Human-Computer Interaction & Emerging Technologies, Vol. 233, 2026, 325–332
https://doi.org/10.54941/ahfe1007532

Deep Learning for Eye-Gaze Event Detection for Personalized Gaze-Based Interaction in Real-World Settings

Ivan A. Basyul and Boris B. Velichkovsky

Institute of Psychology, Russian Academy of Sciences, Moscow, Russia

ABSTRACT

The paper presents a review of modern machine learning and, specifically, deep learning approaches to the detection of various oculomotor events. The prospects and limitations imposed by such approaches are discussed. The conditions and tasks in which these approaches prove most productive are described. The described methods can significantly refine the dynamics of visual attention and the perceptual process in general within various experimental psychological tasks. Implementing such methods in research practice will allow for more accurate description and interpretation of results obtained in specific psychological and psychophysiological studies involving the registration of oculomotor activity.

Keywords Perception, Eye tracking, Machine learning, Deep learning

INTRODUCTION

Eye movement analysis, or oculomotorics, is a fundamental tool in neuroscience, experimental psychology, computer science, and clinical diagnostics. Recorded eye movement patterns serve as a quantitative reflection of visual attention processes, cognitive processing, and an indicator of neurological health (Melea and Federici, 2012).

Key types of oculomotor events subject to detection and classification include (Birawo and Kasprowski, 2022):

- fixations: periods when the eye remains relatively stationary, allowing the retina to collect and process detailed visual information;
- saccades: rapid, ballistic movements shifting gaze from one fixation point to another;
- smooth pursuits (SP): movements performed to maintain focus on a slowly moving object;
- glissades: small, corrective movements often following immediately after a saccade (post-saccadic glissades) or preceding it, representing a kind of "drift";
- post-saccadic oscillations (PSO): brief oscillatory movements occurring immediately after the completion of a saccade.

Historically, traditional threshold algorithms, such as I-VT (velocity threshold) and I-DT (dispersion threshold), were used for event detection.

Received February 6, 2026; Revised April 2, 2026; Accepted April 22, 2026; Available online July 20, 2026

These methods require the manual tuning of numerous parameters (Birawoa and Kasprowski, 2022). However, threshold methods possess a significant drawback: their performance depends heavily on data quality and the choice of optimal threshold values, making the comparison of results between different studies difficult. Moreover, they demonstrate low robustness to noise and are unable to reliably isolate subtle events such as glissades or smooth pursuit.

The transition to Machine Learning (ML) and Deep Learning (DL) methods was driven by the need to eliminate dependence on manual threshold tuning and to increase classification robustness. ML/DL models are trained on labeled data and are capable of classifying raw eye-tracking data automatically, ensuring higher accuracy and reproducibility of results (Birawo, Kasprowski, 2022).

DATA REQUIREMENTS

Feature Vector Formation. For the successful application of ML and DL algorithms, creating an informative feature vector extracted from raw data is critically important. Beyond basic spatial coordinates (x, y), derived kinematic characteristics are used: velocity, acceleration, and jerk (Birawo and Kasprowski, 2025).

Research shows that using a complex set of features significantly increases classification accuracy. For instance, a model combining velocity, acceleration, jerk, and movement direction achieved the highest accuracy. Velocity and acceleration are key parameters for differentiating high-speed saccades. However, including only acceleration does not ensure effective distinction between fixations and smooth pursuit, highlighting the need to integrate directional features and temporal context (Birawo and Kasprowski, 2025).

Recording. Characteristic Requirements Detection accuracy depends on recording quality. High temporal resolution (sampling rate) is an important condition for the reliable isolation of various eye movement types. Different event types require different minimum frequencies: frequencies of 60 to 90 Hz are sufficient for studying fixations; for accurate analysis of saccades, a frequency of 120 to 250 Hz or higher is necessary. Eye movements used in deep clinical diagnostics (e.g., pathological saccades, glissades, PSO) are very subtle and rapid. A low sampling rate (less than 120 Hz) leads to the smoothing of velocity peaks, distortion of the saccade kinematic profile, and makes reliable isolation of events like glissades and PSO impossible. Since ML and DL models are trained on kinematic features (velocity, acceleration, jerk), any loss of precision in these features caused by a low sample rate inevitably reduces overall classification accuracy. Consequently, for the effective application of ML/DL in the analysis and classification of all distinguishable eye movement types, eye trackers with a sampling rate of at least 250 Hz are required.

Data Labeling. A significant number of ML and DL methods applied in eye movement analysis belong to supervised learning methods and require

verified, well-labeled data. This labeled data serves as a ground truth, or "correct answer," for training classifiers and evaluating their ability to generalize to new data. The quality and standardization of labeled datasets, such as GazeCom, are of great importance for developing robust algorithms and enabling comparative evaluation of different approaches (Melnyk et al., 2024). The accuracy of coordinate annotation (pupil, screen, or camera coordinates) determines the reliability of the labeling and, consequently, the maximum achievable accuracy of the trained models (Chhimpa et al., 2025).

Comparative Analysis of Detection Algorithms

Threshold Methods. A key disadvantage of traditional algorithms like I-VT and I-DT is their sensitivity to threshold selection. There is no standard optimal velocity threshold, and varying it critically affects performance. For example, the I-DT algorithm in one evaluation showed maximum recall for fixations (99%) at a dispersion threshold of 7 px, but minimal fixation precision (39%) at a 1 px threshold. At the optimal point (dispersion threshold 3.5 px), I-DT provides high fixation recall (95%) and precision (98%), yet the F1-score for saccades is only 66%. The I-VT algorithm, using an optimal velocity threshold of 0.5 px/ms, demonstrates an even lower F1-score for saccades — 60% (Birawo and Kasprowski, 2022). These low F1-scores for saccades indicate that threshold methods struggle to accurately determine the boundaries of rapid oculomotor events.

Machine Learning and Deep Learning Methods. Approaches based on ML (Random Forest, RF; Support Vector Machine, SVM; Logistic Regression) and DL (Convolutional Neural Networks, CNN; Recurrent Neural Networks, RNN/LSTM/GRU) offer classification that is significantly more robust and accurate, as they learn to isolate movement patterns without the need for manual threshold setting. Overall, deep learning algorithms, such as CNNs, and machine learning algorithms, such as RF, surpass traditional threshold methods across most metrics. Specifically, RF and CNN classifiers show comparable results in detecting fixations and saccades, with both significantly outperforming I-VT and I-DT in all performance evaluation metrics, except for saccade recall in some cases (Birawo and Kasprowski, 2022).

The F1-score is the most reliable metric for evaluating algorithms, especially with unbalanced classes where the number of fixations usually exceeds the number of saccades. The fact that ML/DL models can achieve F1-scores for saccades up to 91%, whereas traditional methods do not exceed 66% (Birawo and Kasprowski, 2022), demonstrates their fundamental superiority in accurately determining the kinematic boundaries of rapid events.

The Random Forest (RF) method acts as an optimal compromise. It ensures accuracy comparable to DL models, surpassing traditional approaches, while being a tree-based method that is inherently more interpretable than complex deep neural networks. In research and clinical settings where feature significance analysis is important, RF is often chosen for its balance between high accuracy and explainability. Among DL approaches, hybrid architectures combining CNNs and recurrent models

(GRU/LSTM) are the most promising (Chirag, Divya, Keyan, 2025). CNNs are effective in extracting local, spatial movement features, while recurrent layers capture temporal dependencies in sequential data, which is necessary for classifying events requiring consideration of temporal history (Chirag and et al., 2025).

Table 1: Comparison of classification accuracy (F1-Scores).

Algorithm/Model	F1-score (Fixations)	F1-score (Saccades)	Manual Tuning Required	Temporal Dependency Handling
I-VT (0.5 px/ms)	94%	60%	High	No
I-DT (3.5 px)	96%	66%	High	No
Random forest (RF)	Exceeds I-DT	Exceeds I-DT	Training only	Moderate
CNN/CNN+RNN	Up to 99%	Up to 91%	Training only	High (esp. CNN+GRU)

Detection of "Complex Events": Smooth Pursuit, Glissades, Post-Saccadic Oscillations

Smooth Pursuit (SP). Smooth pursuit requires accurate classification based on low but steady velocities, creating a risk of confusing these events with fixations (Birawo and Kasprowski, 2025). Traditional algorithms like IVDT may use a combination of velocity and dispersion thresholds for SP classification, but they do not account for PSO (Birawo and Kasprowski, 2022). DL models, especially those using recurrent architectures (CNN+RNN/GRU), show significant accuracy improvements for SP, reaching up to 88% (Birawo, Kasprowski, 2025). Their advantage lies in the ability to capture steady movement direction and long-term temporal dependencies, effectively distinguishing SP from random drift associated with fixation.

Glissades and Post-Saccadic Oscillations (PSO). Glissades and PSOs represent the most difficult events for automatic detection. Traditional algorithms detect glissades based solely on their duration, often defining them as movement occupying half the duration of a saccade (Birawo and Kasprowski, 2022). This approach leads to high rates of misclassification: short saccades may be erroneously labeled as glissades, and long glissades as saccades (Birawo and Kasprowski, 2022). Furthermore, such algorithms often fail to account for glissades preceding saccades or other events like PSO. DL models trained on a full set of kinematic features allow for the classification of PSO and glissades by considering them in the context of fixations and saccades (Birawo and Kasprowski, 2022). However, PSO detection remains a challenging task: even advanced DL models achieve a PSO detection accuracy of only 73% (Birawo and Kasprowski, 2025). This indicates that these subtle movements are likely not yet fully separated from noise or residual fixation movements in current feature vectors.

A successful example of a non-threshold method for detecting complex events is demonstrated by an algorithm that combines saccade detection in the acceleration domain with specialized onset and offset criteria for PSO. This method demonstrated high agreement with manual annotation,

achieving a Cohen's kappa coefficient of approximately 0.8 (Larsson, Nystrom, Stridh, 2013).

Application of ML/DL methods in psychological and clinical research

The high sensitivity and objectivity provided by ML/DL are necessary for using oculomotor data as accurate biomarkers in complex cognitive and neurological tasks (Graham et al., 2024).

Reading and Cognitive Load Research. In visual search studies, increased saccade amplitude reflects improved search efficiency in the presence of meaningful visual cues (Sun and Jiang, 2025). ML analysis allows for high-precision measurement of these subtle changes in eye movement parameters. ML detection acquires special significance in studies of reading disorders, such as dyslexia. Glissades and altered saccade amplitude are important markers of this condition (Szalma and Weiss, 2020). ML-based classifiers using behavioral and oculomotor features achieved a classification performance of 76.25% based on oculomotor features alone (and 92.91% based on behavioral ones) in detecting reading difficulties (Szalma and Weiss, 2020). Since traditional threshold methods cannot reliably isolate glissades, and glissades are critical for characterizing dyslexia (Szalma and Weiss, 2020), the application of high-precision ML/DL approaches (F1 > 90%) for their detection becomes a cornerstone in the research and early diagnosis of dyslexia.

Challenges. The most difficult events to detect remain "subtle" events like PSO and glissades (PSO accuracy approx. 73%) (Birawo and Kasprowski, 2025). Future research must focus on optimizing architectures and developing features sensitive to low-amplitude oscillations. Further development depends on the availability of large, high-quality annotated datasets. Hybrid DL architectures (CNN+RNN) provide the most accurate solutions. Their application is a necessary condition for high-sensitivity research, including dyslexia diagnosis and clinical neurodiagnostics. Successful integration requires strict data collection standards (sampling rate ≥ 250 Hz).

Beyond Detection: ML in Personalized Learning, UI/UX, and HCI

While the translated manuscript provides a robust foundation for detecting oculomotor events using ML, the next logical step is applying these detected events to solve complex real-world problems. The following analysis expands on the manuscript by exploring how these precise metrics drive innovation in education, interface design, and human-computer interaction.

1. Personalized Learning and Intelligent Tutoring Systems (ITS)

The manuscript highlights the ability of ML to detect fixation durations and saccadic velocities with high precision. in Educational Technology (EdTech), these metrics are not just physiological data points but direct indicators of Cognitive Load.

Real-time Cognitive State Estimation: Research indicates that fixation duration, blink rate, and pupil diameter are significant predictors of cognitive

load. By feeding these features into ML models (such as Random Forests or SVMs), systems can classify a student's state as "bored," "engaged," or "confused" (Cognitive Overload) in real-time.

Adaptive Content Delivery: Intelligent Tutoring Systems (ITS) utilize this data to adapt the curriculum dynamically. If a student's gaze pattern indicates frequent regression (reading the same line repeatedly—a specific sequence of saccades and fixations), the system can infer difficulty and automatically simplify the text or offer a hint.

Predicting Learning Outcomes: Deep learning models analyzing scanpaths (the sequence of fixations) can predict comprehension scores before a student even answers a test question. This allows for proactive intervention rather than reactive grading.

2. Interface Optimization (UI/UX) and Predictive Saliency

The manuscript discusses "saliency" implicitly through visual search tasks. In modern UI/UX, ML transforms eye tracking from a retrospective analysis tool into a predictive design tool.

Predictive Saliency Modeling: Traditional eye tracking requires testing a design with human participants. However, modern Deep Learning models (specifically Generative Adversarial Networks - GANs and CNNs) can now be trained on vast datasets of human eye movements. These models can generate "synthetic heatmaps" for a new website design instantaneously, predicting where users will look with 80–90% accuracy without needing a physical eye tracker.

Automated Usability Testing: By detecting "glissades" and erratic saccades (which the manuscript identifies as noise or corrective movements), ML models can flag UI elements that cause visual confusion. If a user's eye has to make many corrective movements (glissades) to land on a button, the button is likely poorly placed or designed. This allows for automated A/B testing where the "winner" is chosen based on the lowest visual effort required.

3. Human-Computer Interaction (HCI) and Real-World Interaction

The high-precision detection of "subtle events" like Smooth Pursuit (SP) mentioned in the manuscript is critical for advanced HCI.

Foveated Rendering in VR: Virtual Reality requires massive computing power. By detecting the exact fixation point with low latency (using the high-frequency tracking >250Hz recommended in the text), VR systems can use Foveated Rendering. This technique renders only the center of the gaze in high resolution while blurring the periphery, saving up to 60% of GPU resources. This relies entirely on the robust prediction of saccade landing points to avoid "pop-in" artifacts.

Assistive Technology: For individuals with motor impairments (e.g., ALS), the eye becomes the primary input device. The "Midas Touch" problem (unintentional clicking by staring) is a classic HCI challenge. The manuscript's discussion on separating fixations from smooth pursuit is vital here. ML classifiers can distinguish between a "gaze for observation" (looking at a button) and a "gaze for intent" (wanting to click), enabling more natural communication interfaces.

Driver Safety Systems: In automotive HCI, detecting Post-Saccadic Oscillations (PSO) and slow eyelid closure (PERCLOS) are key biomarkers for drowsiness. The high-precision ML models described in the text are essential here, as the difference between an alert driver's blink and a drowsy driver's microsleep is a matter of milliseconds and subtle kinematic changes.

CONCLUSION

The transition from threshold-based methods to Machine Learning in eye tracking, as detailed in the manuscript, is not merely a technical upgrade; it is an enabling technology. By reliably detecting complex events like glissades and distinguishing smooth pursuit from fixations, we unlock the ability to model human cognition, predict visual attention in interfaces, and build responsive, gaze-aware environments.

ACKNOWLEDGMENT

This paper was supported by Russian Science Foundation, Grant No. 25-18-00757.

REFERENCES

Birawo B, Kasprowski P. Review and Evaluation of Eye Movement Event Detection Algorithms. Sensors (Basel). 2022 Nov 15;22(22):8810. doi: 10.3390/s22228810. PMID: 36433407; PMCID: PMC9699548.

Birawo, B. A., Kasprowski, P. (2025). Performance Analysis of Eye Movement Event Detection Neural Network Models with Different Feature Combinations. Applied Sciences, 15(11), 6087. https://doi.org/10.3390/app15116087

Chhimpa GR, Kumar A, Garhwal S, Kumar D, Wani NA, Wani MA, Shakil KA. A Comprehensive Framework for Eye Tracking: Methods, Tools, Applications, and Cross-Platform Evaluation. J Eye Mov Res. 2025 Sep 23;18(5):47. doi: 10.3390/jemr18050047. PMID: 41149949; PMCID: PMC12564957.

Chirag S., Divya N., Keyan Lin. A deep learning approach to track eye movements based on events. Computer Vision and Pattern Recognition (cs.CV). 2025 eprint 2508.04827. doi.org/10.48550/arXiv.2508.04827

El Hmimdi AE, Kapoula Z. Can Saccade and Vergence Properties Discriminate Stroke Survivors from Individuals with Other Pathologies? A Machine Learning Approach. Brain Sci. 2025 Feb 22;15(3):230. doi: 10.3390/brainsci15030230. PMID: 40149752; PMCID: PMC11940339.

Graham, L., Vitorio, R., Walker, R., Barry, G., Godfrey, A., Morris, R., & Stuart, S. (2024). Digital Eye-Movement Outcomes (DEMOs) as Biomarkers for Neurological Conditions: A Narrative Review. Big Data and Cognitive Computing, 8(12), 198. https://doi.org/10.3390/bdcc8120198

Larsson L., Nystrom M., Stridh M. (2013). Detection of Saccades and Postsaccadic Oscillations in the Presence of Smooth Pursuit. Biomedical Engineering, IEEE Transactions on. 60. 2484-2493. 10.1109/TBME.2013.2258918.

Mele ML, Federici S. Gaze and eye-tracking solutions for psychological research. Cogn Process. 2012 Aug;13 Suppl 1:S261-5. doi: 10.1007/s10339-012-0499-z. PMID: 22810423.

Melnyk, Kateryna & Friedman, Lee & Komogortsev, Oleg. (2024). What can entropy metrics tell us about the characteristics of ocular fixation trajectories?. PLOS ONE. 19. e0291823. 10.1371/journal.pone.0291823.

Sun N, Jiang Y (2025) Eye movements and user emotional experience: a study in interface design. Front. Psychol. 16:1455177. doi: 10.3389/fpsyg.2025.1455177

Szalma J., Weiss B. 2020. Data-Driven Classification of Dyslexia Using Eye-Movement Correlates of Natural Reading. In ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers). Association for Computing Machinery, New York, NY, USA, Article 40, 1–4. https://doi.org/10.1145/3379156.3391379