Investigating User’s Attentional Focus in Computational Environments

A Literature Review with Emphasis on Webcam Data

Cassiano da Silva Souza
cassiano.souza@edu.pucrs.br
Pontifical Catholic University of Rio Grande do Sul and Federal Institute Sul-rio-grandense
Porto Alegre, Brazil

Milene Selbach Silveira
milene.silveira@pucrs.br
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil

Isabel Harb Manssour
isabel.manssour@pucrs.br
Pontifical Catholic University of Rio Grande do Sul
Porto Alegre, Brazil

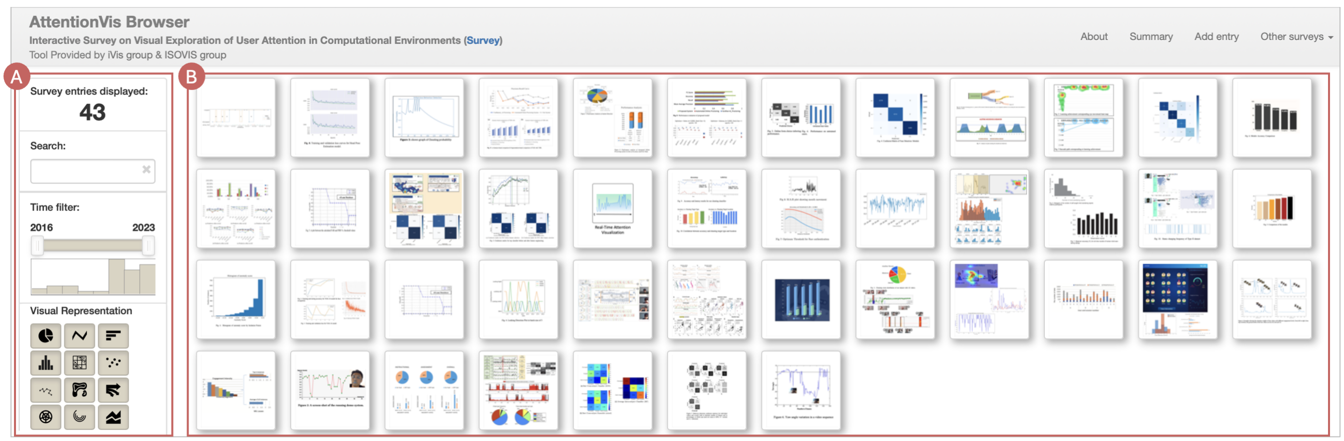

Figure 1: *AttentionVis Browser*: a web-based user interface of our visual survey, which is composed of (A) the interaction panel, including the search field and filters by category, and (B) the main panel - thumbnails representing each paper.

ABSTRACT

Maintaining the user’s attentional focus has become a recurring concern in recent years. This is due to the consolidation of remote and hybrid models for study and work, which were widely experienced during the social distancing caused by COVID-19. This paper presents a review of works that address this problem by analyzing webcam data, a promising device for behavioral studies. The literature review from 2013 to 2023 was carried out using a hybrid search strategy, through which we selected and analyzed 57 papers. The summary of this study is presented in an interactive visual survey format called the *AttentionVis Browser* tool. As additional contributions, we provide a list of lessons learned, a list of work limitations, and possibilities for future research.

KEYWORDS

attention monitoring, webcam, data analysis, literature review.

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia’2024). Juiz de Fora, Brazil. Porto Alegre: Brazilian Computer Society, 2024.

ISSN 2966-2753

1 INTRODUCTION

The advancement of information and communication technologies (ICT) has significantly impacted our daily lives. According to the Internet Steering Committee in Brazil [10], with the advent of the COVID-19 pandemic and given the barriers imposed by social isolation to contain the spread of the virus, the demands for such resources proved to be fundamental for the continuity of work, education, and social interactions. In this context, remote and hybrid models stand out, from temporary solutions to consolidated practices in the post-pandemic scenario.

Recent surveys on the work environment show a growing adoption of remote work, with significantly higher numbers than before the pandemic [40]. Although services provided on company premises still predominate (66.5%), hybrid and remote models already represent 33.5% of activities [2]. Barrero et al. [7] suggest this is a trend, with the prediction that at least one working day per week will be conducted remotely in the coming years. Likewise, educational institutions are also adapting to new circumstances. The growing acceptance of distance learning (DL) is evidenced by the 166.4% increase in course enrollments in these modalities between 2015 and 2021, according to the Higher Education Map in Brazil [19]. On the other hand, enrollment in face-to-face courses decreased by 20.6%, indicating a significant change in the Brazilian educational landscape with an increasing focus on DL.

However, these models accompany a recurring challenge that negatively impacts both school [12, 57] and work performance [6]: maintaining attentional focus. Nakayama et al. [37] and Wang et al. [57] point out digital distractions - such as messages, notifications, and social networks, among others - along with the multitasking environment, as the main factors that harm our ability to maintain focus on essential tasks. This occurs because our brain cannot process many perceptual stimuli simultaneously, depending on the complex cognitive process of "attention" to select relevant information and discard irrelevant information [34, 52].

With the growing demand for remote and hybrid activities, seeking effective strategies to mitigate distractions and promote greater concentration becomes increasingly important. One approach adopted is attentional analysis based on eye tracking data because, according to Posner [42], there is a direct relationship between eye movement and changes in attention. Thus, this research aims to **synthesize and organize existing knowledge about attentional analysis conducted based on data obtained via webcam** - a low-cost eye movement capture device, with possibilities for large-scale study [9]. To do this, we carried out a literature review from 2013 to 2023, using a hybrid search strategy, through which we selected 57 papers. These papers were analyzed and categorized, offering a comprehensive overview of application domains, data, and techniques, identifying research gaps, and pointing out possible future directions. The main contributions of this work are:

- Presentation of the state of the art related to the topic, identifying features that enable attentional analysis;
- Development and availability of the tool "*AttentionVis Browser*" that presents this study's results in a visual and interactive format, as demonstrated in Figure 1.
- Presentation of identified lessons learned, allowing researchers to identify promising areas, avoid errors, and contribute to the development of this field;
- Presentation of a set of limitations of current solutions;
- Identification of research opportunities on attentional analysis.

The following sections present the methodological process, analyze the results, and discuss lessons learned, limitations, and research opportunities.

2 METHODOLOGY

For the literature review, we adopted a hybrid search strategy based on Mourão et al. [36] guidelines to ensure greater efficiency in retrieving relevant studies, as corroborated by Wohlin et al. [60]. The process involves preparing a search string based on the proposed research questions, conducting a database search for studies in a single digital library, applying inclusion and exclusion criteria, and then using the selected studies as a seed set for applying the Backward Snowballing (BS) and Forward Snowballing (FS) techniques.

The BS technique involves reviewing all references of papers selected in the seed set to identify additional studies relevant to the research. The FS seeks to identify new studies referencing the publications that make up the seed set [59]. The previously defined selection criteria must be strictly followed when analyzing these new papers. In this hybrid search strategy, papers obtained via BS are not submitted to FS, and vice versa, avoiding overlaps as indicated by Mourão et al. [36] and Wohlin et al. [60]. The papers are analyzed for data extraction and synthesis at the end of the process. Figure 2 summarizes the steps that are detailed below.

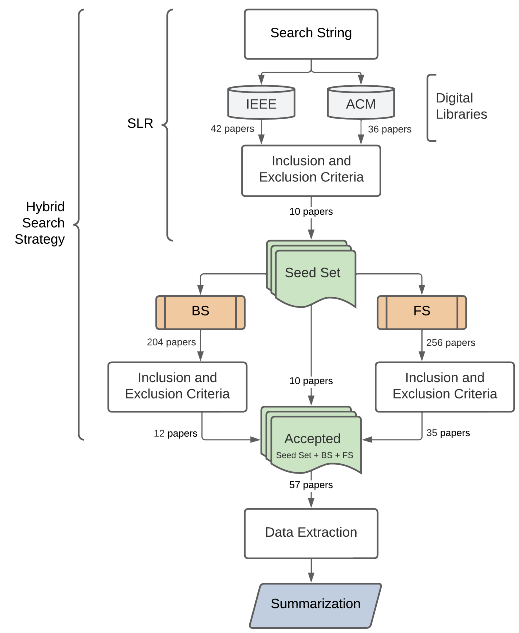

Figure 2: Selection process carried out using the hybrid search strategy.

As a way of guiding this study, employing the PICO (Population, Intervention, Comparison, Outcome) [27] criteria, we defined the following questions that this research aims to answer:

- **RQ1:** What are the most prominent areas of concentration in attention and user behavior studies using webcam data, and what practical applications are derived from?
- **RQ2:** What information can be obtained through a webcam while a user performs their tasks, and which features are relevant for behavioral analysis?
- **RQ3:** What insights can be obtained from analyzing data to identify user attention and behavior patterns, and how are these presented?
- **RQ4:** How are data visualization techniques applied to convey the insights resulting from this analysis clearly?

Although Mourão [36] suggests the use of a single digital library, we expanded our searches to include the IEEE Xplore Digital Library[1] and the ACM Digital Library[2], due to their significance in the computing area and the relevance of their publications. Furthermore, these libraries include publications from journals and

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://dl.acm.org

conference proceedings that are relevant to our research, such as ETRA[3], CVPR[4], and ICMI-MLMI[5]

After defining the search databases, we identified the most relevant terms to compose a search string to return papers related to our research topic. The resulting string, composed of the terms ((“eye track*” OR “gaze" AND (“attention” OR “attentiveness” OR “engagement” OR “monitoring” OR “behavior”) AND (“webcam”)), was applied exclusively in the abstract field. When using the search string in the selected databases, 78 studies were identified. The inclusion (IC) and exclusion (EC) criteria expressed in Table 1 were applied to keep only relevant works.

The initial stage of the filtering process eliminated studies that did not meet the basic criteria, such as IC1, IC2, EC1, EC2, and EC4. We only considered the last ten years of research to analyze the current context. The resulting papers were analyzed based on the title to provide a basis for removing duplicate works (EC3). For a better understanding of the focus of each study and the application of the EC5, EC6, EC7, EC8, EC9, and EC10 criteria, we used the Three-pass [25] method. As a result, we arrived at a seed set consisting of ten papers.

Table 1: Inclusion and exclusion criteria.

Criteria	Description
Inclusion criteria	
IC1	Published between January 2013 and October 2023.
IC2	Should be published in a conference, workshop, or journal.
Exclusion criteria	
EC1	Full-text is not available online.
EC2	The study that is not written in English.
EC3	Duplicated studies returned by different search engines.
EC4	Books, book chapters, abstracts, and gray literature.
EC5	The studies that are reviews or mapping studies.
EC6	Do not use data from webcams as the primary data source.
EC7	Do not provide user attention, behavior, or engagement insights.
EC8	Studies focused on comparative analysis (algorithms, methods, or devices)
EC9	Focused on the emotions or sentiment analysis.
EC10	The study does not answer at least one research question.

Following the hybrid search strategy, the next stage of the process involved applying the BS technique to the papers that make up the seed set. From the 204 references identified, the selection processes were applied as in the previous stage, resulting in 12 accepted papers. Afterward, we used the FS technique to deepen our analysis further. As suggested by Mourão et al. [36], Google *Scholar*[6] was used to search for studies that cite the papers contained in the seed set. In this process, carried out in November 2023, 256 studies were identified. In the same way, as in the BS, new studies were only accepted if they met the selection criteria. Thus, 35 more papers were identified.

Applying the hybrid search strategy in this literature review resulted in the selection of 57 relevant papers. We analyzed them thoroughly, observing aspects such as the application domain, relevant features, Machine Learning (ML) algorithms and methods, datasets, and visualization techniques. This data was collected and organized in an electronic spreadsheet. We then summarize this information to provide an overview of the state of the art based on the directions presented in the research questions.

3 ANALYSIS OF THE RESULTS

This section analyzes the results obtained from the previous step. This information serves as a basis for answering the research questions, identifying the contributions and limitations of studies, and suggesting research opportunities.

3.1 Research Topics Overview

The analysis shows a significant increase in publications on the research topic from 2021 onwards, as illustrated in Figure 3. The growing interest in investigating webcam data to analyze user attention in digital environments can be attributed to the impacts of the COVID-19 pandemic. The need to adapt to a new reality, marked by the widespread adoption of remote work and distance learning, brought challenges directly related to maintaining focus [37]. Additionally, the wide availability of webcams integrated with laptops and mobile devices and their capability to serve as low-cost eye trackers likely contribute to this investigation.

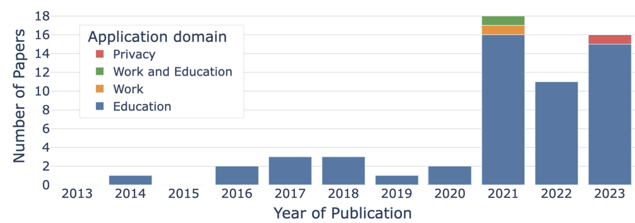

Figure 3: Number of papers per year, categorized by domain application.

The Figure 3, in addition to illustrating the number of papers per year of publication, also classifies them into four areas of study: **Education**, **Work**, **Privacy**, and the intersection of **Work and Education**. Among them, Education stands out with 95% of the publications, demonstrating the importance of this area for society and the academic interest in mitigating attentional challenges related to the learning context. The intersection of Work and Education is only explored in the study by Ozgen et al. [38], which analyzes cheating behaviors in job interviews and online exams. Likewise, it should be observed that the areas related to Work [39] and Privacy [49] appear in just one post each publication, suggesting a minimal representation of these topics.

Online activities represent 91% of studies, reflecting the adaptation to digital transformations, including remote classes, online assessments, collaborative games, and digital documents. The remaining 9% are directed to face-to-face interactions, even though in computing environments, such as computer labs [48]. Thus, four study topics stand out:

(1) **Cheating detection** (37 papers). Involve user analysis during online educational assessments or selection processes, preventing fraud;
(2) **Attention monitoring** (11 papers). Focused on observing and analyzing how users direct their attention during specific activities;

[3]Eye Tracking Research and Applications
[4]Computer Vision and Pattern Recognition
[5]Multimodal Interfaces and Machine Learning for Multimodal Interaction
[6]https://scholar.google.com

(3) **Engagement level** (8 papers). Involve aspects such as interest, motivation, emotions, and active participation during task performance;
(4) **User privacy** (1 paper). Based on observing where and how users direct their attention, including time spent in specific areas, to infer their preferences.

3.2 Webcam Feature Extraction

Data extracted from videos are essential for identifying patterns related to human behavior. Therefore, capturing videos with adequate resolution and frame rate is essential, ensuring data quality and, consequently, accuracy and effectiveness in subsequent analysis [29]. Through these data, it is possible to extract features such as **Facial expressions** (analyzing facial landmarks, such as eyes, mouth, nose, chin, or any relevant structure [41]), **Head position**, **Eye movements**, **Body pose**, and **Objects** present in the environment. By analyzing them - individually or combined - it is possible to understand many details about the user, from how they are positioned and where they are looking to what their facial expressions can reveal about their reactions or emotional state. Figure 4 presents the features explored by the selected studies.

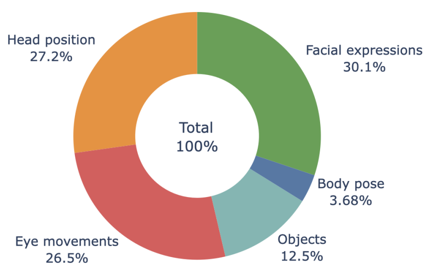

Figure 4: Webcam features.

Image processing and computer vision techniques extract this information from each frame. There are several tools for this purpose; however, the most cited were: OpenCV (in 32% of papers), Dlib (26%), and Yolo (15%). One of the purposes of these extracted features is to provide relevant information to ML algorithms, allowing them to learn and perform tasks such as classification, pattern recognition, and prediction.

3.3 Used Datasets

We explored the datasets used to identify the main sources of data available to investigate aspects such as engagement, attention, emotional state, and related factors. Although many studies do not provide clear information, we identified that 38% of the studies conducted chose to create their own datasets. This approach offers greater control over characteristics and metrics but requires a long labeling process to categorize the data correctly.

The datasets made available on public platforms or shared by the scientific community represent a valuable resource for advancing research in the area. These datasets (shown in Table 2), composed of images or videos, facilitate immediate study application. Furthermore, because they are already labeled, it significantly reduces the effort required to manually categorize data, optimizing researchers' time and increasing efficiency in conducting experiments.

Table 2: Open-source datasets and some use cases

Dataset	Purpose	Use case
PAFE [29]	Predicting attention and mind-wandering	[29]
LFW [18]	Facial recognition for cheat detection	[51] [44]
AFLW2000-3D [64]	Facial landmarks for exam integrity	[20]
EngageWild [24]	Engagement detection	[24] [63]
OEP [5]	Webcam and wearcam data for cheat detection	[5] [62]
FER-2013 [15]	Emotion recognition in proctoring systems	[56]
ImageNet [46]	Object detection	[53]
COCO [31]	Person and/or object detection	[56] [38] [39]
Pandora [8]	Head pose, and shoulder estimation	[17]
CASIA-WebFace [61]	Facial verification and identification	[43]
BIWI [14]	Head pose estimation	[17]

3.4 Machine Learning

In the reviewed literature, we observed a variety of ML approaches adopted for the analysis of attention, which reflects the diversity of techniques and strategies used. Thus, the following emerge: **Convolutional Neural Network (CNN)**, **Support Vector Machine (SVM)**, **Deep Neural Network (DNN)**, **Random Forest (RF)**, **Recurrent Neural Network (RNN)**, and **Decision Tree (DT)**. The first two algorithms mentioned are the most used for attentional feature analysis, with 21 and 12 studies, respectively. It is necessary to emphasize that seven papers do not specify the algorithm used, limiting themselves to using the general term "Machine Learning" as the applied approach. The most used algorithms are illustrated in Figure 5, and those less conventional, mentioned in only two words or less, are not included in the graph.

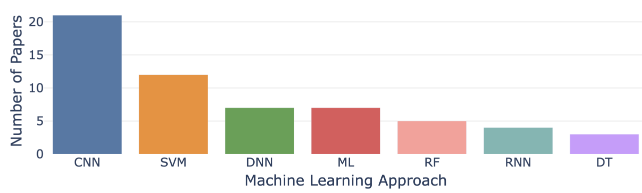

Figure 5: Main ML algorithms used by the studies.

We also identified that four studies do not mention the use of ML in their methods, which suggests a possible application of more traditional image processing techniques or statistical methods. On the other hand, 16 studies used multiple algorithms to compare their performance, emphasizing that the effectiveness of ML algorithms varies with the approach, features, and context. This makes it difficult to claim one algorithm as universally superior.

3.5 Combination with Multimodal Data

In addition to the analyses carried out using webcam features, an interesting approach to complementing and enhancing the results is the combination with multimodal data. This approach is particularly used in cheating detection outcomes and was considered in 47% of studies. Figure 6 presents the data sources identified in these studies: **Microphone**, **Screenshot/screenshare**, **System logs**, **Eye tracker**, **Keyboard**, **Mouse**, and **Form**. Among them, the most used device was the microphone attached to webcams (30%), which allows voice recognition, external noises, and parallel discussions. Additionally, 26% of papers use screenshots or screenshare features. The first one can be used to identify the active window [20] and screen sharing to support teachers and supervisors during the exams [35]. System logs are considered in 15% of the papers and can come from learning platforms [4], chats [13] and operational systems [45].

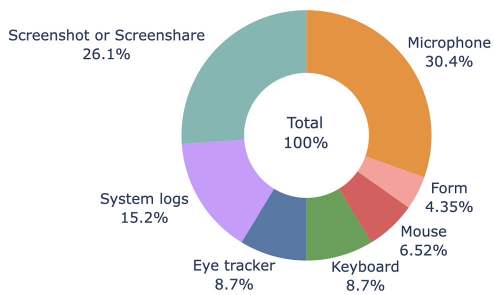

Figure 6: Multimodal data sources.

3.6 Outcome Data

The outputs derived from the analysis of attentional data can be categorized in two ways. The first approach is more simplified, using binary outputs that classify the user's states directly, such as attentive or inattentive/distracted [26, 32, 48]; focused or not focused [29]; and, cheating or not [5]. The second way involves additional information, expressed in intensity levels, such as sleeping, drowsy or awake [54]; or "Not engaged at all", "Nominally engaged", "Engaged in task", and "Very engaged" [24, 58, 63]. This detailed approach offers more opportunities to analyze the individual's behavior and/or emotional state, allowing for an in-depth analysis.

The way these data are displayed can change depending on the purpose of the analysis. Studies such as those by Shata et al. [50] and Jadi [21] choose to display warning messages on the screen to alert about potential cheating during exams. Ozgen et al. [39] use labels to do it in this same topic. These approaches are useful for quick interpretation without presenting the details or basis for such information. On the other hand, Li et al. [30]'s study uses data visualization techniques to help synthesize the different available outputs into more understandable formats.

3.7 Visualization Techniques

To understand how authors present the complex information related to attention analysis - from input and processing to data output - we explore how different studies employ data visualization and examine their application methods. Thirteen different visualization techniques were used to express some information. Line graphs, typically used to represent time series data, were the most used, appearing in 23 studies. Next, histograms, used for distribution data, were found in 16 studies. The confusion matrix, applied in 10 studies, indicates the assertiveness levels of ML models.

The purposes for using visualizations in the selected studies were classified into six main categories: **Pattern Recognition**, **Insights**, **Evaluation Metrics**, **Addicional Details**, **Comparison**, and **Other purposes**. Each reflects a distinct set of goals, as seen in Table 3.

Table 3: Purpose of data visualization usage in selected studies.

Category	Purpose of Visualizations	Total papers
Pattern Recognition	Head pose	1
	Facial expressions	4
	Multimodal data	3
	Eye movements or gaze directions	8
Insights	Show results to the user	8
	Attention, engagement or behavior level	7
	Cheating behaviors probability	4
Evaluation metrics	Accuracy	11
	Precision	3
	Recall	1
	F-1 score	1
Additional details	Distribution of data collected	7
	Results of interviews	5
Comparison	Comparative analysis of ML classifiers	9
	Data-driven experiment analysis	3
Other purposes	Correlation between features	2
	Outliers	2
	System architecture	1

While data visualization is a powerful tool for effectively communicating insights derived from data analysis [22], Table 3 underscores that its use is mainly associated with aspects that demonstrate, compare, and analyze ML models. The objective of using these visual resources is primarily to offer support and clarity to the paper's reader, thus synthesizing, in a graphic form, complex information related to the analyses. The use of visual techniques to demonstrate results to the end user of the proposed solution is found in only eight papers.

4 INTERACTIVE VISUAL SURVEY

Based on the *TextVis Browser* project, developed by Kostiantyn et al. [28], we propose an interactive visual survey of attentional data analysis, called *AttentionVis Browser*, available at:

https://davintlab.github.io/AttentionVis-Browser

This tool, developed in HTML and JavaScript, was designed to quickly and intuitively summarize and present the results of this study, including only those papers that present visualization techniques (43 papers). Figure 1 illustrates the user interface, consisting of a main and interaction panels.

In the main panel, thumbnails of the visualization techniques representing each paper are organized in a grid format. They are ordered by publication year (in descending order) and then by the primary author's surname. Clicking on a specific thumbnail displays details of the selected study, including the complete bibliographic reference, a URL link to access the full paper, a BiBTeX file link, and a list of categories assigned, as illustrated in Figure 7.

Figure 7: Details of a survey entry.

The interaction panel allows the user to filter the content displayed on the main panel, selecting specific works through textual search or restricting the results by year of publication or category. A summarized view of the papers can be consulted using the menu option “Summary”. Additionally, we provide a form for authors who wish to contribute with additional entries. The information will then be verified and added.

5 DISCUSSION

This section presents the research questions and their answers, allowing insights into the topic. After that, lessons learned and limitations of the research are also presented.

5.1 Research Questions

Next, we discuss the main findings of our literature review, organized according to the research questions.

RQ1:*“What are the most prominent areas of concentration in attention and user behavior studies using webcam data, and what practical applications are derived from?”*

We found that 95% of the publications are concentrated in the Education area, highlighting its importance in research on attentional focus in digital environments. Although in smaller numbers, other areas were also identified, such as Work, with two publications, and Privacy, with one publication. The data indicate that user attention can be explored as a central point or part of a broader context. The greatest focus is on cheating detection systems (63%), aiming to ensure integrity in educational [5, 30] and work contexts [39]). Other applications aim to improve the educational process, such as classes [47] and teaching materials [47, 48]). They also address issues such as mind wandering [29], tiredness [1], and drowsiness [54] during the execution of activities.

RQ2:*“What information can be obtained through a webcam while a user performs their tasks, and which features are relevant for behavioral analysis?”*

Through a webcam, it is possible to extract several crucial pieces of information about the user's behavior while performing tasks. This includes **facial expressions**, which encompass the movement of the lips, eyebrows, and other aspects to understand the user's emotions and reactions; **eye movements**, which allow identifying areas of visual focus through coordinates; the assessment of **body posture** and the observation of **head movement** (roll, pitch and yaw); and the detection of **objects** in the environment. These features play a fundamental role in understanding the user's behavioral patterns. They can be analyzed in isolation or combination, allowing the development of solutions related to attentional focus, engagement, and emotional states.

RQ3:*“What insights can be obtained from analyzing data to identify user attention and behavior patterns, and how are these presented?”*

The analysis of data obtained through a webcam, using ML algorithms, makes it possible to identify patterns, trends, and behaviors of users while interacting in digital environments. This analysis can reveal information about the occurrence of cheating [5], levels of attention [26] and focus [29], indicators of distraction [26, 48] or fatigue [54], degree of engagement [24, 58, 63] or user preferences [33]. Information is presented through alert [21], labels [35], flags [21, 50], prompt [51], tables [58], and graphs [63].

RQ4:*“How are data visualization techniques applied to convey the insights resulting from this analysis clearly?”*

Table 3, in the “insights” category, highlights how visualizations are used to present the results obtained. Visualization techniques are applied to (I) present results to the end users of the application, (II) illustrate levels of attention, engagement, and behavior, and (III) indicate the probability of cheating. Although this category represents 33% of the studies, we observed that these visualizations lack details or explanations that would help the end user to better understand the results.

5.2 Lessons Learned

This section describes the knowledge acquired while conducting the research. To this end, we created a list of lessons learned, as summarized in Figure 8 and detailed below.

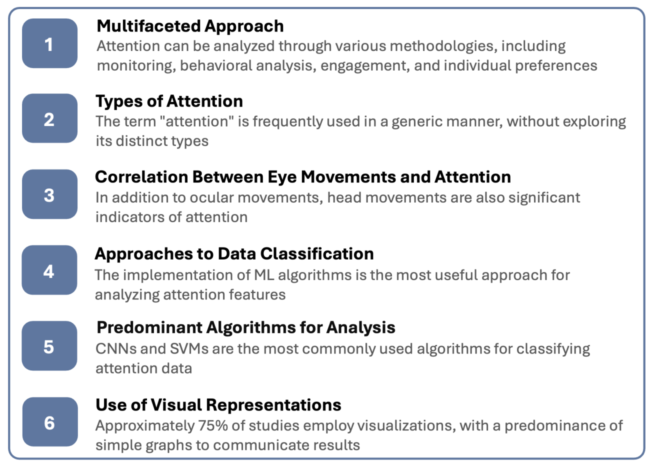

Figure 8: Summary list of Lessons Learned.

Multifaceted Approach. The literature approaches attention in different ways, depending on the objective of the study: it can be used to identify inappropriate behaviors during assessment activities, analyze the level of engagement during studies, help educators in decision-making, or even discover certain user preferences by analyzing the direction of their gaze while filling out a simple online form. .

Types of attention. The term “attention” is broad without delving into its different types. Even so, the evidence presented by Tzeng et al. [55], for example, shows how eye patterns differ depending on the type of task being performed. This can contribute to investigations involving the concept of “divided attention” (focusing on different tasks simultaneously) and “sustained attention” (prolonged periods without distractions). Likewise, the various approaches to identifying cheating, as presented by Irfan et al. [20], contribute to the analysis of “alternating attention” by enabling the identification of external noises, parallel conversations, and head movements, evidencing the change in focus.

Correlation between Eye Movements and Attention. Despite the relationship between eye movements and attention described by Posner [42], this feature is not essential for attentional predictions and classifications. The approaches adopted by Alyuz et al. [3], Irfan et al. [20], and Cote et al. [11], for example, consider only the user's head movements to indicate their attentional focus, obtaining satisfactory results within their objectives.
Approaches to Data Classification. The selected works demonstrated the importance of ML for the analysis of data extracted from a webcam since 89% of the presented solutions clearly express its use. Khan et al.[26] demonstrated the effectiveness of machine learning models in automatically classifying attention using eye-tracking metrics. Furthermore, Seiden et al. [49] emphasize the accuracy of the algorithms in predicting the location of the visual focus on the screen, which is essential for understanding where the user is directing their attention.
Predominant Algorithms for Analysis. There is a predominance of certain algorithms in attention analysis, such as CNNs, which have been widely used to identify patterns in data from webcams. This popularity is due to the effectiveness of CNNs in classifying audio and video data [16], which makes them especially suitable for analyzing eye movements and facial expressions. Another approach that stands out is using SVMs, especially in contexts involving smaller-scale datasets.
Use of Visual Representations. Approximately 75% of the studies use visualization techniques to elucidate information or results. One trend observed is using these resources to express data related to ML processes, such as metrics, comparisons, the influence of features, and model performance.

6 LIMITATIONS AND RESEARCH OPPORTUNITIES

This section presents the limitations of this work regarding the reviewed studies and the research conducted. These limitations, in turn, represent opportunities that can be investigated in future research and are also presented.

The identified **limitations regarding the reviewed studies** are described below.

Limited scope. The studies are predominantly focused on Education, highlighting the importance of broadening the scope and encompassing professional environments, which is equally important in the current context. In addition, the analyses are restricted to a single type of task, excluding the possibility of simultaneously monitoring the development of other types of activities.
Aspects related to attention. When there is a change of focus to an activity outside the investigated scope, this transition is seen only as a distraction or, in certain contexts, as cheating. This opens the opportunity to identify whether this change is, in fact, a distraction or whether the user chose to redirect his attention to another activity relevant to his work or study, characterizing 'alternating attention' [23].
Detailing of methods for analyzing and extracting patterns. The replicability of the proposed solutions is compromised by the lack of detailing of the versions of the tools used, especially for the extraction of features (only 19% of the studies provide this essential information). This makes it difficult for other researchers to identify, analyze, and reproduce the results. In addition, there are challenges in identifying methods and algorithms due to the lack of such information.
Analyses Report. In the studies reviewed, we often found generic descriptions of attention monitoring, such as normal or abnormal behavior, attentive or distracted state, and the possibility of cheating. Sometimes, data are communicated only through labels or flags, hindering users' understanding.
Uninformative visualizations. The limitations of the visualizations lie in the lack of details about the data presented, the direction of focus at each moment, and the precise definition of what is considered a distraction. Only in the studies by Li et al. [30] and Ozgen et al. [38] do we find a more detailed approach (focused on detecting cheating). The user must be aware of the periods of distraction and the reasons associated with these moments, as this can help them identify behavioral patterns and implement strategies to mitigate these distractions, promoting self-regulation.

The identified **limitations regarding the conducted research** are listed below.

Specific devices for data capture. This research builds on studies primarily using webcams to collect visual data. While these are widely available and widely used, it is important to note that alternative devices, such as electroencephalography (EEG) and commercial eye trackers (Tobii[7], and SR Research[8], for example), offer more accurate and detailed measurements.
Comprehensive approach. The broad approach adopted in our work offers global understanding and contextualization advantages but may lack detailed depth on specific topics.
Focus on solutions and applications. This review is defined by the exclusion criteria presented in Table 1, which delimit the scope due to the large volume of related works. The main focus is on practical solutions, excluding comparative studies (between devices, methods, and techniques), wearable devices, and EEG, among others.

Throughout our investigation, we identified areas that are under-explored or not addressed by current studies. Thus, some **research opportunities** are described below.

We suggest the development of more comprehensive solutions, expanding attention analysis beyond the educational scope; the development of evaluation frameworks that assist in measuring the effectiveness and usability of these tools; the application of narrative visualizations in the results obtained through the ML classifiers on attentional analysis, aiming at a more intuitive communication, facilitating understanding by the user and providing subsidies that contribute to self-regulation. Furthermore, we consider promising the classification of different "types of attention" during data analysis, such as transitions of focus between different activities, distinguishing distractions from deliberate choices, and characterizing "alternating attention", for example.

7 FINAL REMARKS

The research presented here aimed to synthesize and organize the existing knowledge related to data analysis, especially data captures

[7] http://www.tobii.com
[8] http://www.sr-research.com

from webcams. In this sense, it offers a comprehensive view of application domains, data, features, techniques, trends, gaps, and opportunities. Finally, as our investigation's result, we identified five **contributions**:

State of the art mapping. We explore a decade of related studies providing an overview of the state of the art related to attentional analysis using webcam data.

AttentionVis Browser. The interactive visual tool built upon the review results offers an overview of the field. It can be used by the general community (for educational purposes) and the scientific community (to aid in searching for related works and extend it).

Presenting lessons learned. The lessons learned highlight the knowledge acquired with this study and aim to provide researchers with support for the continued development of the topic. Pondering on these lessons allows for adjusting and improving methods and processes, avoiding repeating errors, and supporting new hypotheses for future investigation.

Presenting limitations research. Examining the limitations identified in current studies and our own research, we aim to encourage further investigations in areas that require better development.

Presenting research opportunities. The identified opportunities can inspire the exploration of new directions, approaches, and solutions to the challenges that permeate the analysis of attention.

ACKNOWLEDGMENTS

This paper was supported by the Ministry of Science, Technology, and Innovations, with resources from Law No. 8.248, dated October 23, 1991, within the scope of PPI-SOFTEX, coordinated by Softex. Isabel H. Manssour would like to thank the financial support of the CNPq Scholarship - Brazil (303208/2023-6).

REFERENCES

[1] Andrea F. Abate, Lucia Cascone, Michele Nappi, Fabio Narducci, and Ignazio Passero. 2021. Attention monitoring for synchronous distance learning. *Future Generation Computer Systems* 125 (2021), 774–784. https://doi.org/10.1016/j.future.2021.07.026

[2] Cevat Giray Aksoy, Jose Maria Barrero, Nicholas Bloom, Steven J Davis, Mathias Dolls, and Pablo Zarate. 2023. *Working from home around the globe: 2023 Report.* Technical Report. EconPol Policy Brief.

[3] Nese Alyuz, Eda Okur, Utku Genc, Sinem Aslan, Cagri Tanriover, and Asli Arslan Esme. 2017. An unobtrusive and multimodal approach for behavioral engagement detection of students. In *Proceedings of the 1st International Workshop on Multimodal Interaction for Education.* ACM, Glasgow, UK, 26–32. https://doi.org/10.1145/3139513.3139521

[4] Sinem Aslan, Nese Alyuz, Cagri Tanriover, Sinem E. Mete, Eda Okur, Sidney K. D'Mello, and Asli Arslan Esme. 2019. Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, Glasgow, Uk, 1–12. https://doi.org/10.1145/3290605.3300534

[5] Yousef Atoum, Liping Chen, Alex X. Liu, Stephen D. H. Hsu, and Xiaoming Liu. 2017. Automated Online Exam Proctoring. *IEEE Transactions on Multimedia* 19, 7 (July 2017), 1609–1624. https://doi.org/10.1109/TMM.2017.2656064

[6] Charles E. Bailey. 2007. Cognitive Accuracy and Intelligent Executive Function in the Brain and in Business. *Annals of the New York Academy of Sciences* 1118, 1 (2007), 122–141. https://doi.org/10.1196/annals.1412.011

[7] Jose Maria Barrero, Nicholas Bloom, and Steven J Davis. 2021. *Internet access and its implications for productivity, inequality, and resilience.* Technical Report. National Bureau of Economic Research.

[8] Guido Borghi, Marco Venturelli, Roberto Vezzani, and Rita Cucchiara. 2017. Poseidon: Face-from-depth for driver pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society, Los Alamitos, US, 5494–5503. https://doi.org/10.48550/arXiv.1611.10195

[9] Benjamin T. Carter and Steven G. Luke. 2020. Best practices in eye tracking research. *International Journal of Psychophysiology* 155 (2020), 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

[10] Comitê Gestor da Internet no Brasil. 2022. *Pesquisa sobre o uso das tecnologias de informação e comunicação nas escolas brasileiras - TIC Educação 2021* (1 ed.). Núcleo de Informação e Coordenação do Ponto BR, São Paulo, BR.

[11] Melissa Cote, Frederic Jean, Alexandra Branzan Albu, and David Capson. 2016. Video summarization for remote invigilation of online exams. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision.* IEEE, Lake Placid, US, 1–9. https://doi.org/10.1109/WACV.2016.7477704

[12] Adele Diamond and Kathleen Lee. 2011. Interventions Shown to Aid Executive Function Development in Children 4 to 12 Years Old. *Science* 333, 6045 (2011), 959–964. https://doi.org/10.1126/science.1204529

[13] Fahmid Morshed Fahid, Seung Lee, Bradford Mott, Jessica Vandenberg, Halim Acosta, Thomas Brush, Krista Glazewski, Cindy Hmelo-Silver, and James Lester. 2023. Effects of Modalities in Detecting Behavioral Engagement in Collaborative Game-Based Learning. In *Proceedings of the 13th International Learning Analytics and Knowledge Conference.* ACM, Arlington, US, 208–218. https://doi.org/10.1145/3576050.3576079

[14] Gabriele Fanelli, Thibaut Weise, Juergen Gall, and Luc Van Gool. 2011. Real Time Head Pose Estimation from Consumer Depth Cameras. *Pattern Recognition* 6835 (2011), 101–110. https://doi.org/10.1109/3DV.2014.54

[15] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, Yingbo Zhou, Chetan Ramaiah, Fangxiang Feng, Rui Li, Xiaojie Wang, Dimitris Athanasakis, John Shawe-Taylor, Maximilian Milakov, John Park, Radu Ionescu, Marius Popescu, Cristian Grozea, James Bergstra, Jingjing Xie, Lukasz Romaszko, Bing Xu, Zhang Chuang, and Yoshua Bengio. 2015. Challenges in representation learning: A report on three machine learning contests. *Neural Networks* 64 (2015), 59–63. https://doi.org/10.1016/j.neunet.2014.09.005

[16] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. 2017. CNN architectures for large-scale audio classification. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing.* IEEE, IEEE Press, New Orleans, US, 131–135. https://doi.org/10.48550/arXiv.1609.09430

[17] Basavaraj N Hiremath, Anushree Mitra, Aman Thapa, S Amoolya, and A Tameem. 2023. Proctoring using Artificial Intelligence. *SSRN Electronic Journal* (2023). https://doi.org/10.2139/ssrn.4411332

[18] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. 2007. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments.* Technical Report 07-49. University of Massachusetts, Amherst.

[19] Instituto Semesp. 2023. *Mapa do Ensino Superior no Brasil.* Semesp, São Paulo, BR. https://www.semesp.org.br/mapa/edicao-13/download/ 13ª ed.

[20] Mohamed Irfan, Mohammed Aslam, Ziyan Maraikar, Upul Jayasinghe, and Mohamed Fawzan. 2021. Ensuring Academic Integrity of Online Examinations. In *Proceedings of the IEEE 16th International Conference on Industrial and Information Systems.* IEEE, Kandy, LK, 295–300. https://doi.org/10.1109/ICIIS53135.2021.9660737

[21] Amr Jadi. 2021. New Detection Cheating Method of Online-Exams during COVID-19 Pandemic. *International Journal of Computer Science and Network Security* 21, 4 (April 2021), 123–130. https://doi.org/10.22937/IJCSNS.2021.21.4.17

[22] Marcos Kalinowski, Tatiana Escovedo, Hugo Villamizar, and Hélio Lopes. 2023. *Engenharia de Software para Ciência de Dados: Um guia de boas práticas com ênfase na construção de sistemas de Machine Learning em Python.* Casa do Código, São Paulo, BR.

[23] Virgínia Kastrup. 2007. O funcionamento da atenção no trabalho do cartógrafo. *Psicologia e Sociedade* 19, 1 (Jan 2007), 15–22. https://doi.org/10.1590/S0102-71822007000100003

[24] Amanjot Kaur, Aamir Mustafa, Love Mehta, and Abhinav Dhall. 2018. Prediction and Localization of Student Engagement in the Wild. In *Proceedings of the Digital Image Computing: Techniques and Applications.* IEEE, Canberra, AU, 1–8. https://doi.org/10.48550/arXiv.1804.00858

[25] S. Keshav. 2007. How to Read a Paper. *Special Interest Group on Data Communication da Association for Computing Machinery* 37, 3 (jul 2007), 83–84. https://doi.org/10.1145/1273445.1273458

[26] Ahsan Raza Khan, Sara Khosravi, Sajjad Hussain, Rami Ghannam, Ahmed Zoha, and Muhammad Ali Imran. 2022. EXECUTE: Exploring Eye Tracking to Support E-learning. In *Proceedings of the IEEE Global Engineering Education Conference.* IEEE, Tunis, TN, 670–676. https://doi.org/10.1109/EDUCON52537.2022.9766506

[27] Barbara Kitchenham and Stuart Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering.* Technical Report EBSE 2007-001. University of Durham.

[28] Kostiantyn Kucher and Andreas Kerren. 2015. Text visualization techniques: Taxonomy, visual survey, and community insights. In *Proceedings of the IEEE Pacific Visualization Symposium.* IEEE, Hangzhou, CN, 117–121. https://doi.org/10.1109/PACIFICVIS.2015.7156366

[29] Taeckyung Lee, Dain Kim, Sooyoung Park, Dongwhi Kim, and Sung-Ju Lee. 2022. Predicting Mind-Wandering with Facial Videos in Online Lectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops.* IEEE, New Orleans, US, 2103–2112. https://doi.org/10.1109/CVPRW56347.2022.

00228
[30] Haotian Li, Min Xu, Yong Wang, Huan Wei, and Huamin Qu. 2021. A visual analytics approach to facilitate the proctoring of online exams. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, New York, US, 1–17. https://doi.org/10.1145/3411764.3445294
[31] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Proceedings of the 2014 Conference on Computer Vision*, Vol. 8693. Springer, Zurich, CH, 740–755. https://doi.org/10.48550/arXiv.1405.0312
[32] Jens Madsen, Sara U. Júlio, Pawel J. Gucik, Richard Steinberg, and Lucas C. Parra. 2021. Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences* 118, 5 (2021). https://doi.org/10.1073/pnas.2016980118
[33] Rachna Maithani. 2023. Securelens: Enhancing Trustworthiness in Remote Exams with Opencv Protective Eye. *Journal of Modernization in Engineering Technology and Science* 05, 6 (2023), 1857–1861.
[34] Margaret W Matlin. 2004. *Psicologia cognitiva, 5ª ed.* LTC, Rio de Janeiro, BR.
[35] Samuel Monteiro, Rutuja Bhate, Lav Sharma, and Phiroj Shaikh. 2022. Proct-Xam – AI Based Proctoring. In *Proceedings of the 2nd Asian Conference on Innovation in Technology*. IEEE, Ravet, IN, 1–6. https://doi.org/10.1109/ASIANCON55314.2022.9908817
[36] Erica Mourão, Marcos Kalinowski, Leonardo Murta, Emilia Mendes, and Claes Wohlin. 2017. Investigating the Use of a Hybrid Search Strategy for Systematic Reviews. In *Proceedings of the 2017 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, Toronto, CA, 193–198. https://doi.org/10.1109/ESEM.2017.30
[37] Makoto Nakayama and Charlie C Chen. 2022. Digital Distractions and Remote Work: A Balancing Act at Home. *Information Resources Management Journal* 35, 1 (2022), 1–17. https://doi.org/10.4018/IRMJ.308675
[38] Azmi Can Ozgen, Mahiye Uluyagmur Ozturk, and Umut Bayraktar. 2021. Cheating Detection Pipeline for Online Interviews and Exams. *ArXiv* abs/2106.14483 (2021), 4. https://doi.org/10.48550/arXiv.2106.14483
[39] Azmi Can Ozgen, Mahiye Uluyagmur Ozturk, Orkun Torun, Jianguo Yang, and Mehmet Zahit Alparslan. 2021. Cheating Detection Pipeline for Online Interviews. In *Proceedings of the 29th Signal Processing and Communications Applications Conference*. IEEE, Istanbul, TR, 1–4. https://doi.org/10.1109/SIU53274.2021.9477950
[40] Kim Parker. 2023. About a third of U.S. workers who can work from home now do so all the time. https://www.pewresearch.org/short-reads/2023/03/30/about-a-third-of-us-workers-who-can-work-from-home-do-so-all-the-time Accessed: 2024-02-07.
[41] Leandro Persona, Fernando Meloni, and Alessandra Alaniz Macedo. 2023. An accurate real-time method to detect the smile facial expression. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web* (Ribeirão Preto, Brazil) *(WebMedia '23)*. Association for Computing Machinery, New York, USA, 46–55. https://doi.org/10.1145/3617023.3617031
[42] Michael I Posner. 1980. Orienting of attention. *Quarterly journal of experimental psychology* 32, 1 (1980), 3–25. https://doi.org/10.1080/00335558008248231
[43] Tejaswi Potluri, Venkatarama Phani Kumar, and K Venkata Krishna Kishore. 2023. An automated online proctoring system using attentive-net to assess student mischievous behavior. *Multimedia Tools and Applications* 82, 20 (Aug. 2023), 30375–30404. https://doi.org/10.1007/s11042-023-14604-w
[44] Tejaswi Potluri, Venkatarama Phani Kumar Sistla, and Drkv Krishna. 2023. Detecting autism of examinee in automated online proctoring using eye-tracking. *Journal of Theoretical and Applied Information Technology* 101, 3 (2023).
[45] Swathi Prathish, Athi Narayanan S., and Kamal Bijlani. 2016. An intelligent system for online exam monitoring. In *Proceedings of the International Conference on Information Science*. IEEE, Kochi, IN, 138–143. https://doi.org/10.1109/INFOSCI.2016.7845315
[46] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. 2015. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252. https://doi.org/10.1007/s11263-015-0816-y
[47] Andrew Sanders, Bradley Boswell, Andrew Allen, Gursimran Singh Walia, and Md Shakil Hossain. 2022. Development and Field-Testing of a Non-intrusive Classroom Attention Tracking System (NiCATS) for Tracking Student Attention in CS Classrooms. In *Proceedings of the IEEE Frontiers in Education Conference*. IEEE, Uppsala, SE, 1–9. https://doi.org/10.1109/FIE56618.2022.9962447
[48] Andrew Sanders, Bradley Boswell, Gursimran Singh Walia, and Andrew Allen. 2021. Non-Intrusive Classroom Attention Tracking System (NiCATS). In *Proceedings of the IEEE Frontiers in Education Conference*. IEEE, Lincoln, US, 1–9. https://doi.org/10.1109/FIE49875.2021.9637411
[49] Steven Seiden, Long Huang, and Chen Wang. 2023. Poster: Snooping Online Form Choice Privacy in Video Calls. In *Proceedings of the IEEE Symposium on Security and Privacy*. IEEE, San Francisco, US, 1–2.
[50] Abdelrahman Shata, Zineddine N. Haitaamar, and Abdsamad Benkrid. 2023. Baseline Estimation in Face Detection for AI Proctored Examinations through Convoluted Neural Networks. In *Proceedings of the International Conference on IT Innovation and Knowledge Discovery*. IEEE, Manama, BH, 1–5. https://doi.org/10.1109/ITIKD56332.2023.10100328
[51] Merzoug Soltane and Mohamed Ridda Laouar. 2021. A Smart System to Detect Cheating in the Online Exam. In *Proceedings of the International Conference on Information Systems and Advanced Technologies*. IEEE, Tebessa, DZ, 1–5. https://doi.org/10.1109/ICISAT54145.2021.9678418
[52] Robert J Sternberg. 2008. *Psicologia cognitiva* (4ª ed.). artmed, Rio de Janeiro, BR.
[53] P. Tejaswi, S. Venkatramaphanikumar, and K. Venkata Krishna Kishore. 2023. Proctor net: An AI framework for suspicious activity detection in online proctored examinations. *Measurement* 206 (Jan. 2023), 112266. https://doi.org/10.1016/j.measurement.2022.112266
[54] Shogo Terai, Shizuka Shirai, Mehrasa Alizadeh, Ryosuke Kawamura, Noriko Takemura, Yuki Uranishi, Haruo Takemura, and Hajime Nagahara. 2020. Detecting Learner Drowsiness Based on Facial Expressions and Head Movements in Online Courses. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*. ACM, Cagliari, IT, 124–125. https://doi.org/10.1145/3379336.3381500
[55] Jian-Wei Tzeng, Cheng-Yu Hsueh, Chia-An Lee, and Wei-Yun Shih. 2023. Identifying the Correlation Between Online Exam Answer Trajectory and Test Behavior Based on Artificial Intelligence and Eye Movement Detection Technology. In *Proceedings of the International Conference on Consumer Electronics*. IEEE, PingTung, TW, 503–504. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226745
[56] Puru Verma, Neil Malhotra, Ram Suri, and Rajesh Kumar. 2024. Automated smart artificial intelligence-based proctoring system using deep learning. *Soft Computing* 28, 4 (Feb. 2024), 3479–3489. https://doi.org/10.1007/s00500-023-08696-7
[57] Chenghao Wang et al. 2022. Comprehensively summarizing what distracts students from online learning: A literature review. *Human Behavior and Emerging Technologies* 2022, 1 (2022), 1483531. https://doi.org/10.1155/2022/1483531
[58] Jacob Whitehill, Zewelanji Serpell, Yi-Ching Lin, Aysha Foster, and Javier R. Movellan. 2014. The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions. *IEEE Transactions on Affective Computing* 5, 1 (Jan. 2014), 86–98. https://doi.org/10.1109/TAFFC.2014.2316163
[59] Claes Wohlin. 2014. Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering. In *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering* (London, England, United Kingdom) *(EASE '14)*. Association for Computing Machinery, New York, US, 10 pages. https://doi.org/10.1145/2601248.2601268
[60] Claes Wohlin, Marcos Kalinowski, Katia Romero Felizardo, and Emilia Mendes. 2022. Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology* 147 (2022), 106908. https://doi.org/10.1016/j.infsof.2022.106908
[61] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z Li. 2014. Learning face representation from scratch. *arXiv preprint* 1411.7923 (2014), 9 pages. https://doi.org/10.48550/arXiv.1411.7923
[62] Intan Nurma Yulita, Fauzan Akmal Hariz, Ino Suryana, and Anton Satria Prabuwono. 2023. Educational Innovation Faced with COVID-19: Deep Learning for Online Exam Cheating Detection. *Education Sciences* 13, 2 (Feb. 2023), 194. https://doi.org/10.3390/educsci13020194
[63] Cheng Zhang, Cheng Chang, Lei Chen, and Yang Liu. 2018. Online Privacy-Safe Engagement Tracking System. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ACM, Boulder, US, 553–554. https://doi.org/10.1145/3242969.3266295
[64] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. 2016. Face alignment across large poses: A 3D solution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, US, 146–155. https://doi.org/10.1109/CVPR.2016.23