Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Eye-tracking-based Risk Perception Prediction for Adaptive Hazard Recognition Training

Mohammad Rezaeiashtiani[1], Meesung Lee[1], Ali Golabchi[1,2], Vicente A. Gonzalez[1], Syed N. Ahmed[3], and Gaang Lee[1]

[1] Department of Civil & Environmental Engineering, University of Alberta, Edmonton, T6G 2W2, Canada
[2] EWI Works International Inc., Edmonton, Alberta, T6G 1H9, Canada
[3] Department of Medicine, University of Alberta, Edmonton, Alberta, T6G 2B7, Canada

ABSTRACT: Hazard recognition is a critical cognitive skill that directly impacts construction site safety, yet conventional safety training methods often fail to account for individual differences in risk perception. This study explores the application of eye-tracking-based biosignal analysis for real-time risk perception assessment in adaptive virtual reality (VR) training. By capturing eye-tracking metrics such as fixation duration, gaze velocity, and pupil dilation, we developed a machine learning model to classify workers' risk perception levels. Among the evaluated classifiers, Gradient Boosting achieved the highest performance (accuracy = 76.81%, AUC = 0.81), demonstrating superior discrimination between high and low-risk perception groups. Compared to traditional physiological signals like EEG, EDA, and PPG, eye-tracking provides a non-invasive, real-time alternative for monitoring hazard recognition. These findings highlight the potential of integrating eye-tracking with adaptive VR training to improve risk perception assessment, which could inform personalized safety interventions in future research and applications. Further studies should validate these results in larger, more diverse cohorts and refine adaptive training algorithms for real-world implementation.

1. INTRODUCTION

Construction workers may overlook as much as 40% to 50% of potential hazards during construction projects (A. Albert et al., 2014; Namian et al., 2016), and the failure to detect hazard is one of the primary causes of construction accidents (Jeelani et al., 2017). Hazard recognition is a fundamental cognitive skill that enables workers to identify and respond to potential dangers, directly influencing them on-site safety (A. Albert et al., 2017). However, a lack of awareness of these potential hazard is often related to shortcomings in individual cognitive abilities such as perception, attention, vigilance, and decision-making (Choi et al., 2020; Perlman et al., 2014). Research indicates that workers who exhibit weaker risk detection abilities in detecting risk tend to miss hazards more frequently, increasing their risk of accidents and injuries (Sun & Liao, 2019). As the construction workforce becomes increasingly multicultural, understanding and assessing individual differences in risk perception abilities are becoming more crucial (Gómez-Bull et al., 2023). Studies have shown that immigrant workers may exhibit distinct cognitive biases that affect their risk perception, leading to higher accident rates among these groups (Casey et al., 2015).

Safety training can be used to prevent unsafe behaviors caused by individual differences in risk perception abilities and to promote proactive safety practices (Aderamo et al., 2024). However, traditional safety training methods are often uniform and fail to consider individual differences in risk perception (Eiter & Bellanca, 2020). One critical limitation of existing training approaches is their inability to adapt to workers' varying levels of risk perception abilities, both in terms of initial competency and learning progress (Duffy, 2003). Traditional training programs often assume that all workers start from the same baseline, which is not the case in real-world construction environments (Aderamo et al., 2024). As a result, training levels that do not align with workers' needs can reduce motivation and engagement (L. Albert & Routh, 2021a), and this lack of engagement can diminish the overall effectiveness of safety training programs (Júnior et al., 2023). Realistic training scenarios are crucial for preparing workers for actual high-risk situations while considering differences in hazard recognition abilities, but traditional training programs often lack this consideration (Cooper & Cotton, 2000).

Recently, research has grown in using VR simulation technology for safety training to bridge this gap (Avveduto et al., 2017). VR-based safety training allows workers to engage in realistic scenarios that enhance their risk perception abilities without the risks associated with real-world dangers (Luo et al., 2025). Furthermore, training methods that generate VR scenarios tailored to individuals' risk perception abilities can overcome the limitations of traditional training methods (Tichon, 2007). However, for VR training to be truly effective, it must be adaptive—meaning it should dynamically adjust to each worker's real-time risk perception level (Avveduto et al., 2017). Risk perception serves as a key indicator of an individual's ability to recognize hazards, making it a crucial metric for adaptive training systems (Gómez-Bull et al., 2023). Implementing training programs that are sensitive to individual cognitive differences can foster a more inclusive safety culture, ultimately enhancing overall site safety by ensuring that all workers feel adequately prepared to identify hazards relevant to their experiences (Sokas et al., 2009).

Existing VR-based adaptive safety training has primarily captured risk perception levels using wearable biosignals such as electroencephalography (EEG), photoplethysmography (PPG), electrodermal activity (EDA), and skin temperature (Noghabaei & Han, 2020). Chae et al. (2024) investigated the habituation effect of auditory warnings in construction equipment operators by quantifying response reduction over time using behavioral data, EEG, and EDA. Their study highlighted that repeated warnings led to increased reaction times. Also, Lee et al. (2021) developed a machine learning framework that leveraged physiological signals from wearable sensors (including EDA, PPG, and skin temperature) to continuously assess construction workers' perceived risk levels during on-site activities, demonstrating that biosensor-derived data can provide objective, real-time safety insights (Lee et al., 2021). These previous studies are meaningful in that they quantified participants' risk perception level using biosignals.

However, biosignals such as EEG, PPG, EDA, and skin temperature have significant limitations when used for real-time risk perception monitoring in adaptive training. EEG provides valuable cognitive insights, but it is highly sensitive to movement, which restricts natural body motion and is counterproductive in realistic training environments (Zhu et al., 2020). PPG, EDA, and skin temperature, on the other hand, are not sufficiently rapid in detecting moment-to-moment fluctuations in risk perception (Kang & Kim, 2022). Moreover, they are influenced by external stimuli that activate the sympathetic nervous system, reducing their specificity for assessing hazard recognition (Jimenez-Molina et al., 2018). In contrast, eye-tracking offers unique advantages as a risk perception monitoring tool. Compared to other biosignals, eye-tracking enables high-resolution, real-time measurements of visual attention patterns, fixation durations, and saccadic movements, which are directly linked to hazard detection and recognition (Peysakhovich et al., 2019). Despite this potential, research on evaluating or predicting individuals' risk perception levels using eye-tracking remains limited.

The objective of this study is to predict individuals' risk perception levels using eye-tracking signals during VR training sessions. The study aims to establish a framework for real-time risk perception assessment, which can serve as a foundation for adaptive VR-based hazard recognition training. To achieve this, the study collects data using a protocol that reduces observational bias and develops a machine learning model to predict risk perception levels. This research enables more precise adjustments to adaptive VR safety training by improving the accuracy of risk perception level predictions through biosignals. Ultimately, this study contributes to enhancing VR safety training in high-risk industries and creating safer work environments.

2. RESEARCH DESIGN

The research framework of this investigation is methodically delineated into four distinct stages as shown in Fig. 1. Stage 1 entailed the creation of a VR scenario encompassing multiple struck-by hazards. Within this VR context, a worker engaged in the construction of a wooden structure may encounter potential collisions with forklifts, the hook of a tower crane, or other construction apparatus. This incident of collision afforded the opportunity to gather hazard recognition data from the study's participants. Stage 2 concentrated on the acquisition of eye-tracking metrics as participants navigated the VR scenario unencumbered. The eye-tracking data collection process was designed to minimize the Hawthorne effect, and the measures taken to achieve this are explained in more detail in Section 2.2. Significant metrics documented included fixation duration, gaze velocity, gaze acceleration, and pupil diameter, which acted as indicators of participants' cognitive processing capabilities. Stage 3 encompasses the labeling process, which is essential for quantifying response time; although the initial timing of the warning sound is established, it remains imperative to meticulously document the timing of hazard detection as experienced by the participants. This time-stamped data is derived from their reactions within the VR environment, which include alterations in their visual perspective or modifications in their directional movements. Stage 4 entails the construction of a machine learning model capable of classifying risk perception levels into two categories: high and low, predicated on the labeled eye-tracking metrics. This systematic methodology offers a holistic approach to comprehending how individuals perceive and react to hazards within a simulated environment.

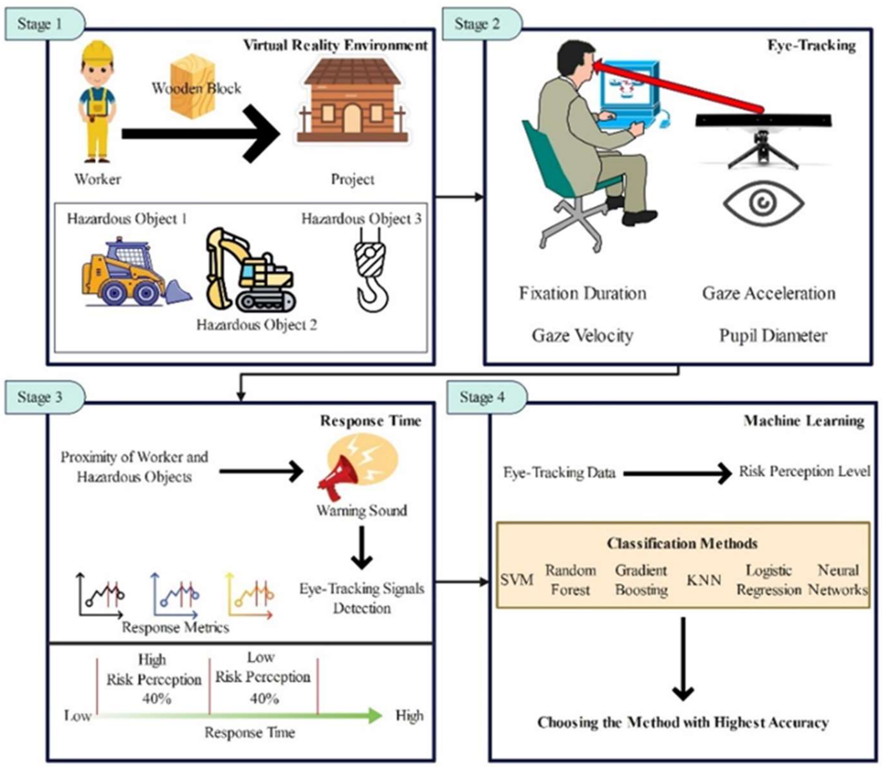


Figure 1: Research framework

2.1 VR Scene Design

The simulated VR construction site was developed using Unreal Engine. The hazard of being struck by equipment has a substantial fatality rate within construction sites (OSHA, 2022). Understanding the fundamental reasons behind this fatal hazard through cognitive processes is crucial; therefore, workers were exposed to multiple struck-by hazards as the focus of this experiment. Participants experienced these struck-by hazards while performing construction tasks. Diverse construction equipment was positioned within this virtual construction setting, including forklifts, rollers, trucks and a tower crane. In addition to construction equipment performing different tasks, the VR environment includes various construction site elements such as buildings, and materials as shown in Fig. 2. To ensure that subjects experience a variety of hazardous situations while performing construction tasks, the VR scenario was designed to incorporate multiple hazard events (including forklift collisions, among others) as shown in Fig. 3. Also, a health bar is set up that deducted when a collision occurred to encourage participants to avoid any collision.

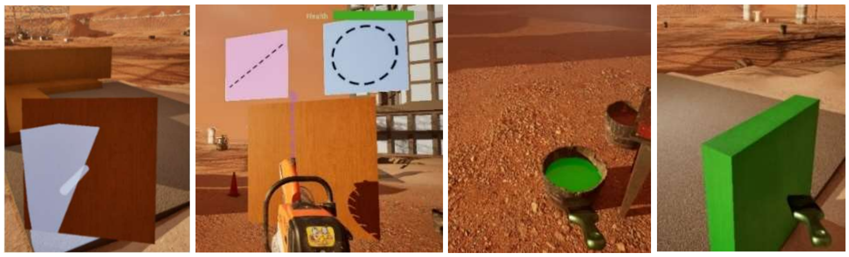


Figure 2: Construction VR environment illustrating workers engaged in wood panel tasks

Figure 3: Potential collision events in the VR scene: forklift approach and descending tower crane hook

2.2 Hazard Recognition Experiment

Existing research methods often involve biases, such as the Hawthorne effect, due to unnatural behavioral responses stemming from the awareness of being observed (McCambridge et al., 2014). This bias occurs when participants alter their behavior as a result of knowing they are being watched (Jimenez-Molina et al., 2018). Therefore, there is a need for studies that minimize the Hawthorne effect through carefully designed experimental protocols. In this study, efforts were made to reduce this bias by ensuring that participants were immersed in the VR training environment for an extended period before data collection began, allowing them to acclimate and behave more naturally. Additionally, no direct observation or verbal interaction occurred during the training tasks to further mitigate awareness-related behavioral changes.

A desktop computer system featuring the Smart Eye Aurora eye tracking sensor was employed in this experiment. It provided participants with a realistic and immersive experience, allowing for precise measurement of their visual attention and decision-making processes during the simulated construction tasks. As outlined in the Introduction, our eye-tracking choice was driven by the limitations of alternatives (EEG, PPG, EDA, and skin temperature) that either restrict natural movement or lack real-time temporal resolution. Eye-tracking, by contrast, offers high-resolution, real-time insights into visual attention and hazard detection. Thus, we selected the Smart Eye Aurora device for its seamless desktop integration and reliable 60 Hz tracking. Fig. 4 displays the layout of the experimental setup, including the positioning of the eye-tracking sensor beneath the desktop display. The distance between the participants and the display was approximately 25 cm, calibrated individually using iMotions software to ensure optimal sensor alignment. Four individuals pursuing an engineering degree, with a mean age of 25.6 ± 1.9 years, were enlisted for the study. This study was approved by the University of Alberta Ethics Board under Ethics ID Pro00141652. Upon arrival, the participants were greeted and asked to complete a consent form prior to the initiation of the experiment. Subsequently, they were provided with a comprehensive overview of the experimental protocols, wherein they were classified as laborers in the construction sector tasked with the assembly of a residential structure utilizing wooden panels while engaging in the processes of locating, cutting, and applying color to them. Participants were instructed to consider safety issues, such as maintaining a safe distance from construction equipment. They were also educated on the importance of being productive and adhering to time management principles throughout the assembly process, as these factors would significantly influence the overall efficiency and success of their tasks. The visual indicators were implemented as health and progress metrics within the virtual reality environment to furnish instantaneous feedback regarding participants' performance, thereby enabling them to modify their actions as necessary and improve their experience. The entirety of the eye tracking information amassed by the iMotions software throughout the experimental sessions was subsequently organized for analytical purposes using this software as shown in Fig. 5.

a b

Figure 4: Experimental setup: a) smart eye aurora eye tracking sensor, b) desktop setup

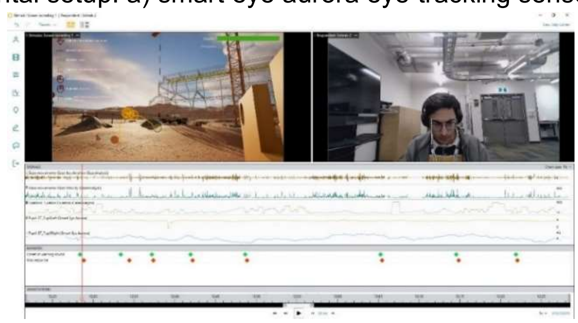


Figure 5: Eye-tracking data processing and analysis using iMotions software

2.3 Data Processing and Labeling

Parameters for predicting hazard recognition levels were extracted from eye-tracking metrics. These attributes included measurements such as fixation duration, gaze velocity, gaze acceleration, and pupil dilation, which provided significant insights into the participants' levels of risk perception(Glöckner & Herbold, 2011). Fixation refers to the eye remaining relatively still at a specific point, while gaze velocity measures the speed at which the eye moves from one point to another (Otero-Millan, 2019). Gaze acceleration, on the other hand, indicates how quickly the change in gaze direction occurs, providing insights into cognitive processes and attentional shifts during task performance. These metrics are crucial for understanding how individuals’ risk perception processes are influenced by hazards such as construction warning sounds (Downing et al., 2004). Pupil diameter is another important metric in eye-tracking studies, as it can reflect cognitive load and emotional responses (Kiefer et al., 2016). Changes in pupil size often correlate with varying levels of attention and arousal, offering further insights into how individuals react to potential risks in their environment (Daniels et al., 2012). By analyzing pupil diameter alongside gaze metrics, researchers can gain a comprehensive understanding of how cognitive and emotional factors interact in risk assessment scenarios (Hershaw & Ettenhofer, 2018). For each parameter, eight distinct statistical features were computed, yielding a total of 32 features per time window. Specifically, these metrics included the minimum, maximum, mean, standard deviation, median, skewness, coefficient of variation, and interquartile range.

Researchers labeled eye-tracking signals based on reaction time following the onset of warning sounds. The labeling process was manually conducted by construction safety experts, who identified when participants recognized the warning sound and exhibited hazard recognition behaviors, such as rotating or shifting their view. To evaluate risk perception, the dataset was initially categorized based on the response times of the participants. More specifically, the fastest 40% of responses were classified as belonging to the high-risk perception category, while the slowest 40% were classified as belonging to the low-risk perception category. The remaining 20% of responses were omitted from classification at this juncture to mitigate potential ambiguity in the training procedure.

2.4 Hazard Recognition Prediction with Machine Learning Models

Employing this categorized dataset, a variety of machine learning classifiers were trained. The machine learning models in this study included support vector machine (SVM), random forest, gradient boosting, k-nearest neighbors (KNN), logistic regression and neural networks (Kotsiantis et al., 2006). The continuous data, which were divided based on window size, were fixed along with shifting time to move the window size forward through the historical signal data. Based on our previous experience, we selected a 2-second window size with a 1-second shift as the optimal configuration. For each model, a grid search was conducted to optimize 3 to 4 key hyperparameters across multiple configurations, and the best-performing combination was selected based on validation performance. To rigorously assess these models, a five-fold cross-validation methodology was executed without shuffling (Hijmans, 2012). Specifically, the dataset was partitioned into five contiguous segments: in each fold, one segment functioned as the test set, while the remaining four segments were utilized for training. This methodology preserves the chronological or sequential integrity of the data, thereby preventing any potential data leakage from future observations and facilitating a more realistic evaluation of model efficacy.

The model's performance was assessed based on key evaluation metrics, including the confusion matrix, accuracy, precision, recall, and F1 score (Ramasubramanian et al., 2017). The confusion matrix provided an overview of the model's classification results by displaying the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Accuracy measured the overall correctness of the model's predictions by calculating the proportion of correctly classified instances out of the total number of samples. Precision evaluated the reliability of positive classifications by determining the proportion of correctly predicted positive cases among all predicted positives. Recall assessed the model's ability to correctly identify positive cases by calculating the proportion of true positive instances out of all actual positive samples. The F1 score, which combines precision and recall into a single metric, was also used to provide a balanced measure of the model's effectiveness. Furthermore, the Receiver Operating Characteristic (ROC) curve was utilized as an additional performance evaluation tool. The ROC curve illustrated the relationship between the false positive rate (FPR) and true positive rate (TPR) across varying classification thresholds, allowing for a more comprehensive assessment of the model's performance beyond a single threshold value (Grunkemeier & Jin, 2001).

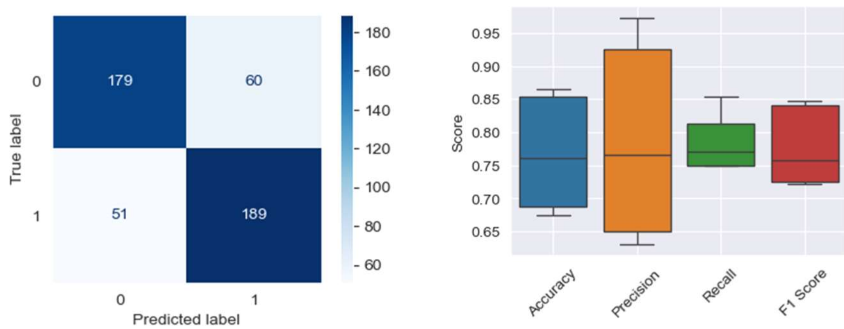
3. RESULTS

As demonstrated in Table 1, the classification models exhibited varying levels of performance, with gradient boosting achieving the highest accuracy (76.81%), followed by random forest (73.25%) and SVM (72.22%). Gradient boosting also outperformed other models in precision (78.89%), recall (78.75%), and F1 score (77.87%), making it the most effective model for classification. In contrast, KNN demonstrated the lowest performance across all metrics, with an accuracy of 67.62%, indicating its limitations in handling the dataset's complexities. Logistic regression and neural networks showed comparable results, with accuracy values of 70.76% and 69.73%, respectively. The consistently superior performance of gradient boosting across all key metrics highlights its robustness and reliability for risk perception classification. A comprehensive overview of each model's performance is presented in Table 1.

Table 1: Types of ML models and their performance metrics

	SVM	Random forest	Gradient boosting	KNN	Logistic regression	Neural networks
Accuracy	72.22	73.25	76.81	67.62	70.76	69.73
Precision	74.39	74.50	78.89	68.72	72.36	70.64
Recall	72.22	73.25	78.75	67.62	70.76	69.73
F1 score	70.85	72.32	77.87	66.77	69.99	69.39

Building on the findings, the gradient boosting classifier was chosen as the final model due to its superior accuracy compared to other candidates. The model's performance was evaluated using a confusion matrix and key classification metrics. Fig. 6-a presents the confusion matrix, which details the distribution of correct and incorrect classifications. The model successfully identified 179 instances as negative and 189 instances as positive. However, 60 negative instances were misclassified as positive, and 51 positive instances were misclassified as negative. While these results demonstrate strong discriminative power between the two classes, there is a slight tendency toward false positives, indicating a more cautious classification approach. Fig. 6-b illustrates a box-and-whisker plot of accuracy, precision, recall, and F1-score across cross-validation folds. The visualization highlights high precision with some variability, while accuracy, recall, and F1-score maintain stable distributions, further supporting the model's reliability. Overall, the gradient boosting classifier exhibits robust classification capabilities, making it the most suitable choice in this study.



a b

Figure 6: a) confusion matrix and b) performance metrics

The ROC curve provides a clear visualization of a model's ability to differentiate between positive and negative classes. The area under the curve (AUC) condenses this performance into a single value, where 0.5 represents random guessing and 1.0 indicates perfect discrimination. Higher AUC values generally signify stronger classification performance. Fig. 7 presents the ROC curves for all evaluated models, including SVM, random forest, gradient boosting, KNN, logistic regression, and neural networks. The gradient boosting classifier achieves the highest AUC (0.81), demonstrating its superior ability to distinguish between risk perception categories. In comparison, random forest (0.79), neural networks (0.77), logistic regression (0.77), SVM (0.75), and KNN (0.73) follow in performance. These results confirm that gradient boosting is the most effective model, as it consistently outperforms others in both accuracy and discrimination ability. The relatively lower AUC of SVM, despite its solid performance in previous evaluations, further reinforces gradient boosting as the optimal choice for risk perception classification in this study.

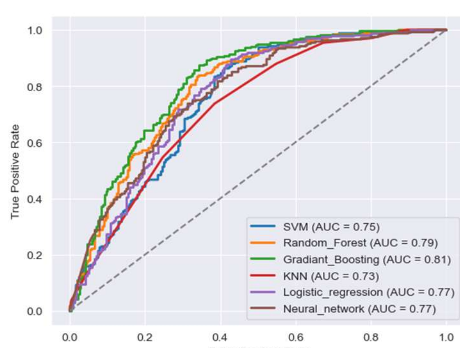


Figure 7: Comparison of ROC curves for different models

4. DISSCUSSION

The results of this study reinforce the potential of eye-tracking-based biosignal analysis for risk perception assessment in adaptive hazard recognition training. Our findings indicate that gradient boosting achieved the highest classification performance among the tested machine learning models, demonstrating superior accuracy, precision, recall, and F1-score. The strong discriminative power of this model suggests that eye-tracking metrics can effectively capture differences in risk perception, making it a valuable tool for real-time, individualized training adjustments. The ROC analysis further supports our findings, as gradient boosting exhibited the highest AUC (0.81), confirming its robustness in distinguishing between high and low-risk perception groups. These results suggest that integrating eye-tracking metrics with adaptive VR-based training may enhance workers' hazard recognition abilities, improving overall construction safety.

These findings align with prior research emphasizing the advantages of biosignal-based assessments over traditional self-report measures (Noghabaei & Han, 2020; Lee et al., 2021). Previous studies utilizing EEG, PPG, and EDA have demonstrated that physiological signals can objectively quantify workers' risk perception levels in VR-based training environments (Chae et al., 2024; Jimenez-Molina et al., 2018). However, unlike EEG, which is highly sensitive to motion artifacts, or EDA and PPG, which have limited responsiveness to momentary attentional shifts, eye-tracking enables high-resolution, real-time monitoring of visual attention and hazard recognition behaviors (Peysakhovich et al., 2019). Our results extend this body of work by showing that eye-tracking data, when analyzed through advanced machine learning models, can provide a reliable, non-invasive, and scalable alternative for adaptive safety training.

Despite similarities with previous studies, our research diverges in key aspects. Many biosignal-based risk perception assessments rely on stationary, controlled environments with minimal distractions (Zhu et al., 2020), which may not reflect natural hazard detection behavior. In contrast, our study emphasized ecological validity by capturing eye-tracking data during a dynamic VR construction environment, ensuring that participants' risk responses were more authentic and representative of real-world hazard recognition challenges. Furthermore, prior research has primarily focused on binary risk perception classifications (i.e., identifying risk vs. no risk), whereas our approach incorporates machine learning-driven predictive modeling, enabling more nuanced differentiation of individual risk perception levels. However, our small, age-homogeneous sample may limit generalizability, warranting further validation with larger, more diverse groups.

5. CONCLUSION

This study demonstrates the effectiveness of biosignals, particularly eye-tracking data, as objective indicators of risk perception in VR-based safety training. By integrating machine learning, we developed a classification model capable of distinguishing risk perception levels, offering a real-time, unbiased alternative to traditional self-reported measures. Our findings highlight the potential of biosignal-based assessments for adaptive VR training, allowing scenarios to dynamically adjust based on individual risk perception, thereby enhancing hazard recognition. Future research should focus on expanding the sample size, incorporating additional biosignals, and developing fully adaptive VR training systems that respond to real-time physiological feedback, further advancing personalized and data-driven safety training methodologies in high-risk industries.

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by WorkSafeBC’s Proof of Principle and Prototyping Program (RS2023-P309) and the Research Exploration Fund at the University of Alberta. Their generous contributions were instrumental in facilitating this research and in advancing our project objectives.

7. REFERENCES

Aderamo, A. T., Olisakwe, H. C., Adebayo, Y. A., & Esiri, A. E. (2024). Behavioral safety programs in high-risk industries: A conceptual approach to incident reduction. *Comprehensive Research and Reviews in Engineering and Technology*, *2*(1), 64–82.

Albert, A., Hallowell, M. R., & Kleiner, B. M. (2014). Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study. *Journal of Construction Engineering and Management*, *140*(2), 04013042.

Albert, A., Hallowell, M. R., Skaggs, M., & Kleiner, B. (2017). Empirical measurement and improvement of hazard recognition skill. *Safety Science*, *93*, 1–8. https://doi.org/10.1016/J.SSCI.2016.11.007

Albert, L., & Routh, C. (2021a). Designing impactful construction safety training interventions. *Safety*, *7*(2), 42.

Albert, L., & Routh, C. (2021b). Designing impactful construction safety training interventions. *Safety*, *7*(2), 42.

Avveduto, G., Tanca, C., Lorenzini, C., Tecchia, F., Carrozzino, M., & Bergamasco, M. (2017). Safety training using virtual reality: A comparative approach. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10324 LNCS*, 148–163. https://doi.org/10.1007/978-3-319-60922-5_11/FIGURES/10

Bhatia, S. (2019). Predicting Risk Perception: New Insights from Data Science. *Https://Doi.Org/10.1287/Mnsc.2018.3121*, *65*(8), 3800–3823. https://doi.org/10.1287/MNSC.2018.3121

BLS. (2024). *BLS. 2024. "A look at falls, slips, and trips in the construction industry." Bureau of Labor Statistics, U.S. Department of Labor, The Economics Daily. Accessed December 17, 2024.* https://www.bls.gov/opub/ted/2024/a-look-at-falls-slips-and-trips-in-the-constructionindustry.htm.

Casey, T. W., Riseborough, K. M., & Krauss, A. D. (2015). Do you see what I see? Effects of national culture on employees' safety-related perceptions and behavior. *Accident Analysis & Prevention*, *78*, 173–184.

Chae, J., Hwang, S., & Kang, Y. (2024). Measuring Habituation to Auditory Warnings Using Behavioral and Physiological Data. *Journal of Construction Engineering and Management*, *150*(7). https://doi.org/10.1061/JCEMD4.COENG-14450

Choi, M., Ahn, S., & Seo, J. (2020). VR-Based investigation of forklift operator situation awareness for preventing collision accidents. *Accident Analysis & Prevention*, *136*, 105404.

Cooper, M., & Cotton, D. (2000). Safety training–a special case? *Journal of European Industrial Training*, *24*(9), 481–490.

Daniels, L. B., Nichols, D. F., Seifert, M. S., & Hock, H. S. (2012). Changes in pupil diameter entrained by cortically initiated changes in attention. *Visual Neuroscience*, *29*(2), 131–142. https://doi.org/10.1017/S0952523812000077

Downing, P. E., Dodds, C. M., & Bray, D. (2004). Why does the gaze of others direct visual attention? *Visual Cognition*, *11*(1), 71–79. https://doi.org/10.1080/13506280344000220

Duffy, V. G. (2003). Effects of training and experience on perception of hazard and risk. *Ergonomics*, *46*(1–3), 114–125. https://doi.org/10.1080/00140130303524

Eiter, B. M., & Bellanca, J. L. (2020). Identify the Influence of Risk Attitude, Work Experience, and Safety Training on Hazard Recognition in Mining. *Mining, Metallurgy and Exploration*, *37*(6), 1931–1939. https://doi.org/10.1007/S42461-020-00293-8/TABLES/3

Gao, Y., González, V. A., Yiu, T. W., Cabrera-Guerrero, G., Li, N., Baghouz, A., & Rahouti, A. (2022). Immersive virtual reality as an empirical research tool: exploring the capability of a machine learning model for predicting construction workers' safety behaviour. *Virtual Reality*, *26*(1), 361–383. https://doi.org/10.1007/S10055-021-00572-9/FIGURES/6

Glöckner, A., & Herbold, A. K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/BDM.684

Gómez-Bull, K. G., Ibarra-Mejía, G., & Vargas-Salgado, M. M. (2023). Risk perception in the construction industry: A literature review and future research directions. *Work*, *76*(4), 1333–1344. https://doi.org/10.3233/WOR-220379/ASSET/IMAGES/10.3233_WOR-220379-FIG2.JPG

Grunkemeier, G. L., & Jin, R. (2001). Receiver operating characteristic curve analysis of clinical risk models. *The Annals of Thoracic Surgery*, *72*(2), 323–326. https://doi.org/10.1016/S0003-4975(01)02870-3

Hershaw, J. N., & Ettenhofer, M. L. (2018). Insights into cognitive pupillometry: Evaluation of the utility of pupillary metrics for assessing cognitive load in normative and clinical samples. *International Journal of Psychophysiology*, *134*, 62–78. https://doi.org/10.1016/J.IJPSYCHO.2018.10.008

Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, *93*(3), 679–688. https://doi.org/10.1890/11-0826.1

Jeelani, I., Albert, A., & Gambatese, J. A. (2017). Why do construction hazards remain unrecognized at the work interface? *Journal of Construction Engineering and Management*, *143*(5), 04016128.

Jimenez-Molina, A., Retamal, C., & Lira, H. (2018). Using Psychophysiological Sensors to Assess Mental Workload During Web Browsing. *Sensors 2018, Vol. 18, Page 458*, *18*(2), 458. https://doi.org/10.3390/S18020458

Júnior, R. R., de Classe, T. M., dos Santos, R. P., & Siqueira, S. W. M. (2023). Current risk situation training in industry, and games as a strategy for playful, engaging and motivating training. *Journal on Interactive Systems*, *14*(1), 138–156.

Kang, D. H., & Kim, D. H. (2022). 1D Convolutional Autoencoder-Based PPG and GSR Signals for Real-Time Emotion Classification. *IEEE Access*, *10*, 91332–91345. https://doi.org/10.1109/ACCESS.2022.3201342

Kiefer, P., Giannopoulos, I., Duchowski, A., & Raubal, M. (2016). Measuring cognitive load for map tasks through pupil diameter. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9927 LNCS*, 323–337. https://doi.org/10.1007/978-3-319-45738-3_21/FIGURES/5

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, *26*(3), 159–190. https://doi.org/10.1007/S10462-007-9052-3/METRICS

Lee, B. G., Choi, B., Jebelli, H., & Lee, S. H. (2021). Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach. *Journal of Building Engineering*, *42*, 102824. https://doi.org/10.1016/J.JOBE.2021.102824

Luo, Y., Seo, J., & Hasanzadeh, S. (2025). Understanding Hazard Recognition Behavior Using Eye-tracking Metrics in a VR-Simulated Environment: Learning from Successful and Failed Conditions. *KSCE Journal of Civil Engineering*, 100173.

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. *Journal of Clinical Epidemiology*, *67*(3), 267–277. https://doi.org/10.1016/j.jclinepi.2013.08.015

Namian, M., Albert, A., Zuluaga, C. M., & Behm, M. (2016). Role of safety training: Impact on hazard recognition and safety risk perception. *Journal of Construction Engineering and Management*, *142*(12), 04016073.

Noghabaei, M., & Han, K. (2020). Hazard Recognition in an Immersive Virtual Environment: Framework for the Simultaneous Analysis of Visual Search and EEG Patterns. *Construction Research Congress 2020: Computer Applications - Selected Papers from the Construction Research Congress 2020*, 934–943. https://doi.org/10.1061/9780784482865.099

OSHA. (2022). *Accident Search Results | Occupational Safety and Health Administration osha.gov*. https://www.osha.gov/ords/imis/accidentsearch.search?acc_keyword=

Otero-Millan, J. (2019). Fixational Eye Movements in Visual, Cognitive, and Movement Disorders. *Contemporary Clinical Neuroscience*, 107–122. https://doi.org/10.1007/978-3-030-31407-1_6/FIGURES/2

Perlman, A., Sacks, R., & Barak, R. (2014). Hazard recognition and risk perception in construction. *Safety Science*, *64*, 22–31.

Peysakhovich, V., Dehais, F., & Duchowski, A. T. (2019). Why is Eye Tracking an Essential Part of Neuroergonomics? *Neuroergonomics: The Brain at Work and in Everyday Life*, 27–30. https://doi.org/10.1016/B978-0-12-811926-6.00004-X

Ramasubramanian, K., Singh, A., Ramasubramanian, K., & Singh, A. (2017). Machine Learning Model Evaluation. *Machine Learning Using R*, 425–464. https://doi.org/10.1007/978-1-4842-2334-5_7

Sokas, R. K., Jorgensen, E., Nickels, L., Gao, W., & Gittleman, J. L. (2009). An intervention effectiveness study of hazard awareness training in the construction building trades. *Public Health Reports*, *124*(4_suppl1), 161–168.

Sun, X., & Liao, P.-C. (2019). Re-assessing hazard recognition ability in occupational environment with microvascular function in the brain. *Safety Science*, *120*, 67–78.

Tichon, J. (2007). Training Cognitive Skills in Virtual Reality: Measuring Performance. *Http://Www.Liebertpub.Com/Cpb*, *10*(2), 286–289. https://doi.org/10.1089/CPB.2006.9957

Zhu, L., Ma, Q., Bai, X., & Hu, L. (2020). Mechanisms behind hazard perception of warning signs: An EEG study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *69*, 362–374. https://doi.org/10.1016/J.TRF.2020.02.001