Visualizing Eye-Tracking Dynamics for Predicting Human Error in Virtual Reality

Tanaz Ghahremani
VR/AR Lab
University of Luxembourg
Esch-sur-Alzette, Luxembourg
tanaz.ghahremani.001@student.uni.lu

Sahar Niknam
VR/AR Lab
University of Luxembourg
Esch-sur-Alzette, Luxembourg
sahar.niknam@uni.lu

Berin Venedik
University of Luxembourg
Esch-sur-Alzette, Luxembourg
berin.venedik.001@student.uni.lu

Jean Botev
VR/AR Lab
University of Luxembourg
Esch-sur-Alzette, Luxembourg
jean.botev@uni.lu

Abstract

Human error remains a pervasive risk across safety-critical domains, motivating research on predictive approaches that enable selective automation and adaptive training. This paper presents a longitudinal Virtual Reality (VR) study for cognitive failure prediction, collected over 18 months from a participant solving mental arithmetic problems. We propose a spatio-temporal representation that transforms eye-tracking features into compact visualizations, and evaluates its effectiveness using convolutional, recurrent, and fusion neural models. These models leverage both spatial and temporal dynamics, achieving 87% accuracy in identifying missed responses. However, distinguishing wrong from correct answers proved more difficult, likely due to similar ocular behavior during confident but incorrect responses and the impact of class imbalance. Overall, the results highlight the feasibility of personalized AI systems trained on rich within-subject data, and position VR combined with deep learning as a platform for real-time monitoring and adaptive training to reduce human error.

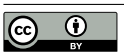
CCS Concepts

• **Computing methodologies** → **Neural networks**; • **Human-centered computing** → **Empirical studies in ubiquitous and mobile computing**; **Virtual reality**; • **Information systems** → **Personalization**.

Keywords

Personalized neural networks, Cognitive failure, Cognitive load, Eye-tracking, Spatio-temporal representation, Data conversion, Virtual reality, Deep learning

ACM Reference Format:
Tanaz Ghahremani, Sahar Niknam, Berin Venedik, and Jean Botev. 2026. Visualizing Eye-Tracking Dynamics for Predicting Human Error in Virtual Reality. In *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems (CHI EA '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3772363.3798798



CHI EA '26, Barcelona, Spain

ACM ISBN 979-8-4007-2281-3/26/04
https://doi.org/10.1145/3772363.3798798

1 Introduction

Human error is a major cause of adverse incidents across domains, from medical malpractice and manufacturing to transportation and nuclear disasters [10, 13, 15, 24]. As an unavoidable aspect of human performance, error must be addressed systemically in work design. Many such errors arise from cognitive failures, such as lapses in attention or memory [20]. Advances in automation and AI reduce reliance on manual performance and therefore lower the likelihood of human error in routine or precise tasks [30, 32, 33]. However, socially acceptable automation systems require minimal, conditional delegation, and predicting human error enables such selective intervention. Studies show that errors are not only detectable via post-error neural activity [4, 5, 34], but also predictable from physiological and behavioral data [12, 17, 27]. Achieving such predictive capacity, however, demands large-scale datasets that capture diverse physiological and behavioral patterns. Virtual reality (VR) offers a unique opportunity to address this challenge. Within virtual environments, we can integrated biosensors that provide extensive physiological and behavioral data. Furthermore, such data can be gathered in scenarios that would be difficult—if not impossible—to reproduce in real-world settings, such as repeated high-risk decision-making in war zones or the study of a single-factor effect under otherwise controlled cognitive workload. Moreover, VR is already widely adopted as a training environment, from surgical simulation and medical education to flight training and industrial safety [6, 16, 23, 25], serving as a platform for studying human performance and error through continuous and objective monitoring of cognitive markers, such as attentional focus and engagement level. Eye-tracking, for example, is a reliable method for assessing cognitive states [14, 22, 26]. Many studies have tested neural networks trained on eye-tracking data as an effective means of modeling cognitive processes and markers [7–9, 11, 21].

Eye-tracking data are often used as time series or aggregated measures (e.g., mean pupil size). In contrast, our work contributes by re-expanding quantitative eye-tracking data into spatio-temporal

visualizations of eye behaviors. By transforming numerical eye-tracking features into a minimal visualization of the eyes, we reintroduce the missing spatial context—for example, how pupil dilation relates to gaze direction—helping models to capture subtle signatures such as individual gaze habits. We then use these visualizations as spatio-temporal inputs to a Convolutional Neural Networks (CNN), a Long Short-Term Memory (LSTM), and a fusion model, capturing spatial patterns, temporal dynamics, and their interactions to predict cognitive failure.

2 Related Work

CNNs have gained popularity for time-series and sequential data, supplementing traditional recurrent models. CNN architectures enable parallelization, efficient parameter sharing, and convolution and pooling operations, lowering the complexity of the model while preserving crucial temporal-spatial data structures. Recent Human-Computer Interaction (HCI) studies have used image-based representations of time-series data to use CNNs for sequence modeling, demonstrating competitive or superior performance to RNNs while improving training efficiency [1–3, 19, 31]. Empirical studies have shown that CNNs can capture long-range dependencies [1], achieve state-of-the-art results in domains such as text classification [36], and outperform hand-crafted methods in human-activity recognition from raw sensor streams [35]. A common extension of this approach converts time series into image representations and applies 2D CNNs. Wang and Oates [31] introduced Gramian angular fields (GAF) and Markov Transition Fields (MTF), using tiled CNNs on the resulting images and achieved highly competitive classification accuracy. The same methodology was applied successfully across domains, including emotion recognition [19], user identification [2], physiological signal analysis [18], hand-gesture recognition [3]. In eye-tracking research, imaging-based representations of gaze features have similarly been shown to outperform traditional summary-feature classifiers in attentional-state tasks [29].

This study proposes a practical alternative to complex time-series encodings by transforming quantitative eye-tracking sequences into temporal eye frames and utilizing 2D CNNs for predicting cognitive failure. Our motivation is twofold: image encodings for enhancing exploratory data analysis and human-in-the-loop validation, and investigating whether a conceptually directly rendered temporal eye representation, rather than engineered transformations like GAFs or MTFs, can achieve similar predictive efficacy while prioritizing interpretability and the preservation of spatial intuition within the feature space.

3 Methodology

In this longitudinal study, we collected eye-tracking data from a participant who completed 300 VR sessions solving multiplication questions over 18 months. The participant was hired as a research assistant, and the study received ethics approval from [–BLINDED–]. The study was conducted in VR as a series of 10-minute sessions. During each session, we collected eye-tracking data at 120 Hz using the integrated Tobii tracker in the HTC VIVE Pro Eye headset (Figure 2). Sessions began with eye-tracking calibration, followed by a test 1×1 question to verify system functionality. The environment

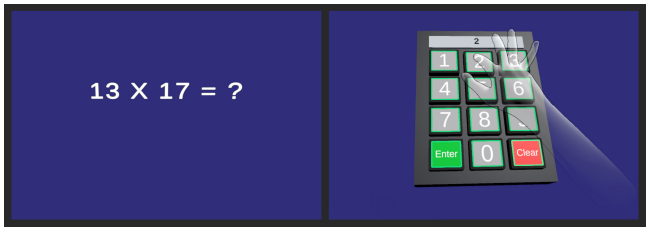


Figure 1: Virtual environment.

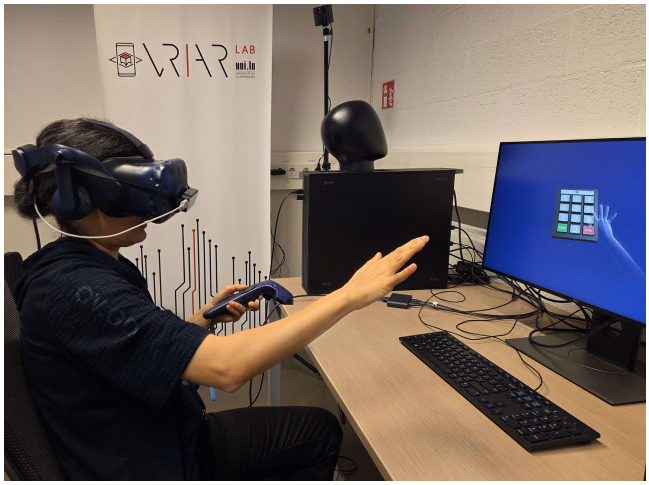


Figure 2: Study setup.

was designed to be minimal, with questions displayed in white on a solid blue background to control for confounds such as pupil light reflex and attention-related cognitive load (Figure 1).

Multiplication problems were displayed for a maximum of 15 seconds. If the participant gave no indication of finding an answer during the interval, the problem was immediately replaced by the next one. When ready to respond, she pressed a controller button to activate a virtual keypad, which was operated through hand gestures tracked by a Leap Motion Controller (Figure 2). The question set evolved across sessions. In the first session, the participant was presented with a random set of about 1, 200 two-digit multiplication problems. Thereafter, a personalized subset of the questions was selected, consisting of problems the participant judged manageable yet challenging—demanding genuine cognitive effort without being so difficult as to cause disengagement or so simple as to require only superficial processing. Whenever the participant made fewer than five errors in five consecutive sessions, the personal subset was updated with more difficult questions. Upon completion of 300 sessions, we collected 17, 343 eye-tracking samples, which included *eye openness*, *pupil diameter*, *pupil relative location* (2D vector), and *gaze direction* (3D vector). Samples were labeled as *Correct* for accurate responses (15, 217 samples), *Missed* when no response was given within 15 seconds (1, 100 samples), and *Wrong* for inaccurate responses (1, 026 samples).

3.1 Preprocessing

Because response times varied, samples differed in length and were standardized by truncation to the final 30 frames—corresponding to 250 ms—an interval associated with high-level cognitive processing [28]. The sequential eye-tracking data were then converted into spatio-temporal eye frames (Figure 3). To generate these frames, eye outlines were positioned at fixed horizontal offsets. Eyelid movement was represented by openness signals. Pupil diameter, illustrated with green circles, was discretized into four levels, allowing for a stable representation without noise amplification. Pupil position was transferred from sensor data, and the iris was rendered for size. Gaze direction was indicated by arrows originating from the pupil. Furthermore, a fading movement trace was added to frames to retain temporal structure and help convolutional neural networks recognize temporal patterns. This visualization reflects actual anatomical scales, including proportional eye form and distance between eyes. Finally, to address class imbalance, we applied majority-class downsampling followed by minority-class augmentation via controlled noise injection prior to image generation. All images were subsequently normalized and resized to a fixed resolution for CNN training.

Figure 3: Example of an eye representation frame.

3.2 Model Architectures and Training

To evaluate different eye-tracking representations, we trained a CNN, an LSTM, and a fusion model combining both. All models used a batch size of 32, up to 250 epochs, with a learning rate of 10^{-3}, weight decay of 10^{-4}, and early stopping (patience 50). Data were split 80:10:10 for training, validation, and testing. The CNN was trained on temporal eye images, leveraging spatial and short-term temporal features by concatenating all frames into a 30-channel input (250 ms at 120 Hz). The CNN architecture, as shown in Figure 4, consists of convolutional layers for spatial feature extraction, followed by a temporal attention module to highlight informative frame dynamics. Dropout was used throughout the network to prevent overfitting. The final classifier is a stack of fully connected layers with normalization and non-linear activations that produce the class predictions. This CNN architecture used local connectivity, parameter sharing, and hierarchical feature abstraction while focusing on short-term temporal cues. It served as a baseline image-based model for comparing sequential and fusion approaches. The second approach used a recurrent neural network with LSTM units to model raw quantitative eye-tracking time series. This bidirectional LSTM architecture incorporates both forward and backward temporal contexts, improving its capacity to recognize patterns before cognitive failures. The model output was condensed into a classifier head with Softmax activation and dropout regularization, providing a complementary perspective to the CNN-only approach. The Fusion model, as shown in Figure 5, integrates per-frame visual embeddings from temporal eye frames with parallel quantitative feature sequences to learn joint spatio-temporal representations. The Fusion architecture is composed of three modules: a CNN-based image encoder, a recurrent feature encoder, and a fusion LSTM followed by a classifier. The image encoder extracts spatial representations from the eye-image sequence, while the feature encoder uses a bidirectional recurrent layer to capture temporal patterns in the numerical attributes. Their outputs were combined and processed by a fusion LSTM to model cross-modal temporal dependencies. The resulting representation is then passed to a fully connected classifier with nonlinear activation and dropout for the final prediction. The Fusion model was designed to capture complementary patterns, such as instantaneous visual cues and long-term sequential trends, in order to outperform single-modality baselines.

4 Results

We evaluated our models on the dataset with three classes: *Correct*, *Missed*, and *Wrong* answers. The *Missed* class showed a clear and consistent ocular pattern, allowing models across all architectures to identify these samples with relatively high accuracy. However, performance was lower when discriminating between *Correct* and *Wrong* responses, as their eye-tracking signatures appeared highly similar. None of the models achieved satisfactory classification performance on the *Wrong* class, although the *Missed* class was detected with reasonable recall (Table 1). The LSTM-only model performed best overall, particularly on the *Missed* class. The CNN-only model maintained its competitiveness, while the Fusion model produced balanced results, but neither addressed the poor separability of the *Wrong* class. In the three-class setup, models struggled to identify the *Wrong* class, although *Missed* responses were detected more reliably, and *Correct* responses performed relatively well. The difficulties appeared to originate not from limited training capacity, but from the behavioral similarities between *Correct* and *Wrong* trials. These results prompted a change to a binary task contrasting the *Correct* and *Wrong* classes.

The binary classification task resulted in substantially improved performance (Table 2). Overall, the LSTM-only model outperformed the CNN-only model in both three-class and binary classification settings. The loss curves in Figure 6 and the confusion matrix in Figure 7 illustrate the stability of training and the class separation achieved by the model. However, the CNN-only model showed highly competitive results, suggesting that convolutional architectures are a promising alternative for longer time sequences or complex feature sets. The Fusion model provided the most balanced performance across the two classes, although it is more complex and computationally more costly than the single-modality models.

5 Discussion

This work investigates the potential of semantic visualizations of eye-tracking time series for predicting cognitive failure in VR. We trained a CNN-only, an LSTM-only, and a Fusion model on spatio-temporal data representations to evaluate how well different architectures capture patterns indicative of impending failure.

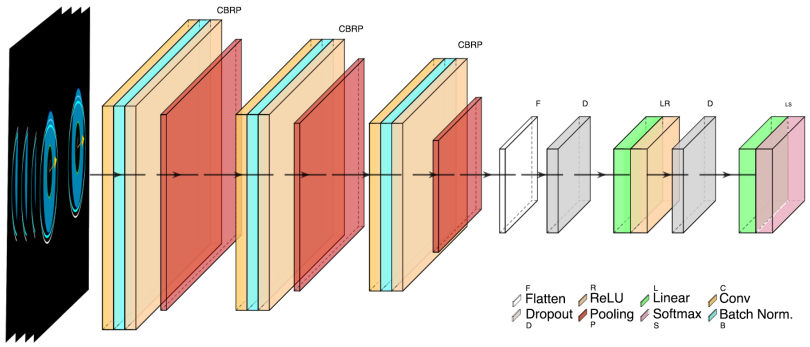


Figure 4: CNN model architecture.

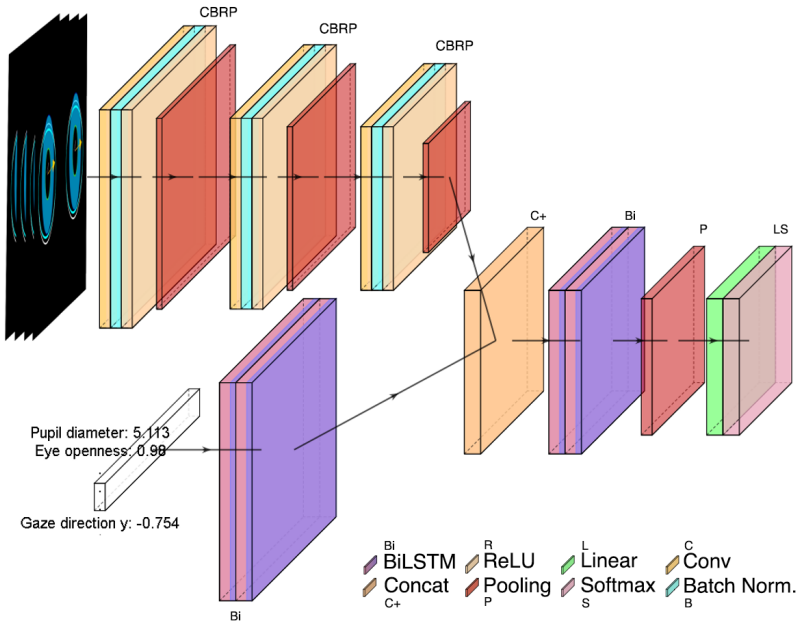


Figure 5: CNN and LSTM Fusion model architecture.

Table 1: Precision (P), recall (R), F1-score, and overall accuracy (Acc.) of the models on the three-class test set.

	Correct			Missed			Wrong			Acc.
Model	P	R	F1	P	R	F1	P	R	F1	
CNN-only	0.572	0.553	0.563	0.647	0.647	0.643	0.348	0.364	0.356	0.522
LSTM-only	0.607	0.773	0.680	0.626	0.902	0.739	0.458	0.100	0.164	0.605
Fusion	0.624	0.707	0.662	0.638	0.814	0.716	0.387	0.218	0.279	0.588

Table 2: Model performance on the two-class test set.

	Correct			Missed			Acc.
Model	P	R	F1	P	R	F1	
CNN-only	0.865	0.813	0.838	0.750	0.816	0.781	0.814
LSTM-only	0.932	0.827	0.876	0.783	0.913	0.843	0.862
Fusion	0.920	0.847	0.882	0.800	0.893	0.844	0.866

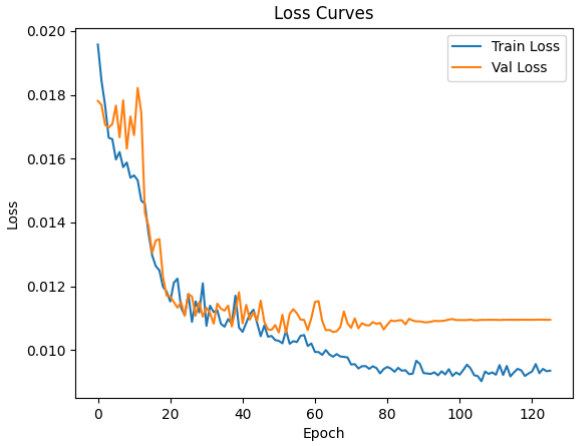


Figure 6: LSTM-only model's loss on the test dataset.

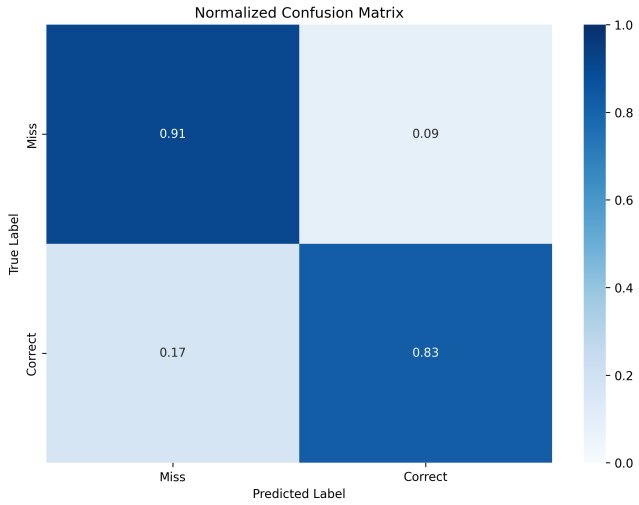


Figure 7: LSTM-only model's normalized confusion matrix on the test dataset.

While all models reliably distinguished *Missed* from *Correct* responses (Table 2), separating *Wrong* from *Correct* samples remained unresolved (Table 1).

As highlighted in Table 1, in the three-class task, all models achieved their highest scores on the *Missed* class, suggesting that this class exhibits pronounced and learnable ocular patterns. Performance on *Correct* responses was moderate, while *Wrong* responses remained poorly detected across architectures. The slightly better performance on *Correct* samples can be attributed to the class imbalance. Despite downsampling the majority class and augmenting the minorities, *Correct* samples still offered greater variation and information content, biasing the training process toward classifying ambiguous cases as *Correct*. Overall accuracy in the three-class task remained modest (0.52–0.60), reflecting limited robustness, with the LSTM-only model performing best overall, particularly on the *Missed* class, due to its ability to describe short sequences dominated by pupil diameter dynamics. Given the quick convergence of the models (Figure 6), this limitation is unlikely to stem from insufficient training capacity, but rather from subtle, difficult-to-capture nuances in ocular behavior. Reviewing the wrong answers revealed frequent examples of attentional slips (*e.g.*, `468` instead of `668`, or `281` instead of `261`). These observations suggest that the participant was not struggling with the arithmetic, and produced incorrect results with confidence. This false confidence likely explains the similarity in cognitive and ocular behavior between the *Correct* and *Wrong* classes. Building on this intuition, we next focused on the two-class task, contrasting *Correct* against *Missed* responses. As expected, all models surpassed 81% accuracy in the binary task, with LSTM-only and Fusion architectures outperforming CNN-only (Table 2). The LSTM-only model achieved the highest precision on *Correct* and the highest recall on *Missed* responses. The Fusion model achieved marginally higher overall accuracy (0.866 vs. 0.862 for LSTM) and a very similar F1 score on the Missed class. The Fusion model provided the most balanced performance across the two classes, although it is more complex and computationally more costly than the single-modality models. This suggests that while combining spatial and temporal features can yield comparable or slightly improved performance, the simpler LSTM-only model, leveraging the strong temporal signal, remains highly effective and efficient for this specific task.

5.1 Limitations and Future Work

The main limitation of this study is its single-participant sample, which restricts generalizability and renders the findings exploratory. Future work should test how such individualized models scale to larger and more diverse cohorts. The inherent class imbalance in the dataset is another limitation. Despite applying downsampling and augmentation, the uneven response distribution exposed the models to substantially greater variability in the *Correct* class than in *Missed* or *Wrong*. This imbalance reduced generalizability—particularly in the three-class setting—and biased predictions toward *Correct* for ambiguous cases. Another challenge lies in the design of the arithmetic task, which was deliberately simple and personalized to sustain engagement without discouragement. While this ensured genuine cognitive effort, it also resulted in predominantly easy trials. Consequently, *Missed* and *Wrong* responses largely reflected attentional lapses rather than substantive cognitive errors. Finally, future work will explore more complex tasks in realistic, real-world scenarios to further improve the applicability and generalizability of the results.

6 Conclusion

A central contribution of this work is a semantic spatio-temporal representation of eye-tracking data, a compact and interpretable visualization method that captures cognitive dynamics, enabling exploratory data analysis and supporting human-in-the-loop validation. While we did not achieve acceptable performance in predicting slips and lapses, our findings indicate that deep learning models can predict cognitive inconclusiveness. In time-sensitive decision contexts, awareness of such an inconclusive mental state justifies the socially acceptable delegation of tasks to automated systems.

As headsets become lighter and biosensors more integrated, large-scale physiological data collection in naturalistic settings becomes feasible. Combined with advancing AI, this enables extraction of subtle cognitive markers and the development of personalized failure models to enhance safety and performance in everyday and high-risk contexts.

References

[1] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv preprint arXiv:1803.01271* (2018). doi:10.48550/arXiv.1803.01271 v2.

[2] Carmen Camara, Pedro Peris-Lopez, Masoumeh Safkhani, and Nasour Bagheri. 2023. ECG Identification Based on the Gramian Angular Field and Tested with Individuals in Resting and Activity States. *Sensors* 23, 2 (2023), 937. doi:10.3390/s23020937

[3] Chana Chansri and Jakkree Srinonchat. 2024. Utilizing Gramian Angular Fields and Convolution Neural Networks in Flex Sensors Glove for Human–Computer Interaction. *IEEE Transactions on Human–Machine Systems* 54, 4 (Aug 2024), 475–483. doi:10.1109/thms.2024.3404101

[4] Rebecca J Compton, Dylan Gearinger, Hannah Wild, Danielle Rette, Elizabeth C Heaton, Stephanie Histon, Pablo Thiel, and Marc Jaskir. 2021. Simultaneous EEG and pupillary evidence for post-error arousal during a speeded performance task. *European Journal of Neuroscience* 53, 2 (2021), 543–555. doi:10.1111/EJN.14947

[5] Claudia Danielmeier and Markus Ullsperger. 2011. Post-error adjustments. *Frontiers in psychology* 2 (2011), 233. doi:10.3389/fpsyg.2011.00233

[6] Pawel Dymora, Bartosz Kowal, Miroslaw Mazurek, and Sliwa Romana. 2021. The effects of Virtual Reality technology application in the aircraft pilot training process. In *IOP conference series: materials science and engineering*, Vol. 1024. IOP Publishing, 012099. doi:10.1088/1757-899X/1024/1/012099

[7] Bulat Ibragimov and Claudia Mello-Thoms. 2024. The use of machine learning in eye tracking studies in medical imaging: a review. *IEEE journal of biomedical and health informatics* 28, 6 (2024), 3597–3612. doi:10.1109/JBHI.2024.3371893

[8] Monika Kaczorowska, Małgorzata Plechawska-Wójcik, and Mikhail Tokovarov. 2021. Interpretable machine learning models for three-way classification of cognitive workload levels for eye-tracking features. *Brain sciences* 11, 2 (2021), 210. doi:10.3390/brainsci11020210

[9] Michal Krol and Magdalena Krol. 2017. A novel approach to studying strategic decisions with eye-tracking and machine learning. *Judgment and Decision Making* 12, 6 (2017), 596–609. doi:10.1017/S1930297500006720

[10] CM La Fata, L Adelfio, R Micale, and G La Scalia. 2023. Human error contribution to accidents in the manufacturing sector: A structured approach to evaluate the interdependence among performance shaping factors. *Safety science* 161 (2023), 106067. doi:10.1016/j.ssci.2023.106067

[11] Jia Zheng Lim, James Mountstephens, and Jason Teo. 2022. Eye-tracking feature extraction for biometric machine learning. *Frontiers in neurorobotics* 15 (2022), 796895. doi:10.3389/fnbot.2021.796895

[12] Cheng-Jhe Lin, Changxu Wu, and Wanpracha A Chaovalitwongse. 2014. Integrating human behavior modeling and data mining techniques to predict human errors in numerical typing. *IEEE Transactions on Human-Machine Systems* 45, 1 (2014), 39–50. doi:10.1109/THMS.2014.2357178

[13] Martin A. Makary and Michael Daniel. 2016. Medical error—the third leading cause of death in the US. *BMJ* 353 (2016). doi:10.1136/bmj.i2139

[14] Sandra P Marshall. 2007. Identifying cognitive state from eye metrics. *Aviation, space, and environmental medicine* 78, 5 (2007), B165–B175.

[15] Zhang Meihui, Dai Licao, Chen Wenming, and Pang Ensheng. 2025. Analysis of human errors in nuclear power plant event reports. *Nuclear Engineering and Technology* 57, 10 (2025), 103687. doi:10.1016/j.net.2025.103687

[16] Jose E Naranjo, Diego G Sanchez, Angel Robalino-Lopez, Paola Robalino-Lopez, Andrea Alarcon-Ortiz, and Marcelo V Garcia. 2020. A scoping review on virtual reality-based industrial training. *Applied Sciences* 10, 22 (2020), 8224. doi:10.3390/app10228224

[17] Yeon Ju Oh, Yong Hee Lee, and Jong Hun Yun. 2011. A Study on the Operator's Erroneous Responses to the New Human Interface of a Digital Device to be introduced to Nuclear Power Plants. In *International Conference on Human-Computer Interaction*. Springer, 337–341. doi:10.1007/978-3-642-22098-2_68

[18] Jiahao Qin, Feng Liu, et al. 2025. GAF-FusionNet: Multimodal ECG Analysis via Gramian Angular Fields and Split Attention. *arXiv preprint arXiv:2501.01960v1* (2025). doi:10.32388/I182MY Preprint.

[19] Jie-Lin Qiu, Xin-Yi Qiu, and Kai Hu. 2018. Emotion Recognition Based on Gramian Encoding Visualization. In *Brain Informatics (Lecture Notes in Computer Science, vol. 11309)*. Lecture Notes in Computer Science, Vol. 11309. Springer, Cham, 3–12. doi:10.1007/978-3-030-05587-5_1

[20] James Reason. 1990. *Human Error*. Cambridge university press. doi:10.1017/cbo9781139062367

[21] Antonio Rizzo, Sara Ermini, Dario Zanca, Dario Bernabini, and Alessandro Rossi. 2022. A machine learning approach for detecting cognitive interference based on eye-tracking data. *Frontiers in Human Neuroscience* 16 (2022), 806330. doi:10.3389/fnhum.2022.806330

[22] Darrell S Rudmann, George W McConkie, and Xianjun Sam Zheng. 2003. Eye-tracking in cognitive state detection for HCI. In *Proceedings of the 5th international conference on Multimodal interfaces*. 159–163. doi:10.1145/958432.958464

[23] Greg S Ruthenbeck and Karen J Reynolds. 2015. Virtual reality for medical training: the state-of-the-art. *Journal of Simulation* 9, 1 (2015), 16–26. doi:10.1057/jos.2014.14

[24] Paul M Salmon, MA Regan, and Ian Johnston. 2005. Human error and road transport: phase one: a framework for an error tolerant road transport system.

[25] Romain Seil, Claude Hoeltgen, Hervé Thomazeau, Hermann Anetzberger, and Roland Becker. 2022. Surgical simulation training should become a mandatory part of orthopaedic education. *Journal of Experimental Orthopaedics* 9, 1 (2022), 22. doi:10.1186/s40634-022-00455-1

[26] Vasileios Skaramagkas, Giorgos Giannakakis, Emmanouil Ktistakis, Dimitris Manousos, Ioannis Karatzanis, Nikolaos S Tachos, Evanthia Tripoliti, Kostas Marias, Dimitrios I Fotiadis, and Manolis Tsiknakis. 2021. Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE reviews in biomedical engineering* 16 (2021), 260–277. doi:10.1109/RBME.2021.3066072

[27] Nahoko Takada, Tipporn Laohakangvalvit, and Midori Sugaya. 2022. Human error prediction using heart rate variability and electroencephalography. *Sensors* 22, 23 (2022), 9194. doi:10.3390/s22239194

[28] Anne E Urai, Anke Braun, and Tobias H Donner. 2017. Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature communications* 8, 1 (2017), 14637. doi:10.1038/ncomms14637

[29] Lisa-Marie Vortmann, Jannes Knychalla, Sonja Annerer-Walcher, Mathias Benedek, and Felix Putze. 2021. Imaging Time Series of Eye Tracking Data to Classify Attentional States. *Frontiers in Neuroscience* 15, 664490 (2021). doi:10.3389/fnins.2021.664490

[30] Jack Ng Kok Wah. 2025. Revolutionizing surgery: AI and robotics for precision, risk reduction, and innovation. *Journal of Robotic Surgery* 19, 1 (2025), 47. doi:0.1007/s11701-024-02205-0

[31] Zhiguang Wang and Tim Oates. 2015. Imaging time-series to improve classification and imputation. In *Proceedings of the 24th International Conference on Artificial Intelligence* (Buenos Aires, Argentina) *(IJCAI'15)*. AAAI Press, 3939–3945. doi:10.5555/2832747.2832798

[32] Nicholas J Ward. 2000. Automation of task processes: An example of intelligent transportation systems. *Human Factors and Ergonomics in Manufacturing & Service Industries* 10, 4 (2000), 395–408.

[33] Thomas Wendler, Fijs W. B. van Leeuwen, Nassir Navab, and Matthias N. van Oosterom. 2021. How molecular imaging will enable robotic precision surgery: the role of artificial intelligence, augmented reality, and navigation. *European Journal of Nuclear Medicine and Molecular Imaging* 48, 13 (2021), 4201–4224. doi:10.1007/s00259-021-05445-6

[34] C Wirth, PM Dockree, S Harty, E Lacey, and M Arvaneh. 2019. Towards error categorisation in BCI: single-trial EEG classification between different errors. *Journal of neural engineering* 17, 1 (2019), 016008. doi:10.1088/1741-2552/ab53fe

[35] Jian Bo Yang, Minh Nhut Nguyen, Phyo Phyo San, Xiao Li Li, and Shonali Krishnaswamy. 2015. Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)*. 3995–4001. doi:10.5555/2832747.2832806

[36] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *Advances in neural information processing systems* 28 (2015). doi:10.5555/2969239.2969312