Applying Computer Vision and Machine Learning Techniques in STEM-Education Self-Study

Rustam Abdrahmanov¹, Assyl Tuimebayev², Botagoz Zhussipbek³, Kalmurat Utebayev⁴, Venera Nakhipova⁵, Oichagul Alchimbayeva⁶, Gulfairuz Makhanova⁷, Olzhas Kazhybayev⁸

International University of Tourism and Hospitality, Turkistan, Kazakhstan¹
Boston University, Boston, USA²
Korkyt Ata Kyzylorda University, Kyzylorda, Kazakhstan³, ⁷
M. Auezov South Kazakhstan University, Shymkent, Kazakhstan⁴, ⁶
South Kazakhstan Pedagogical University named after O. Zhanibekov, Shymkent, Kazakhstan⁵
Astana IT University, Astana, Kazakhstan⁸

Abstract—In this innovative exploration, "Applying Computer Vision Techniques in STEM-Education Self-Study," the research delves into the transformative intersection of advanced computer vision (CV) technologies and self-directed learning within Science, Technology, Engineering, and Mathematics (STEM) education. Challenging traditional educational paradigms, this study posits that sophisticated CV algorithms, when judiciously integrated with modern educational frameworks, can profoundly augment the efficacy of self-study models for students navigating the increasingly intricate STEM curricula. By leveraging state-of-the-art facial recognition, object detection, and pattern analysis, the study underscores how CV can monitor, analyze, and thereby enhance students' engagement and interaction with digital content, a pioneering stride that addresses the prevalent disconnect between static study materials and the dynamic nature of learner engagement. Furthermore, the research illuminates the critical role of CV in generating personalized study roadmaps, effectively responding to individual learner's behavioral patterns and cognitive absorption rhythms, identified through meticulous analysis of captured visual data, thereby transcending the one-size-fits-all educational approach. Through rigorous qualitative and quantitative research methods, the paper offers groundbreaking insights into students' study habits, proclivities, and the nuanced obstacles they face, facilitating the creation of responsive, adaptive, and deeply personalized learning experiences. Conclusively, this research serves as a clarion call to educators, technologists, and policy-makers, emphatically demonstrating that the thoughtful application of computer vision techniques not only catalyzes a more engaging self-study landscape but also holds the latent potential to revolutionize the holistic STEM education ecosystem.

Keywords—Load balancing; machine learning; server; classification; software

I. INTRODUCTION

In the rapidly evolving educational landscape, traditional teaching methodologies are incessantly being re-evaluated and challenged, particularly in Science, Technology, Engineering, and Mathematics (STEM) disciplines. The advent of digital technology has reshaped pedagogical strategies, heralding new approaches like self-directed learning, which has gained prominence for fostering students' autonomy and responsibility in the learning process [1]. However, maximizing the efficacy of self-study in STEM education requires addressing intrinsic complexities and diverse student engagement methodologies [2]. This research aims to bridge this gap by harnessing computer vision (CV) techniques, offering a transformative perspective on enhancing self-study's effectiveness in STEM education.

STEM fields, inherently multifaceted and dynamic, demand educational approaches that not only convey complex concepts but also adapt to individual cognitive styles and paces [3]. Traditional self-study, while offering flexibility, often falls short of this adaptability, leading to learner frustration and sub-optimal learning outcomes [4]. Computer vision's potential in education, particularly in monitoring and responding to student engagement and facilitating personalized learning trajectories, remains largely underexplored [5].

Computer vision, a field that grants computers a high-level understanding of digital images and videos, is traditionally aligned with applications in security, surveillance, and detection [6-7]. However, its implications extend profoundly into educational realms. Through detailed visual data analysis, CV holds the promise of decoding student engagement patterns, providing educators with nuanced insights into the often imperceptible aspects of self-study behaviors that either catalyze or hinder learning. This research pivots around the innovative application of CV in capturing and analyzing these intricate behavioral nuances, thereby guiding the development of more responsive and adaptive self-directed learning models.

Integrating CV into education, especially within STEM, poses unique challenges and opportunities. The precision required in STEM subjects translates to the necessity for educational resources to adapt in real-time to students' understanding, ensuring concepts are neither misinterpreted nor oversimplified [8]. By employing CV techniques, such as facial recognition and object detection, it becomes feasible to analyze students' interaction with educational content, thereby tailoring materials and study paths that resonate with individual learning approaches, an advancement far beyond the capabilities of traditional educational software [9].

Moreover, the role of CV in tracking and analyzing engagement brings a new dimension to educators' understanding of student performance. Conventional assessment methods offer only summative feedback, often...
neglecting the formative aspects of a learner's journey [10]. Through continuous and non-intrusive monitoring, CV provides a wealth of formative feedback, empowering educators to make informed, timely interventions and students to gain awareness of their learning habits.

Significantly, the ethical considerations of utilizing CV in education are paramount, entailing careful navigation. Privacy concerns, data security, and the consent of the involved parties are crucial factors that educators and technologists must prioritize. Establishing robust ethical protocols and transparent operational guidelines ensures the responsible use of CV in educational settings, safeguarding participants' fundamental rights while harnessing technology's benefits [11].

In light of the above, this study ventures into an interdisciplinary examination of how CV can revolutionize self-study within STEM education. It builds on existing literature that outlines the theoretical frameworks of self-directed learning and delves into empirical evidence supporting the integration of advanced technologies in education [12]. By establishing a symbiotic relationship between CV technology and educational pedagogy, this research underscores a forward-thinking approach to cultivating STEM competencies, proposing a model that respects individual cognitive differences and celebrates personalized educational journeys.

Through this paper, we invite educators, technologists, and policy-makers to envision a future where technology and education converge to offer enriched, student-centric learning experiences. By navigating the technical, pedagogical, and ethical terrains of this integration, we aim to construct a comprehensive understanding that could fundamentally transform the way STEM education perceives and leverages self-study.

II. RELATED WORKS

The synthesis of technology and education, especially in self-directed STEM learning, has instigated a plethora of research, with various studies corroborating the transformative potential of integrating advanced technological frameworks, such as computer vision (CV), into educational models. These scholarly pursuits, encompassing a diverse range of insights and findings, lay the groundwork for understanding the trajectory and implications of utilizing CV in self-regulated learning environments.

Starting with the broader impacts of technology in education, studies have indicated a paradigm shift in instructional strategies, emphasizing the need for more learner-centric approaches facilitated by technology [13]. In the context of STEM education, researchers have highlighted the necessity for innovative methods that cater to enhancing students' critical thinking and problem-solving skills, proposing that digital technologies can bridge the gap between theoretical knowledge and practical application [14].

The concept of adaptive learning, pivotal to this discussion, leverages technology to tailor educational experiences to individual needs. One study [15] provides insights into adaptive learning systems' role in promoting cognitive growth, arguing that these systems accommodate diverse learners' profiles, thereby fostering a more inclusive learning environment. However, the challenge remains in effectively tracking and interpreting individual learner interactions and responses in real-time, a gap that computer vision promises to address [16].

Computer vision’s foray into educational strategies marks a relatively new venture. Its application has been predominantly explored in surveillance, recognition systems, and user interaction tracking in various sectors [17]. Within educational research, studies have often circumscribed their focus to online learning environments, utilizing simple CV techniques for user log-in and basic interaction [18]. However, more nuanced applications of CV, such as emotion recognition, behavioral analysis, and engagement tracking, are emerging themes in contemporary literature [19].

Next study [20] delve into the potential of CV in recognizing and interpreting human emotions, an aspect crucial for personalizing learning experiences. They argue that emotional states play a significant role in learning efficiency, with certain emotional conditions favoring the absorption and retention of new information. Incorporating CV into educational platforms could thus allow for real-time adaptation based on learners' emotional cues, providing immediate feedback or altering content presentation to enhance learning efficacy [21].

Furthering the discourse on personalized learning, researchers have explored data-driven approaches. For example, [22] highlights the importance of learning analytics in understanding students' learning processes. They discuss how data obtained from students' digital footprints on learning platforms can inform more personalized and effective teaching strategies. This data-centric approach aligns with the capabilities of CV in capturing and analyzing extensive datasets of learner interactions and behaviors [23].

In the realm of self-directed learning, especially in online and digital contexts, maintaining student engagement and motivation is paramount. Studies by Chen, Lambert, and Guidry [24] underscore the challenge educators face in keeping students engaged with digital platforms. CV's potential to monitor visual cues and physical responses presents unprecedented opportunities for understanding and enhancing student engagement on a much finer, more personalized scale [25].

The integration of CV in education also extends to practical skill-based learning in STEM. For instance, research on laboratory learning indicates that CV can significantly enhance remote laboratory experiences, a critical component of STEM education. Gravier, Fayolle, Bayard, Ates, and Lardon [26] have explored these prospects, emphasizing that CV can facilitate more interactive and hands-on experiences in a virtual environment.

Despite the promising advancements, the literature consistently echoes the ethical implications of employing CV in educational settings. Privacy concerns, particularly with facial recognition and behavioral tracking, are prevalent [27]. Next research [28] stresses the need for robust privacy protection frameworks, emphasizing informed consent, data security, and transparency in how monitoring technologies are used in education. These considerations are crucial in ensuring
the ethical integrity of integrating any form of surveillance or tracking technology into learning environments.

Conclusively, the body of work surrounding the integration of computer vision in education outlines a landscape ripe with potential yet requires careful navigation concerning ethical, technical, and pedagogical constraints. This research contributes to this ongoing scholarly dialogue, contextualizing the application of CV within the specific challenges and opportunities presented by self-directed STEM education [29-32]. Through an interdisciplinary lens, this study seeks to build upon the foundations laid by existing literature [33-37], proposing an innovative convergence of CV technology and educational pedagogy to enhance the quality and effectiveness of self-study in STEM disciplines.

III. MATERIALS AND METHODS

In order to investigate the central research query, several "STEM Workshops: Python and Raspberry Pi Practical Activity" were organized as a precursor to the main experimental procedure. These preliminary sessions were instrumental in gathering the necessary data for the creation and subsequent validation of the RASEDS, directly contributing to the resolution of the initial research query. Once the efficacy of RASEDS was confirmed, the data derived from the system were harnessed to develop a predictive model for student performance in STEM subjects. Subsequently, this predictive mechanism was designed to suggest customizable learning resources, tailored to forecasted performance trends.

Our research adopted a quasi-experimental design to ascertain whether introducing personalized educational resources, recommended through RASEDS, could significantly improve student involvement and confidence in STEM-related tasks, thereby providing comprehensive answers to the second and third research inquiries. The sequence of research activities is graphically represented in Fig. 1.

To meticulously record the nuances of each learner's practical engagement, we strategically installed cameras to film their hands-on interaction with the educational materials, a critical component for the RASEDS's engagement detection mechanism. Care was taken in choosing camera perspectives that would clearly record the learners' hands and the instructional tools they used. Mindful of the ethical considerations when filming individuals, especially those underage, we established rigorous measures to secure informed consent from all attendees or their legal guardians (for those younger than 18). This measure was pivotal in maintaining ethical standards concerning the visual content that included identifiable participant imagery.

In the aftermath of these sessions, we collected 4,515 photographs. These were methodically divided into primary and secondary datasets, following an 80:20 split. Consequently, we allocated 3,612 photographs for initial training purposes and reserved 903 as a subsequent test collection. These images are integral to the training phase of the YOLOR model, representing a significant stride towards achieving our research's overarching goals.

![Flowchart of the proposed system](image-url)
Following the workshop's end, the participants showcased their STEM initiatives, firmly rooted in the Internet of Things (IoT) sphere. Utilizing the insights gained about sensors and coding principles throughout the workshop, the students embarked on devising creative approaches to tangible issues through the application of IoT. Their ventures spanned various concepts, from intelligent domestic setups and energy-saving configurations to automated methods promoting greener lifestyles and operational spaces.

Upon the conclusion of the project presentations, each was subjected to a thorough analysis conducted by two connoisseurs within the STEM domain. The assessment protocol was grounded in the principles specified by the Creative Product Analysis Matrix (CPAM) approach, involving three broad categories and nine evaluative markers, elaborately itemized in Table I. This technique of appraisal, validated in its efficacy by Besemer in 1998, guaranteed an exhaustive and precise examination of the students' endeavors.

| TABLE I. DILOG TESTS IN CLASSROOM |
|---|---|
| Scale | Indicator |
| Novelty | Original |
| | Amazing |
| Resolution | Valuable |
| | Useful |
| | Understandable |
| Synthesis | Organic |
| | Elegant |
| | Good |

Evaluation was carried out utilizing a five-point Likert scale, enabling a detailed interpretation of each project's merits and areas for improvement. The consistency of scoring between the two experts was confirmed, with a correlation coefficient marking between 0.68 and 0.84. This high degree of concordance underscored the substantial agreement in their assessments, bolstering the integrity of the evaluation phase. Such a metric reinforced the consistency and trustworthiness of the ratings given, laying a dependable groundwork for the predictive model of STEM learning outcomes.

In response to the intricate and ever-evolving facets of STEM activity-based learning, we pioneered a system known as the Real-time Automated STEM Engagement Detection System. This system is designed to autonomously and instantaneously gauge students' engagement levels. At its core, RASEDS utilizes cutting-edge object detection, particularly the YOLOR method, to pinpoint the presence of students' hands and all associated educational materials engaged during the tasks. This interaction between the students' hands and the educational tools is documented, reflecting direct insights into the students' immediate actions. These consequential behaviors are then aligned with the parameters set by the ICAP framework, serving as a robust metric for evaluating student engagement throughout STEM-centric tasks.

This study engages with the SHAP (SHapley Additive exPlanations) methodology, an advanced technique within the realm of interpretable artificial intelligence, to critically analyze the contributory features inherent in the academic performance prediction model. Concurrently, an intriguing observation emerges from the C1 cohort, exhibiting a marginal enhancement in predictive accuracy relative to the established baseline, which is preliminarily set at 50%. This nuanced increment, albeit minimal, signals a critical inference: the interactive dynamics encapsulated within the online classroom environment exert a relatively insubstantial influence on the academic trajectories associated with non-STEM coursework. This revelation underscores the necessity for a differential pedagogical approach, potentially customized to the distinct educational exigencies of STEM and non-STEM curricula.

IV. EXPERIMENTAL RESULTS

A. Evaluation of Emotional Expression of Students

We implemented a quasi-experimental approach to investigate the impact of using RASEDS for recommending adaptive learning materials in STEM education, particularly in enhancing student engagement and self-confidence. This experiment was integrated into the 'Networks Embedded System and Application' course, spanning two academic terms. While students undertook the course independently, collaboration and dialogue were encouraged during the project development phase. Emphasizing IoT and AI, the course required students to leverage their understanding of both software and hardware to devise solutions for real-world challenges, thereby resonating with the fundamental tenets of STEM education (see Fig. 2).

The experimental phase of the study was meticulously structured and spanned duration of seven weeks. This phase was critically segmented into two distinct assessment periods, wherein participants from diverse groups were engaged in comprehensive evaluations. The primary objective of these assessments was to ascertain and quantify two fundamental dimensions: the degree of participant involvement and the perception of personal competence.

At the outset of the experimental phase, in the first week, an initial assessment was conducted. This preliminary evaluation served as a baseline measurement, establishing the initial state of participant engagement and their self-assessed competence. This was crucial for providing a reference point against which any subsequent changes could be measured. The initial assessment was designed to be comprehensive, ensuring that all relevant aspects of involvement and personal competence were adequately captured.

As the program progressed, participants continued their engagement in the designed activities and interventions. This progression was systematically documented and is visually represented in Fig. 3 of the study. The figure illustrates the temporal flow of the program, marking key milestones and the transition from the initial to the final stages.
Fig. 2. STEM education process in Application

Fig. 3. Pretest and posttest engagements of self-efficacy.
In the concluding phase of the program, during the fifth week, a final assessment was conducted. This assessment mirrored the initial one in structure but aimed to capture the evolved state of participant involvement and competence. The comparison between the initial and final assessments was pivotal in determining the effectiveness of the program. It enabled the researchers to quantify the changes in the levels of involvement and personal competence, attributing these changes to the interventions and activities experienced by the participants.

In summary, the experimental phase, with its well-defined and strategically placed assessments, provided a robust framework for evaluating the impact of the program on participant involvement and personal competence. The assessments, anchored at the beginning and end of the program, offered critical insights into the developmental trajectory of the participants, as detailed in Fig. 3.

The outcomes derived from the confusion matrix, presented in Fig. 4, enable the computation of precision, recall, and F1 score for each category of engagement as detected by RASEDS. In the realm of machine learning model evaluation, a confusion matrix serves as a pivotal tool, offering nuanced insights into the classification prowess of a model across various categories [38]. The presented matrix delineates the performance of a classifier in segregating data into five distinct classes: Interactive, Constructive, Active, Passive, and Other. The matrix’s structure, with rows representing actual classes and columns depicting predicted classes, provides a comprehensive view of both the model’s accuracy and its errors in classification.

A closer inspection of the matrix reveals intricate details about the model’s performance. For the 'Interactive' class, there is a prominent diagonal element of 81, indicative of a high rate of correct predictions. However, there is a noticeable misclassification with the 'Passive' and 'Constructive' categories, as evidenced by the presence of 8 and 4 instances, respectively, in these columns. This pattern suggests a certain level of ambiguity or overlap in the defining characteristics of these classes as interpreted by the model.

The 'Constructive' class exhibits impressive prediction accuracy, with 89 instances correctly classified. Nonetheless, there are marginal confusions with the 'Interactive', 'Active', and 'Passive' classes, albeit to a lesser extent than observed in the 'Interactive' class. This points to a generally robust model performance in this category, with room for improvement in differentiating finer nuances between certain classes.

For the 'Active' and 'Passive' classes, the model demonstrates commendable predictive accuracy, as indicated by 91 and 87 correct predictions, respectively. Misclassifications in these categories are relatively lower, suggesting that the model effectively captures the distinct features of these classes. The 'Other' category, with a high correct prediction count of 91, confirms the model’s capacity to accurately identify instances that do not conform to the primary classes.

In sum, the confusion matrix provides an invaluable quantitative assessment of the model’s classification abilities, highlighting areas of strength and pinpointing aspects that warrant further refinement [39]. Through this detailed analysis, researchers can gain a profound understanding of the model’s behavior across varied classifications, guiding targeted improvements in its predictive accuracy.

Intricately woven into this analysis are six pivotal variables, each derived from a comprehensive aggregation of the absolute values of corresponding interactive or emotional metrics within a specific interactive phase. For instance, the variable 'summary_interaction' is computed by summing the absolute SHAP values of various interactive categories during the summary stage, represented formulaically as:

\[ \text{summary_interaction} = |ics| + |ims| + |ios| + |ccs| + |cms| + |cos| \]

Analogously, ‘summary_emotion’ encapsulates the emotional undertones of the summary phase, calculated as:

\[ \text{summary_emotion} = |ips| + |cps| + |ins| + |cns| \]

V. DISCUSSION

In this study, a comprehensive literature review was conducted, focusing on the application of machine learning and computer vision techniques across various domains, as cited in References [40-42]. These techniques have been noted for their diverse utility, ranging from healthcare to educational applications. Building on this foundation, the current study specifically applies machine learning and computer vision methods within the realm of STEM education, aiming to explore and expand the educational potential of these innovative technologies.

In the context of STEM education, the integration of computer vision and machine learning (ML) offers transformative potential [43-45]. This research paper has explored how these technologies can be leveraged to enhance self-study methodologies in STEM subjects, with a focus on personalized learning, engagement, and improved learning outcomes.
Personalization of Learning: One of the most significant contributions of ML in STEM education is the ability to tailor educational content to individual students' needs. By analyzing student performance and learning behaviors, ML algorithms can adaptively modify the curriculum, presenting topics in a manner that aligns with each student's unique learning style and pace. This personalization is crucial in self-study environments, where learners often lack the direct guidance of an instructor.

Enhanced Engagement through Computer Vision: The application of computer vision in educational tools has been shown to increase student engagement. By incorporating interactive visual elements and real-time feedback systems, computer vision can make abstract STEM concepts more tangible and comprehensible. This visual interactivity is particularly effective in self-study scenarios, keeping students motivated and engaged in the absence of traditional classroom dynamics.

Data-Driven Insights: ML algorithms provide valuable insights into student learning patterns, identifying areas of difficulty and success. This data can inform the design of future educational content, ensuring that it addresses common challenges and reinforces key concepts. In self-study, these insights become crucial for students to monitor their progress and for educators to understand the efficacy of the learning material.

Overcoming Challenges: Despite the advantages, the integration of computer vision and ML in STEM education is not without challenges. Concerns regarding data privacy, the digital divide, and the need for robust and unbiased algorithms are paramount. Ensuring that these technologies are accessible and equitable for all students, regardless of background or resources, is essential for their successful implementation in educational settings.

Future Directions: Looking forward, the continued development and refinement of ML and computer vision technologies promise even greater advancements in STEM education. The potential integration of augmented reality (AR) and virtual reality (VR) technologies, combined with ML-driven personalized learning paths, could revolutionize the way STEM subjects are taught and learned. Additionally, the ongoing improvement of algorithmic transparency and fairness will be crucial in ensuring that these technologies serve all students effectively.

In conclusion, the application of computer vision and machine learning in STEM-education self-study represents a significant step forward in educational technology. These tools offer the potential for highly personalized, engaging, and effective learning experiences. However, careful attention must be paid to the challenges and ethical considerations that come with the implementation of such advanced technologies. As the field progresses, continuous evaluation and adaptation will be necessary to fully realize the benefits of these innovations in STEM education.

VI. CONCLUSION

This study's journey into the realms of advanced technology's application within educational settings, particularly through the Real-time Automated STEM Engagement Detection System (RASEDS), unveils new horizons in STEM-related pedagogies. The evidence presented underscores the potential of such innovative intersections between technology and education, where systems like RASEDS are not mere analytical tools but catalysts for transformative educational experiences. The ability of RASEDS to discern engagement levels accurately heralds a future where learning can be genuinely individualized, responding in real-time to students' engagement fluctuations. Moreover, the observed enhancement in self-efficacy among learners signals a profound impact on learners’ psychological resources, potentially influencing their academic trajectories and career paths in STEM fields.

However, the journey does not conclude here. While the findings affirm the positive trajectories, they also cast light on the complexities and multi-dimensional challenges within technology-integrated education. Future research needs to navigate these sophisticated dynamics, including the nuanced understanding of engagement and self-efficacy, ethical considerations surrounding data security, and the psychological safety of learners. Furthermore, pedagogical strategies must evolve in tandem with these technological advancements, ensuring that human-centric learning remains at the core of educational endeavors. As we stand on the brink of this new era, the responsibility is collective—educators, technologists, policymakers, and researchers must collaborate to ensure these innovations are harnessed responsibly, ethically, and with the holistic development of learners in mind.

REFERENCES


