Educational Journal
Volume-V, Issue-I, May, 2026
DOI: https://doi.org/10.3126/ej.v5i1.95837
Research Management Cell,
Tribhuvan University, Sanothimi Campus, Sanothimi, Bhaktapur

Tracking Students' Unseen Status in the Virtual Classroom: A Systematic Review

Suresh Bahadur Diyal
M.Phil. in ICT Scholar,Faculty of Science, Health & Technology
Nepal Open University, Manbhawan, Lalitpur, Nepal
suresh.diyal@sac.tu.edu.np
https://orcid.org/0009-0005-8329-4829

Dr. Bhoj Raj Ghimire
Assistant Professor, Faculty of Science, Health and Technology
Nepal Open University, Manbhawan, Lalitpur, Nepal
bghimire@nou.edu.np
https://orcid.org/0000-0001-9391-6752

Abstract

Online learning has broadened learning opportunities but also brought with it a number of challenges to authentic student learning. A significant issue is that of the "invisible learner", the one who is attending the class but is not easy to evaluate in regard to attention and participation. This review presents an overview of world literature to identify the phenomenon of 'unseen students', to discuss methods for tracking engagement in low-resource settings and to explore technological solutions for engagement detection. This study followed the PRISMA framework, where 150 studies were first identified, followed by 15 studies that were included in the final study. The results suggest that the more sophisticated engagement detection models (CNN and LSTM) that use visual information are highly accurate. Stability, on the other hand, is a continuing problem with the internet, power, and limited device access to today's invisible learners. The review also finds that a combined monitoring method is necessary, which combines deep learning, lightweight recognition technique approaches and multimodal analytics. However, ethical and pedagogical principles need to inform the implementation of such systems to ensure that monitoring benefits educational quality and inclusivity.

Keywords: *virtual classrooms, unseen learners, student engagement detection multimodal approach, low-resources education context*

Introduction

COVID-19 caused the global education systems to undergo a paradigm shift that necessitated the use of online teaching-learning practices to replace the traditional classroom learning. Across the world, universities and schools adopted virtual platforms to maintain the continuity of their academic activities with the assistance of learning management systems and video conferencing software (Dhawan, 2020; Greenhow, Staudt Willet, and Galvin, 2021). This was a swift and broad transition that occurred in Nepal. Educational institutes schools transitioned to Zoom, Microsoft Teams, Google Meet, and Moodle to participate in formal courses, formative and summative assessments. This institutionalized online learning as an alternative and in certain instances as the only possible alternative to face-to-face learning because of this sudden change to digital migration (Pangeni, 2016; Rijal, 2022).

Although online learning brought in a new possibility of access to a larger number of learners, it also caused special problems related to the compliance with the engagement and participation of learners. It has always been found that the success of online learning is not only based on the delivery of the content but also on the confirmation of the presence of the students, their attention, and communication (Li, Li, and Chen, 2021). In Nepal, the internet connectivity was not stable, there were frequent power failures, the ICT infrastructure was not developed, and digital literacy among the teachers and students was not elevated, which obstructed the prospects of online education (Ghimire et al., 2022; Rana and Rana, 2020). Such a confusion between the physical attendance and intellectual interaction poses very important questions concerning the credibility of online education.

These problems are contained in international studies. In another study, published in Computers and Education by Li et al. (2021) it was revealed that deep learning models are effective to detect real-time student engagement during an online lesson. Likewise, in the article Smart Learning Environments Dewan, Murshed and Lin (2019). Additional studies also highlight the promise of multimodal interaction recognition, which incorporates both physiological cues and interaction history, as well as affective computing to have more information on student attention (Bosch, 2016; Whitehill et al., 2014).

Although these developments are being witnessed all over the world, Nepal offers different gaps. The lack of access to quality devices and internet services is one of the digital inequalities that have increased the gap between rural and urban students (Shrestha et al., 2022). Educators in isolated locations cite a lack of technical expertise, exposure to professional education, and the institutional support of perennial incorporation of online education (Gharti, 2023). although international research confirms the possibility of tracking invisible student behaviour with the application of artificial intelligence (AI) and computer vision, these solutions have not been developed and implemented in Nepal. This brings into serious consideration: *What is the status of the student behind the camera in real-time class? How can unseen student engagement be monitored in low-resource environments? What forms of technology can accurately be adapted to solve these problems?*

The problem of the so-called virtual classroom students remains especially acute due to the fact that engagement has been recognized as a multidimensional construct and includes behavioral, emotional, and cognitive aspects (Fredricks, Blumenfeld, and Paris, 2004). In physical classroom, a teacher can use such cues as eye contact, posture, and verbal involvement to deduce attentiveness (Christenson et al., 2012). On the contrary, there is a lack of visibility in online atmosphere, particularly when cameras are deliberately switched off by students. This generates what can be described as digital invisibility where, learners can be seen in a classroom, yet they are not in practice. This disengagement, which is hidden, is not only damaging the short-term learning results but also infringing on the legitimacy and efficacy of the virtual education systems (Nguyen, 2015).

Across the world, researchers have emphasized the importance of systematic frameworks in measuring and monitoring virtual learning environments engagement. Dewan et al. (2019) described the following taxonomy of engagement detection techniques: manual techniques (including self-reports and observational checklists); semi-automatic techniques (including engagement tracing); and automatic techniques (based on computer vision, sensor data, and learning analytics). Of these, the use of computer vision techniques that analyze the facial expression, gestures, and eye movements has been of specific interest because it is not intrusive and does not require any costs (Monkaresi et al., 2017). Likewise, multimodal techniques based on a combination of facial recognition with physiological measurements, behavioral information, and contextual evidence are more accurate and reliable when it comes to predicting student engagement (Bosch et al., 2015; Vail et al., 2016).

But, with technological potentiality, people should also acknowledge the ethical issues. The problem of privacy, consent, and data security are raised as the main concerns when it comes to tracking the facial features and behaviours of students in real-time (Whitehill et al., 2014). Therefore, any solution in the future should strike a balance between technology and moral responsibility since engagement detection should not pose any threat to the rights of students or increase suspicions of surveillance in education.

Although online education has been institutionalized in the global environment, as well as in Nepal, the actual engagement of the student is a problematic issue. This problem of students being invisible in virtual classes indicates the inadequacy in existing practices and the importance of creating effective systems to monitor student engagement. This synthesis essay is a review paper done on the world and local studies on engagement detection, challenges in tracking invisible students, and the ways to incorporate technological and pedagogical innovation. In the process, it aims to empower digital pedagogy, increase the accountability of education, and reinforce the credibility of virtual education in Nepal and similar developing situations.

Objectives

This Study aims to explore strategies for monitoring unseen students engagement in low-resource environment and examine students' real-time status in the virtual classroom.

Literature Review

Student Engagement in online Environment

Student engagement is also known to be a key determinant of learning outcome in both traditional and electronic environments. The engagement has been conceptualized as being behavioral, emotional and cognitive (Fredricks, Blundenfeld and Paris, 2004). In traditional classroom, the teacher uses direct indicators like body language, body position, and facial expression to measure attention (Christenson, Reschly, and Wylie, 2012). Nevertheless, these signals cannot be seen in the context of online learning, especially when the cameras are off or the microphones are muted, which results in the emergence of so-called digital invisibility (Nguyen, 2015). This invisibility leaves the question as to whether the students are being actively engaged in the learning process or are just sitting on the system without being involved.

Global studies have shown that technological solutions could be used to deal with these issues. Li, Li, and Chen (2021) employed the deep learning-based classification of real-time engagement in online classes and demonstrated that facial micro-expressions and patterns of behavior can be effective proxies of attention. The taxonomy of engagement detection methods offered by Dewan, Murshed, and Lin (2019) includes both manual (self-report, observational checklists) and computer vision sensor-based, and log-file analysis methods of engagement detection. Of these, machine learning and deep learning methods that are based on automatic means have become particularly popular because they are non-intrusive and scale well.

Machine Learning and Deep Learning Approaches for Engagement Detection

Deep learning models have improved the process of engagement detection by automating the analysis of the high-dimensional features such as facial features, eye movements, and gesture detection. Convolutional Neural Networks (CNNs), are used extensively in computer vision problems aiming to identify the type of emotions and interest levels in webcam derivations. Indicatively, hybrid models that merge CNN to transfer learning of previous architecture (e.g., VGG16, ResNet) have been reported to be over 90 percent engagement recognition accurate (Ayari et al., 2025). CNNs are strong in the sense of capturing space features (patterns in the facial image) and therefore are useful in a scenario that will allow the student to see part or even parts of the face.

Long Short-Term Memory (LSTM) networks, in their turn, are structured to recognize time-dependencies and are especially appropriate when studying engagement in time (Monkaresi et al., 2017). LSTMs can watch Likes and Ubekamism in attention to determine when a pattern leaves the usual trend through sequential frame processing of video or behavioral logs and differentiate temporary distractions and long-term disengagement. It has been demonstrated that CNNs and LSTMs outperform when combined because CNNs perceive spatial features whereas LSTMs perceive temporal dynamics of how learners behave (Bosch et al., 2015).

Classical neural network algorithms are also still applicable, especially when there is a limitation on resources, and the deep architecture is infeasible. The Local Binary Pattern Histograms (LBPH) and Fisherfaces are used as an illustration of algorithms that can be used in facial recognition and emotion recognition tasks with comparatively low processing requirements (Ojala et al., 2002; Belhumeur et al., 1997). The algorithms are grayscale image features that reveal discriminating features that allow one to identify simple forms of engagement like mouth motion, facial orientation, and eye openness. Although not as advanced as CNN-based models, they offer viable solutions to the situation kept within low-resource settings in which bandwith and matched hardware constraints limit the application of computationally expensive techniques.

Challenges in Developing and Low-Resource Contexts

Regardless of these technological innovations, there are still huge challenges in the developing world like Nepal. CNNs, LSTMs and neural network algorithms are effective when there is a reliable hardware, steady internet connection and constant supply of electricity. The rural schools in Nepal, however, do not always have such basic structures (Rana & Rana, 2020). Moreover, digital divide increases disparities in engagement monitoring. It is true that students in the urban areas use high-speed internet and personal gadgets, and those in rural or lowincome families may need to share one mobile phone with multiple learners (Shrestha, Koirala, and Poudel, 2022). Even, lightweight engagement detection solutions like LBPH or Fisherfaces have challenges in such settings because of poor access to webcams or workable devices.

The Challenge of "Unseen" Students

It is the issue of hidden students in Nepal that the popularization of such virtual platforms as Zoom and Google Meet during the COVID-19 demonstrated. It was common that teachers reported that students turned off cameras because of poor bandwidth or because they had no attetion and it was impossible to check whether the students were attentive (Ghimire, Adhikari, and Poudel, 2022). Although policy documents, including the ICT Master Plan of Nepal (20132017), envisaged greater integration of digital learning into practice, it has not been done in reality (MoE, 2016). The outcome is a disconnect between the viability of adoption of international technological solutions (e.g., CNN-LSTM models) in the Nepalese context and their availability.

Emerging Directions and Adaptable Solutions

The new body of knowledge has indicated that the means of catering to invisible students should be based on nimbleness in technology and practicality on context, as well. Lightweight face recognition methods based on LBPH and Fisherfaces can be applied in rural schools because of their low computing capabilities, and CNN-LSTM hybrid models might be applicable in urban schools with better facilities. On the same note, low-resource engagement of activity monitoring can be provided by use of log-based analytics that can track student interaction in chat, polls, and quizzes (Cocea and Weibelzahl, 2011). Hybrid methods, which involve the automated detecting methods as well as human teachers monitoring, could be more practical in Nepalese schools as compared to fully automated. Meanwhile, ethical issues including privacy and consent and surveillance risks should be considered. Whitehill et al. (2014) believe that engagement detection models must be built on delivering practical insights to educators, instead of surveilling the learners. The engagement detection needs to be incorporated into future systems together with the pedagogical enhancements, including gamified tasks, group activities, and personalized feedbacks, so that technology is going to enhance instruction, and not to substitute it.

Methodology

In this paper, the researcher used the Systematic Literature Review (SLR) approach guided by the PRISMA 2020 framework (Page et al., 2021) to ensure clearness and reproducibility for the paper selection process. The paper selection process is shown below in the figure. An extensive literature search was performed in the large academic databases, such as Scopus, Web of Science, IEEE Xplore, Springer and Google Scholar. To put the developing country view into perspective, the policy documents and reports on the same by UNESCO, UNICEF and Nepal Ministry of Education were also referred. The search targeted peer-reviewed articles released between 2019 to 2025 to obtain post-pandemic news on the topic of virtual education and engagement monitoring technologies. Relevant studies were retrieved using Boolean search strings based on the combination of the following keywords: student engagement detection, camera-off online learning, and unseen students, CNN, LSTM, deep learning, LBPH, Fisherfaces, and learning analytics. These keywords were narrowed down in combination with online learning and virtual classroom. The inclusion criteria consisted of the following: (1) the study must examine the topic of student engagement detection in an online or blended learning setting, (2) the study must employ artificial intelligence or machine learning methods Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Local Binary Pattern Histogram (LBPH), Fisherfaces, or multimodal learning analytics, (3) the study must be based on empirical evidence or systematic analysis, (4) the study must examine developing-country or low-resource educational settings, either explicitly or comparatively. The studies were eliminated when they concentrated on just the face-to-face classroom interaction, these studies had not

been empirically validated, were just opinion type articles and also when they were duplicates of the same records.

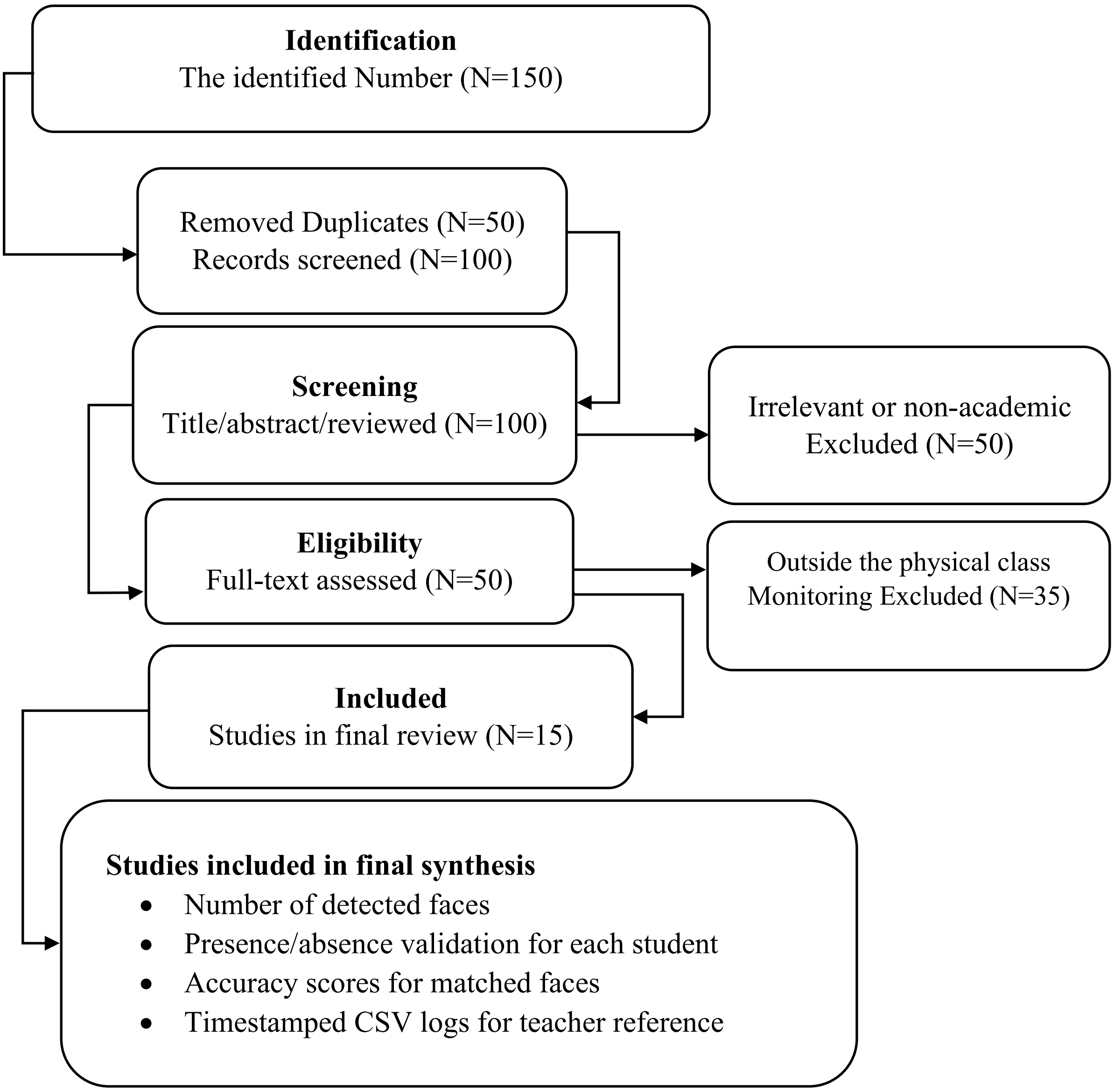

Figure 1: PRISMA flow diagram of the literature selection process.

The first search resulted in the retrieval of 150 records. The 50 duplicate records were removed followed by screening 100 left out records on the basis of titles and abstracts. Out of these, 50 were eliminated due to the irrelevancy or failure to meet the inclusion criteria. The articles were then reviewed in full to see whether they were eligible or not and 35 of these articles were discarded because they were not directly related to the research area of virtual classroom monitoring or engagement detection. Finally, 15 peer-reviewed articles were chosen and included in the final synthesis. To ensure transparency and minimize the selection bias, the selection process was followed by using the PRISMA flow diagram. The following analytical variables were among those included: type of algorithm (like: CNN, LSTM, LBPH, Fisherfaces), type of data to perform the analysis (visual, behavioral log- or multimodal), accuracy measures reported, computational requirements, infrastructure limitations and ethical implications. The above variables were systematically integrated into data extraction process.

Thematic synthesis approach was used in order to categorize the findings into five domains of analysis: (1) Efficacy of visual-based recognition of engagement; (2) viability of lightweight recognition techniques in low-resource setting; (3) structural constraints in camera-off contexts; (4) multimodal analytics; and (5) ethical and pedagogical consequences of real-time monitoring systems. The methodological process is consistent with the aim of the study and the evaluation of technological effectiveness and situational suitability is possible due to this systematic synthesis.

Results and Discussion

The results are presented in thematic areas, which are directly related to the proposed system workflow and the research objectives and questions.

Theme 1: The Real-Time Status of Students' real-time status behind the camera

Students who are not visible during virtual classes is one of the main research questions of this study. The results show that global engagement detection models are largely based on the premise of having continuous visual information. Facial and behavioral cues, when available, have been shown to be very accurate predictors of attentive, distracted, and disengaged learners through the use of deep learning techniques like Convolutional Neural Network (CNNs) and Long Short-Term Memory (LSTM) networks. The research of Li et al. (2021) shows that analyzing facial expressions at both spatial and temporal levels is capable of reliably inferring engagement states from expressions. Likewise, Monkaresi et al. (2017) show that temporal modeling decreases the misclassification due to short-term distractions.

However, while conducting the review, it is found there is a structural weakness; if the students turn off the camera, then the first level image capture of the system is not working. However, in these situations, an algorithmic model may not be able to tell the difference between technical and intentional disengagement. This means that the status of unseen students is not clear in the camera-off situation. Therefore, although AI models can detect engagement, they are only effective if there is a continuous stream of visual information an issue that might not be possible in low-resource settings, such as Nepal.

Theme 2: Technological Strategies for Monitoring Unseen Students

The first research goal that examined the methods of monitoring invisible engagement in low-resource settings finds that a hybrid algorithmic approach would be most feasible. CNN-based models have been successful in high-resource settings for extracting spatial features, including eye movement, head orientation, and facial micro-expressions, for engagement classification. The system, when coupled with LSTM networks that process sequences of frames over time, increases reliability by differentiating between a short distraction and a prolonged disengagement. This hybrid is meant to enhance the validity check and measurement of accuracy in the proposed workflow of the system.

In applications where bandwidth, processing power, and the quality of the available devices are not high, however, lightweight algorithms like Local Binary Pattern Histogram (LBPH) and Fisherfaces are more practical alternatives. Studies based on the texture-based recognition approaches (Ojala et al., 2002) show that LBPH is an approach that can be used for identity verification with a small demand on the computational power. Similarly, Belhumeur et al. (1997) demonstrate the robustness of Fisherfaces under varying light conditions, especially useful for Nepalese homes which suffer from electricity instability.

The results thus reflect the need to have algorithm selection that is context-sensitive. CNN, LSTM architectures excel at increasing the detection accuracy, whereas LBPH and Fisherfaces offer more flexible solutions under low-resource conditions. This is an affirmation that technology solutions cannot be universal and must be tailored to the infrastructures.

Theme 3: The Camera-Off Problem as a Structural Research Gap

One of the main findings of the PRISMA guided literature review is that there is a lack of studies that specifically explore the camera-off issue. Many studies have been conducted on the detection of student engagement through visual analytics, but very few studies have focused on cases where students deliberately go out of view. This is important especially in the developing world where bandwidth limitation, sharing devices or privacy concerns are driving camera-off participation rather than disengagement.

It indicates the data acquisition stage as the place of the breakdown. When it is not able to find a face, the system may declare the student absent, despite the fact that the learner might be cognitively engaged. This indicates a conceptual shortcoming of a purely vision-based system. For this reason, it's clear that the findings indicate that unseen student monitoring cannot only be based on facial recognition algorithms.

Theme 4: Multimodal Monitoring as a Contextual Solution

The results support highly the multimodal approach to engaging the student in the learning task, especially in light of the limitations of the visual-only approach. Log-based analytics, such as chat usage, quiz responses, polls, and all interactions logged with times (date, time, etc.), can also be used as a complement to facial detection systems. This holistic approach takes care of the validation of engagement even in the absence of visual input.

The use of multimodal data streams in the reporting part of the system flow can lead to a more balanced profile of attendance and engagement of instructors. This would be a form of monitoring rather than surveillance as it is the pedagogical responsibility. Based on the results, it can be inferred that the hybrid monitoring system that integrates AI-based detection with behavioral interaction logs is a more holistic and context-aware approach to the challenge of invisible learners.

Theme 5: Ethical and Pedagogical Issues

In this thematic finding is regarding the ethical aspects of real-time monitoring. Though algorithmic detection is beneficial for accountability, it can be detrimental to trust and autonomy of students if there is too much surveillance. The continuous tracking of the face might be problematic in terms of privacy issues, especially in a socio-cultural setting where some home spaces will be visible in online classes. Much of the literature focuses on the importance of monitoring having pedagogical purposes, not disciplinary ones.

Thus, the results indicate that technological frameworks need to be intertwined with pedagogical methods that involve interaction like formative assessment, collaborative learning, and gamified activities. This kind of integration is more about incentivizing people to be seen, than having to be seen.

Synthesis of Findings in Relation to Research Objectives

Based directly on the study's goals, the results are found to be:

- For low resource settings, an effective technological solution for monitoring students' unseen engagement must strike the balance between heavy computation and high accuracy of deep learning.
- The real-time engagement of students behind the camera is still unknown in camera-off situations, and there is a significant structural limitation in the existing engagement detection model.
- Combining the multimodal analytics of the face recognition with interactions based logs gives a more reliable solution and a more context-sensitive solution.
- Inclusive and responsible use of technology requires the accompaniment of ethical safeguards and pedagogical innovation.

The study shows that the challenge of unseen learners is not only technical but systemic in nature and is related to infrastructural, algorithmic, pedagogical and ethical aspects. This can only be resolved in the context of Nepal and other developing countries by implementing flexible, people-centered, and mixed monitoring systems which guarantee accountability and inclusiveness of virtual education.

Discussion

This study aimed at investigating how virtual classrooms can monitor students' unseen engagement in low-resource settings, and how students' 'real-time status' in virtual classrooms can be identified. The results clearly show that there are great advances in the technology of detecting engagements, but that these have limitations of practice and are structurally limited, especially in developing countries like Nepal.

Many studies have shown that deep learning algorithms perform well in determining a student's engagement, particularly Convolutional Neural Networks (CNNs) and Long Short-

Term Memory (LSTM) networks when visual data is available. CNNs have been adeptly used in extracting spatial information like facial orientation, eye gaze and micro expressions, which have been used to accurately classify attentiveness (Li et al., 2021). These models can be used in conjunction with LSTMs, which process temporal sequences, to distinguish between distraction over a short period and complete disengagement (Monkaresi et al., 2017). This type of hybrid architectures are well suited to the suggested system flow in this research, especially for the validation and measurement of accuracy. CNN–LSTM models can thus be integrated into high-resource education systems to enable the real-time monitoring of student presence and engagement with the material, while offering high reliability.

The situation is more complicated however when the camera-off phenomenon is taken into account. Most engagement detection studies in the world have been conducted in the context of continuous availability of face data, however this does not apply to many developing scenarios. Bandwidth constraints, power fluctuations, device sharing restrictions/privacy concerns are common reasons for students to disable cameras in Nepal (Rana & Rana, 2020; Shrestha et al., 2022). The system fails at the first stage of image capture when visual data is not available, and the subsequent algorithmic processes are not effective. This structural decomposition is revealing that a major restriction in vision-based engagement detection models is its inability to directly handle unseen learners and a major research gap is the handling of unseen learners directly. Technological capacity is present but contextual feasibility is unknown. The results also indicate that the recognition algorithms used in this work are more flexible in the low resource scenario like Local Binary Pattern Histogram (LBPH) and Fisherfaces. The use of texture-based image processing (Ojala et al., 2002) has made a reasonable LBPH algorithm, which requires little computation and uses grayscale images. Also, due to the fragility of lighting situations, Fisherfaces obtained by Linear Discriminant Analysis (Belhumeur et al., 1997) are a desirable characteristic for homes with unreliable power. These techniques do not always achieve the best level of classification accuracy as deep neural networks but provide viable alternative techniques that are appropriate for Nepal's infrastructural limitations. This means that the effectiveness of the algorithmic solutions is not guaranteed and they need to be “tuned” to the socio-economic conditions.

Discussions aren't only about the algorithmic feasibility, but also conceptual interpretations of engagement. The general view is that student engagement is a multi-dimensional construct (Fredricks et al., 2004; behavioral, emotional and cognitive).

The main advantage of vision-based systems is that they can measure both behavioral (e.g. gaze direction, facial expression) and cognitive indicators, which can be less accurate. A student who seems to be "out of it" could be thinking about what he or she sees on the screen, and a student who is looking at the screen could be "out of it" mentally. This helps to reinforce some of the arguments that highlight that engagement detection systems should not use visibility as learning. Thus, using only facial analytics could mean the answer to complex learning processes is simplified.

Log-based evidence of engagement is possible through the chat, the responses to quizzes, time in the platform and by interactions with the platform (Cocea & Weibelzahl, 2011). The hybrid method, which integrates these information sources, improves the presence verification and addresses the negative aspect of misclassifying absentees as presences, using facial recognition. This multimodal logic is particularly helpful in situations where there is no camera, and the interaction data can be leveraged instead of the visual data. Hence the focus of the conversation is on moving from single modality AI models to learning analytics ecosystems.

Real-time monitoring regimes are also fraught with ethical issues. Facial tracking is an ongoing process and this poses privacy, consent and surveillance issues, particularly if the students are tracked at home. With this research on affective computing, the use of engagement detection systems for surveillance purposes should be avoided and the systems should deliver pedagogical actionable feedback (Whitehill et al., 2014).

The unseen learner issue discussed is that technology is not a cure-all. Engaged learning doesn't require cameras on; quizzes, activities that can be completed collaboratively with peers, break out conversations and gamified learning can be used to generate voluntary engagement. This fits with the general conceptual models of digital pedagogy in design and not surveillance as engagement. Hence, it does not intend to be a disciplinary system, but rather a supportive system is to be considered.

As these ideas are synthesized, the nature of the problem of the unseen student is revealed to be a multi-dimensional challenge with technological, infrastructural, pedagogical and ethical aspects. In controlled environments, the accuracy can be improved by using more sophisticated AI models, but in Nepal, digital inequality and contextual factors restrict their relevance. Lightweight algorithms give some answers and aren't very deep. Multimodal analytics and pedagogical innovation are the way forward. Finally, this research presents the need for a blended model to cover the needs of the invisible student, one that is accurate through algorithms as well as viable through infrastructure and ethical. Future research should thus focus on camera-off engagement scenarios, extend their multimodal system to scale, and evaluate their effectiveness over time on learning outcomes. If AI can't do it within this context, the idea of detecting engagement will be a monumental, but incomplete, technical achievement.

Conclusion

With the advent of rapid virtual learning environments, the methods of delivering education globally are also changing, with new challenges arising in the realm of real engagement of pupils, which is hard to prove. This study aims to understand what is meant by “unseen” learners in a virtual classroom and what this means in terms of technological and contextual means of making learners "visible" in a virtual classroom, particularly in a low-resource country such as Nepal. The study aimed at achieving two main goals: understanding the actual state of the student behind the camera and determining possible methods of engagement monitoring in limited settings with a structured system workflow framework to guide the systematic review.

The results show that with ongoing visual information, advanced deep learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks provide a high accuracy of student engagement detection. These models can model the spatial and temporal aspects of these, and distinguish between a short interruption and a longer dissociation. However, they can only operate when the camera is operating. The system doesn't detect when students often switch off their camera for bandwidth issues, power problems or someone else is using it, or maybe because the students don't want anyone to see them. This shows the limitation of the structure of vision-based engagement monitoring methods and the significant research gap in dealing directly with the camera-off situation.

In less resourceful situations, however, the techniques like Local Binary Pattern Histogram (LBPH) and Fisherfaces can be more flexible. The above techniques are not exhaustive and can be used for identity verification if computing power is limited and the light sources change over time. But full solutions to the "unseen learner" problem cannot be obtained with algorithmic monitoring. Engagement has a more than a mere facial visibilities or gaze detection, it has a behavioural, emotional and cognitive dimension.

Therefore, in this research, it is revealed that hybrid and multimodal combination has to be employed to be able to monitor the unseen students effectively. Combining the power of face recognition with log-based data like chat activity, quiz answers, and timestamps of interactions gives a more holistic and context-rich view of engagement. This integration helps to minimize the risk of students who have technical difficulties being categorized as "disengaged" students, and it helps to improve accountability without the use of a more surveillance-based approach.

Importantly, the study makes it clear that technological interventions need to take account of ethical and pedagogical issues. Facial monitoring that is continuous poses questions about privacy, consent and student autonomy, especially in the home settings of learning. Hence, the role of systems to detect engagement must be supplementary to teaching and learning and not a system of control. The pedagogical innovation of interactive teaching, formative assessment, cooperative learning and student-centered teaching has always been the focus of how to achieve truly participative teaching in virtual classrooms.

Finally, it is important to realize that the unseen learner problem is not a technical problem, per se, but a systemic problem that is influenced by infrastructure inequity, access inequity, pedagogical design, and ethical responsibility. Scalable and inclusive solutions should be focused on contextual feasibility as well as technological innovation, especially in developing countries like Nepal. Future studies are needed to explore camera-off engagement modeling, follow-up impact evaluation, and the creation of ethically informed and multimodal learning analytics. These are the only integrated and humanistic approaches that can help virtual education to be accountable and inclusive in the changing landscape of digital learning.

Research Gap and Future Recommendation

Most works concur that the Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks and lightweight algorithms (LBPH, Fisherfaces) all agree on the continuous visual input being a predetermination. However, camera-off, infrastructure instability and device sharing that are common in developing countries is not a well-researched aspect in laboratory classroom environments. This suggests a strong empirical need: the lack of integration of multimodal analysis in the real classroom with the previously developed context-sensitive frameworks of engagement monitoring systems. Therefore, the recent studies should be moved out of the stage of conceptual synthesis into the stage of experimental implementation and demonstration of a hybrid system of engagement monitoring in the practice of real-time virtual classes. An empirical study would allow the evaluation of performance in different bandwidth scenarios, test the algorithms' consistency in camera-out scenarios and meditate on the pedagogical and ethical implications of such practice. This gap between theory and experimental practice will make it easier to close this gap and will be the basis for the development of scalable, inclusive and contextualized engagement detection systems which can be used in low resource learning environments.

Acknowledgement

I would like to express my sincere gratitude to Nepal Open University for providing the academic framework and support necessary to complete this study.

Conflict of Interest

In this article, the author declares that there are no conflicts of interest regarding the publication.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not for profit sectors.

AI Disclosure

Though the main concepts, analysis, and writing style belong to us, we made use of AI assistants to refine our language skills and ensure that there were no grammatical mistakes in the paper. AI tools were not used beyond basic editing purposes. We, the authors, are fully responsible for this paper’s originality and its correctness.

References

Ayari, A., Chaabouni, M., & Ben Ghezala, H. (2025). A deep learning approach for automatic detection of learner engagement in educational context. In *Proceedings of the 17th International Conference on Computer Supported Education* (pp. 372–379). SCITEPRESS – Science and Technology Publications. https://doi.org/10.5220/0013283200003932

Basnet, S., Basnet, H. B., & Bhattarai, D. K. (2021). Challenges and opportunities of online education during Covid-19 situation in Nepal. *Rupantaran: A Multidisciplinary Journal*, *5*(1), 89-99.

https://doi.org/10.3126/rupantaran.v5i01.39867

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711–720. https://doi.org/10.1109/34.598228

Bosch, N. (2016, July). Detecting student engagement. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (pp. 317–320). ACM. https://doi.org/10.1145/2930238.2930371

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., & Zhao, W. (2015, March). Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 20th International Conference on Intelligent User Interfaces* (pp. 379–388). ACM. https://doi.org/10.1145/2678025.2701397

C, B. V. (2016). A descriptive study on e-governance. *International Journal of Computational Science and Information Technology*, *4*(1), 67–74. https://doi.org/10.5121/ijcsity.2016.4107

Christenson, S. L., Reschly, A. L., & Wylie, C. (Eds.). (2012). *Handbook of research on student engagement*. Springer US. https://doi.org/10.1007/978-1-4614-2018-7

Cocea, M., & Weibelzahl, S. (2011). Disengagement detection in online learning: Validation studies and perspectives. *IEEE Transactions on Learning Technologies*, *4*(2), 114–124. https://doi.org/10.1109/TLT.2010.14

Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, *6*(1), 1. https://doi.org/10.1186/s40561-018-0080-z

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5–22. https://doi.org/10.1177/0047239520934018

Elazab, S., & Alazab, M. (2016). The effectiveness of the flipped classroom in higher education. In *Proceedings - 2015 5th International Conference on e-Learning, ECONF 2015* (pp. 207–211). IEEE. https://doi.org/10.1109/ECONF.2015.34

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gharti, L. (2023). Challenges of online learning in the post Covid-19 era: Lived experiences of teachers in remote Nepal. *English Language Teaching Perspectives*, *8*(1-2), 80–96. https://doi.org/10.3126/eltp.v8i1-2.57861

Government of Nepal. (2018). *2018 Digital Nepal Framework: Unlocking Nepal's growth potential*. Ministry of Communication and Information Technology. http://mocit.gov.np/application/resources/admin/uploads/source/Others/Digital%20Nepal%20Framework%20Book%20updated.pdf

Huawei. (2018). *Accelerating SDGs through ICT digital solutions: Next steps recommendations for accelerating digitally enabled-sustainable development*.

Jones, P., Wynn, M., Hillier, D., & Comfort, D. (2017). The Sustainable Development Goals and information and communication technologies. *Indonesian Journal of Sustainability Accounting and Management*, *1*(1), 1–14. https://doi.org/10.28992/ijsam.v1i1.22

Klapper, L., El-Zoghbi, M., & Hess, J. (2016). *Achieving the Sustainable Development Goals: The role of financial inclusion*. CGAP.

Kostoska, O., & Kocarev, L. (2019). A novel ICT framework for sustainable development goals. *Sustainability*, *11*(7), 1961. https://doi.org/10.3390/su11071961

Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2017). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, *8*(1), 15–28. https://doi.org/10.1109/TAFFC.2016.2515084

Nguyen, T. (2015). The effectiveness of online learning: Beyond no significant difference and future horizons. *MERLOT Journal of Online Learning and Teaching*, *11*(2), 309–319.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 971–987. https://doi.org/10.1109/TPAMI.2002.1017623

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, Luke A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*(n71). https://doi.org/10.1136/bmj.n71

Pangeni, S. K. (2017). Open and distance learning: Cultural practices in Nepal. *European Journal of Open, Distance and E-Learning*, *19*(2), 32–45. https://doi.org/10.1515/eurodl-2016-0006

Rayamajhi, S., & Rana, K. (2024). Reflecting on online learning experiences of secondary school teachers and students during COVID-19. *KMC Journal*, *6*(2), 171–190. https://doi.org/10.3126/kmcj.v6i2.68900

Tang, Y. M., Chen, P. C., Law, K. M. Y., Wu, C. H., Lau, Y., Guan, J., He, D., & Ho, G. T. S. (2021). Comparative analysis of student's live online learning readiness during the coronavirus (COVID-19) pandemic in the higher education sector. *Computers & Education*, *168*, 104211. https://doi.org/10.1016/j.compedu.2021.104211

Tjoa, A. M., & Tjoa, S. (2016). The role of ICT to achieve the UN Sustainable Development Goals (SDG). *IFIP Advances in Information and Communication Technology*. https://doi.org/10.1007/978-3-319-44447-5_1

Vail, A. K., Grafsgaard, J. F., Wiggins, J. B., Lester, J. C., & Boyer, K. E. (2014, November). Predicting learning and engagement in tutorial dialogue. In *Proceedings of the 16th International Conference on Multimodal Interaction* (pp. 255–262). ACM. https://doi.org/10.1145/2663204.2663276

Welchman, L. (2015). *Managing chaos: Digital governance by design*. Rosenfeld Media.

Whitehill, J., Serpell, Z., Lin, Y. C., Foster, A., & Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing*, *5*(1), 86–98. https://doi.org/10.1109/TAFFC.2014.2316163