Towards Collaborative and Intelligent Learning Environments Based on Eye Tracking Data and Learning Analytics: A Survey

YUEHUA WANG[1], **(Member, IEEE), SHULAN LU**[2], **AND DEREK HARTER**[1], **(Member, IEEE)**

[1]Department of Computer Science, Texas A&M University–Commerce, Commerce, TX 75428, USA
[2]Department of Psychology, Texas A&M University–Commerce, Commerce, TX 75428, USA

Corresponding author: Yuehua Wang (yuehua.wang@tamuc.edu)

ABSTRACT The current pandemic has significantly impacted educational practices, modifying many aspects of how and when we learn. In particular, remote learning and the use of digital platforms have greatly increased in importance. Online teaching and e-learning provide many benefits for information retention and schedule flexibility in our on-demand world while breaking down barriers caused by geographic location, physical facilities, transportation issues, or physical impediments. However, educators and researchers have noticed that students face a learning and performance decline as a result of this sudden shift to online teaching and e-learning from classrooms around the world. In this paper, we focus on reviewing eye-tracking techniques and systems, data collection and management methods, datasets, and multi-modal learning data analytics for promoting pervasive and proactive learning in educational environments. We then describe and discuss the crucial challenges and open issues of current learning environments and data learning methods. The review and discussion show the potential of transforming traditional ways of teaching and learning in the classroom, and the feasibility of adaptively driving learning processes using eye-tracking, data science, multimodal learning analytics, and artificial intelligence. These findings call for further attention and research on collaborative and intelligent learning systems, plug-and-play devices and software modules, data science, and learning analytics methods for promoting the evolution of face-to-face learning and e-learning environments and enhancing student collaboration, engagement, and success.

INDEX TERMS Artificial intelligence, attention measurement, data science, human–computer interaction, learning, multimodal learning analytics.

I. INTRODUCTION

The COVID-19 pandemic is an ongoing disruptive vent across our world. It has changed education entirely and resulted in many primary and higher education institutions switching to in-person classes in 2020 and Spring 2021. There are over 1.2 billion children out of the classroom in 186 countries and about two-thirds of colleges and universities remaining fully or primarily online [1]. Both teachers and researchers have noticed that students face a learning and performance decline as a result of this urgent and mandatory shift in instruction format [1], [2]. Actually, online teaching and distance e-learning is not new as a teaching method.

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

With advances in networking and edge computing, learning does not necessarily have to be in traditional classrooms, at schools, sitting with other classmates. Remote learning has been used and becoming more popular before the current events, but the pandemic has greatly accelerated their use. Learning activities in online settings can have many advantages. Remote learning can be self-initiated and happen in and outside of classrooms, and can be more convenient for nontraditional students and those continuing adult learning. It is a common experience that if we don't pay attention to learning materials or missing central parts of teaching lectures, the quality of learning, performance, and even retention and application of the learned knowledge can be greatly degraded. To some extent, how well learners have engaged during learning has emerged as a key indicator of whether or not
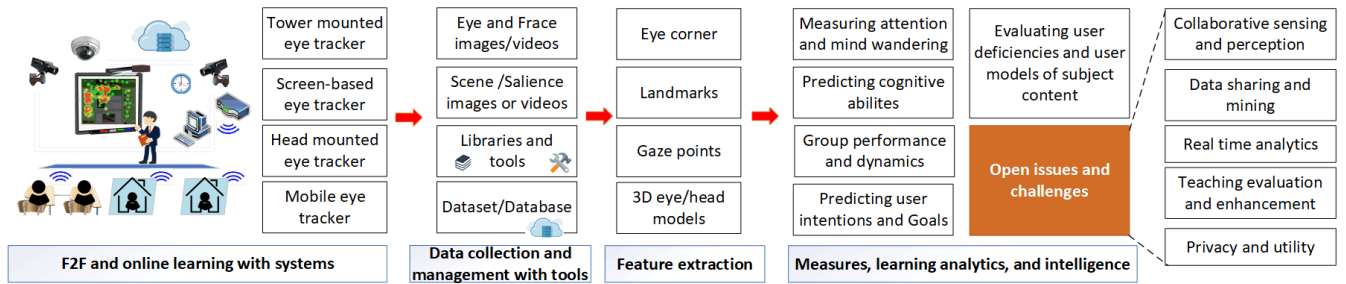
FIGURE 1. A graphical summary of eye-tracking enabled learning systems and analytics with essential components and techniques.

they are mastering the learning materials in order to achieve personal goals, accomplishment, satisfaction, as well as self-improvement. We argue that apart from learning management system software such as D2L Brightspace, Blackboard, or Moodle, emerging technologies for promoting proactive and collaborative teaching and learning of complex STEM topics like computer science have become an urgent need to ensure continuous quality learning and transform teaching and learning both in the classroom and online.

Typically, in learning environments, there is a wide variety of software and technology available to conduct instruction and learning activities. Examples include projectors, cameras, microphones, computers, and eye trackers. Eye trackers are devices, hardware platforms, or systems that capture and track the movement of the students' eyes to determine where and how long they are gazing. Essentially, the collected eye-tracking data serves as a huge potential resource that can be mined to discover new insights into when students are learning well or poorly, and how to craft teaching sessions to improve the learning experience. Learning and data analytics can be then applied to these environments to better understand, analyze, and improve the learning experience. Over the past few years, extensive research efforts [3]–[7], learning systems with eye-tracking devices [8]–[12], and learning analytics [13]–[16] have been conducted and devoted to understanding student learning and assist student learning in real-world learning environments by capturing, monitoring, estimating, and evaluating student engagement in a variety of learning tasks.

Meantime, recent studies and a few surveys [17]–[22] reviewed eye-tracking methodologies [17], major decision-making process factors [18], calibration methods, accuracy, evaluation [22], and the use of popular commercial and eye-tracking applications [19]. Particularly, Alemdag and Cagiltay [17] investigated the use of eye-tracking technology in multimedia learning. This study found that eye tracking assessments of cognitive progress in multimedia learning include selecting, organizing, and integrating information about the learning process. Orquin and Loose [18] found that saliency, surface size, visual clutter, and position are four major identified factors throughout the decision-making process. They also went through four decision theories: rational models, constrained rationality, evidence accumulation, and parallel constraint satisfaction models. Obaidellah *et al.* [20]

conducted a review of the application of eye tracking to programming tasks by describing the most prevalent eye-tracking metrics and commercial product characteristics such as sampling rate, resolution, duration, and dependent variables. This research reported that the Tobii eye-tracker is the most commonly used device. In [19], Al-Rahayfeh and Faezipour examined the use of prominent commercial products and eye tracking tools. Methods for eye tracking and head movement detection were reviewed in [19] and [21], which we have described with multi-sensor eye-tracking systems and summarized in Table 1. [22] researched the degree accuracy and pixel metrics achieved by Personal computers, TV panels, head-mounted, automotive, and head-held device platforms. Those reviews can be excellent resources for technical information, calibration methods and experiment settings, accuracy, evaluations, product specifications, and applications. The surveys on educational data classification, data fusion, and learning analytics were performed in [23] and [24].

Despite there has been extensive research on eye-tracking and learning analytics for improving student learning, they are detached, broadly spreading over the field with different focuses. There is no review in relation to fundamental resources including open projects, accessible sensing data and library/tools, or frameworks to be compatible with existing sensors, devices, and tools while solving scientific problems. The variety of eye-tracking and learning analytics, the boundary between hardware, and the dynamic nature of teaching and learning further increase the complexity of reviewing existing studies and research work in the field. To fill the research gaps and eliminate the complexity, in this paper, we review the literature for promoting pervasive, student collaboration, engagement, and success based on our previous work presented in [25] and [26]. In [25] and [26], we studied and classified eye-tracking devices and platforms into four categories with a detailed comparison. Based on these findings, the main contributions of this paper are summarized as follows:

- We present a comprehensive survey by emphasizing the existing eye-tracking enabled learning systems and analytics and combining them with data science, multimodal learning, and artificial intelligence.
- We provide a graphical summary of the current eye-tracking-enabled learning systems and analytics with

TABLE 1. Four categories of eye-tracking devices and platforms.

Categories	Setup (Camera, monitors, Shelf)	Operating conditions	Commercial devices/ platforms	Features	Typical applications	Eye tracking method
Tower-mounted eye tracker [8, 11, 27–31]	Infrared (IR) illumination camera, monitor, chin/forehead rest	Head fixed/ restricted (Laboratory setting)	SR Eyelink 1000, EyeLink2000, SMI Hi-Speed stationary, Dual Purkinje image tracker	Sensitive, complex system with optics and servo motors, easy to keep a constant distance away from the screen, high eye tracking accuracy	Reading, psychophysics, neuroscience, human's ocular hemodynamics	Electro-Oculography, scleral Search Coils, infrared oculography
Screen-based eye tracker [9, 11, 32–35]	Web/industry camera monitors(screen), infrared illuminators, host machines	Head fixed (optional), remote (Laboratory setting)	Eyegaze, Tobii 1750, Tobii Pro, Gazepoint,Smart Eye, SMI RED, Mirametrix S2, Tobii TX300	Easy to be installed, configured, and transported, well-suited for tracking eyes' motion, limited resolution and accuracy	Neuroscience, ophthalmology, reading studies, psychology , gaze-contingency studies, infant and child research	Infrared oculography or video oculography, multi-camera-based eye tracking
Head-mounted eye tracker [10, 30, 36–38]	Helmet/hat/glass frame, cameras, base station/ storage device, IR light and mirror	Free head movement with limited pose variation (Daily Life)	Tobii Pro Glasses, Dikablis Glasses, ASL head-mounted Eye Tracker, EyeLink I,OpenEyes	Easy to transport; track with the head free-to-move or with head supported	Gaming, marketing studies, website analysis, neuro/psychological research, and virtual reality	Video oculography, multi-camera-based eye tracking
Mobile eye tracker [12, 39–41]	Tablet/ Smart phone	Free movement (Daily Life)	EyeTab, GazeSpeek, TabletGaze, GazeCapture	Fast and easy setup; ideal for close-up stimuli, touchscreen and reaching	Neuroscience, virtual reality,marketing, website analysis, disability assistance	Video oculography, on-board camera-based gaze detection and estimation

essential components and techniques to lead this survey and simplify the understanding of the systems and methodologies in the literature, as shown in Figure 1.

- We investigate the learning analytics and methodologies to model student learning, capture and study fluctuations, detect gaps between teaching and learning, create personalized curriculum and content, provide and deliver learning resources and learning activities to make education more accessible for students and teachers.
- We identify and discuss the challenges, open issues, and opportunities of enabling collaborative and intelligent teaching and learning in real-world learning environments.

The remaining paper is organized as follows. Section II presents the methodologies used for conducting this literature review and the graphical summary. Section III provides an in-depth review of the most recent available datasets and tools in an attempt to assist current and future studies with more insightful and precise guidance. Section IV states the eye-tracking learning data collection and management. Section V discusses challenges and open issues. Finally, we conclude this survey and suggest future research directions based on current research gaps in Section VI.

II. METHODOLOGY

A set of major terms, such as eye-tracking devices and platforms, gaze tracking, student attention, mind wandering, intentions, cognitive ability, interaction, learning outcomes, learning dataset, learning data mining, and intelligence, were derived in order to conduct a comprehensive literature review. The search expressions for these significant search terms were built using Boolean operators and interchangeable search strings. IEEE (Institute of Electrical and Electronics Engineers), ACM (Association for Computing Machinery), Wiley Online Library, ScienceDirect, Springer-Link, JSTOR,

Scopus, and Web of Science were used to apply the search terms and expressions to digital online databases that are well established and maintained in the field of computing, sensing, networking, and education. Additional publications relevant to the study questions we're engaged in were retrieved using Google Scholar and Google. The systematic reviews' references [17], [18], [23], [36], [42]–[50] were also examined to ensure that no relevant papers were neglected. Only English-language articles from conference proceedings, transactions, magazines, books, essays, technical reports, white papers, and manufacturer's technical guides were studied.

A preliminary literature search yielded 295 results. 67 duplicate articles were found in multiple databases and were eliminated. The remaining 228 items that appeared to be acceptable for consideration were carefully reviewed and filtered using the following criteria: C1. The magazine covered cutting-edge technology and use cases for encouraging student learning and engagement; C2. The publication covered eye tracking and learning analytics technologies and related applications for enhancing student learning success; C3. The paper used experimental trials based on either laboratory data or open-source datasets to explain, analyze, show, or assess the technology. A total of 98 papers were gathered at the conclusion of the study selection process.

Data extraction and synthesis were then performed. For ease of explanation, we provide a graphical summary of the current eye-tracking enabled learning systems and analytics with essential components and techniques. It can serve as a generic foundation and strong support for using different types of eye-tracking hardware, analytic modules, algorithms, and tools for future research. A general learning system has five main components, namely (1) face-to-face (F2F) and online learning with eye-tracking systems, (2) data collection and management with tools, (3) feature extraction,

(4) measurement, learning analytics, and intelligence, and (5) open issues and challenges, as shown in Figure 1.

F2F and online learning with eye-tracking systems: The use of eye-tracking sensors and hardware platforms has been transforming the traditional way of teaching and learning in physical and virtual classrooms. Both sensors and platforms can assist researchers and educators in data collection, learning analytics, and various application scenarios. Based on the hardware setup, in our previous work [25], we classified the existing eye-tracking systems into four main types: tower-mounted eye tracker [8], [11], [27]–[31], screen-based eye tracker [9], [11], [32], [33], [35], head-mounted/wearable eye tracker [10], [30], [36]–[38], and portable/mobile eye tracker [12], [39]–[41], as listed in Table 1. A detailed discussion on those four categories of eye-tracking devices and platforms can be found in [25].

The details of components (2) - (5) with their corresponding modules or technologies will be described in the rest of this survey. Specifically, the studies for (2) data collection and management with tools and (3) feature extraction will be reviewed in Section III. The technologies for (4) measures, learning analytics, and intelligence will be examined in Section IV followed by a discussion on (5) open issues and challenges.

III. DATA COLLECTION AND MANAGEMENT WITH TOOLS

Data is the most important element and part of all data science, analytics, machine learning, and artificial intelligence. Without data, we have nothing to train models, and all related collaboration and intelligence will remain futile. Data can be any fact, value, eye/face/scene/salience images, sound, videos that are collected during teaching and learning in F2F classrooms and online learning with different eye-tracking hardware and systems. For the purposes of recording and promoting data-driven scientific research in eye-tracking, a set of public eye-tracking and gaze datasets has been published and become available in recent years. Even though existing work [51]–[55] that we analyzed has provided a summary of eye-tracking datasets, they are only used for comparison in terms of the size of the dataset, the number of subjects, gaze targets, head poses, and collection duration without considering hardware features and environmental setup. They, however, don't appear opposite to test algorithms and applications with low-cost webcams or mobile devices in classrooms or online learning.

These observations motivate us to re-examine and take a closer look at the publicly available datasets with their real-world setting and applications, with an emphasis on unique resource requirements for enabling face-to-face or online learning to better benefit from technologies and student learning datasets. We concentrate mostly on the increasing use of eye tracking and eye gaze monitoring systems for learning analytics in real learning environments including classrooms and online learning. Table 2 presents a summary of publicly available student eye tracking datasets and saliency detection benchmarks during student learning.

Saliency is what stands out and attracts great human visual attention in a scene (like photo, website, slide, and video). In essence, our brain and eyes are always unconsciously and automatically focus on the most salient (important) regions during viewing. To detect saliency and predict users' visual attention, eye-tracking data and analysis are often measured and provided as objective ground truth in different studies.

In MIT saliency benchmark [55], there are collections of saliency benchmark, saliency, and saliency-related datasets with eye tracking data captured by different commercial-off-the-shelf eye trackers. To take EyeTrackUAV [56] as an example, raw gaze data, fixation, and cascade events of fourteen(14) participators are collected using EyeLink 1000 Plus when they are watching UAV (Unmanned Aerial Vehicles) videos as visual stimuli to study their visual behavior. EMOd(EMotional attention dataset) [57] is created to help investigate how image sentiment influences human attention and visual perception. TurkerGaze [58] constructs a webcam-based system for crowd-sourced eye tracking data collection from 200 participants across the United States via Amazon. Wandering eyes [59] collects and studies 77 students' eye tracking data using an Eye Tribe tracker in a laboratory environment to examine eye movement patterns of mind-wandering and investigate how mind wandering affects attention during video lectures. A detailed overview of these datasets is available in [54] and [55]. Due to space constraints, we only focus here on datasets collected with mobile or cost-effective eye trackers that can be easily deployed in our daily life for face-to-face, online, and mobile learning.

PoG [60] dataset comprises 3D coordinations of display and Mobile-eye tracker, predefined target pixel locations, and eye tracking videos containing eye movements of 20 subjects sitting in front of the display and looking at the target pixels. It is an early eye-tracking dataset for evaluating point-of-gaze detection algorithms, but severely limited with respect to variabilities in body and head poses, illumination, and visual stimulus. In [61], eight (8) color cameras are fastened on a display to capture images from different views in a synchronized manner. The images are then manually annotated with facial landmarks to define head pose, reconstruct the 3D shape of eye regions, and synthesize more eye images for a person and pose independent 3D gaze estimation. All data including raw multi-view eye images, 3D eye shape models with annotations, and the synthesized eye images based on 3D models are contained in the UT multi-view gaze dataset.

Recent years have seen an explosion of using self-developed mobile eye trackers and smart portable devices as hardware for pervasive eye tracking in daily life settings. The main characteristics of this type of hardware are its abilities of image acquisition, data processing, and computing. These characteristics facilitate the rapid prototyping of new eye tracking configurations, algorithms, and applications in classrooms without requiring to purchase(repurchase) eye tracking systems or limit user movements and illumination changes. MPIIegoFixation [62], LPW (Labelled Pupils in the Wild) [63], GazeSim [64], MPIIMobileAttention [66] are the

TABLE 2. Public eye-tracking datasets and saliency benchmark.

Name	Hardware	Tools	# Subjects	Pose	Illumination	Data Type	Data Size	Field of Use	Year
Point of Gaze (PoG) [60]	Mobile eye XG (head mounted, infrared monocular)	None	20	Sitting	Laboratory setting	Video, CSV, mat, dat	Data(6.4 GB)	Point of Gaze detection	2012
UT Multi-view Gaze dataset [61]	1 chinrest and 8 cameras	None	50	Sitting	Laboratory setting	Images, modes(.obj), gaze data(.csv)	Data (20GB) including 3D eye shape models with annotations, and synthesized eye images	3D Gaze Estimation, video coding	2014
TurkerGaze [58]	Web camera	clmtrackr	200	Sitting	Laboratory setting	Image, JSON	20 K Images with object/region annotation and recorded eye tracking data	Crowdsourcing saliency	2015
MPIIego-Fixa-tion [62]	PUPIL	None	5	Free move-ment	Daily Life	Video with annotations	Data (3.1M)	Fixation detection in mobile settings	2015
Gaze Cap-ture [41]	Mobile phones and tablets with front-facing camera	itracker	1474	Uncon-strained	Daily Life	Video, images and JSON	images (2.5M) and detection data (1.5 M)	Eye tracking	2016
LPW [63]	PUPIL	None	22	Free move-ment	Daily Life	Video with annotations	Data (2.4 GB) including 130K images	Pupil detection	2016
GazeSim [64]	PUPIL	ArUco and VPython	14	Standing	Laboratory setting	Images with 3D information of pupil	Data (81.4M)	3D gaze estimation	2016
Invisible-Eye [65]	Self-developed mobile eye tracker with 3 cameras	None	17	Sitting/stand-ing	Laboratory setting	Video and images with labels	Data (49.2GB) including training data and test data	low-cost eye tracker development and learning based gaze estimation	2017
TabletGaze [40]	Tablet with front-facing camera	Tablet	51	4 types	Daily Life	Video	816 videos (122.6 GB)	Gaze estimation	2017
Wandering Eyes [59]	Eye Tribe tracker	OpenSesame and PyGaze	77	Sitting	Laboratory setting	Video, CSV, Matlab (.m)	Data (342.1MB)	Student learning assessment	2017
EMOd [57]	EyeLink 1000	CNN model and LabelME platform	16	Sitting	Laboratory setting	Images with fixation annotations	Stimuli images (82.3 MB), fixation maps(13.7 MB), raw eye-tracking data(2.3 MB), image-level Annotation(64 MB)	Image sentiment and visual Attention	2018
MPIIMobile-Atten-tion [66]	PUPIL	Python programs in the dataset	20	Free move-ment	Daily Life	Video with annotations	Data (2.4GB)	Fixation detection, attention forecasting	2018
EyeTrack UAV [56]	EyeLink 1000 Plus	Eyelink toolbox	14	Sitting	Controlled using scripts	Video, CSV	UAV stimuli videos and raw eye tracking data	Monitoring human visual behavior	2018
MPIIGaze [51]	Laptop with front-facing camera	Gazenet	15	Uncon-strained	Daily Life	Images with annotations	Images (213K) including 37K images(B) with annotations	Gaze estimation	2019
MIT saliency bench-mark [55]	Commercial eye trackers	Various tools	8∼72	Sitting	Laboratory setting	Video,image, CSV, eye-tracking data	Depends on specific dataset	Saliency detection, human fixation and behavior study	2006-present

datasets collected by exploring the design of pupil (mobile eye tracker [38]) with freedom in selecting cameras and additional parts. MPIIEgoFixation [62] collects data when participants walk around in the presence of head movements and scene dynamics. There are five (5) participants attending experiments and over 2,300 fixations and 40,000 frames with fine-grained annotations for evaluating the proposed fixation detection method based on the visual similarity of gaze targets.

LPW [63] contains 66 eye region videos from 22 participants with a head-mounted eye tracker in indoor and outdoor environments. Without loss of generality, the participants are allowed to wear glasses, contact lenses, and make-up for robust pupil detection as PuRe [67] demonstrates. GazeSim [68] is provided for studying 2D-to-3D gaze estimation. In experiments, 14 participants attend the data collection and need to stand at 5 different distances to the

display. The distance is then recorded as the depth information for 2D-to-3D mapping. MPIIMobileAttention [66] contains annotated videos recorded from 20 participants engaged in common activities (like reading books, using mobile phones, working on computers, and walking) in indoor and outdoor environments. InvisibleEye [65] develops its mobile eye tracker with three low-resolution cameras embedded into a normal glass frame. The basic idea behind it is to use low-resolution eye images captured in different views as input to train a neural network for learning-based gaze estimation. The dataset contains 200,000 synthesized eye images, 280,000 real eye images from 17 participants under laboratory setting during calibration, and 240,000 eye images from 4 participants when looking at targets from different angles in a mobile setting.

Several other datasets [40], [41], [51] are recorded and published using modern portable commodity hardware like

laptops, tablets, and smartphones in real-world settings. MPIIGaze [51] includes 213,000 images from 15 participants using laptops with a front-facing camera over several months in their daily life and offers 37,667 images with manually annotated facial landmarks. The data in MPIIGaze covers a wider range of background, illumination, eye appearance, and poses. GazeCapture dataset [41] is a large-scale public dataset, containing data from a total of 1474 participants using mobile phones/tablets via crowdsourcing. It trains iTracker that we have analyzed in this study for real-time robust eye tracking and gaze prediction on mobile devices. TabletGaze [40] contains 816 videos, timestamps, and locations of 5×7 predefined gaze points. During data collection, 51 participants can have four kinds of body postures: standing, sitting, slouching, and lying with a tablet in an unconstrained mobile environment. They may wear glasses if necessary and have different head poses and eye appearances while looking at predefined locations on the tablet screen.

We argue that open-source projects and datasets have unique abilities to support both commercial and low-cost eye trackers, easily alter experiments to specific scenarios, quickly prototype ideas, and enable research and applications without major restrictions. Development and maintenance of open-source projects and datasets can help build and support a community of student education, researchers, and developers. We include the most recent available datasets in this study in an attempt to assist future studies with more insightful and precise guidance. This is a new dimension beyond just depth and width. However, a number of open-source projects and datasets (like OGAMA, OpenEyes, EyeTab, OpenEyes, and EyeRecToo) are no longer maintained by original developers. There is a huge need for more efforts and contributions made by us in the community to label, format, and pre-process the data so that the dataset can be easily downloaded and used by other studies.

Moreover, the data mining strategies and associated tools are also in need of a substantial amount of research attention. The goal is to facilitate the integration and extension of existing research studies and to promote collaboration, idea sharing, and innovation work. As there is still no widely accepted open-source library or toolkit available for convenient data processing and mining, current strategies [16], [17], [69]–[75] mainly deal with disparate sources of data and very specific situations. This makes it very difficult to understand student learning and develop practical teaching/learning strategies, especially for interested novices as well as interdisciplinary researchers. We expect that both teaching-learning hardware, software, and datasets will be pervasively used in our daily learning activities and incorporated into the next generation of education and human-computer interactions.

IV. MEASURES, LEARNING ANALYTICS, AND INTELLIGENCE

In this section, we will review the studies that have been accomplished in conjunction with the collected data to characterize, study, and improve the learning process. We break down the discussion by generalizing measures or approaches that researchers have used in applying data science, learning analytics, and machine learning (see Table 3).

A. MEASURING ATTENTION AND MIND WANDERING

When monitoring a learning task, how well students are paying attention to the materials is often a good indication of whether or not they are comprehending the materials. Several researchers have attempted to gather information about the attention of students during a learning task, and apply data analytics to build models that can automate the classification and detection of attention and inattention while a learning task is being conducted. Faber *et al.* [78], [86] have applied the idea to classifying and modeling mind wandering during comprehension tasks where the goal is to study and learn a body of materials.

Participants could self-report anytime they became aware that they had not comprehended a previous part of what they had been reading. Also, explicit probes were sometimes given, here the student had to either state they were currently paying attention, or their mind was wandering. Indeed, this type of self-report and explicit probing can alter the labeling of attention vs. non-attention itself, just by the act of the participants being meta-cognitively more aware of their focus of attention. However, it does provide a set of labeled eye gaze data where time periods or states of inattention can be associated with eye gaze behavior. The authors in this study applied several standard machine learning classifiers, such as simple logistic regression, Bayes nets, SVM, and decision trees, to build a binary classifier that could take summary eye gaze data and predict whether the student was attentive or mind wandering during a particular segment. The authors showed that it was possible to build a model of attention for the system that could potentially be applied to building an automated system capable of detecting and classifying when student's minds are wandering during the learning task.

In another study, Gomes *et al.* [73], [87] captured eye-tracking data from high school students working on a task designed to teach some basic engineering skills. Unsupervised machine learning clustering was used to cluster student eye data into several groups based on their gaze pattern. When compared to post-task evaluation metrics, highest performing students tended to cluster together with similar gaze patterns, and likewise, lower-performing students had different clusters of gaze behavior identifiable from the eye tracking data.

Raca & Dillenbourg [71] demonstrated a more general system for detecting the attention of a whole group of students in a classroom. In the system, regular videos were captured and video imaging techniques were used to estimate the focus of attention of individuals detected within the view of the cameras. A rough estimation of gaze direction could be calculated for individuals detected in the camera's field of view. Other measures that were extracted from the camera data included when slides were changed, slide durations, and when questions/discussions occurred from students and

TABLE 3. Multimodal learning analytics with eye-tracking data.

Study	Hardware	Software	Environment	Sensors	Measures	ML Models
Bendarik[76, 77]	Tobii ET 1750[1]	Libsvm (Matlab)	Individual	Think Aloud	Action Codes (Cognition, Evaluation) Activity Performance (High, Medium, Low)	SVM[4]
Bixler[78]	EyeTribe[1]	Open Gaze and Mouse Analyzer (OGAMA) and Weka	Individual	Self-Report Probe	Attention (Mind Wandering)	NB[5], SVM, LR[6], DT[7]
Edwards[79]	SMI iView[1]	SMI Experiment Center, Openface, Affdex SDK	Individual	GSR[8], Camera (Facial Expressions)	Activity Performance (Quiz Score), Cognitive Load	—
Kardan[80]	Tobii T120[1]	Weka	Individual	No extra sensors	Attention, Activity Performance (High Achievers, Low Achievers)	DTC, SVM, LRR[9], LR, NN[10]
Mangaroska[81]	SMI RED 250[1]	SMI software and SDK	Individual	No extra sensors	Activity Performance, Expertise	Linear Model
Ooms[69]	SR EyeLink1000[2]	SR Dataviewer	Individual	No extra sensors	Expertise (Expert, Novice)	—
Prieto[82]	SMI[3]	SMI Begaze	Classroom	EEG[11], Accelerometer, Audio, smartphone	Teacher Activity (Explaining, Monitoring), Social Plane (Group, Individual)	RF[12], DT, SVM
Raca[50, 83]	SMI[3]	SMI software and SDK	Classroom	Camera	Attention, Teacher Activity	—
Zemblys[84]	SR EyeLink1000[2]	Scikit-Learn, Seaborn, Lme4, Psychopy	Individual	No extra sensors	Eye Movement Events (Fixations, Saccades, PSO)	RF
Sharma[85]	SMI RED 250[2]	SMI software and SDK	Individual	No extra sensors	Content coverage, scanpath , students/ teacher co-attention	GPM[13], SVM, GAM[14]

[1] Tower-mounted eye tracker   [2] Screen-based eye tracker   [3] Head-mounted/ wearable eye tracker   [4] SVM - Support Vector Machine   [5] NB - Naive Bayes
[6] LR - Simple or binary Logistic Regression   [7] DTC - Decision Tree / Decision Tree Classifier   [8] GSR - Galvanic Skin Response   [9] LRR - Linear Ridge Regression
[10] NN - Neural Network (Multilayer Perception)   [11] EEG - Electroencephalogram   [12] RF - Random Forest   [13] GPM - Gaussian process models
[14] GAM - Generalized additive models

other properties. In addition, some self-report data was also collected from the students on their level of attention, sometimes during the class, and sometimes in questionnaires afterward. These researchers later went on to refine their system with more sensors and capabilities [47]. In general, this work demonstrated the possibilities that non-intrusive measures of attention might have in building learning management systems for real-time classroom feedback to instructors in more traditional lecture classroom settings.

B. PREDICTING COGNITIVE ABILITIES
Another useful measure is to try and use eye-tracking data to directly predict the level of cognitive ability, or the expertise level of the student or user of the learning system [16], [17], [70]. This is related to the concept of user modeling in user interface design. But here researchers are interested in modeling more fundamental cognitive abilities, such as working memory capacity. In general, this type of model of the users' cognitive abilities might be useful in providing different content or materials that are known to be more effective for a user with, for example, more preference for visual learning or verbal materials.

For example, [70] reports using eye-tracking data to predict some basic cognitive abilities, such as perceptual speed, and visual and spatial working memory capacities. Several good instruments are available in psychology for measuring individuals along these dimensions. So again the goal is to see if standard machine learning and modeling techniques can be used to classify subjects based on gathered eye-tracking data into the correct level of a particular cognitive ability type. In the study, Conati *et al.* gathered eye-tracking data of people using an interface designed to learn about and make a decision

for a target problem domain. They also applied standard measurements to determine cognitive ability scores on the subjects after the fact of five separate cognitive measures. Again basic machine learning classifiers were tested, and it was shown that they are potentially able to predict these types of cognitive measures from eye-tracking data along and automatically, with enough accuracy to be used as a model of the users to help design better decision making systems.

As another example, some researchers are interested in predicting the performance levels of subjects from sensor data. Bilkstein [16] used eye-tracking data to perform a classification task on labels generated from behavioral codes based on think-aloud protocols. In this research, a simple puzzle (an 8-tiles puzzle game) was used as the task to be solved. So another measure used to build machine learning predictors was an after-the-fact assessment of the skill level of the puzzle solvers based on their time/speed of solving the puzzle. The researchers tried several approaches to building models of subject skill levels. In one approach described, subjects were broken down into high, medium, and low performers based on their skill at the task, and a classifier was trained to predict the user skill level from the eye-tracking data gathered while solving the puzzle. The model worked well enough to roughly identify skill level from eye-tracking data of people on this puzzle task.

In [69], the authors used gathered eye-tracking data to model behavioral differences between novice and expert map users. In this paper, the authors showed that there were statistical differences in various derived measures of where and how novice and expert map readers used a program to search for and answer questions about information from a map. This type of real-time prediction of the user level

on task-based eye-tracking or other gathered sensor data is obviously a very powerful tool. Open questions remain on if general cognitive abilities are sufficient for building learning systems that are more reactive to the abilities of the learners, or if more specific models need to be developed based on the particular task or subject domain. Or in other words, could a general eye-tracking model be developed that might be applicable to many different learning domains to detect and classify user skill level while learning, or do models need to be built for each individual learning domain and learning system?

C. ANALYZING GROUP PERFORMANCE AND DYNAMICS

Another promising approach to using eye-tracking and multimodal analytics is in studying group dynamics and performance of teams in active learning and hands-on group training [71], [72], [74], [75]. Often one measure that has been found useful is the amount of synchronization of activity and behavior of team members when performing the learning task. The overall approach is to analyze the synchronization of two or more sets of eye-tracking data and compute the number of times a pair or team achieves joint visual attention (JVA). JVA has been studied in many contexts and disciplines and is known as a strong predictor of the quality of a group's interaction and success. For example, [71] have shown that in pair programming, good programmers tend to have higher JVA compared to less proficient programmers. Also, there is some evidence that you can actually try to influence and encourage JVA in groups to get them to improve their performance [74].

Spikol *et al.* [75] research focused on project-based learning in a group setting. The questions asked were which features of a student group are good predictors of team success. A truly multimodal set of sensor data was gathered from teams working on engineering training tasks, including data from cameras and microphones, embedded sensors on components used in the learning task, and mobile computing devices. One of the key predictors was sensor data from image processing that detected the focus of attention on other team members, or shared focus of attention on learning artifacts. Shared focus and amount of interaction between team members ended up being important features that machine learning classifiers could use to build predictive models of the performance of teams in the lessons based on the multimodal sensory data gathered. Huang *et al.* [88] proposed to collect eye-tracking data, body postures, and motion and leverage k-means clustering to recognize the collaborative learning states when students programmed robots to solve maze problems. The results indicated that learning was significantly and negatively correlated with the probability of remaining in the non-collaborative state. The more collaboration there is among students, the more learning they gain.

D. PREDICTING USER INTENTIONS AND GOALS

In many types of systems, especially collaborative situations, predicting user intention is useful in mediating better interactions. For example, [89] showed that SVM classifiers can accurately predict intentions for a request before users make them for a collaborative customer-worker system. [85] presented a study with 40 university students who watching massive open online course (MOOC) lecture videos to envision the student motivation and learning gain in terms of content coverage, AOI, perceptual and conceptual attention by using the Gaussian process models (GPM), SVM, Generalized additive models (GAM).

In general, predicting user intentions and goals is very valuable in the context of usability and user interface (UI) design. The general paradigm is that you can easily capture and study eye gaze behavior when given a task to complete using a user interface. You can then measure the ease or difficulty users have with completing the task. For usability testing, the model of eye gaze behavior can then be applied to determining good and usable interfaces from poor ones [90].

For learning analytics, the opposite application is often more useful. We would like to infer what the current goal or intention of the learner might be from the given set of eye gaze or sensor data currently being gathered from the user. This information would allow one to determine when students are on task with what they are currently expecting from the lesson or are off on a tangential task that might not be most conducive to completing the current goal.

E. EVALUATING USER DEFICIENCIES AND USER MODELS OF SUBJECT CONTENT

If we can more accurately model and predict students' skill levels as demonstrated from measures gathered during their learning session, the information can be used to directly influence the learning activities and direct students towards more appropriate tasks or concepts for their current level of mastery. In this type of study, researchers are interested in questions of the domain knowledge of a subject area. Thus building models of the content area (such as a concept map), and mapping or estimating which concepts students have mastered, and which they are deficient in is useful in planning appropriate study activities.

In the field of intelligent tutoring systems, learner modeling is the area by which researchers attempt to model the domain knowledge (and deficiencies) of a learner in order to better guide the learning session [91]. For example, [92] describes work on leveraging eye tracking data to improve student modeling. Most of these efforts look at ways to build classifications of eye gaze patterns that map onto data where it was clear students were deficient or did not comprehend a particular needed concept at a given moment. While much of the research has demonstrated the possible advantages of being able to model high-level student understanding of concepts, it is still very difficult to integrate this type of modeling into working learning systems.

V. OPEN CHALLENGES AND FUTURE WORK

This section presents a range of challenging open issues and future research directions. As mentioned in the previous

sections, the data gathered during learning can be used to predict mind wandering and sculpt user cognitive processing profiles, and assess the dynamics of group collaboration on problem-solving with data science and machine learning. Despite its great advantages, there are novel and emerging challenges that must be solved in the future for enabling pervasive and proactive learning with more interactive, personalized, engaging learning systems. Due to this, our discussion is conducted, which could serve as a guide for future studies.

A. COLLABORATIVE SENSING AND PERCEPTION

Instead of single-subject-based eye tracking, a collaborative sensing and networked learning system is expected to be built as a fundamental basis for facilitating face-to-face teaching/learning, and making eye tracking, attention measuring, intention predicting, group cooperation and performance analysis easy tasks for instructors. The system can comprise many multimodal sensory elements including eye tracking devices (e.g., wearable eye tracker), desktops/laptops, microphones, tablets and digital pens, world cameras, a control terminal/station, and a server farm. The majority of elements are deployed in the classroom and connected to the server farm via a wired/wireless network. During lectures, the server is responsible for gathering and analyzing real-time data streaming from the eye trackers, microphones, tablets, and world cameras to provide feedback, active monitoring, and intervention of learning analytics to the lecturer/instructor with the control terminal/station. The more devices that are involved for cooperation, the more information can be collected for enabling collaborative sensing and perception. However, the system becomes more complex with higher construction costs and maintenance overhead.

B. DATA SHARING AND MINING

One of the main barriers to the widespread adoption of eye tracking and MMLA technology is data collection and sharing. In essence, the multimodal data collected during teaching and learning from different media/sensors (camera, computer, tablet, digital pen) should provide ground truth with detailed annotations against the eye tracking devices, models and techniques used for evaluation. By offering the data, any interested educator or researcher in the community can easily start their studies and research to understand how to improve students' learning, shift the day-to-day teaching activities and promote teachers' professional development in classrooms or online learning. Nevertheless, it is rather costly. While there is a wealth of open-access datasets [40], [41], [51], [56]–[66], [93], [94], it still remains unsatisfactory, mainly due to that most of the datasets were model/application-oriented, asynchronous, or uncorrelated across modalities (audio, video, eyetracking, and gestures, digital interaction logs). Some datasets like PoG [60], UT Multi-view Gaze dataset [61], TurkerGaze [58], and MPIIEgoFixation [41] have even turned to inadequate as a result of the lack of fidelity annotations, periodical follow-up management, and maintenance. It rightfully illuminates one of the promising paths for future work.

C. REAL-TIME EYE TRACKING ANALYTICS

Over the past five years, as instructors, we have observed that students, particularly beginners, often give up prematurely after encountering initial, sometimes even minor, problems on a learning task, wrongly concluding they lack aptitude for the task area. The advancement of real-time eye tracking analytics makes it possible to detect when students are not progressing sufficiently on a problem, are sidetracked, or focus on irrelevant details. It is also noteworthy that every student is special with a unique educational background, learning capability, and objectives. And students deserve the opportunity to learn in their own ways. Although experienced teachers might be capable of noticing and managing student distractions in the classroom easily, it doesn't mean that experienced teachers understand every student and his/her learning difficulties/observables. For students, inappropriate instructions might cause plenty of anxieties, frustrations, and vexations. There is an urgent need to conduct real-time eye tracking analytics for assisting educators/teachers to come to know more about each student and discover effective ways to engage both the individual students and the entire class in active learning in real time. Additionally, appropriate learning support should be provided to students so that they can easily make progress and gain confidence in their own pathways of learning.

D. TEACHING PRACTICE EVALUATION AND ENHANCEMENT

It is very important to conduct appropriate teaching activities at the right time for enabling effective learning. Teaching practice (activity) assessment and evaluation to measure teaching efficacy and student success, however, is intensely difficult in face-to-face classrooms, where each session is unique and fulfilled with various activities and open-ended tasks [95], [96]. The activities may include speeches, writing, discussion, group work, demos, exercises, games, case studies, and problem-based learning. Teachers are also allowed to make necessary changes to their teaching activities when different types of multimodal sensory elements such as computers, laptops, tablets, PDAs, smart cell phones, and digital pens are available for assistance. The open-ended tasks could be data formatting, problem definition, solution design, program implementation and debugging, data analysis, and design optimization. Even with a slight change in environment or session, the teaching and learning results can be significantly different. What is worse, the possible combinations of activities and tasks increase the complexity of teaching practice evaluation and enhancement. Such inherited nature and physical system limitations in sensing and data collection have made it infeasible to conduct real-time teaching practice evaluation and reflect instant formative feedback from students.

To solve it, novel approaches to characterize teaching practices and discover patterns of effective teaching practices in face-to-face classrooms are desired with fine-grained data collection and analysis. That is, both traditional and

innovative teaching practices rely more than ever on highly complex, interactive teaching-learning environments that operate beyond the reach of current learning analytics. Accompanied by the rapid advances in multimodal sensory elements in the environments, the fine-grained data could be collected effortlessly, which would provide new opportunities to develop efficient approaches for characterizing teaching practices and discovering patterns. Particularly, throughout specific teaching activities, students' learning trajectories might be clearly micro-scoped. It would allow teachers to build up better teaching activities and learning materials for improving student learning experiences and promoting student success.

E. PRIVACY AND UTILITY

Students' privacy is one of the biggest concerns of collecting eye-tracking data and conducting learning data analytics during F2F learning in a similar system as shown in Figure 1. Students' interests, reading/writing/web browsing contents, and watching videos displayed on screens can easily expose personal information (e.g., gender, name, address, age, and health status) or reveal political, sexual, cultural, or other personal preferences. It is also proven [22], [97], [98] that eye movement/gaze pattern is as unique as an iris or fingerprint and can be used for user authentication. A key challenge is to find the right trade-off between privacy and utility. A certain amount of random noise can be introduced to hide subjects and reduce the potential loss without affecting the utility of the eye-tracking data and degrading system performance. We expect that a series of advanced eye-tracking data privacy-preserving mechanisms for erasing sensitives data from the raw dataset, aggregating data, enabling differential privacy, and providing AOI (Areas of Interest) metrics and summary data of eye movement events could be developed in the near future.

VI. CONCLUSION

In this study, motivated by the common problems and technical constraints faced by both students and teachers in daily learning and teaching activities, we have examined the current trends in hardware and research that are being applied to using the collected data for classroom and online learning analytics and prediction. Eye-tracking data has certainly been shown to be an attractive source of information for F2F and group settings, as it provides a window into student attention, and allows us to monitor important measures of the learners and group dynamics. We have also reviewed different types of eye-tracking and learning systems, software, datasets, and related studies with a particular emphasis on their respective advantages and limitations. Furthermore, we have identified and discussed a string of open issues and challenges that are needed to be addressed for enabling and promoting pervasive and proactive learning. Those open issues and challenges by no means represent all problems of previous research and studies in progress. But they are identified and highlighted in this study as promising future research directions.

We will continue to witness extensive research studies and activities in this field in the future. Specifically, our finding in this study calls for further attention and research on addressing one or more of the open issues and challenges that we have discussed by combining existing pedagogical and human-computer interaction (HCI) theories. Additionally, future research should focus on developing new multi-sensory elements, systems, and tools, learning analytics methods, machine learning, and artificial intelligence algorithms that can be used as functional modules added or removed as necessary. This can facilitate F2F and online teaching as well as learning and make eye-tracking, attention measuring, intention predicting, group cooperation analysis easy tasks. There is no doubt that as the field matures, it becomes more common to have collaborative sensing and networked learning with positive motivation in pervasive learning.

REFERENCES

[1] (2020). *Coronavirus Impacts on Students and Online Learning*. [Online]. Available: https://www.bestcolleges.com/blog/coronavirus-impacts-on-students/

[2] E. R. Wester, L. L. Walsh, S. Arango-Caro, and K. L. Callis-Duehl, "Student engagement declines in STEM undergraduates during COVID-19–driven remote learning," *J. Microbiol. Biol. Educ.*, vol. 22, no. 1, p. 2385, 2021.

[3] J. Zaletelj and A. Kosir, "Predicting Students' attention in the classroom from kinect facial and body features," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, p. 80, 2017.

[4] C. Thomas, "Multimodal teaching and learning analytics for classroom and online educational settings," in *Proc. 20th ACM Int. Conf. Multimodal Interact.*, Oct. 2018, pp. 542–545.

[5] S. Park, X. Zhang, A. Bulling, and O. Hilliges, "Learning to find eye region landmarks for remote gaze estimation in unconstrained settings," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, Jun. 2018, pp. 1–10.

[6] P. Blikstein and M. Worsley, "Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks," *J. Learn. Anal.*, vol. 3, no. 2, pp. 220–238, 2016.

[7] D. Daniele, S. Maren, D. Hendrik, B. Dirk, T. Stefaan, and S. Marcus, "Learning pulse: A machine learning approach for predicting performance in self-regulated learning using multimodal data," in *Proc. LAK*, 2017, pp. 188–197.

[8] INITION London. (2019). *Motion Capture & Tracking—Sensomotoric Instruments iView X Hi-Speed*. [Online]. Available: https://www.inition.co.uk/product/sensomotoric-instruments-iview-x-hi-speed/

[9] S. E. Gaither, K. Pauker, and S. P. Johnson, "Biracial and monoracial infant own-race face perception: An eye tracking study," *Develop. Sci.*, vol. 15, no. 6, pp. 775–782, Nov. 2012.

[10] J. J. MacInnes, S. Iqbal, J. Pearson, and E. N. Johnson, "Wearable eye-tracking for research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices," *bioRxiv*, Jan. 2018, Art. no. 299925.

[11] UCLA Language Processing Lab. (2019). *SR Eyelink 1000*. [Online]. Available: http://processing.linguistics.ucla.edu/resources.html

[12] X. Zhang, H. Kulkarni, and M. R. Morris, "Smartphone-based gaze gesture communication for people with motor disabilities," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2017, pp. 2878–2889.

[13] P. Blikstein, "Using learning analytics to assess Students' behavior in open-ended programming tasks," in *Proc. 1st Int. Conf. Learn. Anal. Knowl. (LAK)*, 2011, pp. 110–116.

[14] S. Scherer, N. Weibel, L.-P. Morency, and S. Oviatt, "Multimodal prediction of expertise and leadership in learning groups," in *Proc. 1st Int. Workshop Multimodal Learn. Anal. (MLA)*, 2012, p. 1.

[15] F.-Y. Yang, C.-Y. Chang, W.-R. Chien, Y.-T. Chien, and Y.-H. Tseng, "Tracking learners' visual attention during a multimedia presentation in a real classroom," *Comput. Educ.*, vol. 62, pp. 208–220, Mar. 2013.

[16] P. Blikstein, "Multimodal learning analytics," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl. (LAK)*, 2013, pp. 102–106.

[17] E. Alemdag and K. Cagiltay, "A systematic review of eye tracking research on multimedia learning," *Comput. Educ.*, vol. 125, pp. 413–428, Oct. 2018.

[18] J. L. Orquin and L. S. Mueller, "Attention and choice: A review on eye movements in decision making," *Acta Psychol.*, vol. 144, no. 1, pp. 190–206, Sep. 2013.

[19] A. Al-Rahayfeh and M. Faezipour, "Eye tracking and head movement detection: A state-of-art survey," *IEEE J. Transl. Eng. Health Med.*, vol. 1, 2013, Art. no. 2100212.

[20] U. Obaidellah, M. Al Haek, and P. C.-H. Cheng, "A survey on the usage of eye-tracking in computer programming," *ACM Comput. Surv.*, vol. 51, no. 1, pp. 1–58, Apr. 2018.

[21] H. R. Chennamma and X. Yuan, "A survey on eye-gaze tracking techniques," 2013, *arXiv:1312.6410*. [Online]. Available: http://arxiv.org/abs/1312.6410

[22] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495–16519, 2017.

[23] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1355, May 2020.

[24] S. Mu, M. Cui, and X. Huang, "Multimodal data fusion in learning analytics: A systematic review," *Sensors*, vol. 20, no. 23, p. 6856, Nov. 2020.

[25] Y. Wang, S. Lu, and D. Harter, "Eye tracking and learning analytics for promoting proactive teaching and learning in classroom: A survey," in *Proc. 4th Int. Conf. Educ. E-Learning*, Nov. 2020, pp. 156–160.

[26] Y. Wang, S. Lu, and D. Harter, "Multi-sensor eye-tracking systems and tools for capturing Student attention and understanding engagement in learning: A review," *IEEE Sensors J.*, early access, Aug. 18, 2021, doi: 10.1109/JSEN.2021.3105706.

[27] J. V. P. De Graef, K. Verfaillie, and F. Germeys. (1998). *Dual Purkinje Eyetrackers_Laboratory of Experimental Psychology*. [Online]. Available: https://ppw.kuleuven.be/home/english/research/lep/resources/purkinje

[28] M. J. Mendel, V. V. Toi, C. E. Riva, and B. L. Petrig, "Eye-tracking laser Doppler velocimeter stabilized in two dimensions: Principle, design, and construction," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 10, no. 7, pp. 1663–1669, 1993.

[29] P. M. V. Diepen, "New data-acquisition software for the Leuven dual-pc controlled purkinje eye-tracking system," Lab. Exp. Psychol., Univ. Leuven, Leuven, Belgium, Psyc. Rep. 246, 1998.

[30] *iView X System Manual: Version 2.7*, SensoMotoric Instrum., Berlin, Germany, 2009.

[31] Cognitive Neuroscience-The University of Edinburgh. (2019). *Eye Tracking Lab and Equipment*. [Online]. Available: http://www.cns.ppls.ed.ac.uk/eyetrackingtechnicaldetails

[32] T. E. Halverson, "An 'active vision' computational model of visual search for human-computer interaction," Ph.D. dissertation, Dept. Comput. Inf. Sci., Univ. Oregon, Eugene, OR, USA, 2008.

[33] University of Rochester. (2015). *Eye Tracking Shows Use of Speech Disfluencies Among Young Children*. [Online]. Available: https://www.tobiipro.com/fields-of-use/infant-child-research/customer-cases/university-rochester/

[34] S. Hutt, C. Mills, S. White, P. J. Donnelly, and S. K. D'Mello, "The eyes have it: Gaze-based detection of mind wandering during learning with an intelligent tutoring system," in *Proc. Int. Conf. Educ. Data Mining, Int. Educ. Data Mining Soc.*, 2016, pp. 86–93.

[35] Avotec. (2017). *Visual Solutions: In-Bore Inert Systems are Indistinguishable From Background Noise*. [Online]. Available: https://www.avotecinc.com/visual-products

[36] M. Vidal, J. Turner, A. Bulling, and H. Gellersen, "Wearable eye tracking for mental health monitoring," *Comput. Commun.*, vol. 35, no. 11, pp. 1306–1311, 2012.

[37] D. Rosengrant. (2015). *Tobii Pro Eye Tracking Reveals Student Attention in Classroom*. [Online]. Available: https://www.tobiipro.com/fields-of-use/education/customer-cases/kennesaw-state-university/

[38] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. UbiComp*, 2014, pp. 1151–1160.

[39] E. Wood and A. Bulling, "Eyetab: Model-based gaze estimation on unmodified tablet computers," in *Proc. Symp. Eye Tracking Res. Appl.*, 2014, pp. 207–210.

[40] Q. Huang, A. Veeraraghavan, and A. Sabharwal, "TabletGaze: Dataset and analysis for unconstrained appearance-based gaze estimation in mobile tablets," *Mach. Vis. Appl.*, vol. 28, nos. 5–6, pp. 445–461, 2017.

[41] K. Krafka, A. Khosla, P. Kellnhofer, H. Kannan, S. Bhandarkar, W. Matusik, and A. Torralba, "Eye tracking for everyone," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2176–2184.

[42] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht, "Augmenting the senses: A review on sensor-based learning support," *Sensors*, vol. 15, no. 2, pp. 4097–4133, Feb. 2015.

[43] D. Davis, G. Chen, C. Hauff, and G.-J. Houben, "Activating learning at scale: A review of innovations in online learning strategies," *Comput. Educ.*, vol. 125, pp. 327–344, Oct. 2018.

[44] J. Roberts. (2017). *Trusted Reviews—Daydream VR: Everything You Need to Know About Google's VR Platform*. [Online]. Available: https://www.trustedreviews.com/news/daydream-vr-3148686

[45] R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[46] C. Ching, J. K. Hwee, and T. Chin, "A review of technological pedagogical content knowledge," *J. Educ. Technol. Soc.*, vol. 16, no. 2, pp. 31–51, 2013.

[47] M. Raca, "Camera-based estimation of student's attention in class," Ph.D. dissertation, I&C-School Comput. Commun. Sci., Ecole Polytechnique Federale De Lausanne, Switzerland, 2015.

[48] J. López-Zambrano, J. A. Lara, and C. Romero, "Towards portability of models for predicting Students' final performance in university courses starting from moodle logs," *Appl. Sci.*, vol. 10, no. 1, p. 354, 2020.

[49] J. López-Zambrano, J. A. Lara, and C. Romero, "Improving the portability of predicting students' performance models by using ontologies," *J. Comput. Higher Educ.*, pp. 1–19, Mar. 2021.

[50] M. Raca and P. Dillenbourg, "Classroom social signal analysis," *J. Learn. Anal.*, vol. 1, no. 3, pp. 176–178, Nov. 2014.

[51] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "MPIIGaze: Real-world dataset and deep appearance-based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 162–175, Jan. 2019.

[52] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4511–4520.

[53] O. Ferhat and F. Vilariño, "Low cost eye tracking: The current panorama," *Comput. Intell. Neurosci.*, vol. 2016, Mar. 2016, Art. no. 8680541.

[54] S. Winkler and R. Subramanian, "Overview of eye tracking datasets," in *Proc. 5th Int. Workshop Qual. Multimedia Exper. (QoMEX)*, Jul. 2013, pp. 212–217.

[55] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" 2016, *arXiv:1604.03605*. [Online]. Available: http://arxiv.org/abs/1604.03605

[56] V. Krassanakis, M. P. Da Silva, and V. Ricordel, "Monitoring human visual behavior during the observation of unmanned aerial vehicles (UAVs) videos," *Drones*, vol. 2, no. 4, p. 36, Oct. 2018.

[57] S. Fan, Z. Shen, M. Jiang, B. L. Koenig, J. Xu, M. S. Kankanhalli, and Q. Zhao, "Emotional attention: A study of image sentiment and visual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7521–7531.

[58] P. Xu, K. A Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao, "TurkerGaze: Crowdsourcing saliency with webcam based eye tracking," 2015, *arXiv:1504.06755*. [Online]. Available: http://arxiv.org/abs/1504.06755

[59] M. Pillan, "Wandering eyes: Reframing ethnography and collecting hints for the design of products and systems for domestic environments," in *Proc. Cumulus REDO*, 2017, pp. 520–530.

[60] C. D. Mcmurrough, V. Metsis, J. Rich, and F. Makedon, "An eye tracking dataset for point of gaze detection," in *Proc. Symp. Eye Tracking Res. Appl. (ETRA)*, 2012, pp. 305–308.

[61] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3D gaze estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1821–1828.

[62] J. Steil, M. X. Huang, and A. Bulling, "Fixation detection for head-mounted eye tracking based on visual similarity of gaze targets," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, Jun. 2018, p. 23.

[63] T. Marc, Z. Xucong, S. Yusuke, and B. Andreas, "Labelled pupils in the wild: A dataset for studying pupil detection in unconstrained environments," in *Proc. ETRA*, 2016, pp. 139–142.

[64] Y. S. Pai, B. Tag, B. Outram, N. Vontin, K. Sugiura, and K. Kunze, "GazeSim: Simulating foveated rendering using depth in eye gaze for VR," in *Proc. ACM SIGGRAPH Posters*, 2016, p. 75.

[65] T. Marc, S. Julian, S. Yusuke, and B. Andreas, "Invisibleeye: Mobile eye tracking using multiple low-resolution cameras and learning-based gaze estimation," *ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 106, 2017.

[66] J. Steil, P. Müller, Y. Sugano, and A. Bulling, "Forecasting user attention during everyday mobile interactions using device-integrated and wearable sensors," in *Proc. 20th Int. Conf. Human-Comput. Interact. with Mobile Devices Services*, Sep. 2018, p. 1.

[67] T. Santini, W. Fuhl, and E. Kasneci, "PuRe: Robust pupil detection for real-time pervasive eye tracking," *Comput. Vis. Image Understand.*, vol. 170, pp. 40–50, May 2018.

[68] M. Mansouryar, J. Steil, Y. Sugano, and A. Bulling, "3D gaze estimation from 2D pupil positions on monocular head-mounted eye trackers," in *Proc. 9th Biennial ACM Symp. Eye Tracking Res. Appl.*, Mar. 2016, pp. 197–200.

[69] K. Ooms, P. De Maeyer, and V. Fack, "Study of the attentive behavior of novice and expert map users using eye tracking," *Cartography Geographic Inf. Sci.*, vol. 41, no. 1, pp. 37–54, Jan. 2014.

[70] C. Conati, S. Lallé, M. A. Rahman, and D. Toker, "Further results on predicting cognitive abilities for adaptive visualizations," in *Proc. IJCAI*, 2017, pp. 1568–1574.

[71] J. Patrick, M. Dejana, N. Marc-Antoine, and D. Pierre, "Collaborative gaze footprints: Correlates of interaction quality," in *Proc. CSCL*, 2001, pp. 184–191.

[72] M.-J. Tsai, H.-T. Hou, M.-L. Lai, W.-Y. Liu, and F.-Y. Yang, "Visual attention for solving multiple-choice science problem: An eye-tracking analysis," *Comput. Educ.*, vol. 58, no. 1, pp. 375–385, 2012.

[73] J. Gomes, M. Yassine, M. Worsley, and P. Blikstein, "Analysing engineering expertise of high school students using eye tracking and multimodal learning analytics," in *Proc. Int. Conf. Educ. Data Mining (EDM)*, Memphis, TN, USA, 2015, pp. 375–377.

[74] B. Schneider and P. Blikstein, "Unraveling Students' interaction around a tangible interface using multimodal learning analytics," *J. Educ. Data Mining*, vol. 7, no. 3, pp. 89–116, 2015.

[75] D. Spikol, E. Ruffaldi, G. Dabisias, and M. Cukurova, "Supervised machine learning in multimodal learning analytics for estimating success in project-based learning," *J. Comput. Assist. Learn.*, vol. 34, no. 4, pp. 366–377, Aug. 2018.

[76] B. Roman, E. Shahram, and V. Hana, "A computational approach for prediction of problem-solving behavior using support vector machines and eye-tracking data," in *Eye Gaze in Intelligent User Interfaces*. London, U.K.: Springer, 2013, pp. 111–134.

[77] B. Kalantar, B. Pradhan, S. A. Naghibi, A. Motevalli, and S. Mansor, "Assessment of the effects of training data selection on the landslide susceptibility mapping: A comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)," *Geomatics, Natural Hazards Risk*, vol. 9, no. 1, pp. 49–69, 2018.

[78] B. Robert and D. Sidney, "ACM UMAP," in *Proc. Int. Conf. User Modeling, Adaptation, Personalization*, vol. 9146, 2015, pp. 31–43.

[79] E. Ashley, M. Anthony, S. Srinivas, G. Joe, W. Linwei, B. Reynold, and A. Cecilia, "Sensor-based methodological observations for studying online learning," in *Proc. SmartLearn*, 2017, pp. 25–30.

[80] S. Kardan and C. Conati, "Exploring gaze data for determining user learning with an interactive simulation," in *User Modeling, Adaptation, and Personalization* (Lecture Notes in Computer Science), vol. 7379. Berlin, Germany: Springer, 2012, pp. 126–138.

[81] K. Mangaroska, K. Sharma, M. Giannakos, H. Traeteberga, P. Dillenbourg, and P. Dillenbourg, "Gaze-driven design insights to amplify debugging skills: A learner-centred analysis approach," *J. Learn. Anal.*, vol. 5, no. 3, pp. 98–119, 2018.

[82] P. L. Prieto, K. Sharma, P. Dillenbourg, and M. Jesús, "Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors," in *Proc. LAK*, 2016, pp. 148–157.

[83] K. Sharma, Z. Papamitsiou, and M. Giannakos, "Building pipelines for educational data using AI and multimodal analytics: A 'grey-box' approach," *Brit. J. Educ. Technol.*, vol. 50, no. 6, pp. 3004–3031, 2019.

[84] R. Zemblys, D. C. Niehorster, O. Komogortsev, and K. Holmqvist, "Using machine learning to detect events in eye-tracking data," *Behav. Res. Methods*, vol. 50, no. 1, pp. 160–181, Feb. 2018.

[85] K. Sharma, M. Giannakos, and P. Dillenbourg, "Eye-tracking and artificial intelligence to enhance motivation and learning," *Smart Learn. Environments*, vol. 7, no. 1, pp. 1–19, Dec. 2020.

[86] M. Faber, R. Bixler, and S. K. D'Mello, "An automated behavioral measure of mind wandering during computerized reading," *Behav. Res. Methods*, vol. 50, no. 1, pp. 134–150, Feb. 2018.

[87] X. Ochoa and M. Worsley, "Augmenting learning analytics with multimodal sensory data," *J. Learn. Anal.*, vol. 3, no. 2, pp. 213–219, 2016.

[88] K. Huang, T. Bryant, and B. Schneider, "Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data," in *Proc. Int. Conf. Educ. Data Mining Soc. (EDM)*, Eric, 2019, pp. 318–323.

[89] C.-M. Huang, S. Andrist, A. Sauppé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Frontiers Psychol.*, vol. 6, p. 1049, Jul. 2015.

[90] J. B. Linden, A. Sasse, C. Sas, T. C. Ormerod, A. Dix, P. Bagnall, and T. McEwan, "People computer XXI-HCI.. but not as we know it," in *Proc. HCI*, vol. 1, 2007, pp. 77–86.

[91] A. S. Robert, G. Arthur, H. Xiangen, and H. Heather, *Design Recommendations for Intelligent Tutoring Systems: Learner Modeling*, vol. 1. Adelphi, MD, USA: US Army Research Laboratory, 2013.

[92] C. Cristina, A. Vincent, and M. Antonija, "Eye-tracking for Student modelling in intelligent tutoring systems," in *Design Recommendations for Intelligent Tutoring Systems*, vol. 1. Adelphi, MD, USA: US Army Research Laboratory, 2013, pp. 227–236.

[93] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "DAiSEE: Towards user engagement recognition in the wild," 2016, *arXiv:1609.01885*. [Online]. Available: http://arxiv.org/abs/1609.01885

[94] S. Poria, E. Cambria, A. Hussain, and G. B. Huang, "Towards an intelligent framework for multimodal affective data analysis," *Neural Netw.*, vol. 63, pp. 104–116, Mar. 2015.

[95] M. Ivanović, A. K. Milićević, V. Aleksić, B. Bratić, and M. Mandić, "Experiences and perspectives of technology-enhanced learning and teaching in higher education–serbian case," *Proc. Comput. Sci.*, vol. 126, pp. 1351–1359, Mar. 2018.

[96] J. Brosens, F. Adebesin, and R. Kruger, "In the eye of the beholder: Teaching user-centered design to information and communication technology Students with the help of eye tracking," in *Handbook of Research on Diverse Teaching Strategies for the Technology-Rich Classroom*. Hershey, PA, USA: IGI Global, 2020, pp. 296–318.

[97] P. Kasprowski and J. Ober, "Eye movements in biometrics," in *Proc. Int. Workshop Biometric Authentication*. Berlin, Germany: Springer, 2004, pp. 248–258.

[98] V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio, "GANT: Gaze analysis technique for human identification," *Pattern Recognit.*, vol. 48, no. 4, pp. 1027–1038, Apr. 2015.

YUEHUA WANG (Member, IEEE) is currently an Assistant Professor in computer science with Texas A&M University–Commerce. Her research interests include artificial intelligence, human–machine interaction, software programming and engineering, vehicular sensing and control, vehicle communication systems, resource virtualization, cyber-physical systems, autonomous driving, the Internet of Things, ubiquitous, and mobile computing.

SHULAN LU received the Ph.D. degree in cognitive science from The University of Memphis. She is currently a Professor in psychology with Texas A&M University–Commerce. Her research interests include embodied cognition, learning and instruction, human centered computing, and event perception.

DEREK HARTER (Member, IEEE) is currently an Associate Professor of Computer Science at Texas A&M University–Commerce. His primary research interests are in machine learning and deep learning applied to human cognition and instruction. He also has research interests in HCI, virtual reality, and augmented reality.

• • •