logistic classification (SL) that uses the LogitBoost algorithm, Instance based classification with IBk algorithm (IBk), Attribute Selected Classification using J48 classifier, Bagging using REP (reduced-error pruning) tree classifier (B), Classification via Regression (CvR) and Decision Trees with J48 classifier. The predicted outputs were engaged, disengaged and neutral variables. They compared three datasets based on number of features. One data set containing 30 features, second data set containing 10 features and the third dataset containing 6 features.

They applied the dataset containing the minimum number of features which is dataset containing 6 features. The detail of the features is listed in Table 2.2. Moreover, [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.applied the rule as learner is considered to be engaged when the logged data showed that users were focused on reading pages, taking tests or both, as well as performing other actions and spending a reasonable time on these actions and a learner was considered to be disengaged when they were browsing quickly through pages or when spending a long time on the same page or test. The time scale to predict disengagement is 10 minutes. The engagement levels that may occur at 5 minutes scale could not be predicted by their model.

2.5.2. Engagement Prediction Models from Collaboration Factors

[Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).developed logistic regression model to predict continued participation in an online health forum. They applied features such as the number of threads in a post, the number of replies, the number of days from the time of the last post or reply on discussion forum. They applied 16 features, listed in detail in Table 2.2, to predict continued participation. However, the prediction time interval they used is 1-month time which is very long time scale compared to prediction time of 5 minutes. However, the features of the discussion forum occurred in health related discussion, not e-learning related. To the best of our knowledge, the effect of these features in the discussion forum in the e-learning environment on the engagement level has never been studied. Moreover, they discussed that relations between features such as number of replies to someone's post and the time between someone's post and replies he got and engagement are unknown.

2.5.3. Engagement Prediction Models from Emotional Factors

[Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences.proposed an affective model that measured student engagement based on their emotions. They mapped different emotions to five levels of engagement. These levels are strong engagement, high engagement, medium engagement, low engagement and disengagement. They used observation of facial expression from recorded videos and self-reporting method to detect the emotions. They applied self-reporting method to detect the level of engagement of participants. They analysed 22 emotions, listed in detail in Table 2.2, in each level of engagement to detect strong emotions. The emotion that was felt by the largest number of participants indicated that the emotion was strongest. That strongest emotion will be mapped to strong engagement level. The time interval used to predict the engagement level was between 7 and 12 minutes, which is not as small time scale as 5 minutes. [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913. combined information about the movements of the eyes, head, and facial emotions to produce a concentration index with three classes of engagement: “very engaged”, “nominally engaged” and “not engaged at all”. The model they built recognized a dominant emotion which is an emotion with highest probability score. The concentration index is calculated by multiplying the dominant emotion probability and emotion weight. They developed a model that detected engagement in real time. [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913. did not consider calculating engagement from different factors other than emotional factor while engagement is a multifaceted construct [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123..

2.5.4. Limitations of the Existing Works

[Kizilcec et al., 2013] Kizilcec, R.F., Piech, C. and Schneider, E., 2013, April. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179). pointed out that constructing a model of engagement with smallest granule of time has not been implemented widely, but implementing it is important as it allows to uncover more subtle patterns. Most existing student engagement prediction models, which we reviewed, did not predict student engagement in smaller time scales such as 5 minutes. Moreover, providing students with support and guidance as soon as possible to reduce the risk of disengagement is critical [Falkner and Falkner, 2012] Falkner, N.J. and Falkner, K.E., 2012, September. A fast measure for identifying at-risk students in computer science. In Proceedings of the ninth annual international conference on International computing education research (pp. 55-62)..

In their works, [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018., [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789 and [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124. predicted student engagement from behavioural factors alone. However, engagement need to be defined as multi factor construct to ensure that the richness of real human experience is understood [Henrie et al., 2015] Henrie, C.R., Halverson, L.R. and Graham, C.R., 2015. Measuring student engagement in technology-mediated learning: A review.*Computers& Education*, *90*, pp.36-53.. Previously known are three main factors of engagement: “behavioural, emotional and cognitive” [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.. The existing works did not consider building student engagement prediction models from three factors namely behavioural, collaboration and emotional factors. They neglected adding the element of collaborative engagement factor as explained by [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204. which asserts that individuals' interactions with teachers or other students have been identified as key influencer of engagement. Moreover, [Calvo and D'Mello, 2010] Calvo, R.A. and D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing, 1(1), pp.18-37.remarked that affect detection systems that integrate data from different factors have been widely advocated but rarely implemented.

There are works in the literature, after building the engagement prediction models they visualize the predicted engagement levels on dashboards. We discussed these works in the next section, which visualized engagement levels from prediction models. We discussed also works which visualized engagement levels trends without applying engagement prediction models.

Table 2.2 below summarizes the student engagement prediction models used with input features, predicted output and prediction time scales for the three factors namely behavioural, collaboration and emotional.

References	Factors	Models Applied	Input Features	Predicted Outputs	Prediction Time Scale
[Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student	Behavioural factors	Decision Tree, J48, JRIP and gradient boosted algorithm	Highest education level, final results, score on the assessment and the number of clicks	High and low engagement levels	1 week

Course Assessment Scores.Computational intelligence and neuroscience, 2018.,					
[Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of		logistic regression	Time on asgmt pages (m), Avg time between first access & asgmt deadline (h), Avg session duration with asgmt views (h), Avg page views / session with asgmt views (c), Visits to 'Files' after an asgmt view (c), Visits to other 'Assignments' after an asgmt view (c), Visits to 'Modules' after an asgmt view (c), Visits to static 'Pages' after an asgmt view (c), Total asgmt views with no subsequent visit (c) , Visits to other Canvas tools after an asgmt view (c) , Number of asgmt submissions 6am-6pm (c), Number of asgmt submissions 6pm-midnight (c) , Number of asgmt submissions midnight-6am (c) , Total number of	engaged or not engaged	1 semester

activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789			submissions (c), Total visits to asgmt pages before deadline (c), Total visits to asgmt pages after deadline (c), Number of unique sessions with site visits (c), Visits to Canvas's 'Calendar' of assignments (c) , Longest period of inactivity within the site (h)		

[Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.		Bayesian Nets with K2 algorithm and a maximum of three parent nodes (BN), Logistic regression (LR), Simple logistic classification (SL) that uses the LogitBoost algorithm, Instance based classification with IBk algorithm (IBk), Attribute Selected Classification using J48 classifier, Bagging using REP (reduced-error pruning) tree classifier (B), Classification via Regression (CvR) and Decision Trees with J48	Number of pages, average time on pages, number of tests, average time on tests, number of correctly answered tests, number of incorrectly answered tests	Low, high and neutral levels of engagement	10 minutes

[Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health	Collaboration	logistic regression	PostCount, ReplyCount, SelfReplyCount, OtherReplyCount, TimeGap1, TimeGap2, AvgDays, Age, Gender, HasLocation, HasImage, PosUnigrams, NegUnigrams, TotalUnigrams, Question, Url	engaged vs not engaged	1 month

Text Mining and Information Analysis (pp. 12-20).					
[Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's	Emotional	The emotion that was felt by the largest number of participants indicated that the emotion was strongest than others. That strongest emotion will be mapped to strong engagement level.	Surprise, Enthusiastic, Excited, Angry, Ashamed, Fearful, Nervous, Happy, Content, Delighted, Joyful, Satisfied, Disgusted, Disappointed, Sad, Bored, Depressed, Tired, Sleepy, Relaxed, Still, Quiet	5 levels of engagement: strong, high, medium, low and disengagement	Between 7 and 12 minutes

engagement. Journal of King Saud University-Computer and Information Sciences.					
[Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using		The concentration index is calculated by multiplying the dominant emotion probability and emotion weight	Emotion shown in the facial expression which can be one of the seven categories: Angry, Disgust, Fear, Happy, Sad, Surprise or Neutral	Three levels of engagement: very engaged, nominally engaged and not engaged at all.	Real time

Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.					

Table 2.2: Summary of the student engagement prediction models

2.6. Visualization of Engagement Level

We reviewed papers concerning two issues: i. visualizing the student engagement levels from the prediction model or algorithm ii. visualizing student engagement level without using prediction models. We explain these two issues below.

2.6.1. Visualizing from the Prediction Model or Algorithm

[Liu et al., 2014] Liu, M., Calvo, R.A., Pardo, A. and Martin, A., 2014. Measuring and visualizing students' behavioral engagement in writing activities. IEEE Transactions on learning technologies, 8(2), pp.215-224.estimated student engagement through two algorithms they developed before visualizing it. The algorithms are point-based algorithm (pbA) and intensity-based algorithms (ibA). They implemented three types of visualizations: Point-Based visualization, Line-based visualization and Height-Based visualization. [Liu et al., 2014] Liu, M., Calvo, R.A., Pardo, A. and Martin, A., 2014. Measuring and visualizing students' behavioral engagement in writing activities. IEEE Transactions on learning technologies, 8(2), pp.215-224.used a different colour on height-based visualization to represent different tasks. They described a learning analytic system called Tracer, which derives behavioural engagement measures and creates visualizations of behavioural patterns of students writing on a cloud-based application and a novel learning analytic (LA) system that collects behavioural data of users writing, estimates the level of engagement. In [Liu et al., 2014] Liu, M., Calvo, R.A., Pardo, A. and Martin, A., 2014. Measuring and visualizing students' behavioral engagement in writing activities. IEEE Transactions on learning technologies, 8(2), pp.215-224., the minimum time they used for the interaction before visualizing engagement was two minutes. In our case, the visualizer that we built visualizes engagement levels every minute. [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.visualized low and high engagement levels after predicting the engagement levels using a model. They developed a dashboard to visualize the number of activities in VLE, individual student engagement in each assessment and the percentage of low and high-level engagement using a chart. [Coffin et al., 2014] Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G., 2014, March. Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 83-92).clustered students into subpopulation to gain additional insight. The subpopulations or the groups were: auditors, active and qualified students based on learner grade performance within the first weeks of the course. They were able to reveal a more detailed story of student engagement after they divided the students' related data into the groups. They implemented a histogram to visualize weekly student participation. They also applied a cumulative distribution plot which does not require a fixed number of bars, unlike histograms.

2.6.2. Visualizing without Prediction Models

[Carrillo et al., 2016] Carrillo, R., Lavoué, É. and Prié, Y., 2016, April. Towards qualitative insights for visualizing student engagement in web-based learning environments. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 893-898).visualized only indicators of engagement but not measured engagement from a prediction model. They used the number of logins on a learning application by time period, the number of times a learning resource was accessed, and the time spent on a learning document for behavioural indicators and cognitive indicators such as which node(s) or link(s) did s/he delete? when did s/he do it? how did s/he modify the structure of the mind map document? [Pesare et al., 2016.] Pesare, E., Roselli, T. and Rossano, V., 2016. Visualizing student engagement in e-learning environment. In 22th International Conference on Distributed Multimedia Systems (DMS) (pp. 26-33).did not apply student engagement prediction model to predict the engagement before visualizing it. [Pesare et al., 2016.] Pesare, E., Roselli, T. and Rossano, V., 2016. Visualizing student engagement in e-learning environment. In 22th International Conference on Distributed Multimedia Systems (DMS) (pp. 26-33).implemented scatterplot visualization to display trends and relationships in a cloud of points and a linear visualization, a time series to visualize details of a particular student in terms of trends and distribution. Moreover, a pie chart to display the distribution of interaction. [Ginda et al., 2019] Ginda, M., Richey, M.C., Cousino, M. and Börner, K., 2019. Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. PloS one, 14(5), p.e0215964.implemented a bar graph that shows the difference between an instructor's predictions and the average time learners spent on the same set of modules. In our work, we build on the visualizations from the works of [Liu et al., 2014] Liu, M., Calvo, R.A., Pardo, A. and Martin, A., 2014. Measuring and visualizing students' behavioral engagement in writing activities. IEEE Transactions on learning technologies, 8(2), pp.215-224. and [Pesare et al., 2016.] Pesare, E., Roselli, T. and Rossano, V., 2016. Visualizing student engagement in e-learning environment. In 22th International Conference on Distributed Multimedia Systems (DMS) (pp. 26-33)..

2.6.3. Limitations of the Existing Work

[Coffin et al., 2014] Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G., 2014, March. Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 83-92).applied a model and clustered students into three categories: auditors, active and qualified and visualized the outputs of the engagement prediction. However, they did not consider further classifying or filtering students in one of these categories. Moreover, we proposed to visualize the least engaged learner after classifying the student into the 4 classes of engagement levels using the model we built. After classifying a learner as very low or low, there may be a large number of students in the category of very low or low. Then, the task is to identify the least engaged learner from such larger numbers of students in this category. We have not seen any work that reported such a finding.

2.7. Chapter Summary

In this chapter, we have presented the exiting works that tried to detect student engagement from behavioural or interaction factors, from collaboration factors and from emotional factors. All of these works applied many features. [Sarsa and Escudero, 2016] Sarsa, J. and Escudero, T., 2016.A Roadmap to Cope with Common Problems in E-Learning Research Designs.*Electronic Journal of E-learning*, *14*(5), pp.336-349.remarked that the high number of features involved in e-learning processes complicates and masks the identification and isolation of the intervening factors. Most existing student engagement prediction models, which we reviewed, did not predict student engagement in smaller time scales such as 5 minutes. Moreover, all the papers reviewed neglected adding the element of collaborative engagement factor as part of the multifactor (behavioural, collaboration and emotional factors) even if it was explained by [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.which asserts that individuals' interactions with teachers or other students have been identified as key influencer of engagement. After classifying a learner's engagement as very low or low, there may be a large number of students in the category of very low or low. Then, the task is to identify the least engaged learner from such larger numbers of students in this category. We have not seen any work that reported such a finding.

In order to address the limitations, we performed empirical study to determine the significant features that affected a given level of engagement. The significant features can be applied in building the student engagement prediction model. We can visualize the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner. The detail works are described in subsequent chapters.

Chapter 3

Factors Affecting Student Engagement

3.1. Introduction

[Chao et al, 2008] Chao, Y.C.E., Zhao, Y., Kupper, L.L. and Nylander-French, L.A., 2008. Quantifying the relative importance of predictors in multiple linear regression analyses for public health studies.*Journal of occupational and environmental hygiene*, *5*(8), pp.519-529.implemented techniques for quantifying the relative importance of features for public health studies. We wanted to contribute by implementing one of these techniques particularly the Pratt's index to determine the relative importance of features in online student engagement detection area. We selected Pratt's index because it has been a useful and practical strategy. Moreover, Pratt's index requires simple computation and is easy to understand and interpret [Liu et al., 2014] Liu, Y., Zumbo, B.D. and Wu, A.D., 2014. Relative importance of predictors in multilevel modeling. *Journal of Modern Applied Statistical Methods, 13*(1), p.2.. We have done a work that reported empirical evidence for determining the most important factors whose features affected student engagement levels. We analysed the most important features from three factors: behavioural, collaboration and emotional factors in asynchronous online learning environment to determine the most important factors affecting student engagement levels. There were two research questions in this study. These were: Can we identify the relationship between features of the three factors which are behavioural, collaboration and emotional factors and student engagement levels, that is whether there is a positive relationship or negative relationship? Can we determine which ones are the most important features?

We set up an experiment for our objective. There were 12 participants in the experiment. The task involved interacting with learning activities in a LMS and interaction with facial emotion recognition tool. The interactions were saved as log files. After collecting the log file data, we analysed the log files to determine the most important features affecting student engagement levels from three factors: behavioural, collaboration and emotional factors in asynchronous online learning environment using Pearson correlation analysis and Pratt's index. The steps of the experiment have been depicted as a block diagram as follows in Figure 3.6:

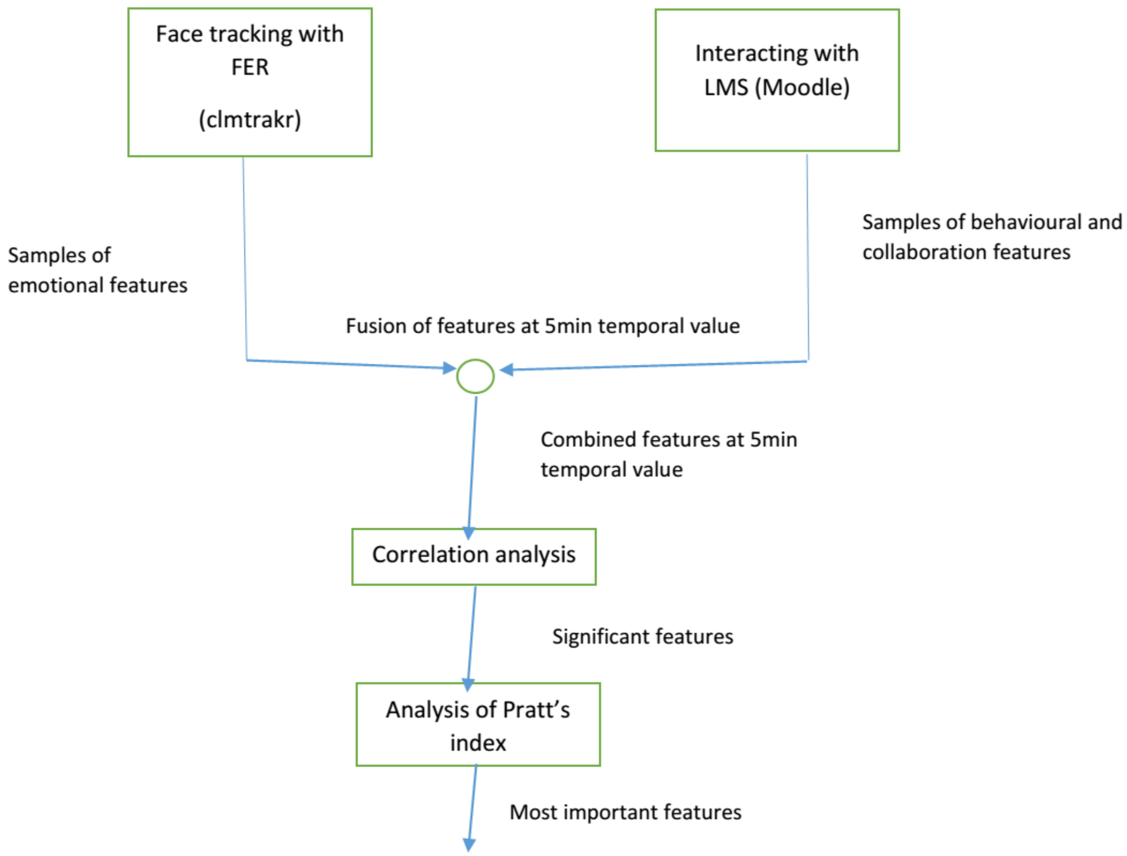


Figure 3.6: The steps of the experiment as a block diagram

In this chapter, we will describe data collection in detail. We will also present analysis of data. We will also discuss the results and implications of the analysis.

3.2. Data Collection Details

3.2.1. Experimental Setup

- Interaction with LMS

In our study, Moodle was chosen as a learning management system to allow students to interact with learning activities. It can be accessed free of charge. We accessed the zipped file from the URL: https://download.moodle.org/ . The version we installed was version 3.5 on Ubuntu 18.04. Its modular design makes it easy to create new courses. It also allows to create interactive course material such as assignments, lesson and quiz. Students can interact with each other through activities such as forum.

Moodle keeps detailed logs of all activities that students perform. It logs every click that students make for navigational purposes and has a log viewing system built into it. Log files can be filtered by course, participant, day and activity. For every task and participant, the log files of the interaction with the LMS are recorded automatically. Sample log file of the LMS (Figure3.7) is shown below. This log file was sampled in 5 minutes interval.

	Time	User full nan	Event context	Component	Event name
2	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	The status of the submission has been viewed.
3	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Course module viewed
4	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	A submission has been submitted.
5	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	File submissi	Submission updated.
6	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	File submissi	A file has been uploaded.
7	11/01/20, 03:25	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Course module viewed
8	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Submission form viewed.
9	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Course module viewed
10	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Submission form viewed.
11	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Course module viewed
12	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	The status of the submission has been viewed.
13	11/01/20, 03:24	Tufa Tufa	Assignment: Assignment to be worked throug	Assignment	Course module viewed
14	11/01/20, 03:24	Tufa Tufa	Course: Introduction to Descriptive Statistics	System	Course viewed
15	11/01/20, 03:21	Tufa Tufa	Forum: Discussion forum to do the assignmer	Forum	Post created
16	11/01/20, 03:21	Tufa Tufa	Forum: Discussion forum to do the assignmer	Forum	Some content has been posted.
17	11/01/20, 03:20	Tufa Tufa	Forum: Discussion forum to do the assignmer	Forum	Discussion viewed
18	11/01/20, 03:20	Tufa Tufa	Forum: Discussion forum to do the assignmer	Forum	Discussion viewed
19	11/01/20, 03:20	Tufa Tufa	Forum: Discussion forum to do the assignmer	Forum	Course module viewed
20	11/01/20, 03:20	Tufa Tufa	Course: Introduction to Descriptive Statistics	System	Course viewed
21	11/01/20, 03:18	Tufa Tufa	Quiz: Quiz2	Quiz	Quiz attempt reviewed
22	11/01/20, 03:18	Tufa Tufa	Quiz: Quiz2	Quiz	Quiz attempt submitted
23	11/01/20, 03:18	Tufa Tufa	Course: Introduction to Descriptive Statistics	System	User graded
24	11/01/20, 03:15	Tufa Tufa	Quiz: Quiz2	Quiz	Quiz attempt viewed

Figure 3.7: Log file saved in the database of moodle

- Interaction with Facial Emotion Recognition Tool (Clmtrakr)

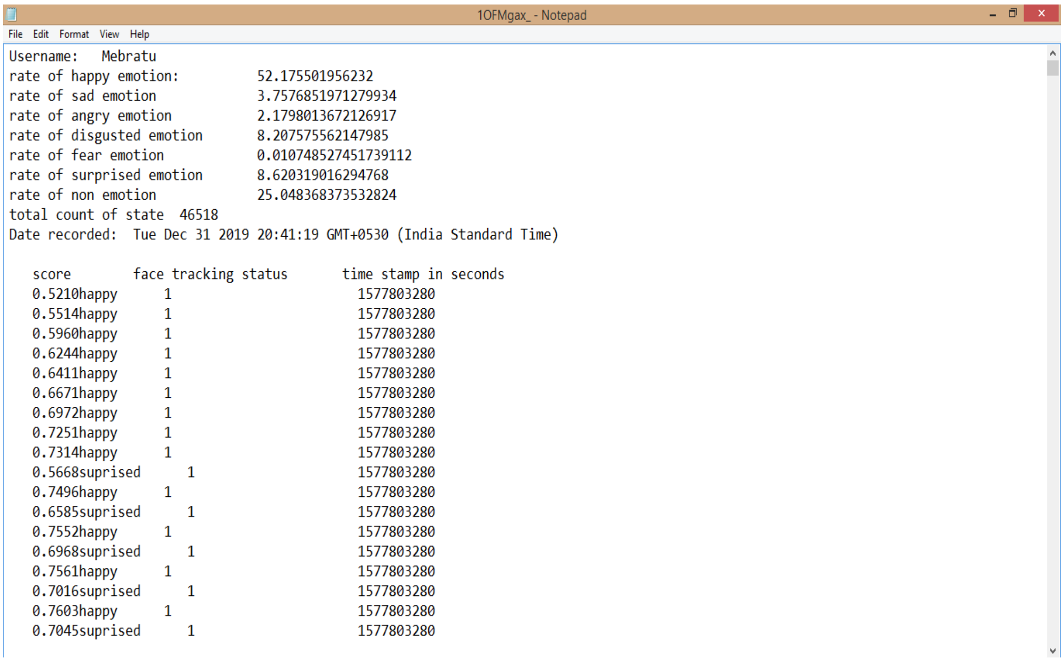
In our study, we applied a facial emotion recognition tool called clmtrackr. We chose clmtrackr as it is a software-based framework that can be integrated into web based learning environments. It performs all its processing on the user's computer without the need for a server infrastructure or additional browser plug-in. [Robal et al., 2018] [Khazan, (2014] Olga Khazan (2014) This App Reads Your Emotions on Your Face. The Atlantihttps://www.theatlantic.com/technology/archive/2014/01/this-app-reads-your-emotions-on-your-face/282993/ Accessed 25 November 2019.. It is an open source JavaScript face tracking library to detect engagement where no video frame or image will be captured. The emotion from the detected facial expression will be classified by the software in real time. This has two advantages: (i) it will make the system efficient as there is no transporting of video images to a server

[GhasemAghaei et al., 2016] GhasemAghaei, R., Arya, A. and Biddle, R., 2016, June. A dashboard for affective e-learning: data visualization for monitoring online learner emotions. In EdMedia+ Innovate Learning (pp. 1536-1543). Association for the Advancement of Computing in Education (AACE)., (ii) because there is no capturing of video image while classifying the facial emotion, it will not create discomfort on the user who worries that his/her privacy may be violated [Robal et al., 2018] . Clmtrackr fits facial models to faces in videos or images. It can be used for getting precise positions of facial features in an image, or precisely tracking faces in video. It also tracks a face and outputs the coordinate positions of the face model as an array, following the numbering of the model. The algorithm fits the facial model by using 70 small classifiers, i.e. one classifier for each point in the model. The models were trained on annotated data from the MUCT database plus some self-annotated images [Øygard, 2017] Øygard A. M.(2017), Released on Sep 12, 2017 on GitHub. https://github.com/auduno/clmtrackr.. MUCT stands for "Milborrow / University of Cape Town" [Milborrow et al., 2010] Milborrow, S., Morkel, J. and Nicolls, F., 2010. The MUCT landmarked face database. Pattern Recognition Association of South Africa, 201(0).. We accessed the source code from: https://github.com/auduno/clmtrackr. We installed version v1.1.2.

The clmtrackr tool that we used in the study to detect the actual emotion of users from facial expression analysis has the following features: It calculates the recognition rate of six basic emotions: disgust, angry, fear, sad, happy and surprise by getting the classification score when a facial expression on an image is classified as an emotion. It produces downloadable log file as comma separated value (csv) at the end of the session and offers dialog box in the browser window. It uses timer of elapsed time. The following figure (Figure 3.8) shows implementation of Clmtrackr to detect the rate of the six basic emotions.

Figure 3.8: Implementation of clmtrackr to detect the six basic emotions

The six basic emotional features were processed by the tool in every 5 minutes interval automatically and the log file was saved which was used as sample in 5 minutes interval. Sample log file of the rate of facial emotion (Figure 3.9.) is shown below:



```
1OFMgax_ - Notepad
File  Edit  Format  View  Help
Username:   Mebratu
rate of happy emotion:          52.175501956232
rate of sad emotion             3.7576851971279934
rate of angry emotion           2.1798013672126917
rate of disgusted emotion       8.207575562147985
rate of fear emotion            0.010748527451739112
rate of surprised emotion       8.620319016294768
rate of non emotion             25.048368373532824
total count of state  46518
Date recorded:  Tue Dec 31 2019 20:41:19 GMT+0530 (India Standard Time)

   score        face tracking status       time stamp in seconds
   0.5210happy      1                        1577803280
   0.5514happy      1                        1577803280
   0.5960happy      1                        1577803280
   0.6244happy      1                        1577803280
   0.6411happy      1                        1577803280
   0.6671happy      1                        1577803280
   0.6972happy      1                        1577803280
   0.7251happy      1                        1577803280
   0.7314happy      1                        1577803280
   0.5668suprised      1                     1577803280
   0.7496happy      1                        1577803280
   0.6585suprised      1                     1577803280
   0.7552happy      1                        1577803280
   0.6968suprised      1                     1577803280
   0.7561happy      1                        1577803280
   0.7016suprised      1                     1577803280
   0.7603happy      1                        1577803280
   0.7045suprised      1                     1577803280
```

Figure 3.9 : Rate of facial emotion and classification score downloaded as log file

Participants of the experiment accessed the installed and configured learning management system through networked computer equipped with webcam (Figure 3.10. below). Figure 3.10 below shows a student interacting on discussion forum while his face was tracked in real time.

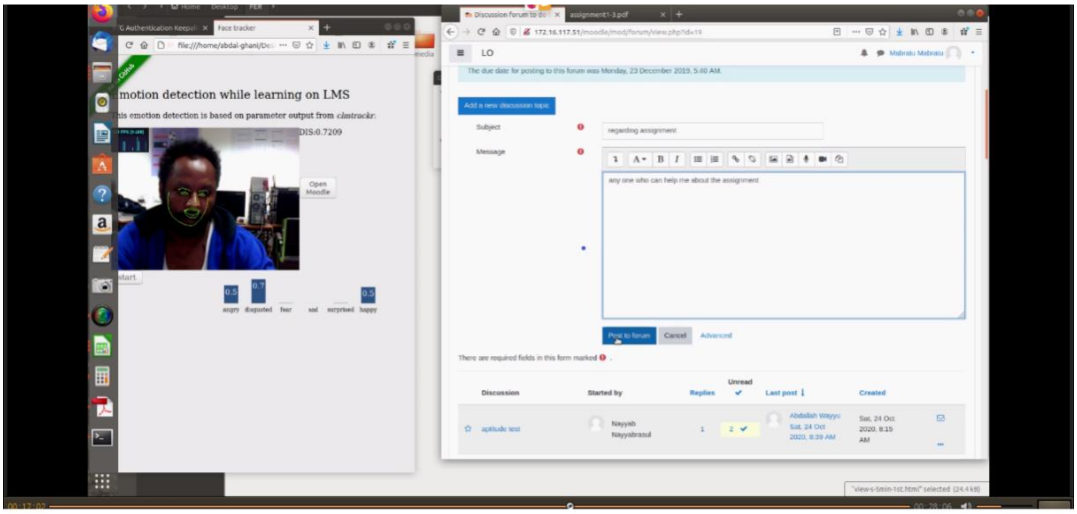


Figure 3.10: A Student interacting on discussion forum while his emotion was recognized from facial expression in real-time

Here, a student was sitting in front of a computer with a webcam and his/her facial expression was detected for emotion recognition in real time while performing the tasks explained in the next section. Engagement can manifest as internal affective states (e.g., interest, positive feelings about the task) [Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143.. The facial emotion recognition was performed to get emotional features. The emotional features were rates of the basic emotions which were: disgust (DIS), angry (ANG), fear (FEA), sad (SAD), surprise (SUR) and happy (HAP).

3.2.2. Task Design

Four tasks were designed and implemented as online tasks. These were content viewing, taking quiz, submitting assignment and posting in discussion forum. Content viewing, taking quiz, and submitting assignment were designed to detect behavioural engagement as explained by [Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143. that engagement can take the form of observable behaviour (e.g., participation in the learning activity, on-task behaviour). According to [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment

Scores.Computational intelligence and neuroscience, 2018., content viewing, discussion forum and quiz were significantly correlated with engagement. Assignment was the most used indicator of engagement according to [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789. We also applied posting to discussion forum to be used as one of the tasks to detect collaborative engagement. According to [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.individual interaction of learners with each other has been main influencer of engagement. Moreover, discussions that take place in face to face classrooms can be mimicked to provide a similar learning experience by using discussion boards or chat rooms using educational technology. The learning experience is enhanced because discussions are captured and can be reviewed later by students and instructors [Salazar, 2010] Salazar, J., 2010. Staying connected: Online education engagement and retention using educational technology tools. *American Society for Clinical Laboratory Science, 23*(3 Supplement), pp.53-58..

- Tasks for Capturing Data on Interaction or Behavioural Features

The tasks chosen for capturing the behavioural features (Table 3.3 below) were three: viewing content, taking quiz and submitting assignment. Features to be recorded as log file for engagement prediction in these tasks is explained in the next section.

- Tasks for Capturing Data on Collaboration Features

The task chosen for capturing the collaboration features was posting to a discussion forum. Features to be recorded as log file for engagement prediction in this task is explained in the next section.

Below is Table 3.3 that shows summary of tasks performed:

Tasks for Capturing Features	**Types of Features**	**Specific Feature**
Viewing content, Taking quiz and Submitting assignment	Behavioural features	Number of content view (NCV), time to read content (TRC), score (SC) of quiz and time to submit assignment (TA)
Posting to a discussion forum	Collaboration features	Time between posts and replies (TPR), time in the forum (TF) and number of replies (NR).

Table 3.3: Summary of tasks performed

3.2.3. Features

There are three factors with a total of 13 features as shown in Table 3.4.

- Behavioural Features

These are four features that are related to interaction with the three tasks: viewing content, taking quiz and submitting assignment as shown in Table 3.4. The features to be recorded as log file for engagement prediction in the task of viewing content were: number of times a student views the content (Number of content view (NCV)) and time to read content (TRC). The feature to be recorded as log file for engagement prediction in the task of taking quiz was: score (SC) of quiz. The feature to be recorded as log file for the engagement prediction in the task of submitting assignment was time to submit assignment (TA) which was calculated by subtracting download time of the file containing assignment question from the upload time of the file containing assignment answer.

- Collaboration Features

These are three features that are related to interaction of a learner with another learner through the task of posting in discussion forum [Rabbany et al., 2014] Rabbany, R., Elatia, S., Takaffoli, M. and Zaïane, O.R., 2014. Collaborative learning of students in online discussion forums: A social network analysis perspective. In Educational data mining (pp. 441-466).Springer, Cham. as shown in Table 3.4.The features to be recorded as log file for the engagement prediction were time between posts and replies (TPR), time in the forum (TF) and number of replies (NR). [Rabbany et al., 2014] Rabbany, R., Elatia, S., Takaffoli, M. and Zaïane, O.R., 2014. Collaborative learning of students in online discussion forums: A social network analysis perspective. In Educational data mining (pp. 441-466).Springer, Cham. remarked that discussion forums within Course Management Systems provide the basis for collaborative learning.

- Emotional Features

These are features that are related to the rates of the six basic emotions displayed such as disgust, anger, fear, sad, surprise and happy as shown in Table 3.4.

Type of feature	Feature name	Feature description
Behavioural features	Number of content view(NCV)	The number of content view that is stored in the log file categorized or sampled in 5 minutes
	Time of reading content(TRC)	The time spent measured in minutes while reading the content
	Time to submit assignment(TA)	Total time spent on assignment as calculated by subtracting the download time from upload time measured in minutes
	Score(SC)	Final mark a student obtained after taking a quiz
Collaborative feature	Number of replies(NR)	Number of messages sent by any participant to someone's post as replies (Number of replies one gets to his/her posts)
	Time between post and replies(TPR)	Time spent between someone's post and a reply to it measured in minutes
	Time in the forum(TF)	Total time used in the forum measured in minutes
Emotional feature	Disgust(DIS)	Rates of basic emotions as recorded through facial emotion recognition tool to measure levels of engagement. The log file of the rate of facial emotion was downloaded every 5 minutes automatically.
	Anger(ANG)	
	Fear(FEA)	
	Sad(SAD)	
	Surprise(SUR)	
	Happy(HAP)	

Table 3.4: Three types of features: behavioural features, collaboration features and emotional features

While the student is interacting on the "Moodle" LMS platform, the Facial Emotion Recognition tool need to run in parallel on each system. That is how we conducted our experiments also.

The synchronization issue between the Moodle and clmtrackr was addressed in such a way that the samples of the behavioural and collaboration features from the Moodle were recorded every 5 minutes. At the same time, the rate of facial emotions is downloaded at every 5 minute automatically which gives a sample of emotional features. Database queries were utilized to get the sample of the behavioural and collaboration features from the log file, which are frequencies of interactions (number of clicks) or time elapsed during interactions. The features were fussed at feature fusion level which requires temporal characteristics. In our study the temporal value was 5 minutes. It is therefore straightforward to apply correlation analysis.

3.2.4. Participants

The participants were 12 postgraduate students of Indian Institute of Technology Guwahati. Their average age was 33.6, with minimum age of 24 and maximum age of 39. There were 10 males and 2 females. The participants signed a consent form. The participants interacted with the learning management system (LMS) in two sessions, minimum of 15 minutes each. Their face was being tracked with facial emotion recognition tool. The following Table 3.5 summarizes the profile of the participants.

	Gender		Age			LMS experience	
	Male	Female	Min	Max	Average	Yes , I have	No, I don't have
Quantity	10	2	24	39	33.6	7	5

Table 3.5: Summary of the profile of the participants

3.2.5. Procedure

The participants were given instruction on what they will do, before starting interaction. They were working on desktop with webcam which were connected to the Internet. The experiment took place in User Centric Computing and Networking (UCCN) lab of Computer Science and Engineering (CSE) department. There were two sessions, each session took 15 minutes. They interacted with the four tasks explained above on the LMS while their facial expressions were being tracked. The log file of the interaction with the LMS was later extracted at the end of the two sessions. This log file was categorized in 5 minutes sample. At the same time, the log file of the rate of facial emotion was downloaded every 5 minutes. Each interaction of the participants in the experiment was recorded with screen recording software. The recorded video interaction was used for labelling the engagement levels of each participant for analysis purpose. The content of Descriptive Statistics with worked examples was used.

3.3. Analysis of Data

After collecting the data, we performed detailed analysis. We correlated the interactions data to four levels of engagement applying Pearson correlation analysis. We also calculated Pratt's index to determine features that are most important for a given level of engagement, which will indicate the most important factors. We also determined the significant features that affected a given level of engagement. We applied the significant features in building the student engagement prediction model, which will be discussed in chapter 4. The first step to determine the most important features as well as build the engagement model was to define engagement level, which is presented below.

3.3.1. Engagement Levels

In order to determine the most important features as well as build the engagement prediction model, we are proposing four levels of engagement. These are based on the works of [Whitehill et al., 2014] Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A. and Movellan, J.R., 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), pp.86-98. and [Kaur et al, 2018] Kaur, A., Mustafa, A., Mehta, L. and Dhall, A., 2018, December.Prediction and localization of student engagement in the wild. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE.. These levels are very high (VH) engagement level, high (H) engagement level, low (L) engagement level and very low (VL) engagement level. Student engagement is about students putting time, energy, thought, and effort and to some extent feelings into their learning [Dixson, 2015] Dixson, M.D., 2015. Measuring student engagement in the online course: The Online Student Engagement scale (OSE). *Online Learning*, *19*(4), p.n4.. The engagement levels are meanings to clearly understand what constitutes engagement while the student is interacting with the LMS. These levels can be predicted reliably from 10-second video clips [Whitehill et al., 2014] Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A. and Movellan, J.R., 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), pp.86-98.. When the student seems uninterested and looks away from the screen frequently, we classify that engagement as VERY LOW (VL). When the student is clearly not into the task and moves restlessly in the chair, we classify that engagement as LOW (L). When the student seems to like the content, and requires no admonition to stay in the task, we classify that engagement as HIGH (H). When the student stared at the screen and was focused, and could be highly praised for his/her level of engagement in the task, we classify that engagement as

VERY HIGH (VH). In the next section, the experimental detail was presented which was implemented to collect features for the model building.

The categories of the engagement for labelling the recorded video of interaction are shown below (Table 3.6). The labelling task was performed by viewing the video frame of 10 seconds length and giving number to rate each clip or video frame.

Engagement intensity	**Meaning**	**Point given**
VERY LOW (VL)	Not engaged at all	1
LOW (L)	Barely engaged	2
HIGH (H)	Engaged in the content	3
VERY HIGH (VH)	Very engaged	4

Table 3.6: Categories of the engagement for labelling the recorded video of interaction

3.3.2. Pre-processing

For each participant, we captured 30 minutes of video, recording the interaction with the LMS. The last 5 minutes of the 30 minutes video recorded was not uniform for all participants. This is due to the fact that some participants took less than 5 minutes to complete the activities. Accordingly, labelling the screen recorded interaction video for four engagement levels was done only for the first 25 minutes. Engagement labels of 10-second video clips can be reliably predicted [Whitehill et al., 2014] Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A. and Movellan, J.R., 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), pp.86-98.. The labelling was done in 10 seconds interval (frame) as one of the four engagement levels by human observer. Here we assumed that student behaviour remains consistent between each of these intervals.

There were 30 samples or frames of 10 seconds in 5 minutes interaction. So, in our labelling task, we used the frames in the first 25 minutes (5 samples of 5 minutes each) of the recorded video. Thus, when we counted the points given, each level had a chance to get points or scales between 0-30 in 5 minutes interval. Table 3.7 shows points for 4 levels of engagement for a single participant:

Participant	Time Interval	Engagement Levels			
		ENG -VL	ENG -L	ENG -H	ENG -VH
Participant 1	0-5min	0	0	5	25
	5-10min	0	0	0	30
	10-15min	0	0	2	28
	15-20min	0	0	9	21
	20-25min	0	0	25	5

Table 3.7: The Four levels of engagement data for a single participant.

For example, 25 in this table (1^{st} row and 4^{th} column) indicate 25 frames where this participant has been found to be at very high levels of engagement (ENG-VH) in that particular time interval

We calculated the average of the levels of engagement for each participant for the whole 25 minutes as shown in Table 3.8.

Participant	Time Interval	ENG-VL	ENG-L	ENG-H	ENG-VH
Participant 1	0-25min	0	0	8.2	21.8
Participant 2	0-25min	0	0	0.4	29.6
Participant 3	0-25min	0.4	0	6.8	22.8
Participant 4	0-25min	0	0.6	1.4	28
Participant 5	0-25min	0	0.4	5.6	24
Participant 6	0-25min	0	0.2	1.4	28.4
Participant 7	0-25min	0	1.2	8.4	20.4
Participant 8	0-25min	0.4	1.2	2.8	25.6
Participant 9	0-25min	0	0	4.4	25.6
Participant 10	0-25min	0	0	0.8	29.2
Participant 11	0-25min	0.6	2.2	4	23.2
Participant 12	0-25min	0	0	1.6	28.4

Table 3.8: Average of the four levels of engagement labelled for all participants

Corresponding to the labelling, categorizing the log file in sample of 5 minutes was done for the three factors consisting of 13 features. For each participant, and for each of the 13 features in the three factors, we would have 5 samples of 5 minutes length in 25 minutes long interaction. Table 3.9 shows the 5 samples of the three factors for a single participant:

Participant	Time Interval	Feature						
Participant 1	0-5 min	Collaboration	**NR**	**TPR**	**TF**			
			0	0	0			
		Behavioural	**TA**	**NCV**	**SC**	**TRC**		
			0	13	1.43	4		
		Emotional	**ANG**	**DIS**	**FEA**	**SAD**	**SUR**	**HAP**
			10.3	59.67	1.16	6.15	1.16	9.88
	5-10 min	Collaboration	**NR**	**TPR**	**TF**			
			0	0	2			
		Behavioural	**TA**	**NCV**	**SC**	**TRC**		
			0	7	0	2		
		Emotional	**ANG**	**DIS**	**FEA**	**SAD**	**SUR**	**HAP**
			16.36	54.17	0.35	2.26	1.14	4.44
	10-15 min	Collaboration	**NR**	**TPR**	**TF**			
			0	0	0			
		Behavioural	**TA**	**NCV**	**SC**	**TRC**		
			0	4	7.14	1		
		Emotional	**ANG**	**DIS**	**FEA**	**SAD**	**SUR**	**HAP**
			5.81	18.19	0.25	10.5	16.54	36.64
	15-20min	Collaboration	**NR**	**TPR**	**TF**			
			0	0	0			
		Behavioural	**TA**	**NCV**	**SC**	**TRC**		
			0	4	0	3		
		Emotional	**ANG**	**DIS**	**FEA**	**SAD**	**SUR**	**HAP**
			5.5	17.28	0.24	10.12	16.94	38.49
	20-25min	Collaboration	**NR**	**TPR**	**TF**			
			1	8689	0			
		Behavioural	**TA**	**NCV**	**SC**	**TRC**		
			0	4	8.57	2		
		Emotional	**ANG**	**DIS**	**FEA**	**SAD**	**SUR**	**HAP**
			4.48	14.2	0.2	9.32	17.81	38.05

Table 3.9: 5 Samples of 5 minutes length of the collaboration features, behavioural features and emotional features for one participant.

The captions for Table 3.9 are put as: NR=Number of Replies, TPR=Time between Post and Replies, TF=Time in the Forum, TA=Time of Assignment Submission, NCV=Number of Content View, SC=Score of quiz, TRC=Time to Read Content, ANG=Anger, DIS=Disgust, FEA=Fear, SAD=sad, SUR=Surprise, HAP=Happy

After averaging the samples for each participant, in each factor, the results we obtained were shown in Table 3.10a,b and c.

Participant	NR	TPR	TF
Participant 1	0.2	1737.8	0.4
Participant 2	0.8	5710.8	1
Participant 3	0.2	1730.4	1
Participant 4	0.6	2916.4	1
Participant 5	0.4	2883.4	1.6
Participant 6	0.2	512.6	2.4
Participant 7	0	0	0.4
Participant 8	0.4	1044	1.6
Participant 9	0.4	515.6	1.4
Participant 10	0.4	5774.8	0.4
Participant 11	0.6	824.6	1.4
Participant 12	0.4	1665	0.8

Table 3.10a: Average of 5 samples of 5 minutes length of the collaboration features for all participants

(TF and TPR were measured in minutes)

Participant	TA	NCV	SC	TRC
Participant 1	0	6.4	3.428	2.4
Participant 2	4303.6	3.4	3.142	2
Participant 3	926.2	9.2	3.428	2.6
Participant 4	1214.2	5.8	5.428	3
Participant 5	4343.4	4.6	3.428	2.4
Participant 6	2402.2	3.8	5.428	1.8
Participant 7	2306.6	6.8	5.428	3.4
Participant 8	0	3	4	2
Participant 9	4061.8	3	2.286	2
Participant 10	2390.4	4.8	1.142	2.8
Participant 11	1211	4	5.142	3.4
Participant 12	4639	3.8	3.142	2.2

Table 3.10b: Average of 5 samples of 5 minutes' length of the behavioural features for all participants (TA and TRC were measured in minutes)

Participant	ANG	DIS	FEA	SAD	SUR	HAP
Participant 1	8.49	32.702	0.44	7.67	10.718	25.5
Participant 2	3.692	18.62	13.976	4.842	9.198	47.602
Participant 3	0.47	0.448	0.03	2.914	4.954	33.124
Participant 4	2.536	9.286	0.02	4.642	7.978	47.7166
Participant 5	2.22	2.828	19.538	11.44	28.696	15.868
Participant 6	6.978	7.29	0.002	0.13	7.47	76.854
Participant 7	1.738	6.064	1.548	20.796	60.158	9.56
Participant 8	15.22	16.07	2.992	13.89	25.888	25.112
Participant 9	3.888	7.534	13.998	46.29	5.614	22.598
Participant 10	38.084	43.442	0.004	1.768	2.912	13.612
Participant 11	1.542	5.432	0.03	3.818	45.128	43.382
Participant 12	0.196	0.868	0	1.288	4.428	93.206

Table 3.10c: Average of 5 samples of 5 minutes length of emotional features for all participants

3.3.3. Correlation Analysis

The correlation coefficient, denoted by r, is a measure of the strength of the straight-line or linear relationship between two variables. The correlation coefficient takes on values ranging between +1 and –1 [Bruce Ratner, 2003] Bruce Ratner - Statistical Modeling and Analysis for Database Marketing_ Effective Techniques for Mining Big Data-Chapman and Hall_CRC (2003). Pearson correlation coefficient was used in our study. We assumed values between 0.3 and 0.7 (-0.3 and − 0.7) indicate a moderate positive (negative) linear relationship. Values between 0.7 and 1.0 (− 0.7 and − 1.0) indicate a strong positive (negative) linear relationship as reported in [Bruce Ratner, 2003] Bruce Ratner - Statistical Modeling and Analysis for Database Marketing_ Effective Techniques for Mining Big Data-Chapman and Hall_CRC (2003). We used MS Excel 2010 for the analysis.

The correlation was computed between very low (VL) engagement (ENG-VL) levels, low (L) engagement levels (ENG-L**)**, high (H) engagement levels (ENG-H) and very high (VH) engagement levels (ENG-VH) as shown in Table 3.8, and features in Table 3.10a, b and c. The following table (Table 3.11) shows the correlation analysis result.

		ENGAGEMENT LEVELS			
		ENG-VL	ENG-L	ENG-H	ENG-VH
Collaboration features	**NR**	-0.12	0.08	-0.68	0.59
	TPR	-0.29	-0.41	-0.52	0.59
	TF	0.22	0.15	-0.30	0.23
Behavioural features	***TA***	-0.53	-0.38	-0.32	0.42
	NCV	0.10	-0.12	0.62	-0.55
	SC	0.23	0.60	0.16	-0.30
	TRC	0.29	0.60	0.35	-0.48
Emotional features	***ANG***	-0.11	-0.16	-0.34	0.35
	DIS	-0.24	-0.25	-0.17	0.23
	FEA	-0.29	-0.22	0.01	0.06
	SAD	-0.16	-0.00	0.33	-0.29
	SUR	0.33	0.82	0.49	-0.66
	HAP	-0.06	-0.16	-0.53	0.53

Table 3.11: The Correlation analysis result of very low (VL), low (L), high (H) and very high (VH) engagement level

Table 3.12 below summarizes the significant features which were identified from the correlation analysis result (found to indicate a moderate or strong relationship) with respect to the engagement levels.

Engagement level	Type of factor	Significant feature	Total number of significant features
Very low (VL)	Collaboration		2
	Behavioural	Time of assignment submission (TA)	
	Emotional	Surprise (SUR)	
Low (L)	Collaboration	Time between post and reply(TPR)	5
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Score of quiz (SC)	
	Emotional	Surprise (SUR)	
High (H)	Collaboration	Number of replies(NR), Time between post and reply(TPR) and Time in the forum (TF)	10
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Number of content view (NCV)	
	Emotional	Anger (ANG), Sad emotion(SAD), Surprise (SUR) and Happy emotion(HAP)	
Very high (VH)	Collaboration	Number of replies(NR), Time between post and reply(TPR)	9
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Number of content view (NCV) and Score of quiz (SC)	
	Emotional	Anger (ANG), Surprise (SUR) and Happy emotion(HAP)	

Table 3.12: Summary of the significant features

It may be observed from Table 3.11, based on the rule in [Bruce Ratner, 2003] Bruce Ratner - Statistical Modeling and Analysis for Database Marketing_ Effective Techniques for Mining Big Data-Chapman and Hall_CRC (2003), we found that Time of assignment submission (TA) to be significantly and negatively and SUR (Surprise) to be significantly and positively correlated with very low (VL) level of engagement. No collaboration feature was found to be significantly correlated with very low (VL) level of engagement. Time of assignment submission (TA), Time between post and reply (TPR) were found to be significantly and negatively correlated, Time of reading content (TRC), score (SC) and surprise (SUR) emotions significantly and positively correlated with low (L) level of engagement. We also found the three collaboration features to be significantly and negatively correlated with high (H) level of engagement. Time of assignment submission (TA) was behavioural feature, anger (ANG), and happy (HAP) were emotional features to be significantly and negatively correlated with high (H) level of engagement. Time of reading content (TRC), number of content view (NCV), sad (SAD) and surprise (SUR) were found to be significantly and positively correlated with high (H) level of engagement. We found number of replies (NR), Time between post and reply (TPR), Time of assignment submission (TA) anger (ANG) and happy (HAP) to be significantly and positively correlated with very high (VH) level of engagement. number of content view (NCV), Time to read content (TRC), score (SC), and surprise(SUR) were found to be significantly and negatively correlated with very high (VH) level of engagement.

3.4. Post Analysis using Pratt's Index

The Product Measure $B_j r_{yxj}$ quantifies the relative importance of a predictor (x_j) as the product of its standardized regression coefficient (B_j) in the full model with p predictors and its zero-order correlation (r^2_{yxj}) with the dependent variable y. The equation to calculate Product Measure is given by Eq.(1) as follows:

$$\mathbf{B_j r_{yxj} = (B_j)(r^2_{yxj})} \qquad (1)$$

Where B_j is standardized regression coefficient for p predictors and r^2_{yxj} is zero-order correlation of p predictors with the dependent variable y. In our study, the p predictors are the features and the dependent variable y is the level of engagement. Product measure (or Pratt's Index) $B_j r_{yxj}$ has been advocated as producing meaningful measures of relative importance when the predictors are mutually correlated [Chao et al, 2008] Chao, Y.C.E., Zhao, Y., Kupper, L.L. and Nylander-French, L.A., 2008. Quantifying the relative importance of predictors in multiple linear regression analyses for public health studies.*Journal of occupational and environmental hygiene*, *5*(8), pp.519-529.. A theoretical discussion of its use in multiple linear regressions was given by Pratt, and $B_j r_{yxj}$ is also called the Pratt's Index by many researchers [Chao et al, 2008] Chao, Y.C.E., Zhao, Y., Kupper, L.L. and Nylander-French, L.A., 2008. Quantifying the relative importance of predictors in multiple linear regression analyses for public health studies.*Journal of occupational and environmental hygiene*, *5*(8), pp.519-529.. The larger the value, the bigger the impression of that feature on the output.

In our experiment, we applied Pratt's index to the correlational analysis results (Table 3.11). We wanted to determine which feature, for each level, is the most important based on the extent to which each feature contributes to the prediction of the dependent variable (student engagement level). We first computed the mutual correlation of those features whose relative importance is to be calculated. Then, we calculated the Pratt's index. We did these for each level of engagement. We started with very low level (VL) of engagement. Behavioural factor has also a single feature (TA) affecting very low level of engagement. Emotional factor has one feature: SUR affecting very low level of engagement. Collaboration factor has no any feature affecting very low level of engagement. Thus, we wanted to determine which of the two features is the most important. To do this, we first calculated the mutual correlation coefficients of these two features as below in Table 3.13.

	TA	*SUR*
TA	1	
SUR	-0.15531	1

Table 3.13: The Mutual correlation coefficient of two features affecting very low (VL) level of engagement.

Similarly, we first calculated the mutual correlation coefficients of five features affecting low level of engagement (L) as shown in Table 3.14.

	TPR	*TA*	*SC*	*TRC*	*SUR*
TPR	1				
TA	0.264314	1			
SC	-0.55907	-0.34108	1		
TRC	-0.07332	-0.30334	0.33643	1	
SUR	-0.43231	-0.15531	0.541772	0.619425	1

Table 3.14: The Mutual correlation coefficients of five features affecting low (L) engagement level

Then, we first calculated the mutual correlation coefficients of ten features affecting high level of engagement (H) as shown in Table 3.15.

	NR	*TPR*	*TF*	*TA*	*NCV*	*TRC*	*ANG*	*SAD*	*SUR*	*HAP*
NR	1									
TPR	0.58141	1								
TF	0.128225	-0.33067	1							
TA	0.267596	0.264314	0.097381	1						
NCV	-0.52012	-0.07694	-0.46648	-0.42232	1					
TRC	-0.10689	-0.07332	-0.49112	-0.30334	0.479258	1				
ANG	-0.00681	0.507847	-0.23565	-0.18952	-0.15562	-0.02678	1			
SAD	-0.15696	-0.38689	0.036146	0.220485	-0.22884	-0.12285	-0.14757	1		
SUR	-0.2457	-0.43231	-0.05341	-0.15531	0.069183	0.619425	-0.25189	0.166637	1	
HAP	0.227282	-0.09754	0.340979	0.263819	-0.27627	-0.35227	-0.34217	-0.46067	-0.38633	1

Table 3.15: The Mutual correlation coefficients of ten features affecting high level of engagement

And we first calculated the mutual correlation coefficients of nine features affecting very high level of engagement (VH) as shown in Table 3.16.

	NR	*TPR*	*TA*	*NCV*	*SC*	*TRC*	*ANG*	*SUR*	*HAP*
NR	1								
TPR	0.58141	1							
TA	0.267596	0.264314	1						
NCV	-0.52012	-0.07694	-0.42232	1					
SC	-0.15396	-0.55907	-0.34108	0.143077	1				
TRC	-0.10689	-0.07332	-0.30334	0.479258	0.33643	1			
ANG	-0.00681	0.507847	-0.18952	-0.15562	-0.55878	-0.02678	1		
SUR	-0.2457	-0.43231	-0.15531	0.069183	0.541772	0.619425	-0.25189	1	
HAP	0.227282	-0.09754	0.263819	-0.27627	0.253585	-0.35227	-0.34217	-0.38633	1

Table 3.16: The Mutual correlation coefficients of nine features affecting very high level of engagement (VH)

After calculating the mutual correlation coefficients, we computed the Pratt's index. We put their correlation coefficient, standard regression coefficient and the product of both (Pratt's index) for the four levels of engagement.

There were two features to be compared to determine which one was the most important affecting very low (VL) level of engagement as shown in Table 3.17:

Features	**Correlation coefficients**	**Standardized regression Coefficients**	**Pratt's index**
TA	-0.531018465	-0.491661742	0.261081463
SUR	0.329763821	0.253402245	0.083562892

Table 3.17: Correlation coefficients, standard regression coefficients and Pratt's indices of two features affecting very low (VL) level of engagement

From Table 3.17, TA was with the highest Pratt's index. Thus, TA which is a behavioural feature was the most important feature affecting very low (VL) engagement level.

Next, we applied the same step for identifying the most important feature affecting low level of engagement (L) as shown in Table 3.18. Here, we compared five feature of which three features were from behavioural factors (TA, TRC, SC) and one feature were from emotional factors (SUR) and one feature from collaboration factor (TPR) affecting low level of engagement (L).

Features	Correlation coefficients	Standardized regression Coefficients	Pratt's index
TPR	-0.414028395	0.019535623	-0.008088303
TA	-0.384366355	-0.207186657	0.07963558
SC	0.603720497	0.161885706	0.097733719
TRC	0.605051734	0.085024141	0.051444004
SUR	0.816563279	0.652458606	0.532773739

Table 3.18: Correlation coefficients, standard regression coefficients and Pratt's indices of five features affecting low (L) engagement level

It is observed from Table 3.18 that SUR is with the highest value of Pratt's index. Hence, SUR is the most important feature affecting low level of engagement.

Our next task was to determine the most important feature for the high level of engagement (H) as shown in Table 3.19. Here, we compared ten features of which four were emotional features (ANG, SAD, SUR and HAP) and three were collaboration features (TPR, NR and TF) and three were behavioural features (TA, NCV and TRC) to determine the most important feature affecting high level (H) engagement.

Features	Correlation coefficients	Standardized regression Coefficients	Pratt's index
NR	-0.67592279	-0.253072009	0.171057138
TPR	-0.520521546	-1.031165042	0.536743622
TF	-0.302549004	-0.504666967	0.152686488
TA	-0.321772115	0.264468368	-0.085098546
NCV	0.622208124	-0.416001047	-0.258839231
TRC	0.353098413	-0.04439475	-0.015675716
ANG	-0.336395221	-0.760585708	0.255857398
SAD	0.330538559	-0.972103807	-0.321317792
SUR	0.489212962	-0.570001029	-0.278851892
HAP	-0.532934782	-1.532534457	0.816740917

Table 3.19: Correlation coefficients, standard regression coefficients and Pratt's indices of ten features affecting high level (H) engagement

Happy (HAP) was with the highest value of Pratt's index as shown in Table 3.19. Hence, happy (HAP) was the most important feature affecting high level of engagement.

There were nine features to be compared to determine which one was the most important feature affecting very high level of engagement as shown in Table 3.20. The nine features were: TPR and NR, which are collaboration features, TRC, NCV and SC, which are behavioural features and three emotional features (ANG, SUR and HAP) affecting very high level of engagement (VH).

Features	Correlation coefficients	Standardized regression Coefficients	Pratt's index
NR	0.593835274	0.727270355	0.43187879
TPR	0.591543282	-0.295611957	-0.174867267
TA	0.418700473	0.659575402	0.276164533
NCV	-0.551922229	0.390817362	-0.21570079
SC	-0.302621121	0.59623226	-0.180432475
TRC	-0.480492665	-0.307064478	0.147542229
ANG	0.35325218	1.02227991	0.361122606
SUR	-0.655829459	-0.312397432	0.204879439
HAP	0.529729058	0.239300677	0.126764522

Table 3.20: Correlation coefficients, standard regression coefficients and Pratt's indices of nine features affecting very high level of engagement (VH)

We noted from Table 3.20 that Number of replies (NR) was with the highest value of Pratt's index. Hence, Number of replies (NR) was the most important feature affecting very high level of student engagement.

3.5. Discussion

The finding of the analysis indicated that from 13 features, only 11 were significant for the four levels of student engagement. From the 11 significant features, only 4 were found to be the most important features. Two of the emotional features: disgust (DIS), and fear (FEA) were never correlated with any level of student engagement. The collaboration feature which was Time between post and reply (TPR) correlated with low, but not with very low levels of engagement.

In our study, we found two features that were correlated significantly with students' very low level of engagement (VL): Time of assignment submission (TA) and surprise (SUR). Time of assignment submission (TA) was a behavioural feature that affected very low engagement negatively. As an asynchronous online learner keeps doing the assignment for longer time, his disengagement decreases very highly and the time of assignment submission was the most important feature to predict the very low engagement level.

We also found that five features were significant for predicting students' low level of engagement (L). These were the Time of assignment submission (TA), Number of content view (NCV), Score of quiz (SC), Surprise emotion (SUR) and Time between post and reply (TPR). Surprise emotion (SUR) was the most important feature that affected low level of engagement. Surprise emotion (SUR) affected low level of engagement positively. As an asynchronous online learner felt surprise emotion (SUR) for longer time, his/her disengagement increases highly. In the attention level, positive affect seems to reduce resources available for effortful processing [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.. The surprise emotion was the most important feature to predict the low engagement level.

We found ten features to affect high level of student engagement (H) significantly. These were: The Number of replies (NR), Time between post and reply (TPR), Time in the forum (TF), anger (ANG) emotion, surprise (SUR) emotion, sad (SAD), happy (HAP) emotion, Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV). The happy (HAP) feature in our study was found to affect high level of engagement negatively. As an asynchronous online learner kept feeling happy emotion for longer time, his engagement decreases highly. This was confirmed by [Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R. and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA. in that positive affect leads to behavioural disengagement because it signals that all is well and there is no need to engage. The happy emotion was the most important feature to predict the high engagement level. However, there was no work in the literature that reported happy (HAP) was the most important emotional feature that affected high level of student engagement when compared with other three emotional features such as sadness (SAD), anger (ANG) and surprise (SUR) emotions and other collaboration and behavioural features.

Nine features were found to affect very high level of engagement (VH) significantly. These were Number of replies (NR), Time between post and reply (TPR), Time in the forum (TF), anger (ANG) emotion, surprise (SUR) emotion, happy (HAP) emotion, Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV). Number of replies (NR) affected very high level of engagement positively. As an asynchronous online learner keeps getting many replies, his engagement increases very highly. The Number of replies (NR) was the most important feature to predict the very high engagement level.

In our study surprise (SUR) emotion was related to very high engagement level. Surprise (SUR) emotion was also mapped to strong engagement according to [Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement.

Journal of King Saud University-Computer and Information Sciences.. However, we found out that surprise emotion affected student engagement very highly and negatively. This negative relationship may be due to the fact that in the attention level, positive affect seems to reduce resources available for effortful processing [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.. Happy (HAP) was emotional feature that affected very high level of engagement positively. This was confirmed by [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press. who stated that positive emotions may broaden the scopes of attention, cognition, and action, widening the array of percept, thoughts, and actions.

Other contributions of our study was the finding that the most important feature to affect very low level of engagement was the Time of assignment submission (TA) which was a behavioural feature when compared to emotional feature that was surprise (SUR). The most important feature to affect low levels of student engagement was surprise (SUR) which was emotional feature when compared with another collaboration feature that was time between post and reply (TPR) and three behavioural features which were time of assignment submission (TA), time of reading content (TRC) and score of quiz (SC). Happy (HAP) was the most important emotional feature that affected high level of student engagement when compared with other three emotional features such as sadness (SAD), anger (ANG) and surprise (SUR) emotions and other collaboration and behavioural features. The most important feature to affect very high level of student engagement was number of replies (NR) which was collaboration feature when compared with other emotional features that were anger emotion (ANG), surprise (SUR) and happy emotion (HAP and four behavioural features which were Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV) and score of quiz (SC), and one collaboration feature which was Time between post and reply (TPR). Moreover, the collaboration features which were Time between post and reply (TPR) and number of replies (NR) were correlated with high and very high levels of engagement. The collaboration feature which was time in the forum (TF) correlated with high, but not with low, not with very low and not with very high levels of engagement. This has implication that it confirms that individual interaction of learners with each other has been main influencer of engagement [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.. Furthermore, the previously unknown relationships between the features such as number of replies to someone's post and the time between someone's post and replies he/she got and engagement levels as reported by [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20). are now known.

We found that anger (ANG) was emotional feature that affected high level of engagement. This is confirmed by [Li et al., 2016] Li, J., Ngai, G., Leong, H.V. and Chan, S.C., 2016. Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics.ACM SIGAPP Applied Computing Review, 16(3), pp.37-49.who stated that anger is a measure of high level of attention. The anger (ANG) feature in our study was found to be inversely related to high level of engagement. This finding confirmed the finding in [Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R. and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA.that achievement-related anger correlated negatively with academic performance.

We also found out that sad affected high level of engagement positively. Similar result was reported by [Teimouri, 2018] Teimouri, Y., 2018. Differential roles of shame and guilt in L2 learning: How bad is bad?. The Modern Language Journal, 102(4), pp.632-652.. Embarrassment or guilt, anxiety, fear, nervousness, sadness, and shame are negative emotions that are named as language anxiety. Guilt had positive effects on the motivation and language achievements of second language learners [Teimouri, 2018] Teimouri, Y., 2018. Differential roles of shame and guilt in L2 learning: How bad is bad?. The Modern Language Journal, 102(4), pp.632-652..

Time to read content (TRC) is a behavioural feature that affected low and very high level of engagement. As Time to read content (TRC) increases, student disengagement increases highly. Similar result was also reported by [Cocea and

Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.that long time spent on the same page was associated with disengagement. Another behavioural feature affecting low and very high engagement was score of quiz. As score increases, student disengagement increases highly. This was unexpected result, but [Woolf et al., 2009] Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. and Picard, R., 2009. Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), pp.129-164.reported that when problems are easy, a student gets bored.

The following Table 3.21 summarizes the most important features.

Engagement level	Type of factor	Significant feature	Total number of significant features	Most important feature
Very low (VL)	Collaboration		2	Time of assignment submission (TA)
	Behavioural	Time of assignment submission (TA)		
	Emotional	Surprise (SUR)		
Low (L)	Collaboration	Time between post and reply(TPR)	5	Surprise (SUR)
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Score of quiz (SC)		
	Emotional	Surprise (SUR)		
High (H)	Collaboration	Number of replies(NR), Time between post and reply(TPR) and Time in the forum (TF)	10	Happy emotion(HAP)
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Number of content view (NCV)		
	Emotional	Anger (ANG), Sad emotion(SAD), Surprise (SUR) and Happy emotion(HAP)		
Very high (VH)	Collaboration	Number of replies(NR), Time between post and reply(TPR)	9	Number of replies(NR)
	Behavioural	Time of assignment submission (TA), Time to read content (TRC), Number of content view (NCV) and Score of quiz (SC)		
	Emotional	Anger (ANG), Surprise (SUR) and Happy emotion(HAP)		

Table 3.21: Summary of the most important features

One of the expected outcomes of our study was that it is of great importance to implement a prediction model that is effective and economical [Whittaker et al., 2002] Whittaker, T.A., Fouladi, R.T. and Williams, N.J., 2002. Determining predictor importance in multiple regression under varied correlational and distributional conditions. *Journal of Modern Applied Statistical Methods*, *1*(2), p.44.. For implementing a prediction model that is effective and economical, determining the relative importance of the predictors is important [Whittaker et al., 2002] Whittaker, T.A., Fouladi, R.T. and Williams, N.J., 2002. Determining predictor importance in multiple regression under varied correlational and distributional conditions. *Journal of Modern Applied Statistical Methods*, *1*(2), p.44.. Similarly, for effectiveness and economic benefits, it is important to determine the relative importance of features that are used in student engagement prediction models. The other outcomes would be that the findings of this study enables the e-learning designers to be aware of the features of the factors affecting student engagement that need to be altered to increase student engagement and learning [Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143.. In general, the outcome of this study is important as a critical aspect of many intervention efforts aimed at reducing dropout rates through increasing student engagement [Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143..

[Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143. remarked that researchers must specify the dimensions of engagement and ensure that their measures align properly with these descriptions of engagement. Moreover, [Henrie et al., 2015] Henrie, C.R., Halverson, L.R. and Graham, C.R., 2015. Measuring student engagement in technology-mediated learning: A review.*Computers& Education*, *90*, pp.36-53.reported that the term engagement must include multiple components to ensure that the richness of real human experience is understood. We depended on the definition of engagement by Redmond et.al.. Accordingly, we dealt with three factors of engagement which were behavioural, collaboration and emotional engagement. Behavioural and collaboration were measured through interactions with the learning activities in LMS using mouse clicking modality. The features were measured by counting the frequencies of interactions and time taken for the interactions through log file analysis. 7 features were extracted through this modality. The emotional engagement was measured through facial emotion recognition using face tracking tool. Here also 6 features were extracted through this modality. The total 13 features from the above modalities were combined at the feature fusion level. Feature fusion requires temporal characteristics to be similar. Accordingly, we made the temporal characteristic to be of 5 minutes to fuse the features of the modalities.

In our experiment, we allowed students to sit in front of networked computers in one lab together. In the distance learning, two or more students may happen to be in one household. However, it would have been better if some students sat in separate rooms to simulate distance learning where isolation could be felt easily. Regarding content, we just used Elementary Statistics course which was not real course taken by the participants. Allowing the participants to take real course which will have grades as part of the syllabus could have more accurate result.

One limitation of this study was that it was challenging to convince the subjects to participate in the study. This led to small sample size and biased sample. Thus, the samples may not be representative of the population under study. The bigger the sample size is the better reliability it will have. The other limitation may also be bias in engagement labelling. One problem related to labelling was how to rate short events such as brief eye closure or glancing to the side. Should the labeller overlook such events as normal behaviour or label them as very low engagement level is a limitation.

3.6. Chapter Summary

In this chapter, the finding of the analysis indicated that from 13 features, only 11 were significant for the four levels of student engagement. From the 11 significant features, only 4 were found to be the most important features.

In our study, we found two features were significant for students' very low level engagement (VH): Time of assignment submission (TA) and surprise (SUR). Time of assignment submission (TA) was a behavioural feature that affected very low engagement negatively. As an asynchronous online learner keeps doing the assignment for longer time, his disengagement decreases very highly and the time of assignment submission (TA) was the most important feature to predict the very low engagement level.

We also found that five features were significant for students' low level of engagement (L). These were the Time of assignment submission (TA), Number of content view (NCV), Score of quiz (SC), surprise emotion (SUR) and Time between post and reply (TPR). Surprise emotion (SUR) was the most important feature that affected low level of engagement. Surprise emotion (SUR) affected low level of engagement positively. As an asynchronous online learner felt surprise emotion (SUR) for longer time, his/her disengagement increases highly. In the attention level, positive affect seems to reduce resources available for effortful processing [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.. The surprise emotion was the most important feature to predict the low engagement level.

We found ten features to affect high level of student engagement (H) significantly. These were the Number of replies (NR), Time between post and reply (TPR), Time in the forum (TF), anger (ANG) emotion, surprise (SUR) emotion, sad (SAD), happy (HAP) emotion, Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV). The happy (HAP) feature in our study was found to affect high level of engagement negatively. As an asynchronous online learner kept feeling happy emotion for longer time, his engagement decreases highly. This was confirmed by [Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R.

and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA.in that positive affect leads to behavioural disengagement because it signals that all is well and there is no need to engage. The happy emotion was the most important feature to predict the high engagement level.

Nine features were found to affect very high level of engagement (VH) significantly. These were Number of replies (NR), Time between post and reply (TPR), Time in the forum (TF), anger (ANG) emotion, surprise (SUR) emotion, happy (HAP) emotion, Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV). Number of replies (NR) affected very high level of engagement positively. As an asynchronous online learner keeps getting many replies, his engagement increases very highly. The Number of replies (NR) was the most important feature to predict the very high engagement level. This has implication that it confirms that individual interaction of learners with each other has been main influencer of engagement [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204..

This study implies that the features: time of assignment submission (TA) and the surprise emotion (SUR) should be monitored to allow intervention at appropriate times. This finding has also implication that the two features: happy emotion (HAP) and number of replies (NR) should be supported to lead students to high and very high levels of student engagement in asynchronous online learning.

Moreover, we have described the implications of the study in general as follows. The finding presented in this chapter can help evaluate and improve understanding of asynchronous online student engagement. It informs how online learning systems might be used to enhance student engagement. Information about engagement levels can help teaching staff use online systems to manage and lead student learning. This can help reduce the 'transactional distance' between teachers and students and lead to learning. If a teacher keeps track of the engagement level of students, the learning process will be more effective.

In future work, allowing students to sit in separate rooms to simulate distance learning where isolation could be felt easily would be very important. Regarding content, allowing the participants to take real course which will have grades as part of the syllabus could have more accurate result.

The building and validation processes of the predictive model as well as the visualizations of the predicted engagement levels have been presented in the subsequent chapters.

Chapter 4

Student Engagement Level Prediction Model

4.1. Introduction

In this chapter, we described student engagement prediction model from three factors: behavioural, collaboration and emotional factors in an asynchronous online learning environment. For achieving our objective, we collected empirical data related to interaction with learning activities in LMS, that is described in chapter 3.

There were two research questions in this study. These were: i. can we build student engagement prediction model from three factors: behavioural, collaboration and emotional factors across micro level time scale such as 5 minutes? ii. Will collaborative features as a result of interaction in the discussion forum in e-learning such as number of replies to someone's post and the time between someone's post and replies he got impact student engagement levels?

In our work, we used the result we obtained in chapter 3. The result was obtained after the user interacted with learning activities in a Learning Management System (LMS) and with facial emotion recognition tool. The interactions were saved as log files. After collecting the log file data, we analysed the log files to determine the features that correlated significantly with the levels of engagement from three factors: behavioural, collaboration and emotional factors in an asynchronous online learning environment. We developed a student engagement prediction model using non-linear regression techniques from those features.

The subsequent sections present the model building, validating the prediction model and the discussion part.

4.2. Model Building

The student engagement prediction model that we built was based on the significant features identified for the four levels of engagement as described in chapter 3.

4.2.1. Non-Linear Regression Analysis and Result

When we examined the linear regression diagnostic, we observed that the values of R^2 were less than 0.5. The less value of R^2 implied that the linear regression line did not fit the data [Peter Bruce and Andrew Bruce, 2017] Peter Bruce, Andrew Bruce - Practical statistics for data scientists_ 50 essential concepts-O'Reilly (2017). We also plotted the scatter plot of the data related to the significant features with the corresponding engagement levels. We observed from the scatter plot that the data were not described by a linear function. Thus, it was necessary to implement a technique that fitted a non-linear function to the data. In our data, we used a polynomial regression (quadratic) to capture the data in non-linear relationship. Adding in higher order terms such as cubic polynomial often leads to undesirable twists in the regression equation [Peter Bruce and Andrew Bruce, 2017] Peter Bruce, Andrew Bruce - Practical statistics for data scientists_ 50 essential concepts-O'Reilly (2017). A method that is appropriate for fitting a non-linear function to the data is called iterative least square fitting [Brown,2001] Brown, A.M., 2001. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet.Computer methods and programs in biomedicine, 65(3), pp.191-200.. The process minimizes the values of the squared sum of the difference between data and fit. This involves starting with an estimate of parameter values. The sum of squares is described by:

$$SS = \sum_{j=1}^{n} \left[y - y_{fit}\right]^2 \quad (2)$$

where y is the data point (engagement level in our study), y_{fit} is the value of the curve at point y (value of the polynomial function in our study) and SS is the sum of the squares [Brown,2001] Brown, A.M., 2001. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet.Computer methods and programs in biomedicine, 65(3), pp.191-200.. The first iteration involves computing the SS based on the initial parameter values. The second iteration involves changing the parameter values by a small amount and recalculating the SS. This process is repeated many times to ensure that changes in the parameter values result in the smallest possible values of SS. The non-linear functions in our study are the quadratic functions given by:

$$y_{j(fit)} = a_j x_j^2 + b_j x_j + c_j, \quad (3)$$

Where $y_{j(fit)}$ is the value of the curve at point y_j, x_j is the value of the input features, a_j, b_j and c_j are parameters.

We started by taking each feature one at a time for each level of engagement. We obtained a set of p quadratic equations, one for each feature in each level of engagement. We, then, combined these p equations to get a single equation having the form shown in eq 4, for a particular level of engagement.

$$y = \sum_{j=1}^{p} a_j x_j^2 + \sum_{j=1}^{p} b_j x_j + \sum_{j=1}^{p} c_j \quad (4)$$

The Figure 4.11. below shows configuration of initial values of parameters for the non-linear regression of the data for the four levels of engagement.

i. Very low level (VL) of engagement

Non-linear (quadratic) relationship between two features and the very low level (VL) of engagement, using the technique shown in Eq 3 is displayed below.

	A	B	C	D	E	F	G	H	I
1		TA	ENG-VL	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$		
2		0	0	0.3	a	0.00			
3		4304	0	0.0494616	b	0.00			
4		926.2	0.4	0.4337692	c	0.30			
5		1214	0	0.4543831	Mean y	0.12			
6		4343	0	0.0367726	R^2	=1-SUM((C2:C13-D2:D13)^2)/SUM((C2:C13-F5)^2)			
7		2402	0	0.4342061					
8		2307	0	0.4420958					
9		0	0.4	0.3					
10		4062	0	0.1224668					
11		2390	0	0.4352393					
12		1211	0.6	0.4542087					
13		4639	0	-0.063419					
14	Min:	0	0						
15	Max:	4639	0.6						

	K	L	M	N	O	P	Q	R	S	T
1		SUR	ENG-VL	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
2		10.7	0	0.07543	a	-0.0006				
3		9.2	0	0.0632	b	0.02				
4		4.95	0.4	0.01435	c	-0.07				
5		7.98	0	0.05137	Mean y	0.11667				
6		28.7	0	0.00984	R^2	=1-SUM((M2:M13-N2:N13)^2)/SUM((M2:M13-P5)^2)				
7		7.47	0	0.04592						
8		60.2	0	-1.0382						
9		25.9	0.4	0.04565						
10		5.61	0	0.02337						
11		2.91	0	-0.0168						
12		45.1	0.6	-0.3894						
13		4.43	0	0.0068						
14	Min:	2.91	0							
15	Max:	60.2	0.6							

Figure 4.11a: Configuration of initial values of parameters for the non-linear regression of the data for very low level (VL) of engagement.

ii. Low level (L) of engagement

We applied the technique shown in Eq 3 to find non-linear (quadratic) relationship between five features and the low level (L) of engagement.

	A	B	C	D	E	F	G	H	I
1		TA	ENG-L	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$		
2		0	0	0.70	a	0.00			
3		4303.6	0	0.45	b	0.0002			
4		926.2	0	0.83	c	0.7			
5		1214.2	0.6	0.85	Mean y	0.483333333			
6		4343.4	0.4	0.44	R^2	=1-SUM((C2:C13-D2:D13)^2)/SUM((C2:C13-F5)^2)			
7		2402.2	0.2	0.83					
8		2306.6	1.2	0.84					
9		0	1.2	0.70					
10		4061.8	0	0.52					
11		2390.4	0	0.84					
12		1211	2.2	0.85					
13		4639	0	0.34					
14	Min:	0	0						
15	Max:	4639	2.2						
16									

	K	L	M	N	O	P	Q	R	S	T
1		TPR	ENG-L	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
2		1737.8	0	0.446476	a	0.00000008				
3		5710.8	0	1.224739	b	-0.0004				
4		1730.4	0	0.447383	c	0.9				
5		2916.4	0.6	0.413871	Mean y	0.483333333				
6		2883.4	0.4	0.41176	R^2	=1-SUM((M2:M13-N2:N13)^2)/SUM((M2:M13-P5)^2)				
7		512.6	0.2	0.715981						
8		0	1.2	0.9						
9		1044	1.2	0.569595						
10		515.6	0	0.715027						
11		5774.8	0	1.257945						
12		824.6	2.2	0.624557						
13		1665	0	0.455778						
14	Min:	0	0							
15	Max:	5774.8	2.2							
16										

	A	B	C	D	E	F	G	H	I
16									
17									
18		**SC**	**ENG-L**	**y_{fit}**	**Parameter**		**$y_{fit} = ax^2 + bx + c$**		
19		3.428	0	0.38	**a**	0.06			
20		3.142	0	0.25	**b**	0.05			
21		3.428	0	0.38	**c**	-0.5			
22		5.428	0.6	1.54	**Mean y**	0.483333333			
23		3.428	0.4	0.38	**R^2**	=1-SUM((C19:C30-D19:D30)^2)/SUM((C19:C30-F22)^2)			
24		5.428	0.2	1.54					
25		5.428	1.2	1.54					
26		4	1.2	0.66					
27		2.286	0	-0.07					
28		1.142	0	-0.36					
29		5.142	2.2	1.34					
30		3.142	0	0.25					
31	**Min:**	1.142	0						
32	**Max:**	5.428	2.2						
33									

	K	L	M	N	O	P	Q	R	S	T
16										
17										
18		**TRC**	**ENG-L**	**y_{fit}**	**Parameter**		**$y_{fit} = ax^2 + bx + c$**			
19		2.4	0	0.392	**a**	1.7				
20		2	0	0.4	**b**	-7.5				
21		2.6	0	0.592	**c**	8.6				
22		3	0.6	1.4	**Mean y**	0.483333333				
23		2.4	0.4	0.392	**R^2**	=1-SUM((M19:M30-N19:N30)^2)/SUM((M19:M30-P22)^2)				
24		1.8	0.2	0.608						
25		3.4	1.2	2.752						
26		2	1.2	0.4						
27		2	0	0.4						
28		2.8	0	0.928						
29		3.4	2.2	2.752						
30		2.2	0	0.328						
31	**Min:**	0	0							
32	**Max:**	5774.8	2.2							
33										

	A	B	C	D	E	F	G	H	I
34									
35		SUR	ENG-L	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$		
36		10.718	0	0.71	a	-0.002			
37		9.198	0	0.68	b	0.06			
38		4.954	0	0.55	c	0.3			
39		7.978	0.6	0.65	Mean y	0.483333333			
40		28.696	0.4	0.37	R^2	=1-SUM((C36:C47-D36:D47)^2)/SUM((C36:C47-F39)^2)			
41		7.47	0.2	0.64					
42		60.158	1.2	-3.33					
43		25.888	1.2	0.51					
44		5.614	0	0.57					
45		2.912	0	0.46					
46		45.128	2.2	-1.07					
47		4.428	0	0.53					
48	Min:	2.912	0						
49	Max:	60.158	2.2						
50									

Figure 4.11b: Configuration of initial values of parameters for the non-linear regression of the data for low level (L) of engagement

iii. High level (H) of engagement

A technique shown in Eq 3 was used to find non-linear (quadratic) relationship between ten features and the high level (H) of engagement.

	A	B	C	D	E	F	G	H	I	J
1										
2		NR	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
3		0.2	8.2	6.6	a	15				
4		0.8	0.4	6.6	b	-15				
5		0.2	6.8	6.6	c	9				
6		0.6	1.4	5.4	Mean y	3.816667				
7		0.4	5.6	5.4	R^2	=1-SUM((C3:C14-D3:D14)^2)/SUM((C3:C14-F6)^2)				
8		0.2	1.4	6.6						
9		0	8.4	9						
10		0.4	2.8	5.4						
11		0.4	4.4	5.4						
12		0.4	0.8	5.4						
13		0.6	4	5.4						
14		0.4	1.6	5.4						
15	Min:	0	0.4							
16	Max:	0.8	8.4							
17										

	A	B	C	D	E	F	G	H	I	J
16	Max:	0.8	8.4							
17										
18										
19		TF	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
20		0.4	8.2	4.12	a	2				
21		1	0.4	4	b	-3				
22		1	6.8	4	c	5				
23		1	1.4	4	Mean y	3.816667				
24		1.6	5.6	5.32	R^2	=1-SUM((C20:C31-D20:D31)^2)/SUM((C20:C31-F23)^2)				
25		2.4	1.4	9.32						
26		0.4	8.4	4.12						
27		1.6	2.8	5.32						
28		1.4	4.4	4.72						
29		0.4	0.8	4.12						
30		1.4	4	4.72						
31		0.8	1.6	3.88						
32	Min:	0.4	0.4							
33	Max:	2.4	8.4							

	A	B	C	D	E	F	G	H	I	J
37		TRC	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
38		2.4	8.2	6.1272	a	-1.28				
39		2	0.4	4.58	b	9.5				
40		2.6	6.8	6.7472	c	-9.3				
41		3	1.4	7.68	Mean y	3.816667				
42		2.4	5.6	6.1272	R^2	=1-SUM((C38:C49-D38:D49)^2)/SUM((C38:C49-F41)^2)				
43		1.8	1.4	3.6528						
44		3.4	8.4	8.2032						
45		2	2.8	4.58						
46		2	4.4	4.58						
47		2.8	0.8	7.2648						
48		3.4	4	8.2032						
49		2.2	1.6	5.4048						
50	Min:	1.8	0.4							
51	Max:	3.4	8.4							

	A	B	C	D	E	F	G	H	I	J
55		ANG	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
56		8.49	8.2	5.60114	a	-0.002				
57		3.692	0.4	5.861978	b	-0.03				
58		0.47	6.8	5.985458	c	6				
59		2.536	1.4	5.911057	Mean y	3.816667				
60		2.22	5.6	5.923543	R^2	=1-SUM((C56:C67-D56:D67)^2)/SUM((C56:C67-F59)^2)				
61		6.978	1.4	5.693275						
62		1.738	8.4	5.941819						
63		15.22	2.8	5.080103						
64		3.888	4.4	5.853127						
65		38.084	0.8	1.956698						
66		1.542	4	5.948984						
67		0.196	1.6	5.994043						
68	Min:	0.196	0.4							
69	Max:	38.084	8.4							

	A	B	C	D	E	F	G	H	I	J
73		HAP	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
74		25.5	8.2	4.3635	a	0				
75		47.602	0.4	1.644954	b	-0.123				
76		33.124	6.8	3.425748	c	7.5				
77		47.717	1.4	1.630858	Mean y	3.816667				
78		15.868	5.6	5.548236	R^2	=1-SUM((C74:C85-D74:D85)^2)/SUM((C74:C85-F77)^2)				
79		76.854	1.4	-1.95304						
80		9.56	8.4	6.32412						
81		25.112	2.8	4.411224						
82		22.598	4.4	4.720446						
83		13.612	0.8	5.825724						
84		43.382	4	2.164014						
85		93.206	1.6	-3.96434						
86	Min:	9.56	0.4							
87	Max:	93.206	8.4							
88										

	K	L	M	N	O	P	Q	R	S	T
1										
2		TPR	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
3		1737.8	8.2	5.546476	a	0.00000008				
4		5710.8	0.4	6.324739	b	-0.0004				
5		1730.4	6.8	5.547383	c	6				
6		2916.4	1.4	5.513871	Mean y	3.816666667				
7		2883.4	5.6	5.51176	R^2	=1-SUM((M3:M14-N3:N14)^2)/SUM((M3:M14-P6)^2)				
8		512.6	1.4	5.815981						
9		0	8.4	6						
10		1044	2.8	5.669595						
11		515.6	4.4	5.815027						
12		5774.8	0.8	6.357945						
13		824.6	4	5.724557						
14		1665	1.6	5.555778						
15	Min:	0	0.4							
16	Max:	5774.8	8.4							
17										

	K	L	M	N	O	P	Q
19		NCV	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$
20		6.4	8.2	8.9808	a	-0.02	
21		3.4	0.4	3.5688	b	2	
22		9.2	6.8	13.7072	c	-3	
23		5.8	1.4	7.9272	Mean y	3.816666667	
24		4.6	5.6	5.7768	R^2	=1-SUM((M20:M31-N20:N31)^2)/SUM((M20:M31-P23)^2)	
25		3.8	1.4	4.3112			
26		6.8	8.4	9.6752			
27		3	2.8	2.82			
28		3	4.4	2.82			
29		4.8	0.8	6.1392			
30		4	4	4.68			
31		3.8	1.6	4.3112			
32	Min:	0	0.4				
33	Max:	9.2	8.4				

	K	L	M	N	O	P	Q
37		TA	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$
38		0	8.2	9.8	a	0	
39		4303.6	0.4	5.4964	b	-0.001	
40		926.2	6.8	8.8738	c	9.8	
41		1214.2	1.4	8.5858	Mean y	3.816666667	
42		4343.4	5.6	5.4566	R^2	=1-SUM((M38:M49-N38:N49)^2)/SUM((M38:M49-P41)^2)	
43		2402.2	1.4	7.3978			
44		2306.6	8.4	7.4934			
45		0	2.8	9.8			
46		4061.8	4.4	5.7382			
47		2390.4	0.8	7.4096			
48		1211	4	8.589			
49		4639	1.6	5.161			
50	Min:	0	0.4				
51	Max:	4639	8.4				

	K	L	M	N	O	P	Q	R	S	T
55		SUR	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
56		10.718	8.2	3.451807	a	0.003				
57		9.198	0.4	3.34579	b	0.01				
58		4.954	6.8	3.123166	c	3				
59		7.978	1.4	3.270725	Mean y	3.816666667				
60		28.696	5.6	5.757341	R^2	=1-SUM((M56:M67-N56:N67)^2)/SUM((M56:M67-P59)^2)				
61		7.47	1.4	3.242103						
62		60.158	8.4	14.45853						
63		25.888	2.8	5.269446						
64		5.614	4.4	3.150691						
65		2.912	0.8	3.054559						
66		45.128	4	9.560889						
67		4.428	1.6	3.103102						
68	Min:	2.912	0.4							
69	Max:	60.158	8.4							
70										

	K	L	M	N	O	P	Q	R	S	T	U
73		SAD	ENG-H	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$				
74		7.67	8.2	4.327469	a	-0.008					
75		4.842	0.4	3.3945	b	0.43					
76		2.914	6.8	2.685089	c	1.5					
77		4.642	1.4	3.323675	Mean y	3.816666667					
78		11.44	5.6	5.372211	R^2	=1-SUM((M74:M85-N74:N85)^2)/SUM((M74:M85-P77)^2)					
79		0.13	1.4	1.555765							
80		20.796	8.4	6.982491							
81		13.89	2.8	5.929243							
82		46.29	4.4	4.262587							
83		1.768	0.8	2.235233							
84		3.818	4	3.025123							
85		1.288	1.6	2.040568							
86	Min:	0.13	0.4								
87	Max:	46.29	8.4								
88											

Figure 4.11c: Configuration of initial values of parameters for the non-linear regression of the data for high level (H) of engagement

iv. Very high level (VH) of engagement

A non-linear (quadratic) relationship between nine features and the very high level (VH) of engagement was formed using a technique shown in Eq 3.

	A	B	C	D	E	F	G
1		TRC	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$
2		2.4	21.8	23.712	a	-0.8	
3		2	29.6	25.4	b	-0.7	
4		2.6	22.8	22.772	c	30	
5		3	28	20.7	Mean y	25.58333	
6		2.4	24	23.712	R^2	=1-SUM((C2:C13-D2:D13)^2)/SUM((C2:C13-F5)^2)	
7		1.8	28.4	26.148			
8		3.4	20.4	18.372			
9		2	25.6	25.4			
10		2	25.6	25.4			
11		2.8	29.2	21.768			
12		3.4	23.2	18.372			
13		2.2	28.4	24.588			
14	Min:	1.8	20.4				
15	Max:	3.4	29.6				

	A	B	C	D	E	F	G
19		ANG	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$
20		8.49	21.8	24.31396	a	0.002	
21		3.692	29.6	24.1011	b	0.02	
22		0.47	22.8	24.00984	c	24	
23		2.536	28	24.06358	Mean y	25.58333	
24		2.22	24	24.05426	R^2	=1-SUM((C20:C31-D20:D31)^2)/SUM((C20:C31-F23)^2)	
25		6.978	28.4	24.23694			
26		1.738	20.4	24.0408			
27		15.22	25.6	24.7677			
28		3.888	25.6	24.10799			
29		38.084	29.2	27.66246			
30		1.542	23.2	24.0356			
31		0.196	28.4	24.004			
32	Min:	0.196	20.4				
33	Max:	38.084	29.6				

	A	B	C	D	E	F	G
37		SUR	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$
38		10.718	21.8	25.9205	a	-0.001	
39		9.198	29.6	26.08758	b	-0.09	
40		4.954	22.8	26.5296	c	27	
41		7.978	28	26.21833	Mean y	25.58333	
42		28.696	24	23.5939	R^2	=1-SUM((C38:C49-D38:D49)^2)/SUM((C38:C49-F41)^2)	
43		7.47	28.4	26.2719			
44		60.158	20.4	17.9668			
45		25.888	25.6	23.99989			
46		5.614	25.6	26.46322			
47		2.912	29.2	26.72944			
48		45.128	23.2	20.90194			
49		4.428	28.4	26.58187			
50	Min:	2.912	20.4				
51	Max:	60.158	29.6				

	A	B	C	D	E	F	G	H	I	J
55		NR	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
56		0.2	21.8	20.94448	a	0				
57		0.8	29.6	30.11105	b	15.27762				
58		0.2	22.8	20.94448	c	17.88895				
59		0.6	28	27.05552	Mean y	25.58333				
60		0.4	24	24	R^2	=1-SUM((C56:C67-D56:D67)^2)/SUM((C56:C67-F59)^2)				
61		0.2	28.4	20.94448						
62		0	20.4	17.88895						
63		0.4	25.6	24						
64		0.4	25.6	24						
65		0.4	29.2	24						
66		0.6	23.2	27.05552						
67		0.4	28.4	24						
68	Min:	0	20.4							
69	Max:	0.8	29.6							
70										

	A	B	C	D	E	F	G	H	I	J
73		SC	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
74		3.428	21.8	23.99	a	0.4				
75		3.142	29.6	24.38	b	-4				
76		3.428	22.8	23.99	c	33				
77		5.428	28	23.07	Mean y	25.58333				
78		3.428	24	23.99	R^2	=1-SUM((C74:C85-D74:D85)^2)/SUM((C74:C85-F77)^2)				
79		5.428	28.4	23.07						
80		5.428	20.4	23.07						
81		4	25.6	23.40						
82		2.286	25.6	25.95						
83		1.142	29.2	28.95						
84		5.142	23.2	23.01						
85		3.142	28.4	24.38						
86	Min:	1.142	20.4							
87	Max:	5.428	29.6							
88										

	K	L	M	N	O	P	Q	R	S	T
1		TA	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
2		0	21.8	20	a	0				
3		4303.6	29.6	24.3036	b	0.001				
4		926.2	22.8	20.9262	c	20				
5		1214.2	28	21.2142	Mean y	25.58333333				
6		4343.4	24	24.3434	R^2	=1-SUM((M2:M13-N2:N13)^2)/SUM((M2:M13-P5)^2)				
7		2402.2	28.4	22.4022						
8		2306.6	20.4	22.3066						
9		0	25.6	20						
10		4061.8	25.6	24.0618						
11		2390.4	29.2	22.3904						
12		1211	23.2	21.211						
13		4639	28.4	24.639						
14	Min:	0	20.4							
15	Max:	4639	29.6							
16										

	K	L	M	N	O	P	Q	R	S	T	U
37		HAP	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$				
38		25.5	21.8	24.64475	a	-0.001					
39		47.602	29.6	25.01823	b	0.09					
40		33.124	22.8	24.88396	c	23					
41		47.717	28	25.01762	Mean y	25.58333333					
42		15.868	24	24.17633	R^2	=1-SUM((M38:M49-N38:N49)^2)/SUM((M38:M49-P41)^2)					
43		76.854	28.4	24.01032							
44		9.56	20.4	23.76901							
45		25.112	25.6	24.62947							
46		22.598	25.6	24.52315							
47		13.612	29.2	24.03979							
48		43.382	23.2	25.02238							
49		93.206	28.4	22.70118							
50	Min:	9.56	20.4								
51	Max:	93.206	29.6								

	K	L	M	N	O	P	Q	R	S	T
55		TPR	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
56		1737.8	21.8	23.60687	a	0.0000001				
57		5710.8	29.6	24.977	b	-0.0004				
58		1730.4	22.8	23.60727	c	24				
59		2916.4	28	23.68398	Mean y	25.58333333				
60		2883.4	24	23.67804	R^2	=1-SUM((M56:M67-N56:N67)^2)/SUM((M56:M67-P59)^2)				
61		512.6	28.4	23.82124						
62		0	20.4	24						
63		1044	25.6	23.69139						
64		515.6	25.6	23.82034						
65		5774.8	29.2	25.02491						
66		824.6	23.2	23.73816						
67		1665	28.4	23.61122						
68	Min:	0	20.4							
69	Max:	5774.8	29.6							
70										

	K	L	M	N	O	P	Q	R	S	T
73		NCV	ENG-VH	y_{fit}	Parameter		$y_{fit} = ax^2 + bx + c$			
74		6.4	21.8	21.0192	a	0.02				
75		3.4	29.6	26.4312	b	-2				
76		9.2	22.8	16.2928	c	33				
77		5.8	28	22.0728	Mean y	25.58333333				
78		4.6	24	24.2232	R^2	=1-SUM((M74:M85-N74:N85)^2)/SUM((M74:M85-P77)^2)				
79		3.8	28.4	25.6888						
80		6.8	20.4	20.3248						
81		3	25.6	27.18						
82		3	25.6	27.18						
83		4.8	29.2	23.8608						
84		4	23.2	25.32						
85		3.8	28.4	25.6888						
86	Min:	0	20.4							
87	Max:	9.2	29.6							

Figure 4.11d: Configuration of initial values of parameters for the non-linear regression of the data for very high level (VH) of engagement

After finding the configuration of initial values of parameters for the non-linear regression of the data for the four levels of engagement, the next step was to find the better fit using SOLVER.

4.2.2. Finding of Better Fit Using SOLVER

We used SOLVER function in MS Excel to fit data with the non-linear functions. We implemented SOLVER to fit a curve to the data. Solver tries to maximize the value of R^2. The program will iteratively cycle through the fitting routine, changing the parameter values of a, b and c until the largest value of R^2 is calculated. The R^2 value, the correlation index or coefficient of determination, is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[y - y_{fit}]^2}{\sum_{i=1}^{n}[y - y_{mean}]^2} \tag{5}$$

and is calculated by entering and expressing it as an array formula as:

=1−SUM ((C2:C13−D2:D13)^2)/SUM((C2:C13−Mean–of–y)^2)

The following Table 4.22 displays the fit as calculated by SOLVER. The table illustrates the best fit and an improvement over the fit provided by the initial parameter values.

	Features											
		NR	TPR	TF	NCV	TRC	TA	ANG	SUR	HAP	SAD	SC
Engagement levels	**Parameters**											
Very high level(VH)	a	0	0.0000001		0.02	-0.8	0	0.002	-0.001	-0.001		0.4
	b	15.28	-4.00E-04		-2.00	-0.7	0.001	0.02	-0.09	0.09		-4
	c	17.89	24.00		33.00	30	20	24	27	23		33
	R^2	1	0.96		0.98	0.97	0.95	1.0	0.93	0.99		1.0
High level(H)	a	15	8E-08	2	-0.02	-1.28	0	-0.002	0.003	0	-1.28	
	b	-15	-0.0004	-3	2	9.5	-0.001	-0.03	0.01	-0.12	9.5	
	c	9	6	5	-3	-9.3	9.8	6	3	7.5	-9.3	
	R^2	0.99	1.0	0.97	0.99	0.91	0.99	0.97	0.99	1.0	0.91	
Low level(L)	a		0.00000008			1.7	-6E-08		-0.002			0.06
	b		-0.0004			-7.5	0.0002		0.06			0.05
	c		0.9			8.6	0.7		0.3			-0.5
	R^2		0.98			0.99	0.80		0.91			0.92
Very low level(VL)	a						-6E-08		-0.0006			
	b						0.0002		0.02			
	c						0.3		-0.07			
	R^2						0.90		0.99			

Table 4.22: The fit as calculated by SOLVER for the four levels of engagement

4.2.3. Proposed Model

The student engagement prediction model that we built was based on the significant features identified for the four levels of engagement in chapter 3. Thus, we obtained Eq. 6-9 as our final proposed model. After we obtained Eq 6-9, we recomputed the values of y for the given features and got the engagement level ranges or the prediction intervals. The equations and the engagement level ranges were presented in Table 4.23 as follows.

Eq.#	Equation	Range	
		Min.	Max.
6	$VL = 0.02 \times SUR + 0.23$ Where $2.9 \le \text{SUR} \le 11$	=0	=2
7	$\text{L} = 0.06 \times \text{SC}^2 + 0.05 \times \text{SC} + 1.7 \times \text{TRC}^2 - 7.5 \times \text{TRC} + 0.06 \times \text{SUR} + 10$, Where, $\text{TRC} \ge 2.4$ and $2.9 \le \text{SUR} \le 11$	>2	=7
8	$\text{H} = 15 \times \text{NR}^2 - 15 \times \text{NR} + 2 \times \text{TF}^2 - 3 \times \text{TF} - 0.02 \times \text{NCV}^2 + 2 \times \text{NCV} - 1.28 \times \text{TRC}^2 + 9.5 \times \text{TRC} - 0.03 \times \text{ANG} + 0.01 \times \text{SUR} - 0.123 \times \text{HAP} + 0.43 \times \text{SAD} + 35.5$, Where $0 \le \text{NR} \le 0.4$, $0.4 \le \text{TF} \le 0.8$, and $0.13 \le \text{SAD} \le 21$	>7	=62
9	$\text{VH} = 15.3 \times \text{NR} + 0.4 \times \text{SC}^2 - 4 \times \text{SC} + 0.02 \times \text{NCV}^2 - 2 \times \text{NCV} - 0.8 \times \text{TRC}^2 - 0.7 \times \text{TRC} + 0.02 \times \text{ANG} - 0.09 \times \text{SUR} + 0.09 \times \text{HAP} + 232$, Where, $3 \le \text{NCV} \le 9.2$, $1.8 < \text{TRC} \le 3.4$, and $0 \le \text{HAP} \le 43$	>62	=254

Table 4.23: The equations and the engagement level ranges

4.3. Validating the Proposed Model

4.3.1. Purpose

In this experiment, we aim to validate the student engagement prediction model. For achieving our objective, we collected empirical data related to interaction with learning activities in LMS. We also applied self-reporting as ground truth data. We then compared the prediction of the model with the self-reporting. Further details of data collection and analysis are provided in the following subsections.

4.3.2. Participants

There were 12 participants and all signed consent forms. The participants were new and took part on the validation experiment, not on the model building experiment. Their minimum age is 23, their maximum age is 41 and their average age is 31.2. The gender of all was male. The participants interacted with the learning management system (LMS) in one session, for 25 minutes. The face of each participant was being tracked with facial emotion recognition tool. The following Table 4.24 summarizes the profile of the participants.

	Gender		Age			LMS experience	
	Male	Female	Min	Max	Average	Yes , I have	No, I don't have
Quantity	12	0	23	41	31.2	4	8

Table 4.24: Summary of the profile of the participants

4.3.3. Procedure

The participants were given instruction on how to interact with the system for the validation experiment. The experiment took place in User Centric Computing and Networking (UCCN) lab of Computer Science and Engineering (CSE) department.

There was only one session of 25 minutes duration. They were working on LMS which was Moodle that was implemented for model building. They interacted with four learning activities. These were content reading for Descriptive Statistics course, submitting assignment, taking part in the discussion forum and taking quiz. Each of the 12 participants took 25 minutes for the interaction. At the end of the 25 minutes, they filled a questionnaire taken from the work of [Dixson, 2015] Dixson, M.D., 2015. Measuring student engagement in the online course: The Online Student Engagement scale (OSE). *Online Learning*, *19*(4), p.n4.for self-reporting the engagement levels they experienced. Their faces were being tracked with face tracking tool called clmtrackr while interacting with the system to classify the six basic emotions from facial expressions. The data collected after the interaction with the LMS and averaged for 25 minutes from the log file and face tracking tool during the validation experiment was given in Table 4.25 below. This log file was categorized in 5 minutes sample. At the same time, the log file of the rate of facial emotion was downloaded every 5 minutes. [Dixson, 2015] Dixson, M.D., 2015. Measuring student engagement in the online course: The Online Student Engagement scale (OSE). *Online Learning*, *19*(4), p.n4. used 19 items with 5 levels scale for each of the 19 items. We also applied the model built on the collected data to classify each of the participants in to one of the four engagement levels after computation. The classified engagement levels after computed by the model were given below in Table 4.26. We then matched the engagement levels classified by the model and the self-report. We classified the participants in to the four engagement levels from the self-report based on the works of [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context.

Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184.and [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE.. Inspired by the work of [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184.we mapped scale 1 or 2 to very low (VL) engagement levels, and scale 3 was mapped to low (L) engagement level, scale 4 was mapped to high (H) engagement level, and scale 5 was mapped to very high (VH) engagement level. We also used the maximum value of the response to classify the participant in to one of the four engagement levels based on the work of [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE.. [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE. remarked that it would suffice to take the maximum activation value of ‘feeling’ in response to each of the scale from the self-reporting. The responses given through the self-report and the classified engagement levels based on the works of [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE. and [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184. is given in Table 4.27 below.

S/no	Participant	TRC	NCV	TA	TF	SC	NR	TPR	ANG	SAD	HAP	SUR
1	Participant 1	1.4	2.4	1.8	1.4	2	0	0	4.82	37.38	38.01978	6.92
2	Participant 2	1	2.4	1.4	1	2	0	0	36.62	5.12	24.16027	10.34
3	Participant 3	1.2	2.4	2.4	2	1.2	0	0	30.42	0.48	32.96322	1.48
4	Participant 4	1.6	2.4	1.8	1.4	1.6	0.8	576	35.16	1.06	13.67896	0.68
5	Participant 5	1.6	2.4	1.2	0.8	2	0	0	0.26	16.68	79.19	3.42
6	Participant 6	1	2.4	1.4	1.8	2	0.2	288	3.04	10.36	57.76176	24.24
7	Participant 7	1.8	2.8	1.4	1	2	0.2	48	10.06	2.12	41.79586	23.98
8	Participant 8	1.6	1.6	2	1.6	2	0.4	576	20.38	2.2	47.06668	9.64
9	Participant 9	2.6	4.6	1.8	2.2	1.6	0.8	864	24.1	2.96	28.2219	16.2
10	Participant 10	3.4	4.4	1.6	1.2	2	0.8	864	3.72	0.06	60.21171	1.74
11	Participant 11	0.8	2.4	1.4	1.6	2	0.8	864	16.6	0.94	28.09114	23.06
12	Participant 12	1.4	2.8	3.2	3.2	2	0.4	1440	9.48	17.82	48.35905	7.6

Table 4.25: The data collected after the interaction with the LMS

The data in Table 4.25 was sampled in 5 minutes and averaged for 25 minutes from the log file and face tracking tool during the validation experiment. NR=Number of Replies, TPR=Time between Post and Replies, TF=Time in the Forum, TA=Time of Assignment Submission, NCV=Number of Content View, SC=Score of quiz, TRC=Time to Read Content, ANG=Anger, SAD=sad, SUR=Surprise, HAP=Happy. All time related features were measured in minutes.

S/no	Participant	Y_{VL}	Y_L	Y_H	Y_{VH}	Decision after Model Computation
1	Participant 1	0.3684	10.7552	46.22417	228.4954	VH
2	Participant 2	0.4368	10.9604	46.63949	227.5762	VH
3	Participant 3	0.23	10.1464	44.99572	231.2179	VH
4	Participant 4	0.23	10.2336	49.83329	240.7371	VH
5	Participant 5	0.2984	10.5452	48.44643	225.2974	VH
6	Participant 6	0.23	10.34	43.5061	226.5392	VH
7	Participant 7	0.23	10.34	47.20471	226.6126	VH
8	Participant 8	0.4228	10.9184	41.6138	231.26	VH
9	Participant 9	0.23	2.2256	57.56451	224.4232	VH
10	Participant 10	0.2648	4.5964	53.94156	217.717	VH
11	Participant 11	0.23	10.34	43.64719	238.6248	VH
12	Participant 12	0.382	10.796	49.64044	231.2256	VH

Table 4.26: The classified engagement levels after computed by the model based on the algorithm in Figure 4.11.

In the Table 4.26, y_{VL} is the value of y for the given features for the very low (VL) level of engagement, y_L is the value of y for the given features for the low (L) level of engagement, y_H is the value of y for the given features for the high (H) level of engagement, y_{VH} is the value of y for the given features for the very high (VH) level of engagement

S/no	Participant	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Decision Based on [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A

							cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE. and [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184.
1	Participant 1	0	1	6	4	8	VH
2	Participant 2	2	3	8	5	1	L
3	Participant 3	0	0	1	4	14	VH
4	Participant 4	3	0	0	3	13	VH
5	Participant 5	2	1	3	4	9	VH
6	Participant 6	0	3	3	3	10	VH
7	Participant 7	2	3	2	5	7	VH
8	Participant 8	1	1	3	4	10	VH

9	Participant 9	0	0	5	5	9	VH
10	Participant 10	0	0	4	7	8	VH
11	Participant 11	2	5	7	5	0	L
12	Participant 12	0	0	0	3	16	VH

Table 4.27: The Responses given through the self-report and the classified engagement levels

The data in Table 4.27 was based on the works of *[Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE.*and *[Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184..* 8 under scale 5 means that participant 1 selected scale 5 for 8 items or questions out of the 19 items. And according to *[Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE.*, because 8 is the maximum of the responses given by participant 1, we decided that the engagement level of participant 1 was VH as scale 5 was mapped to very high (VH) level of engagement inspired by *[Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184..*

We determined the engagement levels using the proposed model according to the algorithm listed below in Figure 4.12: There are four ranges as shown in section 4.2.3.

```
FOR each student s
    COMPUTE the engagement level, y
    IF (0 <= y <= 2) THEN //0 <= y <= 2 is the range of very low (VL) engagement
        DETERMINE s to be with very low (VL) engagement level
    ELSEIF (2 < y <= 7) THEN
        DETERMINE s to be with low (L) engagement level
    ELSEIF (7 < y <= 62) THEN
        DETERMINE s to be with high (H) engagement level
    ELSEIF (62 < y <= 254) THEN
        DETERMINE s to be with very high (VH) engagement level
    ELSEIF s=VL and s=L and s=H and s=VH THEN
        DETERMINE s to be with higher engagement level exhibited by majority of students
    ELSE
        DETERMINE s to be with no engagement level
    END IF
```

Figure 4.12: Algorithm to determine the engagement levels using the proposed model

Thus from the analysis, the proposed model was able to correctly predict the engagement levels of 10 participants out of 12. The accuracy of the model was found to be 83.3%.

4.4. Discussion

We proposed a student engagement prediction models using 9 features out of 13 that were significant to affect the levels of student engagement and emerged in the final models. We built a student engagement prediction model that predicts 4 levels of engagement using these features through non-linear regression techniques. The 4 student engagement levels were very low engagement level (VL), low engagement level (L), high engagement level (H) and very high engagement level (VH). The features were also of three categories. These were behavioural, collaboration and emotional features. The features were from interaction with an LMS and facial emotion recognition tool. Initially we proposed a linear model for the features in each engagement level. However, the linear regression line did not fit the data. Thus, we implemented a technique that fit a non-linear function to the data. We used a polynomial regression (quadratic) to capture the data in non-linear relationship. Adding in higher order terms such as cubic polynomial often leads to undesirable twists in the regression equation [Peter Bruce and Andrew Bruce, 2017] Peter Bruce, Andrew Bruce - Practical statistics for data scientists_ 50 essential concepts-O'Reilly (2017).

We performed validation of the results of the study. We determined the accuracy of identifying students with discrete levels of engagement. The proposed model was able to correctly predict the engagement levels of 10 students out of 12. The accuracy of the model was found to be 83.3%. However, the accuracy was not greater than 83.3%, because of the fact that the students were unable to accurately distinguish and report their actual level of engagement through the self-report questionnaire [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184.. Moreover, [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123.explained that agreement between external observer used for annotation while building the model and self-reporting used for annotation while validation purpose is very low. They also remarked that it is difficult to specify exact bounds on what constitutes "good" accuracy. However, at a minimum, the accuracy should exceed random guessing (chance).The students may consciously or unconsciously conceal his or her real emotions as shown by observable cues like facial, however will still reveal their internal feelings by invisible cues like bio signals [Gunes and Pantic, 2010] Gunes, H. and Pantic, M., 2010. Automatic, dimensional and continuous emotion recognition. International Journal of Synthetic Emotions (IJSE), 1(1), pp.68-99.. Thus, implementing brain signal reader to compare with the prediction of the model can be done in future work to get better accuracy of the model. The restrictions of the domains in the model are from the monotonic properties of the features with respect the engagement levels obtained from the correlation analysis result.

One of the contributions of this study is that we built a student engagement prediction model from three factors namely behavioural, collaboration and emotional factors as engagement is a multifaceted construct. Moreover, our student engagement prediction model predicted student engagement levels in smaller time scale that is 5 minutes with more than 83% accuracy. Providing students with support and guidance as soon as possible to lessen the danger of disengagement is critical [Falkner and Falkner, 2012] Falkner, N.J. and Falkner, K.E., 2012, September. A fast measure for identifying at-risk students in computer science. In Proceedings of the ninth annual international conference on International computing education research (pp. 55-62)..

The other contribution of this study was the finding that two collaborative features which are Time in the forum (TF) was significant in predicting high and Number of replies (NR) was significant in predicting both high and very high levels of engagement. This finding suggests that these two collaborative features should be supported to lead students to high and very high levels of student engagement in asynchronous online learning. Time between post and reply played little part in predicting student engagement.

The final contribution was that surprise was an emotional feature that affected very low and low engagement levels. As surprise emotion increases, student disengagement increases very highly. In the attention level, positive affect seems to reduce resources available for effortful processing [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.. Time to read content (TRC) is a behavioural feature that affected low level of engagement. As Time to read content (TRC) increases, student disengagement increases highly. Similar result was also reported by [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124. that long time spent on the same page was associated with disengagement. Another behavioural feature affecting low engagement was score of quiz. As score increases, student disengagement increases highly. This was unexpected result, but [Woolf et al., 2009] Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. and Picard, R., 2009. Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), pp.129-164.reported that when problems are easy, a student gets bored. This study suggests that these features should be monitored to allow intervention at appropriate times. Two emotional features, disgust and fear did not correlate with any of the engagement levels.

The model presented in this paper can help evaluate and improve understanding of asynchronous online student engagement. It informs how online learning systems might be used to enhance student engagement. Information about engagement levels can help teaching staff use online systems to manage and lead student learning. This can help reduce the 'transactional distance' between teachers and students and lead to learning [Hamish Coates (2007)] Hamish Coates (2007) A model of online and general campus-based student engagement, Assessment & Evaluation in Higher Education, 32:2, 121-141, DOI: 10.1080/02602930600801878. If a teacher keeps track of the engagement level of students, the learning process will be more effective [Thomas and Jayagopi, 2017] Thomas, C. and Jayagopi, D.B., 2017, November. Predicting student engagement in classrooms using facial behavioral cues. In Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education (pp. 33-40)..

The model is based on 9 significant features, but not the most important features. The relative importance of the features was not determined. Further study can be done to determine which ones are the most important features. Moreover, further study can be performed to determine if same prediction results can be obtained with the most important features. Future research could consider other technologies such as mobile devices. Future research might also analyse other factors of engagement such as cognitive and social engagement factors.

This research was limited since the study was conducted with few participants. Results would be more generalizable if more participants were considered. Another limitation of the current study was that labelling the recorded interaction into levels of student engagement was done by the researcher. The results may have been affected by the interpretations of the researcher.

4.5. Chapter Summary

This chapter presents a student engagement prediction model using 9 features that were significant out of 13 to affect the levels of student engagement and emerged in the final models. We built the student engagement prediction model using the features through non-linear regression technique. The 4 student engagement levels were very low engagement level (VL), low engagement level (L), high engagement level (H) and very high engagement level (VH). The three factors were behavioural, collaboration and emotional, and measured from interaction with an LMS and facial emotion recognition tool.

Moreover, we built a student engagement prediction model from three factors namely behavioural, collaboration and emotional factors as engagement is a multifaceted construct. Moreover, our student engagement prediction model predicted student engagement levels in smaller time scale that is 5 minutes with more than 83% accuracy. Providing students with support and guidance as soon as possible to lessen the danger of disengagement is critical.

One emotional feature that is surprise (SUR) and two behavioural features which are Time to read content (TRC) and score of quiz (SC) were found to be indicators of lack of engagement. This finding suggests that these features should be monitored to allow intervention at appropriate times.

The other finding of this study was that two collaborative features which are Time in the forum (TF) was significant in predicting high and Number of replies (NR) was significant in predicting both high and very high levels of engagement. This finding suggests that these two collaborative features should be supported to lead students to high and very high levels of student engagement in asynchronous online learning. Time between post and reply played little part in predicting student engagement.

We performed validation of the results of the study. We determined the accuracy of identifying students with discrete levels of engagement. The proposed model was able to correctly predict the engagement levels of 10 students out of 12. The accuracy of the model was found to be 83.3%. However, the accuracy was not greater than 83.3%, because of the fact that the students were unable to accurately distinguish and report their actual level of engagement through the self-report questionnaire.

The model was based on 9 significant features, but not the most important features. The relative importance of the features was not determined. Further study can be done to determine which ones are the most important features. Moreover, further study can be performed also to determine if same prediction results can be obtained with the most important features. Future research could consider other technologies such as mobile devices. Future research might also analyze other factors of engagement such as cognitive and social engagement factors.

Implementation of student engagement awareness system that has been developed by us through the implementation of this model is presented in the subsequent chapter.

Chapter 5

A Student Engagement Awareness Dashboard

5.1. Introduction

In this chapter, we propose a student engagement visualization dashboard that visualizes the instantaneous engagement levels every minute, visualizes trends of student engagement levels and filters and displays the least engaged learner. The dashboard is based on a student engagement prediction model, which we also developed, and presented in chapter 4. We also performed the validation of these proposed visualizers in controlled experiment.

The subsequent sections described the proposed visualizer, the validation of the visualizer, and the discussion part.

5.2. The Proposed Visualizer

We visualized student engagement levels after predicting it using the model we built. We built three types of visualizations after classifying the student into the 4 classes of engagement levels. These visualizations are: i. instantaneous engagement level visualization, ii. Engagement level trend visualization and iii. Least engaged learner visualization. In the subsequent section, we explain each of these visualizations.

5.2.1. Instantaneous Engagement Level Visualization

Instantaneous engagement level visualization refers to displaying the number of students with a particular engagement level every minute. Visualizing data helps stakeholders easily see trends to understand what has been happening [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18.. Moreover, tracking engagement while the student is learning allows intervention at an appropriate time [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.. We used a pie chart to visualize the instantaneous engagement levels of the students. The size of the sector is proportional and varies with the number of students at a particular level of engagement. For instance, if there are larger numbers of students with very high levels of engagement, then the size of the sector allocated for very high levels of engagement will be bigger. There are four sectors of the pie chart as there are four levels of engagement. Each of the sectors also displayed four different colours. The variation of the colours was according to the intensity of engagement levels. Accordingly, for very high levels of engagement we used deep green as this level is very accepted and aspired level. We used light green for a high level of engagement. For a low level of engagement, we used light red and we used dark red for very low levels of engagement. The colours are commonly used in the traffic signals around the world, with similar meaning. Therefore, the use of colours is consistent with our everyday knowledge. Let’s use case study to explain the pie chart visualization of instantaneous engagement level every minute. There were 20 students studying online. It is assumed that the students were in asynchronous distance learning, which means the students may not enter the system at the same time. However, it is observed at a particular time t=07:29pm, which we say, the first minute, the dashboard visualizes the instantaneous engagement level of one student (Figure 5.13a). In the Figure 5.12a, VL=1 is the number of student with very low level of engagement. When the time t=07:31 pm, after two minutes, the dashboard visualizes the instantaneous engagement level of two students (Figure 5.13b). In the Figure 5.12b, VL=2 is the number of students with very low level engagement. At time t= 07:39pm, after ten minutes, the dashboard visualizes the instantaneous engagement level of three students (Figure 5.13c). In the Figure 5.12c, VL=2 is the number of students with very low level engagement and L=1 is the number of student with low level of engagement. At time t=07:49 pm, after twenty minutes, the dashboard visualizes the instantaneous engagement level of three students (Figure 5.13d). In the Figure 5.12d, VL=1 is the number of student with very low-level engagement, L=1 is the number of student with low-level engagement and VH=1 is the number of student with very high-level engagement. At time t =07:59pm, after thirty minutes, the dashboard visualizes the instantaneous engagement level of four students (Figure 5.13e). In the Figure 5.12e, VL=1 is the number of student with very low-level engagement, L=1 is the number of student with low-level engagement, H=1 is the number of student with high-level engagement and VH=1 is the number of student with very high-level engagement.

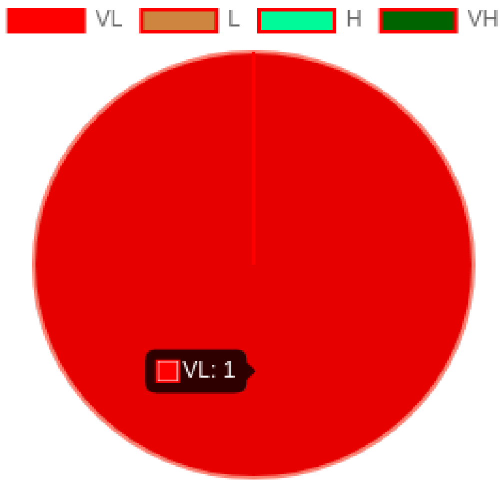


Figure 5.13a: Pie chart visualization of instantaneous engagement level of one students

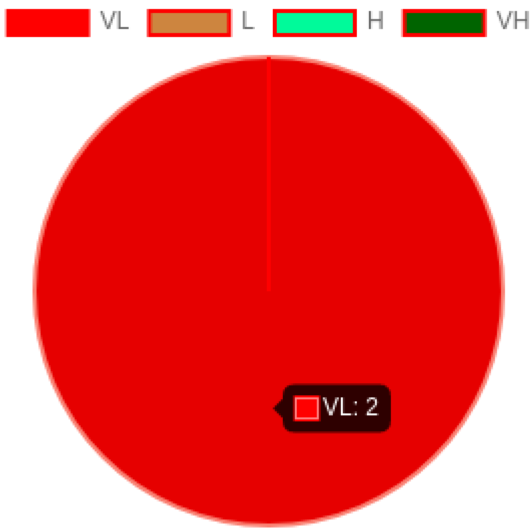


Figure 5.13b: Pie chart visualization of instantaneous engagement level of two students

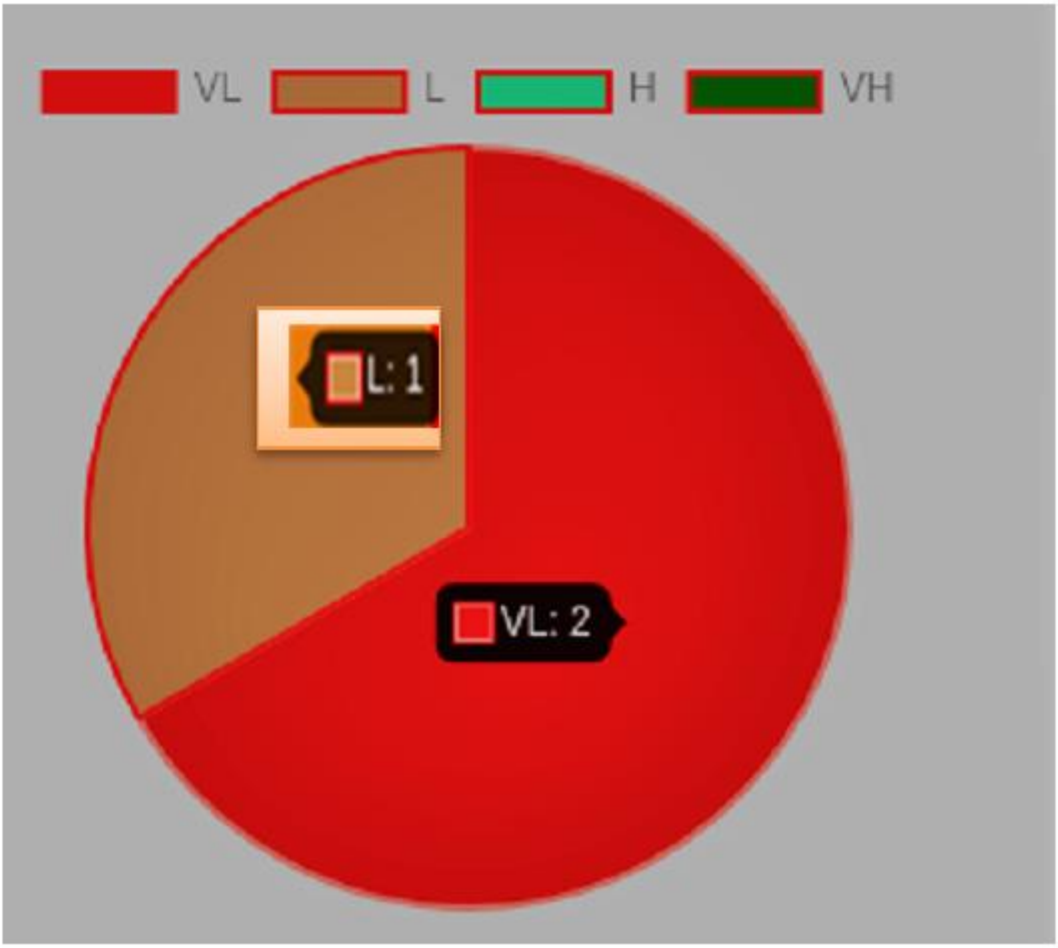


Figure 5.13c: Pie chart visualization of instantaneous engagement level of three students

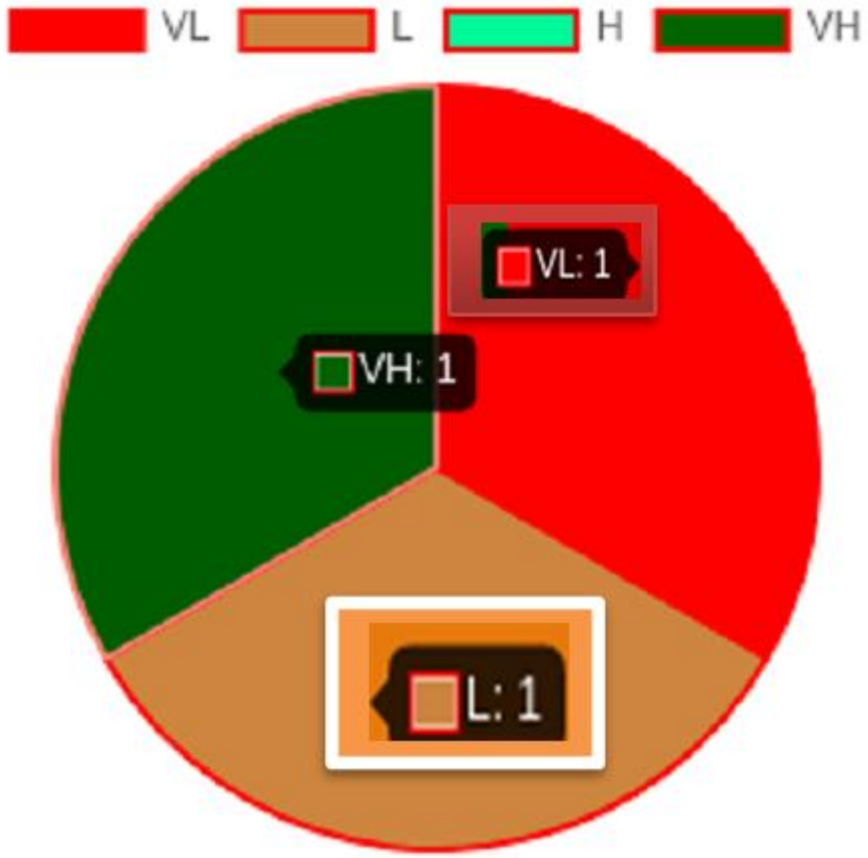


Figure 5.13d: Pie chart visualization of instantaneous engagement level of three students

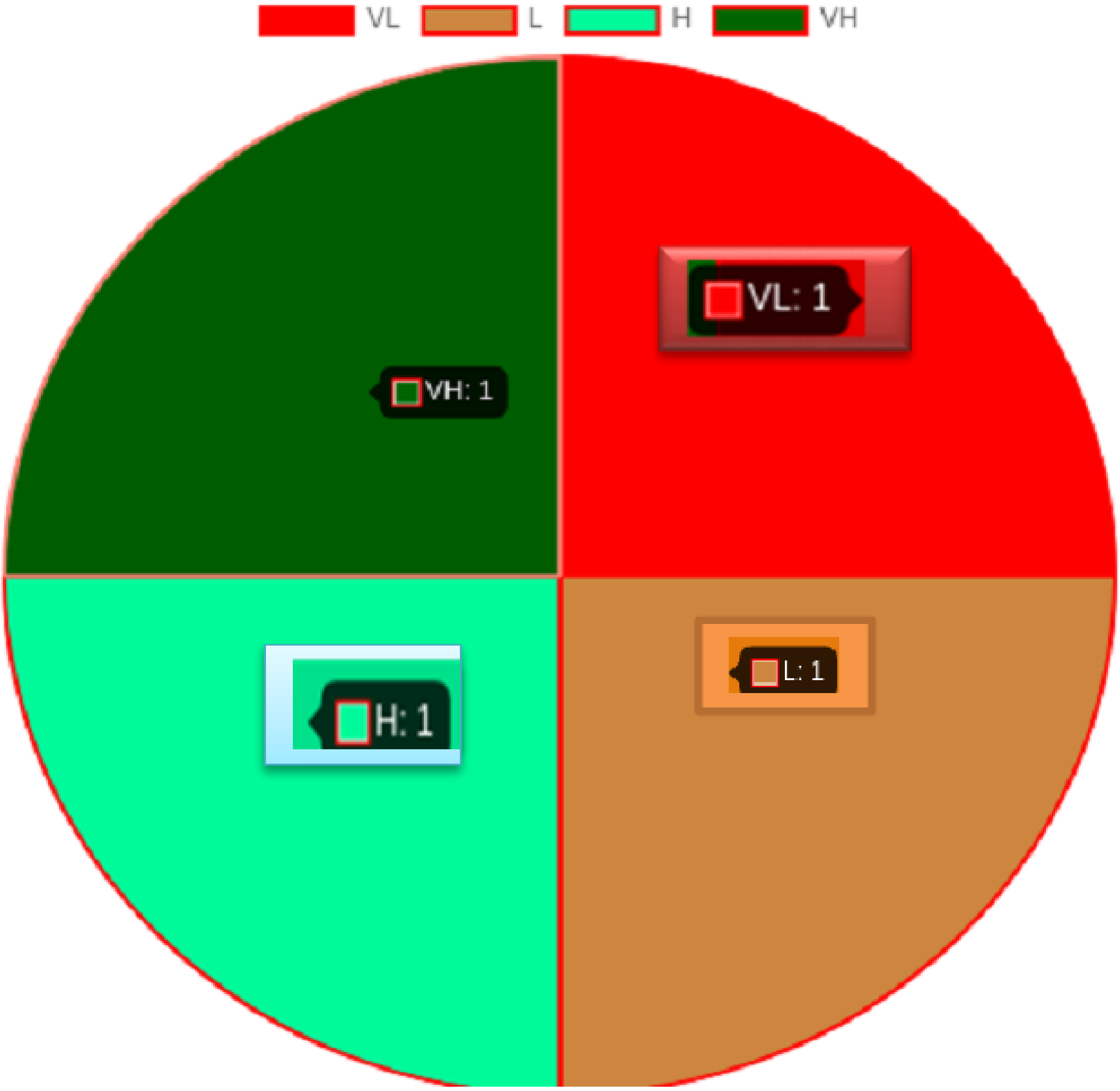


Figure 5.13e: Pie chart visualization of instantaneous engagement level of four students

A pie chart is recommended to visualize a limited number of proportions to avoid readability issues [Sedrakyan et al., 2019] Sedrakyan, G., Mannens, E. and Verbert, K., 2019. Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages, 50, pp.19-38.. The pie chart we implemented visualized the number of students experiencing the four engagement levels every minute dynamically. This allows the teacher to gain insight into the number of students experiencing the engagement levels at a glance.

5.2.2. Engagement Level Trend Visualization

Engagement level trend visualization refers to displaying the number of students experiencing the four engagement levels for the last 30 minutes. We used a line graph to visualize the number of students with the engagement levels. The line graph is appropriate for visualizing trend over time. Line graph visualization can be used to support awareness of progress during specified periods of time [Sedrakyan et al., 2019] Sedrakyan, G., Mannens, E. and Verbert, K., 2019. Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages, 50, pp.19-38.. We plotted the number of students with the engagement levels in the y-axis and the time of visualization in a minute on the x-axis. The visualization of the number of students with the engagement levels is shown in Figure 5.14 below.

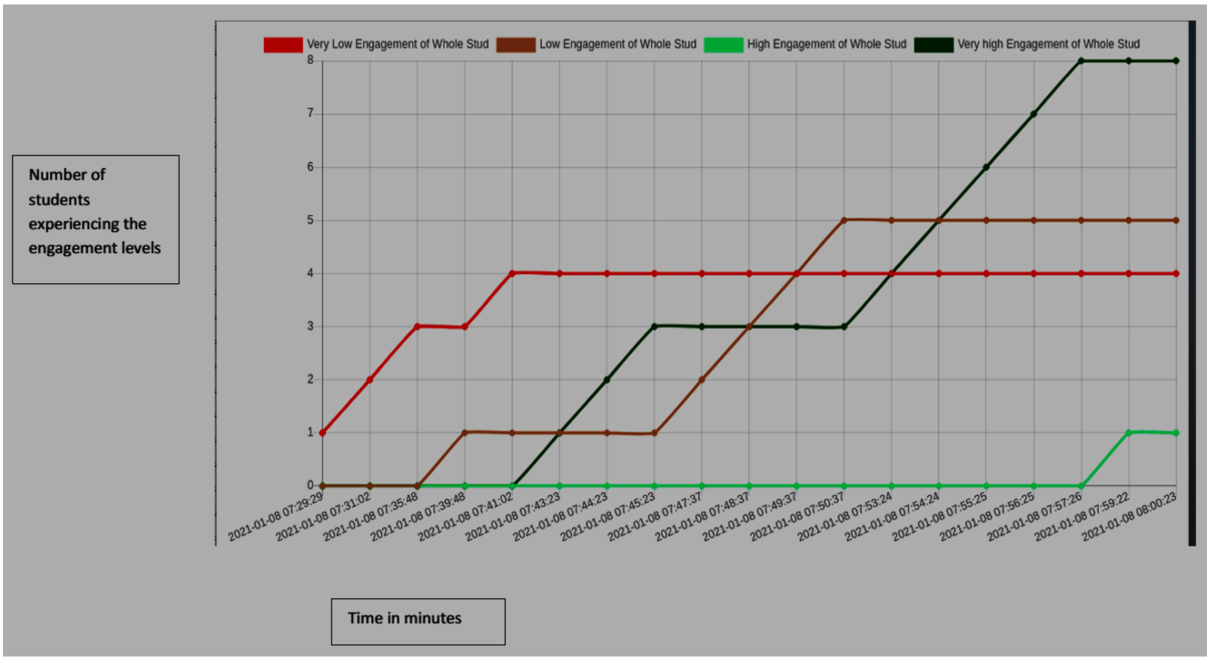


Figure 3.14: The visualization of the number of students with the engagement levels for the last 30 minutes

As it is observed from the visualization of the number of students experiencing the engagement levels in Figure 5.14, at the first minute (07:29pm), there is only one student and his engagement level was predicted to be very low(VL). Thus we can read this information from the Figure 5.13a as VL=1 at 07:29. After two minutes (07:31), the graph in Figure 5.14 shows there are two students. However, both of them were experiencing very low engagement. Thus we can see VL=2. After ten minutes (07:39), we can observe that three students were found to be experiencing very low engagement and one student experiencing low engagement. After twenty minutes (07:49), we could observe that three students were experiencing very high, four students were experiencing very low and another four low engagement levels. After 30 minutes of interaction (07:59), we could observe from the graph that there were 8 students with very high level of engagement, 5 students with high level of engagement, 4 students with very low level of engagement, and one student with high level of engagement. Moreover, the number of students experiencing very high level of engagement increases from 3 at 7:50 to 8 at 7:57. We need to notice the difference between the number of students visualized after thirty minutes with the instantaneous engagement visualization (4 students) and the trend visualization (18 students). Because trend visualization visualizes what has been stored from the last thirty minutes. However, the instantaneous visualization visualizes the instant engagement within one minute.

5.2.3. Least Engaged Learner Visualization

Least engaged learner visualization refers to displaying through filtering the lists of those students who are critically disengaged for the last 30 minutes for further investigation. The need to visualize the most disengaged learners through filtering is because instructors are overwhelmed by data reports provided to them in the online courses [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18.. The visualization was done through textual format [Mazza et al., 2012] . The list of the least engaged learners is in the order of decreasing disengagement level. The top ones are the most disengaged learners who need immediate intervention. We counted the number of the very low (VL) state and low (L) engagement states of each of the learners whose engagement levels were found in these two levels for 30 minutes. We then sorted the frequency of the engagement states. The one with the maximum was assumed to be the least engaged learner. The textual form of visualization for the least engaged learner was shown below in Figure 5.15b.

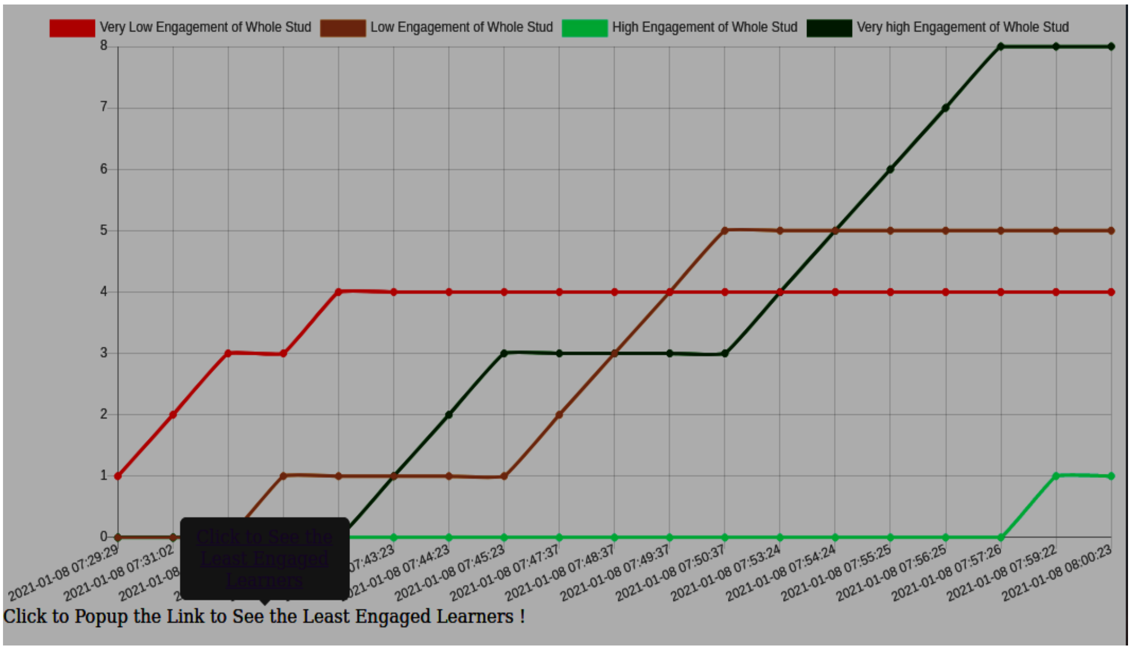


Figure 5.15a: Visualizing a popup and a link to see the least engaged learners through textual format

Userid	Firstname	Lastname	Engagement Level Predicted	Number of Low or Very low engagement levels(States)(Frequency)
21	MOHAMMAD GOLAM	ABDULAUADIR	L	7
13	Sibu	Sibu	VL	5
7	Sayyed	Nayyabrasul	L	4
22	Tufa	Feyisa	L	4
17	Abdulwahid	Abdulwahid	VL	3
14	Abdisa	Abdisa	VL	2
20	Masresha	Masresha	L	2
23	Mabratu	Mabratu	VL	1

Figure 5.15b: Visualizing least engaged learners through textual format

5.2.4. Dashboard Implementation

The student engagement visualization was developed using technologies such as chart.js, jQuery and PHP. We explained each of these technologies below.

- **Display Data in Pie Chart Using PHP and jQuery**

A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. It is generally used when one has to display comparison or difference of some data in a graphical manner. So in this subsection we will show the steps how to display data in pie chart using PHP, jQuery and chartJS.

Step 1: Make a PHP file to display data in pie chart

We make a PHP file to connect to the Moodle database in the MariaDB, retrieve relevant data and encode the array using json_encode() function.

Step 2: Add chartJS API and jQuery

AJAX request will be sent to the PHP by the jQuery to read student engagement data from the Moodle database. This helps to display pie chart using predefined function to load packages and then gets our encoded array and then create an object of chartJS pie chart to draw pie chart. The packages are the chartJS and jQuery libraries which were installed by us.

• **Chart.js**

Chart.js is an open-source, tiny, fast, easy to use, library supporting six chart types: doughnut, pie, polar, line, bar and radar. Chart.js uses HTML5 Canvas [Caldarola and Rinaldi, 2017] Caldarola, E.G. and Rinaldi, A.M., 2017. Big Data Visualization Tools: A Survey. *Research Gate*..

We accessed it from https://github.com/chartjs/Chart.js/releases/tag/v2.9.3 and installed the version: v2.9.3. A simplified API is one benefit to the canvas-based chart.js [Heckel, 2013] Tim Heckel, Canvas-Based Chart.js Version 0.1 Released, accessed from https://www.infoq.com/news/2013/03/chartjs-v.0.1-released.

• jQuery

jQuery is a lightweight, "write less, do more", JavaScript library. The purpose of jQuery is to make it much easier to use JavaScript on our website. jQuery takes a lot of common tasks that require many lines of JavaScript code to accomplish, and wraps them into methods that we can call with a single line of code. jQuery also simplifies a lot of the complicated things from JavaScript, like AJAX calls and DOM manipulation. We installed the jQuery library of version v3.5.0. The jQuery library contains the following features: HTML/DOM manipulation, CSS manipulation, HTML event methods, Effects and animations, AJAX and Utilities.

• AJAX

Ajax is a set of web development techniques using many web technologies on the client-side to create asynchronous web applications. With Ajax, web applications can send and retrieve data from a server asynchronously without interfering with the display and behaviour of the existing page. AJAX stands for Asynchronous JavaScript And XML. AJAX is not a programming language. AJAX just uses a combination of: A browser built-in XMLHttpRequest object to request data from a web server and JavaScript and HTML DOM to display or use the data. AJAX allows web pages to be updated asynchronously by exchanging data with a web server behind the scenes.

The purpose of the XMLHTTPRequest object is to allow JavaScript to formulate HTTP requests and submit them to the server. Traditionally programmed web applications normally make such requests synchronously, in conjunction with a user-initiated event such as clicking on a link or submitting a form, resulting in a new or updated page being served to the browser.

Using XMLHTTPRequest, however, you can have your page make such calls asynchronously in the background, allowing you to continue using the page without the interruption of a browser refresh and the loading of a new or revised page. This capability underpins all Ajax applications, making the XMLHTTPRequest object the key to Ajax programming.

In addition to the visualization of instantaneous engagement level, we applied AJAX for communicating the home page of our Learning Management System (Moodle) with the facial emotion recognition tool. The ajax application allowed to transfer the rate of emotion calculated every minute to a PHP file where it will be made to be stored in the database of Moodle. We created our own table in the Moodle database to store the rate of emotion which will be retrieved later for predicting the engagement levels.

- **PHP**

The PHP Hypertext Preprocessor (PHP) is a programming language that allows web developers to create dynamic content that interacts with databases. PHP is basically used for developing web based software applications. PHP is free to download and use. We installed PHP 7.2.24. The PHP code we implemented was used to connect to MariaDB server for storing the rate emotion every minute which is accessed through the AJAX application. We also used the PHP code to calculate the engagement level ranges based on the prediction model we developed in chapter 4. The prediction ranges for a particular user is stored in to Moodle's database, in a table created by us.

• MariaDB

MariaDB is a community-developed, commercially supported fork of the MySQL relational database management system (RDBMS). MaraiDB is free to download and use. We downloaded and installed MariaDB 10.1.47 version. The MariaDB is used with PHP. MariaDB intended to maintain high compatibility with MySQL. We installed the LMS(Moodle) which used the database built with Mysql (moodle) into our MariaDB server. We used the database moodle in the MariaDB server. We used the database of Moodle(LMS) which is named moodle in our system. Our data are stored in tables of moodle database. A table is a collection of related data, and it consists of columns and rows. We used one table of moodle database, which is mdl_user that stores the users of Moodle (LMS). The remaining five tables were created by us in the moodle database. These tables were: a table to store the rate of emotions, a table that stored the prediction ranges, three tables that store visualization related data. The relevant codes are put in Appendix A.

5.3. Validation of the Visualizer

To measure the usability of the visualizer in this study, the System Usability Scale (SUS) explained by [Barnum, 2011] Barnum, C.M., 2011. Usability testing essentials: ready, set--test/Carol Barnum. Burlington, MA: Morgan Kaufmann Publishers,.was utilized due to its simplicity. The scale in Table 5.28 is made up of ten items on a five-point Likert scale from Strongly Disagree to Strongly Agree, with a high score indicative of greater usability. This scale results in a single number representing an overall level of usability perceived by the end-user. The even-numbered items are positive statements, and the odd-numbered items are negative statements. This alternation is done to balance the responses.

System Usability Score (SUS) statements used in the study
I think that I would like to use this visualizer frequently during the online engagement monitoring as a teacher.
I found the visualizer unnecessarily complex during the online engagement monitoring.
I thought the visualizer was easy to use during the online engagement monitoring.
I think that I would need the support of a technical person to be able to use this visualizer during the online engagement monitoring.
I found the various functions in this visualizer were well-integrated.
I thought there was too much inconsistency in this visualizer.
I would imagine that most people would learn to use this visualizer very quickly.
I found the visualizer very cumbersome to use
I felt very confident using the visualizer during the online engagement monitoring.
I needed to learn a lot of things before I could get going with this visualizer during the online engagement monitoring.

Table 5.28. The SUS-based questionnaires used to collect ratings from the participants

5.3.1. Participants

There were 10 university students who participated in this study. All were males. The participants' age ranged from 24 to 42. The average age was 32. All of them were volunteers.

5.3.2. Procedure

We created a simulated online learning setting to perform the study. In the setting, one participant acted as a “teacher” and the remaining nine participants were the “students”. Each “teacher” took fifteen minutes in front of the screen, interacting with the three visualizers. The role of the “students” was to interact with an LMS and with a face tracking tool. At the end of the fifteen minutes interaction, the teacher filled the System Usability Scale (SUS) test for measuring satisfaction.

5.3.3. Results and Analysis

To measure the user satisfaction of the proposed dashboard, we collected and analysed the ratings by the participants on the SUS questionnaire, shown in Figure 5.16. We analysed the ratings according to the rules explained by Barnum, (2011). As Figure 5.16 shows, the minimum SUS score is 75. A System Usability Scale (SUS) test for measuring satisfaction achieved an average score of 89.5. The score indicates that the user satisfaction of the visualizer was high. Based on these findings, we concluded that this visualizer appeared to be useful for teachers during the learning process.

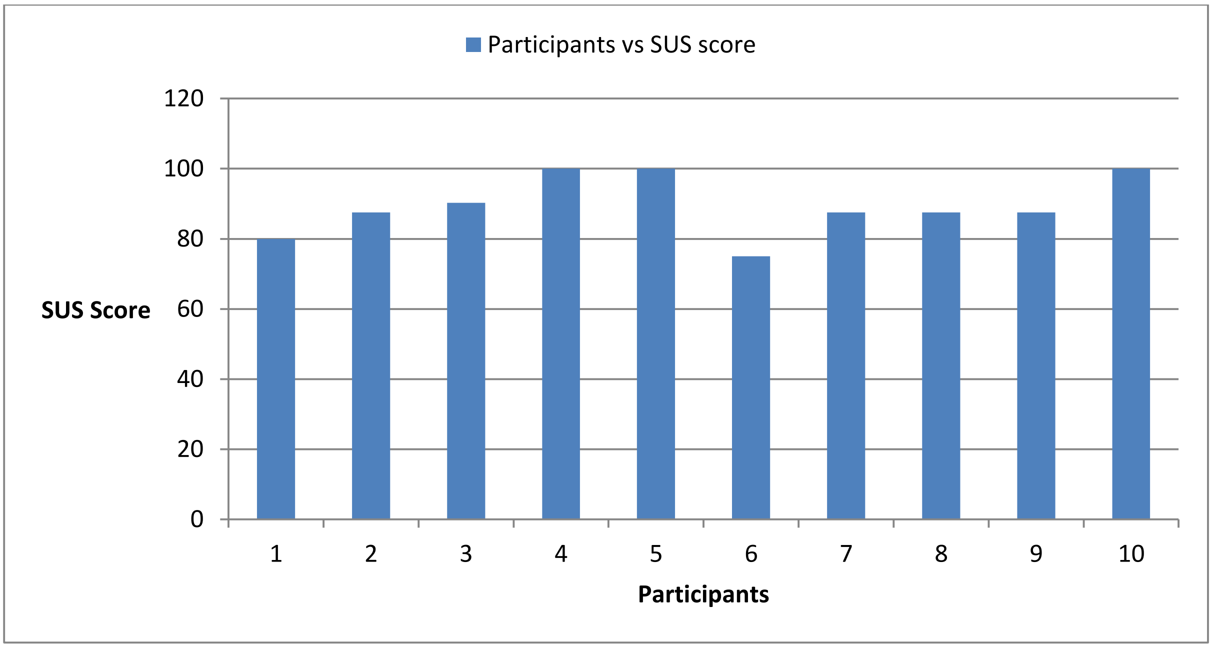


Figure 5.16: The SUS ratings obtained in the experiment

5.4. Discussion

In our study, we visualized engagement levels that were predicted from a model. The model predicted a student engagement into one of the four engagement levels every minute. These levels were very high (VH), high (H), low (L) and very low

(VL) engagement levels. One of the benefits of the visualizer that we built was it worked seamlessly with the existing LMS which is Moodle. This is a benefit because people prefer to use tools they know [Ward et al., 2010] Ward, M.O., Grinstein, G. and Keim, D., 2010. Interactive data visualization: foundations, techniques, and applications. CRC Press.. There were three types of visualizations implemented.

The first was instantaneous student engagement visualization. We visualized the number of students experiencing each of the four engagement levels every minute. We used pie chart visualization and displayed the instantaneous engagement levels of all students every minute dynamically. The size of the sector is proportional with the number of students experiencing the engagement levels. There are four sectors of the pie chart as there are four levels of engagement. Each of the sectors also displayed four different colours. The variation of the colours was according to the intensity of engagement levels. Accordingly, for very high levels of engagement we used deep green as this level is very accepted and aspired level. We used light green for high level of engagement. For low level of engagement, we used light red and we used dark red for very low levels of engagement.

The second visualization was engagement levels trend visualization. We visualized the trend of engagement levels of all students for the last 30 minutes. We used line graph to visualize the trend of the engagement levels of all students. We plotted the number of students experiencing the engagement levels in the y-axis and the time of visualization in minute on the x-axis. The visualization of the trend of the engagement levels is shown in terms of the number of students experiencing the engagement levels.

Instructors are overwhelmed by data reports provided to them in the online courses [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18.. Such challenges were solved in the literature by classifying the student in to different classes of engagement levels using the model they built. [Coffin et al., 2014] Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G., 2014, March. Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 83-92).applied a model and classified students into three categories: auditors, active and qualified and visualized the outputs of the model predictions. However, they did not consider further classifying or filtering students in one of these categories. After classifying a learner as auditors, there may be a large number of students in this particular category. In our work, we identified the least engaged learner from such larger numbers of students in the categories of low and very low engagement levels. We have not seen any work that reported such finding. We visualized the least engaged learners through filtering the lists of those students who are critically disengaged for the last 30 minutes for further investigation. The visualization was done through textual format [Mazza et al., 2012] . The list of the least engaged learners is in the order of decreasing disengagement level. The top ones are the most disengaged learners who need an immediate intervention. We counted the number of the very low (VL) state and low (L) engagement states of each of the learners whose engagement levels were found in these two levels for a period 30 minutes. We then sorted the frequency of the engagement states. The one with the maximum was assumed to be the least engaged learner.

The goal of the instantaneous visualization was allowing teachers to gain insight about the engagement levels of all students at a glance. This will allow the teacher to take immediate action. The goal of the trend visualization was to allow the teacher to develop greater understanding of the patterns of the engagement levels in the last 30 minutes or so. It will help him/her capture the changes in the engagement levels. The goal of the least engaged learner visualization was to allow the teacher gain insight about the e-learners who were critically disengaged so that a teacher will not be overwhelmed by the huge volume of engagement levels data. According to [Sedrakyan et al., 2019] Sedrakyan, G., Mannens, E. and Verbert, K., 2019. Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages, 50, pp.19-38., the visualizers will help the teachers to adapt or improve an instructional design. Our visualizer also will help the instructor to improve an instructional design to reengage the students. For instance, the teacher shall gain insight into the engagement patterns of the students, such that many students fall in the category of very low or low engagement levels. This can help the teacher to improve the instructional strategy so that he/she can adapt their style for better understanding and learning, leading to a better teaching as well as learning experience.

We also evaluated the usability of these visualizations in controlled experiment and found out that the perceived usefulness by the teachers was high. A System Usability Scale (SUS) test for measuring satisfaction achieved an average score of 89.5.

Our choice to visualize the engagement levels each minute has brought a challenge. We were forced to ignore the behavioural and collaboration features as they tend to be equal to zero as the sampling becomes every minute.

One limitation of this study is that the effect of this visualizer on learning outcomes was not evaluated. Moreover, the real-life case studies that evaluate the impact of this visualizer on learning effectiveness and efficiency have not been carried out.

In future work, we will investigate the impact of this visualizer on learning outcomes and on learning effectiveness and efficiency. We also plan to study how to reengage a disengaged student in future.

5.5. Chapter Summary

In this chapter, there were three types of visualizations implemented. The first was instantaneous student engagement visualization. We visualized the number of students experiencing each of the four engagement levels every minute. We used pie chart visualization and displayed the instantaneous engagement levels of all students every minute dynamically. The size of the sector is proportional with the number of students experiencing the engagement levels. Each of the sectors also displayed four different colours. The second visualization was engagement levels trend visualization. We visualized the trend of engagement levels of all students for the last 30 minutes. We used line graph to visualize the trend of the engagement levels of all students. We plotted the number of students experiencing the engagement levels in the y-axis and the time of visualization in minute on the x-axis. We visualized the least engaged learners through filtering the lists of those students who are critically disengaged for the last 30 minutes for further investigation. We also evaluated the usability of this visualizer in controlled experiment and found out that the perceived usefulness by the teachers was high. A System Usability Scale (SUS) test for measuring satisfaction achieved an average score of 89.5. In future work, we will investigate the impact of this visualizer on learning outcomes and on learning effectiveness and efficiency. We also plan to study how to reengage a disengaged student in future. Conclusion and Future Scope is presented in the subsequent chapter.

Chapter 6

Conclusion and Future Scope

We performed empirical study to determine the significant features that affected a given level of engagement. The significant features were applied in building the student engagement prediction model with high accuracy. We implemented visualization of the instantaneous engagement levels every minute, visualization of trends of student engagement levels and filtering and displaying the least engaged learner. This chapter concludes the thesis with the summary of the thesis, besides a discussion on the limitation and opportunities for future research.

6.1. Summary of Thesis

The finding of the analysis from the empirical study indicated that from 13 features, only 11 were significant for the four levels of student engagement. From the 11 significant features, only 4 were found to be the most important features.

The most important feature to affect very low level of engagement was the Time of assignment submission (TA) which was a behavioural feature when compared to emotional feature that was surprise (SUR). The most important feature to affect low levels of student engagement was surprise (SUR) which was emotional feature when compared with another collaboration feature that was time between post and reply (TPR) and three behavioural features which were time of assignment submission (TA), time of reading content (TRC) and score of quiz (SC). Happy (HAP) was the most important emotional feature that affected high level of student engagement when compared with other three emotional features such as sadness

(SAD), anger (ANG) and surprise (SUR) emotions and other collaboration and behavioural features. The most important feature to affect very high level of student engagement was number of replies (NR) which was collaboration feature when compared with other emotional features that were anger emotion (ANG), surprise (SUR) and happy emotion (HAP) and four behavioural features which were Time of assignment submission (TA), Time to read content (TRC), and Number of content view (NCV) and score of quiz (SC), and one collaboration feature which was Time between post and reply (TPR).

This thesis also presents a student engagement prediction model using 9 features that were significant out of 13 to affect the levels of student engagement and emerged in the final models. We built the student engagement prediction model using the features through non-linear regression technique. The 4 student engagement levels were very low engagement level (VL), low engagement level (L), high engagement level (H) and very high engagement level (VH). The three factors were behavioural, collaboration and emotional, and measured from interaction with an LMS and facial emotion recognition tool.

Moreover, we built a student engagement prediction model from three factors namely behavioural, collaboration and emotional factors as engagement is a multifaceted construct. Our student engagement prediction model predicted student engagement levels in smaller time scale that is 5 minutes with more than 83% accuracy. Providing students with support and guidance as soon as possible to lessen the danger of disengagement is critical.

One emotional feature that is surprise (SUR) and two behavioural features which are Time to read content (TRC) and score of quiz (SC) were found to be indicators of lack of engagement. This finding suggests that these features should be monitored to allow intervention at appropriate times.

The other finding of this study was that two collaborative features which are Time in the forum (TF) was significant in predicting high and Number of replies (NR) was significant in predicting both high and very high levels of engagement. This finding suggests that these two collaborative features should be supported to lead students to high and very high levels of student engagement in asynchronous online learning. Time between post and reply played little part in predicting student engagement.

We performed validation of the results of the study. We determined the accuracy of identifying students with discrete levels of engagement. The proposed model was able to correctly predict the engagement levels of 10 students out of 12. The accuracy of the model was found to be 83.3%.

There were three types of visualizations implemented. The first was instantaneous student engagement visualization. We visualized the number of students experiencing each of the four engagement levels every minute. We used pie chart visualization and displayed the instantaneous engagement levels of all students every minute dynamically. The size of the sector is proportional with the number of students experiencing the engagement levels. Each of the sectors also displayed four different colours. The second visualization was engagement levels trend visualization. We visualized the trend of engagement levels of all students for the last 30 minutes. We used line graph to visualize the trend of the engagement levels of all students. We plotted the number of students experiencing the engagement levels in the y-axis and the time of visualization in minute on the x-axis. We visualized the least engaged learners through filtering the lists of those students who are critically disengaged for the last 30 minutes for further investigation. We also evaluated the usability of this visualizer in controlled experiment and found out that the perceived usefulness by the teachers was high. A System Usability Scale (SUS) test for measuring satisfaction achieved an average score of 89.5.

We have identified a few limitations, further investigation of which may strengthen our work, and at the same time, may open up new directions for future research. We discuss the identified shortcomings and future research scope as follows.

6.2. Future Research Scopes

In future work, allowing students to sit in separate rooms to simulate distance learning where isolation could be felt easily would be very important. Regarding content, allowing the participants to take real course which will have grades as part of the syllabus could have more accurate result.

We determined the accuracy of identifying students with discrete levels of engagement. The proposed model was able to correctly predict the engagement levels of 10 students out of 12. The accuracy of the model was found to be 83.3%. However, the accuracy was not greater than 83.3%, because of the fact that the students were unable to accurately distinguish and report their actual level of engagement through the self-report questionnaire. For future, identifying the four levels of engagements namely VERY LOW (VL), LOW (L), HIGH (H), VERY HIGH (VH) from video frames could be done automatically by software. One of the finding of our work is that the emotional feature “fear” did not correlate with any of the engagement levels. In future, the above conclusion may be re-verified when a participant will take a real course which has grades as part of syllabus.

The model was based on 9 significant features, but not the most important features. Further study can be performed to determine if same prediction results can be obtained with the most important features. Future research could consider other technologies such as mobile devices. Future research might also analyse other factors of engagement such as cognitive and social engagement factors.

In future work, we will investigate the impact of this visualizer on learning outcomes and on learning effectiveness and efficiency. We also plan to study how to reengage a disengaged student in future. Moreover, in future the following could be performed: i) Considering use of Brain signal reader (e.g. EEG signals) to compare with the prediction of the model to get better accuracy of the model. ii) Considering 'Teacher-Student Interaction' to measure engagement of a student in e-learning context. iii) Considering Social and Cognitive engagement states of students (in addition to the Behavioural, Collaboration and Emotional states).

Furthermore, the usability of the whole system (in terms of learnability, memorability, effectiveness, efficiency, and user satisfaction) should also be tested empirically once it will be deployed for academic purposes in real-time. These may also lead to open new research avenues for multidisciplinary researchers in the future.

Bibliography

1. [Al-Alwani, 2016] Abdulkareem Al-Alwani, 2016,A Combined Approach to Improve Supervised E-Learning using Multi-Sensor Student Engagement Analysis,American Journal of Applied SciencesAccessed on 25 July 2019

2. [Altuwairqi et al., 2018] Altuwairqi, K., Jarraya, S.K., Allinjawi, A. and Hammami, M., 2018. A new emotion–based affective model to detect student's engagement. Journal of King Saud University-Computer and Information Sciences.

3. [Motz et al., 2019] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA, 10 pages. https://doi.org/10.1145/3303772.3303789

4. [Badge et al., 2012] Badge, J.L., Saunders, N.F. and Cann, A.J., 2012. Beyond marks: new tools to visualise student engagement via social networks. Research in Learning Technology, 20.

5. [Bahraini et al.,2016] Bahreini, K., Nadolski, R. and Westera, W., 2016. Towards multimodal emotion recognition in e-learning environments. Interactive Learning Environments, 24(3), pp.590-605.

6. [Baker and Rossi, 2013] Baker, R.S. and Rossi, L.M., 2013. Assessing the disengaged behaviors of learners. Design recommendations for intelligent tutoring systems, 1, p.153.

7. [Barnum, 2011] Barnum, C.M., 2011. Usability testing essentials: ready, set--test/Carol Barnum. Burlington, MA: Morgan Kaufmann Publishers,.

8. [Bastian et al. ,2011] Bastian M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media

9. [Beck, 2004] Beck, J.E., 2004, August. Using response times to model student disengagement.In*Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments* (Vol. 13, p. 20).

10. [Bodily et al., 2017] Bodily, R., Graham, C.R. and Bush, M.D., 2017. Online learner engagement: Opportunities and challenges with using data analytics. *Educational Technology*, pp.10-18.

11. [Bosch, 2016] Bosch, N., 2016, July. Detecting student engagement: human versus machine. In Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization (pp. 317-320). ACM.

12. [Bosse et al., 2013] Bosse, T., Gerritsen, C., de Man, J. and Treur, J., 2013, November. Learning emotion regulation strategies: A cognitive agent model. In 2013

IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 2, pp. 245-252). IEEE.

13. [Brooks and Thompson, 2017] Brooks, C. and Thompson, C., 2017. Predictive modelling in teaching and learning. Handbook of learning analytics, pp.61-68.

14. [Brown,2001] Brown, A.M., 2001. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet.Computer methods and programs in biomedicine, 65(3), pp.191-200.

15. [Bruce Ratner, 2003] Bruce Ratner - Statistical Modeling and Analysis for Database Marketing_ Effective Techniques for Mining Big Data-Chapman and Hall_CRC (2003)

16. [Calvo and D'Mello, 2010] Calvo, R.A. and D'Mello, S., 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. IEEE Transactions on affective computing, 1(1), pp.18-37.

17. [Caldarola and Rinaldi, 2017] Caldarola, E.G. and Rinaldi, A.M., 2017. Big Data Visualization Tools: A Survey. *Research Gate*.

18. [Carrillo et al., 2016] Carrillo, R., Lavoué, É. and Prié, Y., 2016, April. Towards qualitative insights for visualizing student engagement in web-based learning

environments. In Proceedings of the 25th International Conference Companion on World Wide Web (pp. 893-898).

19. [Chao et al, 2008] Chao, Y.C.E., Zhao, Y., Kupper, L.L. and Nylander-French, L.A., 2008. Quantifying the relative importance of predictors in multiple linear regression analyses for public health studies.*Journal of occupational and environmental hygiene*, *5*(8), pp.519-529.

20. [Cocea and Weibelzahl, 2011] Cocea, M. and Weibelzahl, S., 2011. Disengagement detection in online learning: Validation studies and perspectives. IEEE transactions on learning technologies, 4(2), pp.114-124.

21. [Coffin et al., 2014] Coffrin, C., Corrin, L., de Barba, P. and Kennedy, G., 2014, March. Visualizing patterns of student engagement and performance in MOOCs. In Proceedings of the fourth international conference on learning analytics and knowledge (pp. 83-92).

22. [D'Errico et al., 2016] D’Errico F., Paciello M., Cerniglia L. (2016), When emotions enhance students’ engagement in e-learning processes, Journal of e-Learning and Knowledge Society, v.12, n.4, PP. 9-23. ISSN: 1826-6223, e-ISSN:1971-8829
23. [Dewan et al., 2019] Dewan, M.A.A., Murshed, M. and Lin, F., 2019. Engagement detection in online learning: a review. Smart Learning Environments, 6(1), pp.1-20.

24. [Dixson, 2015] Dixson, M.D., 2015. Measuring student engagement in the online course: The Online Student Engagement scale (OSE). *Online Learning*, *19*(4), p.n4.

25. [D'Mello et al., 2017] D'Mello, S., Dieterle, E. and Duckworth, A., 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning.Educational psychologist, 52(2), pp.104-123.

26. [Economides, 2013] Economides, T., 2013, October. The state of the art in educational technology. In 2013 IEEE Global Humanitarian Technology Conference (GHTC) (pp. 285-287). IEEE.

27. [Edmunds and Hartnett, 2014] Edmunds, B. and Hartnett, M., 2014.Using an online Learning Management System to personalise learning for primary students.Journal of Open, Flexible, and Distance Learning, 18(1), pp.11-29.

28. [Torun, 2013] Emel Dikbas Torun, 2013, Synchronous Interaction in Online Learning Environment with Adobe Connect Pro,Procedia-Social and Behavioural Science 106(2013)2492-2499,Science Direct

29. [Falkner and Falkner, 2012] Falkner, N.J. and Falkner, K.E., 2012, September. A fast measure for identifying at-risk students in computer science. In Proceedings of the ninth annual international conference on International computing education research (pp. 55-62).

30. [Garrison, 2011] Garrison, D.R., 2011. E-learning in the 21st century: A framework for research and practice. Taylor & Francis.

31. [GhasemAghaei et al., 2016] GhasemAghaei, R., Arya, A. and Biddle, R., 2016, June. A dashboard for affective e-learning: data visualization for monitoring online learner emotions. In EdMedia+ Innovate Learning (pp. 1536-1543). Association for the Advancement of Computing in Education (AACE).

32. [Ginda et al., 2019] Ginda, M., Richey, M.C., Cousino, M. and Börner, K., 2019. Visualizing learner engagement, performance, and trajectories to evaluate and optimize online course design. PloS one, 14(5), p.e0215964.

33. [Gunes and Pantic, 2010] Gunes, H. and Pantic, M., 2010. Automatic, dimensional and continuous emotion recognition. International Journal of Synthetic Emotions (IJSE), 1(1), pp.68-99.

34. [Hamish Coates (2007)] Hamish Coates (2007) A model of online and general campus-based student engagement, Assessment & Evaluation in Higher Education, 32:2, 121-141, DOI: 10.1080/02602930600801878

35. [Hancock and Zubrick, 2015] Hancock, K.J. and Zubrick, S., 2015.Children and young people at risk of disengagement from school.Commissioner for Children and Young People, Western Australia.

36. [Harry, 2002] Harry, K. ed., 2002. Higher education through open and distance learning. Routledge.

37. [Henrie et al., 2015] Henrie, C.R., Halverson, L.R. and Graham, C.R., 2015. Measuring student engagement in technology-mediated learning: A review.*Computers& Education*, *90*, pp.36-53.

38. [Hu and Li, 2017] Hu, M. and Li, H., 2017, June. Student engagement in online learning: A review. In *2017 International Symposium on Educational Technology (ISET)* (pp. 39-43).IEEE.

39. [Hussain et al., 2018] Hussain, M., Zhu, W., Zhang, W. and Abidi, S.M.R., 2018. Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores.Computational intelligence and neuroscience, 2018.

40. [Jeon, 2017] Jeon, M., 2017. Emotions and affect in human factors and human–computer interaction: Taxonomy, theories, approaches, and methods. In *Emotions and affect in human factors and human-computer interaction* (pp. 3-26). Academic Press.

41. [Jerk-O-Meter - MIT Media], Jerk-O-Meter - MIT Media https://groupmedia.media.mit.edu

42. [Kaur et al, 2018] Kaur, A., Mustafa, A., Mehta, L. and Dhall, A., 2018, December.Prediction and localization of student engagement in the wild. In *2018 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE.

43. [Kizilcec et al., 2013] Kizilcec, R.F., Piech, C. and Schneider, E., 2013, April. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179).

44. [Kung-Keat and Ng, 2016] Kung-Keat, T. and Ng, J., 2016. Confused, bored, excited? An emotion based approach to the design of online learning systems. In *7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings* (pp. 221-233).Springer, Singapore.

45. [Leo et al., 2009] Leo, T., Manganello, F., Pennacchietti, M., Pistoia, A. and Chen, N.S., 2009, July. Online synchronous instruction: Challenges and solutions. In 2009 Ninth IEEE International Conference on Advanced Learning Technologies (pp. 489-491).IEEE.

46. [Li et al., 2016] Li, J., Ngai, G., Leong, H.V. and Chan, S.C., 2016. Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics.ACM SIGAPP Applied Computing Review, 16(3), pp.37-49.

47. [Liu et al., 2014] Liu, Y., Zumbo, B.D. and Wu, A.D., 2014. Relative importance of predictors in multilevel modeling. *Journal of Modern Applied Statistical Methods*, *13*(1), p.2.

48. [Liu et al., 2014] Liu, M., Calvo, R.A., Pardo, A. and Martin, A., 2014. Measuring and visualizing students' behavioral engagement in writing activities. IEEE Transactions on learning technologies, 8(2), pp.215-224.

49. [Manseras et al., 2018] Manseras, R., Eugenio, F. and Palaoag, T., 2018, March. Millennial Filipino Student Engagement Analyzer Using Facial Feature Classification. In IOP Conference Series: Materials Science and Engineering (Vol. 325, No. 1, p. 012006). IOP Publishing.

50. [Manwaring et al., 2017] Manwaring, K.C., Larsen, R., Graham, C.R., Henrie, C.R. and Halverson, L.R., 2017. Investigating student engagement in blended learning settings using experience sampling and structural equation modeling.The Internet and Higher Education, 35, pp.21-33.

51. [Mazza et al., 2012] Mazza, Riccardo and Bettoni, Marco and Faré, Marco and Mazzola, Luca (2012) MOCLog – Monitoring Online Courses with log data. In: 1st Moodle Research Conference (MRC2012).PP. 132-139

52. [Milborrow et al., 2010] Milborrow, S., Morkel, J. and Nicolls, F., 2010. The MUCT landmarked face database. Pattern Recognition Association of South Africa, 201(0).

53. [Moubayed et al., 2018] Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. IEEE Access, 6, pp.39117-39138.

54. [Nortvig et al., 2018] Nortvig, A.M., Petersen, A.K. and Balle, S.H., 2018. A Literature Review of the Factors Influencing E-Learning and Blended Learning in Relation to Learning Outcome, Student Satisfaction and Engagement. *Electronic Journal of e-Learning*, *16*(1), pp.46-55.

55. [Khazan, (2014] Olga Khazan (2014) This App Reads Your Emotions on Your Face. The Atlantihttps://www.theatlantic.com/technology/archive/2014/01/this-app-reads-your-emotions-on-your-face/282993/ Accessed 25 November 2019.

56. [Øygard, 2017] Øygard A. M.(2017), Released on Sep 12, 2017 on GitHub. https://github.com/auduno/clmtrackr.

57. [Park & Bonk, 2007] Park & Bonk, 2007,Synchronous Learning Experiences: Distance and Residential Learners' Perspectives in a Blended Graduate Course, Journal of Interactive Online Learning

58. [Pekrun and Linnenbrink-Garcia, 2012] Pekrun, R. and Linnenbrink-Garcia, L., 2012.Academic emotions and student engagement. In *Handbook of research on student engagement* (pp. 259-282).Springer, Boston, MA.

59. [Pesare et al., 2016.] Pesare, E., Roselli, T. and Rossano, V., 2016. Visualizing student engagement in e-learning environment. In 22th International Conference on Distributed Multimedia Systems (DMS) (pp. 26-33).

60. [Peter Bruce and Andrew Bruce, 2017] Peter Bruce, Andrew Bruce - Practical statistics for data scientists_ 50 essential concepts-O'Reilly (2017)

61. [Pireva et al., 2015] Pireva, K., Imran, A.S. and Dalipi, F., 2015, August. User behaviour analysis on LMS and MOOC. In 2015 IEEE Conference on e-Learning, e-Management and e-Services (IC3e) (pp. 21-26). IEEE.

62. [Rabbany et al., 2014] Rabbany, R., Elatia, S., Takaffoli, M. and Zaïane, O.R., 2014. Collaborative learning of students in online discussion forums: A social network analysis perspective. In Educational data mining (pp. 441-466).Springer, Cham.

63. [Ray and Chakrabarti, 2012] Ray and Chakrabarti (2012), Design and Implementation of Affective E-Learning Strategy Based on Facial Emotion Recognition,Proceedings of the InConINDIA 2012, AISC 132, pp. 613–622.

64. [Razuri et al., 2013] Rázuri, J.G., Sundgren, D., Rahmani, R. and Cardenas, A.M., 2013, November. Automatic emotion recognition through facial expression analysis in merged images based on an artificial neural network. In 2013 12th Mexican International Conference on Artificial Intelligence (pp. 85-96). IEEE.

65. [Redmond et al.,2018)] Redmond, P., Abawi, L.A., Brown, A., Henderson, R. and Heffernan, A., 2018. An online engagement framework for higher education. Online learning, 22(1), pp.183-204.

66. [Reeve et al., 2004] Reeve, J., Jang, H., Carrell, D., Jeon, S. and Barch, J., 2004. Enhancing students' engagement by increasing teachers' autonomy support. Motivation and emotion, 28(2), pp.147-169.

67. [Robal et al., 2018] Robal, T., Zhao, Y., Lofi, C. and Hauff, C., 2018. Towards Real-time Webcam-based Attention Tracking in Online Learning. In ACM Annual Meeting of Interactive User Interfaces (IUI). PP. 189–197

68. [Ruhe and Zumbo, 2008] Ruhe, V. and Zumbo, B.D., 2008. Evaluation in distance education and e-learning: The unfolding model. Guilford Press.

69. [Rumberger and Rotermund, 2012] Rumberger, R.W. and Rotermund, S., 2012. The relationship between engagement and high school dropout. In Handbook of research on student engagement (pp. 491-513). Springer, Boston, MA.

70. [Sadeque et al., 2015] Sadeque, F., Solorio, T., Pedersen, T., Shrestha, P. and Bethard, S., 2015, September. Predicting continued participation in online health forums. In Proceedings of the Sixth International Workshop on Health Text Mining and Information Analysis (pp. 12-20).

71. [Salazar, 2010] Salazar, J., 2010. Staying connected: Online education engagement and retention using educational technology tools. *American Society for Clinical Laboratory Science*, *23*(3 Supplement), pp.53-58.

72. [Samara et al., 2019] Samara, A., Galway, L., Bond, R. and Wang, H., 2019. Affective state detection via facial expression analysis within a human–computer interaction context. Journal of Ambient Intelligence and Humanized Computing, 10(6), pp.2175-2184.

73. [Sarder, 2014] Sarder, M. B. (2014, June), Improving Student Engagement in Online Courses Paper presented at 2014 ASEE Annual Conference & Exposition, Indianapolis, Indiana. 10.18260/1-2—20611. PP. 24.719.1 - 24.719.17

74. [Sarsa and Escudero, 2016] Sarsa, J. and Escudero, T., 2016.A Roadmap to Cope with Common Problems in E-Learning Research Designs.*Electronic Journal of E-learning*, *14*(5), pp.336-349.

75. [Sedrakyan et al., 2019] Sedrakyan, G., Mannens, E. and Verbert, K., 2019. Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. Journal of Computer Languages, 50, pp.19-38.

76. [Shailaja and Sridaran, 2014] Shailaja, J. and Sridaran, R., 2014, March. Taxonomy of e-learning challenges and an insight to blended learning. In 2014 International Conference on Intelligent Computing Applications (pp. 310-314).IEEE.

77. [Sharma et al, 2019] Sharma, P., Joshi, S., Gautam, S., Filipe, V. and Reis, M.J., 2019.Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning.arXiv preprint arXiv:1909.12913.

78. [Soloviev, 2018] Soloviev, V., 2018. Machine learning approach for student engagement automatic recognition from facial expressions. Scientific Publications of the State University of Novi Pazar Series A: Applied Mathematics, Informatics and mechanics, 10(2), pp.79-86.

79. [Stott, 2016] Stott, P., 2016. The perils of a lack of student engagement: Reflections of a “lonely, brave, and rather exposed” online instructor. *British Journal of Educational Technology, 47*(1), pp.51-64.

80. [Teimouri, 2018] Teimouri, Y., 2018. Differential roles of shame and guilt in L2 learning: How bad is bad?. The Modern Language Journal, 102(4), pp.632-652.

81. [Thomas and Jayagopi, 2017] Thomas, C. and Jayagopi, D.B., 2017, November. Predicting student engagement in classrooms using facial behavioral cues. In Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education (pp. 33-40).

82. [Heckel, 2013] Tim Heckel, Canvas-Based Chart.js Version 0.1 Released, accessed from https://www.infoq.com/news/2013/03/chartjs-v.0.1-released

83. [Wang and Degol, 2014] Wang, M.T. and Degol, J., 2014. Staying engaged: Knowledge and research needs in student engagement. Child development perspectives, 8(3), pp.137-143.

84. [Ward et al., 2010] Ward, M.O., Grinstein, G. and Keim, D., 2010. Interactive data visualization: foundations, techniques, and applications. CRC Press.

85. [Whitehill et al., 2014] Whitehill, J., Serpell, Z., Lin, Y.C., Foster, A. and Movellan, J.R., 2014. The faces of engagement: Automatic recognition of student engagementfrom facial expressions. IEEE Transactions on Affective Computing, 5(1), pp.86-98.

86. [Whittaker et al., 2002] Whittaker, T.A., Fouladi, R.T. and Williams, N.J., 2002. Determining predictor importance in multiple regression under varied correlational and distributional conditions. *Journal of Modern Applied Statistical Methods*, *1*(2), p.44.

87. [Woolf et al., 2009] Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D. and Picard, R., 2009. Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3-4), pp.129-164.

88. [Woodfine et al., 2008] Woodfine, B.P., Nunes, M.B. and Wright, D.J., 2008. Text-based synchronous e-learning and dyslexia: Not necessarily the perfect match!.Computers& Education, 50(3), pp.703-717.

89. [You, 2016] You, J.W., 2016. Identifying significant indicators using LMS data to predict course achievement in online learning.*The Internet and Higher Education*, *29*, pp.23-30.

90. [Zhang and Nunamaker, 2003] Zhang, D. and Nunamaker, J.F., 2003. Powering e-learning in the new millennium: an overview of e-learning and enabling technology. Information systems frontiers, 5(2), pp.207-218.

Appendix A: Dashboard Implementation Source Code

I. AJAX code for sending rate of emotion to PHP file

```
    var str=0;//for id of the user
    var sco=0;//score of surprise
    var sco2=0;//score of happy
    var sco3=0;//score of sad
    var sco4=0;//score of anger

function sendUseridwzEmo(str) {
  if (str == "") {
    document.getElementById("txtHint").innerHTML = "";
    return;
  } else {
    if (window.XMLHttpRequest) {
      // code for IE7+, Firefox, Chrome, Opera, Safari
      xmlhttp = new XMLHttpRequest();
    } else {
      // code for IE6, IE5
      xmlhttp = new ActiveXObject("Microsoft.XMLHTTP");
    }
    xmlhttp.onreadystatechange = function() {
      if (this.readyState == 4 && this.status == 200) {
        document.getElementById("txtHint").innerHTML = this.responseText;
      }
    };
   xmlhttp.open("GET","data.php?qDIS="+str+"&s3="+sco3  +"&s2="+sco2  +
"&s1="+sco + "&s4="+sco4,true);
    xmlhttp.send();
  }
}
```

II. PHP file for storing rate of emotion, determining and storing engagement prediction ranges and storing visualization related data

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN">

<meta http-equiv="expires" content="Sun, 01 Jan 2014 00:00:00 GMT"/>
<meta http-equiv="pragma" content="no-cache" />
<html>
<body>
<?php
header("Cache-Control: no-store, no-cache, must-revalidate, max-age=0");
header("Cache-Control: post-check=0, pre-check=0", false);
header("Pragma: no-cache");

$qDIS = intval($_GET['qDIS']);//for recieiving userid for inserting into
emotionscore
echo "id for sur, happy, anger or sad:  \n".$qDIS;
$s1=$_GET['s1'];//for accepting emotion score of surprise
echo "Surprise:  \n".$s1  ;

$s2=$_GET['s2'];//for accepting emotion score of happy
echo "Happy:  \n".$s2;

$s3=$_GET['s3'];//for accepting emotion score of sad
echo "Sad:  \n".$s3;

$s4=$_GET['s4'];//for accepting emotion score of anger
echo "Anger:  \n".$s4;
$userideng=$qDIS;

$engpredVL=0;
$engpredVL=0.02*$s1+0.23;//VL engagement calculated from surprise
$engpredVL=(float)($engpredVL);

echo "engpredVL:  \n".$engpredVL  ;

$engpredL=0;
$engpredL=0.06*$s1+0.3;//because L engagement is also due to surprise only
$engpredL=(float)($engpredL);
echo "engpredL:  \n".$engpredL;//because L engagement is also due to surprise
only

$engpredH=0;
$engpredH=0.03*$s4+0.01*$s1-0.123*$s2+0.43*$s3+7.2;//because H
engagement is also due to 4 emotions-sur-hap-sad-anger
$engpredH=(float)($engpredH);
```

```
echo "engpredH:  \n".$engpredH;//because H engagement is also due to 4
emotions-sur-hap-sad-anger

$engpredVH=0;
$engpredVH=0.02*$s4-0.09*$s1+0.09*$s2+74;//because VH engagement is also
due to 3 emotions-sur-hap-anger
$engpredVH=(float)($engpredVH);

echo "engpredVH:  \n".$engpredVH;//because VH engagement is also due to 3
emotions-sur-hap-anger

$hostname = "localhost";
$username = "moodleuser";
$password = "Allahuakbar";
$db = "moodle";
$dbconnect=mysqli_connect($hostname,$username,$password,$db);

if ($dbconnect->connect_error) {
  die("Database connection failed: " . $dbconnect->connect_error);
}

?>
<table border="1" align="center">
<tbody valign="top">
<tr>

  <td>Username</td>
  <td>Firstname</td>
  <td>Lastname</td>
</tr>
</tbody>
<?php

$queryuser = mysqli_query($dbconnect, "SELECT username,firstname,lastname
FROM mdl_user WHERE id='$qDIS'")
   or die (mysqli_error($dbconnect));

while ($row = mysqli_fetch_array($queryuser)) {
  echo
  "<tr>

  <td>{$row['username']}</td>
   <td>{$row['firstname']}</td>
   <td>{$row['lastname']}</td>
   </tr>\n";

}
```

```
$query = "INSERT INTO
emotionscore2(userid,emoscore,emotiontype,created,recstarttime,recendtime)
  VALUES ('$qDIS', '$s1', '$s2', '$s3', '$s4', '$s1')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting emotion score');
    } else {
      echo "emotion score successfully inserted.\n";
    }

 $t=time();
echo($t . "<br>");
echo(date("h:i:sa",$t));

if($engpredVL>=0 && $engpredVL<=2)
{
$engpred="";
$engpredval=0;
$engvar1="VL";
$engvar1val=$engpredVL;
$engpred="VL";
echo $engpred;
  $engpredval=$engpredVL;
    echo  $engpredval;

}

else if($engpredL>2 && $engpredL<=7)
{
$engpred="";
$engpredval=0;
$engvar2="L";
$engvar2val=$engpredL;
$engpred="L";
echo $engpred;
  $engpredval=$engpredL;
   echo  $engpredval;

}
else if($engpredH>7 && $engpredH<=62)
{
$engpred="";
$engpredval=0;
$engvar3="H";
$engvar3val=$engpredH;
$engpred="H";
echo $engpred;
   $engpredval=$engpredH;
   echo  $engpredval;
```

```
}

else if($engpredVH>62 && $engpredVH<=254)
{
$engpred="";
$engpredval=0;
$engvar4="VH";
$engvar4val=$engpredVH;
$engpred="VH";
echo $engpred;
  $engpredval=$engpredVH;
    echo  $engpredval;
}

else
{

$engpred="NOENG";
$engpredval=0;
    echo $engpred;
   echo  $engpredval;
}

$query = "INSERT INTO
engagementprediction(userid,predstarttime,predendtime,dateofpred,engpred,engpr
edval,engvar1,engvar1val,engvar2,engvar2val,engvar3,engvar3val,engvar4,engvar
4val)
  VALUES ('$qDIS', '$t', '$s1', '$s2', '$engpred', '$engpredval', '$engvar1',
'$engvar1val', '$engvar2', '$engvar2val', '$engvar3', '$engvar3val', '$engvar4',
'$engvar4val')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting engagementprediction');
    } else {
      echo "Engagementprediction successfully inserted.\n";
    }

$query = mysqli_query($dbconnect, "SELECT * FROM engagementprediction ")
   or die (mysqli_error($dbconnect));
   echo
  "<tr>
  <td>epid</td>
  <td>userid</td>
   <td>time of eng</td>
   <td>emoscore-happy</td>
   <td>emoscore-ang</td>
   <td>engpred</td>
   <td>engpredval</td>
```

```
  <td>engvar1</td>
  <td>engvar1val</td>
  <td>engvar2</td>
  <td>engvar2val</td>
  <td>engvar3</td>
  <td>engvar3val</td>
  <td>engvar4</td>
  <td>engvar4val</td>
  </tr>\n";

while ($row = mysqli_fetch_array($query)) {
 echo
 "<tr>
 <td>{$row['epid']}</td>
 <td>{$row['userid']}</td>
  <td>{$row['predstarttime']}</td>
  <td>{$row['predendtime']}</td>
  <td>{$row['dateofpred']}</td>
  <td>{$row['engpred']}</td>
  <td>{$row['engpredval']}</td>
  <td>{$row['engvar1']}</td>
  <td>{$row['engvar1val']}</td>
  <td>{$row['engvar2']}</td>
  <td>{$row['engvar2val']}</td>
  <td>{$row['engvar3']}</td>
  <td>{$row['engvar3val']}</td>
  <td>{$row['engvar4']}</td>
  <td>{$row['engvar4val']}</td>
  </tr>\n";

}

//this is for the instantaneous visualization i.e. number of stud experiencing each
of the four levels in pie chart

$countstudVH=0;$countstudH=0;$countstudL=0;$countstudVL=0;$totcountstud=
0;$engpredid=[];

$query = mysqli_query($dbconnect, "SELECT DISTINCT userid,engpred FROM
engagementprediction,mdl_user WHERE
engagementprediction.userid=mdl_user.id AND
FROM_UNIXTIME(predstarttime) > now()-interval 10 minute")
  or die (mysqli_error($dbconnect));

//$data=array();
while ($row = mysqli_fetch_array($query)) {

  if($row['engpred'] =="VH")
```

```
  $countstudVH=$countstudVH+1;
   if($row['engpred'] =="H")
  $countstudH=$countstudH+1;
   if($row['engpred'] =="L")
  $countstudL=$countstudL+1;
   if($row['engpred'] =="VL")
  $countstudVL=$countstudVL+1;
  $totcountstud=$totcountstud+1;

     $engpredid=$row['epid'];
 }

$t=time();
//here also insert number of students
$query = "INSERT INTO engpercentstud2(timecreated,engpred,percentstud,epid)
  VALUES ('$t', 'VH', '$countstudVH', '$engpredid')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting engpercentstud2');
    } else {
      echo "engpercentstud2 successfully inserted.\n";
    }

    $query = "INSERT INTO
engpercentstud2(timecreated,engpred,percentstud,epid)
  VALUES ('$t', 'H', '$countstudH', '$engpredid')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting engpercentstud2');
    } else {
      echo "engpercentstud2 successfully inserted.\n";
    }
    $query = "INSERT INTO
engpercentstud2(timecreated,engpred,percentstud,epid)
  VALUES ('$t', 'L', '$countstudL', '$engpredid')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting engpercentstud2');
    } else {
      echo "engpercentstud2 successfully inserted.\n";
    }
    $query = "INSERT INTO
engpercentstud2(timecreated,engpred,percentstud,epid)
  VALUES ('$t', 'VL', '$countstudVL', '$engpredid')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting engpercentstud2');
    } else {
      echo "engpercentstud2 successfully inserted.\n";
    }
```

```
  $query = mysqli_query($dbconnect, "SELECT * FROM engpercentstud2 ")
 or die (mysqli_error($dbconnect));

  //this is for line graph of trend of stud number with each of the engagement
level for the whole students(number of students vs time)

$countstudVH=0;$countstudH=0;$countstudL=0;$countstudVL=0;$totcountstud=
0;$timecreated=[];
  $studnumberVL=0; $studnumberL=0;$studnumberH=0;$studnumberVH=0;
  $engpredVH=""; $engpredH="";$engpredL="";$engpredVL="";

$query = mysqli_query($dbconnect, "SELECT DISTINCT
userid,predstarttime,engpred FROM engagementprediction,mdl_user WHERE
engagementprediction.userid=mdl_user.id AND
FROM_UNIXTIME(predstarttime) > now()-interval 30 minute")
 or die (mysqli_error($dbconnect));

//$data=array();
while ($row = mysqli_fetch_array($query)) {
  if($row['engpred'] =="VH")
 $countstudVH=$countstudVH+1;
  if($row['engpred'] =="H")
 $countstudH=$countstudH+1;
  if($row['engpred'] =="L")
 $countstudL=$countstudL+1;
  if($row['engpred'] =="VL")
 $countstudVL=$countstudVL+1;

  $timecreated=$row['predstarttime'];
}

    {
    $studnumberVH=$countstudVH;
    $engpredVH="VH";
    }

  {
    $studnumberH=$countstudH;
    $engpredH="H";
    }

  {
    $studnumberL=$countstudL;
    $engpredL="L";
    }

  {
```

```
    $studnumberVL=$countstudVL;
    $engpredVL="VL";
    }

$query = "INSERT INTO
wholestudeng3(engpredVL,studnumberVL,engpredL,studnumberL,engpredH,stud
numberH,engpredVH,studnumberVH,timecreated)
  VALUES ('$engpredVL', '$studnumberVL', '$engpredL',
'$studnumberL','$engpredH', '$studnumberH','$engpredVH',
'$studnumberVH','$timecreated')";
   if (!mysqli_query($dbconnect, $query)) {
       die('An error occurred when inserting wholestudeng3');
   } else {
     echo "wholestudeng3 successfully inserted.\n";
   }


   $query = mysqli_query($dbconnect, "SELECT * FROM wholestudeng3 ")
  or die (mysqli_error($dbconnect));

 //this is for line graph of emotion vs engaement for the whole students(emotion vs
engagement in time)
  $timecreated=[];$engpredval=0;
   $emosur=0; $emohap=0;$emosad=0;$emoanger=0;
   $engpred="";  $evseuserid=0;
$query = mysqli_query($dbconnect, "SELECT
emotionscore2.userid,predstarttime,engpred,engpredval,emoscore,emotiontype,cre
ated,recstarttime FROM engagementprediction,emotionscore2 WHERE
engagementprediction.userid=emotionscore2.userid AND
FROM_UNIXTIME(predstarttime) > now()-interval 15 minute")
   or die (mysqli_error($dbconnect));

while ($row = mysqli_fetch_array($query)) {
   $engpred=$row['engpred'];
   $engpredval=$row['engpredval'];
   $evseuserid=$row['userid'];
   $timecreated=$row['predstarttime'];

   $emosur=$row['emoscore'];
   $emohap=$row['emotiontype'];
  $emosad=$row['created'];
 $emoanger=$row['recstarttime'];

}
```

```
//insert the data retrived from engagementprediction and emotionscore2 into
emotionvsengagement table

$query = "INSERT INTO
emotionvsengagement(timecreated,engpred,engpredval,emosur,emohap,emosad,e
moanger,userid)
  VALUES ('$timecreated', '$engpred', '$engpredval', '$emosur','$emohap',
'$emosad','$emoanger', '$evseuserid')";
   if (!mysqli_query($dbconnect, $query)) {
        die('An error occurred when inserting emotionvsengagement');
    } else {
      echo "emotionvsengagement successfully inserted.\n";
    }

    $query = mysqli_query($dbconnect, "SELECT * FROM emotionvsengagement
")
   or die (mysqli_error($dbconnect));


echo
  "<tr>
  <td>evseid</td>
  <td>timecreated</td>
   <td>engpred</td>
   <td>engpredval</td>
   <td>emosur</td>
   <td>emohap</td>
   <td>emosad</td>
   <td>emoanger</td>
   <td>userid</td>

     </tr>\n";

while ($row = mysqli_fetch_array($query)) {
  echo
  "<tr>
  <td>{$row['evseid']}</td>
  <td>{$row['timecreated']}</td>
  <td>{$row['engpred']}</td>
  <td>{$row['engpredval']}</td>
   <td>{$row['emosur']}</td>
  <td>{$row['emohap']}</td>
   <td>{$row['emosad']}</td>
  <td>{$row['emoanger']}</td>
  <td>{$row['userid']}</td>

      </tr>\n";
```

```
}

//$query->close();
//$dbconnect->close();
//header("refresh: 1");
?>

</table>
</body>
</html>
```

III. HTML code to load the jQuery and chartJS libraries and invoke the jQuery code that is written to send ajax request to PHP file and to visualize the engagement levels

```
<meta http-equiv="expires" content="Sun, 01 Jan 2014 00:00:00 GMT"/>
<meta http-equiv="pragma" content="no-cache" />
<!DOCTYPE html PUBLIC"-//W3C//DTD XHTML 1.0 Strict//EN">
<html>
<head>
<meta content="text/html;charset=utf-8" http-equiv="Content-Type">
<meta content="utf-8" http-equiv="encoding">
<meta http-equiv="refresh" content="60" >
<title>Visualize Instantaneous Engagement -Circle Graph-</title>
<style type="text/css">
.chart-container{
width:640px;
height:auto;
}
</style>
</head>
<body>
<H1>Instantaneous Engagement Visualization</H1>
<h2>Please wait to see the instantaneous engagement dynamically every
minute!</h2>
```

```
<div id="chart-container">
<canvas id="mycanvas"></canvas>
</div>
<!--javascript-->
<script type="text/javascript" src="js4chart/jquery.min.js"></script>
<script type="text/javascript" src="js4chart/Chart.min.js"></script>
<script type="text/javascript" src="js4chart/visualizepercent-3.js"></script>
</body>
</html>
```

IV. The jQuery code that is written to send ajax request to PHP file and to visualize the engagement levels

```
$(document).ready(function(){
$.ajax({
url:"http://localhost/moodle/my/visualizepercent2.php",
method:"GET",
success:function(data){
console.log(data);
var engpred=[];
var percentstud=[];
var coloR = [];

for(var i in data){
engpred.push(data[i].engpred);
percentstud.push(data[i].percentstud);
}
var dynamicColorsLG = function() {
            return "#00FA9A";
    };
    var dynamicColorsLR = function() {
            return "#CD853F";
    };
    var dynamicColorsDG = function() {
            return "#006400";
    };
    var dynamicColorsDR = function() {
            return "#FF0000";
    };
coloR.push(dynamicColorsDR());
coloR.push(dynamicColorsLR());
coloR.push(dynamicColorsLG());
coloR.push(dynamicColorsDG());
```

```
var chartdata={
labels: engpred,
datasets:[
{
label:"Instantaneous engagement of the four levels",
backgroundColor: coloR,//dark green'#006400',//dark green
borderColor:'#FF0000',//dark red
hoverBorderColor:'#FA8072',//light red

data:percentstud
}
]
};
var ctx=$("#mycanvas");
var bargraph= new Chart(ctx,{
type:'pie',
data: chartdata
});

},
error:function(data){
console.log(data);
}
});
});
```

V. PHP file to retrieve visualization related data and encode into JSON

```
<?php
header('Content-Type:application/json');
$hostname = "localhost";
$username = "moodleuser";
$password = "Allahuakbar";
$db = "moodle";
$dbconnect=mysqli_connect($hostname,$username,$password,$db);
if ($dbconnect->connect_error) {
  die("Database connection failed: " . $dbconnect->connect_error);
}
```

```
 $queryrows = mysqli_query($dbconnect, "SELECT perid,engpred,percentstud
FROM engpercentstud2 ORDER BY perid DESC LIMIT 4;
")
   or die (mysqli_error($dbconnect));

   $data=array();
   foreach($queryrows as $row){
   $data[]=$row;
   }
    print json_encode($data);
   ?>
```

Publications

We published two papers and submitted one for publications with the journals and a conference below.

1. Wakjira A, Bhattacharya S, *predicting student engagement in the online learning environment*, International Journal of Web-Based Learning and Teaching Technologies (IJWLTT) (Published)
2. Wakjira A, Bhattacharya S, *Student engagement awareness dashboard in an asynchronous e-learning environment*, Lect. Notes in Networks, Syst., Vol. 321, Milan Tuba et al. (Eds): ICT Systems and Sustainability, 978-981-16-5986-7, 511567_1_En, (Chapter 74)) (Passed e-proofing)
3. Wakjira A, Bhattacharya S, Identifying the most important factors affecting student engagement in e- learning environment, CSI transaction on ICT (Revised and resubmitted)

Vitae

1. Personal Information

Name:	Abdalganiy Kebede Wakjira
Place of Birth:	Horro Guduru Wallagaa, Oromia, Ethiopia
Date of Birth:	January 5, 1979
Marital Status:	Married
Nationality:	Ethiopian
Contact Address:	Cell phone: +251937705583/+251900726686
Email:	aabiidk@gmail.com

2. Education Background

- PhD Researcher (Thesis submitted): June 2021
  - Indian Institute of Technology Guwahati
  - User centric design, e-learning, student engagement

- M. Sc. in Information Technology from Oct 2009-Oct 2011
  - Sikkim Manipal University, India in A grade

- B. Sc. in Computer Science from Sep 2004 –Feb 2008
  - Haramaya University with GPA of 3.94 out of 4

- B.Ed. degree in Mathematics from Sep 1997 –July 2001
  - Bahir Dar University with GPA of 2.31 out of 4

- Post graduate diploma in Higher Education Teaching from Sep 2013-July 2014
  - Haramaya University

3. Work Experience

3.1. Haramaya University
From November 5, 2012 – Present
- **Lecturer in the department of Software Engineering**

Taught courses in the undergraduate program such as Data Structure and Algorithms**,** Computer Networking and Data Communication**,** Java Programming, Client Server Computing, Software Project Management

3.2. Addis Ababa Institute of Technology (AAiT)
From January 2012 to November 2012
- **ICT Head**

Major duty of the position

Support AAiT in the introduction of Software tools for modern administration, organize staff training, Lead the team of ICT administrators and computer lab attendants, Control the overall network administration, Supervise the staff of the unit.

3.3. Finchaa Sugar Factory

From January 2011 to November 2011

- **Management Information System Service Head**

Major duty of the position

Coordinate the divisional IT support, Manage the LAN, Work as database administrator, Work as system administrator, Manage and coordinate the ongoing implementation of integrated MIS

3.4. Finchaa Sugar Factory

From July 2008 to December 2010

- **Network administrator**

Major duty of the position

Install and configure workstations, Troubleshoot basic LAN problems, Repair desktop computers, Work as system administrator, Work as database administrator

3.5. Haramaya University

From September 2007 to July 2008

- **Network Administrator**

Major duty of the position

Manage day to day computer support and networking problems, Install and configure workstations, Troubleshoot basic LAN problems

3.6. St.Theresa School, Diredawa

From September 2003 to August 2007

- **Mathematics and physics Teacher (Grades 5, up to 10)**

From September 2004 to August 2007

- **Mathematics Department head**

3.7. Amuru Secondary School (Grade 9 and 10)

From September 2001 to July 2003

- **Mathematics and physics teacher**

4. Training

- Microsoft Windows and Solaris System Administration training conducted by Addis Ababa University, ICT department for 2 months
- Microsoft Windows System Administration, A+ hardware, managing Cisco devices training conducted by National IT Solutions for 2 months
- Oracle database administration conducted by National IT Solutions for 2 weeks
- Introduction to FOSS and the Linux training for 3 days
- Mobile Application Programming offered by Merita Technologies on January 28 and January 29,2013

- Strategic planning and Management conducted by Jethro Leadership and Management Institute from July 26-August 1,2012

5. Hobbies

- Programming
- Reading

6. Computer Skills

Well versed in HTML,Javascript,Dreamweaver,Java,vb.net 2010, PHP, Mysql, oracle lOG client and server, crystal report 8.5, sql server, JSP, Net Beans IDE, android java

7. Language Skills

Language	Speaking	Listening	Writing	Reading
English	Excellent	Excellent	Excellent	Excellent
Afaan Oromo	Excellent	Excellent	Excellent	Excellent
Amharic	Excellent	Excellent	Excellent	Excellent

8. Projects Done

- Health Extension Management System developed with Microsoft Visual Basic 2010 and Sql Server for Harar City woreda Administration
- Java Application for health monitoring
- Web based Maths tutoring system developed through PHP, Mysql and apache
- Location aware advertising app using android
- Developed a medical service management system for Finachaa Sugar Factory using vb.net 2005 and oracle 10G
- Online sales management for Finchaa Sugar Factory using php and Mysql, through WAMP server