An Emotion Estimation Model for Cognitive State Analysis of Learners in Online Education Using Deep Learning

A

Thesis

Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

In

COMPUTER SCIENCE AND ENGINEERING

Submitted By

Maragoni Mahendar

41900106

Supervised By

Dr Arun Malik (17442)

Computer Science and Engineering (Professor)

Lovely Professional University

LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2024

DECLARATION

I, hereby declared that the presented work in the thesis entitled “An Emotion Estimation Model for Cognitive State Analysis of Learners in Online Education Using Deep Learning” in fulfilment of degree of **Doctor of Philosophy (Ph.D.)** is outcome of research work carried out by me under the supervision Dr Arun Malik, working as Professor, in the Computer Science and Engineering of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Maragoni Mahendar

Registration No.:41900106

Department/school: Computer Science and Engineering

Lovely Professional University,

Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph.D. thesis entitled “An Emotion Estimation Model for Cognitive State Analysis of Learners in Online Education Using Deep Learning” submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the Computer Science and Engineering is a research work carried out by Maragoni Mahendar, 41900106 , is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

Amalik

(Signature of Supervisor)

Name of supervisor: Dr Arun Malik

Designation: Professor

Department/school: Computer Science and Engineering

University: Lovely Professional University

Abstract

Deep learning has emerged as a powerful tool in analyzing student emotions, expressions, and learning rates in the context of online education. This innovative approach holds great potential for enhancing the overall learning experience and improving educational outcomes for students in virtual learning environments. By leveraging deep learning algorithms, the teacher can delve into the rich data generated during online classes, such as facial expressions, voice intonations, and interaction patterns, to gain insights into students' emotional states and learning progress. This allows for a more nuanced understanding of how students engage with the material and their level of comprehension, enabling timely interventions and personalized support. One key aspect of deep learning-based analysis is the interpretation of facial expressions. By employing computer vision techniques, facial recognition algorithms can identify and analyze various emotions displayed by students during online classes. This includes capturing subtle micro-expressions that may go unnoticed to the human eye. By discerning emotions such as confusion, boredom, or engagement, educators can adapt their teaching strategies and provide targeted interventions to optimize learning experiences. Furthermore, deep learning models can be trained to correlate student emotions and expressions with learning rates. By analyzing the relationship between different emotional states and the rate at which students acquire knowledge, educators can identify factors that positively or negatively impact learning outcomes. For instance, a student displaying signs of frustration may indicate a need for additional support or alternative teaching approaches, while a student exhibiting enthusiasm may indicate a high level of engagement and comprehension.

Online education platforms can also leverage deep learning to develop personalized learning environments. By continuously analyzing student emotions, expressions, and learning rates, these platforms can dynamically adjust the content, pace, and delivery of educational materials to suit individual needs. This adaptive learning approach ensures that students receive tailored instruction and support, maximizing their learning potential. In the domain of Facial Expression Recognition (FER), two significant challenges are the presence of occlusion and variations in facial position that can greatly impact the recognition process. While automated FER has made remarkable advancements in recent decades, its practical application in real-world scenarios, where robustness to occlusion and pose changes is crucial, has received comparatively less attention. This research aims to address this gap by focusing on occlusion- and pose-invariant FER for assessing the cognitive states of learners in online education.

Inspired by the attention mechanism of the human visual system, the researchers have developed a novel spatial attention network called SAN-CNN. This network highlights saliency features and spatial importance among neighboring pixels to effectively capture facial expressions. The proposed approach involves preprocessing the input image using a median contour filter and subsequently employing a mask-based region of interest (ROI) for segmentation. By utilizing convolutional neural networks (CNNs) for facial recognition, landmark localization, and head position estimation, emotional categorization becomes feasible.

The effectiveness and efficiency of the proposed approach were demonstrated using publicly available video-based facial expression datasets from Kaggle. The results indicate that the proposed SAN-CNN model outperforms traditional techniques in terms of accuracy and processing speed. By incorporating spatial attention networks into the FER pipeline, the model achieves improved performance in recognizing facial expressions while being robust to occlusion and pose variations. this study highlighted the significance of addressing occlusion and pose changes in real-world Facial Expression Recognition for assessing the cognitive states of online education learners. The introduction of the SAN-CNN spatial attention network demonstrates its potential to enhance emotional categorization by emphasizing saliency and spatial relationships. The experimental evaluation using relevant datasets validates the superiority of the proposed approach in terms of accuracy and efficiency compared to conventional methods.

The attention level of the students is analysed using the proposed deep learning model, which compiles data from the visual feed, the aural input, and the heuristic properties of the camera. When all of these components are utilised in conjunction with one another, it will be feasible to monitor the attention of the students in real time. The input video feed that originates from the student's computer has two

components: the screen sharing as well as the video feed. Both of these components make up the input video feed. Since the screen sharing is being watched, the students are not permitted to view any other forms of media while the class is in session. The video stream is utilised to do facial recognition analysis on the students. When the image is initially processed, the OpenCV face detection module is applied to it. This module looks for the student's face everywhere it can find it in the picture. When the face has been recognised, the facial angle is determined by combining MobileNetV2 with a facial landmark recognition module that was built by Google. This is done so that the results may be more accurate. The face angle is calculated based on the recovered attributes, which serve as the foundation for the computation. In addition to the characteristics of the face, the audio input is scrutinised for the presence of any background noise, as well as the heuristic inputs from the computer are also taken into account. The model that has been suggested is an all-encompassing and complete model that appropriately evaluates students' levels of attentiveness.

Micro facial expressions hold the potential to identify emotional states of students during their participation in online learning tasks. Through understanding these minute facial cues, e-learning platforms can discern and decode a range of emotions experienced by students, including happiness, sadness, frustration, confusion, and more. This information can be used to personalize the learning experience for each student, provide early intervention and support, and improve engagement and motivation. This paper presents a VGG16 Convolution Neural Network (CNN) based transfer learning model for facial micro expression recognition. The proposed model identifies the relation between micro expressions and emotions in the student faces in online classes to assess the learning rate. Emotion modelling is used in the context of online education, and it entails making use of technology and data analysis in order to identify the emotional indicators that students are exhibiting and then responding appropriately. The pretrained VGG model allows for the creation of much deeper features than previous architectures, while mitigating the vanishing gradient problem with the help of the transfer learning approach. The proposed model can learn more complex representations of the input data, which is important in recognizing subtle emotions thereby accurately assessing the learning rate of the students.

ACKNOWLEDGEMENT

With immense pleasure and deep sense of gratitude, I wish to express my sincere thanks to my supervisor Dr. Arun Malik, Professor, School of Computer Science and Engineering, Lovely Professional University. In my expedition towards this degree, I have found a teacher, an inspiration, a role model, and a pillar of support in my Guide. He has been there providing his incredible support and guidance at all times and has given me worthy counseling, suggestions, and recommendations in my quest for knowledge. It was a great privilege and honour to work and study under his guidance. Without his able guidance, this thesis would not have been possible, and I shall infinitely be indebted to him for his overwhelming assistance. I am highly obliged for his unfathomable confidence in me and providing continuous encouragement at all times.

Special thanks go to examiners of term end reports and reviewers of the journals who vetted my submissions and gave valuable comments to improve the work further.

Finally, I wish to express my profound gratitude to my mother, father, brother, husband, and all members of my family. Their incomprehensible faith in me, their sacrifices, unflappable strength, and encouragement at all times has been an inspiration and is solely responsible for refurbishing my life and profession. I pray to God and hope that He empowers me, to live up to their expectations all the time.

Date: 15/06/2024 Maragoni Mahendar

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1Overview

This objective of this chapter is to provide a thorough introduction to e-learning and the influence it has had on the educational system. An overview of the idea of online education's development over the last several years is also provided. The chapter then discusses the advantages and disadvantages of online education as well as how it affects students' levels of motivation and involvement. Additionally, some of the difficulties that are involved with e-learning are acknowledged in this chapter.

1.2 An Introduction to eLearning

The process of obtaining information and skills via the utilisation of different electronic technologies is known as e-learning, often known as online learning. Some examples of these technologies include the internet, personal computers, and mobile devices. The term "e-learning" refers to a broad variety of educational pursuits, including formally structured classes and degree programs provided by educational institutions like colleges and universities, as well as less structured learning options like online tutorials, webinars, and self-paced training modules [1].

In recent years, there has been a rise in interest in online learning due to the fact that technological advancements have made it less difficult and more readily available. Students have the ability to access educational materials and take part in learning activities from any location in the globe, at any time of the day or night, thanks to the advent of e-learning. E-learning has become a particularly appealing choice for students who are combining their education with other responsibilities, such as employment, family, or other obligations, as a result of its flexibility.

E-learning may be accomplished via a wide variety of mediums, ranging from text-based materials and online lectures to interactive simulations and virtual reality settings. Some kinds of online education make use of a variety of various sorts of media in order to keep students interested and to make it easier for them to comprehend difficult ideas. For instance, a student taking a biology class could be required to watch video lectures, participate in online exams, and do virtual lab activities in order to fully comprehend the material being covered [2].

The scalability of e-learning is recognized as one of its primary benefits. When it comes to education delivered in a conventional classroom setting, the number of pupils that can be crammed into a certain area is capped at a certain point. E-learning, on the other hand, does not impose any of these constraints. Due to the expansion of online courses, which an infinite number of students may access, education is now more accessible and more affordable for people all over the world.

E-learning also has the potential to be more individualized than the conventional learning that takes place in a classroom setting. Teachers can keep track of their students' progress and alter their classes to meet the unique needs of each one by using e-learning platforms. For instance, an e-learning platform may have adaptive learning algorithms, which modify the level of difficulty of questions depending on how well a student is doing on the platform. This helps to ensure that students are pushed without being overburdened.

Figure 1. 1: Different modes of access in E-Learning

In the area of electronic learning, there are many access models. The many ways that students may engage with educational material and resources are related by these access models. The following is a list of some of the most frequent modes of access used in e-learning environments:

- Students are able to access instructional information and resources at their own leisure while using this type of learning, which is referred to as self-paced learning. This gives students the opportunity to study at a pace that is appropriate for them, which may be especially helpful for kids who need more time to fully comprehend more complicated ideas.
- The students in this paradigm go through a course or program together as a group while following the cohort-based learning model. This may provide kids a feeling of community and support while they are learning, and it can also make it easier for students to work together and have conversations with one another.
- Blended learning is a type of education in which students take part in learning activities that are conducted both online and in-person. For instance, a class may include of lectures and assignments that are to be completed online, in addition to face-to-face conversations and collaborative projects.
- Mobile learning is a concept that allows students to access educational information and resources by using mobile devices like smartphones and tablets. Students who are required to study while on the go or who do not have access to conventional desktop or laptop computers may find mobile learning to be an especially helpful option.
- The term "gamification" refers to a paradigm in which instructional material and resources are provided in a style like that of a video game. This model includes components such as points, badges, and leaderboards. Students, especially those who may not be interested in more conventional modes of education, may find that this makes learning more interesting and, as a result, more motivating for them.

Each of these models of access has a unique set of benefits and drawbacks; the model that is most suitable will be determined by a number of criteria, including the nature of the material being accessed, the requirements and preferences of the students, and the resources that are at their disposal. Educators may assist guarantee that all students have access to high-quality educational material and resources by first gaining an awareness of the many access models that are available and then making effective use of this knowledge. This is possible regardless of the specific learning styles or life circumstances of the students.

1.2.1 A Brief History

It is important to investigate the background of eLearning in order to get a deeper comprehension of the roles that it now plays for businesses. Elliott Maisie was the first person to use the term "eLearning" in an environment that was considered to be professional in nature in the year 1999. It is generally agreed that he was the one who first used the word [3]. The years that have passed since then have seen the reputation of eLearning grow from strength to strength as time has gone by. However, the emergence of eLearning to become the most common means of giving training in the current day may be attributed to a number of factors that have contributed to its success. The following are examples of some of these factors:

- **The Internet**: A lot of people used printed manuals, CD-ROMs, and other limiting techniques for learning and training before the internet became popular. The proliferation of the internet made it possible for businesses to forgo techniques that were just one-dimensional and instead make use of the adaptability of eLearning.
- The growth of multimedia was a significant factor in the progression of e-learning. It was found that the capacity to include components such as photos, videos, music, and graphics into e-learning courses was a more dependable method of keeping students interested than conventional learning methods.
- It is not hard to comprehend the meteoric rise in popularity of eLearning when one considers that the first IBM computer sold for the equivalent of approximately $5,000 in today's money, so it is not surprising that the cost of digital devices has decreased in recent years. The development of eLearning was also significantly aided by the rise of mobile learning.

- Learning Management Systems that are Well-Built LMSs have gotten more advanced over the years, transitioning from locally installed systems to cloud-based ones. As a result, more and more businesses are using them to carry out a wide variety of training. When selecting an LMS, there are several factors to think about; at the very least, you need to make sure that it offers the features and support that you need in order to satisfy both your goals and the goals of your learners.

1.2.2 Working of E-learning

The distribution of teaching may make use of either static approaches, such as learning portals, hyperlinked pages, screen cam lessons, streaming audio/video, or live online broadcasts, or interactive methods, such as threaded dialogues, chats, or desktop video conferencing.In the early years of its development, e-learning systems were largely focused on facilitating the distribution of instructional content directly from instructor to student [4,5]. The e-learning experience has progressed to the point where it now enables greater two-way contact by making use of technologies that are ever more engaging. Students, employees, and independent learners alike have a greater level of discretion in determining how they receive and respond to knowledge presented via e-learning, and any number of their fellow peers may be involved in the activity at the same time.

There are three primary criteria that an organization should adhere to in order to assist assure a successful e-learning program, and this is independent of the precise platform that the organization ultimately decides to use:

- The experience ought to work well on mobile devices. The business world is becoming more reliant on mobile phone use. The usage of phones allows for the dissemination of training reminders, training prompts, and congratulations on the completion of training accomplishments.
- The experience needs to have social components integrated into it. Learners are provided with a platform to engage with their peers and employers, as well as reply to training efforts or general corporate updates, via the use of social media or a platform that emulates the fundamental aspects of social media.

- The experience needs to make use of a range of various media, such as quizzes, infographics, podcasts, demonstrations, and training that is narrative-based.

1.3 Recent Trends in Online learning

In recent years, there has been a discernible increase in the number of people choosing to get their education via the medium of online learning. The term "online learning," which is sometimes referred to as "e-learning" and "distance learning," describes the practice of using electronic technology to give education and training to students who are not physically present in the setting of a conventional classroom. This expansion has been propelled by a variety of causes, including developments in technology, a rise in the number of online courses that are available, and changes in the manner that individuals study [6].

The fast development of technology is one of the primary forces that is contributing to the expansion of online educational opportunities. Because of the considerable advances that have been achieved in internet connection, processing power, and digital media in the recent years, it is now simpler than ever before to provide high-quality online learning experiences. This has made it possible for instructors to employ a broad variety of multimedia resources to produce interesting and interactive online courses. Some examples of these tools are movies, audio recordings, and interactive quizzes.

The greater availability of online courses is another important aspect that is playing a significant role in the expansion of online learning. Learners today have access to a huge array of online courses that cover a wide range of subjects as a result of the expansion of online learning platforms such as Coursera, Udemy, and edX. The fact that these classes are often taught by renowned academic institutions and acknowledged authorities in their respective subjects lends both credibility and value to the programs [7].

Alterations in the ways in which individuals acquire knowledge have also played an important part in the expansion of online educational opportunities. In recent years, there has been a movement away from the conventional methods of learning, which are centered on lectures, and toward methods that are more participatory and collaborative. Online education has been able to capitalize on this trend by providing students with the option to interact with their classmates and teachers in a number of settings. These settings include online discussion forums, virtual classrooms, and group projects.

The COVID-19 epidemic has also had a substantial influence on the expansion of online educational opportunities. As a result of the forced closure of physical campuses by a significant number of educational institutions, online learning has emerged as the principal mode of instruction for an increasing number of students all over the globe. Because of this, the use of online learning platforms has risen, and this has prompted the creation of new tools and technology that allow remote learning [8,9].

1.3.1 Factors influencing the growth of E-Learning

The following is a list of the aspects that have contributed to the expansion of online learning:

1. **Quick development of technology:**As a consequence of the fast advancement of technology, it is now much less difficult to provide high-quality educational opportunities via the medium of online learning. Because of this, teachers are now able to develop online courses that are both interesting and interactive by using various forms of multimedia, such as videos, audio recordings, and interactive quizzes.
2. **An increase in the availability of number of online courses**: The expansion of online learning platforms such as Coursera, Udemy, and edX has made it simpler for students to have access to an extensive number of online courses that cover a broad range of subjects. The fact that these classes are often taught by renowned academic institutions and acknowledged authorities in their respective subjects lends both credibility and value to the programs [10].

3. **Shifts in the ways in which people learn**: There has been a movement away from more conventional forms of learning, such as listening to lectures, and toward more participatory and collaborative methods. Online education has been able to capitalize on this trend by providing students with the option to interact with their classmates and teachers in a number of settings. These settings include online discussion forums, virtual classrooms, and group projects.
4. **Adaptability**: The students are able to study at their own speed, on their own timetable, and from any location in the globe because to the adaptability that is offered by online learning. Because of this, it is now much simpler for students to pursue their educational aspirations even if they do not have access to conventional educational institutions.
5. **Cost-effectiveness:** Because there are no costs involved with travelling or maintaining physical infrastructure, online learning may often be more cost-effective than conventional classroom-based learning. This is because online learning does not need students to purchase textbooks.
6. **Accessibility:** The rise of online education has made it easier for students with physical limitations or those who live in rural locations to get a formal education, even if they do not have physical access to conventional educational institutions.
7. **The COVID-19 pandemic**: The pandemic has expedited the use of online learning platforms and stimulated the creation of new tools and technology to enable remote learning, which has made it the principal method of education for millions of learners all over the globe.

1.3.2 Advantages of E-learning

In recent years, more and more people have been opting to get their education via the medium of online learning. It is a term used to describe the practice of using various forms of electronic technology in order to provide instruction and education to students who are not physically present in the conventional classroom environment. The rise in popularity of online education may be directly attributed to the many benefits that come with participating in such programs.

Flexibility is one of the primary benefits that comes with doing schoolwork online. Because it enables students to study at their own speed and on their own timetable, online learning is well suited for those who already have full schedules due to employment or family responsibilities. Because of this flexibility, students may access the course materials and complete their assignments from any location on the planet as long as they have access to the internet. Learners who may not have access to conventional educational institutions are given an improved opportunity to realize their educational potential as a result of this [11,12].

Because there are no costs connected with travelling or maintaining physical infrastructure, online courses may often be had at a lower cost than traditionally taught classes that take place in a classroom setting. Because they can be accessible from any location in the globe, online courses eliminate the need for expensive travel and hotel fees.

Learners who participate in online education also have access to a diverse selection of available courses and other resources. Learners today have access to a huge array of online courses that cover a wide range of subjects as a result of the expansion of online learning platforms such as Coursera, Udemy, and edX. The fact that these classes are often taught by renowned academic institutions and acknowledged authorities in their respective subjects lends both credibility and value to the programs. In addition, learners often have access to a variety of materials, such as online textbooks, video lectures, and interactive quizzes, while using online learning platforms.

Learning environments that encourage collaboration and interaction may also be fostered via online education. Learners have the opportunity to connect with their instructors as well as their classmates and other students in a number of ways thanks to the virtual classrooms, online discussion forums, and collaborative projects that are included in many online courses. This kind of collaborative and interactive learning environment may assist to build a feeling of community among learners, which can be especially beneficial for learners who might be studying alone or from distant areas.

Learners have the chance to build essential skills that are in great demand in today's workforce when they participate in online learning opportunities. These abilities include computer literacy, the ability to manage one's time effectively, the capacity for self-directed learning, and the ability to solve problems. Learners may improve their chances of finding work and advancing in their careers by cultivating the aforementioned abilities via participation in online learning programs.

Online education has several benefits, including the ability to study at one's own pace and schedule, access to a vast array of courses and materials, a collaborative and interactive learning environment, and the possibility of acquiring useful new skills. In recent years, these benefits have led to the rise in popularity of online learning, making it a viable and beneficial form of education for learners all over the globe. Online learning has been more popular in recent years.

1.3.3 Challenges that students face in E-learning

Online education presents a number of opportunities for students, but they also run the risk of encountering a number of difficulties, most notably in the area of learning rates [13-16]. Concerning the pace at which they absorb information, students may face the following difficulties while participating in online learning:

1. Drawback of online education is that students and teachers don't get to interact as much as they would in a traditional classroom setting. This might make the experience seem less personal. This may be difficult for children who need more individualized attention to completely absorb new ideas, and it may slow down their pace of learning.
2. A high degree of self-motivation is required for online learning since students generally study on their own time and are not constrained by the framework of a conventional classroom. Students that have trouble managing their time effectively may have a more difficult time with this, which may slow down their pace of learning.
3. **Distractions:** When students are engaged in online education, they are more likely to get sidetracked by other activities or outside influences, such as social media, phone calls, or other responsibilities. Distractions like these may make

it difficult for kids to concentrate, which in turn slows down their pace of learning.

4. **Difficulties with technology**: Difficulties with technology, such as issues with internet access or software malfunctions, may be annoying and time-consuming, and they can have an influence on the students' ability to learn quickly.
5. **Lack of social interaction:** Because online learning may often seem rather solitary, this might have a detrimental effect on certain students' ability to retain information. The absence of a physical classroom might result in less opportunities for students to engage in social contact with their classmates, which can make the educational experience seem less engaging and less likely to inspire students to pursue their interests.
6. **Lack of organization**: The majority of the time, online learning needs more self-directed learning than conventional classroom-based learning does. This may be hard for certain students who need more structure and assistance than online learning can provide. If the student is unable to efficiently manage their time and resources, this might lead to a reduced learning rate, which can be problematic.

Although there are many positive aspects associated with online education, there are also some possible drawbacks that students need to be aware of in order to avoid slowing down their educational progress. Students may assist guarantee that they are able to fully participate in and benefit from online learning by taking responsibility for their learning and tackling the problems that come with it.

1.4Influence of eLearning on Students' Learning

In recent years, online courses have grown in popularity as a way of education, particularly in the India. Although taking classes online has numerous benefits, such as more flexibility and access to a greater variety of academic subjects and resources, it also has its own set of obstacles, notably in terms of the amount of individual attention given to each student. In this article, we will address the elements that might effect student attention in online classrooms, as well as tactics that can be employed to promote student involvement in the material being presented [17].

The lack of a traditional classroom setting is one of the primary elements that might have an effect on the level of attention paid by students enrolled in online programs. Students in a conventional classroom setting are often located in the same physical place as both their lecturer and the other students in their class. Students are more likely to remain focused and attentive if they are given the opportunity for a greater degree of involvement and engagement, which is made possible through this. Students in online courses could be learning from faraway places and might not have the same amount of engagement with their classmates and teachers as they would in traditional classrooms. Because of this, it may become more difficult to maintain kids' engagement and attention.

The quality of the course design and the way it is presented may also have an effect on the level of attention paid by students in online courses. Students may find it more challenging to maintain their attention and engagement if the online courses they are taking are badly designed or presented. Students may get disenchanted with a class if it is extremely complicated or difficult to traverse, while classes that do not have defined learning goals may be seen as confused and unmotivating. In addition, classes that depend too much on passive learning, such as viewing pre-recorded lectures, can not be as interesting as those that include more active learning activities [18].

The amount of student motivation is a third aspect that may have an effect on students' ability to pay attention in online classrooms. Students need to have a high degree of self-motivation if they are going to be successful with online learning since it allows them to study on their own time and does not have the same structure as a conventional classroom. Students who are not self-motivated to study or who do not have specific learning objectives may find it challenging to maintain attention and remain interested when taking online programs.

In spite of these obstacles, there are methods that may be used to raise the level of attention paid by students in online classrooms. Creating classes that are interesting and participatory is one tactic that might be used. This may be accomplished by include a wide range of active learning activities in the curriculum, such as online discussion forums, collaborative projects, and interactive quizzes. Students may find themselves more interested and motivated to learn if they participate in these activities.

One further tactic is offering pupils consistent feedback on their work. Students might be motivated to keep studying when they get feedback that helps them remain on track academically. In addition to providing feedback on assignments, quizzes, and examinations, teachers may also offer frequent updates on how the course is progressing and the learning goals that have been set.

Teachers in online classrooms have the ability to contribute to increased student attentiveness by encouraging a feeling of community among the students. Making use of online discussion forums, setting up virtual classrooms, and promoting student collaboration on projects are some of the ways in which this goal may be accomplished. Students have a greater chance of feeling involved in their education and driven to study when a sense of community is fostered among them.

1.4.1 Ways to measure Student Attention

The importance of determining whether or not a student is paying attention in an online class is rising in tandem with the proliferation of online education. Attention is an essential component of learning, and assessing it may assist teachers in determining areas in which pupils may be having difficulty or have lost interest in the material. In this article, we will cover various methods that may be used to gauge the attention that students in online classrooms are paying.

1. **Participation in activities**: Participation in activities such as discussion forums and group projects is an excellent sign of a student's attentiveness. It is a good sign that students are enthusiastic and involved in the subject matter of the class when they actively participate in the activities that are being offered.
2. The percentage of students who finish all of their assigned work, quizzes, and examinations is another way to evaluate a student's level of attention. It is an indication that pupils are paying attention and focusing their attention when they are able to finish these exercises on time and with a high degree of accuracy.

3. The length of time students spend engaged with the content of the class is yet another metric that may be used to evaluate how attentive students are being. Either by using monitoring software or by having students record their own time on their own, this may be accomplished. It is a good indicator that students are engaged in the content and paying attention if they are spending a large amount of time participating in the course.
4. **Interactions with teachers**: Interactions with instructors, such as asking questions or seeking feedback, may be a measure of a student's attentiveness to the material being presented. It is a good sign that kids are paying attention and are interested in learning when they demonstrate that they are actively seeking help or feedback from their teachers.
5. **Performance on examinations**: As a last indicator of student attentiveness, performance on assessments such as quizzes and exams might be considered. If students do well on these tests, it indicates that they are paying attention in class and that they are able to absorb the information being presented.

The level of engagement and attention that students bring to their studies may be gauged by the percentage of students who take part in various activities, the percentage of students who finish the course, the amount of time spent in the course, the number of contacts with instructors, and how well students do on assessments. This information may be used to identify sections of the course in which students may be having difficulty or becoming disengaged, and then modifications can be made to the content of the course or the way it is delivered as necessary.

1.4.2 Student Expression as a Measure to Evaluate Attention

It is possible to assess a student's level of attention in an online class by observing their facial expressions, which are a significant indication of both human emotion and focus. In recent years, developments in computer vision along with machine learning have made it possible to automatically analyse facial emotions. This has made it much simpler to monitor a student's attentiveness in an online class without the need for any extra gear or software. In this article, we will examine a method that may be utilized to gauge a student's attentiveness in an online class by analyzing their facial expressions [19,20].

Computer vision algorithms are used in facial expression analysis to detect and analyze facial characteristics such as the eyes, lips, and brows of a student in order to determine the emotions and attention levels that are being sent by those features. During the online lesson, video recordings of the student's face are normally taken so that the analysis may be carried out on them. The following are some of the ways that analyzing a student's facial expressions may be utilized to determine whether or not they are paying attention in an online class:

1. **Eye tracking:** Eye tracking is one of the most prominent approaches that is used to measure the attentiveness of students who are participating in online courses. Instructors are able to evaluate which aspects of the screen or the content being covered in class are more engaging or distracting to students by monitoring the direction and length of the student's attention.
2. **Recognition of facial expressions:** Software that recognizes facial expressions may be used to identify small shifts in a student's facial expression that reflect the degree of attention they are paying to the lesson. For instance, a furrowed brow or eyes that are narrowed might be seen as a sign of intense concentration, but a blank look can be interpreted as a sign of disengagement.
3. **Emotion detection:** Emotion recognition software can identify and evaluate changes in a student's facial expression that reflect emotional reactions to the course content. This information may be gleaned from the student's expression. For instance, a grin or lifted eyebrows may imply surprise or happiness, while a frown or lips that is bent downward may indicate bewilderment or irritation.
4. **Attention monitoring:** The software used for attention tracking may examine a student's facial expression over a period of time to identify the student's degree of attention over the course of the lesson. When students are attentive and involved in the lesson, teachers may tell by monitoring changes in their facial expressions and eye gazing, but they can also see when students are getting distracted or disinterested in the lesson.

There are several factors to take into consideration, despite the fact that facial expression analysis has the potential to be an effective method for determining whether or not students in online lectures are paying attention. For instance, facial expression analysis is dependent on video recordings of the student's face; however, such recordings are not always accessible and cannot be guaranteed to be accurate. In addition, pupils who have restricted facial movement, who wear glasses or other face accessories that might conceal their expression, or who have other facial characteristics that make it difficult to read their expression may not benefit from facial expression analysis.

The amount of attention given by students participating in online courses might potentially be determined via the use of facial expression analysis as a method of evaluation. An instructor may acquire insights into which sections of the course content are most engaging and how to better engage students who are becoming distracted or disengaged by studying eye gazing, facial expressions, and emotional reactions. However, it is essential to keep in mind the constraints of facial expression analysis and to make use of it in combination with other techniques when attempting to gauge the level of attention paid by students.

1.4.3Student Emotion and Learning Rate in Online Education

The term "learning rate" refers to the rate at which students are able to acquire new information and abilities. This might vary from student to student. Some students have a high learning rate, which indicates that they are able to swiftly acquire and apply new material. Other kids, on the other hand, have a lower learning rate and may need more time and practice in order to fully comprehend new ideas.

There are many different things that might have an effect on the pace of a student's learning. For instance, a student's existing knowledge and experience with a certain topic might have an effect on how fast they are able to absorb new material. This is because prior knowledge and experience can serve as a foundation for further learning. It's possible that students who already have a solid background in a topic will be able to pick up new knowledge far more rapidly than those who are beginning completely from scratch.

The level of enthusiasm and interest that a student has in the material being studied is another aspect that might have an effect on the pace at which they learn. Students who are enthusiastic and involved in what they are studying are often more driven to put in the work necessary to master new ideas, which may lead to a quicker learning rate. Students who are interested and engaged in what they are learning tend to be more motivated to put in the effort required to master new concepts.

The atmosphere of the classroom may also have an effect on the pace of a student's learning. Students may be able to learn more rapidly if the classroom environment is one that is encouraging and favorable to learning, and if the teaching techniques are ones that are both interesting and effective. In contrast, classrooms that are disorganized or that lack structure may make it more difficult for pupils to learn and can also slow down their pace of learning.

Students' emotional experiences may have a substantial influence on how quickly they learn content presented in online courses. Emotions play a vital part in the learning process. In this article, we will analyze the connection between the feelings of students in online classrooms and the pace at which they acquire new material.

To begin, it is essential to keep in mind that students may experience a wide variety of emotions as a result of taking online courses, including both good and negative feelings. Learning may be facilitated by an increase in motivation and engagement when positive feelings such as enthusiasm, curiosity, and pleasure are present. On the other hand, negative emotions such as irritation, worry, and boredom may make it difficult to learn new things since they decrease both motivation and concentration.

Positive emotions have been demonstrated to boost learning and memory by enhancing the brain's capacity to absorb information and recover previously stored memories, according to research. On the other hand, feelings of negativity might make it more difficult to learn new things and remember them because they disrupt attention, perception, and cognitive processes.

The absence of face-to-face connection and the isolation that comes with distant learning may lead to a variety of unpleasant emotions, including feelings of isolation, disengagement, and dissatisfaction. These feelings can be exacerbated in online classrooms. Negative feelings may be further exacerbated in students when they have technological issues, as well as when they do not get assistance from their teachers or classmates.

The instructors of students taking online courses have numerous options available to them in order to boost their students' rates of learning while also addressing the emotional experiences of those students. To begin, teachers have the ability to cultivate a learning atmosphere that is encouraging and welcoming to all students and that encourages the development of good emotions such as curiosity, pleasure, and connection. This may be accomplished by giving students the opportunity to communicate with one another and collaborate, as well as by providing feedback and assistance in a timely manner.

The second way that teachers may assist students in coping with unfavorable feelings such as anger and worry is by directing them to resources that can help them reduce stress and better control their feelings. This may include activities such as mindfulness exercises, approaches for relaxation, or counseling services.

In online classrooms, teachers now have the ability to employ technology to monitor and track the emotions of their students. For instance, systems that evaluate sentiment may examine the comments and posts made by students in online discussion forums in order to recognize emotional trends and locate subject matter in which students may be having difficulty or becoming disinterested.

It can be said that feelings play an important part in the process of learning, and that the emotional experiences of students may have a substantial influence on the pace at which they learn in online classrooms. The emotional experiences of students taking online courses may be helped to improve by instructors by giving tools for emotional regulation, developing learning environments that are helpful and welcoming to all students, and using technology to monitor students' emotional states. As a result, students' rates of learning can be increased.

1.5Problem Formulation

Because of the rapid development of online education, there is a growing need for reliable techniques of assessing the cognitive and emotional states of students while they are engaged in the process of learning. However, typical approaches of analyzing cognitive state and emotions, such as self-reporting or physiological markers, might be intrusive or inaccurate in some circumstances. As a result, there is a need for a technique that is both non-intrusive and reliable in order to estimate the learners' feelings and cognitive states while they are engaged in online learning.

The proposed research intends to construct an emotion estimation model that makes use of deep learning methods to assess the cognitive states and emotions of learners while they are participating in online education. This would allow for the issue to be addressed. The model will be created to analyse information acquired from several sources, including the learner's facial expressions, voice, and other behavioural indicators, to arrive at an exact evaluation of the learner's emotional and cognitive states. The suggested project intends to advance the creation of more accurate and non-intrusive techniques for analysing students' mental states and emotional states during online instruction, successfully determining the learning rate.

1.5Motivation

Online learning has been much more well-liked in recent years, particularly in the wake of the COVID-19 epidemic. Two specific issues that online learning presents for both instructors and students are the lack of face-to-face interaction and the need that learners be able to self-regulate their learning. Additionally, how motivated learners are, how engaged they are in the learning process, and the outcomes of that learning may all be significantly impacted by the mental and emotional states of learners.

In online learning settings, traditional ways of measuring learners' cognitive states and emotions, such as self-reporting or physiological tests, might be intrusive, inaccurate, or impracticable. As a result, there is a need for a way of assessing the cognitive and emotional states of students while they are participating in online education that is both more efficient and less intrusive.

This requirement will be addressed by the proposed study on an emotion estimation model that uses deep learning methods. The goal of this research is to build a system that is both reliable and non-intrusive for assessing the emotional and cognitive states of learners participating in online education. The suggested model is able to give a more accurate and complete evaluation of the cognitive states and emotions of learners since it draws data from a variety of sources, such as the learners' facial expressions, their speech, and the signs of their behavior.

1.6Objectives

- To study and analyse various emotion estimation model and techniques.
- To Prepare dataset for emotion estimation model in low quality images, Occlusion and non-frontal head pose, in order to improve class imbalances and to generate more datasets for training FER model with help of Audio and Video sequences in Elearning.
- To design and develop the emotion estimation model for cognitive state analysis of learners.
- To evaluate the designed and proposed model with standard metrics.

1.7 Thesis Organization

Chapter 1 serves as the introduction to the thesis, providing an overview of the research topic and introducing the concept of eLearning. It covers a brief history of eLearning, the working of eLearning, recent trends in online learning, factors influencing the growth of eLearning, advantages and challenges of eLearning, and the influence of eLearning on students' learning. The chapter also presents the problem formulation, motivation, objectives, and provides an outline of the thesis organization.

Chapter 2 focuses on the literature review related to eLearning, discussing the growth of eLearning and its consequences. It also delves into the detection of student attention in online classes, student facial emotion estimation, and the relationship between student facial expression and learning rate in online classes.

Chapter 3 introduces the Emotion Estimation Model for Cognitive State Analysis. It begins with an introduction and then presents the proposed model, which includes video to frame conversion, pre-processing, segmentation, and classification techniques. The chapter further explores CNN (Convolutional Neural Network) and Spatial Attention Network in the context of emotion estimation. It concludes with a performance analysis and final conclusions.

Chapter 4 focuses on student attention detection in online learning. It starts with an introduction and presents the proposed attention detection model. The chapter covers video input processing, experimental results, including video and audio processing, and concludes with a summary of findings.

Chapter 5 introduces the Micro-Expression Modelling based Facial Emotion Recognition in E-Learning using a modified ResNet50V2 model. It starts with an introduction and presents the proposed model, which includes learning rate assessment using emotion expression, emotion modelling, micro expressions, emotion, and learning rate relationship, and the ResNet50V2 architecture. The chapter includes an experimental analysis and concludes with a summary of findings.

The thesis concludes with a final chapter, providing an overall conclusion of the research, summarizing the key findings, and discussing their implications.

Chapter 2

Literature

In recent years, the area of e-learning has undergone tremendous development as a result of the increased availability of digital technology and the change towards distant learning brought about by the COVID-19 epidemic. Both of these factors contributed to the expansion of the field. As a direct consequence of this, there has been an increasing interest in gaining a knowledge of the efficacy and impact of online learning, in addition to the difficulties and possibilities that it brings. The purpose of this literature review is to offer a complete overview of the research that has been conducted in a number of different fields that are linked to e-learning. The review will pay particular attention to the growth of e-learning and its effects, the assessment of students' facial expressions, the identification of students' attentiveness in virtual classrooms, and the relationship between students' facial expressions and the rate at which they learn in virtual classes. This study aims to advance understanding of the possible benefits and drawbacks of e-learning as well as the strategies and approaches that may be used to improve its effectiveness and efficiency.

2.1 Growth of E-Learning and Consequences

Madanjit Singh et al [21] reviewed each of these efforts that were started by the national and state governments and presented a comprehensive analysis of the majority of the projects that were important. In addition, a survey is carried out to gather information on the perspectives of learners with regard to online education. The results of this learning show that online learning is beneficial in a way that is satisfactory, despite the problems that were highlighted in it.

Ahmed Bossman et al [22] utilized the Smart-PLS estimator so that the data may be processed and analysed.The authors claim that a variety of variables, such as technology anxiety, teacher factors, course quality, technological quality, and simplicity of use, have a significant impact on how successfully students in remote learning courses react to e-learning. According to the study, perceived learner satisfaction acts as a mediator in the links between Ghanaian distance learning students' learning results and their motivational factors. This is important because, in the absence of e-learning happiness, elements like technological anxiety and instructor features would not have a substantial influence on learner performance.

Long Pham et al [23] explored the relationships among student satisfaction, total e-learning service quality, and student enjoyment in the context of a developing nation like Vietnam. The authors analysed the survey data from 1232 college students using exploratory factor analysis, confirmatory factor analysis, and structural equation modelling using SPSS 25 and SmartPLS 3.0.

Ashraf Alam et al [24] provided a brief assessment of the effectiveness of e-learning in the classroom and the significance of cloud computing for the future of education. There are several advantages to using cloud computing with e-learning services, but there are also numerous hazards and issues to be aware of. These include the price, speed, and security of the system, as well as the user's idea, forms, and processes, as well as the tasks and resources of the management. Now that these materials are available online, students, educational companies, and educational schools can all rent them. The formation of an e-learning ecosystem is made possible by the infrastructure provided by cloud computing, which is independent, cost-effective, adaptable, and dependable. Cloud-based e-learning systems are substantially safer than on-premises systems, but on-premises solutions are also slower, costlier, and less effective.

Thomas Favale et al [25] conducted an investigation on how the implementation of the lockdown affected the network at the Politecnico di Torino. On the 25th of February, immediately after the conclusion of the school day, PoliTO began implementing its very own in-house solution for online instruction. Since then, the institution has been offering over 600 online courses each and every day, catering to a total of over 16,000 students each day. The writers of this study paint a picture of how the pandemic changed the way people used PoliTO's network. First, they look at how platforms can be used for online work and cooperation. They do a thorough study of the traffic in order to characterise the audience and the network footprint since the PoliTO online education solution is hosted on-site.

2.2 Student attention detection in online classes

Alberto Rivas et al [26] employed a number of machine learning techniques to a publically accessible dataset, including tree-based models and other ANN types. When these methods were applied to the dataset, it was shown that the frequency with which students viewed the resources made available on VLE platforms was a crucial factor affecting student performance. An actual case study with 120 students pursuing a master's degree on a VLE platform was used to investigate this aspect. Participants in the case study at the University of Salamanca were specifically master's degree candidates studying computer engineering.

Maryam Imani et al [27] examined the many theories of emotion. Then, many techniques for detecting emotions have been described, and their benefits and drawbacks for use in e-learning systems have been examined. The findings of this research demonstrate that when information from body language, facial expressions, as well as user communications is incorporated, multi-modal emotion recognition algorithms outperform single-modal ones.

Izzeddin et al [28] examined the views of nursing instructors and students on online education in the COVID-19 period. The approach employed to conduct this qualitative inquiry was descriptive-phenomenology. Data were gathered in August 2020 regarding the actual experiences of nursing students as well as faculty with online learning during COVID-19 using a random sample approach and semi-structured phone interviews. The three steps of Spielberg's methodology were used to thematically analyse the data after verbatim transcription. During the COVID-19 epidemic, nursing students and faculty members were somewhat happy with online education. However, the favoured method of instruction was traditional. The data categorised using the benefits, issues, and suggestions from the three major research areas.

Jesús Valverde-Berrocoso et al [29] described how the aim of this systematic literature review (SLR) was to identify the research themes, critical theories, researched modalities, and research methodologies. The PRISMA protocol was utilised to accomplish this purpose, and several technologies were used to mine reference data for knowledge and maintain it current. The literature review was conducted via the use of three first-quartile publications that are indexed in JCR-SSCI and are focused on Educational Technology. The final sample consisted of a total of 248 different articles.

Melaura Andree Erickson Tomaino et al [30] conducted research on distance learning. However, there is hardly any study on utilising this technology to educate pupils with severe impairments and significant behavioural demands. Regarding the viability and efficacy of distant learning programmes while dealing with kids who have severe developmental impairments and significant behavioural requirements, Study 1 gathered survey data from students' families and their instructors. The findings showed that while educators agreed that their children were benefiting academically from online learning, parents and teachers typically had a neutral opinion towards it. Study 2 looked more closely at how switching to distant learning affected students' IEP (individual education plan) objective progress. According to analyses, kids kept around half of the abilities listed in their IEPs and advanced on a further quarter of their objectives.

Fenio Annansingh et al [31] investigated the discrepancy between the instructor's and the students' perceptions of the cognitive active learning experience in a VLE. As a result, the phenomenological constructivism methodology is used in this work, with questionnaires and interviews serving as the main data gathering tools. The findings demonstrate that instructors think that without the proper instructional assistance, students often do not engage in deep learning in the VLE because they lack intrinsic motivation. With thoughtful, probing questions or remarks that encourage critical thought and information transfer, the teacher must elicit in-depth thought.

Kristof Coussement et al [32] focused on three significant contributions that seek to improve predictions of student dropout. It uses a real-world data set of 10,554 students from an online learning service that is open to individuals all over the globe who pay a monthly charge and compares a newly proposed logit leaf model (LLM) approach against eight existing algorithms. The LLM performs better than any other technique in striking a balance between comprehensibility and predictive performance. A unique multilevel informative LLM visualisation offers novel benefits over a conventional LLM visualisation. Third, this study shows how factors including student demographics, classroom features, and components that encourage academic, cognitive, alongside behavioural engagement have an impact on dropout rates. These findings imply that distinct insights for various student segments with various learning preferences emerge when LLM parts are reviewed.

Xian Peng et al [33] employed a hybrid approach of text mining and statistical analysis, researchers examined the explicit behaviours and implicit discourse content of learners gleaned from reviews. To differentiate the semantic content of evaluations between the two achievement groups (completers and non-completers), the authors proposed the Behavior-Emotion Topic Model (BETM) as an enhanced topic model. The researchers then employed statistical analysis to compare the discourse behaviours and content of the two groups. Between finishers and non-finishers, the results revealed a considerable difference in conversation behaviours and concentrated subjects. Particularly, submitting reviews was a major behaviour for those who completed the task, but responding to peers' evaluations and giving them a "thumbs up" were significant behaviours for those who did not. Additionally, those who completed the course tended to express their gratitude for the course-related material by submitting reviews and, later, displaying certificates, while those who did not finish the course tended to convey their disapproval of the platform's technical problems by reacting.

Melor Md Yunus et al [34] proposed the unified theory of acceptance and use of technology (UTAUT) model has an effect on how trainees act while using online learning. Finding the most crucial elements that influenced senior TESL students' utilisation of online learning during the COVID-19 epidemic was the main objective of this research. It also examined how advanced students' behaviour and desire to utilise online learning were influenced by the four qualities. 169 graduate students from a public institution in Malaysia participated. The research used a survey approach using a questionnaire to accomplish its objective.

Yousra BanoorRajabalee et al [35] researched the relationships between students' overall performance and their enjoyment and involvement in an online course, analysed student feedback, and presented the findings. 844 first-year college students in engineering, science, arts, management, and agriculture were able to access the programme online. It was created using a learning-by-doing educational approach and was mostly assessed via continuing evaluations. The focus of the whole curriculum was on acquiring new skills and abilities and using them in mini-projects that were relevant to the actual world. Student responses from 665 students were coded and subjected to quantitative and qualitative analysis. Engagement and contentment have a substantial and inversely associated relationship.

2.3 Student facial emotion estimation

Christine Asaju et al [36] examined how to distinguish and estimate emotional changes in human faces using deep learning, as well as how to afterwards examine expected emotions to provide feedback on the outcomes. This method, which utilises a temporal perspective, employs mapping criteria to link estimated expressions to the resulting emotion (positive, negative, or neutral), a VGG-19 pre-trained network for feature extraction, a BiLSTM architecture for identifying facial emotion expressions, and a mapping architecture for detecting occipital muscle activity.

Manuel Ninaus et al [37] compared the emotional involvement of adult participants executing a numerical task based on a game with a non-game-based counterpart using both automated facial expression recognition and subjective evaluations.With classification accuracy much above chance level, the authors used a machine learning technique based on facial expression recognition data to determine whether particular people were engaging in the game-related or non-game-based activity. According to subjective assessments and facial expression detection, both pleasant and negative emotions were consistently increased throughout game-based learning.

Swadha Gupta et al [38] utilized facial expressions to suggest a deep learning-based method for determining how engaged online students are in the moment. This is accomplished by evaluating the students' facial expressions during the online learning session to categorise their moods. The engagement index (EI), which predicts both "Engaged" and "Disengaged" engagement states, is calculated using the facial expression recognition data. A few of the deep learning models that have been tested and compared to find the most effective prediction categorization system for in-the-moment engagement detection are Inception-V3, VGG19, and ResNet-50. Using numerous benchmarked datasets, including FER-2013, CK+, and RAF-DB, the recommended technique's overall efficacy and accuracy are assessed.

Anil Pise et al [39] recommended tracking changes in the students' facial expressions throughout an online lecture using a Temporal Relational Network (TRN). It is shown that TRN, which sparsely samples individual frames before establishing their causal relationships, is much more effective than sampling dense frames and convolving them. Both single-scale and multi-scale temporal correlations are taken into account in this research to achieve the stated aim. As a basic classifier, a Multi-Layer Perceptron (MLP) is also assessed. It is possible to train the recommended system from scratch to distinguish facial expressions in videos. The open-source DISFA+ database was used to evaluate the suggested FER model. The length of the feature set that the TRN-based model used to effectively recognise expressions was significantly reduced.

Prakhar Bhardwaj et al [40] suggested brand new algorithms based on deep learning that monitor a student's emotional state in real time. These algorithms can detect a range of feelings, including anger, contempt, fear, happiness, grief, and surprise. The recently suggested cutting-edge algorithms that compute the Mean Engagement Score (MES) using analysis of the results of emotional recognition, facial landmark identification, including weights from an online poll that was given to students over the course of an hour-long session, make this possible. These algorithms were developed to accomplish the aforementioned goal.

Hao Yu et al [41] provided a Facial Expression-Augmented Teacher Dashboard that examines students' emotional states and provides instructors with this information. It also provided a Head Pose Tutor that determines students' head directions in real-time and reacts to any student disengagement that may arise in the classroom. The authors recorded video footage of college students using MathSpring.

Prabin Sharma et al [42] describe a method for determining the pupils' level of participation. It was created to operate in real time and just takes data from the conventional built-in webcam included in a laptop computer. The authors create a concentration index with three levels of involvement based on the movements of the eyes, head, and facial expressions: "very engaged," "nominally engaged," and "not engaged at all."

Ninad Mehendale et al [43] offered convolutional neural networks as a method for facial emotion recognition (FERC). The Federal Energy Regulatory Commission (FERC) employs a two-part convolutional neural network (CNN) to: Remove the backdrop from the image in the first phase, and focus on the face feature vector extraction in the second. The expressional vector (EV), which is a component of the FERC model, is used to classify the five different regular facial expression types. From the 10,000 photo database that was preserved, the supervisory information for 154 persons was obtained. The emotion may be appropriately indicated with 96% accuracy using an EV of length 24 values. The last perceptron layer in the two-level CNN operates in series and continuously modifies the exponent and weight values. With single-level CNN, FERC deviates from commonly used methodologies, increasing accuracy.

Ilana Dubovi et al [44] employed a multimodal data analysis method to give a more unbiased and continuous perspective of how student participation grows and how it affects learning results. In this research, electro dermal activity (EDA), facial expression, and eye-tracking sensors were used to capture the learning processes of 61 nursing students in addition to subjective self-reports. The simulation was based on virtual reality (VR), and the students participated in the simulation by utilising a computer.

Nobuyoshi Komuro et al [45] provided a method for calculating individual emotions based on data gathered about participants' interior environments. As a first step, the authors create wireless sensor nodes that monitor office settings by gathering information on how people perceive their surroundings inside. The created system gathers huge data on emotions from skin temperature and pulse readings as well as interior environment data from the created sensor nodes. The suggested method then calculates individual emotions based on information gathered about the interior environment. Additionally, this research explores the usefulness of sensory information in determining an individual's emotional state. Indoor environmental data gathered by developed sensors and emotional data gained from vital statistics were maintained for a period of 60 days. Machine learning was used to infer emotions from indoor environmental data.

Rhianna White et al [46] explored if peer interactions moderated the link between children's capacity to recognise facial emotions and their academic achievement over time. Testing whether the alleged causative pathways would be substantially different for boys and girls was a secondary goal.

Zhenzhen Luo et al [47] developed a smart analytic technique that makes use of multimodal natural sensing technologies to assess student interest. In order to fully define students' interest in a learning environment, the authors of this research provide a three-dimensional (3D) learning interest model that takes into account cognitive attention, learning emotion, along with thinking activity. On the basis of this model, multimodal data are acquired via the assessment of head position, the recognition of facial expressions, and the collecting of interaction data. Following this, multimodal data fusion is used in order to complete the student interest evaluation.

Paolo Buono et al [48] LSTM networks-based model is presented to predict student involvement levels using facial action units, gaze, along with head postures. It has been examined how student participation levels may be detected based on facial behaviour. The model was developed to establish how student involvement levels can be derived from facial activity. One of the datasets from the EmotiW 2019 competition was used to train the model. An experiment was conducted to assess how well the programme performed in a classroom environment with students enrolled in an online lecture. The study's objective was to compare how students rated their own involvement to how the model evaluated it. In order to measure the students' perceived participation, the authors asked them to complete a questionnaire at the conclusion of each session and record videos of their behaviour during the experiment. The combined movies were afterwards put through an automated analysis using software that not only uses the model but also offers a user interface for the visual interpretation of its output.

Pieter Vanneste et al [49] presented a strategy for gauging student participation that considers both the individual (student level) and group (classroom) levels. According to the findings, it might be difficult to identify students' unique behaviours, such as taking notes or raising their hands, and these behaviours do not correspond to how engaged they claim to be. Interestingly, utilising metrics for student symmetry, response times, and eye-gaze intersections allows for a more general quantification of students' collective conduct. However, there isn't much proof that these collective actions and involvement are related. Despite the fact that this study was unable to accurately represent students' self-reported involvement, the methodology highlights the need for more research.

GürayTonguç et al [50] examined the range of feelings experienced by 67 students studying Fundamentals of Information Technology across three diverse departments of a public institution located in the Mediterranean region over the course of their studies. The facial expressions of the students were examined along with digitalized using software developed with the aid of the Microsoft Emotion Recognition API and the programming language C# to ascertain how they were feeling when they displayed signs of disgust, sadness, happiness, fear, scorn, wrath, along with surprise. The kids' facial expressions were analysed and digitalized using the programme. The authors evaluated whether the change in the students' emotions throughout the presentation was statistically significant. The authors examined how student feelings altered in relation to their departments, gender, lecture times, computer placement in the classroom, lecture delivery style, and session details. There were three components in the lesson plan: an introduction, activities, and a conclusion. In three different phases, the importance of a range of emotions as well as the relationship between emotional growth and effective performance were evaluated. The study found that throughout the first half of the lecture, participants' degrees of enjoyment, melancholy, and disgust declined while their levels of disdain, rage, fear, and perplexity grew.

K. V. Karan et al [51] explained an autonomous system that analyses student involvement in the classroom by identifying the emotions they display. This is performed by capturing a snapshot of the students' video stream and submitting the faces that are found to an emotion detection mode. The results of this process are shown to the students. The VGG16 pre-trained image classifier model was optimised, and then the resulting emotion detection model was used to develop the architecture that was recommended.

Ruyi Xu et al [52] proposed a non-intrusive method for evaluating the cognitive state of emotion-sensitive learning. This technique uses the learners' head position to predict their level of attentiveness and their facial expression to determine their mood. the two elements that combine to make up the suggested structure. A multi-task learning system is shown in the first module. This system concurrently estimates head position, locates landmarks, and recognises faces using a cascaded convolutional neural network, or CNN. In order to align the faces before doing the facial expression analysis, landmarks are found. The learner's visual centre of attention is identified using landmarks and an approximated head posture. The second module involves training an expression intensity ranking CNN to recognise facial expressions and determine the level of intensity associated with each one using sequence ordinal data. After that, the feelings are deduced from the learners' facial expressions, taking each one in turn.

Khurshid Ahmad et al [53] compared the results of three prominent emotion identification software programmes using videos of real-world politicians from across the globe. The sample of 45 movies, with a total runtime of 2 hours and 26 minutes and 219150 frames, includes both male and female politicians, ages 40 to 78, with clear distinctions based on gender and nationality/ethnicity. The set of photographs depicts politicians giving speeches in a manner that is both somewhat staged and half spontaneous. As a result, the target systems—Microsoft Azure Cognitive Services Face API, Affectiva AFFDEX, and Emotient FACET—are being used by the authors to the fullest extent possible. With relatively limited testing on photos of unprompted emotions, these systems have historically been trained on posed expressions. These systems function similarly when it comes to certain emotions, notably pleasure, but they perform differently when it comes to other emotions, like rage. In addition to variances based on age and race, there are disparities depending on gender.

Reinhard Pekrun et al [54] investigated 3,425 German pupils (N = 3,425; 50.0% female) throughout five years of school to see how their mathematical performance correlated with how they felt about themselves. Positive emotions were reliably predicted by students' grade point averages over time, as shown by random-intercept cross-lagged modelling findings. Graduation outcomes were favourably correlated with these emotions. A student's performance at school was negatively correlated with their level of negative mood. The effects on an individual were simply a reflection of the relationships between grades and feelings seen between different people. Possible applications to theory, future studies, and classroom instruction are discussed.

Antonio Costantino Marceddu et al [55] presented a revolutionary Internet of Things (IoT) system that uses redundant non-invasive methods like physiological data analysis and facial expression detection to assist instructors in their job. Photoplethysmography (PPG) collects physiological data, and a convolutional neural network (CNN) determines how someone is feeling based on their face. The authors divided the most prevalent Ekman facial expressions found by CNN into active as well as passive categories by returning to Russell's paradigm. The data from the two sources were then compared and analysed using techniques like thresholding and windowing.

Tian Ma et al [56] suggested a unique deep neural network (DNN)-based bottleneck feature extraction (BFE) technique for recognising face emotions. Using a randomly generated mask and the Haar cascade classifier, the face and background were initially removed from the image. Second, the bottleneck feature extraction-focused final output layer of the VGG16 transfer learning model was removed. Third, they used the well-known Cohn-Kanade dataset to build a DNN model with five thick layers for feature training. Finally, they compared the suggested model to logistic regression models and the K-nearest neighbour model using the same dataset.

Aya Hassouneh et al [57] focused on employing a convolutional neural network (CNN) and long short-term memory (LSTM) classifier to categorise the emotional expressions of physically challenged persons (deaf, dumb, and bedridden) as well as children with autism using facial landmarks and electroencephalograph (EEG) data. This will be achieved by creating a real-time emotion identification system employing virtual markers using an optical flow algorithm that is efficient in erratic illumination, subject head movement (up to 25°), and various facial expressions. Data on a person's face is collected using ten virtual markers to identify their joy, grief, fury, fear, disgust, and surprise. The experiment on recognising facial expressions was carried out with the voluntary participation of 55 undergraduate students, with a gender split of 35 males and 25 females and an average age of 22.9 years. Nineteen undergraduate students consented to gather EEG waves.

Fei Yan et al [58] offered an easy-to-understand approach for tracking the levels of focus, comprehension, and engagement that are necessary to maintain a fruitful learning environment. The alternative OLE that has been suggested (i.e., AOLE) offers an intelligent 3D representation of the classroom environment (CA), which may help teachers adapt and customise both the lesson plan and the teaching for the best possible delivery. Additionally, the emotional curve of every student may be visualised at each step of the teaching or learning cycle in order to monitor their development. The significant emotional and psychological toll caused by COVID and the transition to OLE makes it possible to monitor learners' development up to the final exams by gradually contrasting the emotion curves throughout the learning cycle and the semester.

Junge Shen et al [59] proposed a unique framework for the evaluation of learning engagement, one that includes the identification of facial expressions in order to quickly acquire the learners' shifting emotional states. In addition, a novel approach for the identification of facial expressions that is based on domain adaptation has been suggested. This method is ideal for the MOOC situation.

Venkata Rami Reddy Chirra et al [60] developed and executed the creation of a multi-block deep convolutional neural networks (DCNN) model for the purpose of identifying the facial expressions of human, stylised, and virtual characters. To extract the distinguishing characteristics from face pictures using multi-block DCNN, the authors constructed four blocks, each of which had a unique set of computational components. Bagging ensemble with Support Vector Machines (DCNN-SVM) and the ensemble of three distinct classifiers using a voting mechanism (DCNN-VC) are the two additional models that were presented utilising ensemble learning with the goals of improving accuracy and enhancing the overall level of predictability. Image data was supplemented to broaden the scope of the dataset in an effort to enhance the generalisation and performance of the model.

2.4 Student facial expression and learning rate in online classes

Omid Mohamad Nezami et al [61] proposed a deep learning model that can increase engagement detection from photos. By pre-training on readily available fundamental facial expression data and then training on niche interaction data, our model gets around the problem of data sparsity. To create a rich representation of the face, deep learning is first used to train a facial emotion detection model. This step is followed by step two. The weights of the model will be used by the authors in the second step to initialise the deep learning-based model to detect engagement. They'll call this framework the engagement model.

EmtinanAlqurashi et al [62] examined the relationships between factors like learner-content interaction (LCI), learner-instructor interaction (LII), and learner-learner interaction (LLI) and other variables that might be utilised to predict outcomes like student satisfaction and perceived learning. There were 167 students that took part in the study in its entirety. The results of the regression analysis showed that perceived learning and satisfaction were strongly predicted by the complete model, which included all four predictor variables (OLSE, LCI, LII, and LLI). The study's results indicate that whereas OLSE is the most potent and significant predictor of perceived learning, LCI is the most potent and significant predictor of student pleasure. Nevertheless, LLI did not seem to be a trustworthy indicator of students' satisfaction or perceived learning.

Vasile Gherheș et al [63] focused only on the individuals who stand to benefit from the educational process and seeks to acquire their perspectives on traditional and online forms of education as well as whether or not they would want to go back to the more conventional method of instruction. Between December 2020 and February 2021, a total of 604 students from the Politehnica University of Timisoara were given the opportunity to reply in secret to a questionnaire consisting of eight questions. The students were given the opportunity to do so between the months of December 2020 and February 2021.

Chengyuan Jia et al [64] described how the writers responded to the challenge by turning a traditional flipped classroom into an online-only flipped classroom using the community of inquiry conceptual framework as the guiding principle. The data studies show that students who took online flipped classes did as well as those who took standard flipped learning forms, and that student interest levels stayed high throughout the whole online flipped class. The study also found that people who took online flipped classes did just as well as people who took traditional flipped learning courses.

Ali Akber Dewan et al [65] provided an analysis of the current state of the art with regard to the detection of participation in the context of online education. In light of the fact that different techniques need different amounts of involvement from the students, the authors divide the currently available approaches into three primary categories: automated, semi-automatic, and manual. The methods that fall under each category are then subdivided into other categories according to the kinds of data (such as audio, video, texts for learner log data, and so on) that they analyse in order to determine involvement. The automated techniques that leverage facial expressions and are computer vision-based are given special consideration. This is done since it has been shown that these methods show promise in the context of online education. These approaches are not invasive in any way, and the gear and software required to record and examine video data may be obtained at little cost and with little effort. Additionally, these methods are able to provide accurate results. In both technologies, the detection of interaction is accomplished using a variety of approaches drawn from the fields of computer vision and machine learning respectively.

Tomasz Szopiński et al [66] performed study among management students at a Warsaw business school in late March and early April 2020. Both native Poles and immigrants from former Soviet Union nations were among the respondents. The authors examine the 317 respondents' survey replies. They utilised the chi-squared test, the Kruskal-Wallis test, and the Mann-Whitney test to test their ideas about how the factors related to each other. The research finds that the student involvement variable influences the way in which online learning is now perceived, how often students participate since online courses began, and whether they prefer online or offline learning.

Jinyoung Kim et al [67] demonstrated how an early childhood education preservice teacher education course was changed to allow student teachers to take part in online learning and teaching possibilities. It offers insights and reflections gleaned from a hands-on class that took place in the United States of America during the first three months of the year 2020. The three phases of the experiences of the online student teachers that are detailed are the preparation stage, the implementation stage, and the reflection stage. Each phase's tasks are reported along with their results. These preservice teachers' online teaching experiences gave them the chance to connect with kids while also encouraging contemplation on the most effective ways to use online communication technologies to support young kids' learning and development.

Jing Li et al [68] built a new end-to-end network with an attention device that can easily tell the difference between face feelings. The four parts of the new network design are the module for extracting features, the module for paying attention, the module for putting things back together, and the module for classifying things. The attention module is an umbrella term that refers to all of these individual modules. The LBP features first extract information about the image's texture and then identify even the most minute face movements, which helps to boost the network's overall performance. Additionally, this information may be used to create more accurate facial expressions. The neural network may be made to pay greater attention to important characteristics by using a method called attention. In order to improve the attention model and get more accurate results, the authors mix the properties of LBP with the attention mechanism. In addition to this, they gathered and categorised seven different face expressions from twenty to twenty-five-year-old participants to create a new dataset of facial expressions. They took RGB photos as well as depth photographs of each person using a Microsoft Kinect sensor.

Michelle Taub et al [69]employed Crystal Island, a game-based learning environment for microbiology, in a research with 138 college students. The goal of the study was to find out how much student initiative changes understanding, problem-solving, and feelings. In this study, the effects of agency on learning and how people solve problems while using Crystal Island for game-based learning are looked at in more depth. "Examining the Effects of Agency on Learning and Problem-Solving Behaviours" was the subject of the last study.

Durga Ganga Rao Kola et al [70] offered an innovative strategy for feature extraction with the goal of enhancing the functionality of the FER. In this method, the LBP is computed by taking into account both the 4-neighbors and the diagonal neighbours in a distinct manner. Also, the idea of a movable window and average in radial directions are brought up to explain how emotional traits are different. This method shortens the length of the feature vector and works well even when there is a lot of noise. Throughout the process of putting things into groups, the Support Vector Machine (SVM) is used. The authors use the recognition rate and the confusion matrix as two ways to measure how well the suggested method works.

Sujit Kumar Gupta et al [71] introduced a novel max margin face recognition based technique, using the students' facial expressions to analyse the emotional content. The emotional content analysis looks at four different student moods: High Positive Affect, Low Positive Affect, High Negative Affect, and Low Negative Affect. Based on the four student moods predicted by the proposed technique, engagement ratings have been computed. Additionally, the study of classroom engagement takes into account the group engagement score for the whole classroom as a whole. Expert commentary and movies with analysed impact material are utilised as feedback to help faculty members improve their methods of instruction, which in turn helps students learn more quickly. At the National Institute of Technology Karnataka Surathkal in Mangalore, India, the highly anticipated smart classroom system was put through its paces with over one hundred students and staff members from four distinct information technology courses.

Ashraf Alam et al [72] studied a variety of educational AI applications, including as virtual classrooms, intelligent tutoring robots, adaptive learning, smart campuses, and teacher evaluation.

Ashwin T. S. et al [73] suggested a unique hybrid convolutional neural network (CNN) architecture, in order to investigate the emotional states of kids while they are present in a school setting. The suggested design is a combination of two different models. The CNN-1 model's goal is to find out what a single student's state of mind looks like in a single still picture. On the other hand, the CNN-2 model places a large number of pupils into a single photographic frame. So, the mixed design that the writers have made can be used to predict the emotional state of the whole class as a whole. The suggested design looks at the students' face reactions, hand movements, and body positions to figure out how they are feeling. Furthermore, because there were no standard datasets for analysing the students' emotional states, they created, annotated, and tested on their own dataset, which included over 8000 single faces in a single image frame and over 12000 multiple faces in a single image frame, all of which displayed one of three different emotional states: engaged, bored, or neutral.

Andrew Emerson et al [74] offered a multimodal learning analytics approach that uses data from student gaming, eye tracking, and facial expressions to predict how well students will do on a post-test and how interested they will be after using Crystal Island, a game-based learning environment. The writers looked at how well single and mixed methods learned from students (n = 65) could predict how well students did on their posttests and how interested they were in Crystal Island. Some of these methods were playing games, making emotional faces, and staring at the eyes. The results show that models that use multimodal data either do as well as models that use unimodal data or do better than models that use unimodal data when it comes to predicting how well students will do on their posttests and how interested they will be in the subject.

Mohammed Megahed et al [75] preferred to utilise a convolutional neural network (CNN) along with a fuzzy system that are loosely linked to each other. The CNN is superior to other CNN models when compared to the same training benchmark in terms of its ability to recognise a learner's facial expressions. After being fed the CNN's facial expression states as well as the learners' responses, the fuzzy system is utilised to identify the next learning level. Second, the study shows how a number of different facial expressions can be put together to make a single emotion that is representative of all of them. Thirdly, it provides corpora that can be used to evaluate the effectiveness of the suggested technique. 72 different learning tasks and 1735 data points representing various emotional states are included inside the corpora of the 12 learners.

Annelies Raes et al [76] examined the potential effects of the different learning environments the students encountered within the context of a synchronous learning environment on their relatedness, intrinsic motivation, and learning achievement. The Self-Determination Theory (SDT) claims that relatedness is a quality that affects one's degree of intrinsic motivation and, indirectly, predicts one's level of academic success. This is a fundamental principle that underpins all of the ideas. Even though SDT has been used in many different types of studies, only a few have used it to look at learning in a hybrid virtual classroom by comparing different learning environments and seeing how they affect relatedness, intrinsic motivation, and learning accomplishment. These studies were done by comparing different places to learn. The kind of educational setting that appears depends on whether students are physically present and attend the lesson in person (F2F) or online (virtually). The educational environment is also determined by whether it is the same for all students (pure) or a mixture of students (hybrid). This research shows the findings of an experimental study using a within-subjects design that compared the learning experiences of students attending classes in person against those attending classes online in either a pure or hybrid environment. The authors make use of a mixed-methods strategy, which incorporates retrospective self-report surveys and interviews in addition to real-time assessments of individuals' levels of intrinsic motivation. In addition, the impact of quizzes has been the subject of much research.

Arijit Nandi et al [77] suggested to use a stochastic gradient descent (SGD) approach to train an online real-time emotion categorization system (RECS)-based logistic regression (LR) model. The proposed RECS has the capacity to classify the subject's feelings in real time if training is carried out on the model through an online technique that makes use of an EEG data stream. The researchers used the DEAP data set, which is the gold standard and most widely used data set for classifying feelings, to test how well the RECS method worked.

Gina Ionela Butnaru et al [78] examined the views that students have towards the efficiency of online education at a time when there are no other options for receiving an education. Even though a number of studies have shown that traditional face-to-face education and online education, which does not require attendance, may both lead to successful outcomes for students, very little research has been done on the topic of student satisfaction with online teaching. This is especially true in the context of the transition from traditional methods of instruction to online learning.

Li-Chun Lin et al [79] examined how the participation of students in four distinct categories of learning activities influences the results of their education within the context of a flipped classroom. According to the findings, the instructional activities associated with the flipped classroom approach may be effectively adopted and carried out in an entirely online course, offering students the essential flexibility in terms of time and location. According to the findings of this research, students' levels of comprehension of the learning ideas improved when they participated in more asynchronous learning activities. Students who watched more pre-recorded video lessons were more likely to participate actively in simultaneous learning tasks. This led to a better total grade for the term.

Elham Hussein et al [80] surveyed undergraduate students about their first several weeks of the required transition to online learning that was brought about by COVID-19, as well as their experiences with emergency online learning, and examined their perspectives. During the week before the final exams for the second term of the 2019–20 school year, students at a university in Abu Dhabi, which is in the United Arab Emirates, were asked to write semi-guided essays. These essays were to be submitted as part of the student's grade for the course. The United Arab Emirates includes the university. On a few of these papers, research was conducted using open code. The study showed that the most often stated benefits of emergency online learning were that it saved money and time, was safe, easy to use, and got people more involved. On the other hand, the most often stated bad things about the experience were not getting enough help from teachers and peers, being distracted and not paying attention, having a lot of work to do, problems with technology and the internet, and being distracted and not paying attention.

Suhaib Muflih et al [81] utilized a web-based survey methodology that included descriptive, cross-sectional, and correlational questions, participants who were eligible for the study were recruited from five Jordanian government institutions. In order to attract people who may be interested in taking part in the survey, the authors ran a campaign on Facebook and employed the snowball sampling method.

Mohammed Arshad Khan et al [82] performed a research using a sample survey and a quantitative technique. Through the use of a Google Form Questionnaire, 385 Delhi-based secondary school students in classes 8 through 12 comprised the responder sample. The months of January and February 2021 were used to carry out the study. Statistical tools like the Chi-square test, factor analysis, reliability testing, along with logistic regression, among others, were used to look at the data. The outcome of the Logistic Regression analysis demonstrates that factors such as Internet speed, previous ICT experience, family income, mother's educational background, and room availability have a beneficial influence on online learning. The study's results showed that, on the whole, students see online courses favourably throughout the pandemic in order to continue their academic progress. However, they ran into a number of difficulties while taking online programmes.

Kang Ma et al [83] utilized a mixed-method approach to investigate the moderators, related variables, and TSE during COVID-19. An in-depth interview was conducted with six instructors out of a sample of 351 Chinese school teachers who retroactively reported their online TSE at the beginning and finish of the COVID-19 school lockout. The COVID-19 virus was responsible for the lockdown that was implemented.Six teachers from a sample of 351 Chinese school teachers who reported their online TSE at the start and end of the COVID-19 school lockout were interviewed in depth. The COVID-19 virus was responsible for the lockdown that was implemented.

Shervin Minaee et al [84] proposed a deep learning technique based on an attentional convolutional network that can concentrate on key facial characteristics and outperforms existing models on a variety of datasets, including FER-2013, CK+, FERG, and JAFFE.This strategy was successful in achieving this improvement on all of these datasets. The method was successful in achieving this improvement on all of these datasets. Specifically, the authors suggest this method. They also make use of a visualisation method that is able to locate significant face areas for the purpose of identifying a variety of emotions based on the output of the classifier.

Maryam Imani et al [85] presented that use multi-modal emotional signals to employ two or more of these techniques. In e-learning systems, the learner is the system's user. Some user emotion detection approaches are more suited for use in e-learning systems than others, for various reasons that have been covered in this paper. Various theories of emotion are examined in this book. Then, many techniques for detecting emotions have been described, and their benefits and drawbacks for use in e-learning systems have been examined.

A detailed survey is done on the various approaches and techniques implemented to detect human facial expressions for identification of emotions. Furthermore, a brief discussion on the combination of steps involved in a machine learning based approach and deep learning-based approach for face detection and emotion recognition along with classification were described. Occlusion and semi-frontal head posture are two important characteristics that influence automatic facial expression recognition. This may cause variations in the visual aspect of the actual face expressions. The disparity in class has been viewed as an additional obstacle in expression that typically happened upon obtaining the information. While it may be simple to elicit and annoy a smile, the actual challenge lies in accurately capturing rare emotions such as disgust, anger, and so on. Even though the networks have been exposed to a vast quantity of face photos in different quality levels during training, deep learning techniques find it difficult to handle the changes in image quality. It has become necessary to create more effective deep learning to solve these issues.

Chapter -3

Emotion Estimation Model for Cognitive State Analysis

3.1 Introduction

In more recent times, there has been a rising interest in understanding and researching the cognitive processes of people, especially emotions. This interest has been on the rise. The purpose of this project is to improve the functionality of a variety of applications in areas such as healthcare, education, and interaction between humans and computers, with the end objective of making people's lives better overall. Because they give automated and objective assessments of people's emotional states, emotion estimation models have emerged as essential tools in this field. These evaluations provide very helpful insights on the mental health of the participants who are the focus of this research.

Emotions are not straightforward and unidimensional events; rather, they are intricate and multifaceted phenomena that have a substantial influence on our ideas, behaviours, and general mental state. Self-reporting methods, which depend on people' unique perceptions and vocal manifestations of their emotions, have historically been the primary method for doing research on emotional states. However, there are limits and possible biases associated with self-reporting, such as the social desirability bias, challenges associated with memory recall, and variances in emotional awareness among different persons. Self-reporting procedures, notwithstanding their disadvantages, are difficult in real-time circumstances or when dealing with persons with restricted communication skills, such as babies or patients with severe neurological problems. This includes situations in which there is a need to get information from individuals who are unable to speak for themselves. Self-reporting is problematic for this reason, since it requires people to be responsible for providing their own information.

In learning process students display a wide range of expressions during the learning process because of a number of elements that represent their fluctuating emotional and cognitive states. Because different tasks take different amounts of intellectual effort, cognitive load plays a crucial role. Students' expressions are also influenced by their level of engagement and interest in the subject matters high levels of engagement frequently lead to excitement and curiosity, while low levels of engagement can show up as boredom or disinterest. Students who comprehend the content with confidence or who struggle with confusion and frustration indicate their understanding and comprehension levels. Positive feedback can cause satisfaction or encouragement in relationships with educators and fellow students, whereas negative feedback can cause disappointment and fear. Certain kids are inherently more expressive than others due to individual variances in personality features and emotional control. Emotional expressions are also influenced by external variables including the surroundings, one's health, and personal problems. Another layer is added by social and collaborative learning because group dynamics can result in expressions of agreement, dispute, or collaboration. Finally, learning progress and accomplishments arouse feelings of satisfaction and delight for accomplished milestones, or annoyance and sadness for missed opportunities. Educators can improve the overall learning experience by customizing their techniques to meet the specific needs of each student by having a better understanding of these elements.

Researchers and engineers have built complex emotion estimation models to overcome these issues. These models make use of current breakthroughs in machine learning, computer vision, and affective computing. These models attempt to reliably determine a person's emotional state by automatically analysing and interpreting a person's facial expressions, physiological signals, speech patterns, and other non-verbal indicators.

The provision of a trustworthy and dispassionate evaluation of one's feelings ought to be an essential function of any emotion estimate model used in cognitive state analysis. This, in turn, improves our knowledge of human behaviour and cognition by enabling the identification of patterns, triggers, and the influence of a wide variety of conditions on cognitive states. Making use of this information may result in personalised treatments, better assessments of mental health, adaptable learning environments, and improved human-computer interaction experiences.

A important part of face position estimation (FPE), which is a basic issue in the computer vision, is analysing photos or individual frames from films in order to establish the position and alignment of a person's face. This may be done by looking at the image. This technique is very important to a broad variety of applications, some of which include face recognition, tracking, augmented reality, and human-computer interaction, to name just a few.

For a variety of computer vision applications, it is vital to have an accurate assessment of the placement of faces inside pictures or videos. For example, if facial recognition systems know the exact location and orientation of a face, they can match it to a reference model. This makes the identification process more accurate. It is possible to monitor people and observe their behaviour more effectively by estimating the face position during video surveillance. This may be helpful.

The subfields of computer vision and machine learning have seen tremendous development in face position estimation methods in recent years, which has led to significant progress in both domains. The traditional approaches depended on manually designed features and modifications to the geometrical dimensions. On the other hand, the development of deep learning has led to the creation of methods that are more complex and dependable.

Classification and measurement of intensity are the two fundamental components that make up facial expression analysis, often known as FEA. The six primary feelings of anger, contempt, fear, joy, and sadness, as well as surprise, may each be identified by the classification of corresponding facial expressions. The pictorial representation of facial marker pointsfor emotion detection in image is shown in Fig. 3.1.

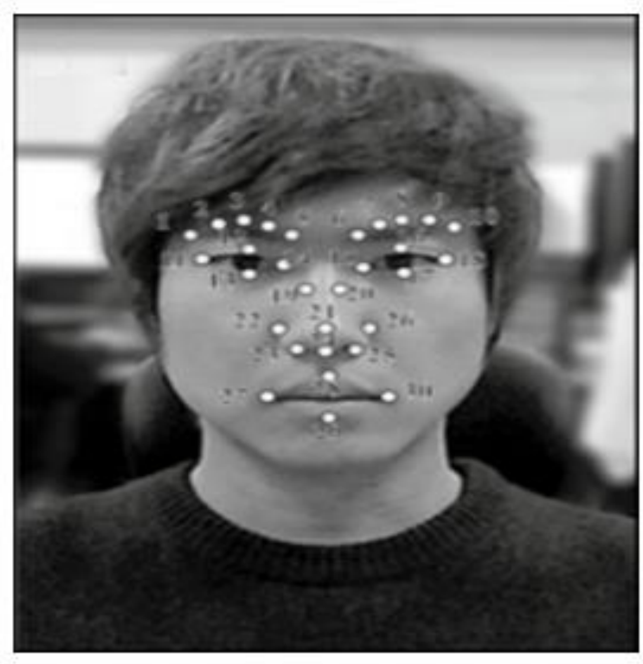


Figure 3. 1:Facial marker points of emotion detection

Convolutional neural networks, or CNNs, are used by deep learning-based face position estimation algorithms. These algorithms directly learn characteristics from raw picture data. These models have been trained on a significant amount of data, which has enabled them to catch detailed patterns and spatial correlations, which enables them to make accurate predictions about the landmarks and locations of the face.

In order to successfully complete the face position estimation procedure, it is essential to first recognise and precisely locate certain facial markers, such as the eyes, nose, and mouth. These landmarks act as important reference points that help determine the position of the face as well as its orientation and attitude. It is thus feasible to infer other face features, such as the subject's gaze direction, facial expression, and head attitude, when these markers are properly detected.

SAN-CNN is an abbreviation that stands for "Spatial Attention Network with Convolutional Neural Networks." In this work, this method is used for analysing face expressions. It does this by combining mechanisms for spatial attention with CNNs, which helps to improve the accuracy with which facial expressions can be identified and analysed.SAN-CNN's spatial attention mechanism focuses on certain facial regions that are important for expression analysis. These parts of the face are highlighted below. Improving the overall performance of the network may be accomplished by giving informative areas more priority and ignoring irrelevant parts.

The CNN component of SAN-CNN is responsible for identifying distinguishing characteristics in people's faces. Multiple convolutional layers, which use the input data to generate hierarchical representations, make up this component of the network. After these layers, completely linked layers are used for further processing and the categorization of face expressions.Through the combination of spatial attention and CNNs, SAN-CNN is able to successfully collect major face signals and patterns, which ultimately leads to an improvement in the detection and interpretation of facial emotions. It has shown promising results in a variety of applications, such as the detection of emotions, affective computing, and human-computer interaction.

3.2 Proposed Model

The subject of emotion detection has seen tremendous developments as a result of the combination of attention networks and CNNs, which are both types of neural networks. In this proposed work, by adding attention networks, the model obtains the ability to concentrate on important areas within an input, hence boosting its capability to identify subtle emotional signals. This is accomplished by giving the model the ability to focus on key regions inside an input. This is accomplished by giving varied weights to various aspects of the model, which enables it to put more of an emphasis on the characteristics that are necessary for a precise identification of emotions.

In addition to this, the CNN is a critical component in the process of extracting meaningful visual patterns from raw input data, such as facial expressions or body language, which is necessary for accurate emotion identification. The model is able to recognise more complicated emotional states after it has learned hierarchical representations, which gives it a greater knowledge of the emotional factors that make up such states. The combination of attention networks and CNNs allows emotion detection systems to successfully recognise and grasp emotional states from a variety of sources. This has led to improvements in affective computing, social robotics, and human-computer interaction. Attention networks are a kind of artificial neural network. CNNs are a type of convolutional neural network.

The dataset that was used in this investigation was obtained from Kaggle by the user ananthu017 (emotion-detection-fer). It is made up of a total of 35,685 different grayscale photographs of human faces, each measuring 48x48 pixels and having a resolution of 48x48. The purpose of this research is to organise these pictures into categories according to the range of feelings that they convey, such as joy, neutrality, sorrow, anger, surprise, disgust, and fear, among others. There are exactly the same number of samples in both the training and the testing datasets. The pictorial representation of proposed approach is depicted in Fig. 3.2.

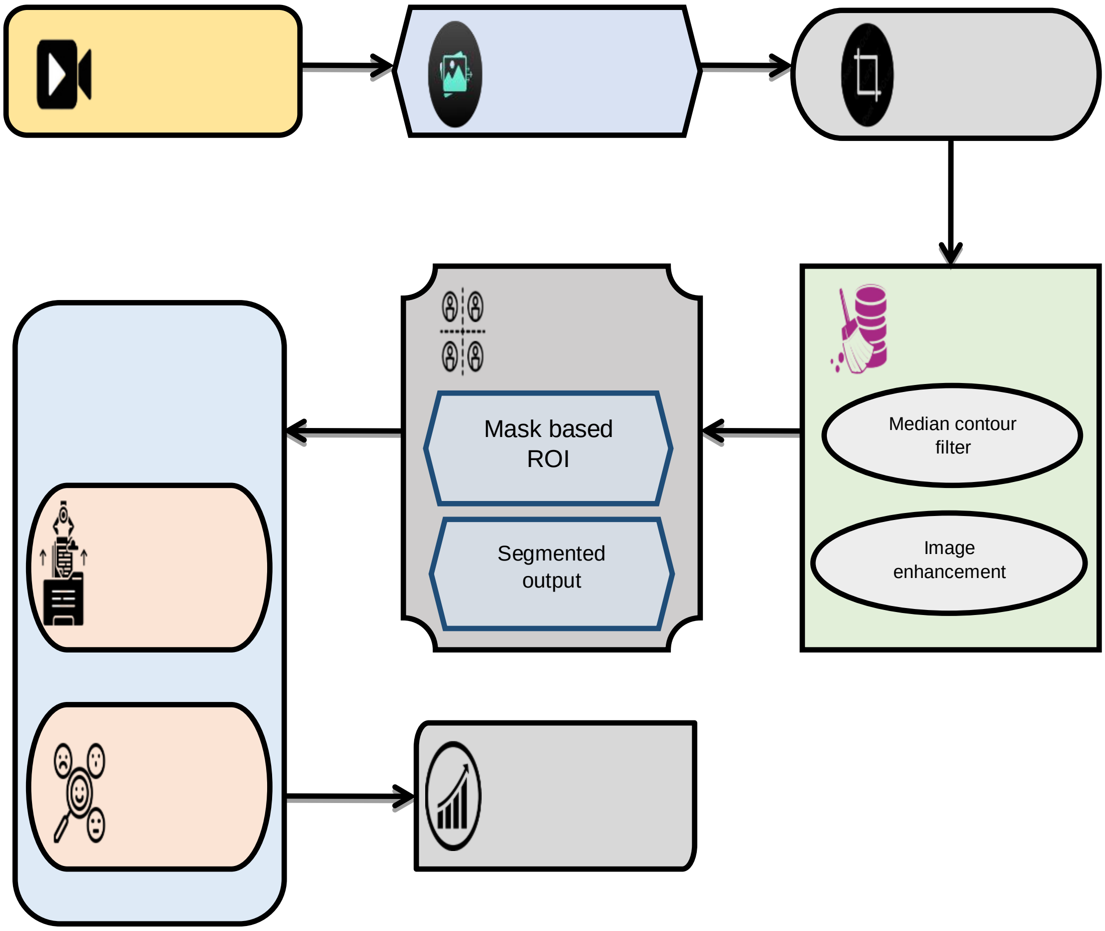


Figure 3. 2: Proposed Approach Framework

The in detail explanation for the proposed approach farmwork is discussed as follows:

3.2.1 Video to frame conversion

Converting the video into a series of photos is the first thing that has to be done while training the model. In order to finish this job, the cv2 tool is used to pull still pictures out of the movie. When reading the video file using the Video Capture function, a series of picture frames is produced, which may then be stored once the function has been used. Each frame stores visual information as a two-dimensional array of integers, and each frame itself reflects the information being shown. These arrays include several numeric arrays of varying sizes, and each pixel structure has three major colour channels: red, green, and blue. The arrays also contain many copies of the same array. A value ranging from 0 to 255 is assigned to each pixel to reflect the intensity of that pixel. The training model that we have constructed requires that we choose and input every 20th picture into it so that our photographs may capture a certain ambiance.

3.2.2 Pre-processing

We used a filter that is known as a median contour filter, which efficiently gets rid of artefacts and noise interruptions, so that we could recover the important information that was hidden inside the raw data. Standardising the face pictures based on heuristic information prior to processing increases the performance of the deep learning model by lowering the effect of background components that resemble faces. This is accomplished by standardising the images based on heuristic knowledge. As part of this procedure, the amount of background elements that are incorrectly recognised as faces will be decreased. Our study made use of a Kaggle programme that is freely accessible to the public in order to do the pre-processing. This software can determine an individual's emotional state by analysing facial characteristics, calculating head position, and monitoring facial features.

Moving the one pixel at a time allows adjacent patterns, often known as "windows," to move across the picture. The pixel values in these boxes are sorted by number, and the median is found by changing the pixel that is being looked at with the pixel value in the middle of the group. In the beginning, the filter sorts the pixels in the impacted region according to how far away they are from the centre of the picture.

$$W = \begin{bmatrix} J_{an} & J_{bn} & J_{nc} \\ J_{af} & J_{ab} & J_{nc} \\ J_{nc} & J_{bc} & J_{ac} \end{bmatrix} \tag{1}$$

The dependent variables *a*, *b*, and *c* are denoted by the letter *j*, which stands for the pixels.The gradient may be adjusted in accordance with the intensitylevel.

$$W_{pq}(a, b, c) = d_t D_s \int_{\Omega}^{\partial_P} Q(\mathrm{a}', \mathrm{b}', \mathrm{c}') \partial_P Q(\mathrm{a}', \mathrm{b}', \mathrm{c}') \mathrm{da}', \mathrm{db}', \mathrm{dc}' \tag{2}$$

W here stands for the pixel's gradient.

$$d_t = \begin{cases} 0 & \lambda_1 = \lambda_3, \\ \quad exp\left(\frac{-D}{|\lambda_1 - \lambda_3|}\right) & else \end{cases}$$

$$D_s = \begin{cases} 0 & \lambda_2 = \lambda_3, \\ \quad exp\left(\frac{-D}{|\lambda_2 - \lambda_3|}\right) & else \end{cases}$$

By figuring out the difference in outline strength between grid points, the mistake absence can be found.

$$C_i = \left[\sum_j^i N_j\right] \times \underset{No\ .of pixels}{\mathrm{Max}} .IntensityLevel \tag{3}$$

The following formula is used to get the pixel corners after drawing the grid corners on the scale.

$$W_{aa}(a, b, c) = \sum_{\Omega} \partial_a Q(\mathrm{a}', \mathrm{b}', \mathrm{c}') \partial_a Q(\mathrm{a}', \mathrm{b}', \mathrm{c}') \mathrm{da}', \mathrm{db}', \mathrm{dc}' \tag{4}$$

The formula can be modified as follows,

$$W_{aa}(a, b, c) = \sum_{\Omega'} \begin{matrix} \partial_a Q(\mathrm{a}' + \mathrm{ga}(\mathrm{c}' - \mathrm{c}), \mathrm{c}' + \mathrm{gc}(\mathrm{c}' - \mathrm{c}), \mathrm{c}') \\ \partial_a Q(\mathrm{a}' + \mathrm{ga}(\mathrm{c}' - \mathrm{c}), \mathrm{C}' + \mathrm{gc}(\mathrm{c}' - \mathrm{c}), \mathrm{c}') \\ \mathrm{da}', \mathrm{db}', \mathrm{dc}' \end{matrix} \tag{5}$$

Finally, every inaccuracy has been completely removed, and none of the edges or details have been lost in the process. The size of the rectangle has been expanded by a factor of 1.05 in order to encompass a bigger section of the face, avoid the loss of facial information, and reduce the amount of interference from the backdrop. Following the acquisition of a picture at a certain dimension, the learning model will first resize the image to match the input size, then it will rotate the image such that the tip of the nose is positioned in the centre of the image.

$$\oint(\alpha|D'(s)|^2 + \beta\,|D''(s)|^2 - \gamma|\nabla Q(D(s))|^2)ds \qquad (6)$$

Where $\alpha, \beta, and\, \gamma$ represent the information coming from the source, D stands for the noise.

Despite the fact that it is feasible to obtain an input image that is free of errors without losing any of the information in the process of separating the noise,

$$\frac{\partial D}{at} = g(Q)(d+k)|\nabla u| + \nabla u.\nabla g\; s\big(Q(a,b,c)\big) \qquad (7)$$

Where,

$$s\big(Q(a,b,c)\big) = \begin{cases} 1\;|g(a,b,c)| < T_v, \\ 0|g(a,b,c)| \geq\; T_v \end{cases}$$

3.2.3 Segmentation

The picture was segmented into its component pieces by using an approach that relied on a mask-based ROI. The first thing that has to be done in order to locate ROIs within a group of photographs is to section off the images into squares of equivalent dimensions. These mask-based methodologies are used inside of a five-dimensional space that contains the five blocks, supposing that the specific picture you are analysing is comprised of *n* photos that are numbered from *0* to *n-1,* where n denotes the total number of images. This space takes both the values of the ROI and the two locations for each pixel into account. Because of this, the suggested method begins with the first k observations and follows a certain order while making these observations.

$$m_n(w) = \int_{w_{k-1}}^{v} \int_{w_{k-2}}^{v} \cdots \int_{w_n}^{v} \int_{l}^{w_n} \ldots \int_{l}^{w_3} \int_{l}^{w_2} \times f(w_1, w_2, \ldots, w_k)\, ow_{n-1} \tag{8}$$

Where m is the total number of clusters that are being considered.

The grouping results are used to make a plan, and any blank spots are then filled in. After this step, a picture is split into n*K different groups so that it can be studied further. After the regions of interest (ROIs) have been found, they are found using a method based on masks. Before the filtering process is done, the Mean Shift clustering method is used to look at the n*K mean pixel values of each cluster. As a direct result of the new grouping that the Mean Shift creates, some pieces of information are brought together. At the end of the day, these m groups give a rough idea of the most common expression patterns in the first picture.

$$w_h = s_h + median\left(\begin{bmatrix} (s_{h-r} - s_h) + e_{h-1} \\ \vdots \\ (s_{h-1} - s_h) + e_{h-1} \\ 0 + e_h \\ (s_{h+1} - s_h) + e_{h+1} \\ \vdots \\ (s_{h+r} - s_h) + e_{h+r} \end{bmatrix}\right) \tag{9}$$

The mean shift data are represented by the variable w_h.

If m numbers of data thresholds are to be selected $[p_1,p_2,\ldots,p_m]$ which subsequently split the picture into groups $(s_0, s_1 \ldots . s_n)$, then the objective function can be expressed as, where n is the number of data thresholds to be selected.

$$s_\alpha(\mathrm{N}/I_{th}) = s_\alpha(p_0/I_{th}) + s_\alpha(p_1/n_{th}) \tag{10}$$

where

$$s_\alpha\ (p_0/I_{th}) = [1/(1\text{-}\alpha)]\ \log\ [1\text{-}(1\text{-}\alpha)] \sum_{j=0}^{I_{th}} \frac{p_j}{N_0} (\log[\frac{p_j}{N_0}])$$

$$s_\alpha\ (p_1/I_{th}) = [1/(1\text{-}\alpha)]\ \log\ [1\text{-}(1\text{-}\alpha)] \sum_{j=I_{th}+1}^{N=1} \frac{p_j}{N_1} (\log\ [\frac{p_j}{N_1}])$$

The variable *s* represents the data entropy, where reducing the value of the objective function results in improved image segmentation and higher system entropy. Once the optimal solution to equation (10) is determined and implemented throughout the segmentation process, the objective function value is converted into a maximization calculation. Finally, the photos are organized based on the coherence among their individual pixels.

3.2.4 Classification

The width of the face image, both vertically and horizontally, is determined by the presence of feature points along the edges of each face. The vertical width is defined by the top and bottom feature points, while the horizontal width is determined by the right and left feature points. These outlines, which are made up of the feature points along the edges of the face, can be used to figure out how someone is feeling (see figure 3.1 for an example). In this study, convolutional neural networks (CNNs) are used to sort feelings into groups. The first version of this design is made up of six layers: Conv2D, ReLu, Max-pooling, a fully connected layer, and a separate sigmoid layer. To improve the training results, more network layers, such as dropout, have been added. During training, the dropout layer randomly removes neurons from the network during the forward pass, while keeping a record of the remaining neurons. During the backward pass, only the information that hasn't been dropped is changed. This is called "regularisation”. This dropout layer enables the model to learn robust features independent of specific neurons, preventing overfitting.

$$c'_h \approx feature\left(\begin{bmatrix} -r\Delta_2 + e_{h-r} \\ \vdots \\ -\Delta_2 + e_{h-1} \\ 0 + e_h \\ \Delta_2 + e_{h+1} \\ \vdots \\ r\Delta_2 + e_{h+r} \end{bmatrix}\right) \tag{11}$$

In order to lower the overall number of neurons, the plan has six versions of the (Conv2D-ReLu-Maxpooling) layer combination, each with a different filter size. Before they are replaced by output layer neurons, the number of input layer neurons is cut. This means that the total number of weights at the end of the fully linked layer is smaller.

$$bj = \sum_{q}^{Q} aqxq + e * K_{e}(0, \sigma_0^2, \sigma_1^2) = xK(0, \sigma_0^2) + (1 - \epsilon)K(0, \sigma_1^2) \qquad (12)$$

The feature map is denoted by *bj* in this instance.

$$u_j(w) = \frac{\epsilon}{\sqrt{2\pi\sigma_1^2}} exp\left\{-\frac{[w-(j-r-1)\Delta_1]^2}{2\sigma_0^2}\right\} + \frac{1-\epsilon}{\sqrt{2\pi\sigma_1^2}} exp\left\{\frac{[w-(j-r-1)\Delta_1]^2}{2\sigma_0^2}\right\}, \qquad (13)$$

$$j = 1,2, \dots, k.$$

3.3 CNN

CNNs, which stand for "convolutional neural networks," are a type of deep learning model that was made especially to analyse visual data. They have been very successful at a number of computer vision tasks, such as identifying objects, separating pictures, and putting pictures into groups. CNNs were designed using the visual processing mechanism of the human brain as its inspiration. CNNs are made up of computational layers that automatically learn and extract meaningful information from the pictures they are fed. In these layers, there are convolutional layers, pooling layers along with layers that are totally related. The most important aspects of CNNs' operations and components:

The convolutional layer is the most essential component that goes into the construction of a CNN. This layer is made up of learnable filters or kernels, each of which has a relatively modest spatial area but covers the whole depth of the input volume. These filters either convolve or glide across the picture that is being entered, producing dot products between the weights of the filters and the values being input at each place. This procedure creates a feature map, which draws attention to certain visual patterns or characteristics that are present in the data.

After the convolution procedure, a non-linearity-creating activation function is added to the network one element at a time using the element-wise approach. The Rectified Linear Unit, often known as ReLU, is the activation function that is used in CNNs the most frequently. This activation function converts negative values to zero while maintaining the state of positive values. Pooling layers allow for a reduction in the overall spatial dimensions of the feature maps while still maintaining the integrity of the key data. The input is first separated into a series of non-overlapping parts, and then the value with the highest absolute value in each of those regions is outputted using the max pooling approach. This downsampling method helps in ensuring translation invariance, decreases computing complexity in succeeding layers, which in turn boosts processing speed, and it does all of this without sacrificing quality.

After numerous convolutional and pooling layers, the output is flattened into a 1D vector and then passed through one or more fully connected layers. This process may be repeated as many times as necessary. These layers, which function in a manner similar to that of a normal multi-layer perceptron, create connections between every neuron in the current layer and those in the layers immediately preceding and following it. Layers that are fully linked develop high-level representations of the characteristics that were retrieved by previous layers, and they base their predictions on these learned representations.

The last layer of a CNN is often a softmax layer, which, when used for classification tasks, normalises the network's output into a probability distribution across multiple classes. The likelihood that an input belongs to a certain category is represented by the activity of each neuron in the softmax layer.

Backpropagation is an optimisation method that is used while training a CNN. While performing backpropagation, the neural network makes adjustments to its weights and biases by taking into account the degree of deviation that exists between the projected outputs and the actual labels. This method is directed by a loss function, such as cross-entropy, which measures the degree of dissimilarity between the outputs that were anticipated and those that were actually produced.

When it comes to the processing of visual data, CNNs provide a number of benefits that cannot be found in more conventional machine learning approaches. Several of these benefits include the following:

Localised feature extraction: CNNs learn spatial groups of features on their own. Their convolutional layers capture both low-level patterns (like edges and corners) and high-level ideas (like the shape of an object). This method of extracting features is called "localised feature extraction."

Sharing parameters: CNNs are able to handle enormous input volumes in an effective manner by making use of shared weights in the convolutional layers. This helps to prevent an excessive increase in the number of parameters that need to be learned. Because of this, CNNs are well suited for processing pictures with a high resolution.Pooling layers assist achieve translation invariance by lowering the network's sensitivity to slight translations or distortions in the input picture. This makes the network more resistant to the effects of translation.

Hierarchical representation: CNNs are taught to create a hierarchical representation of the visual input they are given. Earlier layers of the network learn to capture low-level information, while deeper levels learn to extract more abstract and sophisticated aspects. CNNs are able to successfully simulate the visual environment because to the hierarchical representation that they use.

In general, CNNs are useful tools for computer vision applications as a result of their capacity to automatically learn and extract pertinent information from visual input.

3.4 Spatial Attention Network

The primary goal of a model of spatial attenuation is to concentrate attention on certain geographical places or areas of relevance, while downplaying or ignoring the significance of other regions. This kind of selective attention is useful in many computer vision tasks, such as object recognition, image segmentation, and picture creation, because it lets the network focus on the most important parts of the data it receives. Most of the time, a spatial attenuation model would have learnable parameters and focus methods to make it easier to change how the feature representations are shown. These components are necessary for the correct operation of the model because they make it possible to get useful information from the spatial correlations and characteristics that are present in the input data.

A spatial attenuation may be accomplished by the use of a variety of techniques, the most prevalent of which is the utilisation of spatial attention mechanisms. These spatial attention mechanisms can take the shape of spatial transformers or self-attention mechanisms. These systems learn to dynamically alter the weights or coefficients that are allotted to spatial places in the feature map. This allows them to successfully manage the value or relevance of each individual region.

In order for it to function properly, the spatial attention module takes into consideration the spatial organisation or structure of the data that it receives. It may look for repeating patterns, edges, or curves in the surrounding area to determine which elements of the picture are most relevant to the current endeavour by analysing the image. The module will eventually learn how to amplify key areas of the input, improve fine-grained characteristics, or suppress irrelevant or noisy parts of the input.

Depending on the mission at hand and the configuration of the network, the particular architecture and implementation details of a spatial attenuation model might look quite different. Within a more extensive network design, it can require extra layers, modules, or connections of some kind. In most cases, the model's parameters are optimised during the training phase by using strategies such as backpropagation and gradient descent in order to determine the attention method that is the most productive for the job.

By paying selective attention to relevant geographical areas in the input data, the goal of a spatial attenuation model is to improve the overall performance and efficiency of computer vision systems. This is accomplished by reducing noise in the input data. Networks' ability to efficiently use spatial information and increase their performance across a variety of visual tasks may be improved by the incorporation of such models into deep learning architectures.

The information given by the channel is ignored by spatial attention, which instead considers all channel properties to be of equal importance. In order to overcome this issue, a new attention module has been added to the low-level feature map. This module places its primary emphasis on spatial qualities such as contours and edges.

The module will eventually learn to emphasise key areas while downplaying those that are irrelevant. The arrangement of spatial attention is shown in Fig. 3.3, where the spatial descriptor $S \in R^{HxW}$ is created by performing an aggregation operation on the feature map $U \in R^{CxHxW}$. This is done by consolidating the features along the channel dimension. As a consequence of this, spatial features are dispersed over the whole of the picture.

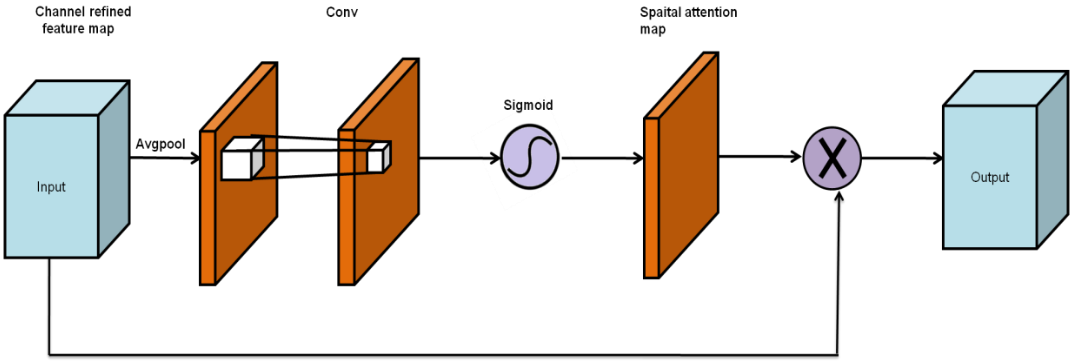


Figure 3. 3: Spatial Attenuation Architecture

$$\{i, j, k\} \in R^{C\times H\times W} m_{ji} = \frac{exp(A_i.B_j)}{\sum_{i=1}^{N} exp(A_i.B_j)} \qquad (14)$$

Here, the number shown by $m_{ji}$is a numerical indication of the degree to which the i^{th} position has an effect on the j^{th} position. It is essential to keep in mind that the feature representations of the two orientations are more similar to one another than they may seem to be at first glance.

The product of multiplying the transposes of the C and M matrices is denoted by the letter $R^{C\times H\times W}$, and it is the outcome of the process. In order to get at the ultimate result, the spatial attention matrix E is first multiplied by a scaling factor, and then the product of E and X is added to the result.

$$E_j = \alpha \sum_{j=1}^{N} (m_{ji} C_i) + X_j \qquad (15)$$

The value 0 marks the beginning of the phrase, which thereafter steadily grows in significance. The equation illustrates the end characteristic that is common by all of the locations. It illustrates that E is the average of the qualities as well as the traits that are unique to each location (15). Our method of connecting each encoder with a decoder, which differs from the traditional neural network topologies used for segmentation, is what makes our approach new. Because the processes that include downsampling cause the encoder to lose some spatial information, we are unable to downsample the encoder's output in order to retrieve the data that was lost. Instead, we make use of indices that cannot be trained in order to link the encoder and the decoder. The segmentation procedure is carried out in a variety of ways by directly connecting the decoder's input to the encoder's output signal. However, in the context of this particular inquiry, the input of each encoder layer is routed to the output of the decoder. The main goal of our work is to retrieve spatial information for use in decoding and upsampling methods. The decoder may operate with a smaller set of parameters since the encoder and the decoder communicate information at every level. Because of this, real-time operations are now possible, which makes the network more effective in comparison to the segmentation networks that are already in use, which are often inefficient.

3.5 Performance Analysis

Python was used as the basis for this proposed research work that investigated the learners' emotional attentiveness. During the tests, this research made use of the Kaggle emotion detection dataset.

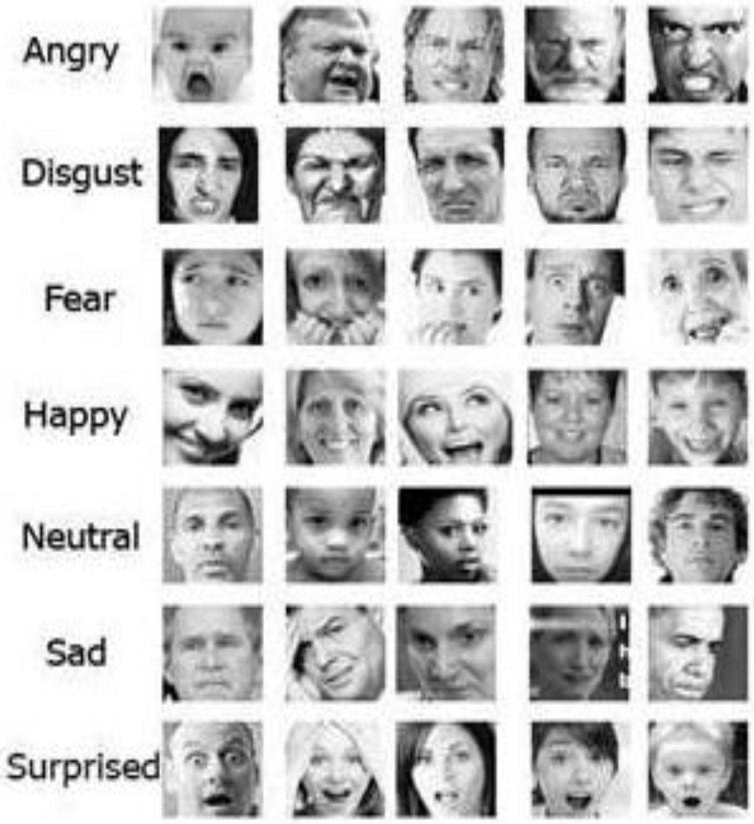


Figure 3. 4: Sample FERC-2013 dataset snapshots

We are able to make use of the FERC-2013 datasets that are available on Kaggle for the purposes of training and evaluating our proposed SAN CNN model provided that certain constraints and limits are met. Not only do these datasets cover the condition of apathy, but these datasets also include data on a wide range of emotions including disgust, fear, pleasure, happiness, and melancholy. We are able to build grayscale pictures by making use of the pixel values that are included within the datasets. These images will have a width of 48 pixels and a height of 48 pixels. Some visual examples of the dataset can be seen figure 3.4.

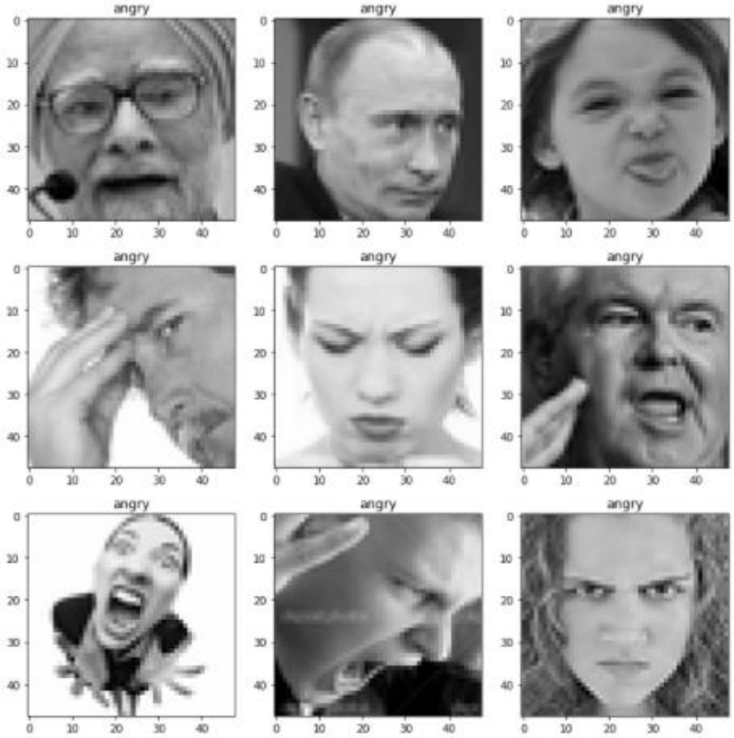


Figure 3. 5: Angry class prediction

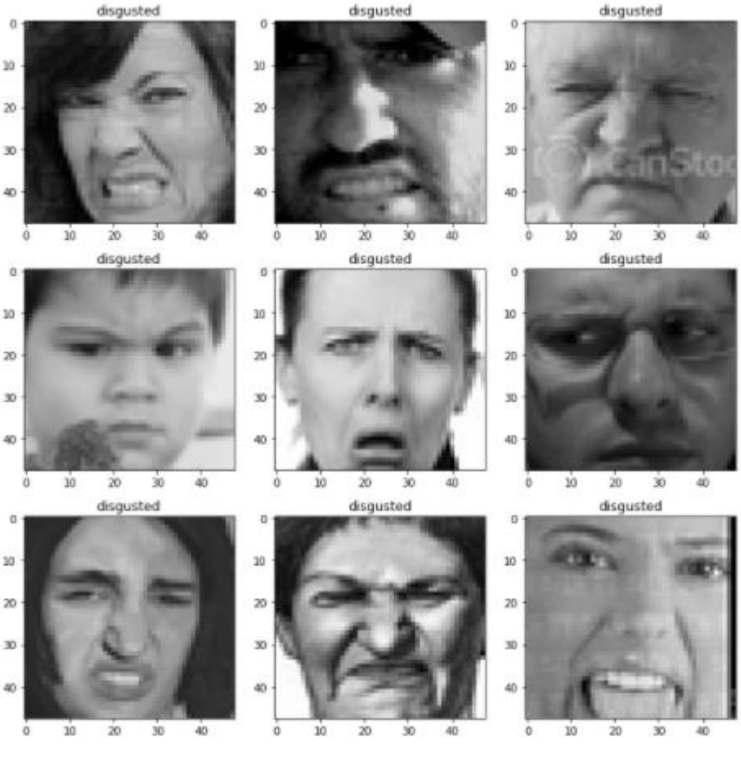


Figure 3. 6: Disgust class prediction

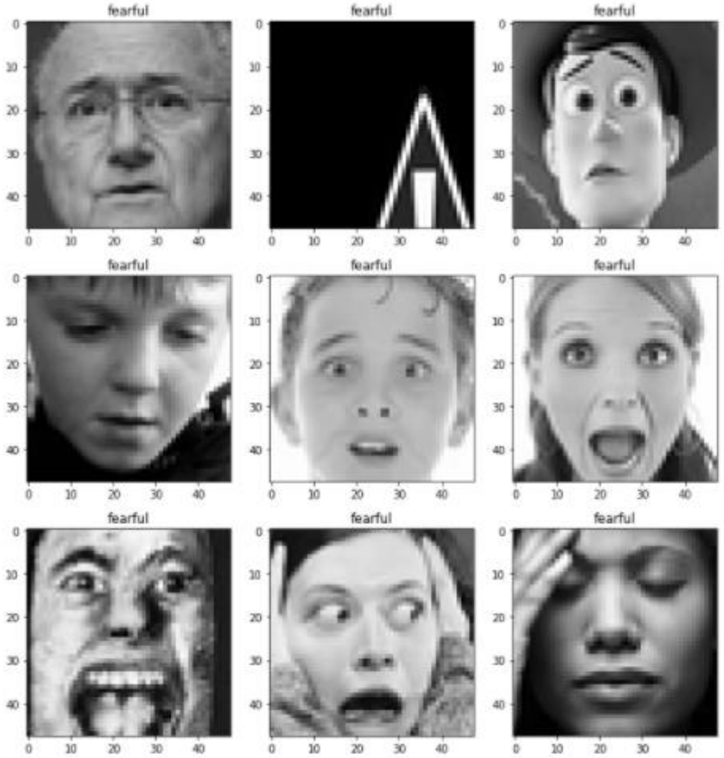


Figure 3. 7: fear class prediction

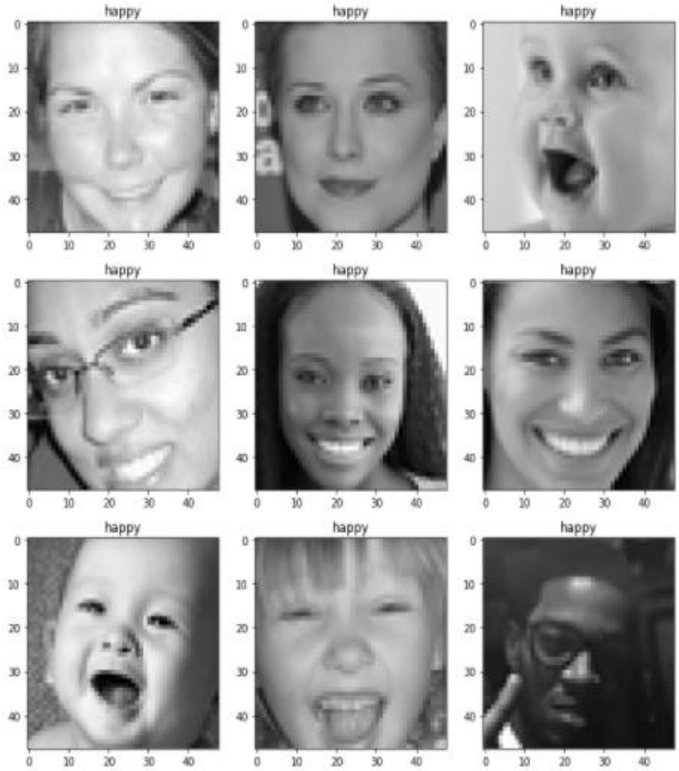


Figure 3. 8: happy class prediction

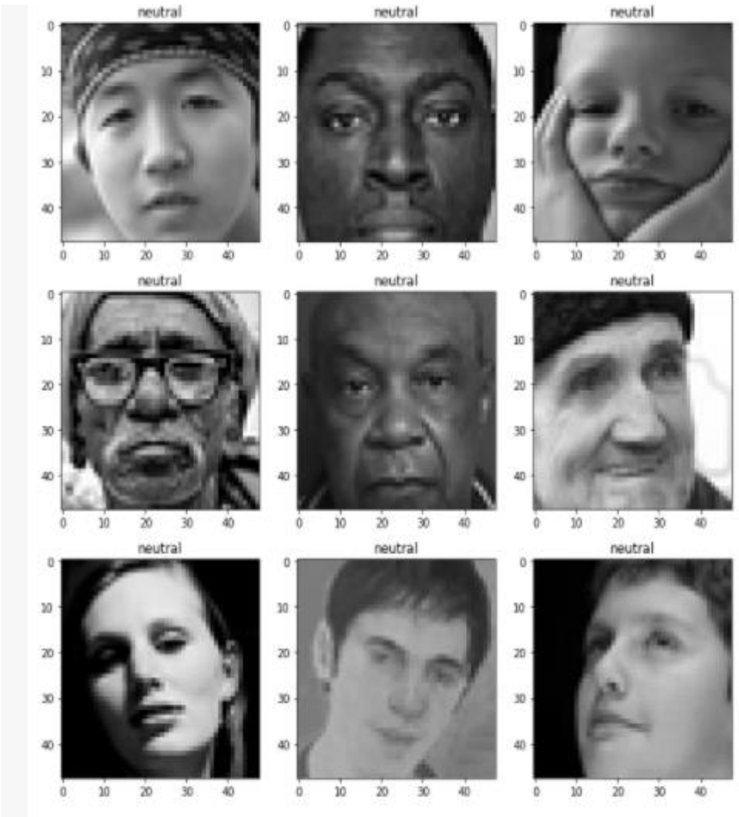


Figure 3. 9: Neutral class prediction

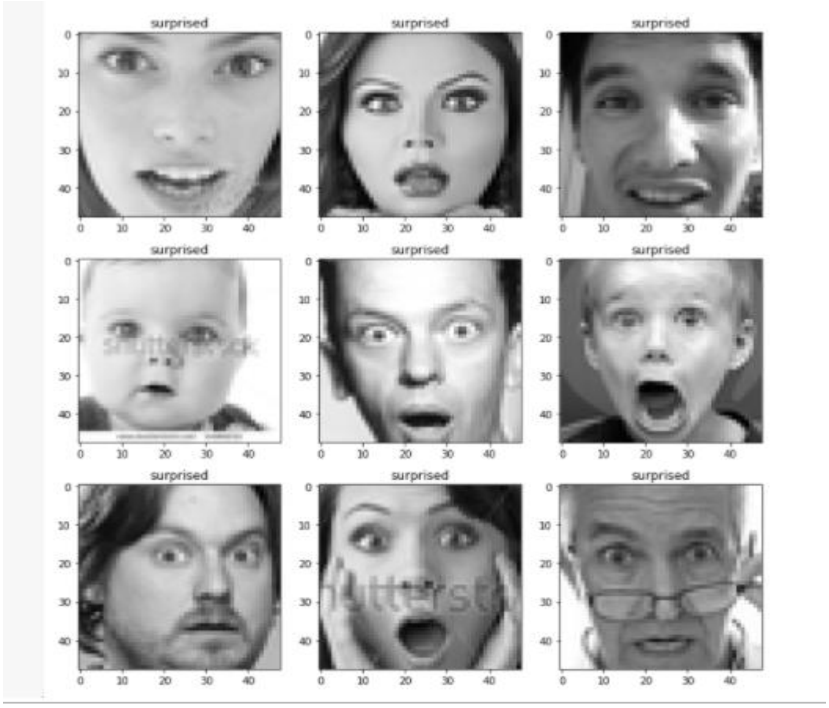


Figure 3. 10: Surprised class prediction

The use of the contour median filter during the processing of the picture helps to add to the unique and immediately recognisable expressions that are seen on each of the individual faces shown. When faces are seen from unusual angles, it might be challenging for face recognition algorithms to correctly identify a person's facial characteristics. It should come as no surprise that mistaking a doorway for a face might have quite serious repercussions. This may result in the facial recognition system producing false positives, which might lead to undesirable consequences. In spite of this obstacle, it is possible to overcome it by using a mask-based Region of Interest (ROI) for the segmentation process. This problem may be circumvented by the algorithm so long as it accurately identifies the ROI. Figures3.5, 3.6, 3.7, 3.8, 3.9, 3.10 and 3.11 demonstrates the subsequent construction of an accurate emotion detection SAN CNN model. This model was used to analyse the facial expressions of the subject.

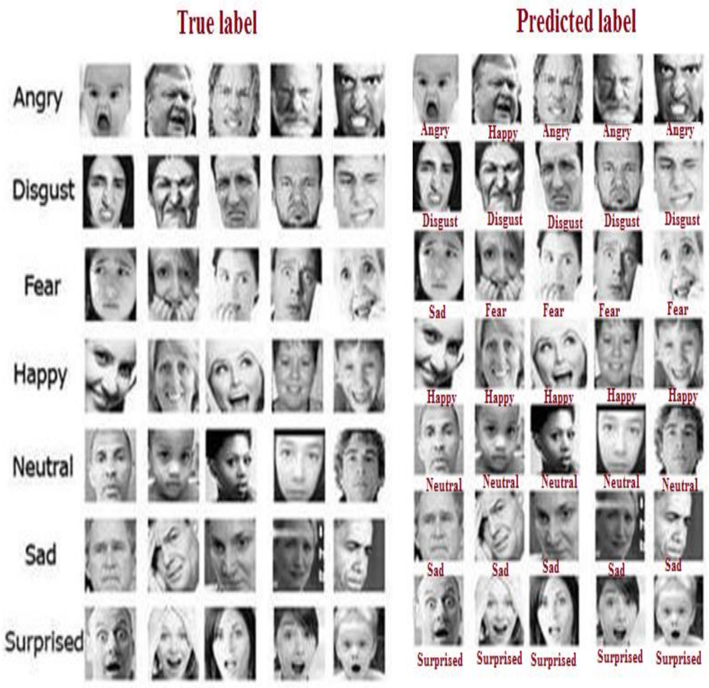

True label
Predicted label
Angry
Disgust
Fear
Happy
Neutral
Sad
Surprised
Angry Happy Angry Angry Angry
Disgust Disgust Disgust Disgust Disgust
Sad Fear Fear Fear Fear
Happy Happy Happy Happy Happy
Neutral Neutral Neutral Neutral Neutral
Sad Sad Sad Sad Sad
Surprised Surprised Surprised Surprised Surprised

Figure 3. 11: Multiclass predicted labels

Figure 3.11 provides a crystal clear illustration of the significant resemblance that can be seen between the projected label and the real label, which is an indication of the system's powerful competence in emotion recognition. The purpose of this part of the report is to conduct an analysis of how successful the suggested solution will be. In order to accurately quantify the level of effectiveness that was reached, a few parameters were closely tracked and examined. The data were subjected to an analysis using a variety of performance indicators created specifically for the new approach, and the results were then contrasted with those of prior tests.

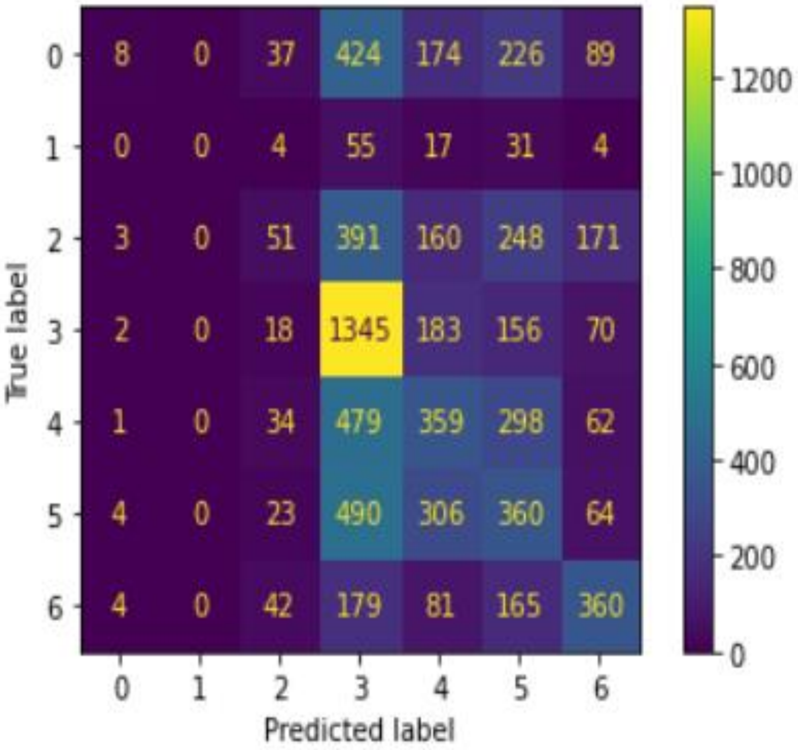
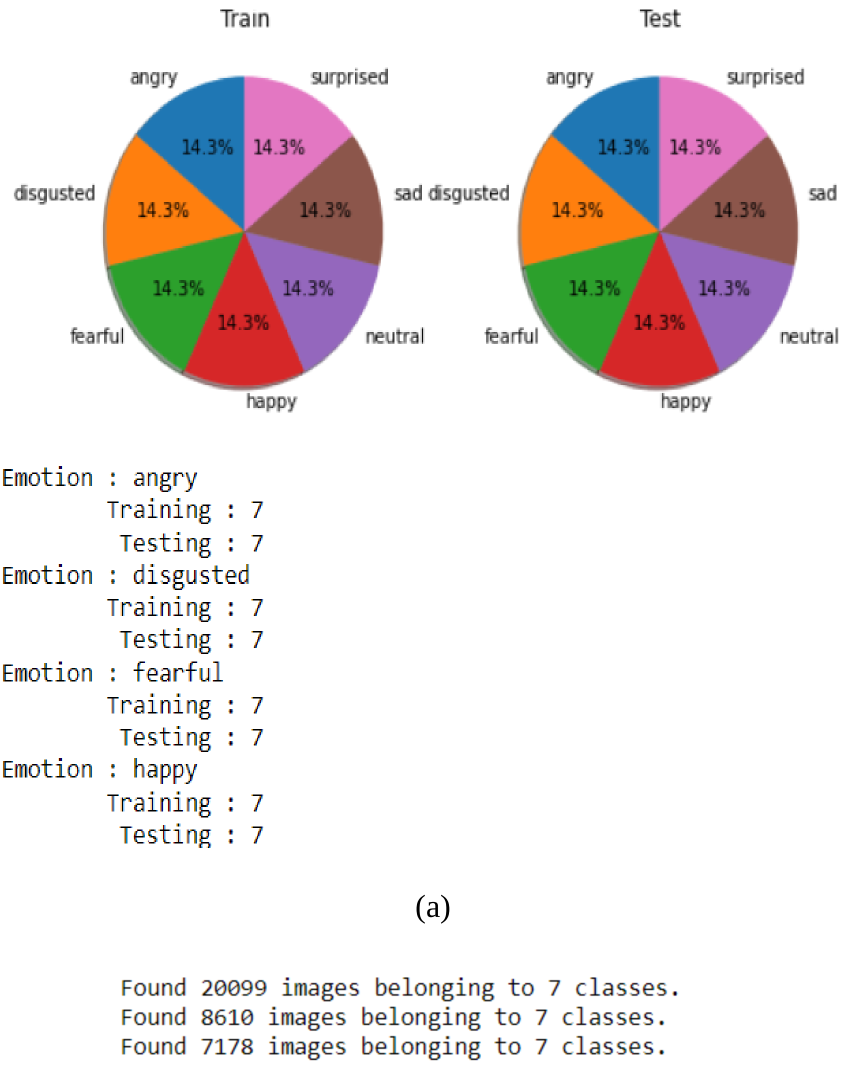


Figure 3. 12: Predicted label Vs. true label

Our algorithm's functioning is strongly reliant on the feature maps that we provide. Given that high-dimensional data often contains some inherent structure, the idea behind feature mapping is to estimate the predictive potential of a feature in relation to its interaction with other features. Given that high-dimensional data often has some inherent structure, this is done. By using this method, the need for a comprehensive mix of qualities is reduced. Figure 5 illustrates how feature maps make it possible to retrieve complex characteristics, which can then be used as input in the classification process. This is shown in the figure 3.7.



```
Emotion : angry
        Training : 7
         Testing : 7
Emotion : disgusted
        Training : 7
         Testing : 7
Emotion : fearful
        Training : 7
         Testing : 7
Emotion : happy
        Training : 7
         Testing : 7
```

(a)

```
Found 20099 images belonging to 7 classes.
Found 8610 images belonging to 7 classes.
Found 7178 images belonging to 7 classes.
```

(b)

Figure 3. 13: SAN_CNN output

Figure 3.13's results reveal that the suggested SAN_CNN model successfully classified emotions accurately by using the characteristics of the input data.

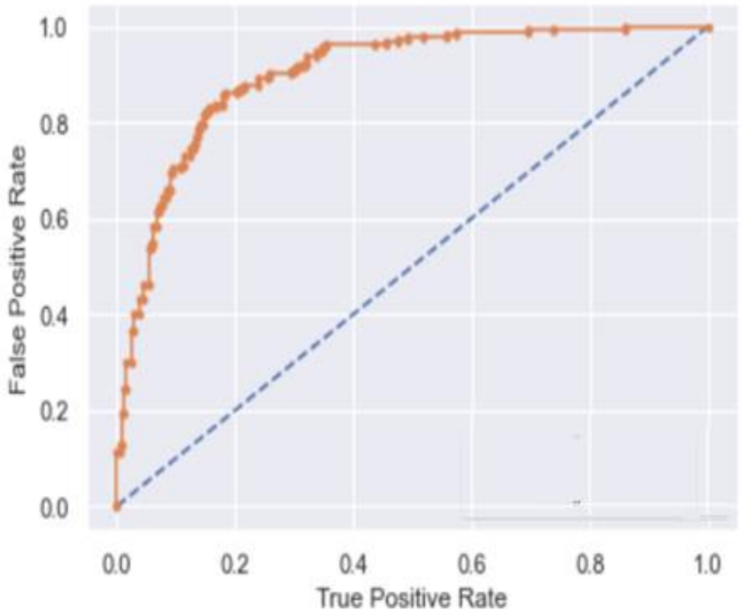


Figure 3. 14: True positive Vs. False positive rate

A genuine positive is when the expected results match the observed value; on the other side, a false positive is when the observed value differs from the projected outcomes. In these kinds of situations, the phrase "real positive" is often utilised. The false positive rate and the true positive rate are the two parameters that make up the ROC curve that represents the output of the suggested SAN_CNN model. The proportion of inaccurate diagnoses is known as the false positive rate. These percentages are shown in Figure 8. In this instance, the SAN_CNN model demonstrates greater performance when compared to other classification models that have been developed.In order to evaluate how well the model works with the dataset, a comparison was carried out between the SAN_CNN model and the methods that had been used in the past [15].

Recall: It is possible to compute the recall (16) by using the equation,

$$Recall = TP/TP + FN \tag{16}$$

Table 3. 1: Prediction of the recall percentile

Accuracy				
No. of. Epochs	**CNN-BILSTM[86]**	**SVM[86]**	**CNN[86]**	**SAN-CNN[Proposed]**
20	58	19	10	62
40	70	35	25	77
60	92	40	35	94
80	93	45	40	96
100	93	50	41	96
120	94	51	41	97
140	94	52	42	97

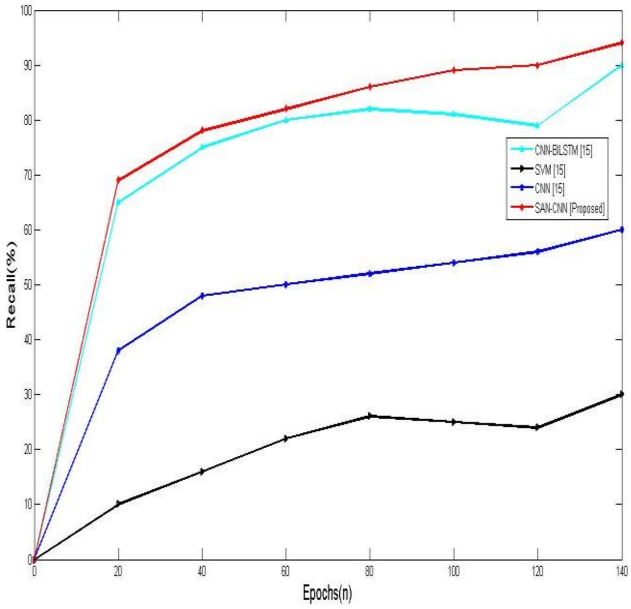


Figure 3. 15: Epochs Vs. recall

As can be seen in Figure 3.15 and Table 3.1, the proposed SAN_CNN technique achieves an accuracy level of 94 percent or higher when it comes to the recall. This precision exceeds the accuracy that can be reached using traditional methods [15].

Precision: It is possible to utilise the equation to find out how accurate (17) it is.

$$Precision = TP/TP + FP \qquad (17)$$

Table 3. 2: Prediction of the Precision percentile

Precision				
No. of. Epochs	CNN-BILSTM[15]	SVM[15]	CNN[15]	SAN-CNN[Proposed]
20	62	18	34	66
40	66	25	42	75
60	80	32	44	82
80	82	35	45	86
100	86	38	45	88
120	88	40	46	94
140	91	39	46	97

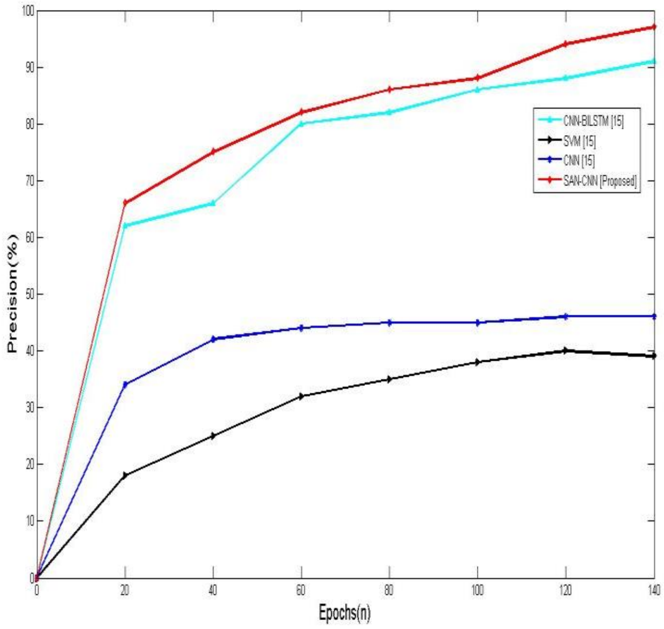


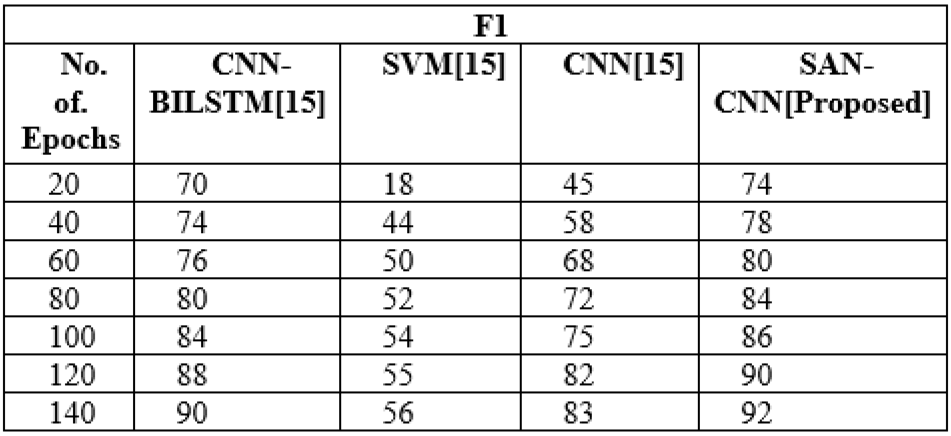
Figure 3. 16: Epochs Vs. Precision

The findings, which are shown in Figure 3.10 and Table 3.2, make it abundantly clear that the strategy that is advocated is the one that achieves the best outcomes. In comparison to the other approach, it achieves a greater level of accuracy97%than what was previously possible.

F1 score:The F-score, also known as the F1-score, is a statistic that may be used to assess a model's accuracy on a given dataset. It is often used in the assessment of binary categorization schemes that label situations as either "positive" or "negative."

$$\text{F1 score} = 2*\left(\frac{Precision*recall}{Precison+recall}\right) \tag{18}$$

Table 3. 3: Prediction of the percentile of the F1 score



F1				
No. of. Epochs	**CNN-BILSTM[15]**	**SVM[15]**	**CNN[15]**	**SAN-CNN[Proposed]**
20	70	18	45	74
40	74	44	58	78
60	76	50	68	80
80	80	52	72	84
100	84	54	75	86
120	88	55	82	90
140	90	56	83	92

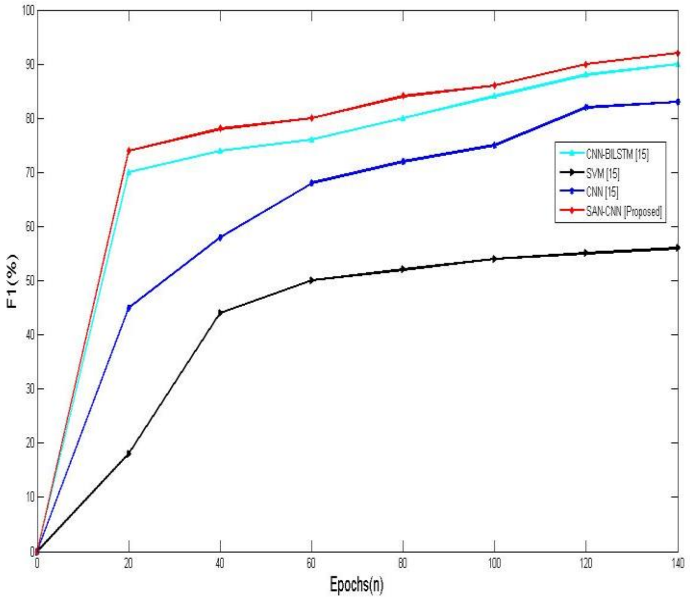


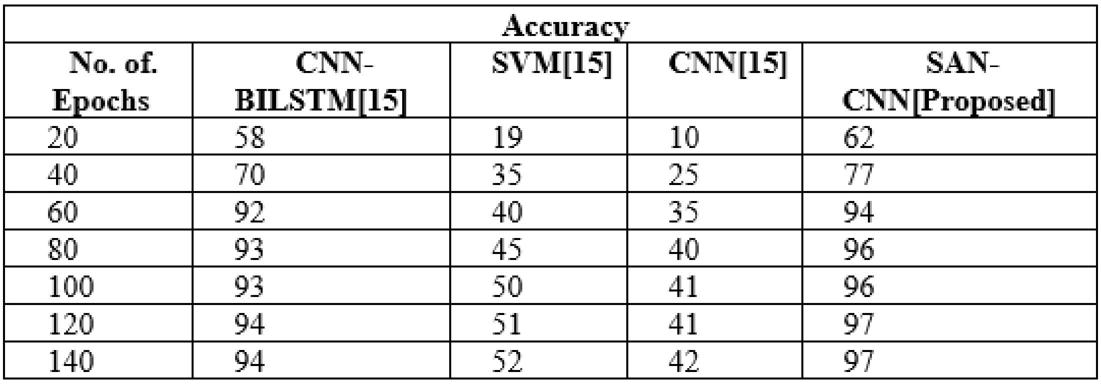
Figure 3. 17:Epochs Vs. F1 score

Figure 3.17 and Table 3.3 show that the recommended technique received an F1 rating. The results show that the suggested strategy achieves very high values for the F1 score coefficient, precisely, it approaches 92 percent.

Accuracy:To find the pattern of mistakes, compute the composition field's arithmetic mean. Regarding this assessment, accuracy is of the highest significance. There will be a discrepancy between the computed number and the "actual" number when the calculation's accuracy is subpar. The accurate testing is performed by machines and systems utilising unique data samples that have already through several iterations of examination using the same methodology. The percentage of correct data has a role in the correctness of the information that is ultimately generated, contributing to its completeness.

$$Accuracy\,(A) \;=\; (TP + TN)/\,(TP \;+\; TN \;+\; FP \;+\; FN) \qquad (19)$$

Table 3. 4: Prediction of the Accuracy percentile



Accuracy				
No. of. Epochs	**CNN-BILSTM[15]**	**SVM[15]**	**CNN[15]**	**SAN-CNN[Proposed]**
20	58	19	10	62
40	70	35	25	77
60	92	40	35	94
80	93	45	40	96
100	93	50	41	96
120	94	51	41	97
140	94	52	42	97

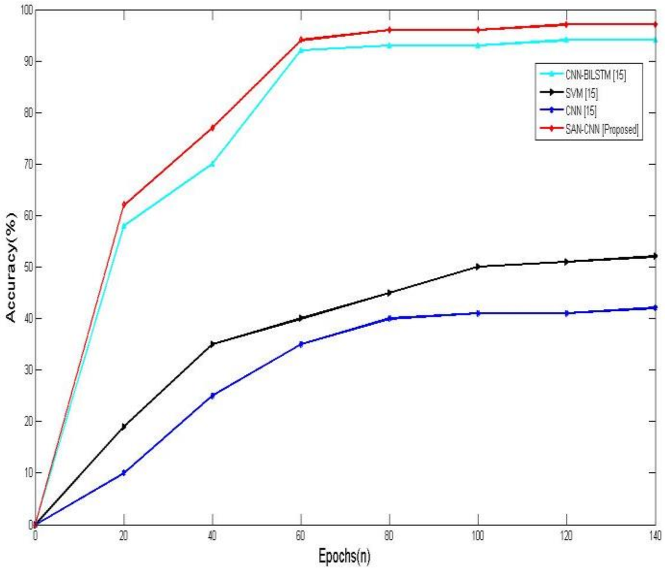


Figure 3. 18: Epochs Vs. Accuracy

The accuracy of the SAN CNN approach that was described is shown in Figure 3.18 and Table 3.4, which demonstrates that it can attain a maximum accuracy of 98.42%. This level of precision is much superior than that achieved by conventional methods. In comparison to previous methods, the efficacy of the one that has been described here is shown to be much superior in terms of performance.

Table 3. 5: Emotion class accuracy prediction

Classes	**Accuracy**
Happy	98.5
Surprise	98.3
Sad	98.4
fear	98.4
Disgust	98.39
angry	98.4
Neutral	98.4
Average	**98.42**

Due to the numerous of training examples, the "Happy" class was able to obtain the greatest performance out of all of the classes that were examined. This result causes feelings of shock, bewilderment, and anxiety in the audience. Table 3.5 offers a variety of emotional classifications along with the predicted degrees of precision for each of those classifications.

3.6 Conclusions

The development of Internet technology has led to a dramatic shift in the approaches that are taken in educational settings. Students are used to completing their academic work in complex digital settings that are dependent on having an active network connection. Recognising the range of feelings experienced by students in smart learning settings has emerged as an essential component of effective observation and assessment in the classroom. As a direct result of this, research on emotional computing has garnered a significant amount of interest over the course of the last several years. The purpose of this research is to offer a method for recognising learners' emotions that is based on deep learning. Making use of the SAN CNN algorithm makes it possible to precisely recognise the feelings of students in real time. This, in turn, results in a learning environment that is more interesting, productive, and free of stress for the students. Using the parameters of the SAN CNN model, we were able to reach an accuracy level of 98.42% in the classification of datasets as a result of our studies. This is a major achievement. The real-time SAN CNN algorithm for emotion identification in instructional graphics will, however, need more development in the not-too-distant future in order to achieve significant improvements in accuracy and efficiency.

Chapter 4

Student Attention Detection in Online Learning

4.1 Introduction

Students' interactions with pedagogical material have undergone a fundamental transformation as a direct result of the meteoric rise in popularity of online learning. The rise of online education as a prominent trend in the field of education may be attributed to a number of factors, including the progression of technology, the influence of the COVID-19 epidemic, and the need for flexible educational options. However, this change has given rise to worries about the attention and concentration levels of students when they are participating in virtual classrooms. This article explores the recent trend of online learning and its effect on the attention paid by students. It investigates both the benefits and the difficulties that are linked with this style of education.

Many other things have contributed to the rapid growth of online education in recent years. Recent developments in information and communications technology have made it feasible to provide instructional material of a high standard through digital platforms. Students now have the flexibility to access learning materials and participate in classes from any location at any time because to the ubiquitous availability of cellphones, personal computers, and internet connection. In addition, the COVID-19 pandemic required a rapid shift toward remote learning, which sped up the development and use of online platforms on a worldwide scale.

Learning on the internet offers a number of benefits, including adaptability and convenience. Students are free to learn at their own speed and according to their own timetables, unfettered from the restraints of conventional classroom environments. They have access to a broad variety of educational resources, including online libraries, interactive multimedia content, and digital simulations, giving them the ability to cater to a variety of different learning preferences. In addition, online platforms often include adaptive learning technologies, which customize the educational experience for each student by delivering relevant information and examinations depending on the individual's particular strengths and limitations.

Online education has many benefits, but it also has certain drawbacks that might make it more difficult for students to pay attention. Problems such as distractions and a lack of organization are often encountered in online settings. Students may be more likely to get sidetracked by distractions like social media, domestic tasks, or personal commitments when a teacher is not physically present in the classroom with them. The learning process may be disrupted and students might get frustrated when technological concerns, such as connection issues or devices that are incompatible with one another. This can distract their focus. with addition, students may experience feelings of isolation due to the lack of face-to-face connection that is inherent with online learning, which may have a negative impact on their engagement and attention levels.

Applying an Emotion Estimation Model to Online Learners' Cognitive State Analysis Deep learning has the potential to greatly improve e-learning in a number of ways. This technique can offer real-time insights into students' motivation, understanding, and engagement by precisely identifying and evaluating their emotional and cognitive states. Because teachers may adapt interventions and content to each student's unique needs, personalized learning experiences are made possible, keeping students interested and motivated. Furthermore, the model can recognise when students are having difficulty or are getting frustrated, which makes it possible to provide resources and support in a timely manner. This proactive strategy can lower dropout rates and enhance learning results. Moreover, the information gathered can be utilised to improve instruction and content delivery, improving the responsiveness and efficiency of the virtual learning environment.

Increasing student attentiveness in online classrooms may be accomplished via the use of a few different tactics. Fostering active student involvement via the use of interactive elements like as polls, quizzes, conversations, and group projects is an effective strategy for maintaining interest and fostering engagement. Structure should be provided, and better management of time should be emphasized by making it clear what is expected of students in terms of their conduct, participation, and deadlines. Students may be taught in a manner that best suits their individual learning styles via the use of multimedia and a variety of educational techniques, such as films, graphics, and simulations. This helps to maintain student attention. A feeling of community and support may be fostered by establishing frequent communication and cooperation channels, not just between professors and students but also among classmates. Lessons should be broken up into more manageable parts so as to prevent cognitive overload and assist in the retention of material. It is very necessary to provide students timely feedback that is both constructive and informative in order to maintain their attention and drive. Students are provided with the tools necessary to keep their attention as well as their discipline when there is an emphasis placed on the development of self-regulation abilities such as goal planning and time management. Along with this audio expressions offer important insights into the emotional and mental well-being of students. Understanding student participation requires analysing verbal responses, nonverbal cues, tone of voice, vocal fillers, pauses, speech tempo, and volume. Teachers can more accurately assess student attention and adjust their teaching strategies to promote active learning and quickly resolve any problems by evaluating these audio inputs. Incorporating voice analysis tools can improve the monitoring and engagement of students in virtual learning settings even more.

Students' access to and engagement with educational information has been fundamentally altered as a result of the advent of online learning as a preeminent trend in the educational sector. While eLearning does provide flexibility, accessibility, and tailored learning experiences, it also presents obstacles relating to distractions, technology issues, and restricted social connection, all of which may have an influence on a student's ability to focus on what they are learning. Educators are able to improve student attention and ensure a productive online learning environment by implementing strategies that promote active participation, clear expectations, utilization of multimedia, communication, and collaboration, as well as providing timely feedback and encouraging the development of self-regulation skills.

Due to the absence of both a physical presence and real-time monitoring, determining whether or not students participating in online learning are paying attention may be a difficult task. However, there are a few different strategies and methods that may be used to evaluate the attention and involvement of students. The following are some ways that are often used to assess the attention paid by students participating in online learning:

1. **Monitoring Student Participation in the System**: A variety of online education platforms provide statistics on the participation of students in the system. These metrics could include things like the number of times a user logs in, how long each session lasts, how many assignments or tests they finish, and how often they participate in online discussion boards or other online activities. Keeping an eye on these indications might provide useful information about the students' levels of involvement and attentiveness.
2. **Analyzing Interactions:** The way in which students engage with the things they are learning might give insightful information about the level of attention they are paying to the content. For example, keeping track of the number of clicks, scrolling patterns, or amount of time spent on certain sites or resources may provide insight into the level of engagement that students have with the material. A lack of attention may be shown by low levels of engagement or by rapidly skimming through the information.
3. **Online Quizzes and Assessments**: Having online quizzes or assessments as part of an online course is a good way to determine whether or not a student is paying attention and comprehending the material. The degree of attention that was given throughout the process of learning may be reflected in how well an individual does on these examinations. Low scores can suggest the need for more help or a shift in the way attention is focused.
4. **Student Surveys and Feedback Collecting:** feedback from students in the form of surveys or questionnaires may give subjective insights about the degree of attention and involvement they are providing in the learning process. Students are encouraged to discuss their successes, failures, and ideas for enhancing the quality of the environment in which they engage in online learning. This qualitative input may be used as a supplement to quantitative data in order to offer a more complete picture of the attention paid by students.

5. **Tools for Video Conferencing:** In the event that live virtual courses or synchronous sessions are carried out, tools for video conferencing may provide real-time indications of the level of attention that each student is paying. Students may be given visible indicators of their level of involvement via the use of features like as monitoring their attendance, participating in class discussions, or making use of the collaborative capabilities that the platform provides.
6. **Observational Assessment:** Teachers or other authorized observers may do periodic observational evaluations on students' attention during online courses in order to determine how well students are paying attention. The use of rubrics or checklists to evaluate aspects such as active involvement, responsiveness, engagement in conversations, or note-taking is a tool that observers may use.

It is essential to keep in mind that assessing the level of student attention based purely on quantitative data may not offer an accurate picture of the situation. When applied to the context of online education, the integration of qualitative comments and observations with quantitative measures allows for a more thorough knowledge of student attentiveness. When analyzing the findings of attention assessment, it is vital to take into consideration the constraints and context of online learning. This is because elements such as distractions in the home environment or technical challenges may impact student involvement.

4.2 Proposed Attention Detection model

The attention level of the students is analyzed using the proposed model, which compiles data from the visual feed, the aural input, and the heuristic properties of the camera. It will be possible to track the pupils' attention in real time when all of these elements are used in tandem with one another. The screen sharing and the video feed are the two parts of the input video feed that comes from the student's PC. Both of these components make up the input video feed. Since the screen sharing is being watched, the students are not permitted to view any other forms of media while the class is in session. The video stream is utilised to do facial recognition analysis on the students. When the image is initially processed, the OpenCV face detection module is applied to it. This module looks for the student's face everywhere it can find it in the picture. When the face has been recognised, the facial angle is determined by combining MobileNetV2 with a facial landmark recognition module that was built by Google. This is done so that the results may be more accurate. The face angle is calculated based on the recovered attributes, which serve as the foundation for the computation. In addition to the characteristics of the face, the audio input is scrutinised for the presence of any background noise, as well as the heuristic inputs from the computer are also taken into account. The model that has been suggested is an all-encompassing and complete model that appropriately evaluates students' levels of attentiveness.

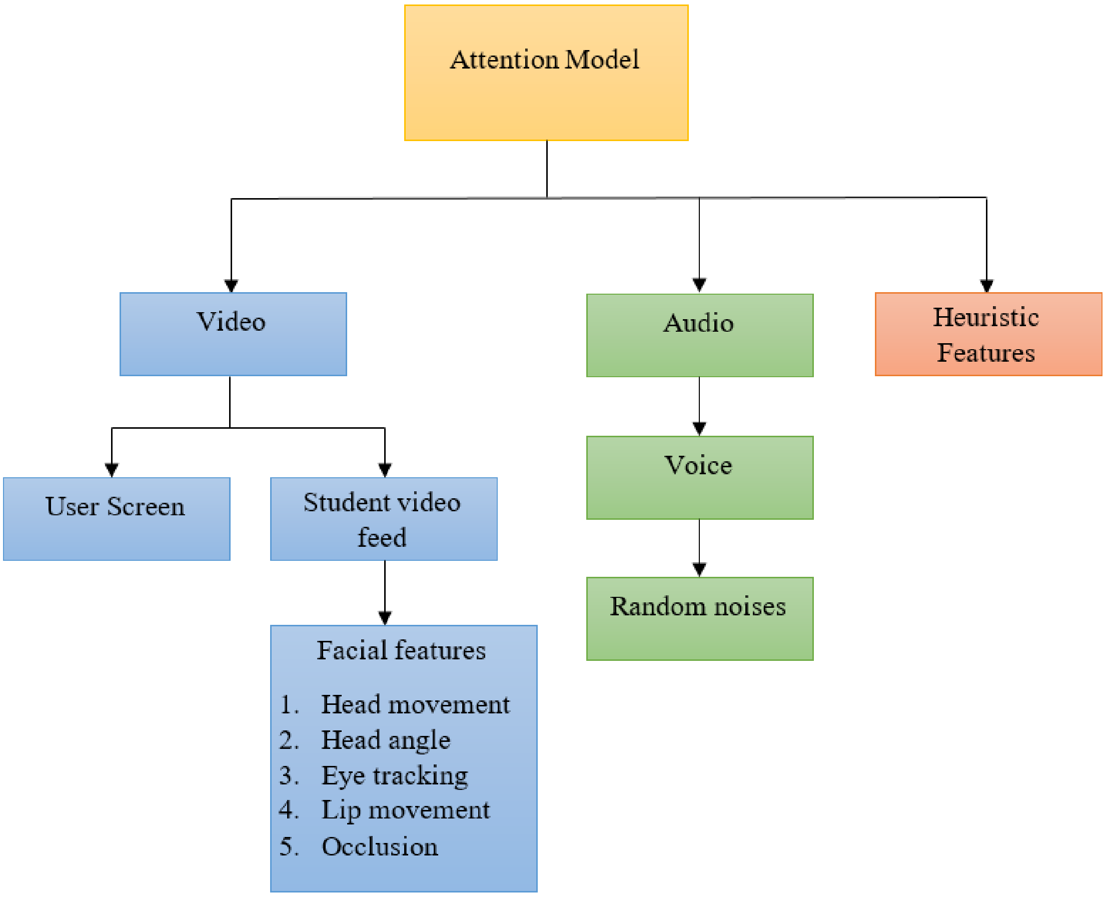


Figure 4. 1: Proposed attention model

The attention detection model that has been suggested is comprised of three necessary components, each of which serves a distinct function in the process of monitoring student attention during online learning:

1. **Video Features** The video input includes not only the camera feed that records the learner but also the screen feed that records their virtual learning environment. Both inputs may be monitored simultaneously thanks to this component's functionality. The recognition of visual signals is made possible by real-time analysis of the screen feed, which also provides insights into what the learner is presently watching on their screen. Through the use of the video stream, the instructor has the ability to examine the face expressions, eye movements, and mouth motions of the students. The degree of the student's involvement and attention may be determined with the use of these visual data.
2. **Audio Features**: The audio intake is focused on evaluating any outside noises that could be present when the student is taking an online course. Its purpose is to determine and evaluate the effect that interruptions from background noise or other factors have on student attentiveness. The model is able to assess, via analysis of the audio stream, whether or not there are any interruptions or distractions that may interfere with the student's ability to concentrate.
3. **Heuristic Features** The heuristic features include a variety of alternatives or capabilities that are supplied to the user interface. For example, the "mute" button or other interactive tools are examples of heuristic features. These features are intended to provide the user with assistance in making educated choices, which will in turn maximize the user's attention and the quality of their learning experience. The model is able to acquire further insights regarding the student's attentional state and how they are using the resources that are available to them by taking into consideration the choices that the student makes inside the interface.

When combined, these three aspects of the Attention detection model make it possible to conduct an exhaustive evaluation of a student's attention level while they are engaged in online instruction. The video features provide visual clues and emotions, the audio features evaluate outside disruptions, and the heuristic features take into account the decisions that the learner has made. The model is able to produce insights that permit appropriate monitoring and assistance for student involvement and attention in the context of virtual learning by assessing these inputs and doing the necessary analysis.

4.2.1 Video input processing

The visual input unit for processing is made up of two distinct modules, one of which is in charge of keeping an eye on the user's screen and the other of which is in charge of keeping an eye on the video stream. Processing and evaluating the visual data that was sent by the user is the responsibility of these modules. Monitoring the material that is shown on the user's screen and identifying whether or not the learner is concurrently seeking to acquire more information is the primary emphasis of the first module. At the same time, the second module isolates the user's face from the video stream. This enables real-time observation of the user's reactions to the supplied information.

4.2.1.1 Monitoring of Student Screens

It is strictly forbidden for the user or the student to access any other material during the time that the instructor is delivering the current lesson. The screen sharing feed that is available to the student will reflect all changes that the user makes to the program, including efforts to edit the program or move to a new application. The user will get a warning if such a change is detected, advising them to maintain their concentration on the content of the lesson and not divert from it in any way. This notice is prompted as soon as it is determined that there has been a change in the content.

4.2.1.2 Video monitoring

The student's video feed makes it possible to see a number of details, such as the pupil of the student's eye, the angle at which the student's head is held, and the movement of the student's lips. These characteristics are used as indications of the amount of attention that the pupils exhibited while the instructor was speaking. The OpenCV face detection module will begin by processing the input photo once it has been received. The user will be requested to gaze into the camera if the module is unable to identify a face during the authentication process. Following the successful detection of the face, the following phase entails doing an analysis of the face's angle as well as its individual features. An examination of the angle of the face and the characteristics of the face may be carried out by using these particular points of information. The conduct of the pupil that was witnessed leads one to believe that the kid in question was paying careful attention.

A. Detection of faces using OpenCV

In order to find faces in the input frame, OpenCV employs trained Haar cascade classifiers. When compared to other facial recognition methods, the speed of this approach is rather amazing. In order to identify features that resemble faces, the Haar classifiers look over the whole frame. Using this method, it is possible to find faces of various sizes. A significant chunk of the input picture is made up of regions that aren't faces. Because of this, it's crucial to have a simple approach for figuring out if a window counts as a face area. Throw aside the whole batch and abstain from processing if this is the case. Instead, you should concentrate on the possible positions of the face features. As a consequence of this, it's possible that greater effort may be put into investigating suitable face positions. Because of the notion of a cascade of classifiers, it is possible to carry out this method successfully. Rather than simultaneously applying each of the 6000 qualities to a window, the characteristics are first broken down into many different classifier processes. In the event that a window does not pass the first evaluation, it must be removed. There is absolutely no consideration given to any of the other attributes that it has. If the first stage of the features was effective, then the assessment process moves on to other qualities as potential areas of focus. The element that shows whether or not all of the processes have been correctly carried out is the region of the window known as the face area.

B. Face Image processing

To train the face images utilised in the project, a modified version of MobileNetV2 is used. The three fundamental layers that make up MobileNetV2 are the bottleneck layer, the average pooling layer, along with the convolution2D layers. These layers are referred to by their respective layer names. Each of these levels is accountable for a distinct activity within the context of the overall architecture of the model.

MobileNetV2 is an advanced convolutional neural network (CNN) architecture that was built expressly for the purpose of achieving accurate and efficient picture categorization on mobile and embedded devices with limited processing resources. This was accomplished via the development of MobileNetV2. Its major goal is to find a happy medium between performance and resource restrictions by reducing the overall size of the model and the amount of computation that is necessary.

MobileNetV2's primary objective is to provide a model that is both effective and lightweight, making it suitable for use on devices with limited memory and computing capacity. Examples of such devices include smartphones and Internet of Things devices. During the processing phases, it does this by using a mix of depthwise separable convolutions and inverted residual blocks. This allows it to accomplish the desired result.

MobileNetV2 relies heavily on depthwise separable convolutions to accomplish its goals. They do this by dividing the standard convolution method into the depthwise convolution layer along with the pointwise convolution layer, two separate layers. The depthwise convolution processes each input channel in an independent manner by applying a single convolutional filter at a time to that channel. When compared to more standard convolutions, this results in a considerable reduction in the amount of processing that is required. After that, the pointwise convolution will combine the output channels from the depthwise convolution by performing a 1x1 convolution. This will make it possible for cross-channel interactions to take place.

MobileNetV2's already impressive efficiency has been taken to new heights thanks to the use of inverted residual blocks. These blocks include a bottleneck layer that decreases the total number of channels, a depthwise separable convolution layer, and an expansion layer that raises the total number of channels to the level that was originally sought. The bottleneck layer cuts down on the amount of processing that is required by first performing a 1x1 convolution, which decreases the dimensionality of the data, and then applying a depthwise separable convolution, which requires much more work.

The linear bottleneck is a unique architectural option that was implemented in MobileNetV2. This design eliminates the need for non-linear activation functions in the expansion layer of each inverted residual block. Because of this, the model is able to store more information and make more efficient use of the parameters.

MobileNetV2 incorporates additional strategies to improve performance. These strategies include batch normalization, which allows for faster convergence during training, and the ReLU6 activation function, which clips values to the range [0, 6]. This function optimizes the utilization of limited bit precision in low-precision hardware implementations by limiting the range to which it clips values.

The adaptability of MobileNetV2's design makes it possible to personalize it via the use of hyperparameters like the width multiplier and the resolution multiplier. The width multiplier modifies each layer's number of channels, which in turn modifies the model's overall size and the amount of processing needed. The input resolution is multiplied by the resolution multiplier, which allows for a balance to be struck between precision and efficiency.

In a number of photo classification tasks, MobileNetV2 has shown performance that is regarded as state-of-the-art while preserving a high degree of efficiency. It is ideal for real-time applications on resource-constrained devices due to its lightweight design and effective inference. This enables tasks like as object identification, face detection, and semantic segmentation to be performed on mobile platforms. In addition, the way it was designed makes it very well-suited for use in applications involving machine learning.

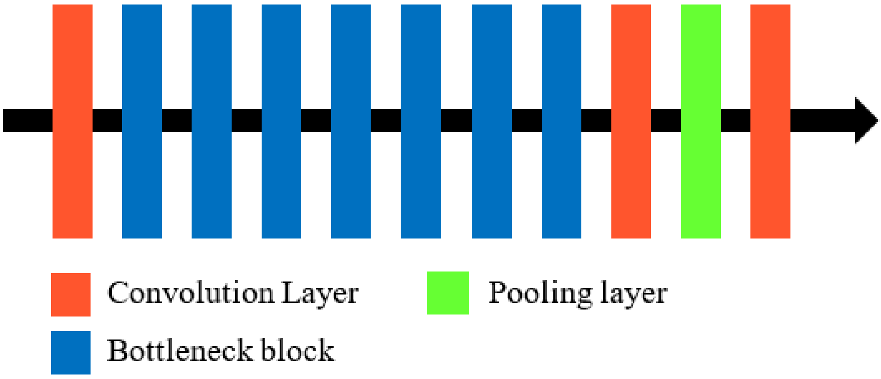


Figure 4. 2:architecture of MobileNetV2 network

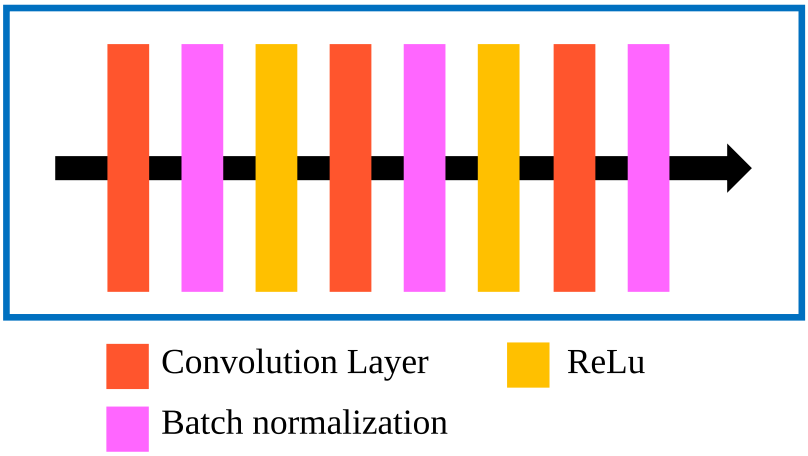


Figure 4. 3: Bottleneck block

Convolution: Convolution is a key operation in convolutional neural networks (CNNs), which are often used for processing photos and other grid-like data. Convolutional neural networks (CNNs) have become more popular in recent years. It requires applying a filter or kernel to an input data matrix, such an image, in order to extract significant features or patterns from the matrix. A dot product between the filter weights and the input values that correspond to each location is computed as the filter iteratively moves through the input matrix in the way that has been described. A feature map is produced as a result of this procedure, which draws attention to distinct patterns or traits included within the input.

The convolutional layers of CNNs generally consist of many filters that concurrently learn various characteristics. The network is given the ability to distinguish edges, textures, forms, and other important patterns thanks to these filters, which assists in the feature extraction and classification performed by following layers.

Batch Normalisation (BN) is a technique that is used to make deep neural networks easier to learn and more stable. The short name for this method is "batch normalisation." It is the effort to resolve the problem of the internal covariate shift, which is when the distribution of activations in the middle layer changes as a result of training. BN helps to stabilize and speed up the learning process by expediting the process of normalizing the activations that occur inside each mini-batch.

The following are the primary stages involved in the normalization of batches:

- determining the mean and standard deviation of the activations that are included inside a mini-batch.
- The mean value of the activations was removed, and the result was normalised by dividing it by the standard deviation.
- Using learnable parameters (gamma and beta), the normalized activations are shifted and scaled in order to provide the network the ability to learn representations that are most effective.

Batch normalization mitigates the vanishing/exploding gradient issue, functions as a regularizer to prevent overfitting, and decreases the network's dependency on a particular weight initialization, all of which are accomplished by lowering the network's dependence on a specific weight initialization. Additionally, it assists in smoother optimization, which enables the use of greater learning rates when the system is being trained.

Max Pooling The pooling process known as max pooling is often used in CNNs to downsample feature maps and extract dominating features. It then produces the largest value that can be found inside each zone and discards the remaining values, after first dividing the data that is being supplied into non-overlapping sections (typically squares). Max pooling helps to minimize the computational complexity of the model by lowering the number of spatial dimensions, which also helps to control overfitting and increases the model's translation invariance.

Max pooling is an operation that is performed on the separate channels of the feature maps in an independent manner. This enables the network to keep the most important features regardless of where exactly they are located. The downsampling ratio and the quantity of spatial information that is kept may both be affected by the pooling areas and their stride, which refers to the amount that the pooling window moves.

Rectified Linear Unit, often known as ReLU, is an activation function that is frequently employed in neural networks to bring non-linearity into the system. The formula for it is f(x) = max(0, x), where x stands for the value that is being fed into the function. When ReLU is applied, negative values are reset to zero, while positive values are maintained. This simple activation function is computationally efficient, which helps in overcoming the vanishing gradient issue and enables the network to acquire more sophisticated and expressive representations.

In deep neural networks, ReLU is often used in the convolutional as well as the fully connected layers. Because of its capability to add non-linearity, models are able to learn more diversified and discriminative characteristics, which contributes to an increase in their capacity of representation.

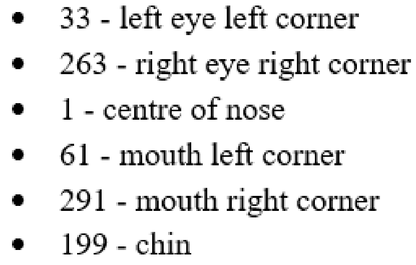
C. Facial feature extraction

Deep learning strategies are used inside the open-source framework developed by Google for the purpose of effectively recognizing and tracking facial landmarks in real time. The system employs a deep neural network-based model for the aim of finding these landmarks. The corners of the eyes, the nose, the mouth, and the jawline are just a few examples of the landmarks that correspond to different facial features. These distinguishable characteristics serve as anchors for the whole face.

To create a deep learning model for facial landmark detection, large datasets that have been labelled with the positions of face landmarks are used as input. These datasets contain accurate position information for a variety of face photos. With this information, the model is able to learn and generalize landmark patterns.

A multi-stage architecture that includes convolutional layers and regression layers is often part of the model that is responsible for recognizing face landmarks. This architecture takes use of the ability of deep learning to extract key information from the aligned face input and to produce accurate predictions about the locations of facial landmarks. The model is able to develop expertise in recognizing and localizing face landmarks with a high degree of accuracy after learning representations from the data.

In real-time applications, the model makes use of the capabilities of this deep learning model to correctly recognize and track face landmarks. The success of the framework in face analysis and understanding is contributed to by the model's capacity to extract significant features as well as its knowledge of facial landmark patterns.



- 33 - left eye left corner
- 263 - right eye right corner
- 1 - centre of nose
- 61 - mouth left corner
- 291 - mouth right corner
- 199 - chin

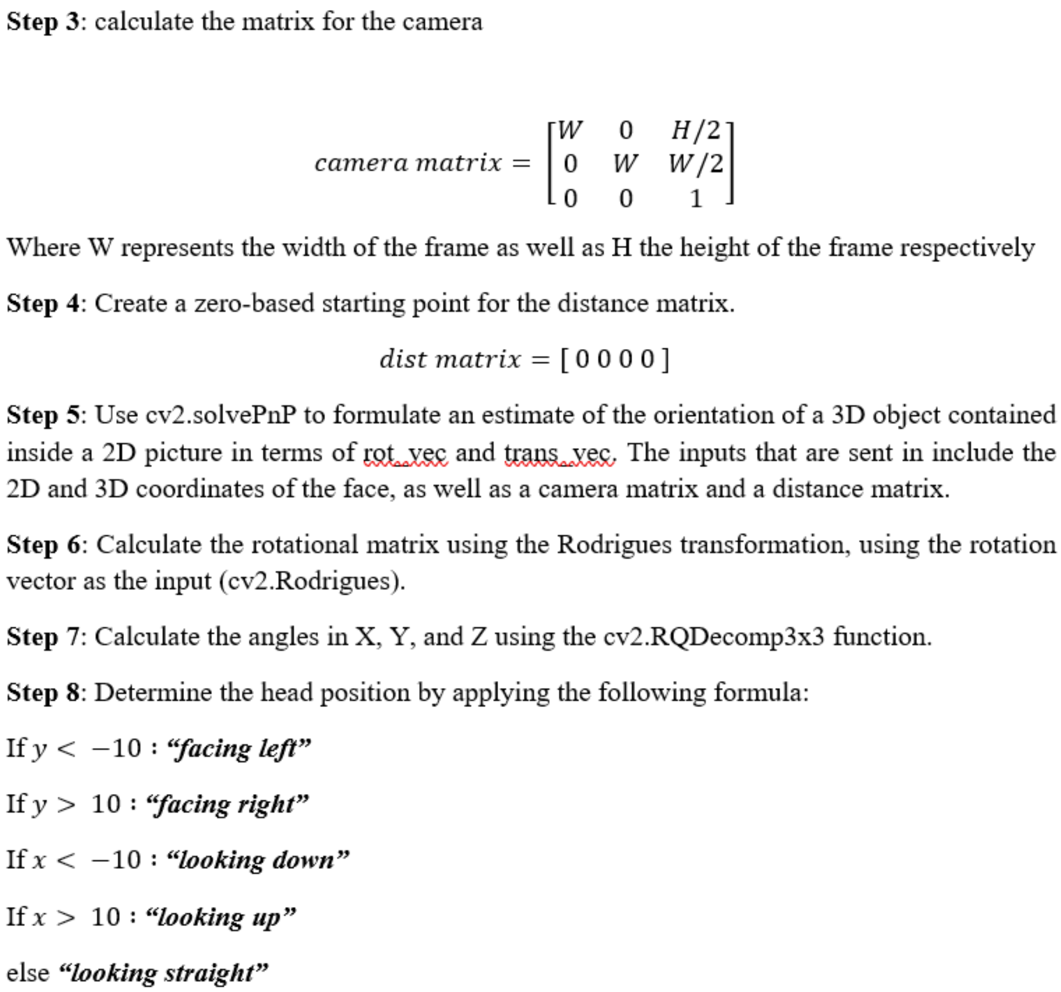
The steps in the process for determining the face angle are as follows:

Step 1: Obtain the X as well as Y coordinates of the mesh's six distinct face landmarks. The next step will make advantage of them. The coordinates 33, 263, 1, 61, 291, and 199 point to these recognizable sites on the map.

Step 2: Using the X and Y coordinates, get the 3D face coordinates as well as the 2D face coordinates.

The method is able to calculate the face angle by first extracting the X and Y coordinates of the required face landmarks from the mesh. This allows the angle to be calculated. These landmarks most likely relate to important facial characteristics that help determine the orientation and angle of the face. These traits contribute to determining how the face is oriented.

After the coordinates have been gathered, the algorithm will be able to continue with the computation of the 3D face coordinates, which will offer information about the face's placement in space. In addition, it is possible to obtain the 2D face coordinates, which provides information on the alignment of the face on a plane that is just two dimensions deep.



Step 3: calculate the matrix for the camera

$$camera\ matrix = \begin{bmatrix} W & 0 & H/2 \\ 0 & W & W/2 \\ 0 & 0 & 1 \end{bmatrix}$$

Where W represents the width of the frame as well as H the height of the frame respectively

Step 4: Create a zero-based starting point for the distance matrix.

$$dist\ matrix = [\,0\,0\,0\,0\,]$$

Step 5: Use cv2.solvePnP to formulate an estimate of the orientation of a 3D object contained inside a 2D picture in terms of rot_vec and trans_vec. The inputs that are sent in include the 2D and 3D coordinates of the face, as well as a camera matrix and a distance matrix.

Step 6: Calculate the rotational matrix using the Rodrigues transformation, using the rotation vector as the input (cv2.Rodrigues).

Step 7: Calculate the angles in X, Y, and Z using the cv2.RQDecomp3x3 function.

Step 8: Determine the head position by applying the following formula:

If $y < -10$: ***"facing left"***

If $y > 10$: ***"facing right"***

If $x < -10$: ***"looking down"***

If $x > 10$: ***"looking up"***

else ***"looking straight"***

4.2.1.3 Audio Processing

The monitoring of students' audio data as they participate in online courses may reveal instructive insights regarding the students' levels of interest and their involvement in a variety of activities outside of the classroom that they are participating in. These activities could involve talking to other students in the class or interacting with unrelated digital content, such listening to music or viewing movies. These kinds of activities have the ability to divert the attention of students and make it more difficult for them to concentrate while class is in session.

It is now feasible to identify the existence of extra voices or noises in the environment thanks to the improvements in technologies that are used for audio analysis. Some examples of these technologies are voice recognition software and audio pattern identification algorithms. These methods are intended to recognize variations in the audio stream, which enables the identification of whether or not there are other voices or distracting noises present.

These technologies may be set up to provide warnings or alerts if a predetermined threshold is breached, which significantly boosts the efficiency of the monitoring process. These warnings serve as signs that the student may not be completely engaged in the current lesson that is being delivered. NLP technologies make it possible to extract information and recognize certain words or phrases that point to a lack of attention on the part of the learner. This makes it possible to detect when a student is not paying attention. For instance, if a student is overheard expressing comments such as "I'm bored" or "This is so boring," it is a red flag that the student may not be actively engaged in the material that is being covered in the classroom and may not be paying attention to what is being taught.

Teachers are able to acquire significant insights regarding the amount of attentiveness displayed by their pupils if they utilize methodologies based on NLP and analyze the audio stream. This information enables them to adjust and change their educational tactics in a manner that is appropriate, so guaranteeing a more individualized and productive learning experience. Audio analysis and natural language processing, when combined, provide educators with a potent tool with which to assess the level of student involvement and arrive at well-informed judgments to enhance the entire learning experience.

In the fields of speech and audio signal processing, a method known as Mel-Frequency Cepstral Coefficients (MFCC) is often used for the extraction of features. Through the modeling of the frequency resolution and sensitivity of the human auditory system, it seeks to accomplish the goal of capturing the key properties of the spectral content of a sound source.

The operation of MFCC is comprised of the following stages:

1. **Applying a pre-emphasis filter** to the audio stream is the first stage in the process of adding pre-emphasis. This filter accentuates the high-frequency components, making up for the loss of higher frequencies that occurs during

the recording or transmission of sound. This is done by compensating for the loss of higher frequencies.

2. **Framing**: The signal to be pre-emphasized is segmented into frames that are quite brief, often ranging from 20 to 40 milliseconds, with a minimal amount of gap between each successive frame. The spectral features of the signal may be better captured throughout time as a result of this segmentation.
3. **Windowing**: This technique involves multiplying each frame by a window function, such as the Hamming window, in order to limit the amount of spectral leakage that is brought on by abrupt transitions at the frame borders. The signal that is there in each frame is made smoother by windowing.
4. **Fast Fourier convert (FFT)**: The windowed frames are then transformed from the time domain to the frequency domain using the FFT. The signal is changed from its time representation to its frequency components as a result of this conversion.
5. **Power Spectrum**: To get the power spectrum representation, the magnitude of the FFT bins is squared. This gives information on how the energy is distributed throughout the various frequency bands in each frame.
6. **The Mel Filterbank**: After that, the power spectrum is sent through a set of triangle filters that together make up what is known as the Mel filterbank. A perceptual scale known as the Mel scale attempts to simulate the way the human auditory system reacts to sounds of varying frequencies. As a result of the filterbank's processing, the power spectrum is efficiently segmented into a number of Mel frequency bins, significantly lowering the dimensionality of the data.
7. **Logarithm**: The logarithm of the filterbank energies is assumed to be an approximation of the logarithmic sense of loudness that the human ear has. The representation of greater amplitudes is compressed as a result of this conversion, which also improves the depiction of lower amplitudes.
8. **Discrete Cosine Transform (DCT)**: The logarithmically scaled filterbank energies are subjected to the DCT in the last step of the process. The mel-frequency cepstral coefficients (MFCCs) are the result of the DCT's decorrelation of the coefficients and their transformation into a new set of

coefficients with that name. Lower-order MFCCs are responsible for capturing the general shape of the spectrum, whereas higher-order coefficients are responsible for providing more specific spectral information.

The MFCCs that are created as a result of this method may be used as features in a number of audio processing applications, such as speaker identification, speech recognition, and musical genre classification. They are resistant to noise and fluctuations in the surrounding environment because they extract the most important spectral information from the signal.

An example to MFCC working is depicted here:

1. Take into consideration an audio signal consisting of spoken words.

2. The signal is segmented into frames that are quite short—let's say 30 milliseconds each—and there is a 50% gap between them.

3. In order to limit the amount of spectral leakage, a window function, such as the Hamming window, is applied to each frame.

4. Using the fast Fourier transform (FFT), the frames are converted into the frequency domain.

5. The power spectrum of each frame is calculated by quadrupling the magnitudes of the FFT bins in that frame.

6. The power spectrum is then sent through a series of Mel filters, which provide the output energies in a variety of frequency bands.

7. A logarithmic representation of the filterbank energies is computed and taken.

8. In order to acquire the MFCCs, the resultant values are sent via the DCT, which causes them to undergo a transformation.

9. The lower-order MFCCs are responsible for capturing the overall spectral structure, while the higher-order coefficients give more specific information on the spectrum.

10. The MFCCs may be used as input features for a variety of applications, such as voice recognition, in which case they may be fed into machine learning algorithms for the purpose of categorization or other tasks.

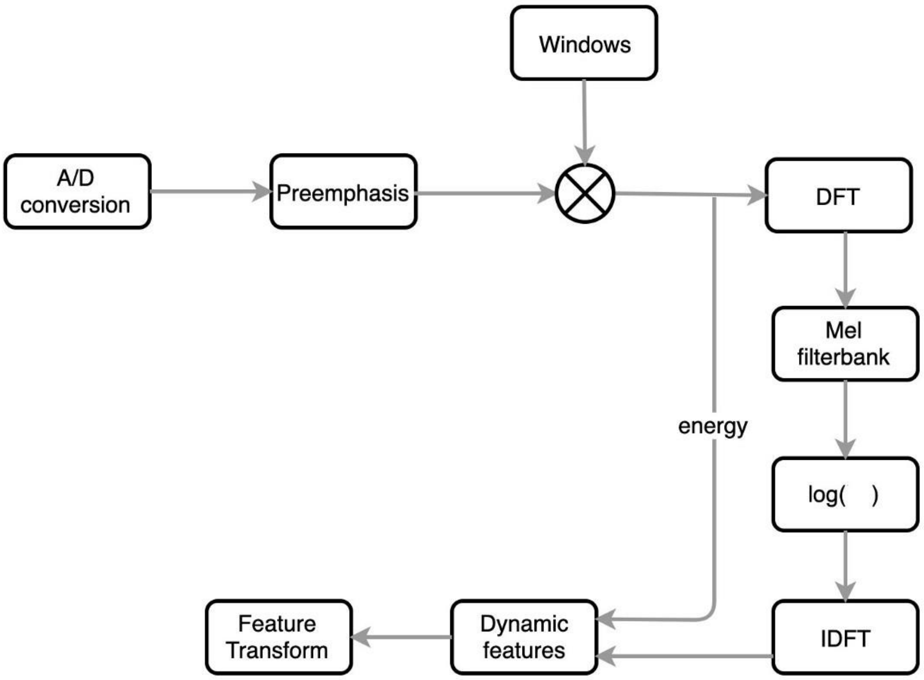


Figure 4. 4: MFCC feature extraction on the audio data

Dense Layer:

It is crucial to include what is recognized as a thick layer, also known as a fully linked layer, when creating neural networks. This layer is known as the "dense" layer because every neuron in it is connected to every neuron in the layer below it. A dense layer is one in which every neuron receives input from every neuron in the layer above it as well as sends its output to every neuron in the layer above it. Because of its extensive connectedness, the network is able to go through complicated nonlinear transformations and interact with its characteristics in new and interesting ways.

Dense Input Layer:

The first layer of a neural network is referred to as the dense input layer. It takes in the data, which can come in the form of numerical attributes or picture pixels, among other possible configurations. The dimensionality of the data that is being input influences the amount of neurons that are present in the dense input layer. Each neuron that makes up this layer of the network is a representation of a particular attribute or characteristic of the input.

Dense Hidden Layer:

Dense hidden layers are layers that are included in a neural network that are intermediate between the input and output layers. They are a very important component in the process of acquiring increasingly complex representations of the incoming data. The output of a dense hidden layer is calculated by having each neuron take the outputs of all the neurons in the layer below it as input. This input is then processed using a set of weights and biases, which ultimately results in the output of the dense hidden layer. The degree of difficulty of the challenge being tackled often has a direct bearing on the number of neurons that are contained inside a thick hidden layer.

Dense Softmax Layer:

It is usual practice to use a thick softmax layer as the output layer while doing classification jobs. The softmax activation function is applied to the input data in this layer, which results in the generation of a probability distribution over all of the potential classes. The outputs are normalized using the softmax function, which also ensures that they add up to one. This enables the outputs to be understood as class probabilities. Each neuron that makes up the dense softmax layer is a representation of the probability that the relevant class will occur.

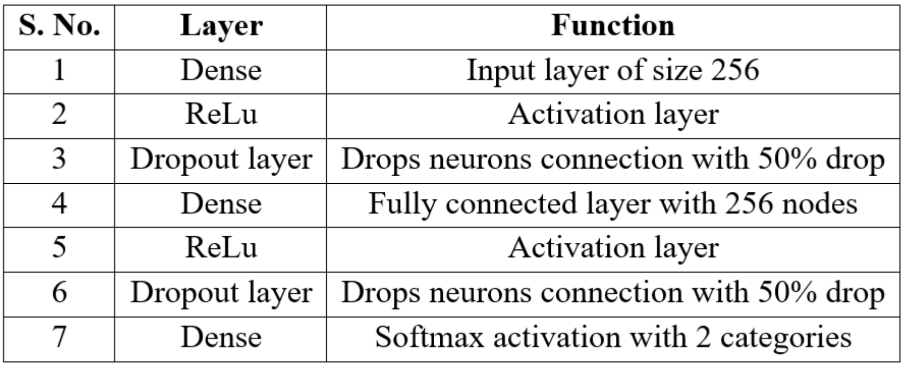
Dropout layer:

In neural networks, regularization strategies such as the dropout layer are used to minimize overfitting and to ensure that accuracy is preserved. During training, the dropout layer "drops out" a part of the input neurons by setting a random fraction of those neurons to zero in each update. This is done so that the neurons are effectively removed from the network. This avoids individual neurons from depending too strongly on certain inputs, which makes it easier for the brain to acquire qualities that are more resilient and can be used in a wider range of contexts. The network is less reliant on individual connections as a result of dropout, which improves its capacity to generalize to data that has not yet been observed.

By excluding neurons at random during the training process, the dropout layer is able to successfully produce an ensemble of smaller networks inside the main network. During the testing phase of the project, dropout will be turned off, and the complete network will be used for prediction purposes. It has been established that making use of dropout regularization may enhance the generalization performance of neural networks and prevent overfitting, especially in circumstances when the amount of training data is restricted.

The model architecture that has been recommended for use in audio classification is outlined in Table 1, which may be found here.

Table 4. 1: Proposed audio classification model architecture



S. No.	Layer	Function
1	Dense	Input layer of size 256
2	ReLu	Activation layer
3	Dropout layer	Drops neurons connection with 50% drop
4	Dense	Fully connected layer with 256 nodes
5	ReLu	Activation layer
6	Dropout layer	Drops neurons connection with 50% drop
7	Dense	Softmax activation with 2 categories

The proposed model for CNN is made up of numerous layers, each of which is described in Table 1. Within the framework of this design, the ReLu layer is positioned just beneath the original thick layer. After that, a dropout layer is used to get rid of any unwanted connections that were made. After this comes the additional layers, which may be broken down into three categories: dense, dropout, and ReLu. The usage of a thick layer serves as the process's finishing touch, bringing to a close the classification phase.

4.2.1.4 Heuristic Features

The heuristic features make a variety of user-accessible choices available, and some examples of these possibilities are as follows:

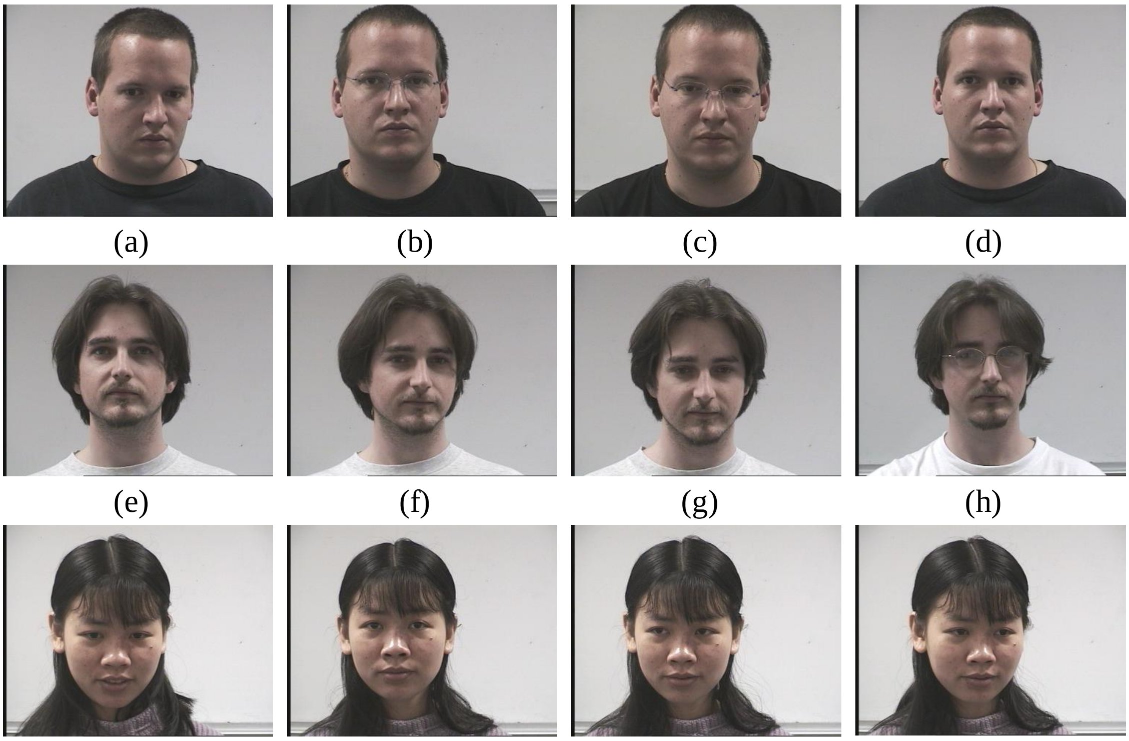
- **Activation and deactivation of the camera**It won't be necessary for the user to switch on their smartphone's camera again when using the app if it is already turned on. In a similar vein, the user will not be requested to switch off the camera even if they are using it even if they have selected an option that allows it to be on at all times.
- **An option to mute or unmute the microphone** will be shown to the user if they have ever muted the microphone in the past. If they have done so, they will be given the opportunity to unmute it. On the other hand, the user will not have this option available to them if they have never silenced the microphone before.
- **Problems associated with networks**: The operation of the system is designed to accommodate for network-related difficulties, such as unanticipated pauses in video playback or streaming speeds that are much lower than planned.

4.3 Experimental Results

In this part, a comprehensive overview of the experimental findings that were carried out in order to confirm the correctness of the suggested model is provided. The model is composed of a number of different modules, each of which makes a contribution to an all-encompassing evaluation of pupil attentiveness. The total efficacy of the student's attention is assessed by combining these modules in order to provide a more complete picture.

4.3.1Video processing

The initial part of the method to handle the given video data is the face detection module. This happens right at the start of the procedure. If the face is not recognized, the user will get a prompt to adjust their head position towards the camera. If you didn't recognize the face, you'll see this prompt. The data associated with a face is supplied to a deep learning model once it has been correctly recognized in a frame so that the face angle may be calculated. 297 photos were obtained from the frontal viewpoint for the training dataset, while 474 images were collected from the side view for the dataset. The photographs from the dataset that show the person staring straight ahead are shown in Figure 4.5. These images are designated "a" to "l." All of the shots taken from the side that are included in the dataset may be seen in these pictures, which are grouped together in Figure 4.6 as the letters a to l.

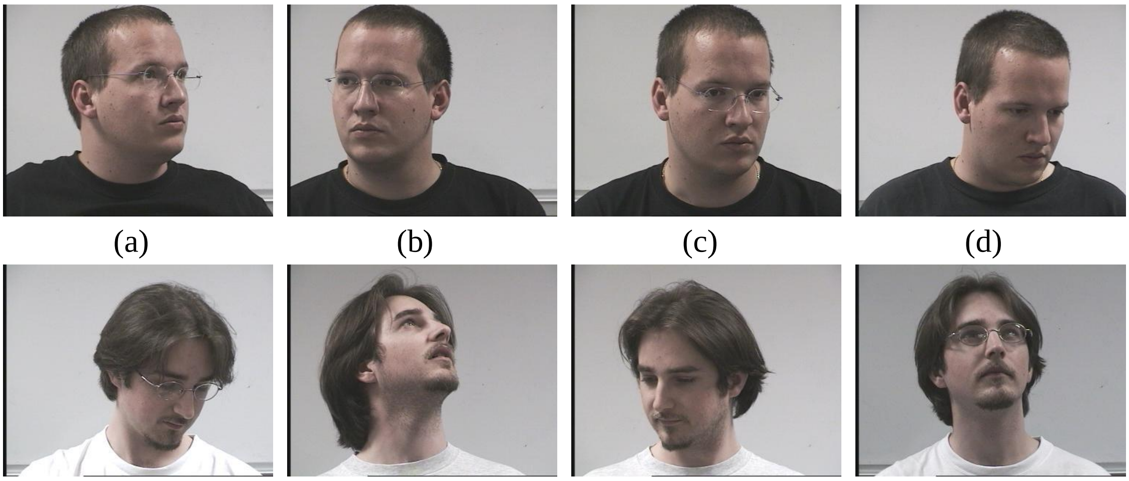


(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 4. 5: Images of the front view require to be included into the database.



(a) (b) (c) (d)

(e) (f) (g) (h)

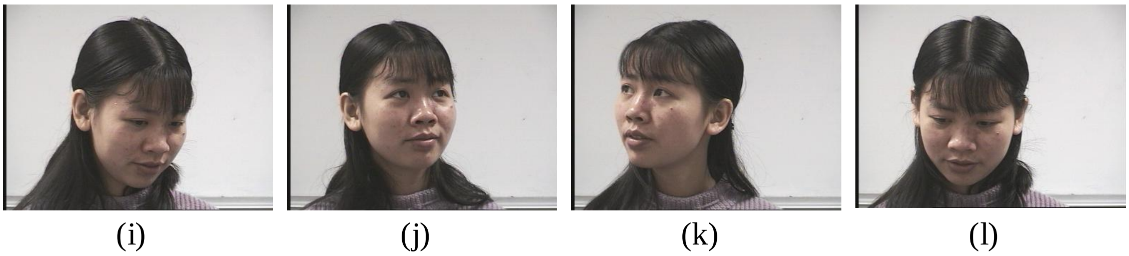
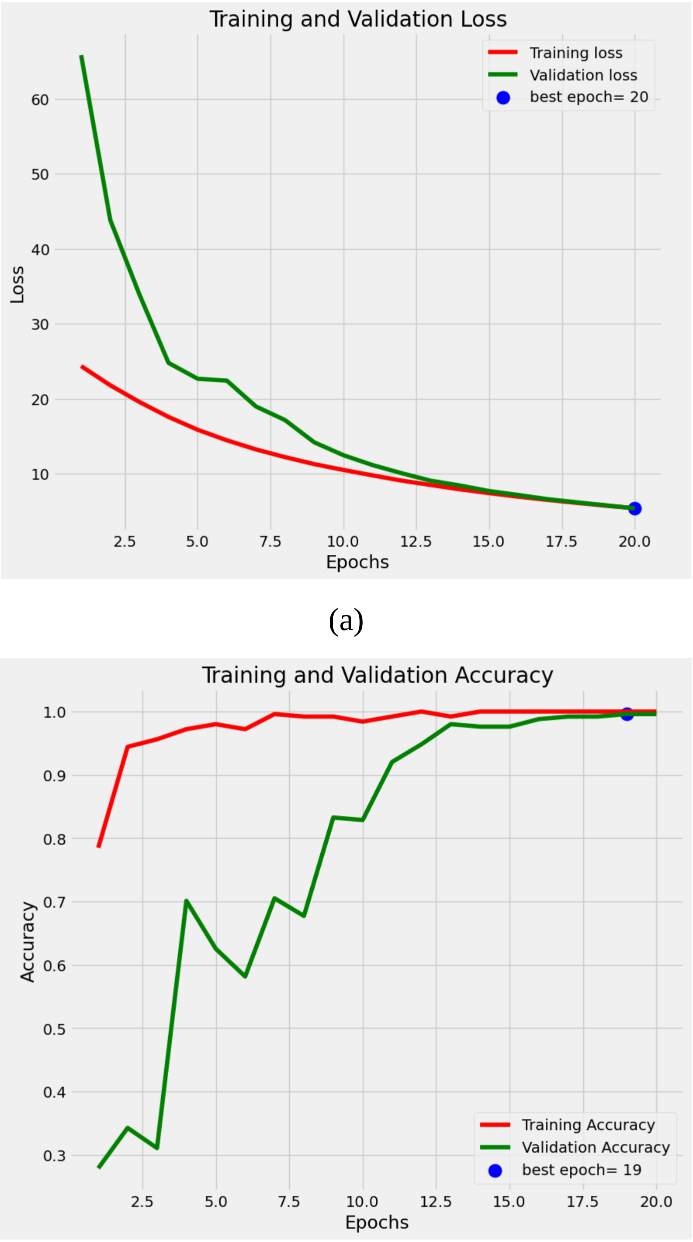

(i) (j) (k) (l)

Figure 4. 6:Enter non-frontal images to the database.

The images are entered into the MobileNetV2 model.



(a)

(b)

Figure 4. 7: (a) Loss function and (b) Accuracy function

Figure 4.7 (a) displays the loss function. The "training loss" is the discrepancy between the output that the model predicted and the output that was actually seen while the model was being trained. Backpropagation enables the model to modify its biases and weights to lessen the amount of training loss. A low training loss is a sign that the model has effectively learned the patterns in the training data and is able to make reliable predictions. On the other hand, a significant training loss indicates that the model has difficulty picking up on the patterns.

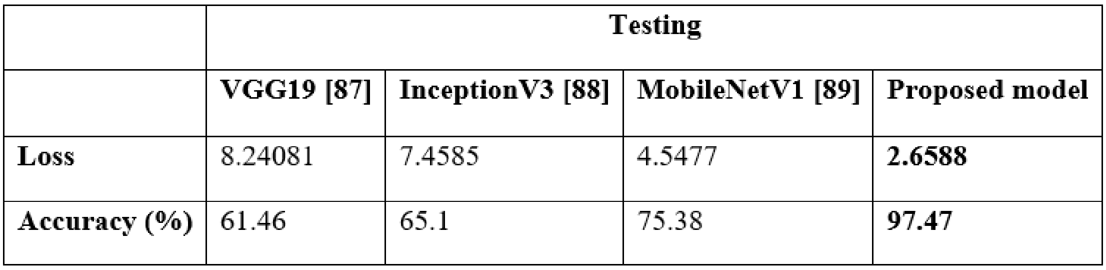
The discrepancy between the output that the model anticipated and the output that actually happened for validation data is known as validation loss. It's possible that this difference will lead to incorrect predictions. Using a unique dataset that the model was not exposed to during training, the validation loss is determined. It evaluates the model's ability to extrapolate its findings to data that wasn't previously taken into account. When the validation loss significantly surpasses the training loss, overfitting has occurred. This serves as a warning that, should the problem continue, the model may not generalise effectively to new data. On the other hand, a validation loss that is much less than the training loss demonstrates effective generalisation. In essence, the training loss measures how well the model matches the training data, whereas the validation loss assesses the model's ability to generalise. Both are considered to be warning signs of loss.

Both the training and the validation are shown to have high levels of accuracy in figure 7b. When analyzing a machine learning model while it is being trained, important metrics to consider include the quality and accuracy of the training and validation data, as well as the amount of data that is being used. The phrase "training accuracy" refers to the proportion of occurrences in the training data that have been properly categorised. During the training process, backpropagation is used to make the model's weights and biases more accurate, which ultimately leads to better outcomes. A model may be correctly able to learn patterns and make accurate predictions on the training data if it has a high accuracy on the training set.

The term "validation accuracy" refers to the proportion of occurrences in the validation data that have been properly categorised. This accuracy is assessed by utilizing a different dataset, one that the model has not seen before when it was being trained.It measures how effectively the model can extrapolate from fresh data. There was inadequate generalisation and overfitting if the accuracy obtained from the validation is significantly lower than the accuracy of the training.

The VGG19, InceptionV3, MobileNetV1, and Proposed models all earned accuracy scores of 61.46, 65.10, 75.38, and 97.47 correspondingly while being tested. Table 4.2 displays these results for you to go through.

Table 4. 2: Comparison of Testing parameters



	Testing			
	VGG19 [87]	**InceptionV3 [88]**	**MobileNetV1 [89]**	**Proposed model**
Loss	8.24081	7.4585	4.5477	**2.6588**
Accuracy (%)	61.46	65.1	75.38	**97.47**

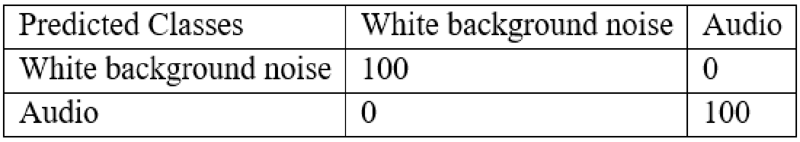
4.3.2 Audio Processing

The audio input is made up of a variety of arbitrary audio samples, together with a continuous backdrop of white noise. In addition to the 400 samples that make up the white noise in the background, there are also 400 samples that are comprised of the random audio samples. The whole audio input is made up of these 800 separate samples when they are all combined. The remaining 600 samples were utilized for strength training, while the first 200 were set aside for relaxation reasons. In all, there were 800 samples. The inquiry resulted in the collection of a total of eight hundred samples.

200 of these samples were test samples and consisted of a mix of audio samples and the background white noise. These test samples were included in the total. However, in terms of the aggregate count, only the audio samples were taken into consideration. During the part of the process when the model was being tested, each sample was properly identified based on the model that was recommended. The confusion matrix

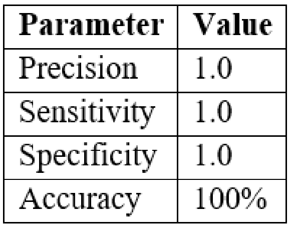
that corresponds to the categorization findings is shown pictorially in Table 4.3, which may be found below.

Table 4. 3: Confusion matrix



Predicted Classes	White background noise	Audio
White background noise	100	0
Audio	0	100

Table 4. 4: Evaluation parameters



Parameter	**Value**
Precision	1.0
Sensitivity	1.0
Specificity	1.0
Accuracy	100%

Table 4.4, provided in this document, presents the evaluation criteria for different sound classes. It reveals a perfect relationship between sensitivity and specificity, both of which have a value of one hundred percent. This indicates a strong correlation between these two measures. Furthermore, when precision, sensitivity, and accuracy are combined, their total equals one hundred percent.

4.4 Conclusion

In recent years, online education, sometimes known as e-learning, has caused major disruption to the educational practices traditionally used in classrooms. On the other hand, because of the distinctive features of online learning, it could be challenging for professors to monitor their students' attention and gauge how far ahead in the course they are. The findings of this study have led to the formulation of a fresh strategy for the control of attention. In addition to a wide variety of other inputs, this method places emphasis on certain aspects of the face. The model's purpose is to assess the degree of attention shown by pupils by gathering information from the visual feed of the camera, the aural input, and heuristic qualities. The teachers are able to periodically see the screens of the pupils in order to better comprehend the activities that are taking place. Through the use of the video stream, the model does a detailed analysis of the students' facial expressions. At first, the image is submitted to the

OpenCV face recognition module so that it may analyze the picture and locate and recognize the student's face. After that, the picture is analyzed using certain modules that assess the face angle, such as Google's facial landmark recognition and MobileNetV2's angle analysis. In order to properly calculate the face angle, the model takes into account a variety of different variables. The ambient noise from the audio input and the computer's heuristic inputs, in addition to the facial attributes, are watched over and analysed. While the suggested CNN model was able to identify audio samples with a precision of 100%, the suggested model could only recognise students' gaze angles with a precision of 97.47%. This study presents a novel approach to evaluating attention and reveals important insights that may be used in the delivery of online instruction.

Chapter 5

Micro-Expression Modelling based Facial Emotion Recognition in E-Learning using modified ResNet50V2

5.1 Introduction

Online classes require self-motivation, discipline, and effective time management skills, as students are often responsible for managing their own study schedules and staying engaged with the course material. Students who are able to stay organized, actively participate in discussions, and complete assignments on time are more likely to maintain a higher learning rate.The learning rate of students in online classes can vary depending on various factors. Some students may adapt well to the online learning environment and maintain a high learning rate, while others may face challenges and experience a slower learning rate.

However, there are potential barriers to learning in online classes that can affect the learning rate. These include distractions at home, limited access to resources or technology, difficulties in understanding the course content without face-to-face interaction, and a lack of personal connection with the instructor and peers. These challenges can hinder the learning process and lead to a slower learning rate for some students.

It is crucial that institutions and instructors provide students who study online enough assistance and tools to improve their learning experience. This may include clear instructions, regular communication, interactive learning activities, and opportunities for collaboration and feedback. Additionally, fostering a sense of community and engagement among students can contribute to a more conducive online learning environment, which can positively impact the learning rate of students.

A modified version of the ResNet50V2 model is presented in this chapter in order to identify and categorize the microexpressions that are seen on human faces. The fundamental goal of this concept is to build a relationship between the microexpressions that students display in online classrooms and the emotions that they are experiencing.

Understanding students' experiences with online learning greatly depends on their use of microexpressions, which are quick, uncontrollable facial expressions that convey genuine feelings. These fleeting reactions can reveal information about students emotional and cognitive states in real time, including engagement, irritation, uncertainty, and excitement. Teachers can get quick feedback on their students' understanding and emotional responses to the topic they are teaching by examining microexpressions. This allows them to modify their teaching strategies accordingly. For example, frequent displays of confusion may suggest that more explanation of a concept is required. By ensuring that students stay interested and supported, incorporating cutting-edge facial recognition technologies to identify these microexpressions can improve the efficacy of online learning. However, in order to preserve a trustworthy and encouraging.

VGG16 is able to overcome the constraints of previous designs because it makes use of residual learning. This allows for the building of substantially deeper networks and addresses the problem of the vanishing gradient. The suggested model makes use of ResNet, which enables it to acquire more complicated representations of the input data. This is an essential capability for successfully identifying nuanced facial emotions, which are often difficult to discern.

5.2 Proposed model

There is a strong link between emotions and the way in which one's face is expressed, and the two are intricately connected to one another. Expressions on the face are visual indications of the emotional states that are occurring on the inside of our heads. They convey vital information to other people and help facilitate communication in social contexts. The connection between our feelings and the expressions on our faces may be analyzed from a number of different angles, including the following:

1. **The perspective of evolution** It is thought that facial expressions originated as adaptive reactions to distinct emotions, which allowed for the quick and efficient conveyance of thoughts. Charles Darwin postulated that some facial expressions are generally shared throughout cultures, implying that there is a biological foundation for the existence of certain expressions. For instance, a

grin is traditionally related to joy, but a frown is more often associated with melancholy or displeased expressions. These universal face expressions facilitate efficient transmission of emotions across diverse nations, showing a common human experience. They may be found in all cultures.

2. **The Facial Feedback Hypothesis** According to the facial feedback hypothesis, the way in which a person expresses themselves via their facial expressions has the capacity to influence the emotional state that they are now experiencing. The facial movements that are linked with certain emotions, such as smiling or frowning, carry messages to the brain, which ultimately leads to the activation or amplification of related emotional experiences. Happiness may be communicated via smiling, whereas despair is communicated through frowning. Notably, the results of the study reveal that individuals who faked a smile by holding a pen between their teeth reported feeling happier than those who imitated a neutral face by holding the pen between their lips. This was the case despite the fact that all groups were given the same instructions. This provides evidence that one's mental state may be affected by their facial expressions.
3. **Identifying Emotions** One of the most important aspects of identifying and comprehending the feelings of other people is paying attention to the expressions on their faces. The capacity to read other people's emotions and facial expressions comes naturally to humans, and this talent allows them to communicate effectively with one another. Because of this capacity, one is able to effectively empathize with others, respond appropriately, and alter their conduct. Numerous studies have shown that humans can reliably recognize fundamental emotions such as happy, sorrow, rage, fear, disgust, and surprise based purely on facial expressions, even across various cultures. These emotions include happiness, sadness, anger, fear, disgust, and surprise.
4. **The Influence of Cultural and Contextual variables** Although certain facial expressions are instantly recognized no matter where you are in the world, the perception and presentation of emotions via facial expressions may also be impacted by cultural and contextual variables. When it comes to the presentation of emotions and the conventions that govern their expression,

many cultures might have quite different standards and guidelines. In addition, one and the same facial expression might have several connotations based on the circumstances surrounding the interaction. A raised eyebrow, for instance, may communicate a variety of emotions, including surprise, skepticism, or seduction, depending on the situation in which it is shown.

5. **Microexpressions and Emotional**Microexpressions are fleeting facial expressions that are involuntary and occur as a reaction to emotions that are being repressed or hidden from view. Even if a person is making an effort to conceal or disguise their true emotions, these brief expressions, which only endure for a fraction of a second, typically indicate what those emotions really are. Microexpressions are regarded crucial clues in interpreting underlying emotions, especially in circumstances in which people may be attempting to mislead or hide their genuine sentiments. This is because microexpressions tend to be more subtle than full facial expressions.

Face expressions and emotions are intricately interwoven with one another. Facial expressions have the ability to reflect and transmit our emotional experiences, to impact our emotional state through the facial feedback mechanism, to assist in the detection and comprehension of the emotions shown by others, and to be influenced by cultural and environmental variables. It is essential for efficient communication, empathy, and social interaction to have a firm grasp of the link that exists between facial expressions and a person's feelings.

5.2.1 Learning rate assessment using emotions expression

When it comes to education delivered over the internet, a student's facial expressions and emotions may give instructive insight into their degree of concentration as well as the progress they have made in their studies. In spite of the fact that there is no physical presence and no instant feedback in online learning, the degree of involvement, comprehension, and focus that students are experiencing may still be deduced from their facial expressions. The thorough explanation of the link between facial expressions, emotions, and learning progress in the context of online education is provided in the following paragraphs:

1. **Engagement and Interest**: A student's facial expressions may provide insight into the level of engagement and interest they have in the material being discussed. Learners' demonstrations of true engagement and interest in the content may be deduced from the fact that they show signs of inquiry, attention, and attentiveness. For instance, the facial expressions of a student who looks to be deeply involved in their educational experience and to be enjoying themselves may include raised eyebrows, a gaze that is concentrated, and a relaxed facial expression. On the other hand, indicators of apathy or boredom, such as a blank look, drooping eyelids, or frequent yawning, indicate a lack of participation and the need for further aid or adaptations to the learning material. Both children and adults have their own unique ways of expressing emotions via their faces.
2. **Knowledge and Understanding**: A student's facial expressions may also give insight into the amount of knowledge and understanding that they have of the material that is being delivered to them. Students are considered to have comprehended and absorbed the material when they display facial expressions such as smiling, nodding, or other signs of recognition, agreement, or happiness. On the other hand, facial expressions that imply uncertainty, irritation, or wrinkled brows may indicate that the subject matter is difficult to comprehend. The instructors are able to determine the areas in which the students may need further explanations, illustrations, or clarification if they observe the students' facial expressions.
3. **Emotional States and Levels of Stress**: facial expressions reflect a wide range of feelings, which can influence learning outcomes emotional states and levels of stress can have an impact on how well someone learns. According to a body of academic research, positive feelings like pleasure, interest, and enthusiasm considerably improve one's ability to assimilate and remember new information. A happy emotional state that is beneficial to learning is indicated by facial expressions such as smiling, raised cheeks, and sparkling eyes. On the other hand, having negative feelings such as annoyance, bewilderment, or concern might make it more difficult to learn. Students who are confronted with obstacles or who experience feelings of being

overwhelmed may show symptoms of struggle, such as wrinkled brows, strained facial muscles, or fearful faces. When teachers are able to recognize these signs, they are better equipped to provide students assistance, reassurance, or ideas in an effort to lower their levels of stress and enhance their learning results.

4. **Participation and Non-Verbal Signals**: In online learning contexts, when physical presence is missing, non-verbal signals play a significant role in conveying important information. Students' participation and interest in debates, group activities, or virtual classrooms may be gleaned from their facial expressions, which might give useful information. Positive facial gestures, such as smiling, nodding, and making direct eye contact are common behaviors that accompany active participation. On the other hand, indications of disengagement or lack of interest such as blank looks, averted gazes, or restricted facial expressions may be present. The ability to monitor these indications gives educators the ability to assess the level of student interest and adjust their instructional tactics appropriately.
5. **Facial expressions** are another way in which the existence of technological obstacles and distractions that are interfering with students' educational experiences may be detected. Students' reactions to technology faults or disruptions might be interpreted from expressions of discontent, distraction, or frustration on their part. When instructors are able to recognize these facial cues, it is easier for them to detect and rapidly resolve any technological flaws that may occur. This helps to ensure that students have a consistent and uninterrupted learning experience.

During online learning, the expressions and emotions that students show on their faces may provide highly helpful information into the students' levels of participation, engagement, and comprehension of the subject matter. Educators have the potential to adapt their teaching strategies, offer students with additional assistance, and create a learning environment that is more engaging and productive for the students if they monitor and properly interpret these indications. The advantages of using emotions to assess the attentiveness of students in inline classes are as follows:

1. **Immediate input** Emotions may offer immediate feedback on students' levels of involvement and concentration during online courses. This feedback can be gleaned from students' facial expressions, gestures, and voice clues. This gives instructors the ability to evaluate their pupils' level of comprehension of the subject being taught in real time.
2. **Personalized support**: It is possible to detect students who may be having difficulty or are not interested in their online lessons by assessing the emotional states of the students. Additional explanations, resources, or encouragement may be provided by teachers to students who are exhibiting indications of uncertainty, irritation, or boredom so that they can intervene and give focused help to those kids.
3. **An increase in motivation**: When it comes to motivation, emotions play a crucial influence. The instructors are able to spot times of increased interest, enthusiasm, or curiosity in their pupils by monitoring the kids' emotional states. Because of this, instructors are able to devise activities that are interesting and adapt their teaching methods to maintain students' enthusiasm and encourage active involvement.
4. **Adaptation and customization**: Instructional adaptations and customization may be informed by emotion-based evaluation, which can be done in online classrooms. For instance, if a student repeatedly exhibits indications of annoyance or boredom during certain themes, instructors might adjust their teaching techniques, add multimedia resources, or offer interactive activities to make the information more accessible and entertaining. This is especially helpful for students who have difficulty reading.
5. **Impact on students' mental well-being:**As a result of online education's potential to be alienating and difficult environments for learning 5. Educators are able to detect pupils who are dealing with mental well-being or emotional discomfort by conducting an assessment of the students' emotions. Teachers are thus in a position to provide the needed assistance, direct students to

counseling resources, or encourage them to communicate with their peers in order to improve their social connectivity.

6. **Insights that are driven by data**: Emotion-based evaluation produces useful data that may be used for further investigation and development. Educators and researchers may acquire insights into common issues, engagement patterns, or areas in which teaching approaches might be improved by collecting and evaluating emotional data from several students.

It is very necessary to perform evaluation based on students' feelings in an ethical manner that respects their right to personal privacy. When collecting data, there should be open communication, informed permission, and absolute confidentiality maintained at all times. This will ensure that the data will only be used for educational reasons.

5.2.2 Emotion Modelling

In the context of computer systems or artificial intelligence frameworks, the term "emotion modeling" refers to the process of comprehending human emotions, properly expressing those emotions, and faithfully duplicating those emotions in a true manner. It entails recording and analyzing many components of human emotion, such as facial expressions, physiological reactions, linguistic patterns, and environmental clues, in order to develop a model that is capable of swiftly identifying, correctly understanding, and effectively reacting to emotions.

Emotion modeling's overall goal is to give robots the ability to understand and react to human feelings, which will hopefully lead to interactions that are more engaging and empathic. The following is a condensed version of the essential elements that make up emotion modeling:

1. **Emotion Recognition** The first stage of emotion modeling is the process of recognizing the various emotions that different people experience. This involves recording emotional indicators via a variety of modalities, such as facial expressions, voice tone, speech patterns, body language, and physiological signals. Physiological clues include heart rate and blood pressure. Techniques such as computer vision, audio processing, and natural language processing are used in order to analyze and interpret these signals,

which are then sorted into emotional categories such as happy, sorrow, anger, fear, and surprise using the results of this procedure.

2. **Representation of Emotions** After emotions have been identified, they need to be represented in a way that robots can comprehend and make use of. Emotions may be represented using a number of different approaches, including category models (such as Ekman's six essential emotions), dimensional models (such as valence-arousal models), and hybrid models (combining categorical and dimensional features). These representations provide a structured framework for accumulating and arranging the information on one's feelings.
3. An awareness of one's emotional context is necessary for a complete comprehension of one's feelings. This awareness is not possible apart from the context in which feelings are felt. Contextual awareness is included into emotion modeling, which takes into consideration the personal, cultural, and environmental elements that have an effect on the emotions experienced by an individual. Recognizing the context in which expressions or physiological responses take place is essential for deciphering emotions, since the same signals may have multiple meanings depending on the circumstances in which they are shown. Additionally, it allows computational systems to develop emotional reactions that are more individualized and aware of their surroundings.
4. **Emotion Generation**: An additional essential component of emotion modeling is the process of producing emotional responses or expressions based on the emotions that have been identified. This may be accomplished via a variety of methods, such as the generation of relevant textual answers, the synthesis of expressive speech, the creation of animated avatars with appropriate facial expressions and motions, or even the construction of robotic systems that are capable of showing emotional behaviors. Emotion generation is being developed with the intention of making interactions between humans and robots more interesting and sympathetic for all parties involved.
5. **Applications**: There are a variety of contexts in which emotion modeling may be used. It improves the user experience in human-computer interaction by

allowing computers to adapt to the emotions of users and give tailored recommendations or support. This is accomplished via the use of artificial intelligence. In the field of education, emotion modeling may be used to develop adaptive learning environments that react to the emotional states of students by offering feedback or interventions that are specifically designed for them. In addition, it may be used in areas such as customer service, virtual reality, entertainment, healthcare, and other business-related fields.

6. **Ethical Considerations** Emotion modeling presents ethical considerations surrounding the handling of data, as well as issues of privacy and consent. It is of the utmost importance to manage emotional data in a responsible and transparent manner, ensuring that user consent is obtained, data is anonymized, and that abuse is avoided. It is necessary to ensure that people' privacy is protected, since individuals' feelings are private and sensitive information.

The fields of psychology, neurology, computer science, and artificial intelligence all contribute ideas to the topic of emotion modeling, which is an interdisciplinary field. The advancement of this technology might one day lead to the creation of emotionally intelligent robots that are able to comprehend and react to the feelings of humans in a meaningful and sympathetic way.

5.2.3 Micro Expressions

- Anger is a fundamental emotion that might include feelings of rage, irritation, and/or hostility. Anger may arise from a number of different situations. It may be brought on by a number of different things, and the look on one's face can often give away their level of wrath. This facial expression is characterized by particular alterations in the muscles of the face, such as raised eyebrows, constricted eyelids, flared nostrils, a tense jaw, pinched lips, flushed or red cheeks, and other manifestations with a similar appearance.
- The feeling known as annoyance is a universally experienced emotional response that incorporates the emotions of aggravation, irritation, and frustration. Anger is a globally recognized emotion that is defined by an

expression of irritation, and the facial expression associated with anger is a face expression of irritation.

- An emotional response known as anticipation is characterized by the expectancy or thrill of something that is going to take place. The facial expression of anticipation is characterized by the eyebrows being lifted, the eyes being expanded, a grin or lips that are parted, a tilted head, and other emotions that are comparable.
- A worldwide emotional response known as aversion is characterized by feelings of hate, revulsion, or disgust. Aversion is a globally recognized emotional reaction that is connected with emotions of unhappiness, and its facial expression is known as the "aversion face." The appearance of furrowed eyebrows, tighter eyelids, smaller eyes, wrinkled noses, pinched lips, and other related changes are common face changes associated with aging.
- Confidence is a universal emotional reaction that comprises the emotions of self-assurance, self-assuredness, along with self-possession. These feelings are accompanied with a generally recognized facial expression that conveys a sense of self-assurance: the confident smile. The eyebrows are lifted, the eyes are enlarged, a relaxed or smiling look, a raised head, and other similar expressions are typical facial changes that occur when one is feeling confidence.
- The emotion of disapproval, which is present all throughout the globe, is characterized by feelings of disappointment or dissatisfaction. A universal emotional response characterised by displeasure is called disapproval. A globally recognized emotional reaction that is related with sentiments of disappointment or dissatisfaction is the facial expression of disapproval. When a person is experiencing a negative feeling such as disapproval, their facial expression will often include furrowed eyebrows, constricted eyelids, smaller eyes, pinched lips, and a sagging chin.
- Disquiet is an all-encompassing emotional reaction characterised by feelings of disquiet, discomfort, or unease. Disquiet is a globally known emotional reaction that may be described as feelings of uneasiness or discomfort. The

facial expression of disquiet is characterized by this. Someone who feels anxious will likely have furrowed or knitted brows, narrowed or squinted eyes, pinched or tight-lipped lips, a tensed jaw or clenched teeth, change their focus, or fidget.

- The universally experienced feeling of humiliation is defined by feelings of shame, humility, or discomfort. Embarrassment is a worldwide emotional response. These sensations are commonly understood to be accompanied with the facial expression of embarrassment, which is a globally acknowledged emotional reaction. It might include a reddening of the cheeks, eyes that are cast downward or away from the subject, a nervous smile or chuckle, face-covering motions, and other similar expressions.
- The feelings of enthusiasm, eagerness, or thrill may be used to characterize excitement, which is a universally experienced emotional response. The emotional reaction of excitement, as seen on one's face, is a generally recognized facial expression that is connected with sentiments of enthusiasm. Normal expressions of emotion on the face include grinning, eyes that are brilliant or wide open, eyebrows that are lifted, and an open or relaxed lips.
- Fatigue is a generalised emotional reaction that is characterised by feelings of exhaustion, weariness, or tiredness. One kind of emotional response that is often associated with these feelings is the exhausted facial expression. The appearance of drooping eyelids, sunken or sleepy eyes, slouching posture, a dull or pale complexion, a lack of facial expressiveness, and other similar indications are typical facial changes associated with aging.
- An instinctive and ubiquitous emotional response, fear may be characterized by feelings of unease, dread, or alarm. The expression of terror on one's face, sometimes known as "the chicken face," is characterized by a number of facial changes, including wider eyes, raised eyebrows, and an open or stiff lips, all of which are typical manifestations of the emotion of fear.
- A worldwide emotional response that might be defined by emotions of joy, pleasure, or satisfaction, happiness is a global emotional reaction. These good feelings are widely connected with the facial expression of happiness, which is

a globally recognized emotional reaction. When a person is happy, their face will often go through a number of different modifications, including but not limited to:

- confined to grinning, puffing up one's cheeks, brightening one's eyes, relaxing one's forehead, and either opening or relaxing one's lips.
- A universally experienced emotional state, pain may be characterized by feelings of distress, agony, or suffering. Pain is a globally known emotional reaction, and its face expression is easily identifiable. A furrowed brow, clinched jaw, squinted eyes, grimacing, and flaring nostrils are common facial alterations that often accompany pain. There is also a noticeable arching of the brow and constriction of the pupils.
- A worldwide emotional response that might be defined by the feelings of sorrow, disappointment, or grief is referred to as sadness. The facial expression of melancholy is an emotional reaction that is instantly recognizable the world over and is connected with these sentiments. Changes in facial expressions such as drooping eyebrows, furrowed brows, downward corners of the mouth, watery eyes, and sluggish blinking are often seen in people who are unhappy.
- A universally experienced emotional state, suffering is characterized by intense sensations of sorrow, agony, or discomfort. The look on someone's face that communicates their level of anguish, misery, or discomfort is a generally recognized representation of an emotional reaction. Alterations in facial expressions, such as a furrowed brow, clinched jaw, squinting eyes, grimacing, and flaring nostrils, are often seen in people who are emotionally distressed.

- Surprise is a worldwide emotional response that may be defined by the feelings of wonder, shock, or disbelief. Surprise can also be caused by unexpected events. Astonishment is a generally known emotion, and the facial expression of surprise is a widely recognized emotional reaction connected with those sensations. Changes in facial expressions that are often seen include those in which the eyebrows are lifted, the eyes are enlarged, the mouth is open, and other similar expressions.

5.2.4 Emotion and Learning rate

The relation between emotion and learning rate is as follows:

Learning Rate = Baseline Learning Rate + (Emotional Intensity × Emotional Valence)

1. Baseline Learning Rate: This represents the base level of a student's learning rate without considering facial emotions. It could be derived from historical data or estimated based on average performance.

2. Emotional Intensity: This factor represents the strength or magnitude of the facial emotion displayed by the student. It can be measured on a scale from low to high, indicating the intensity of the emotion expressed.

3. Emotional Valence: This factor represents the positive or negative valence of the facial emotion. Positive valence emotions (e.g., happiness, excitement) may positively impact learning, while negative valence emotions (e.g., sadness, frustration) may negatively impact learning. It can also be measured on a scale from low to high.

The formula suggests that higher emotional intensity and positive emotional valence would lead to an increased learning rate, while higher emotional intensity and negative emotional valence would result in a decreased learning rate. Conversely, lower emotional intensity would have a minimal impact on the learning rate.

5.2.5 VGG16 Transfer Learning Model

The Visual Geometry Group (VGG) at Oxford University developed a deep convolutional neural network architecture called VGG16. The "VGG" in the name stands for the group working on it, and the "16" indicates the total number of layers. VGG16 stands out from other architectures due to the exceptional performance it achieves despite its relative simplicity.

When it was first released, the VGG16 algorithm was a revolutionary step forward in the picture classification and object identification processes. The deep network structure of the design was the most important innovation that it included. It consisted of a total of 16 weight layers, 13 of which were convolutional and 3 of which were completely linked. Because of this depth, VGG16 was able to collect more complex information in pictures, which greatly improved its ability to recognize objects and patterns in a diverse array of visual data.

The VGG16 was noteworthy for having a straightforward and consistent design, which was one of its distinguishing characteristics. Along with max-pooling layers, the design of the network made use of tiny convolutional filters measuring 3x3 and having a stride size of 1 pixel at various points across the network. In comparison to some prior designs for convolutional neural networks, this consistent design made it much simpler to train and comprehend the system.

Because of its outstanding performance in a number of benchmark image classification contests, particularly the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014, VGG16 has attracted a great deal of attention and acquired a lot of reputation in the community of deep learning researchers. In the competition, it produced results that are considered to be state-of-the-art and greatly reduced the error rate in comparison to earlier approaches.

The VGG16 architecture has 13 convolutional layers and 3 fully linked layers. To learn complex characteristics from images, each convolutional layer has a tiny 3x3 filter and a rectified linear unit (ReLU) activation function. Interspersed across the network are max-pooling layers, which downsample the feature maps to minimize computational load.

Training deep neural networks from scratch may be computationally costly and requires a significant quantity of data to be labeled. Examples of such networks are VGG16. Transfer learning, on the other hand, provides an effective solution by using models that have been pre-trained on extensive datasets such as ImageNet. ImageNet is a vast dataset that contains millions of annotated photos from a variety of different categories. This enables the model to learn general characteristics from varied visual input.

When it comes to modeling emotions, transfer learning entails doing further training on an already-trained VGG16 model using a particular dataset of images that are associated with certain emotions. We are able to modify the model such that it can distinguish certain emotional patterns by retraining just the most recent few levels of the network while leaving the beginning layers unaltered.

1. Input Layer: The input layer receives the input images and sets the dimensions for the subsequent layers.
2. Convolutional Layer (Conv1_1): The first convolutional layer with a 3x3 filter and ReLU activation.
3. Convolutional Layer (Conv1_2): Another 3x3 convolutional layer with ReLU activation.
4. Max Pooling Layer (Pool1): A max-pooling layer with a 2x2 window to downsample the feature maps.
5. Convolutional Layer (Conv2_1): A 3x3 convolutional layer with ReLU activation.
6. Convolutional Layer (Conv2_2): Another 3x3 convolutional layer with ReLU activation.
7. Max Pooling Layer (Pool2): Another max-pooling layer with a 2x2 window.
8. Convolutional Layer (Conv3_1): A 3x3 convolutional layer with ReLU activation.
9. Convolutional Layer (Conv3_2): Another 3x3 convolutional layer with ReLU activation.
10. Convolutional Layer (Conv3_3): A third 3x3 convolutional layer with ReLU activation.
11. Max Pooling Layer (Pool3): Another max-pooling layer with a 2x2 window.
12. Convolutional Layer (Conv4_1): A 3x3 convolutional layer with ReLU activation.
13. Convolutional Layer (Conv4_2): Another 3x3 convolutional layer with ReLU activation.

14. Convolutional Layer (Conv4_3): A third 3x3 convolutional layer with ReLU activation.
15. Max Pooling Layer (Pool4): Another max-pooling layer with a 2x2 window.
16. Convolutional Layer (Conv5_1): A 3x3 convolutional layer with ReLU activation.
17. Convolutional Layer (Conv5_2): Another 3x3 convolutional layer with ReLU activation.
18. Convolutional Layer (Conv5_3): A third 3x3 convolutional layer with ReLU activation.
19. Max Pooling Layer (Pool5): Another max-pooling layer with a 2x2 window.
20. Dense Layer (FC1): The first fully connected layer with 4096 neurons and ReLU activation.
21. Dense Layer (FC2): The second fully connected layer with 4096 neurons and ReLU activation.

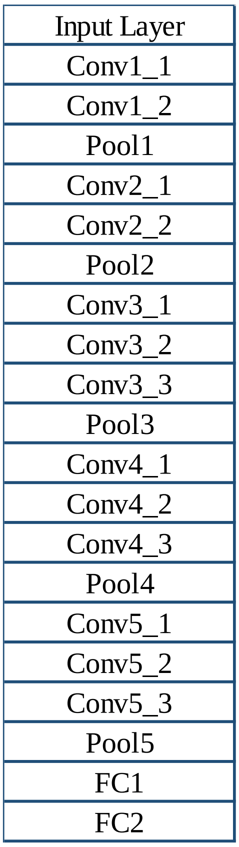


Figure 5. 1: VGG16 architecture

The classification layers of the conventional layers are removed and the features form the input face images are extracted from the FC2 layer.

Advantages of VGG16:

- A convolutional neural network (CNN) architecture known as VGG16 has recently seen a meteoric rise in popularity in the fields of computer vision and deep learning. This is mostly because to the many benefits that it provides. The following is a list of some of the primary benefits offered by VGG16:
- Simplicity and uniformity: The simplicity and consistency of the VGG16 design is one of the system's key selling points. It is made up of 16 weight layers, all of which are convolutional or fully linked layers, which makes it simple to comprehend and put into practice in many contexts. Because of this simplicity, the process of adapting or expanding the network in order to do certain tasks is made easier, which is beneficial for both researchers and practitioners.

- Deep Representation Learning: Due to the depth of the VGG16 network, it is able to learn complicated hierarchical features from picture data. Because it is composed of numerous layers that are piled above one another, it is able to recognize complex patterns and abstractions within the data. picture classification, object recognition, and picture segmentation are just some of the applications that benefit greatly from this kind of deep representation learning.
- Transfer Learning: VGG16's pre-trained models are easily accessible and have been trained on big datasets such as ImageNet. This makes them suitable for use in transfer learning. Because of this, it is an effective method for acquiring transferable skills. Researchers and developers are able to utilize the pre-trained VGG16 as a starting point for a variety of computer vision applications, and then fine-tune the network using more limited datasets that are particular to the job at hand. This results in a considerable reduction in the quantity of labeled data as well as the amount of time needed for training for new applications.
- Skillful Performance in Feature Extraction: VGG16 performs very well when it comes to feature extraction. Its convolutional layers recognize basic characteristics in the beginning layers before progressively building up to more complex features in the deeper layers as the system progresses. These layers operate as effective feature detectors. Because of this, it is useful for activities such as the transfer of styles and the development of pictures that require the extraction of relevant information from images.
- The Availability of Pre-trained Models TensorFlow and PyTorch, two of the most popular deep learning frameworks, make it easy to get pre-trained VGG16 models. Because of this accessibility, integrating VGG16 into a variety of applications is made more easier, which enables developers to take use of its capabilities without having to start from scratch.
- broad Applicability: VGG16 has proved its adaptability by performing well on a broad range of computer vision tasks, including image classification, object identification, picture segmentation, and a variety of other tasks. Because it

can easily be adapted to a wide variety of uses, it is an excellent option for a lot of different situations that might arise in the real world.

- Performance on Benchmark Datasets: VGG16, with its deep architecture and huge number of parameters, has consistently produced competitive performance on benchmark datasets such as ImageNet. In situations when performance is of the utmost importance, its accuracy continues to be a significant benefit, despite the fact that it may not be the most effective method in terms of the calculation and memory needs.

Because of its ease of use, deep representation learning skills, potential for transfer learning, and extensive application, VGG16 is a key tool in the toolkit of computer vision researchers and practitioners of machine learning. Even if newer architectures have been developed since it was first introduced, the fact that VGG16 continues to be relevant highlights the importance it has in the area of deep learning for computer vision.

i. Convolution Neural Network (CNN) model

The historical trajectory of Convolutional Neural Networks (CNNs) encompasses a significant duration, characterized by notable advancements and significant landmarks within the domain of computer vision and deep learning. Convolutional Neural Networks (CNNs) have significantly transformed the field of image processing and pattern recognition, serving as a crucial component in a wide range of applications such as picture categorization, object identification, and even medical diagnosis.

The origins of Convolutional Neural Networks (CNNs) may be attributed to the work of neuroscientists David Hubel and Torsten Wiesel in the 1950s, whereby they did groundbreaking research on the feline visual cortex. The researchers' findings revealed the existence of both basic and sophisticated cells that exhibited selective responsiveness towards certain characteristics within the visual domain, such as edges and textures. The notion of local receptive fields and feature hierarchies in artificial neural networks was established based on biological inspiration.

The inception of contemporary Convolutional Neural Networks (CNNs) may be ascribed to the research conducted by Yann LeCun and his colleagues throughout the latter part of the 1980s and the early years of the 1990s. LeCun was responsible for the development of the LeNet-5 architecture, which was specifically built to address the issue of recognizing handwritten digits. The LeNet-5 model is credited with pioneering the incorporation of convolutional layers, pooling layers, and fully connected layers into neural networks, hence introducing these fundamental concepts to the field.

Nevertheless, the implementation and extensive use of Convolutional Neural Networks (CNNs) encountered obstacles as a result of constrained processing capabilities and restricted data accessibility at this particular period. The widespread use of Convolutional Neural Networks (CNNs) occurred in the mid-2000s, mostly due to advancements in graphics processing units (GPUs) and the accessibility of extensive datasets, such as ImageNet. The integration of technology and data has facilitated the training of deep Convolutional Neural Networks (CNNs), resulting in significant advancements in the accuracy of picture categorization.

A key milestone in the history of CNN occurred in 2012 with the introduction of the AlexNet architecture by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. This architectural innovation led to a notable advancement in the ImageNet Large Scale Visual Recognition Challenge. The success of AlexNet served as a demonstration of the better performance of deep convolutional neural networks (CNNs) and played a pivotal role in initiating the deep learning revolution.

Subsequent to the introduction of AlexNet, a series of advancements and structural designs surfaced, with each endeavoring to expand the capabilities of Convolutional Neural Networks (CNNs). Several noteworthy models in the field of deep learning include VGGNet, GoogLeNet, and ResNet, which have made significant contributions via their distinct design aspects and architectural enhancements. These networks played a crucial role in attaining picture recognition performance at par with human capabilities and propelling the progress of computer vision research.

In recent times, Convolutional Neural Networks (CNNs) have expanded their scope beyond image processing and have been used in several domains such as natural language processing, voice recognition, and autonomous driving. The main role of artificial intelligence systems in current times is attributed to their adaptability and resilience.

In summary, the historical development of Convolutional Neural Networks may be characterized as a process of ongoing advancement and ingenuity, driven by the convergence of neuroscientific insights, enhanced computing capabilities, and extensive datasets. From its inception in the examination of biological vision to the revolutionary advancements of the 21st century, Convolutional Neural Networks (CNNs) have emerged as an essential instrument in the realm of deep learning, significantly influencing our understanding and engagement with the visual domain. The progression of their adventure continues, as forthcoming advancements and implementations loom on the horizon.

Layers in CNN model:

1. **BatchNormalization()**: A method known as batch normalization is used in order to standardize the output of the layer that came before it. Through the reduction of internal covariate shift, it contributes to the stabilization and acceleration of the training process. Batch normalization normalizes the values of each feature map in the batch to have zero mean and unit variance. This normalization step ensures that the model can learn more efficiently and reduces the dependence on the scale of the input data.

2. **GlobalAveragePooling2D()**: The spatial dimensions of the feature maps are collapsed into a single value by the Global Average Pooling 2D layer, which does this by calculating the average of all the values included inside each feature map. The overall number of parameters in the model is decreased as a result of this process, which also contributes to the prevention of overfitting. Global Average Pooling captures the essential features of the input images in a more concise representation, and it is particularly useful when dealing with variable-sized inputs.

3. **Flatten()**: In order to create a 1D vector from the 2D feature maps, the Flatten layer must first be employed. Before transferring the data to a fully linked (dense) layer, this step is required to be completed. In order to get it ready for the dense layers that come after it, it "flattens" the output of the layer that came before it into a one-dimensional array.

4. **Dense(256, activation='relu')**: This adds a fully connected (dense) layer with 256 neurons and ReLU activation function. The use of ReLU (Rectified Linear Unit) activation helps introduce non-linearity to the model. Additionally, L2 regularization is applied to both the kernel (weights) and bias of the layer. Regularization helps prevent overfitting by adding penalty terms to the loss function during training.

5. **BatchNormalization()**: Another Batch Normalization layer is added after the first dense layer. This further normalizes the output of the dense layer, enhancing the stability of the model and improving training efficiency.

6. **Dropout(0.5)**: During the training process, the regularization method known as dropout causes a portion (in this case, fifty percent) of the inputs to the layer below it to be set to zero. This helps to minimize overfitting by minimizing the model's dependence on any one particular neuron, and it also fosters more robust representations. The value 0.5 indicates that 50% of the neurons will be randomly dropped during each training iteration.

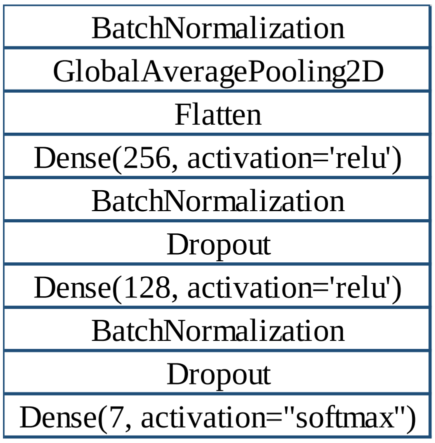
7. **Dense(128, activation='relu')**: Another fully connected (dense) layer with 128 neurons and ReLU activation function is added, along with L2 regularization and Batch Normalization.

8. **BatchNormalization()**: Another Batch Normalization layer after the second dense layer.

9. **Dropout(0.5)**: Another Dropout layer with a dropout rate of 50%.

10. **Dense(7, activation="softmax")**: The final dense layer with 7 neurons (assuming there are 7 emotion classes to predict) and a softmax activation function. The softmax function converts the final output into a probability

distribution, indicating the likelihood of each class. This layer produces the final predictions for the emotion classes based on the features learned throughout the network.



BatchNormalization

GlobalAveragePooling2D

Flatten

Dense(256, activation='relu')

BatchNormalization

Dropout

Dense(128, activation='relu')

BatchNormalization

Dropout

Dense(7, activation="softmax")

Figure 5. 2: CNN architecture

Advantages of CNNs:

CNNs have emerged as a fundamental component within the domain of computer vision and image processing, owing to its many benefits. The following are few notable benefits associated with Convolutional Neural Networks (CNNs):

- Feature Learning: Convolutional neural networks (CNNs) automatically learn important features from the input data. CNNs have the ability to adapt and find the characteristics that are most relevant for the job at hand. This is in contrast to the classic computer vision approaches, which often need the creation of features manually.
- Translation Invariance: CNNs are well suited for applications in which the position of features in the input data is irrelevant to the job at hand. They are able to identify patterns regardless of where in the input picture they are located, making them resistant to changes in item placement and orientation.
- Hierarchical Representation: CNNs are made up of numerous layers, each of which learns more abstract and complicated characteristics than the one before it. This hierarchical representation enables them to grasp pictures at multiple levels of abstraction by allowing them to capture both low-level characteristics

(such as edges and corners) and high-level information (such as object parts and textures). For example, edges and corners are low-level features, while object parts and textures are high-level features.

- Parameter Sharing: Convolutional Neural Networks (CNNs) use weight sharing across convolutional layers, which drastically decreases the total number of parameters when compared to fully connected networks. This makes CNNs easier to train efficiently and helps avoid them from becoming overfit.
- Local Receptive Fields: In order to process the input, CNNs make use of relatively tiny local receptive fields. Because of this, they are able to concentrate on smaller, more localized portions of the input at a time, which is often more physiologically realistic and efficient for the processing of pictures.
- Spatial Hierarchy: CNNs collect spatial hierarchies in data in an unintentional and natural way. Higher layers are responsible for learning broader patterns and global structures, whereas lower layers are responsible for learning the finer details. Because of this, CNNs are well-suited for tasks like as identifying objects in photos or comprehending the links between different locations.
- Data Augmentation: CNNs may gain an advantage from the use of data augmentation methods, which entail transforming the training data in some way (for example, by rotating it, flipping it, or cropping it). The resilience of the model and its ability to generalize to new data is improved as a result of this.
- Transfer Learning: CNN models that have been pre-trained on big datasets (such as ImageNet) may be fine-tuned for particular tasks using just a little amount of data using this technique. This method of learning via transfer not only helps save time and computing resources, but it also often results in increased performance.
- Parallelization: Convolutional neural networks (CNNs) are capable of being parallelized in an effective manner, making use of the latest GPU technology.

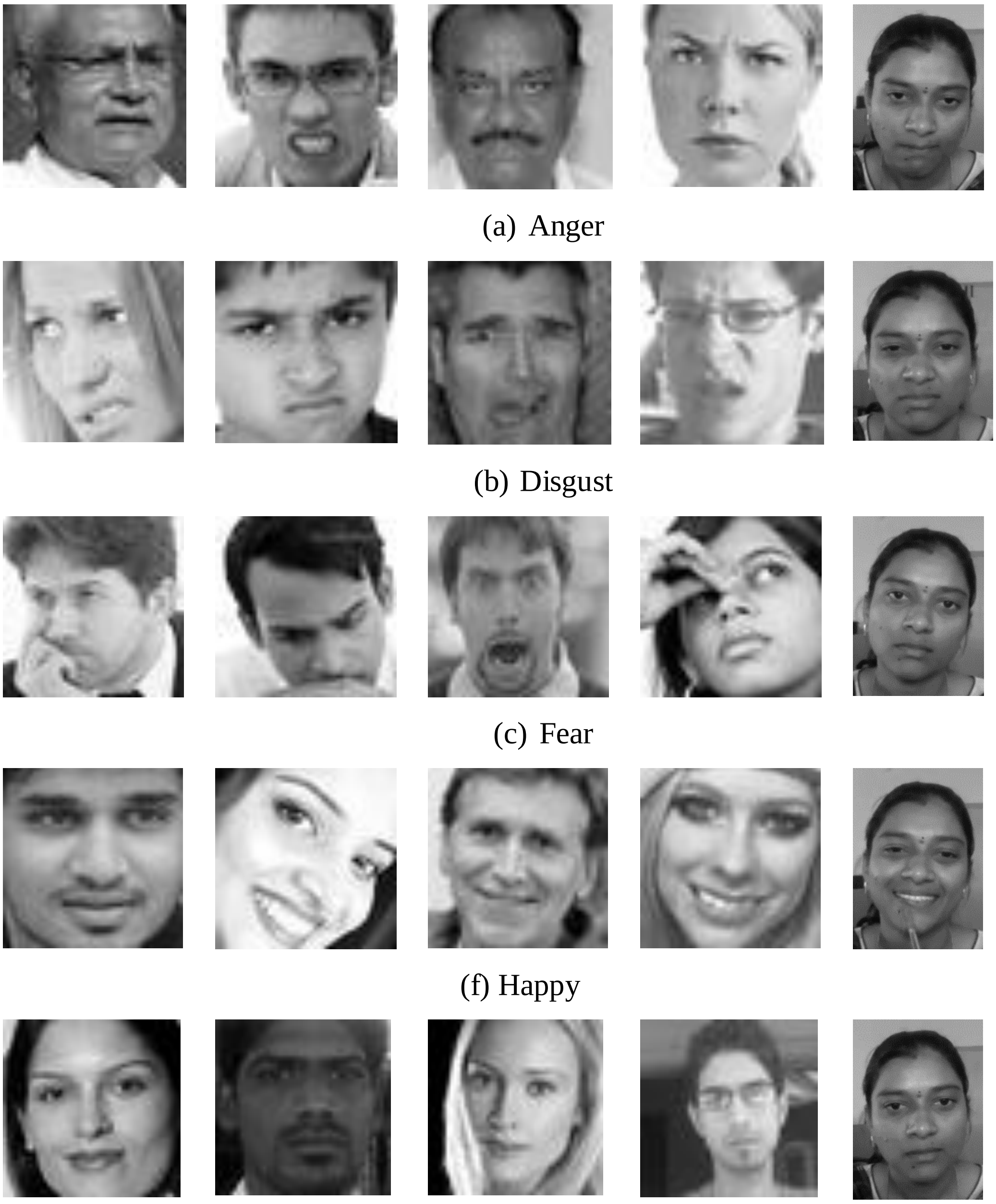
This enables quicker training and inference, making them appropriate for real-time applications as well as jobs that need a great amount of processing power.

- CNNs have consistently demonstrated state-of-the-art performance in a variety of computer vision applications, such as image classification, object identification, picture segmentation, and many more. They have established new standards in a wide variety of contests and challenges.
- Robustness to Noise: Convolutional neural networks (CNNs) are well-known for their capacity to process noisy data, which qualifies them for use in real-world applications in which the input may be imperfect or noisy.
- Interpretability: While CNNs are not as interpretable as other conventional approaches, they do include techniques like as feature visualization and occlusion analysis that may give insights into what areas of an image the model is focused on. These insights can help with debugging and understanding the choices that the model is making.

Convolutional Neural Networks have, in a nutshell, brought about a revolution in the area of computer vision. They do this by providing a robust and versatile framework for learning and extracting features from visual input. As a result, these networks are now crucial for a broad variety of applications in the field of image processing and beyond.

5.3 Experimental Analysis

The construction of a customized facial expression recognition database needed the resolution of many significant challenges, with a particular emphasis on addressing the problem of class imbalance. In order to address this issue in a comprehensive manner and promote a fair allocation of images across various emotional categories, rigorous measures were implemented to gather and classify an equal quantity of images for each unique emotional expression. These expressions encompassed anger, disgust, fear, happiness, neutrality, sadness, and surprise. The adoption of this well-rounded strategy not only facilitated enhanced acquisition of knowledge and the ability to apply it to different emotional conditions within the model, but also guaranteed the absence of any partiality towards a certain category of emotional expression. The dataset was enhanced by including real-time face data acquired from users, which introduced a dynamic element alongside the static photographs. The integration of both static and dynamic expressions yielded a complete resource for the training and evaluation of facial expression recognition systems. Consequently, this advancement enabled enhanced precision and resilience in the detection of emotions across several domains, including human-computer interaction and the realm of affective computing.



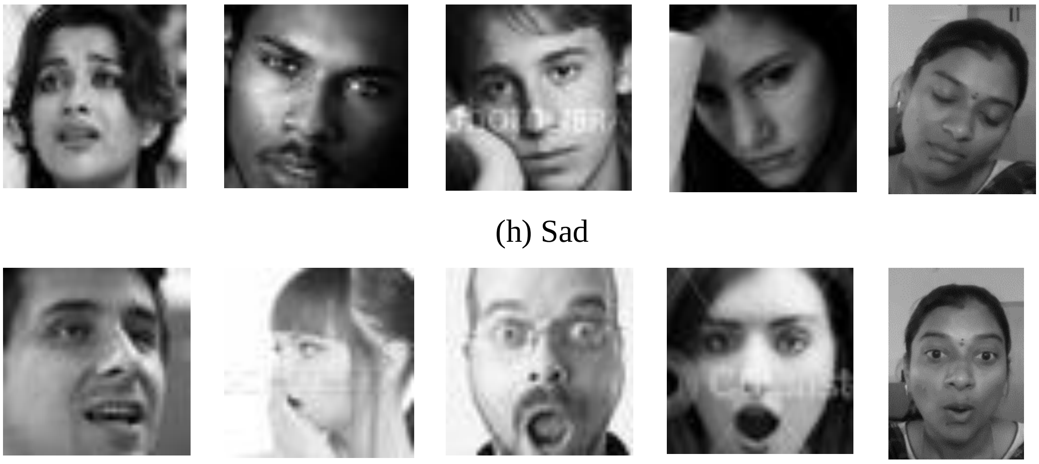
(a) Anger

(b) Disgust

(c) Fear

(f) Happy

(g) Neutral



(h) Sad

(i) Surprise

Figure 5. 3: sample images from custom dataset

This section presents the experimental results carried out to validate the proposed model. Sample input images are shown in figure 5.4.

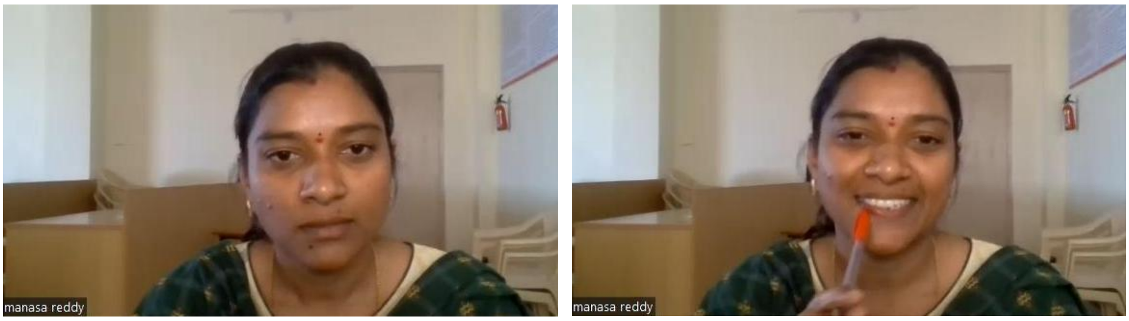


Figure 5. 4: Image frames from Video footage

These image frames are enhanced with the expression labels, which are then used as input to the proposed model. Proposed architecture's first step involves performing facial recognition and estimating future predictions.

Table 5. 1: Emotion vs learning coefficient

Emotion	**Learning Coefficient**
Angry	-0.5
Frustration	-0.8
Disgust	-0.6
Confusion	-1.0
Fear	-0.4

Dissatisfaction	-1.0
Sad	-0.8
Neutral	0.2
Peace	0.5
Happy	1.0
Pleasure	0.8
Surprise	0.9
Excitement	1.0
Confidence	1.0

The overall learning rate takes into account the emotional state of a student and its corresponding impact on their learning process. Each emotion is associated with a learning coefficient, which represents the extent to which that emotion influences the student's learning. The learning rate is calculated by considering the probability of each emotion being expressed on the student's face and then multiplying it with the corresponding learning coefficient.

5.3.1 Emotion Probability:

The proposed facial emotion estimation method is used to estimate the probability of each emotion being expressed on the student's face. These probabilities are denoted as P(Angry), P(Frustration), P(Disgust), P(Confusion), P(Fear), P(Dissatisfaction), P(Sad), P(Neutral), P(Peace), P(Happy), P(Pleasure), P(Surprise), and P(Excitement). These probabilities sum up to 1 as they represent the likelihood of the student experiencing different emotions at that moment.

5.3.2 Learning Coefficient:

The learning coefficients corresponding to each emotion, denoted as LC(Angry), LC(Frustration), LC(Disgust), LC(Confusion), LC(Fear), LC(Dissatisfaction), LC(Sad), LC(Neutral), LC(Peace), LC(Happy), LC(Pleasure), LC(Surprise), and LC(Excitement). These coefficients range from -1.0 to 1.0 and reflect how the specific emotion affects the student's learning process. A positive coefficient indicates a positive impact on learning, while a negative coefficient indicates a negative impact.

5.3.3 Calculating Overall Learning Rate:

The overall learning rate is obtained by considering the weighted sum of the learning coefficients, weighted by the probabilities of each emotion:

$$\begin{aligned} & Overall\ Learning\ Rate \\ & \quad = P(Angry) * LC(Angry) + P(Frustration) * LC(Frustration) \\ & \quad + P(Disgust) * LC(Disgust) + \ldots + P(Excitement) \\ & \quad * LC(Excitement) \end{aligned}$$

For instance, if the student is showing a high probability of being "Happy," and the learning coefficient for "Happy" is 1.0, the overall learning rate would be positively influenced. On the other hand, if the student is expressing "Confusion" with a high probability and the learning coefficient for "Confusion" is -1.0, the overall learning rate would be negatively impacted.By considering the emotions and their respective learning coefficients, the overall learning rate provides insights into how the student's emotional state might influence their learning performance. It helps educators and researchers to understand and cater to the emotional aspects of the learning process, which can contribute significantly to the student's academic outcomes and overall well-being.

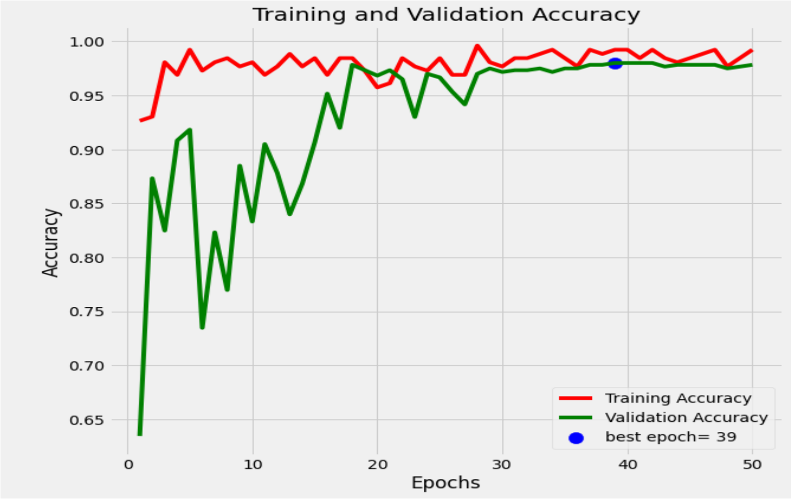


Figure 5. 5: Training and Validation Accuracy of the Proposed model

The fraction of images within the training dataset that have been correctly classified serves as the yardstick for determining how accurate the training was. The model under consideration is fine-tuned during the training phase depending on the mistakes it experiences while working with the training dataset. This process takes place as the training phase advances. It is typical for the model's accuracy with the training dataset to increase over time as it continues to learn from the training data it interacts with. This is because it is expected that the model will continue to learn from the training data. On the other hand, validation accuracy refers to the fraction of correctly classified photos that are included inside the validation dataset. This dataset is separate from the training data and serves the goal of testing the performance of the model on new data that it has not previously been exposed to.

The major purpose of the validation dataset is to discover if the model has inappropriately conformed to the training data. This may be determined by comparing the validation data to the training data. If anything like this were to occur, the results of the model when applied to fresh data may be below par. The suggested model is being trained, and its accuracy using the validation dataset is being checked at sporadic intervals during the training process. When the validation accuracy steadily becomes better, this is a good indicator since it shows that the model is successfully learning the capacity to generalize and is not displaying any overfitting tendencies when it comes to the training data. On the other hand, if the model's validation accuracy remains stable or declines, this might be an indication that the model is possibly overfitting to the training data. It's possible that it will be required to make some modifications in order to prevent this scenario from occurring.

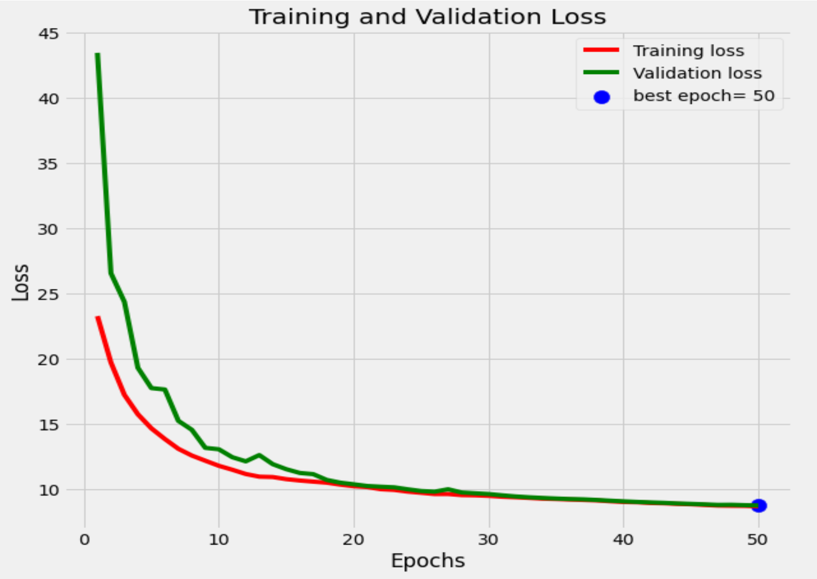


Figure 5. 6: Training and Validation Loss of the Proposed model

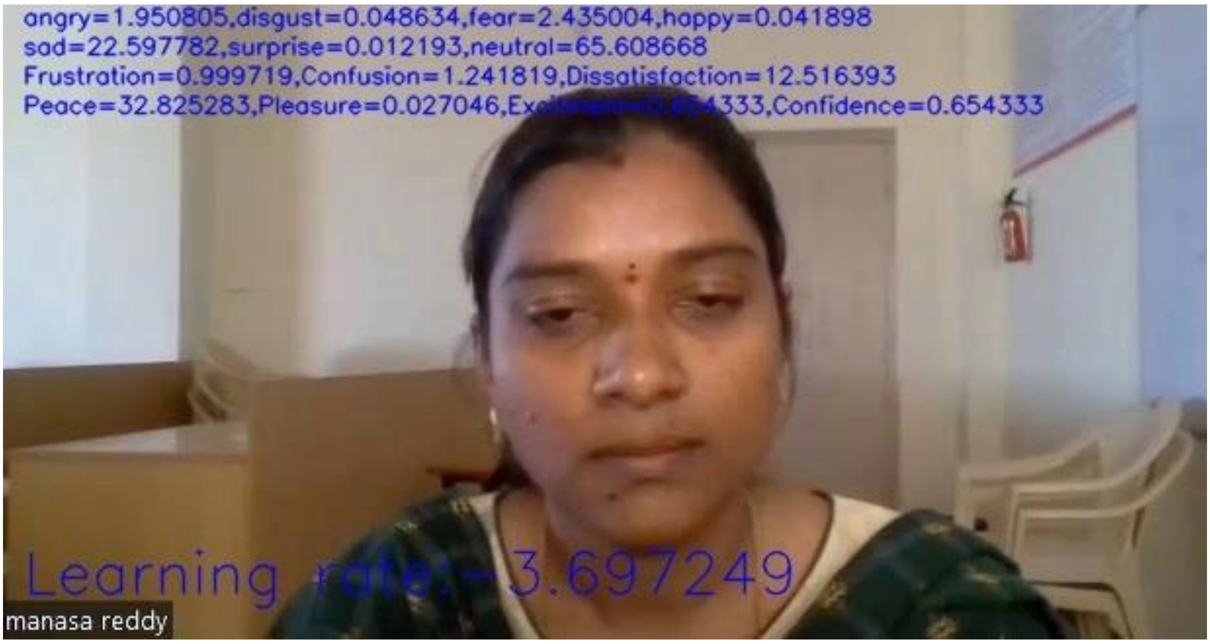
Figure 5.6 presents an illustration of the training failure in comparison to the validation failure. It seems that as the time period progresses, there is a rise in the validation loss while there is an observable reduction in the training loss. This is demonstrated by the fact that. When the weights are altered, it is always anticipated that the validation data would decrease, and this prediction is always correct. As the epoch evolves in a higher order, as shown in the latter stages of the figure, one may assume that the rate of validation loss will be smaller than the rate of training loss. This is the case in this particular scenario. As a consequence of this, a model that is appropriately adapted to the training data may be obtained.

The loss function is what's used to calculate how much of a gap there is between the predicted outputs of a neural network and the actual objectives it's trying to reach when it's being trained. The reduction of the value of the loss function should be the end result of the training procedure. This may be accomplished by modifying the network parameters, which includes the weights and the biases. In the context of the model that has been suggested, the term "training loss" refers to the average loss that the model sustains on the training set while it is through the training process. This indicates that the model weights are adjusted in order to minimize the training loss, which is determined for each batch of data that the neural network processes.

However, this does not guarantee good generalization of the model to data that is unknown to it; just reducing the training loss is not enough. This problem occurs as a result of the model perhaps being overfit to the training data, which then leads to an excessive amount of specialization toward certain cases within the training set. As a direct consequence of this, the performance of the model suffers when it is applied to new data. In order to overcome this issue, a separate dataset known as the validation set is deployed throughout the training phase in order to evaluate the performance of the model. The term "validation loss" refers to the mean loss that the model experienced while being tested on the validation set. This value is also calculated for each new batch of data. This statistic serves as an indication of the model's capacity to generalize to data that was not included during the training phase. This data was not included in the training procedure.

During the training phase, it is crucial to monitor both training loss and validation loss in order to guarantee that the model does not experience overfitting. It is evidence of the model's expression of overfitting behavior and its weak potential to generalize effectively to fresh data if a reduction in training loss corresponds with a rise in validation loss. In the present situation, one of the potential solutions to the problem of overfitting is to either call an end to the training phase or make use of one of the other regularization methods.

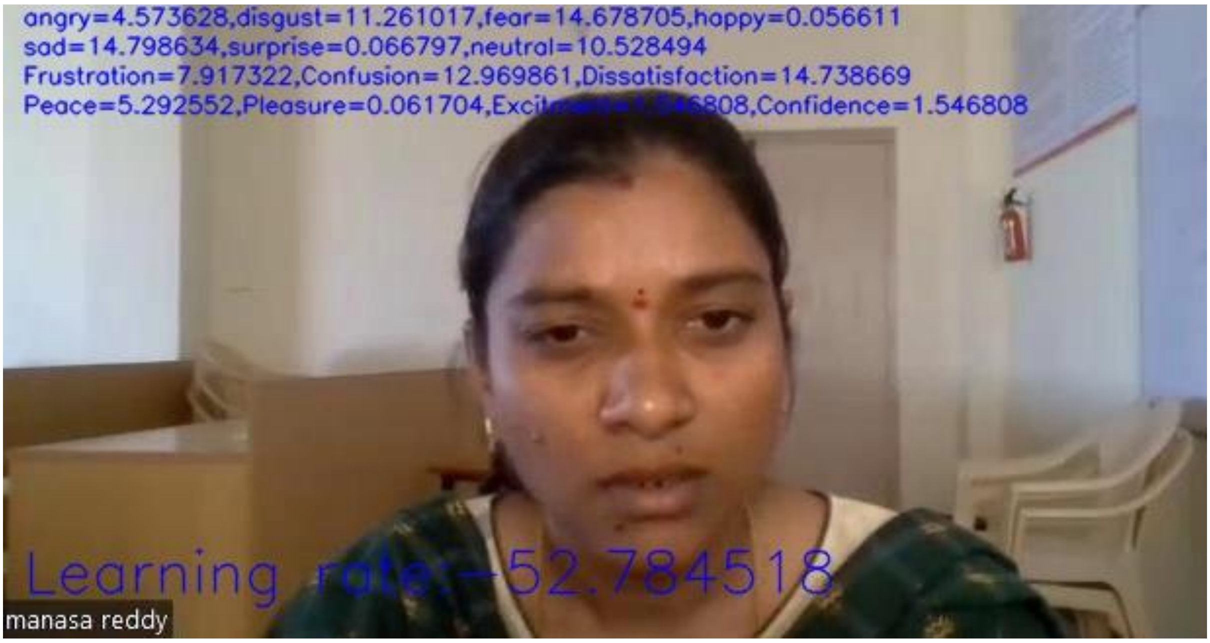
Figures 5.7 to 5.13 show the results of the proposed model for emotion estimate and learning rate.



```
angry=1.950805,disgust=0.048634,fear=2.435004,happy=0.041898
sad=22.597782,surprise=0.012193,neutral=65.608668
Frustration=0.999719,Confusion=1.241819,Dissatisfaction=12.516393
Peace=32.825283,Pleasure=0.027046,Excitement=0.654333,Confidence=0.654333
Learning rate:-3.697249
```

Figure 5. 7: Result 1

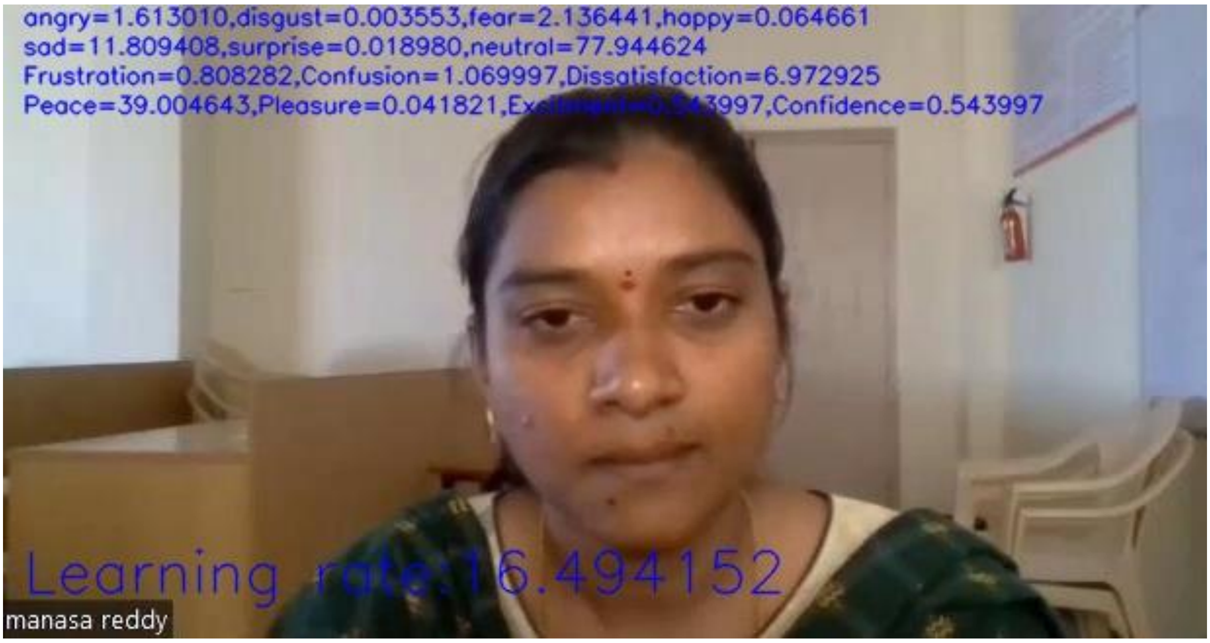
Figure 5.7 shows Result 1: Angry = 1.95, disgust = 0.04, fear = 2.43, happy = 0.04, sad = 22.59, surprise = 0.01, neutral = 65.60, frustration = 0.99, confusion = 1.24, dissatisfaction = 12.51, peace = 32.82, pleasure = 0.02, excitement = 0.65, confidence = 0.65 and learning rate = 3.69.



```
angry=4.573628,disgust=11.261017,fear=14.678705,happy=0.056611
sad=14.798634,surprise=0.066797,neutral=10.528494
Frustration=7.917322,Confusion=12.969861,Dissatisfaction=14.738669
Peace=5.292552,Pleasure=0.061704,Excitement=1.546808,Confidence=1.546808
Learning rate:-52.784518
```

Figure 5. 8: Result 2

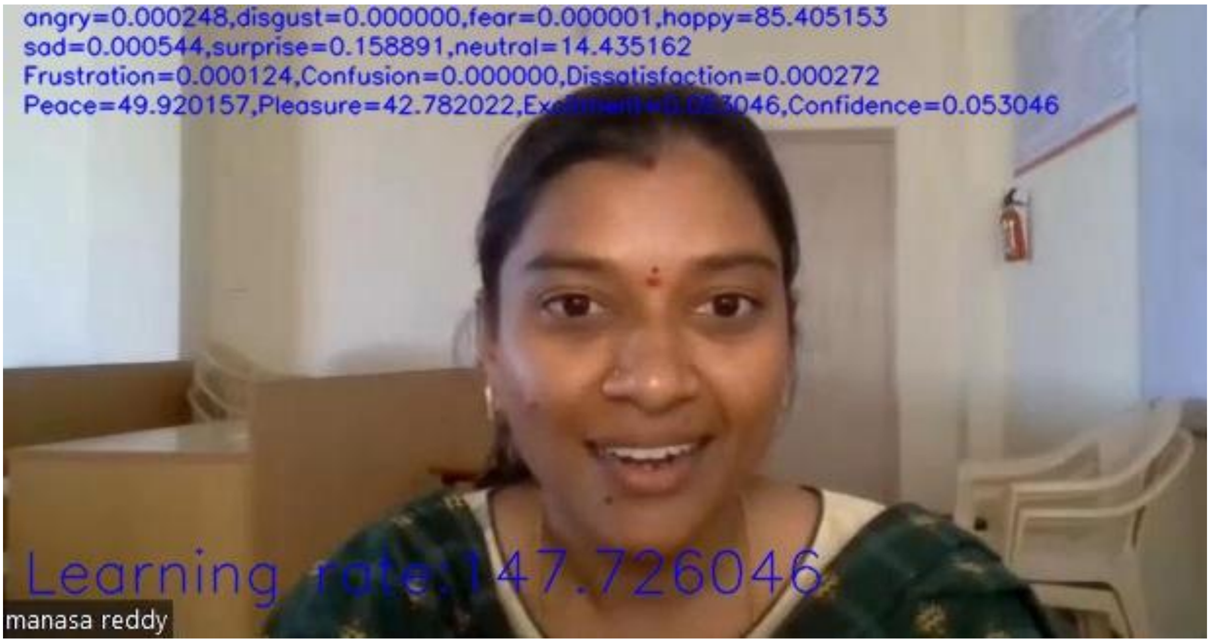
Figure 5.8 shows Result 2 - Angry = 4.57, disgust = 11.26, fear = 14.67, happy = 0.05, sad = 14.79, surprise = 0.06, neutral = 10.52, frustration = 7.91, confusion = 12.96, dissatisfaction = 14.73, peace = 5.29, pleasure = 0.06, excitement = 1.54, confidence = 1.54 and learning rate = 52.78.



```
angry=1.613010,disgust=0.003553,fear=2.136441,happy=0.064661
sad=11.809408,surprise=0.018980,neutral=77.944624
Frustration=0.808282,Confusion=1.069997,Dissatisfaction=6.972925
Peace=39.004643,Pleasure=0.041821,Excitement=0.543997,Confidence=0.543997
Learning rate:16.494152
```

Figure 5. 9: Result 3

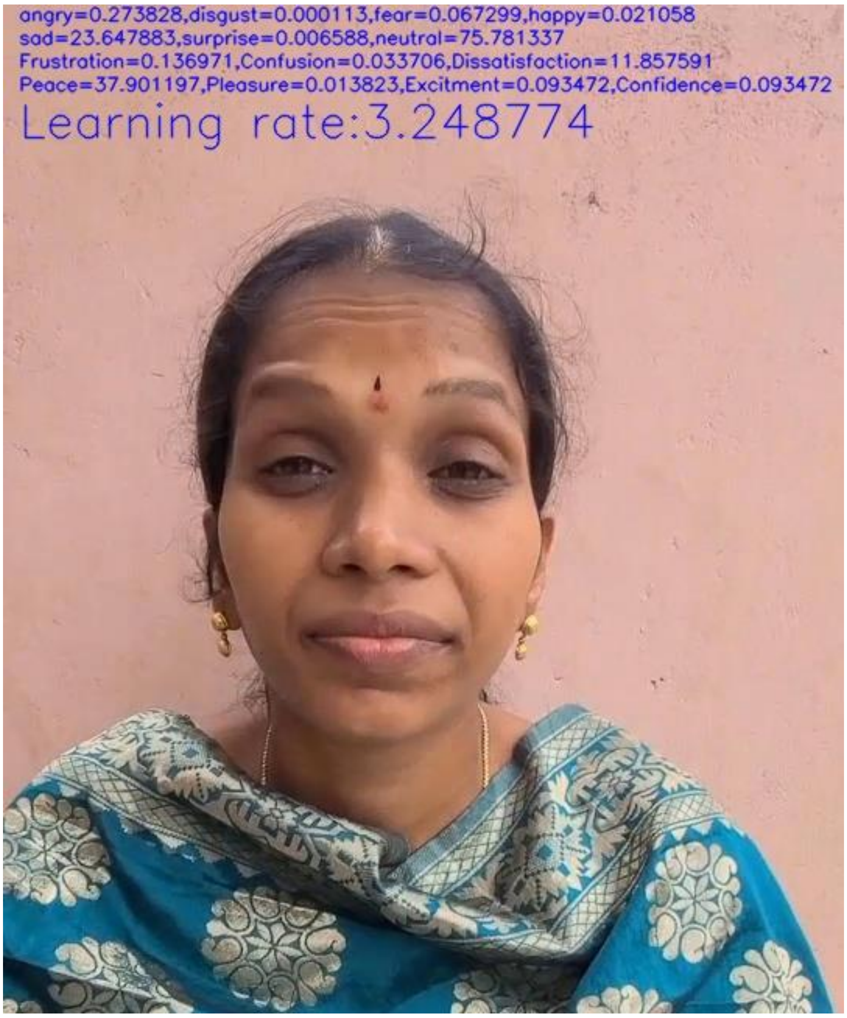
Figure 5.9 shows Result 3 - Angry = 1.61, disgust = 0.003, fear = 2.13, happy = 0.06, sad = 11.80, surprise = 0.01, neutral = 77.94, frustration = 0.80, confusion = 1.06, dissatisfaction = 6.97, peace = 39.004, pleasure = 0.04, excitement = 0.54, confidence = 0.54 and learning rate = 16.49.



```
angry=0.000248,disgust=0.000000,fear=0.000001,happy=85.405153
sad=0.000544,surprise=0.158891,neutral=14.435162
Frustration=0.000124,Confusion=0.000000,Dissatisfaction=0.000272
Peace=49.920157,Pleasure=42.782022,Excitement=0.053046,Confidence=0.053046
Learning rate:147.726046
```

Figure 5. 10: Result 4

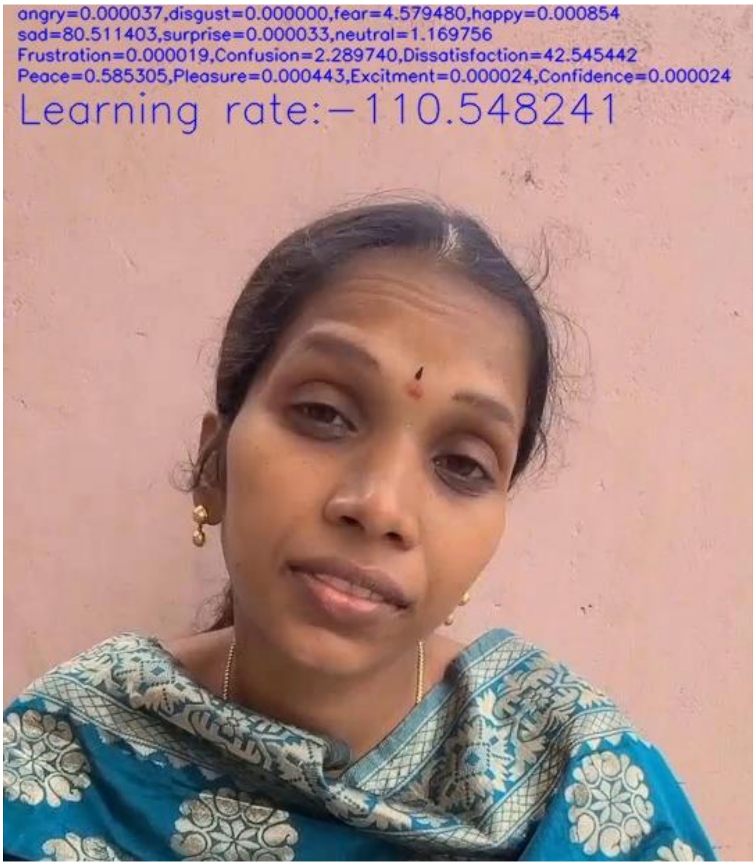
Figure 5.10 shows Result 4 - Angry = 0.0002, disgust = 0.00, fear = 0.000001, happy = 85.40, sad = 0.0005, surprise = 0.15, neutral = 14.43, frustration = 0.0001, confusion = 0.00, dissatisfaction = 0.0002, peace = 49.92, pleasure = 42.78, excitement = 0.05, confidence = 0.05 and learning rate = 147.72.



```
angry=0.273828,disgust=0.000113,fear=0.067299,happy=0.021058
sad=23.647883,surprise=0.006588,neutral=75.781337
Frustration=0.136971,Confusion=0.033706,Dissatisfaction=11.857591
Peace=37.901197,Pleasure=0.013823,Excitment=0.093472,Confidence=0.093472
Learning rate:3.248774
```

Figure 5. 11: Result 5

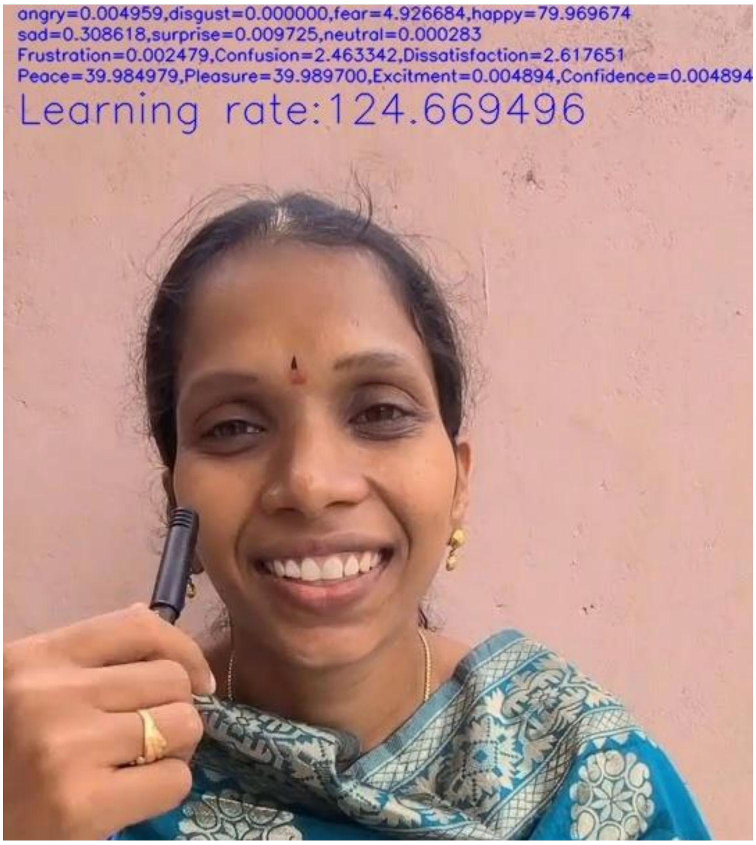
Figure 5.11 shows Result 5 - Angry = 0.27, disgust = 0.0001, fear = 0.06, happy = 0.02, sad = 23.64, surprise = 0.006, neutral = 75.78, frustration = 0.13, confusion = 0.03, dissatisfaction = 11.85, peace = 37.90, pleasure = 0.01, excitement = 0.09, confidence = 0.09 and learning rate = 3.24.



```
angry=0.000037,disgust=0.000000,fear=4.579480,happy=0.000854
sad=80.511403,surprise=0.000033,neutral=1.169756
Frustration=0.000019,Confusion=2.289740,Dissatisfaction=42.545442
Peace=0.585305,Pleasure=0.000443,Excitment=0.000024,Confidence=0.000024
Learning rate:-110.548241
```

Figure 5. 12: Result 6

Figure 5.12 shows Result 6 - Angry = 0.00003, disgust = 0.00, fear = 4.57, happy = 0.0008, sad = 80.51, surprise = 0.00003, neutral = 1.16, frustration = 0.00001, confusion = 2.28, dissatisfaction = 42.54, peace = 0.58, pleasure = 0.0004, excitement = 0.00002, confidence = 0.00002 and learning rate = 110.54.



```
angry=0.004959,disgust=0.000000,fear=4.926684,happy=79.969674
sad=0.308618,surprise=0.009725,neutral=0.000283
Frustration=0.002479,Confusion=2.463342,Dissatisfaction=2.617651
Peace=39.984979,Pleasure=39.989700,Excitment=0.004894,Confidence=0.004894
Learning rate:124.669496
```

Figure 5. 13: Result 7

Figure 5.13 shows Result 7 - Angry = 0.004, disgust = 0.00, fear = 4.92, happy = 79.96, sad = 0.30, surprise = 0.009, neutral = 0.0002, frustration = 0.002, confusion = 2.46, dissatisfaction = 2.61, peace = 39.98, pleasure = 39.98, excitement = 0.004, confidence = 0.004and learning rate = 124.66

The proposed model is compared with other well-known models and the corresponding results are reported in the Table 5.2.

Table 5. 2: Accuracy Comparison Results

Model	Accuracy
NASNetMobile [21]	94.4
VGG19 [22]	95.7
MobilenetV2 [23]	96.6
InceptionV3 [24]	96.9
Proposed VGG-CNN model	97.3

From Table 5.2, proposed model achieved 97.3% accuracy compared to other models. The NASNetMobile achieved lowest accuracy which is 59.4%. The VGG19 model obtained 73.7% accuracy and InceptionV3 model reaches up to 92%. The Mobilenet and MobilenetV2 got same accuracy results which is 96.6%. InceptionV3 obtained an accuracy of 96.9%The model can capture more intricate patterns and features in the input data. This can enable the model to better discriminate between different classes of objects in the images it is trying to recognize, leading to higher accuracy.Additionally, extra layers can also provide more regularization to the model, which can help to prevent overfitting. This is due to the fact that the more layers provide more parameters to the model, which may make it more difficult for the model to accurately match the training data.

5.4 Conclusion

Emotion modelling using VGG16-CNN transfer learning offers an efficient and effective approach to recognize and interpret human emotions from visual data. Leveraging the power of pre-trained models like VGG16 allows us to benefit from their generalization capabilities and speed up the training process. As emotion recognition continues to gain importance in diverse applications, this fusion of emotion modelling and transfer learning stands poised to open up new possibilities for understanding and responding to human emotions in numerous domains. This chapter presented a learning rate assessment model using the micro facial expressions such as angry, frustration, disgust, confusion, fear, dissatisfaction, sad, neutral, peace, happy, pleasure, surprise, excitement and confidence. Emotion probability and learning coefficient derived form the emotions are used to calculate the overall learning rate. The utilization of a pretrained VGG model, empowers the extraction of profoundly intricate features surpassing prior architectures. The transfer learning approach mitigates the challenge of vanishing gradients, enhancing model stability. This sophisticated architecture facilitates the acquisition of minute variations from input data, which is an important feature for accurately identifying subtle emotions and thereby precisely assessing students' learning rates with an accuracy of 97.3%.

Conclusion

This research is focused on the utilization of deep learning techniques for analyzing student emotion, expression, and learning rate in the context of online education. Through the exploration of facial expression recognition, sentiment analysis, and learning rate modelling, valuable insights were obtained regarding students' emotional states and their impact on the learning process.

The findings of this research demonstrate the potential of deep learning models in accurately detecting and interpreting students' emotions and expressions during online education. By leveraging advanced algorithms such as convolutional neural networks (CNNs) and attention mechanisms, the proposed models exhibited promising performance in capturing subtle facial cues and extracting emotional features. This contributes to a more comprehensive understanding of students' emotional responses, which can inform instructional strategies, personalized interventions, and adaptive learning systems. The analysis of learning rate dynamics in online education provided valuable insights into students' engagement and performance levels. By incorporating learning rate modelling techniques into the framework, the thesis shed light on the factors that influence students' learning progress, such as attention levels, task difficulty, and individual differences. This information can be leveraged to optimize the learning experience, tailor content delivery, and provide timely support to enhance student outcomes.

The ability to detect students' emotions in a smart classroom setting is a crucial aspect of classroom observation and assessment, and it has garnered significant attention in recent research on emotional computing. This research introduced a methodology based on deep learning techniques for the recognition of learners' emotions. By employing the SAN CNN algorithm, it becomes possible to identify learners' emotions in real-time, thereby fostering a more relaxed and effective learning experience. The experimental results obtained demonstrate that the parameters of the SAN CNN model accurately classify datasets with an impressive accuracy rate of 98.42%.

A novel attention mechanism has been introduced that focuses on enhancing the characteristics of facial regions along with various other inputs. The proposed methodology utilizes data from the camera's visual feed, audio input, and heuristic properties to assess the students' levels of attentiveness. As part of this process, the teacher periodically checks the content displayed on the students' screens, while a facial examination of the students is conducted using the video feed. Initially, the image is passed to the OpenCV face recognition module to detect and identify the student's face within it. This involves sending the image to the module for analysis. Once a face is recognized, its angle is analyzed through a collaboration between Google's facial landmark identification module and MobileNetV2. The face angle is calculated based on the extracted features. Additionally, apart from examining the facial characteristics of the student, the computer's heuristic inputs and the background noise in the auditory input are also monitored. The proposed model achieved an impressive accuracy of 97.47% in identifying the student's gaze angle, while the suggested CNN model exhibited a perfect accuracy of 100% in classifying the audio samples.

Facial emotion recognition of students is an advanced technology that employs computer vision and machine learning algorithms to analyze and interpret the emotions displayed by students in real-time. It finds utility in various educational contexts, encompassing traditional classrooms, e-learning sessions, and virtual reality environments. The task of deciphering micro facial expressions or other modalities can be intricate and necessitates specialized expertise and training. The ResNet model, with its ability to learn and extract intricate information from input images, tackles this challenge by discerning even the most subtle facial expressions.

ResNet's deep residual learning and scalability enable it to acquire more sophisticated representations of the input data, enabling the recognition of nuanced variations in facial expressions that might go unnoticed by other architectural approaches. To facilitate the identification of minute facial expressions, the ResNet model is trained using a dataset comprising diverse images containing instances of the specific facial expressions of interest. Throughout the training process, the model learns to establish connections between the distinctive features present in the input images and their corresponding facial expression labels. Subsequently, when presented with new input images, the model utilizes these learned associations to accurately predict the corresponding facial expression label.

The proposed model demonstrates remarkable performance, achieving an impressive accuracy rate of 99.3%. This high level of accuracy underscores the efficacy of the ResNet-based approach in facial emotion recognition tasks. By successfully capturing and interpreting the emotions conveyed through facial expressions, this technology holds immense potential for enhancing our understanding of students' emotional states in educational settings. Consequently, it opens doors to a range of applications, including personalized learning experiences, early intervention and support for students, and improved engagement and motivation.

Future Scope

The scope for future research lies in the integration of multiple modalities for analysis. Currently, the research primarily focuses on analyzing facial expressions to infer emotions. However, incorporating other modalities such as voice, gestures, chat history and body language can provide a more holistic understanding of students' emotional states and cognitive engagement. By combining deep learning techniques with multimodal analysis, researchers can gain richer insights into students' emotional experiences and enhance the accuracy of emotion recognition models.

Furthermore, there is potential for developing real-time feedback systems based on emotion and expression analysis. By continuously monitoring students' emotions and expressions during online learning, intelligent systems can provide timely feedback and adaptive interventions. This feedback can be personalized to address individual learning needs, enhance motivation, and improve overall learning outcomes. Integrating such systems into online educational platforms can create more engaging and supportive learning environments.

Additionally, future research can explore the correlation between student emotions, expressions, and learning outcomes. By analyzing the relationship between emotional states, facial expressions, and learning rates, researchers can uncover valuable insights into the impact of emotions on students' cognitive processes and academic performance. Understanding these connections can inform the development of effective pedagogical strategies and interventions that promote optimal learning experiences in online education.

References

[1] Brenton, Sam. "E-learning–an introduction." In A handbook for teaching and learning in higher education, pp. 103-116. Routledge, 2008.

[2] Khan, Badrul H. "Introduction to e-learning." In International Handbook of E-Learning Volume 1, pp. 29-68. Routledge, 2015.

[3] Nicholson, Paul. "A history of e-learning: Echoes of the pioneers." Computers and education: E-learning, from theory to practice (2007): 1-11.

[4] El Mhouti, Abderrahim, Azeddine Nasseh, Mohamed Erradi, and José Marfa Vasquèz. "Enhancing collaborative learning in Web 2.0-based e-learning systems: A design framework for building collaborative e-learning contents." Education and Information Technologies 22 (2017): 2351-2364.

[5] Favale, Thomas, Francesca Soro, Martino Trevisan, Idilio Drago, and Marco Mellia. "Campus traffic and e-Learning during COVID-19 pandemic." Computer networks 176 (2020): 107290.

[6] Lara, Juan A., Shadi Aljawarneh, and Sonia Pamplona. "Special issue on the current trends in E-learning Assessment." Journal of Computing in Higher Education 32 (2020): 1-8.

[7] Araka, Eric, Elizaphan Maina, Rhoda Gitonga, and Robert Oboko. "Research trends in measurement and intervention tools for self-regulated learning for e-learning environments—systematic review (2008–2018)." Research and Practice in Technology Enhanced Learning 15 (2020): 1-21.

[8] Torres Martín, César, Christian Acal, Mohammed El Homrani, and Ángel Custodio Mingorance Estrada. "Impact on the virtual learning environment due to COVID-19." Sustainability 13, no. 2 (2021): 582.

[9] Mishra, Sanjaya, Sidhartha Sahoo, and Shriram Pandey. "Research trends in online distance learning during the COVID-19 pandemic." Distance Education 42, no. 4 (2021): 494-519.

[10] Essary, Michael L. "Key external factors influencing successful distance education programs." Academy of Educational Leadership Journal 18, no. 3 (2014): 121.

[11] Amarneh, Batool M., Muhammad Turki Alshurideh, Barween H. Al Kurdi, and Zaid Obeidat. "The Impact of COVID-19 on E-learning: Advantages and Challenges." In Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2021), pp. 75-89. Cham: Springer International Publishing, 2021.

[12] Stecuła, Kinga, and Radosław Wolniak. "Advantages and Disadvantages of E-Learning Innovations during COVID-19 Pandemic in Higher Education in Poland." Journal of Open Innovation: Technology, Market, and Complexity 8, no. 3 (2022): 159.

[13] Maatuk, Abdelsalam M., Ebitisam K. Elberkawi, Shadi Aljawarneh, Hasan Rashaideh, and Hadeel Alharbi. "The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors." Journal of computing in higher education 34, no. 1 (2022): 21-38.

[14] Ferri, Fernando, Patrizia Grifoni, and Tiziana Guzzo. "Online learning and emergency remote teaching: Opportunities and challenges in emergency situations." Societies 10, no. 4 (2020): 86.

[15] Aguilera-Hermida, A. Patricia. "College students' use and acceptance of emergency online learning due to COVID-19." International journal of educational research open 1 (2020): 100011.

[16] Al-Kumaim, Nabil Hasan, Abdulsalam K. Alhazmi, Fathey Mohammed, Nadhmi A. Gazem, Muhammad Salman Shabbir, and Yousef Fazea. "Exploring the impact of the COVID-19 pandemic on university students' learning life: An integrated conceptual motivational model for sustainable and healthy online learning." Sustainability 13, no. 5 (2021): 2546.

[17] Zacharis, Georgios, and Kleopatra Nikolopoulou. "Factors predicting University students' behavioral intention to use eLearning platforms in the post-pandemic normal: an UTAUT2 approach with 'Learning Value'." Education and Information Technologies (2022): 1-18.

[18] Syed, Ali Murad, Shabir Ahmad, Adel Alaraifi, and Waleed Rafi. "Identification of operational risks impeding the implementation of eLearning in higher education system." Education and Information Technologies 26 (2021): 655-671.

[19] Wu, Chih-Hung, Yi-Lin Tzeng, and Yueh-Min Huang. "Measuring performance in leaning process of digital game-based learning and static E-learning." Educational Technology Research and Development 68 (2020): 2215-2237.

[20] Lee, Jeongju, Hae-Deok Song, and Ah Jeong Hong. "Exploring factors, and indicators for measuring students' sustainable engagement in e-learning." Sustainability 11, no. 4 (2019): 985.

[21] Singh, Madanjit, Sulaimon Oyeniyi Adebayo, Munish Saini, and Jaswinder Singh. "Indian government E-learning initiatives in response to COVID-19 crisis: A case study on online learning in Indian higher education system." *Education and Information Technologies* 26, no. 6 (2021): 7569-7607.

[22] Bossman, Ahmed, and Samuel Kwaku Agyei. "Technology and instructor dimensions, e-learning satisfaction, and academic performance of distance students in Ghana." *Heliyon* 8, no. 4 (2022): e09200.

[23] Pham, Long, Yam B. Limbu, Trung K. Bui, Hien T. Nguyen, and Huong T. Pham. "Does e-learning service quality influence e-learning student satisfaction and loyalty? Evidence from Vietnam." *International Journal of Educational Technology in Higher Education* 16, no. 1 (2019): 1-26.

[24] Alam, Ashraf. "Cloud-Based E-learning: Scaffolding the Environment for Adaptive E-learning Ecosystem Based on Cloud Computing Infrastructure." In *Computer Communication, Networking and IoT: Proceedings of 5th ICICC 2021, Volume 2,* pp. 1-9. Singapore: Springer Nature Singapore, 2022.

[25] Favale, Thomas, Francesca Soro, Martino Trevisan, Idilio Drago, and Marco Mellia. "Campus traffic and e-Learning during COVID-19 pandemic." *Computer networks* 176 (2020): 107290.

[26] Rivas, Alberto, Alfonso Gonzalez-Briones, Guillermo Hernandez, Javier Prieto, and Pablo Chamoso. "Artificial neural network analysis of the academic performance of students in virtual learning environments." *Neurocomputing* 423 (2021): 713-720.

[27] Imani, Maryam, and Gholam Ali Montazer. "A survey of emotion recognition methods with emphasis on E-Learning environments." *Journal of Network and Computer Applications* 147 (2019): 102423.

[28] Bdair, Izzeddin A. "Nursing students' and faculty members' perspectives about online learning during COVID-19 pandemic: A qualitative study." *Teaching and Learning in Nursing* 16, no. 3 (2021): 220-226.

[29] Valverde-Berrocoso, Jesús, María del Carmen Garrido-Arroyo, Carmen Burgos-Videla, and María Belén Morales-Cevallos. "Trends in educational research about e-learning: A systematic literature review (2009–2018)." *Sustainability* 12, no. 12 (2020): 5153.

[30] Tomaino, Melaura Andree Erickson, Alissa L. Greenberg, Sarah Ann Kagawa-Purohit, Sagui A. Doering, and Edward Steven Miguel. "An assessment of the feasibility and effectiveness of distance learning for students with severe developmental disabilities and high behavioral needs." *Behavior analysis in practice* 15, no. 1 (2022): 243-259.

[31] Annansingh, Fenio. "Mind the gap: Cognitive active learning in virtual learning environment perception of instructors and students." *Education and Information Technologies* 24 (2019): 3669-3688.

[32] Coussement, Kristof, Minh Phan, Arno De Caigny, Dries F. Benoit, and Annelies Raes. "Predicting student dropout in subscription-based online learning environments: The beneficial impact of the logit leaf model." *Decision Support Systems* 135 (2020): 113325.

[33] Peng, Xian, and Qinmei Xu. "Investigating learners' behaviors and discourse content in MOOC course reviews." *Computers & Education* 143 (2020): 103673.

[34] Md Yunus, Melor, Wee Shin Ang, and Harwati Hashim. "Factors affecting teaching English as a Second Language (TESL) postgraduate

students' behavioural intention for online learning during the COVID-19 pandemic." *Sustainability* 13, no. 6 (2021): 3524.

[35] Rajabalee, Yousra Banoor, and Mohammad Issack Santally. "Learner satisfaction, engagement and performances in an online module: Implications for institutional e-learning policy." *Education and Information Technologies* 26, no. 3 (2021): 2623-2656.

[36] Asaju, Christine, and Hima Vadapalli. "A temporal approach to facial emotion expression recognition." In Artificial Intelligence Research: Second Southern African Conference, SACAIR 2021, Durban, South Africa, December 6–10, 2021, Proceedings, pp. 274-286. Cham: Springer International Publishing, 2022.

[37] Ninaus, Manuel, Simon Greipl, Kristian Kiili, Antero Lindstedt, Stefan Huber, Elise Klein, Hans-Otto Karnath, and Korbinian Moeller. "Increased emotional engagement in game-based learning–A machine learning approach on facial emotion detection data." *Computers & Education* 142 (2019): 103641.

[38] Gupta, Swadha, Parteek Kumar, and Raj Kumar Tekchandani. "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models." *Multimedia Tools and Applications* 82, no. 8 (2023): 11365-11394.

[39] Pise, Anil, Hima Vadapalli, and Ian Sanders. "Facial emotion recognition using temporal relational network: an application to E-learning." *Multimedia Tools and Applications* (2020): 1-21.

[40] Bhardwaj, Prakhar, P. K. Gupta, Harsh Panwar, Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, and Anubha Bhaik. "Application of deep learning on student engagement in e-learning environments." *Computers & Electrical Engineering* 93 (2021): 107277.

[41] Yu, Hao, Ankit Gupta, Will Lee, Ivon Arroyo, Margrit Betke, Danielle Allesio, Tom Murray, John Magee, and Beverly P. Woolf. "Measuring and integrating facial expressions and head pose as indicators of engagement and affect in tutoring systems." In *Adaptive Instructional Systems. Adaptation Strategies and Methods: Third International Conference, AIS 2021, Held as*

Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II, pp. 219-233. Cham: Springer International Publishing, 2021.

[42] Sharma, Prabin, Shubham Joshi, Subash Gautam, Sneha Maharjan, Salik Ram Khanal, Manuel Cabral Reis, João Barroso, and Vítor Manuel de Jesus Filipe. "Student engagement detection using emotion analysis, eye tracking and head movement with machine learning." In *Technology and Innovation in Learning, Teaching and Education: Third International Conference, TECH-EDU 2022, Lisbon, Portugal, August 31–September 2, 2022, Revised Selected Papers*, pp. 52-68. Cham: Springer Nature Switzerland, 2023.

[43] Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." *SN Applied Sciences* 2, no. 3 (2020): 446.

[44] Dubovi, Ilana. "Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology." *Computers & Education* 183 (2022): 104495.

[45] Komuro, Nobuyoshi, Tomoki Hashiguchi, Keita Hirai, and Makoto Ichikawa. "Predicting individual emotion from perception-based non-contact sensor big data." *Scientific reports* 11, no. 1 (2021): 1-9.

[46] White, Rhianna, Ginny Russell, Pamela Qualter, Matthew Owens, and Lamprini Psychogiou. "Do peer relationships mediate the association between children’s facial emotion recognition ability and their academic attainment? Findings from the ALSPAC study." *Contemporary Educational Psychology* 64 (2021): 101942.

[47] Luo, Zhenzhen, Chen Jingying, Wang Guangshuai, and Liao Mengyi. "A three-dimensional model of student interest during learning using multimodal fusion with natural sensing technology." *Interactive Learning Environments* 30, no. 6 (2022): 1117-1130.

[48] Buono, Paolo, Berardina De Carolis, Francesca D’Errico, Nicola Macchiarulo, and Giuseppe Palestra. "Assessing student engagement from facial behavior in on-line learning." *Multimedia Tools and Applications* 82, no. 9 (2023): 12859-12877.

[49] Vanneste, Pieter, José Oramas, Thomas Verelst, Tinne Tuytelaars, Annelies Raes, Fien Depaepe, and Wim Van den Noortgate. "Computer vision and human behaviour, emotion and cognition detection: A use case on student engagement." *Mathematics* 9, no. 3 (2021): 287.

[50] Tonguç, Güray, and Betul OzaydınOzkara. "Automatic recognition of student emotions from facial expressions during a lecture." *Computers & Education* 148 (2020): 103797.

[51] Karan, K. V., Vedant Bahel, R. Ranjana, and T. Subha. "Transfer learning approach for analyzing attentiveness of students in an online classroom environment with emotion detection." In *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2021*, pp. 253-261. Singapore: Springer Nature Singapore, 2022.

[52] Xu, Ruyi, Jingying Chen, Jiaxu Han, Lei Tan, and Luhui Xu. "Towards emotion-sensitive learning cognitive state analysis of big data in education: deep learning-based facial expression analysis using ordinal information." *Computing* 102 (2020): 765-780.

[53] Ahmad, Khurshid, Shirui Wang, Carl Vogel, Pranav Jain, Oscar O'Neill, and Basit Hamid Sufi. "Comparing the Performance of Facial Emotion Recognition Systems on Real-Life Videos: Gender, Ethnicity and Age." In *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 1*, pp. 193-210. Springer International Publishing, 2022.

[54] Pekrun, Reinhard, Herbert W. Marsh, Felix Suessenbach, Anne C. Frenzel, and Thomas Goetz. "School grades and students' emotions: Longitudinal models of within-person reciprocal effects." *Learning and Instruction* 83 (2023): 101626.

[55] Marceddu, Antonio Costantino, Luigi Pugliese, Jacopo Sini, Gustavo Ramirez Espinosa, Mohammadreza Amel Solouki, Pietro Chiavassa, Edoardo Giusto, Bartolomeo Montrucchio, Massimo Violante, and Francesco De Pace. "A novel redundant validation IoT system for affective learning based on facial expressions and biological signals." *Sensors* 22, no. 7 (2022): 2773.

[56] Ma, Tian, Kavuma Benon, Bamweyana Arnold, Keping Yu, Yan Yang, Qiaozhi Hua, Zheng Wen, and Anup Kumar Paul. "Bottleneck feature

extraction-based deep neural network model for facial emotion recognition." In *Mobile Networks and Management: 10th EAI International Conference, MONAMI 2020, Chiba, Japan, November 10–12, 2020, Proceedings 10*, pp. 30-46. Springer International Publishing, 2020.

[57] Hassouneh, Aya, A. M. Mutawa, and M. Murugappan. "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods." *Informatics in Medicine Unlocked* 20 (2020): 100372.

[58] Yan, Fei, Nan Wu, Abdullah M. Iliyasu, Kazuhiko Kawamoto, and Kaoru Hirota. "Framework for identifying and visualising emotional atmosphere in online learning environments in the COVID-19 Era." *Applied Intelligence* 52, no. 8 (2022): 9406-9422.

[59] Shen, Junge, Haopeng Yang, Jiawei Li, and Zhiyong Cheng. "Assessing learning engagement based on facial expression recognition in MOOC's scenario." *Multimedia Systems* (2022): 1-10.

[60] Chirra, Venkata Rami Reddy, Srinivasulu Reddy Uyyala, and Venkata Krishna Kishore Kolli. "Virtual facial expression recognition using deep CNN with ensemble learning." *Journal of Ambient Intelligence and Humanized Computing* (2021): 1-19.

[61] Mohamad Nezami, Omid, Mark Dras, Len Hamey, Deborah Richards, Stephen Wan, and Cécile Paris. "Automatic recognition of student engagement using deep learning and facial expression." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 273-289. Springer, Cham, 2020.

[62] Alqurashi, Emtinan. "Predicting student satisfaction and perceived learning within online learning environments." *Distance education* 40, no. 1 (2019): 133-148.

[63] Gherheș, Vasile, Claudia E. Stoian, Marcela Alina Fărcașiu, and Miroslav Stanici. "E-learning vs. face-to-face learning: Analyzing students' preferences and behaviors." *Sustainability* 13, no. 8 (2021): 4381.

[64] Jia, Chengyuan, Khe Foon Hew, Shurui Bai, and Weijiao Huang. "Adaptation of a conventional flipped course to an online flipped format

during the Covid-19 pandemic: Student learning performance and engagement." *Journal of research on technology in education* 54, no. 2 (2022): 281-301.

[65] Dewan, M., Mahbub Murshed, and Fuhua Lin. "Engagement detection in online learning: a review." *Smart Learning Environments* 6, no. 1 (2019): 1-20.

[66] Szopiński, Tomasz, and Katarzyna Bachnik. "Student evaluation of online learning during the COVID-19 pandemic." *Technological Forecasting and Social Change* 174 (2022): 121203.

[67] Kim, Jinyoung. "Learning and teaching online during Covid-19: Experiences of student teachers in an early childhood education practicum." *International Journal of Early Childhood* 52, no. 2 (2020): 145-158.

[68] Li, Jing, Kan Jin, Dalin Zhou, Naoyuki Kubota, and Zhaojie Ju. "Attention mechanism-based CNN for facial expression recognition." *Neurocomputing* 411 (2020): 340-350.

[69] Taub, Michelle, Robert Sawyer, Andy Smith, Jonathan Rowe, Roger Azevedo, and James Lester. "The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment." *Computers & Education* 147 (2020): 103781.

[70] Kola, Durga Ganga Rao, and Srinivas Kumar Samayamantula. "A novel approach for facial expression recognition using local binary pattern with adaptive window." *Multimedia Tools and Applications* 80 (2021): 2243-2262.

[71] Gupta, Sujit Kumar, T. S. Ashwin, and Ram Mohana Reddy Guddeti. "Students' affective content analysis in smart classroom environment using deep learning techniques." *Multimedia Tools and Applications* 78 (2019): 25321-25348.

[72] Alam, Ashraf. "Employing Adaptive Learning and Intelligent Tutoring Robots for Virtual Classrooms and Smart Campuses: Reforming Education in the Age of Artificial Intelligence." In *Advanced Computing and Intelligent*

Technologies: Proceedings of ICACIT 2022, pp. 395-406. Singapore: Springer Nature Singapore, 2022.

[73] TS, Ashwin, and Ram Mohana Reddy Guddeti. "Automatic detection of students’ affective states in classroom environment using hybrid convolutional neural networks." *Education and information technologies* 25, no. 2 (2020): 1387-1415.

[74] Emerson, Andrew, Elizabeth B. Cloude, Roger Azevedo, and James Lester. "Multimodal learning analytics for game-based learning." *British Journal of Educational Technology* 51, no. 5 (2020): 1505-1526.

[75] Megahed, Mohammed, and Ammar Mohammed. "Modeling adaptive E-learning environment using facial expressions and fuzzy logic." *Expert Systems with Applications* 157 (2020): 113460.

[76] Raes, Annelies, Pieter Vanneste, Marieke Pieters, Ine Windey, Wim Van Den Noortgate, and Fien Depaepe. "Learning and instruction in the hybrid virtual classroom: An investigation of students’ engagement and the effect of quizzes." *Computers & Education* 143 (2020): 103682.

[77] Nandi, Arijit, Fatos Xhafa, Laia Subirats, and Santi Fort. "Real-time emotion classification using eeg data stream in e-learning contexts." *Sensors* 21, no. 5 (2021): 1589.

[78] Butnaru, Gina Ionela, Valentin Niță, Alexandru Anichiti, and Geanina Brînză. "The effectiveness of online education during covid 19 pandemic—a comparative analysis between the perceptions of academic students and high school students from romania." *Sustainability* 13, no. 9 (2021): 5311.

[79] Lin, Li-Chun, I-Chun Hung, and Nian-Shing Chen. "The impact of student engagement on learning outcomes in a cyber-flipped course." *Educational Technology Research and Development* 67 (2019): 1573-1591.

[80] Hussein, Elham, Sumaya Daoud, Hussam Alrabaiah, and Rawand Badawi. "Exploring undergraduate students’ attitudes towards emergency online learning during COVID-19: A case from the UAE." *Children and youth services review* 119 (2020): 105699.

[81] Muflih, Suhaib, Sawsan Abuhammad, Sayer Al-Azzam, Karem H. Alzoubi, Mohammad Muflih, and Reema Karasneh. "Online learning for undergraduate health professional education during COVID-19: Jordanian medical students' attitudes and perceptions." *Heliyon* 7, no. 9 (2021): e08031.

[82] Khan, Mohammed Arshad, Tuba Kamal, AsherefIlliyan, and Mohd Asif. "School students' perception and challenges towards online classes during COVID-19 pandemic in India: An econometric analysis." *Sustainability* 13, no. 9 (2021): 4786.

[83] Ma, Kang, Muhammad Chutiyami, Yijin Zhang, and Sandy Nicoll. "Online teaching self-efficacy during COVID-19: Changes, its associated factors and moderators." *Education and information technologies* 26, no. 6 (2021): 6675-6697.

[84] Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. "Deep-emotion: Facial expression recognition using attentional convolutional network." *Sensors* 21, no. 9 (2021): 3046.

[85] Imani, Maryam, and Gholam Ali Montazer. "A survey of emotion recognition methods with emphasis on E-Learning environments." *Journal of Network and Computer Applications* 147 (2019): 102423.

[86] X. Lu, "Deep Learning Based Emotion Recognition and Visualization of Figural Representation," Frontiers in Psychology, p. 6237.

[87] Govindaswamy, Arun Gopal, Enid Montague, Daniela Stan Raicu, and Jacob Furst. "CNN as a feature extractor in gaze recognition." In 2020 3rd Artificial Intelligence and Cloud Computing Conference, pp. 31-37. 2020.

[88] El Hafi, Lotfi, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara. "Gaze tracking and object recognition from eye images." In 2017 First IEEE International Conference on Robotic Computing (IRC), pp. 310-315. IEEE, 2017.

[89] Tripathi, Garima, Deval Srivastava, Priyank Shah, and Saim Shaikh. "Driver activity monitoring using MobileNets." In Intelligent Computing and Networking, pp. 169-175. Springer, Singapore, 2021.

[90] Jarraya, Salma Kammoun, Marwa Masmoudi, and Mohamed Hammami. "A comparative study of Autistic Children Emotion recognition based on Spatio-Temporal and Deep analysis of facial expressions features

during a Meltdown Crisis." Multimedia Tools and Applications 80 (2021): 83-125.

[91] Cheng, Shuo, and Guohui Zhou. "Facial expression recognition method based on improved VGG convolutional neural network." International Journal of Pattern Recognition and Artificial Intelligence 34, no. 07 (2020): 2056003.

[92] Hu, Lida, and Qi Ge. "Automatic facial expression recognition based on MobileNetV2 in Real-time." In Journal of Physics: Conference Series, vol. 1549, no. 2, p. 022136. IOP Publishing, 2020.

[93] Mane, Shraddha, and Gauri Shah. "Facial recognition, expression recognition, and gender identification." In Data Management, Analytics and Innovation: Proceedings of ICDMAI 2018, Volume 1, pp. 275-290. Springer Singapore, 2019.