EPiC Series in Built Environment

Volume 5, 2024, Pages 175–183

Proceedings of 60th Annual Associated Schools of Construction International Conference

Automated Gaze Recognition within a Sensor Data Analytics Platform for Construction Education

Mohammad Khalid, Abiola Akanmu, Ph.D., Adedeji Afolabi, Ph.D., and Homero Murzi, Ph.D.
Virginia Tech
Blacksburg, Virginia

Ibukun Awolusi, Ph.D.
University of Texas at San Antonio
San Antonio, Texas

The construction industry is increasingly harnessing sensing technologies to overcome manual data collection limitations and address the need for advanced data analysis. This places an aggravated demand for associated skills to interpret sensor data. Yet, a substantial gap exists between the level of academic preparation and the actual needs of the industry, leading to an underprepared workforce. In this study, ActionSens, a Block-Based Programming Environment, is implemented as an educational tool that combines sensor data from Inertial Measurement Units with machine learning algorithms. This integration enables the classification of construction activities, offering construction students a platform to explore and learn about sensor data analytics. However, in a pedagogical setting, an enhanced learning experience can be achieved through the integration of automated classification models that intelligently detect learners' focus with the potential to provide context-specific support. This study utilizes 19 construction students' eye-tracking data to train and evaluate machine learning models to detect learners' visual focus on specific Areas of Interest within ActionSens. Ensemble, Neural Network, and K-Nearest Neighbor performed the best for both raw and SMOTE-oversampled datasets. The Ensemble had an edge in recognizing Areas of Interest, achieving top precision, recall, F1-score, and AUC in the oversampled data.

Key Words: Machine Learning, Block Programming, Sensor Data, Data Analytics, Eye-tracking.

Introduction

The construction sector is characterized by its dynamic and complex nature, demanding a wealth of information to be analyzed for decision-making and potential performance enhancement. Manual data collection methods hinder modern construction management practices. Therefore, the industry is increasingly adopting sensing technologies to enhance project performance through data-driven approaches, intensifying the demand for workforce skills in effective sensor data management (Khalid et al., 2023). For instance, analyzing Inertial Measurement Unit (IMU) data using machine learning (ML) in construction yields useful insights into worker and equipment states, enabling potential task modifications to address safety and productivity issues (Rashid & Louis, 2020). To immerse students in learning this process, it is imperative to embed the practical analytics processes within the educational platform. Recently, Block-Based Programming Environments (BBPEs) have proven effective in improving learners' data analytics-related skills (Bender, Dziena, & Kaiser, 2022; Tawfik, Payne, & Olney, 2022). However, BBPEs' integration into construction education remains uncharted, resulting in limited avenues to expose students to inquiries and discussions related to construction sensor data. This leaves aspiring graduates inadequately prepared to address authentic construction problems (Khalid et al., 2023). Lack of exposure to such pedagogical environments limits understanding of how BBPI can support learners' engagement which ultimately translates into higher-order learning of sensor data analytics skills. Additionally, in the learning process detecting where learners seek help in human-computer interactions (HCI) holds potential for enhancing overall learning experiences. In construction workforce development, utilizing eye-tracking technology to detect users' eye gaze locations allows for analyzing viewing behavior and provides the potential for personalized support to enhance performance (Jeelani, Han, & Albert, 2018). However, the analysis of eye-tracking data collected during real-world interactions presents a unique challenge, especially when it comes to automatically identifying where a person's gaze is directed within a specific environment, including the objects of their attention. To reinforce the gap with the potential to enhance sensor data analytics learning experience, this paper evaluates the performance of ML models with select gaze features that capitalize on the ability to detect 'where' users are looking at and spending visual attentional resources. The pedagogical BBPE of this study, ActionSens, integrates one of the growing trends in construction research involving cost-effective sensing technology (i.e., IMUs) and ML techniques to classify construction activities, providing students with the means to explore and learn. The study draws upon and contributes to the Aptitude-Treatment Interaction (ATI) theory, which underscores the significance of identifying users characteristics to tailor learning support efficiently. The results indicate that eye movement data contains valuable information, allowing for the automated classification of learners' focus on specific Areas of Interest (AOIs). These classification models hold the potential to recognize context-specific needs and provide support to learners, thereby facilitating the overall acquisition of technical skills in sensor data analytics.

Background

BBPEs in Education

End-user programming (EUP) stands out as a paradigm that enables end-users to create software programs without the requirement for traditional programming expertise. Unlike conventional programming methods, EUP leverages visual and programming by demonstration approaches, allowing users to utilize graphical components such as pages, buttons, links, rules, and database records to develop and iterate their applications (Barricelli, Cassano, Fogli, & Piccinno, 2019). BBPEs are a specific subset within the broader domains of EUP, employing a distinctive block-based, method for programming (Vincur, Konopka, Tvarozek, Hoang, & Navrat, 2017). BBPEs offer a visual interface with interactive blocks representing codes and domain-specific concepts, enabling users to easily grasp and execute complex tasks through simple drag-and-drop functionalities (Barricelli et al., 2019).

Eye-tracking and Machine Learning for Usability

The increasing popularity of BBPEs in domain-specific learning underscores the growing importance of content and usability evaluations for successful instructional outcomes and broad acceptance (Rijo-García, Segredo, & León, 2022). While usability surveys and interviews provide subjective insights, eye tracking offers objective data by tracking users' visual attention, cognitive processes, and decision-making (Dilmen, Kert, & Uğraş, 2023). For example, Dilmen et al. (2023) investigated the efficiency of a BBPE using eye-tracking data to assess task completion times, step counts, and users' focus in guided areas. The eye-tracking data indicated that participants' attention on the system's explanatory field intensifies when they make errors during task execution. Therefore, the study suggested that providing prompt and effective feedback is a crucial element for these BBPE platforms. Tawfik et al. (2022) suggested employing eye-tracking to predict user interactions and help-seeking attributes, aiding designers in refining the learning environment. Additionally, Vincur et al. (2017) highlighted eye tracking for analyzing learners' behavior, while Shah, Ali, Dieker, and Hughes (2023) recommended using ML to detect user aptitudes and provide tailored support based on their activities and emotions. Hence, recognizing interaction variations within specific contexts and focus locations is imperative to facilitate the design of effective instructional settings for situational learning aids. By harnessing eye-tracking technology with ML techniques, learning environments can be integrated with classification features to detect learners' visual focus with the potential to enhance learning performance.

Methodology

This section explains the methodology employed in this research as depicted in Figure 1. Eye-tracking data was collected during the experimental tasks and subsequently, gaze data were mapped onto the BBPE's interface screen. Specific AOIs were identified based on ground-truth information (i.e., key elements of the interface). The collected eye-tracking information was pre-processed and employed in the training of classification models.

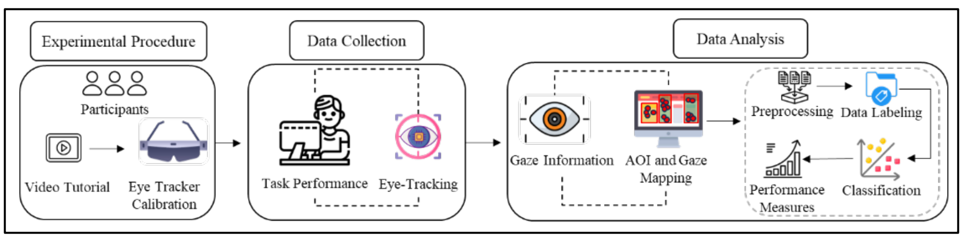

Figure 1. Research methodology overview (icon source- freepik.com).

Experimental Procedure

Participants

To align with the pedagogical goals of the platform, only participants majoring in construction engineering and management, civil engineering, and building construction were recruited through flyers and email invitations. Nineteen eligible participants, comprising 10 males and 9 females, participated in the usability study in exchange for a $25 Amazon gift card as compensation for their time and effort. IRB-approved informed consent was obtained from all participants.

Apparatus

Before taking part in the experiment, all participants were provided with instructional materials containing information about the task procedures, essential components, and platform features. Participants' eye-tracking equipment (Tobii Pro Glasses 3) was prepared by cleaning and adjusting it for a proper fit using the appropriate nasal bridge. Calibration procedures were completed before data recording. A first-person view with pupil fixation overlay was displayed on another computer during the experiment for real-time monitoring. Data was securely transferred to a research computer for Tobii Glasses 3 software analysis after each experiment.

Analytics Platform and Activity

ActionSens aims to provide students with a data-driven learning experience, ultimately enhancing their understanding of sensor data analytics. To facilitate this, ActionSens includes features such as an analytics visualizer for dataset comprehension, a video playback tool for exploring underlying relationships, a code generator, and a primary block workspace with a block menu for data manipulation and subsequent ML model training and evaluation (see Figure 2).

Figure 2. ActionSens interface (left) and participant interacting during usability session (right).

Participants were given a practical demonstration upon arrival to familiarize them with the platform processes, unprocessed sensor data, and construction activity videos (see Figure 3). This aimed to ensure their understanding of the task procedures. Their tasks involved interacting with pre-recorded construction data, including videos and raw IMU data, captured during a simulated construction activity. The main focus of the data analytics tasks was for participants to work with raw sensor data, train multiple machine learning models for classification, and assess prediction performance using confusion matrices and comparisons between predicted and actual data.

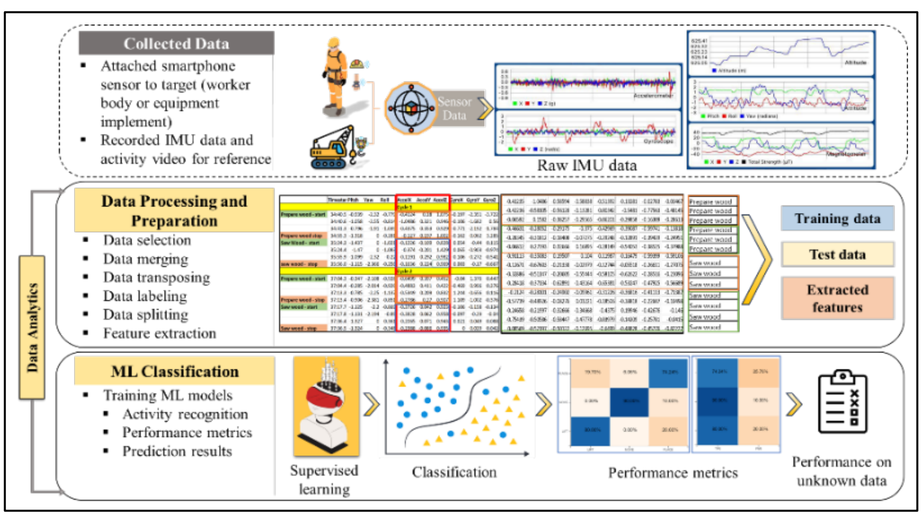

Figure 3. Sensor data analytics workflow overview (icon source- freepik.com).

Data Collection

Eye-tracking Data

The Tobi Pro Glasses 3 eye-tracking system operates at a refresh rate of 50-100 Hz, providing an accuracy of 0.6 degrees. It also includes a scene camera gaze overlay with a resolution of 1920×1080 and a field of view of 95∘×63∘, running at 25Hz. Output data includes gaze position (gaze2D) on the screen meaning when the subject looks at the top left or bottom right corner of the scene video, the gaze2D variable outputs (0, 0). Fixations were classified using the Tobii I-VT Fixation Filter.

Data Analysis

Pre-processing

Twelve key AOIs were identified where interaction behavior could be of significance for formulating pedagogical support for learners (see Figure 4). The block workspace, block menu, analytics visualizer, and code generator served as the main AOIs. The block workspace was subdivided into 8 AOIs for precise gaze analysis during tasks. An 'AOI NONE' category captured instances when participants did not focus on any of the 12 AOIs. Preprocessing included utilizing Tobii Pro Lab's assisted mapping, AOI, and metrics tools. Specific AOIs were defined, and metrics, including mapped fixation points, and gaze durations were extracted. Gaze locations were initially mapped with assisted mapping which utilizes computer vision algorithms. A total of four gaze metrics (x and y-axis fixation points, Euclidean distance, and gaze event duration) were extracted for training classification models.

Data Labeling

The raw data from Tobii ProLab comes in a spreadsheet format that can be filtered for each AOI hit which reports: 0 for AOI active with fixation outside, 1 for AOI active with fixation inside, and empty cells for invisible AOIs. To identify the specific area of the interface users focused on, 'ground-truth' data values were employed for labeling. These categorized user data into 12 different AOIs as mentioned earlier. Gaze data falling outside these AOIs was captured under 'AOI NONE.'

Figure 4. ActionSens interface with on-screen AOIs.

Data Classification and Data Sampling

This research utilized all classifiers available in MATLAB including K-NN, NN, Ensemble, Support Vector Machine (SVM), Naïve Bayes, Decision Tree, Logistic Regression, and Kernel. The default hyperparameter presets in MATLAB were used to train the models. The training employed five-fold cross-validation, a technique that splits the training dataset into five subsets. During each iteration, some data were used for training and others were reserved for testing the model's performance (Bednarik, Vrzakova, & Hradis, 2012). Extracted 28 features for model training were constituted of 7 statistical measures (mean, median, mode, standard deviation, max, min, sum) combined with the 4 eye-tracking metrics (Bednarik et al., 2012). To address the significant class variability, the synthetic minority oversampling technique (SMOTE) was utilized to oversample the training data (Lallé, Conati, & Carenini, 2016), with each minority class oversampled at individual rates to match the majority class, as described in Table 1 (*signifies the majority class). SMOTE addresses class imbalance by generating synthetic data samples within the feature space of the minority class through interpolation among existing instances.

Table 1

Data sample size for classes and SMOTE configuration

Classes	Raw	Sampling Rate	SMOTE
AOI 1	7067	1.936182256	13683
AOI 2	13683*	-	13683
AOI 3	1028	13.31031128	13683
AOI 4	3992	3.42760521	13683
AOI 5	1633	8.37905695	13683
AOI 6	2924	4.679548564	13683
AOI 7	1584	8.638257576	13683
AOI 8	1987	6.886260695	13683
AOI 9	1309	10.45301757	13683
AOI 10	2011	6.804077573	13683
AOI 11	3364	4.067479191	13683
AOI 12	341	40.12609971	13683
AOI NONE	13276	1.030656824	13683

Performance Measures

Five performance metrics were used, namely, accuracy, precision, recall, F1 score, and Area under the ROC Curve (AUC) across all 13 AOIs for evaluating the models' effectiveness in classifying and recognizing important AOIs within the dataset. Accuracy calculates the proportion of correct predictions, while recall measures true positives in relation to true positives and false negatives. Precision quantifies the model's correct positive predictions, and the F1-score combines precision and recall. AUC evaluates the classifier's discrimination power for positive and negative cases under different thresholds.

Results

Model Accuracy

The pre-processed data as mentioned above was trained for multiple models and Ensemble, NN, and K-NN, were among the top three performing models for both raw and the SMOTE-oversampled datasets (see Figure 5).

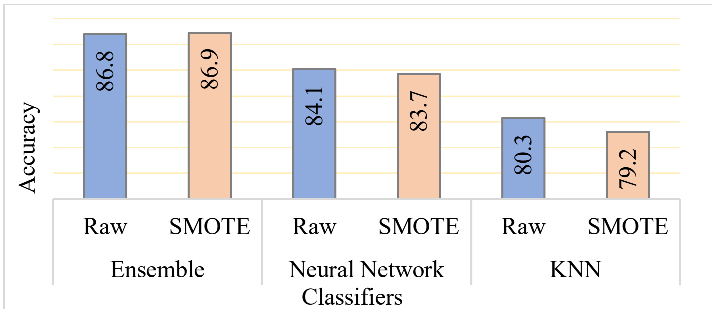

Figure 5. Accuracy comparison of trained models.

For both dataset sampling configurations, the Ensemble algorithm had the highest accuracy in recognizing the visual attention to specific AOIs which is supported by a similar study (Fathy, Mansour, Sabry, Refat, & Wagdy, 2023) that highlighted the effectiveness of Ensemble for classifying users visual attention in particular zones and performed the best in multiclass (i.e., 6 classes) classification. In total 154,583 out of 177,879 were correctly classified. However, for the second and third-best classifiers, NN and K-NN showed higher accuracy for the raw data compared to oversampled datasets.

Precision, recall, F-1 score, and AUC

In Figure 6, all models improved in precision, recall, and F1-score with oversampling, with one exception for AUC. Notably, the Ensemble algorithm achieved the highest precision (0.855), recall (0.858), F1-score (0.856), and AUC (0.981) for the oversampled dataset, outperforming the NN and K-NN models. While the Ensemble had a higher AUC with oversampling, the NN algorithm performed better with raw data in terms of AUC (0.968), while the K-NN algorithm showed consistent AUC values (0.963) across both conditions.

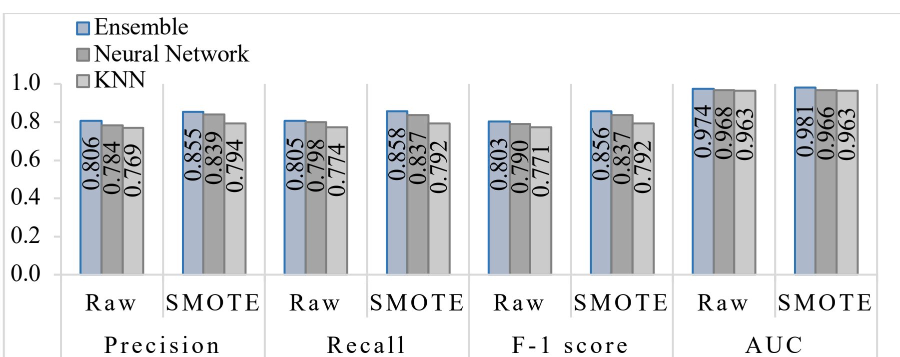

Figure 6. Precision, recall, F-1 score, and AUC comparison.

Discussion

As the construction industry embraces sensing technologies, there is a growing need for pedagogical platforms that prepare future graduates with the anticipated skills. Every student's unique attributes present opportunities to identify and address them, fostering engagement and success through personalized support. Thus, this study explores the use of eye-tracking and ML to analyze eye movements and detect focus in specific AOIs in a BBPE interface. The Ensemble classifier obtained the highest accuracy in correctly recognizing specific AOIs students looked at in oversampled data set configurations. For the oversampled data, Ensemble improved in all performance measures, while the other models reduced in accuracy. These trade-offs between performance while oversampling eye tracking data from HCI interaction have been reported by Lallé et al. (2016). Conversely, oversampling resulted in higher precision, recall, and F1-score than raw data across all three models, with Ensemble achieving the highest performance scores. The AUC score improved for Ensemble, remained constant for NN, and decreased for K-NN. Although performances varied across models and data sampling configurations, the AUC value revealed all models to be above 0.9 which indicates that adopted models are effective in classification (Li et al., 2020). One possible intervention involves identifying the AOIs participants are focusing on and delivering tailored hints related to that content. This could prevent overwhelming learners by avoiding the delivery of all information at once and assisting learners in obtaining specific solutions. Furthermore, this study also incorporated gaze data points outside the designated interface-based AOIs, capturing instances where participants wandered away from required areas during the task. The models' accurate recognition of data points suggests they could detect interface elements and identify participants' Mind Wandering (MW), which can hinder focus and learning. An intervention could involve suggesting hints when MW is detected.

Conclusions and Future Work

This study presents how classification models can effectively recognize where users are visually focusing during sensor data analytics in BBPE by utilizing eye-tracking, which provides behavioral evidence of visual attention. The performance of the Ensemble, NN, and K-NN algorithms showcases the efficacy of AOI recognition during students' activities within a BBPE. The use of classification models is anticipated to boost learners' engagement and results by optimizing their learning experience and managing their cognitive load. Furthermore, analyzing learners' interactions can aid their learning process and enhance their acquisition of sensor data analytics skills in the construction industry. This research also paves the way for future endeavors focused on real-time models for eye-gaze-based detection of user actions, including the potential for smart assistance triggered by specific gaze locations during sensor data analytics. Although, results show there is a trade-off between raw and oversampled data in terms of performance measures, both raw and oversampled data performance were closely spaced. In terms of limitation, the study did not explore the individual impact of each of the 28 features on the classification model. The limited sample size constrains the findings' generalizability in construction. Future research will use a larger sample to improve accuracy and explore optimal feature combinations and hyperparameter tuning for dataset optimization.

Acknowledgments

This work was supported by the National Science Foundation (NSF) [grant numbers 2111003 and 2111045].

References

Barricelli, B. R., Cassano, F., Fogli, D., & Piccinno, A. (2019). End-user development, end-user programming and end-user software engineering: A systematic mapping study. *Journal of Systems and Software, 149*, 101-137. doi:https://doi.org/10.1016/j.jss.2018.11.041

Bednarik, R., Vrzakova, H., & Hradis, M. (2012). What do you want to do next: A novel approach for intent prediction in gaze-based interaction. *Eye Tracking Research and Applications Symposium (ETRA)*. doi:10.1145/2168556.2168569

Bender, J., Dziena, A., & Kaiser, G. (2022). *Learning Computational Thinking Efficiently with Block-based Parsons Puzzles.* Paper presented at the 30th International Conference on Computers in Education (ICCE).

Dilmen, K., Kert, S. B., & Uğraş, T. (2023). Children's coding experiences in a block-based coding environment: a usability study on code.org. *Education and Information Technologies, 28*(9), 10839-10864. doi:10.1007/s10639-023-11625-8

Fathy, F., Mansour, Y., Sabry, H., Refat, M., & Wagdy, A. (2023). Virtual reality and machine learning for predicting visual attention in a daylit exhibition space: A proof of concept. *Ain Shams Engineering Journal, 14*(6), 102098. doi:https://doi.org/10.1016/j.asej.2022.102098

Jeelani, I., Han, K., & Albert, A. (2018). Automating and scaling personalized safety training using eye-tracking data. *Automation in Construction, 93*, 63-77. doi:https://doi.org/10.1016/j.autcon.2018.05.006

Khalid, M., Akanmu, A., Murzi, H., Lee, S. W., Awolusi, I., Manesh, D., & Okonkwo, C. (2023). Industry Perception of the Knowledge and Skills Required to Implement Sensor Data Analytics in Construction. *Journal of Civil Engineering Education, 150*(1), 04023010. doi:10.1061/JCEECD.EIENG-1902

Lallé, S., Conati, C., & Carenini, G. (2016). *Predicting confusion in information visualization from eye tracking and interaction data*. Paper presented at the Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, New York, USA. doi: 10.5555/3060832.3060975

Li, J., Li, H., Umer, W., Wang, H., Xing, X., Zhao, S., & Hou, J. (2020). Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction, 109*, 103000. doi:https://doi.org/10.1016/j.autcon.2019.103000

Rashid, K. M., & Louis, J. (2020). Automated Activity Identification for Construction Equipment Using Motion Data From Articulated Members. *Frontiers in Built Environment, 5*. doi:10.3389/fbuil.2019.00144

Rijo-García, S., Segredo, E., & León, C. (2022). Computational thinking and user interfaces: A systematic review. *IEEE Transactions on Education*. doi:10.1109/TE.2022.3159765

Shah, S., Ali, K., Dieker, L., & Hughes, C. (2023). WAVE: a web-based platform for delivering knowledge-driven virtual experiences. *IEEE Computer Graphics and Applications, 43*(3), 54-60. doi:10.1109/MCG.2023.3260599

Tawfik, A. A., Payne, L., & Olney, A. M. (2022). Scaffolding Computational Thinking Through Block Coding: A Learner Experience Design Study. *Technology, Knowledge and Learning*. doi:10.1007/s10758-022-09636-4

Vincur, J., Konopka, M., Tvarozek, J., Hoang, M., & Navrat, P. (2017). *Cubely: virtual reality block-based programming environment*. Paper presented at the Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden. https://doi.org/10.1145/3139131.3141785