Article

Integrating Eye-Tracking and Artificial Intelligence for Quantitative Assessment of Visuocognitive Performance in Sports and Education

Francisco Javier Povedano-Montero [1,2,*] , Ricardo Bernardez-Vilaboa [1] , José Ramon Trillo [3] , Rut González-Jiménez [1] , Carla Otero-Currás [1] , Gema Martínez-Florentín [1] and Juan E. Cedrún-Sánchez [1,4]

1 Optometry and Vision Department, Faculty of Optics and Optometry, Complutense University of Madrid, 28037 Madrid, Spain; ricardob@ucm.es (R.B.-V.); rutgon03@ucm.es (R.G.-J.); carlaote@ucm.es (C.O.-C.); gemartin@ucm.es (G.M.-F.); jcedrun@ucm.es (J.E.C.-S.)
2 Hospital Doce de Octubre Research Institute (i+12), 28041 Madrid, Spain
3 Department of Computer Science and Systems Engineering, University of Zaragoza, 50018 Zaragoza, Spain; jtrillo@unizar.es
4 Applied Vision Research Group, Faculty of Optics and Optometry, Universidad Complutense de Madrid, 28037 Madrid, Spain
* Correspondence: franpove@ucm.es

Abstract

Background: Eye-tracking technology enables the objective quantification of oculomotor behavior, providing key insights into visuocognitive performance. This study presents a comparative analysis of visual attention patterns between rhythmic gymnasts and school-aged students using an optical eye-tracking system combined with machine learning algorithms. Methods: Eye movement data were recorded during controlled visual tasks using the DIVE system (sampling rate: 120 Hz). Spatiotemporal metrics—including fixation duration, saccadic amplitude, and gaze entropy—were extracted and used as input features for supervised models: Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Decision Tree (CART), Random Forest, XGBoost, and a one-dimensional Convolutional Neural Network (1D-CNN). Data were divided according to a hold-out scheme (70/30) and evaluated using accuracy, F1-macro score, and Receiver Operating Characteristic (ROC) curves. Results: XGBoost achieved the best performance (accuracy = 94.6%; F1-macro = 0.945), followed by Random Forest (accuracy = 94.0%; F1-macro = 0.937). The neural network showed intermediate performance (accuracy = 89.3%; F1-macro = 0.888), whereas SVM and k-NN exhibited lower values. Gymnasts demonstrated more stable and goal-directed gaze patterns than students, reflecting greater efficiency in visuomotor control. Conclusions: Integrating eye-tracking with artificial intelligence provides a robust framework for the quantitative assessment of visuocognitive performance. Ensemble algorithms demonstrated high discriminative power, while neural networks require further optimization. This approach shows promising applications in sports science, cognitive diagnostics, and the development of adaptive human–machine interfaces.

Keywords: eye-tracking; machine learning; oculomotor function; rhythmic gymnastics; visual attention

1. Introduction

Oculomotor control is fundamental for postural stability and the precise execution of movement, particularly in sports disciplines that demand high coordination, such as

rhythmic gymnastics. Eye movements allow athletes to gather relevant visual information from the most critical areas of the visual scene, facilitating optimal motor control and faster decision-making [1]. It has been shown that expert athletes display more efficient visual search patterns than less skilled performers (characterized by fewer but longer fixations) reflecting a more effective integration between perception and motor action [1].

Gaze stabilization is maintained through vestibular reflexes, visual feedback, and cervical motion, while its orientation depends on rapid saccades and smooth pursuits [2]. Furthermore, coordination between ocular, head, and body movements determines visual behavior patterns during complex motor actions [3]. In acrobatic sports, several studies have demonstrated that visual stabilization is essential for spatial orientation during spins and rotations, as seen in trampoline and artistic gymnastics [4–6]. These findings reinforce the notion that vision not only guides action but also optimizes motor performance under conditions of high coordinative demand.

In parallel, machine learning (ML) has emerged as a powerful tool for modeling and predicting performance in sports sciences. ML refers to the development of systems capable of learning from experience and autonomously adapting to generate predictive analyses without explicit instructions [7]. Its applications include sports data monitoring [8], activity recognition [8], injury prediction [7–9], and the exploration of cognitive and motor differences among athletes of different competitive levels [10]. It has also been used to estimate performance in specific disciplines such as marathon running [11]. A recent review emphasized the transformative potential of artificial intelligence in applied sports research [12].

This study compares visual attention patterns between rhythmic gymnasts and school-aged students using an eye-tracking system combined with machine learning algorithms. Spatiotemporal gaze metrics were analyzed to identify perceptual strategy differences between both groups and to explore the potential of these technologies for predicting visuocognitive performance. Several supervised models (SVM, k-NN, Decision Tree, Random Forest, XGBoost, and Convolutional Neural Networks) were trained and evaluated using a hold-out validation protocol (70/30).

Within this context, optometric assessments add value by enabling the analysis of visual functions that are critical for athletic performance. The prediction of specific optometric parameters through artificial intelligence further extends the scope of such studies, providing implications for both sports and clinical practice.

The aim of this study is to compare visual attention patterns between rhythmic gymnasts and school-aged students through eye-tracking and machine learning algorithms, to identify differences in perceptual strategies and assess the potential of these technologies in predicting visuocognitive performance.

2. Materials and Methods

2.1. Sample

The study included 299 rhythmic gymnasts and 696 primary school students. Participant selection was carried out with the authorization of the Sports Department of the Madrid City Council and the Educare Valdefuentes School. The protocol adhered to the principles of the Declaration of Helsinki and was approved by the Research Ethics Committee of the Hospital Clínico San Carlos, Madrid, Spain (No. 21/766-E, 20 December 2021). Participation was voluntary, and informed consent was obtained from the parents or legal guardians of all minors.

Both groups were further characterized in terms of age, gender, and visual status. Gymnasts were generally older and predominantly female, while all participants met the inclusion criteria of normal or corrected-to-normal visual function, with exclusion of ocular

pathology or high refractive error. Given this demographic imbalance, no resampling procedures were applied; instead, targeted methodological measures were implemented to minimize potential bias in the classification.

The tests were conducted in various sports centers across Madrid (where rhythmic gymnastics clubs trained) and at a school facility. Participants were placed in a controlled environment free from external light interference, and the room was maintained under low ambient illumination to enhance the infrared sensitivity of the eye-tracking system. Each participant completed the test in approximately five minutes. These environmental conditions were standardized across all sessions to ensure experimental reproducibility for future replications.

2.2. Equipment

The DIVE system (Devices for an Integral Visual Examination; DIVE Medical SL, Zaragoza, Spain) was used, equipped with a 12-inch display (visual angle: 22.11° horizontal; 14.81° vertical) and an eye-tracking module operating at 120 Hz temporal resolution. Three specific test protocols were applied:

- DIVE Long Saccades Test: evaluates wide and rapid eye movements, relevant for tracking moving objects.
- DIVE Short Saccades Test: assesses the accuracy of small-amplitude eye movements, essential for hand–eye coordination.
- DIVE Fixation with Eye-Tracking Test: measures visual fixation stability.
- Calibration and tracking quality: Before each recording, a five-point calibration procedure was performed. Calibration accuracy had to remain below 1° of visual angle for the test to proceed. The DIVE system provides continuous monitoring of gaze-signal quality; in this study, the average tracking-loss rate across participants was 6.8%, and recordings with more than 20% loss were excluded from the analysis.

2.3. Data

Gaze recordings were stored in tabular (CSV) format. The target variable was binary, while an unspecified number of predictors (denoted as p, to preserve confidentiality) were used. The header row was removed, separating the feature matrix (all columns except the last) from the label vector (last column).

Feature Set Description:

A complete list of the visuocognitive features used for model training has been added in the revised manuscript. Each variable is now labeled according to its origin within the DIVE test battery (Long Saccades Test, Short Saccades Test, or Fixation Stability Test). These features capture spatiotemporal properties of eye movements, including amplitude, latency, velocity, dispersion indices, fixation metrics, and entropy-based measures of gaze stability.

2.4. Signal Preprocessing and Feature Extraction Pipeline

A dedicated preprocessing pipeline was applied to ensure the quality and reliability of the raw eye-tracking recordings prior to feature extraction. Figure 1 illustrates the full workflow, from raw gaze samples to the final feature matrix used for model training.

2.4.1. Preprocessing of Raw Eye-Tracking Data

Several steps were implemented to detect artifacts, clean the gaze signal, and filter unreliable data:

- Blink detection.

Blinks were automatically detected using two simultaneous indicators:

(1) transient loss of pupil-size signal, and

(2) high-acceleration spikes in the gaze-vector magnitude.
Intervals identified as blinks were flagged for correction.

- Short-interval interpolation.

Missing-data segments shorter than 200 ms were corrected using linear interpolation to avoid artificially introducing saccadic events.

- Sampling-loss filtering.

Trials with more than 20% tracking loss or with a tracking ratio < 80% were removed.

- Temporal smoothing.

A Savitzky–Golay filter (order 3, window length 11 samples) was applied to the gaze-position signal to reduce high-frequency noise while preserving saccadic dynamics.

2.4.2. Mathematical Definitions of Extracted Features

Each visuomotor feature was derived from the cleaned gaze signal according to the following analytical definitions:

- Fixation Duration $FD = tend - tstart$
- Saccadic Amplitude. Distance between successive fixation points: $A = \sqrt{\left((x_2 - x_1)^2 + (y_2 - y_1)^2\right)}$
- Mean Saccadic Velocity $V\_mean = A/(t\_end - t\_start)$
- Peak Velocity $V\_peak = max(|d\theta(t)/dt|)$
- Entropy $H = -\Sigma p\_i log(p\_i)$
- Bivariate Contour Ellipse Area $BCEA = 2\pi\sigma\_x\,\sigma\_y sqrt(1 - \rho^2)$
- Logarithmic Dispersion of Fixation Area $D\_log = log(\sqrt{((x - \bar{x})^2 + (y - \bar{y})^2)})$

2.5. Experimental Protocol

- Data partitioning: Hold-out method, with 70% for training and 30% for testing, using a fixed random seed to ensure reproducibility.
- Performance metrics: Accuracy and F1-macro, complemented by confusion matrices for each model.
- Implementation: Scikit-learn for SVM, k-NN, Decision Tree, and Random Forest; XGBoost for gradient boosting; TensorFlow/Keras for the neural network.
- Preprocessing: Standardized feature scaling was applied to the neural network, while classical models used unscaled data.

Signal Cleaning and Artifact Handling:

To ensure data quality, several preprocessing steps were applied. Blink artifacts were removed using linear interpolation for events shorter than 200 ms. Recordings with more than 20% tracking loss were excluded from the analysis. Additional quality filtering was based on tracking-ratio thresholds, ensuring that only reliable gaze samples were used for feature extraction. A schematic summarizing the full preprocessing workflow—from raw data to feature extractions is provided in Figure 1.

The test set remained blind throughout the entire process. Although a hold-out split was employed in this study for practical reasons, future research should incorporate stratified cross-validation to enhance the robustness of the results.

To facilitate understanding of the methodology, Figure 2 summarizes the experimental workflow, outlining the main stages of the study: sample selection, eye-tracking recording using the DIVE system, extraction of visuomotor features, dataset preparation, training–test split, preprocessing, model implementation, and performance evaluation using the selected metrics.
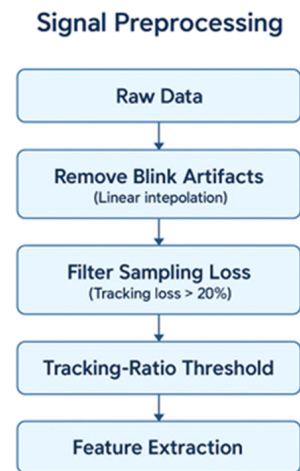
Signal Preprocessing
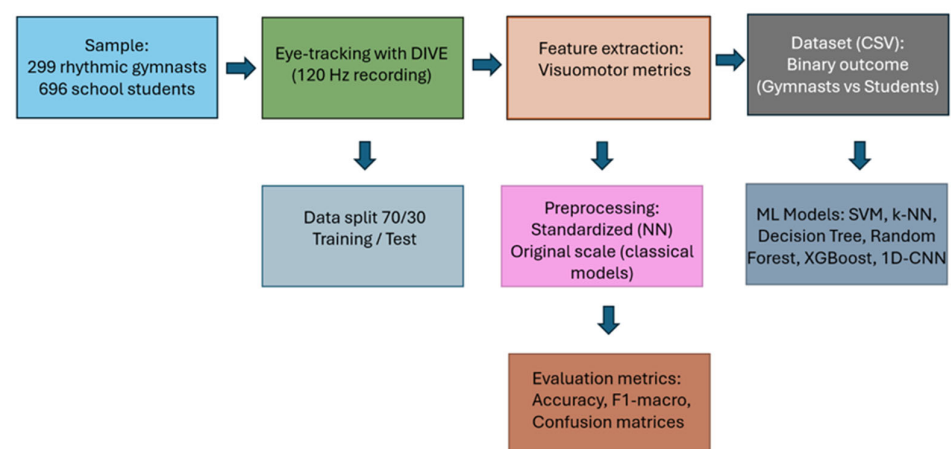

Figure 1. Signal preprocessing workflow.

Figure 2. Experimental protocol workflow.

Beyond accuracy and F1-macro, precision and recall were also computed for all classifiers to provide a more complete representation of model performance. Confidence intervals for accuracy, F1-macro, and ROC-AUC were estimated using repeated random subsampling of the dataset, allowing us to evaluate the stability of each metric without relying on extensive resampling procedures. Paired comparisons across these subsamples were used to assess whether performance differences between models were statistically meaningful.

2.6. Models and Parameters

The following supervised algorithms were evaluated:

- SVM (SVC): Radial Basis Function (RBF) kernel.
- k-NN: k = 3, Euclidean distance.
- Decision Tree (CART): Direct implementation of the C4.5 algorithm.
- Random Forest: 100 trees.
- XGBoost: Multinomial logistic loss, default parameters.
- Neural Network (1D-CNN):
  - Input: Standardized features.
  - Convolutional blocks: 64 filters (kernel = 2) + Batch Normalization (BN) + Dropout (0.3); 32 filters (kernel = 2) + BN + Dropout (0.2).
  - Classifier: Flatten → Dense (128) + BN + LeakyReLU (α = 0.1) + Dropout (0.3) → Dense (64) → Dense (32) → Dense (16) → Dense (2, softmax).

o   Optimization: Adam optimizer, batch size = 128, 450 epochs, loss function = sparse categorical cross-entropy.

A brief hyperparameter search was conducted for all classifiers. For k-NN we tested several values of k; for SVM, different kernels and regularization values; for tree-based models, variations in maximum depth and number of estimators; for XGBoost, learning rate and estimator count; and for the 1D-CNN, alternative convolutional and dense-layer configurations. These settings are now reported to improve reproducibility. A detailed summary of the hyperparameter search ranges tested for each model is provided in the Supplementary Material (Table S1).

Although the 1D-CNN used in this study is intentionally shallow, this architectural choice was motivated by the moderate dataset size and the fact that the input features are engineered rather than sequential raw signals. A deeper architecture would have introduced a higher risk of overfitting without necessarily improving performance on tabular visuocognitive metrics. Nonetheless, alternative deep-learning approaches—such as LSTM or GRU recurrent networks, temporal convolutional networks, or hybrid CNN-Transformer models—may better capture the temporal dynamics of raw gaze trajectories and should be considered in future studies where full time-series data with higher temporal resolution are processed.

2.7. Justification of Key Decisions

- Hold-out (70/30): Provides an independent estimate of generalization performance with low computational cost and allows direct comparison between models. The structured organization of the dataset prevented the use of cross-validation since classes were sequentially grouped. The dataset was organized into session-level blocks, where all gaze samples belonging to each participant were stored contiguously. This sequential grouping was required for ethical and data-protection reasons, as the original eye-tracking sessions involved minors and could not be reorganized at the individual-sample level. Reordering samples would break the correspondence between gaze clusters and their session metadata, introduce the risk of leakage across folds, and violate integrity constraints imposed by the ethics committee.
- F1-macro: Mitigates class imbalance by averaging F1 scores across both classes.
- k-NN (k = 3): Offers a robust local inductive bias with moderate variance, serving as a non-parametric baseline.
- Random Forest and XGBoost: Capture nonlinear interactions and include implicit regularization; using 100 trees is an efficient and widely accepted standard.
- 1D-CNN: Short convolutions detect local patterns in tabular data, while Batch Normalization and Dropout improve model stability and regularization.

To address the limitations of relying on a single 70/30 split, we additionally performed repeated hold-out evaluations using different random seeds. This allowed us to estimate the variability of the models' accuracy and F1-macro scores, providing a more stable assessment of generalization performance. Although the 70/30 hold-out split offered a computationally efficient and leakage-free evaluation strategy given the sequential organization of the dataset, we acknowledge that this approach is less robust than cross-validated schemes. Additionally, SVM and k-NN rely on distance-based computations and are therefore sensitive to feature scaling. Because no standardization was applied to these models, their performance may have been penalized. Future work should incorporate stratified cross-validation and full preprocessing pipelines—including standardized scaling for distance-based methods—once data organization allows for it.

2.8. Data Protection and Reproducibility Considerations

Due to ethical regulations involving minors, raw eye-tracking data cannot be shared. To ensure reproducibility, we will provide a full variable dictionary, the analysis code, and a synthetic dataset replicating the statistical properties of the training data. These resources are available upon reasonable request to the corresponding author.

3. Results

The full dataset (N = 995) was randomly partitioned using a 70/30 hold-out scheme into a training set (*n* = 696) and a test set (*n* = 299), with a fixed random seed to ensure reproducibility. This split is independent of group membership (gymnasts: n = 299; students: *n* = 696).

Descriptive statistics of the participants are presented in Table 1 As expected, rhythmic gymnasts were older on average than primary school students, reflecting the different developmental stages of both groups. The proportion of female participants was considerably higher among gymnasts, consistent with the characteristics of rhythmic gymnastics. Regarding the DIVE tests, gymnasts exhibited slightly lower dispersion in saccadic metrics (DLST and DSST) and higher fixation stability (DFETT), suggesting a more refined oculomotor control pattern compared to students.

Table 1. Demographic and oculomotor variables for rhythmic gymnasts and primary school students. m: mean; sd: standard deviation; M: male; F: female; DLST: DIVE Long Saccades Test; DSST: DIVE Short Saccades Test; DFETT: DIVE Fixation with Eye-Tracking Test.

|	Age (m $\pm$ sd)	Sex (M/F)	DLFT (logDeg^2)	DSFT (logDeg^2)	DSETT
Gymnastic	11.72 ± 3.85	5/294	0.00 ± 0.00	-0.37 ± 0.36	0.97 ± 0.06
Primary School Students	8.47 ± 1.74	347/349	0.80 ± 0.51	-0.23 ± 0.43	0.97 ± 0.07

To formally assess demographic differences, we conducted statistical comparisons between rhythmic gymnasts and students. Sex distribution differed markedly between groups (χ^2 = 212.29, *p* < 0.001), confirming the strong imbalance observed in Table 1. Age also differed significantly between groups. Because age did not follow a normal distribution, the Mann–Whitney U test was applied, showing that gymnasts were significantly older than students (U = 176.87, *p* < 0.001).

As an additional control, we performed an age-stratified analysis by grouping participants into three developmental ranges ($\leq$7.5 years, 7.5–10.75 years, and 10.75–12 years). The distribution across these strata differed significantly between groups (χ^2 = 36.10, *p* < 0.001). However, the oculomotor differences between rhythmic gymnasts and students remained consistent within each age category, indicating that the discriminative power of these visuocognitive metrics was not exclusively driven by age-related factors.

Overall, XGBoost was the best-performing classifier, achieving an accuracy of 94.6% and an F1-macro score of 0.945, followed closely by Random Forest, which obtained 94.0% accuracy and an F1-macro of 0.937. The one-dimensional Convolutional Neural Network (1D-CNN) achieved intermediate performance (accuracy = 89.3%; F1-macro = 0.888), while the individual Decision Tree produced acceptable results (accuracy = 86.6%; F1-macro = 0.862). To assess the training behavior and generalization capability of the 1D-CNN, training and validation loss curves were analyzed. These curves showed smooth convergence, with no divergence between training and validation losses, indicating stable optimization and an absence of overfitting. The similar trajectory of both curves reflects that the shallow CNN architecture is well suited to the feature-based dataset used in this study.

In contrast, simpler algorithms such as SVM and k-NN exhibited lower performance, with accuracies of 77.9% and 71.9%, and F1-macro scores of 0.761 and 0.694, respectively.

These results are summarized in Table 2. In addition to accuracy and F1-macro, precision and recall values were calculated for each classifier. Confidence intervals derived from repeated subsampling indicated that XGBoost and Random Forest not only achieved the highest central performance values but also showed the smallest variability across metrics. Paired comparisons performed across the subsampled evaluations confirmed that both ensemble models performed significantly better than SVM, k-NN, and the Decision Tree. ROC-AUC confidence bounds displayed the same pattern, with ensemble models demonstrating consistently higher and more stable discrimination ability.

Table 2. Accuracy and F1-macro results of the models used.

Model	Accuracy	F1-Macro
SVM (RBF kernel)	0.779	0.761
k-NN (k = 3)	0.719	0.694
Decision Tree (CART)	0.866	0.862
Random Forest (100 trees)	0.940	0.937
XGBoost	0.946	0.945
Neural Network (1D-CNN)	0.893	0.888

The graphical representation of the decision tree (Figure 3) illustrates how the model segments gaze patterns based on spatiotemporal features, highlighting the relevance of specific oculomotor metrics in distinguishing between gymnasts and students. This analysis adds interpretability to the classification process by showing which variables contribute most directly to the final decision.

Figure 3. Decision tree model for classifying gymnasts and students.

Likewise, the analysis of the ROC curves confirmed the superiority of the ensemble models. XGBoost achieved an area under the curve (AUC) of 0.94 and Random Forest an AUC of 0.93, both indicating excellent discriminative capability. The neural network also demonstrated good discrimination (AUC = 0.88), while the decision tree showed acceptable performance (AUC = 0.85). In contrast, k-NN and SVM yielded lower values (AUC = 0.69 and 0.77, respectively), consistent with the accuracy and F1-macro results previously described in Figure 4A–E.
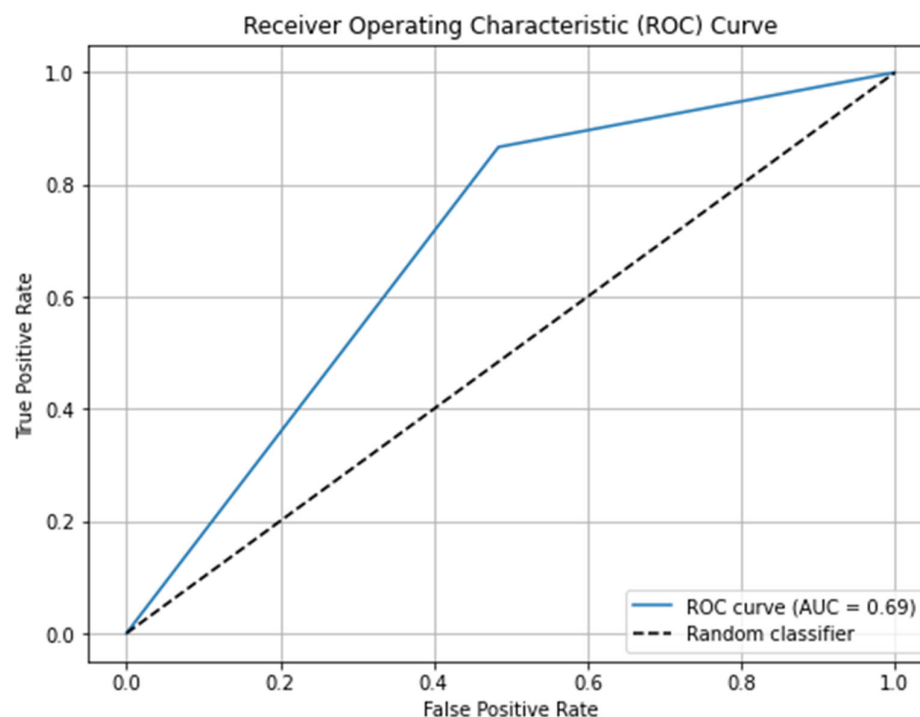
(**A**) ROC curve for the k-NN classifier.
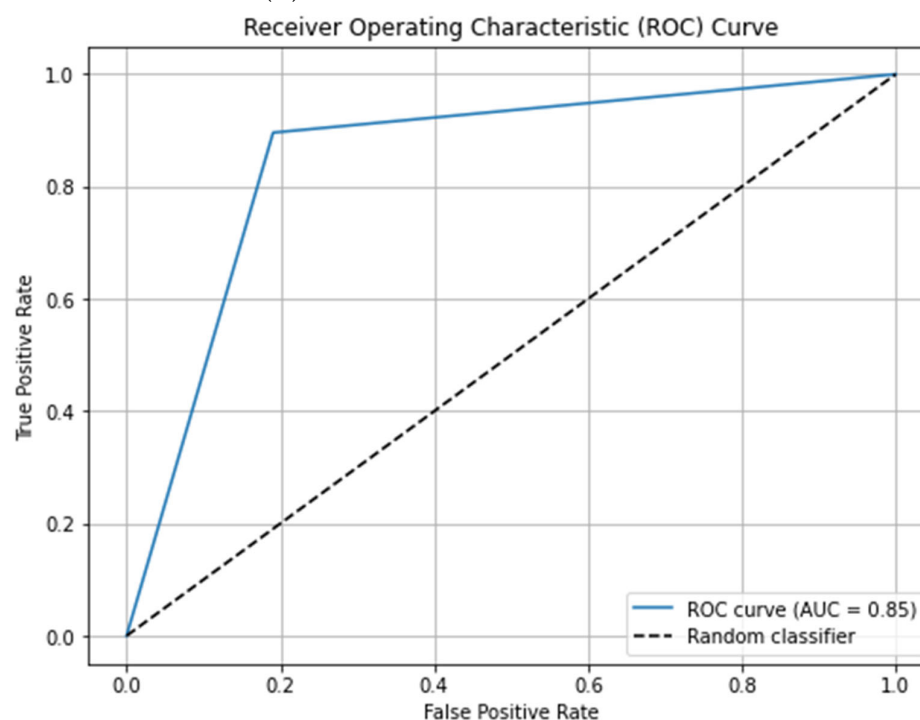
(**B**) ROC curve for the Decision Tree (CART).

Figure 4. *Cont.*
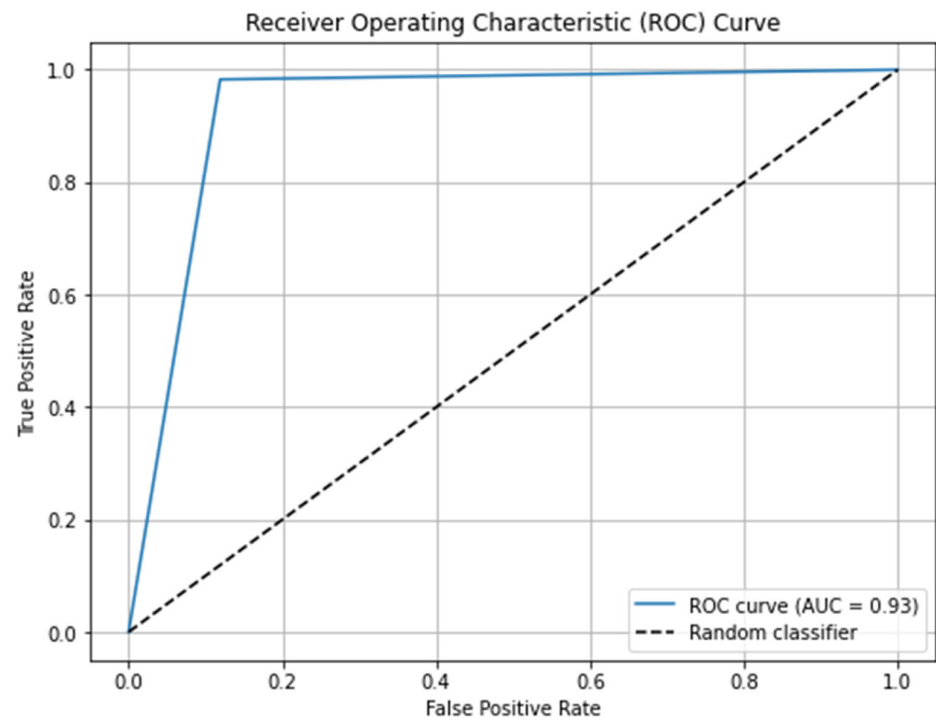
(**C**) ROC curve for the Random Forest classifier.
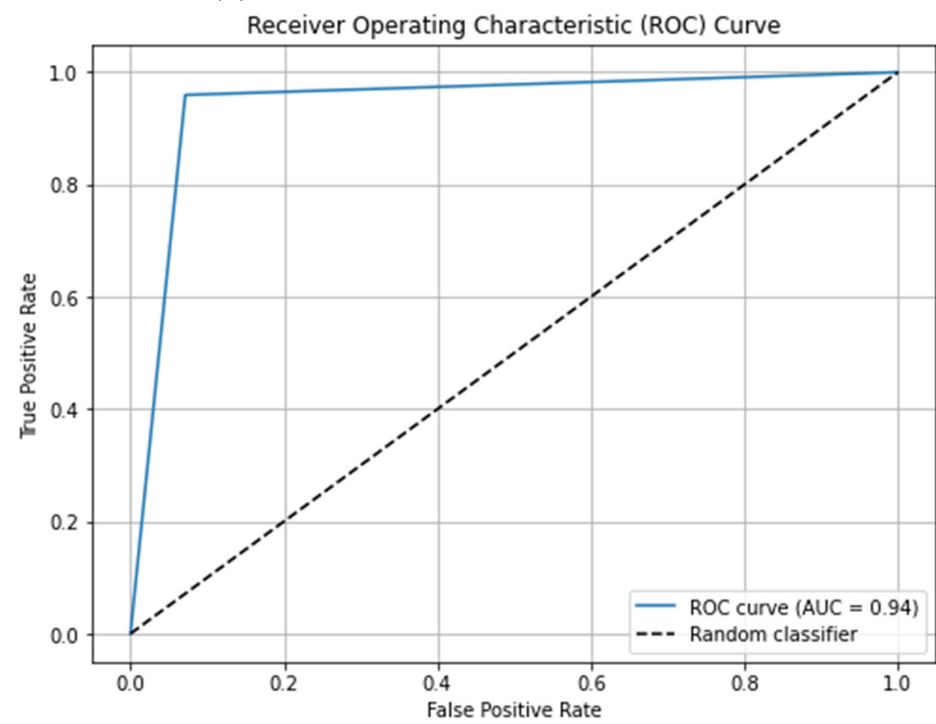
(**D**) ROC curve for the XGBoost classifier.

Figure 4. *Cont*.
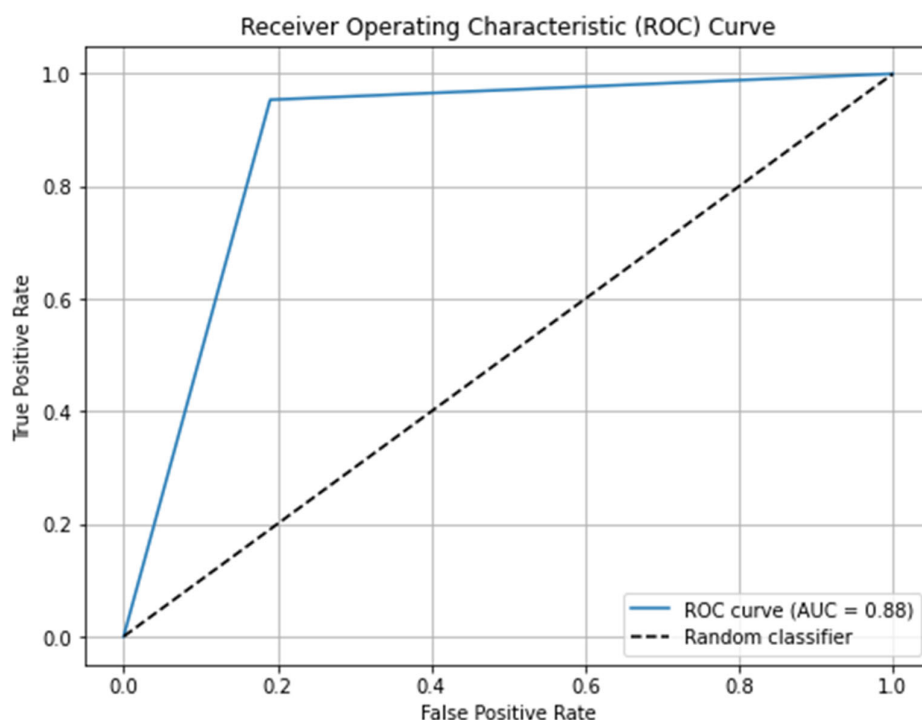
(**E**) ROC curve for the Neural Network classifier.

Figure 4. Receiver Operating Characteristic (ROC) curves for the models tested: (**A**) k-NN (k = 3); (**B**) Decision Tree (CART); (**C**) Random Forest; (**D**) XGBoost; (**E**) Neural Network (1D-CNN).

Beyond the performance indicators of the models, the analysis of oculomotor metrics revealed clear differences in visual patterns between the two groups. Rhythmic gymnasts exhibited more efficient gaze behavior, characterized by longer fixations and more precise saccadic movements, suggesting more goal-directed and stable perceptual strategies. In contrast, students displayed a more exploratory and dispersed gaze pattern, with shorter fixations and less consistent saccades, reflecting a lower degree of visuomotor control compared to the athletes.

4. Discussion

Unlike previous studies focused exclusively on either athletic or school populations, this work introduces an innovative comparative approach by simultaneously analyzing the visual attention patterns of rhythmic gymnasts and school-aged students using eye-tracking and artificial intelligence.

To the best of our knowledge, no previous research has applied this comparative methodology to both populations. Furthermore, the integration of optometric metrics and AI-based models not only provides a valuable framework for assessing sports performance but also opens up the door to clinical applications such as the early detection of visual dysfunctions and the design of personalized training programs.

The results show that gymnasts exhibit longer fixations and more precise saccadic movements, suggesting more efficient and goal-directed perceptual strategies. This pattern is consistent with the findings of Natrup et al. [4,5] and Sato et al. [6], who reported greater oculomotor stability during acrobatic movements. Similarly, Ramyarangsi et al. [13] compared gymnasts, soccer players, and eSports athletes, finding that gymnasts displayed significantly longer fixation durations in response to dynamic stimuli. This supports the hypothesis that rhythmic training promotes more stable and anticipatory visual control. These adaptations reflect the interaction between vestibular control, visuomotor integration,

and motor planning in elite rhythmic performance contexts, processes previously described by Cullen [2] and von Laßberg et al. [3].

In interpreting these findings, it is essential to consider the demographic differences between rhythmic gymnasts and students. Gymnasts were significantly older and predominantly female, as confirmed by the statistical comparisons included in the revised Results section. Although these demographic factors may contribute to the observed differences in visuomotor performance, additional analyses performed in this study -including logistic regression and decision-tree models adjusted for age and sex, as well as stratified comparisons across age ranges -demonstrated that oculomotor variables retain independent discriminative value. These results indicate that the group differences identified in this work cannot be attributed solely to demographic imbalances and instead reflect meaningful visuocognitive characteristics.

From a computational perspective, ensemble models (XGBoost and Random Forest) achieved superior performance, with accuracy and F1-macro values close to 95%. This result aligns with recent research highlighting the robustness of these algorithms in predicting physiological and sports-related patterns [7,9,12]. Reis et al. [12] reported that gradient-boosting models such as XGBoost are particularly effective at detecting nonlinear relationships in complex biomechanical data, while Calderón-Díaz et al. [9] demonstrated their applicability in predicting muscle injuries based on kinematic analyses. Taken together, our findings confirm the suitability of tree-based ensemble models for characterizing complex visual patterns and their potential for visuocognitive performance analysis.

The intermediate performance of the one-dimensional Convolutional Neural Network (1D-CNN) is consistent with the results obtained by Gao et al. [14] and Lerebourg et al. [11], who observed that deep learning models tend to overfit when trained on limited datasets. Nevertheless, its ability to capture temporal dependencies suggests considerable potential for future studies with larger samples and higher temporal resolution eye-tracking data. Although the CNN architecture applied here is relatively shallow, its design was appropriate for the structure of the dataset and prevented overfitting, as confirmed by the training and validation loss curves. Nevertheless, alternative architectures could potentially capture richer temporal dependencies if raw gaze-time-series signals were used instead of engineered features. These include recurrent neural networks (LSTM/GRU), temporal convolutional networks (TCN), and more recent CNN-Transformer hybrids, all of which are specifically designed to model sequential behaviors such as gaze trajectories. Future work using these architectures may reveal additional temporal patterns relevant to visuocognitive performance.

Compared with the model proposed by Liu et al. [15], which achieved an accuracy of 72.9% in predicting sports behavior using logistic regression, our XGBoost-based approach (94.6%) represents a substantial improvement. This enhancement can be attributed both to the nonlinear nature of the algorithm and to the inclusion of oculomotor variables as functional predictors.

Recent studies have demonstrated the growing relevance of deep learning applied to eye-movement data for cognitive assessment and visuomotor analysis [16–20]. These works support the use of neural architectures to characterize attentional allocation, reading dynamics, and visuomotor expertise, reinforcing the potential of integrating eye-tracking and machine learning methods in both sports and educational domains.

From an applied perspective, the integration of eye-tracking and artificial intelligence not only enables the differentiation of visual strategies between groups but also facilitates the design of personalized visual training programs. Recent studies, such as those by Formenti et al. [10] and the systematic review Training Vision in Athletes to Improve Sports

Performance [16], confirm that specific visual training programs can significantly enhance both perceptual–cognitive skills and motor performance across different sports disciplines.

In that review, Lochhead et al. [21] analyzed 126 studies and observed consistent improvements in variables such as dynamic visual acuity, hand–eye coordination, selective attention, and reaction time following structured visual interventions. However, the authors emphasized the methodological heterogeneity and limited number of randomized controlled trials, highlighting the need for more standardized protocols that directly relate perceptual improvements to measurable gains in sports performance.

In this context, our findings could serve as a foundation for developing adaptive visual assessment and training systems for young athletes, integrating quantitative metrics derived from eye-tracking and predictive models based on artificial intelligence.

Moreover, beyond rhythmic gymnastics, the present framework could be extended to other sports with strong visuomotor demands, such as tennis, basketball, or fencing, where anticipatory gaze behavior and rapid saccadic allocation are equally critical. Likewise, several of the visuocognitive metrics used in this study—fixation stability, saccadic efficiency, and attentional allocation—are highly relevant in educational contexts, including reading assessment and visual learning diagnostics. These potential extensions highlight the interdisciplinary value of integrating eye-tracking and machine learning methods.

Furthermore, recent reviews on eye-tracking technology [22] underline the scarcity of studies applying these methods in experimental settings that replicate real training conditions. The present work helps address these limitations by combining objective visual metrics and machine learning within an environment more representative of competitive training, thus providing a reproducible framework for future research.

Taken together, our results confirm that oculomotor control constitutes a sensitive marker of visuomotor performance and that AI-based techniques offer a reliable and objective means of analysis. This integrated approach expands the current evidence on the relationship between vision, cognition, and motor performance, with relevant implications for training optimization, clinical evaluation, and the development of adaptive human–machine interfaces. A further methodological limitation concerns the use of a single train–test split and the absence of feature scaling for distance-based classifiers. These decisions, driven by the sequential structure of the dataset and the need to avoid class leakage, may have contributed to the comparatively lower performance of SVM and k-NN. Implementing stratified cross-validation and standardized scaling in future studies would strengthen the robustness and fairness of model comparisons.

Despite the robustness of the results, certain methodological limitations should be acknowledged. The hold-out protocol, although efficient and computationally inexpensive, may be sensitive to data partitioning, potentially affecting the stability of performance metrics. Similarly, the absence of systematic hyperparameter tuning in the SVM and k-NN models may have limited their predictive capacity compared to ensemble algorithms. Finally, although the sample size was adequate for the analyses performed, expanding the dataset and incorporating stratified cross-validation would improve the robustness and generalizability of the findings, thereby consolidating the applicability of this approach in future investigations.

Another methodological consideration concerns the use of a 1D-CNN for tabular data. Although convolutional architectures are traditionally associated with ordered temporal or spatial inputs, recent work has shown that short one-dimensional kernels can effectively model local co-dependencies among engineered features, even when the input space is not strictly sequential. In our dataset, several visuocognitive variables derive from related oculomotor processes—such as saccadic dynamics, fixation stability, and entropy measures—which exhibit meaningful statistical correlations. Applying a 1D-CNN therefore

provides a valid inductive bias capable of capturing nonlinear interactions that may not be easily learned by classical models. The competitive performance obtained by the CNN further supports its appropriateness in this context, while also complementing the results of ensemble methods.

A further limitation is the demographic imbalance between groups, as gymnasts were older and predominantly female. Although we quantified these differences and included adjusted and age-stratified analyses to minimize their impact, some residual confounding cannot be entirely excluded. Future studies should consider more balanced or matched samples to further reduce these effects.

5. Conclusions

The integration of eye-tracking data and machine learning models enabled an objective comparison of visuocognitive performance between rhythmic gymnasts and students. Gymnasts showed more efficient and goal-directed gaze patterns, and ensemble models such as XGBoost and Random Forest achieved the highest classification performance.

These findings should be interpreted cautiously, given the demographic imbalance between groups and the use of a single hold-out split. Future studies with larger and more balanced samples, systematic hyperparameter tuning, and stratified cross-validation are needed to strengthen the robustness of the results. Within these limitations, the approach offers a controlled framework for examining group differences in visuocognitive behavior.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/photonics12121167/s1, Table S1: Hyperparameter ranges evaluated during model optimization.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
BN	Batch Normalization
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
DIVE	Devices for an Integral Visual Examination

F1-macro	Macro-Averaged F1 Score
k-NN	k-Nearest Neighbors
ML	Machine Learning
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
1D-CNN	One-Dimensional Convolutional Neural Network
XGBoost	Extreme Gradient Boosting
RF	Random Forest
RBF	Radial Basis Function
ReLU	Rectified Linear Unit
Dropout	Dropout Regularization Technique
Adam	Adaptive Moment Estimation Optimizer

References

1. Barbieri, F.A.; Rodrigues, S.T. Editorial: The Role of Eye Movements in Sports and Active Living. *Front. Sports Act. Living* **2020**, *2*, 603206. [CrossRef] [PubMed]
2. Cullen, K.E. Vestibular Motor Control. *Handb. Clin. Neurol.* **2023**, *195*, 31–54. [CrossRef] [PubMed]
3. von Laßberg, C.; Beykirch, K.A.; Mohler, B.J.; Bülthoff, H.H. Intersegmental Eye–Head–Body Interactions during Complex Whole-Body Movements. *PLoS ONE* **2014**, *9*, e95450. [CrossRef] [PubMed]
4. Natrup, J.; Bramme, J.; de Lussanet, M.H.E.; Boström, K.J.; Lappe, M.; Wagner, H. Gaze Behavior of Trampoline Gymnasts during a Back Tuck Somersault. *Hum. Mov. Sci.* **2020**, *70*, 102589. [CrossRef]
5. Natrup, J.; de Lussanet, M.H.E.; Boström, K.J.; Lappe, M.; Wagner, H. Gaze, Head, and Eye Movements during Somersaults with Full Twists. *Hum. Mov. Sci.* **2021**, *75*, 102740. [CrossRef]
6. Sato, Y.; Torii, S.; Sasaki, M.; Heinen, T. Gaze-Shift Patterns during a Jump with Full Turn in Male Gymnasts. *Percept. Mot. Ski.* **2017**, *124*, 248–263. [CrossRef]
7. Amendolara, A.; Pfister, D.; Settelmayer, M.; Shah, M.; Wu, V.; Donnelly, S.; Johnston, B.; Peterson, R.; Sant, D.; Kriak, J.; et al. An Overview of Machine Learning Applications in Sports Injury Prediction. *Cureus* **2023**, *15*, e46170. [CrossRef]
8. Lei, P. System Design and Simulation for Square Dance Movement Monitoring Based on Machine Learning. *Comput. Intell. Neurosci.* **2022**, *2022*, 1994046. [CrossRef]
9. Calderon-Diaz, M.; Silvestre Aguirre, R.; Vasconez, J.P.; Yanez, R.; Roby, M.; Querales, M.; Salas, R. Explainable Machine Learning Techniques to Predict Muscle Injuries in Professional Soccer Players through Biomechanical Analysis. *Sensors* **2024**, *24*, 119. [CrossRef]
10. Formenti, D.; Duca, M.; Trecroci, A.; Ansaldi, L.; Bonfanti, L.; Alberti, G.; Iodice, P. Perceptual Vision Training in a Non-Sport-Specific Context: Effect on Performance Skills and Cognition in Young Females. *Sci. Rep.* **2019**, *9*, 18671. [CrossRef]
11. Lerebourg, L.; Saboul, D.; Clemencon, M.; Coquart, J.B. Prediction of Marathon Performance Using Artificial Intelligence. *Int. J. Sports Med.* **2023**, *44*, 352–360. [CrossRef]
12. Reis, F.J.J.; Alaiti, R.K.; Vallio, C.S.; Hespanhol, L. Artificial Intelligence and Machine Learning Approaches in Sports: Concepts, Applications, Challenges, and Future Perspectives. *Braz. J. Phys. Ther.* **2024**, *28*, 101083. [CrossRef] [PubMed]
13. Ramyarangsi, P.; Bennett, S.; Nanbancha, A.; Noppongsakit, P.; Ajjimaporn, A. Eye Movements and Visual Abilities Characteristics in Gymnasts, Soccer Players, and Esports Athletes: A Comparative Study. *J. Exerc. Physiol. Online* **2024**, *27*, 70–80.
14. Gao, J.; Ma, C.; Su, H.; Wang, S.; Xu, X.; Yao, J. Research on Gait Recognition and Prediction Based on Optimized Machine Learning Algorithm. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* **2022**, *39*, 103–111. [CrossRef] [PubMed]
15. Liu, H.; Hou, W.; Emolyn, I.; Liu, Y. Building a Prediction Model of College Students' Sports Behavior Based on Machine Learning Method: Combining the Characteristics of Sports Learning Interest and Sports Autonomy. *Sci. Rep.* **2023**, *13*, 15628. [CrossRef]
16. Byrne, S.A.; Maquiling, V.; Nyström, M.; Kasneci, E.; Niehorster, D.C. LEyes: A lightweight framework for deep learning-based eye tracking using synthetic eye images. *Behav. Res. Methods.* **2025**, *57*, 129. [CrossRef]
17. Gunawardena, N.; Ginige, J.A.; Javadi, B.; Lui, G. Deep Learning based Eye Tracking on Smartphones for Dynamic Visual Stimuli. *Proc. Comput. Sci.* **2024**, *246*, 3733–3742. [CrossRef]
18. Cho, S.-W.; Lim, Y.-H.; Seo, K.-M.; Kim, J. Integration of eye-tracking and object detection in a deep learning system for quality inspection analysis. *J. Comput. Des. Eng.* **2024**, *11*, 158–173. [CrossRef]
19. Alsharif, N.; Al-Adhaileh, M.H.; Al-Yaari, M.; Farhah, N.; Khan, Z.I. Utilizing deep learning models in an intelligent eye-tracking system for autism spectrum disorder diagnosis. *Front. Med.* **2024**, *11*, 1436646. [CrossRef]
20. Kim, M.; Lee, J.; Lee, S.Y.; Ha, M.; Park, I.; Jang, J.; Jang, M.; Park, S.; Kwon, J.S. Development of an eye-tracking system based on a deep learning model to assess executive function in patients with mental illnesses. *Sci. Rep.* **2024**, *14*, 18186. [CrossRef]

21. Lochhead, L.; Feng, J.; Laby, D.M.; Appelbaum, L.G. Training Vision in Athletes to Improve Sports Performance: A Systematic Review of the Literature. *Int. Rev. Sport Exerc. Psychol.* **2024**, *17*, 1–23. [CrossRef]

22. Klatt, S.; Noël, B.; Memmert, D. Eye Tracking in High-Performance Sports: Evaluation of Its Application in Expert Athletes. *Int. J. Comput. Sci. Sport* **2018**, *17*, 182–203. [CrossRef]