Understanding the Role of Human-Computer Interaction Technologies in Detecting Learners' Attention Ans Engagement: a Systematic Literature Review

Yuchen Zhang

Capital Normal University High School, Beijing, 100000, China

Keywords: Human-computer interaction, Online teaching, Systematic literature review, Eye tracking, Facial recognition, Neural network

Abstract: In the pandemic era, more and more schools or institutions have begun to choose or be forced to conduct education activities online. In China, the current mainstream online teaching model is that teachers are talking endlessly, while most students passively listen. Some educators require cameras to be turned on, while others do not. This makes many people question the effectiveness and quality of online teaching compared to offline teaching since teachers can no longer monitor students' performance in real time. In fact, low-efficiency in online teaching is also a problem that educational institutions are trying to solve. The majority of scholars believe that the relatively low quality of online teaching may be caused by the limitations of current softwares and hardwares monitoring ability. Some researcher in the area Human-Computer Interaction (HCI) dived deep into the problem and tried to develop solutions and supporting systems to facilitate quality online education. In this paper, I conducted a systematic literature review of HCI research applied to detection of learners' attention and engagement on online education settings. Based on the literatures, I discuss the current solutions and future directions from a macro perspective.

1. Introduction

Scientists and scholars in the field of human-computer interaction (HCI) have been doing research on identifying human attention and teaching effects for a long time. I started a computer club at school. During the winter and summer holidays, I usually go to Guilin, a city in southwest China, to teach the local middle school children some computer knowledge, because there are no computer courses in the school, and the students don't know about the principles of computers. However, due to the epidemic, our community had to switch to online teaching activities. However, I found a problem: without the camera, my interaction with the students was significantly reduced, resulting in unsatisfactory teaching content and effect. Therefore, in order to find a solution to this problem, I read some articles about human-computer interaction research. I have been impressed by many clever mathematical models and computer systems, but unfortunately most of these models and systems are not put into practice. This puzzled me and I decided to write this literature review. This paper will more comprehensively consider the application and prospect of human-computer interaction in the context of teaching, that is to say, it will not only consider the technical level of the article.

Teaching is actually a very broad topic. When we want to define the impact of human-computer interaction on teaching and learning, we need to consider this topic as comprehensively as possible. In this article, we chose the method of classifying articles. As the beginning of teaching, we need to decide the target group and the type of teaching. In the process of teaching, the human-computer interaction technology involved in teaching is also indispensable. The attitude towards human-computer interaction technology and the output of teaching are the results of teaching. In this way, we can divide teaching into five categories for more detailed study.

The first direction to focus on is the part about classroom output. What is classroom output? A more clear and easy-to-understand definition is what kind of positive or negative impact the device or method has on the sender (I.e. teachers) and the receiver of education activities(I.e. students). At the same time, classroom output also includes the interactive effects of education givers and

education receivers, such as learners' attention or classroom participation. The second focus is on the technical and model sections. The third direction of research is the attitudes of education recipients and education givers towards human-computer interaction methods, respectively. These attitudes may be negative or positive. We focus on the collection of social data in this direction. The fourth direction is the classification of educational background. The last category of topics is the target learner. For students of different age groups and students of different majors, teachers use different teaching methods and teaching purposes.

We believe that studies in these five directions can help researchers understand the current overview of the field as well as potential future research questions. From this, two key research questions can be constructed;

RQ1. In the field of human-computer interaction, what is the current state of learner attention and classroom engagement in the educational context? In other words, what has been done?

RQ2. What areas of research are underexplored and underexplored? In other words, what should we do next?

2. Related Work

In this section, we will introduce some less common concepts or nouns to help readers better understand the content of this article.

2.1 Literature Reviewing HCI Research

Literature review is a kind of article that makes a comprehensive introduction and elaboration of the current topic, problem or research topic by reading, analyzing, summarizing and sorting out the latest progress, academic opinions or suggestions. In this paper, we seek to provide the some benefits for research on human-computer interaction in the educational context.

We selected articles from the ACM Digital Library (DL). We searched the full database and extracted papers that are related to the topic we are investigating, striving for a systematic approach to the literature review.

2.2 Introduction and Overview of Technical Aspects of Detection of Attention and Engagement

Considering that the backbone of this paper is still technology-oriented, we first introduce the current means and methods of human-computer interaction that have mainstream applications in the educational background. Here I will focus on the two methods mentioned above: eye tracking and face recognition.

At the same time, in order to place the research in the larger context, we will introduce the concept of joint visual attention (JVA). This is a widely studied phenomenon and mechanism. Through it, information about teachers and students will also facilitate their mutual communication. And, Oh reviews previous research that uses eye tracking and visualization of facial recognition to aid communication among peer learners. The literature on differences in expert and novice gaze patterns and facial feature patterns forms the basis for our expectations that these techniques do affect learners. On the other hand, the literature on the use of these techniques for real-time tracking of students' gaze and face states also provides evidence for my expectation that these techniques and methods can indeed help educators in their teaching activities.

2.2.1 Eye Tracking

When people's eyes look in different directions, subtle changes could be observed, and can further produce features that can be extracted through image capture or scanning[1]. More specifically, techniques could be used to track the changes of the eyes in real time, predict the user's state and demand, and respond to the purpose of controlling the device with the eyes. This is the scientific principle of eye tracking technology. Eye tracking devices can be categorized into infrared equipment and image acquisition equipment. Infrared projection is relatively more accurate in small screens, which can be accurate to within 1 cm on a 30-inch screen, supplemented by technologies

such as blink recognition and gaze recognition, which can already replace the mouse and touch pad to a certain extent[2]. In addition, other image acquisition devices, such as cameras on a computer or mobile phone, can also implement eye tracking with supporting so ftwares, with varying accuracy, speed, and stability. Differences in these devices were at one point the focus of scientists' research (5 out of 30 papers covered it). In short, eye tracking is an important way for researchers to explore students' learning processes[3].

2.2.2 Facial Recognition

The process of facial expression recognition generally includes three steps: face detection and localization, facial expression feature extraction, and facial expression classification. First, the input face image or image sequence is detected and located through the face detection algorithm, and image processing is added to minimize the impact of the external environment on the detection results. The second step is to extract the information that can preferentially characterize the facial expression features from the face image or image sequence, and often reduce the dimension of the extracted features to avoid the calculation speed being too slow. The third step is to analyze the expression features extracted in the second step, and the expression features can be divided into six basic expressions, or divided into a combination of facial expression activity units. With facial recognition technology, researchers can more easily monitor the status of students and teachers in class to evaluate the effectiveness of teaching[4].

2.2.3 Joint Visual Attention

Joint visual attention is a central construct in collaboration and communication research. It has been extensively studied by many human-computer interaction researchers. JVA is a tendency of social objects to focus on common references and monitor each other's attention to external entities. In a variety of complex communication scenerios, JVA is considered to be a key component of information sharing. That is, during communication, the speaker uses gaze and gestures to direct attention to relevant objects or events, while the listener also uses gaze to disambiguate and follow the speaker's speech. More importantly, JVA can be quantified in the interactive process and the metric is found to be correlate with attention. For example, in the context of collaborative learning or playing, JVA has been shown to predict task performance or collaboration quality.

3. Methodology

To address two research questions, we conducted a systematic literature review of human-computer interaction research at the ACM Digital Library (DL). As mentioned above, we chose ACM as our corpus because it has been used in past work as a corpus for similar literature reviews in other fields.

3.1 Identifying Search Terms/Corpus Collection

The first step in any review is to identify relevant work. To achieve this, we identified search terms. Human-computer interaction in the educational context has been referred to by many terms, and we tried to build a comprehensive set of keywords including terms that may have been used in the old literature on human-computer interaction in the educational context. Subsequently, I repeated the corpus. Related keywords were iteratively searched. First, we identified search terms based on common terms used in HCI as well as well-known and cited HCI literature monographs. The initial search terms were: “educational outcome”, “technical aspects”, “attitudes towards ... (technologies)”, “educational settings”, “target user”.

Next, we extracted the keywords I defined from the corpus from the first search of ACM DL. I compiled a list of all keywords that appeared in two or more papers and then filtered those that were too vague (e.g. multi-agent systems, communication) or not conceptually related to human-computer interaction Words (eg experimentation, dynamic effects). Afterwards, we considered the remaining recurring keywords to be synonymous or closely related to human-computer interaction in the educational context. Then use this list to search. This process

was performed twice until the search yielded no new papers. A total of three searches were performed. Ultimately, the following keywords were identified (corresponding to those mentioned in the introduction):

*Outcome: “teaching outcome”, “learning outcome”, “learners' attention”, “learns' engagement”.

*Technical aspects: “eye gaze”, “emotional recognition”, “chat message”.

*Attitudes: “teachers' attitudes”, “learners' attitudes”.

*Educational settings: “online”, “offline”, “lectures”, “discussion/seminar”.

*Target user: “K12”, “higher education”, “special education”.

This search process yielded a total of 84 papers, followed. We use specific criteria to filter articles. to ensure their appropriateness.

3.2 Selection Criteria

The first selection criterion is that articles must be written in English due to limitations of personal language proficiency and the general tendency to publish high-quality research in the field of computing in English. Second, we filtered articles to exclude studies that had no real connection to human-computer interaction in an educational context (although it used one of the keywords) or that were below a quality threshold. As it is personal work, there may be some subjective elements to the results, but the vast majority of papers can be considered suitable for further research.

After the second stage of screening, more than 50 papers remained. And these papers are the final corpus of research.

3.3 Data Extraction

We extract from each paper the information I need, including: 1) the statement or research question being addressed, 2) the center or focus of the study, 3) a brief statement about the sample population or subjects, and 4) Key findings related to the research question. In addition, we extracted keywords and coded the methodology employed, including quantitative, qualitative, mixed methods, design, mathematical modelling, or literature review.

3.4 Data Analysis

To identify topic topics in the literature, we encoded each paper in the final corpus. This is beneficial for solving both problems. We used the selection criteria mentioned in 3.2 and 3.3 to extract article content and summarize it into a spreadsheet. It is worth mentioning that since the amount of data in the literature review is not large, Excel was chosen as the software for data analysis and charts were formed to make the trend and overview of low expression more obvious.

4. Result

4.1 Analysis and Results of Research Question 1

In this section, we will categorize the data collected on articles to address the two key research questions mentioned above. First, the focus is what have we already done? According to the final corpus, the macro overview is actually easier to see. We have made a table of keywords and core research content including five major directions (mentioned in the introduction), which can clearly see the past HCI scholars in the field of human-computer interaction in the context of teaching. Tendency and focus.

Because there may be more than one label provided per paper, the total number of data for the graphs in Figure 1 adds up to a bit more than the number of papers in the final corpus.

In the Figure 1, it can be seen that the technical aspects are the most relevant papers. As mentioned above, the update and correction of testing equipment or core technologies is a topic that many researchers are keen on. The second and third places are educational background and teaching output. This is actually a side effect of the unusually large number of technical articles. Because when it comes to newer technologies or models, they must be applied in actual scenarios before they can be accepted by other researchers. Therefore, the research in the above three fields is more detailed and thorough. On the contrary, for the two directions of attitude and target users, the

number of people involved is greatly reduced. In the statistics, the total number of related papers in these two directions combined is not more than half of the technical direction. This shows that previous researchers' research in these two directions is lacking. In addition, it can be found that both the attitudes and the users of goals tend to be sociological, anthropological or pedagogical research areas[5]. It is different from the focus of algorithm research in traditional human-computer interaction. Perhaps this is also the possible reason why these two directions have been neglected by researchers. In fact, among the 14 papers in these two directions, there is no anthropological research directly related to human-computer interaction in teaching. Researchers put more experience on information acquisition in social media.

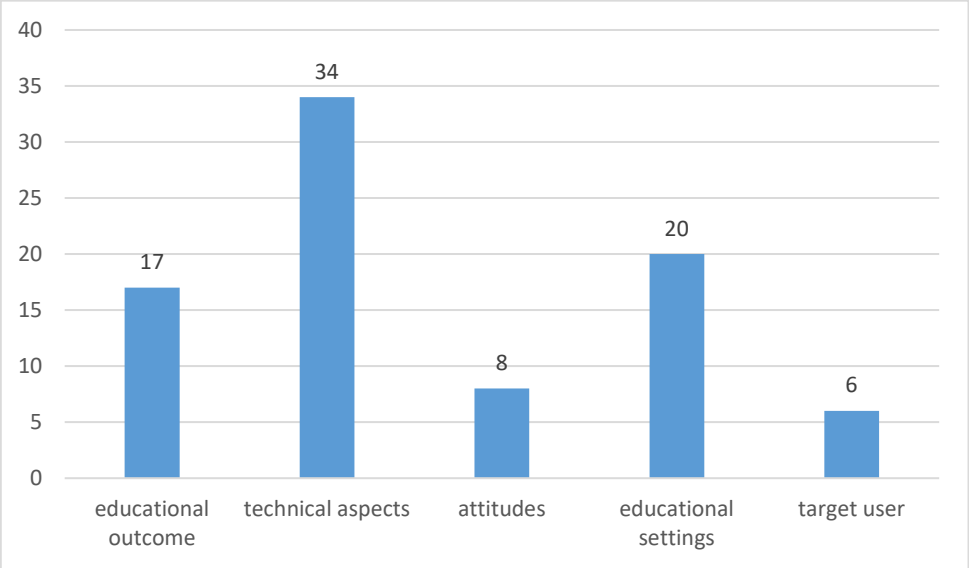

Figure 1. Statistical Chart on Content Classification

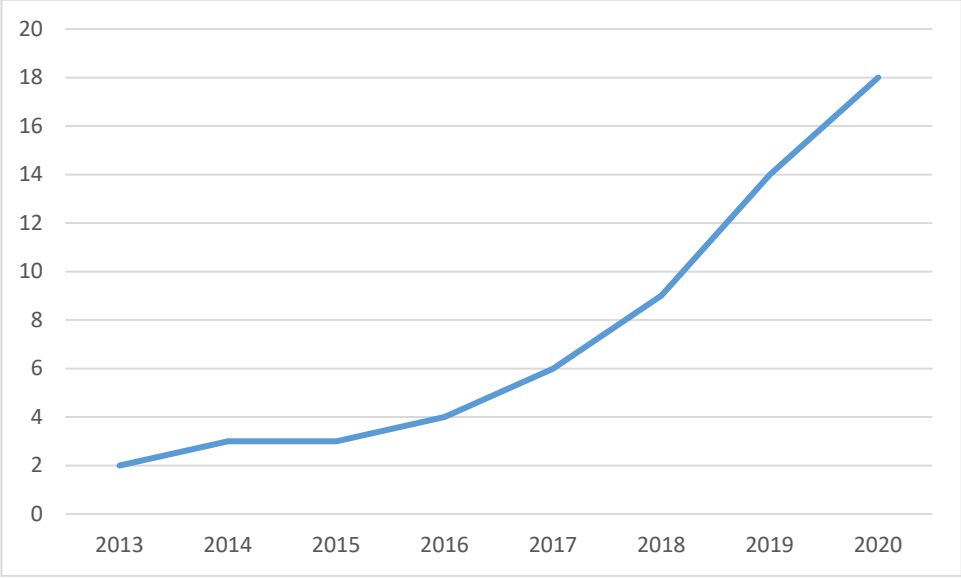

Figure 2 Trend Chart of the Number of Published Hci Papers from 2013 to 2020

In Figure 2, we organize the annual number of publications of relevant papers by time. It can be seen that the trend is still more obvious. With the growth of the year, the number of papers published each year is also increasing, and the growth rate is generally greater than zero, indicating that its growth rate is really accelerating. It can be seen that human-computer interaction in the context of education is indeed a topic that has become more and more popular in recent years. With the continuous development of computer technology and algorithms, this field has also made great achievements in recent years.

In conclusion, we believe that the results of the discussion of the first research question are clear.

With the rapid development of computer technology, artificial intelligence, neural networks and other high-tech technologies, we have done a lot of research on related technologies, and achieved fairly high-level results. However, in some areas of sociology and humanities, such as attitudes of learners or target users, researchers are currently not doing enough.

4.2 Analysis and Results of Research Question 2

In the analysis of this problem, in fact, the results of 4.1 have already enlightened. A little bit of a flaw in the work we've done, maybe that's where more attention and focus will be needed in the future. That said, it is obvious that researchers need to do more research on learners' attitudes or target users, because the implementation of related technologies cannot be separated from people's attitudes and perceptions of them and the relevant target groups. In other words, if people resist a new technology or a new technology has no target population in the current society, then the technology is actually useless or even a waste of resources[6]. For this reason, research in the above two directions is particularly important.

In addition to this, we document the methodology of each article and the nature of the key research questions when processing the literature data. These two pieces of information may serve as a reference for future research by researchers. After all, the methodology used by most researchers may indeed have advantages in this field that no other method can match. Alternatively, some unpopular research questions might be better served by new research. These are all things that we cannot accurately predict at the moment. This is why we need a literature review to sort through this type of article data.

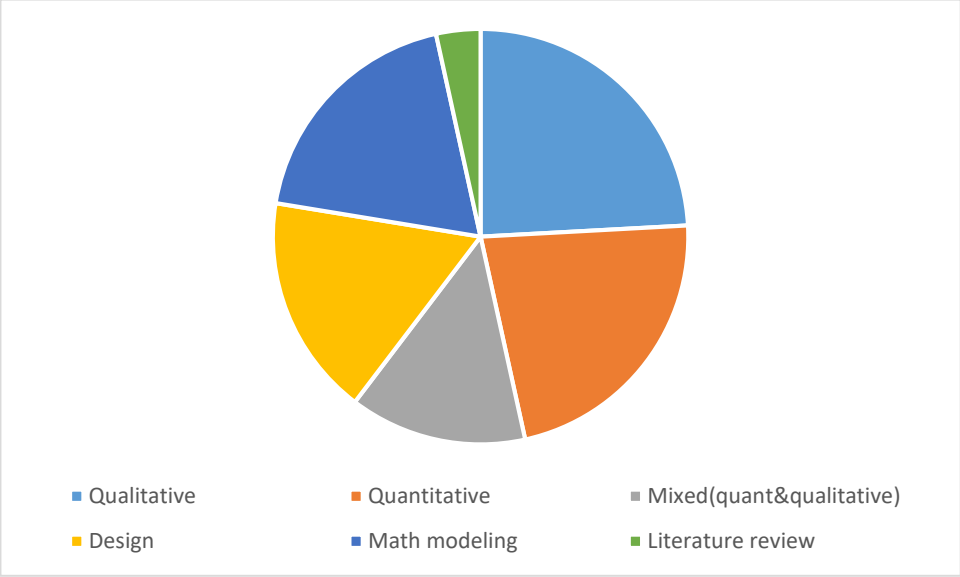

Figure 3. Statistical Chart by Methodology of the Paper

The top two Figure 3 and Figure 4 are the graphical results of the analytical methodology and research questions. Focusing on Figure 3 first, the most noticeable thing is the number of literature reviews. Of the 59 papers counted, there are only two relevant literature reviews. The subject also does not coincide with the purpose of this article. The two articles put more focus on areas such as technology and social media. That is to say, the analysis of the current human-computer interaction research in this paper may be unique, which may be one of the reasons why this paper has reference value. In addition, the number of other methodologies is not very different. The two most common types are quantitative analysis and qualitative analysis. The application of these two methodologies in other fields is also huge. Therefore, the unique research value here may not be displayed. A large number of mathematical modeling papers appeared to make up for the theoretical deficiencies of newly developed techniques. Even the theoretical construction progress has surpassed the actual research equipment and models. This reflects the huge future potential of the field. In other words, researchers who may be technology-oriented in recent years may not have concerns about lack of issues.

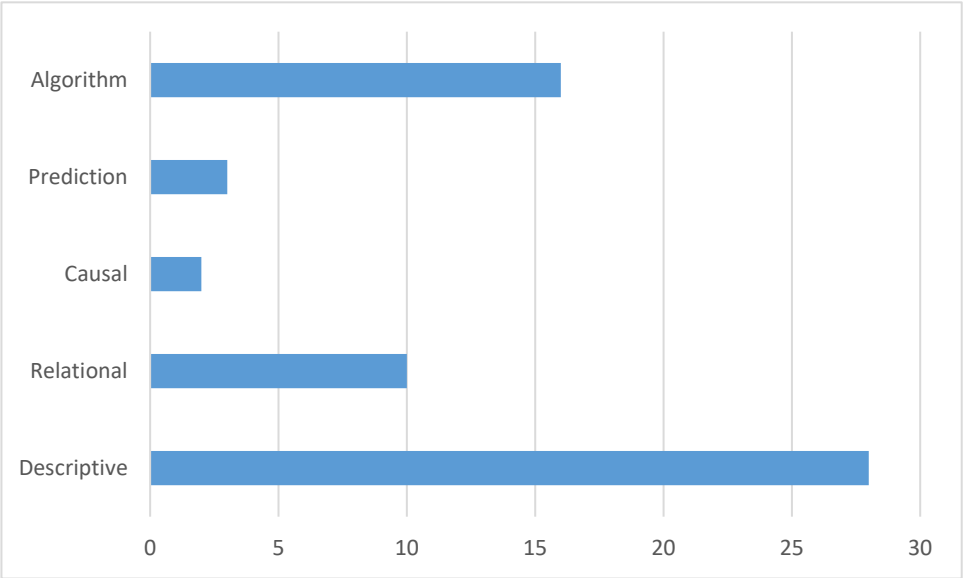

Figure 4. Statistical Chart Categorizing a Paper by Its Research Question

Next, analyze the chart in Figure 4. Descriptive research questions can be found to prevail. The number of papers of 28 indicates a researcher's propensity for this type of descriptive research question. We don't yet know why, but the ease of argument for such questions is one factor worth considering. At the same time, in some sociological studies, descriptive research questions are the first choice of most researchers, because the experimental conditions for causation or association will be harsh. The issue of causality and association is also a topic of preference for technical researchers. If these two types of relationships can be successfully demonstrated, it will have a great positive effect and influence on the updated techniques or models of researchers[7]. Research at the algorithm level also focuses on two key technologies, eye tracking and facial recognition. In recent years, the development of neural networks has spawned a number of new algorithms and artificial intelligence learning models for researchers to choose from. In the papers in the final corpus, newly discovered neural network learning modes such as CNN, RNN, and LSTM have greatly improved the accuracy and system complexity of technologies such as facial recognition[8]. The researchers used these changes to evaluate issues such as the viability of the system and the model. In the 16 algorithm papers, more than half of the topics are in line with our discussion. In addition, there are only three articles of the prediction type, which leads to thinking. At present, the topics of the three prediction-type articles are the type predictions that are more suitable for the target users of the system and software. This also illustrates the lack of research on this type of problem by researchers. In fact, many of the newly developed models are not actually used in everyday life, but only work in the laboratory stage. The reason for this may be the lack of such prediction-type articles, making it difficult to test these techniques that only pass the theoretical test.

To sum up, through the discussion and analysis of the methodology and research questions, we see that there is still great potential for human-computer interaction in the educational context at the technical level. At the same time, some issues related to humanistic care are still lacking. This can lead to issues such as derailment of technology from practical applications or limitations of experimental scenarios.

5. Conclusion

Literature reviews of ACM publications on human-computer interaction are limited by a number of limitations. First of all, ACM DL's search engine has its own set of limitations. For example, although researchers use the same search terms and date ranges, search engines sometimes produce different results. And, while ACM DL claims to be the most comprehensive digital library, it does not contain all possible relevant articles. In addition, articles in the review are limited to papers written in English, which may show different trends from papers published in other languages.

Finally, we provide the results of a systematic review of 59 papers on human-computer interaction in the context of teaching in information processing, thus filling in the gaps left by other related literature reviews. Our review reveals a wealth of current achievements in the technical field and important trends that are increasingly biased towards computer algorithms and neural networks. At the same time, it also reflects the lack of and needs to be supplemented in the intersection of computer and sociology, anthropology and other fields. These results reveal some short- and long-term opportunistic representations of research directions in information processing and human-computer interaction in the educational context. For example, to complement the lesser methodologies of current research, focus on literature reviews and macro representations, and combine the strengths of high-tech equipment and models.

References

[1] Craig Hennessey and Jacob Fiset. “Long range eye tracking: bringing eye tracking into the living room. In Proceedings of the Symposium on Eye Tracking Research and Applications”. Association for Computing Machinery, vol.21, no.8, pp.249-252, 2012.

[2] Ao Liu, Lirong Xia, Andrew Duchowski, Reynold Bailey, Kenneth Holmqvist, and Eakta Jain. “Differential privacy for eye-tracking data”. Association for Computing Machinery, vol.14, no.3, pp.1-10, 2019.

[3] Mike Bartels and Sandra P. Marshall. “Eye tracking insights into cognitive modeling”. Association for Computing Machinery, vol.11, no.7, pp.141-147, 2006.

[4] M. Maithri, U. Raghavendra, Anjan Gudigar, Jyothi Samanth, Prabal Datta Barua, Murugappan Murugappan, Yashas Chakole, and U. Rajendra Acharya. “Automated emotion recognition: Current trends and future perspectives”. Comput, vol.10, no.6, pp.10-16, 2022.

[5] Edward Tse, Johannes Schöning, Yvonne Rogers, Chia Shen, and Gerald Morrison. “Next generation of HCI and education: workshop on UI technologies and educational pedagogy”. Extended Abstracts on Human Factors in Computing Systems, vol.17, no.5, pp.4509-4512, 2012.

[6] Andreas Bollin, Stefan Pasterk, Max Kesselbacher, Elisa Reci, Markus Wieser, and Nina Lobnig. “HCI in K12 Computer Science Education – Using HCI as a Topic and a Didactic Tool”. 14th Biannual Conference of the Italian SIGCHI Chapter (CHItaly '21), pp.1–8, March 2021.

[7] Lei Qu and W. Lewis Johnson. “Detecting the Learner's Motivational States in An Interactive Learning Environment”. In Proceedings of the 2005 conference on Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology, pp.547–554, 2005.

[8] Urban Burnik, Janez Zaletelj, and Andrej Košir. “Video-based learners' observed attention estimates for lecture learning gain evaluation”. Multimedia Tools Appl, vol.77, no.13, pp.16903-16926, 2018.