Proc. of International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA 2025)
7-9 August 2025, Antalya-Türkiye

2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) | 979-8-3315-3562-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACDSA65407.2025.11166437

Predicting NHL Performance from Cognitive-Motor Metrics After Concussion: An Explainable AI Approach

Michael Moschitto
Computer Science and Software Engineering
California Polytechnic State University
San Luis Obispo, California 93407
mmoschit@calpoly.edu

Lauren E. Sergio
Faculty of Health York University,
Toronto, ON, Canada
lsergio@yorku.ca

Marin Litoiu
Lassonde School of Engineering York University,
Toronto, ON, Canada
mlitoiu@yorku.ca

Joydeep Mukherjee
Computer Science and Software Engineering
California Polytechnic State University
San Luis Obispo, California 93407
jmukherj@calpoly.edu

Sumona Mukhopadhyay
Computer Science and Software Engineering
California Polytechnic State University
San Luis Obispo, California 93407
mukhopad@calpoly.edu

***Abstract*—In this work, we introduce a novel approach using cognitive-motor integration (CMI) metrics and concussion history to predict NHL participation. Specifically, we leverage data from the BrDi Test—a validated neuroscience-based assessment involving visuomotor tracing tasks under both standard and non-standard conditions. Using data from the BrDi Test—a neuroscience-based cognitive-motor integration task this research explores the link between mild traumatic brain injury (mTBI), or concussion, and NHL participation. We emphasize both model accuracy and interpretability to support human scouting decisions by comparing explainable machine learning (XAI) models with traditional black-box approaches. We frame the problem as a binary classification task and evaluate a range of machine learning models. Despite the limitations of a small, class-imbalanced dataset, a class-weighted Decision Tree achieved the best average generalization performance, offering strong interpretability alongside competitive performance. Our contributions include the first application of BrDi data to NHL prediction, a comparative analysis of XAI and black-box models for talent evaluation, and a modular ML pipeline adaptable to small-sample, high-dimensional datasets. The results support the potential value of cognitive metrics in augmenting traditional scouting decisions and provide a foundation for future research as more data becomes available.**

***Keywords*—cognitive, motor, concussion, sports analytics, explainable AI**

I. Introduction

Artificial Intelligence (AI) and data science are transforming sports analytics, enabling evaluations that go beyond physical performance metrics to incorporate cognitive and neurological data [1]. In high-speed, cognitively demanding sports like ice hockey, brain function—particularly the integration of thought and action—is crucial to elite-level success. This research explores whether cognitive-motor impairments following concussion can be used to predict National Hockey League (NHL) participation using explainable machine learning (ML) models.

Concussions, or mild traumatic brain injuries (mTBI), are known to impair cognitive-motor integration (CMI)—the coordination between cognitive processes and motor execution [2], [3]. While athletes are typically cleared to return after symptom resolution, research shows that subtle deficits in decision-making and motor control may persist [4], [5]. These residual impairments are often undetectable through traditional medical evaluations but can still impact game performance at the professional level. To capture these subtle deficits, we use the BrDi Test, a validated visuomotor task that measures CMI under both standard and perturbed sensory conditions [6], [7]. Prior studies suggest that athletes with concussion history may show diminished BrDi performance, even after meeting return-to-play criteria. This study is motivated by the hypothesis that:

Hypothesis: Athletes with a history of concussion who exhibit greater impairments in CMI—as measured by BrDi Test performance—are less likely to play in the NHL than those with better post-concussion CMI.

To test this hypothesis, we aim to determine whether BrDi-based visuomotor performance metrics can predict NHL participation among previously concussed athletes using machine learning models that balance predictive accuracy with interpretability. We frame this as a binary classification problem,

where the positive class indicates that a player reached the NHL, and the negative class indicates they did not. We evaluate a suite of models including interpretable approaches (e.g., Decision Tree, Logistic Regression) and black-box models (e.g., XGBoost, Neural Networks) [8], [9]. Despite challenges such as small sample size (93 samples) and class imbalance, our best-performing model—a weighted Decision Tree—achieved an average generalization F1 score of 0.603, suggesting it could identify meaningful post-concussion cognitive patterns linked to NHL success. Importantly, this work emphasizes explainable AI (XAI) in athlete evaluation. While some black-box models offer higher accuracy, XAI ensures transparency, trust, and interpretability for decision-makers such as coaches, scouts, and medical staff [10]. Our contributions are: (1) First application of BrDi-based cognitive-motor performance metrics in NHL draft outcome prediction; (2) A comparative evaluation of explainable AI versus black-box models in the context of sports talent assessment; (3) Development of a scalable, modular ML pipeline suited for small-sample, high-dimensional data common in elite sports research; (4) Preliminary evidence that post-concussion CMI impairments, even if subclinical, may influence elite-level outcomes.

II. Methodology

A. Dataset: BrDi Test

This section gives background to the neuroscience experiment, named the BrDi Test, from which the data set used in this study was collected. The experiment tests the cognitive-motor integration (CMI) of hockey players using a tracing task and includes multiple variations that mimic motor challenges in hockey scenarios. In the BrDi test, participants completed two computer-based CMI tasks on a dual-touch touch-screen laptop. The tasks required participants to slide their index finger along the screen to move a cursor from a central target to one of four peripheral targets. To replicate the dynamic physical scenarios in sports, the study employs standard and non-standard conditions during the trials. The standard condition involved participants looking at and reacting to the target presented on the same vertical screen. This resembles an action such as reaching for a coffee cup or shooting at a goal while looking at it. In contrast, the non-standard condition presents two additional challenges, cue reversal, and plane change. During cue reversal, the cursor moves in the opposite direction of the finger movement (180° reversal). During plane change, targets are presented on a vertical screen, and participants respond on a horizontal screen. This movement is similar to using a brake pedal or a computer mouse. A hockey-specific equivalent of the nonstandard task includes passing the puck to a teammate on one side while navigating an obstacle on the other [11] as shown in Fig. 1. This was conducted on elite hockey athletes and finds potential lingering deficits of concussion specifically during nonstandard conditions. These findings present similar detriments as seen due to the dual task model and may leave athletes vulnerable to future injury and long-term effects.

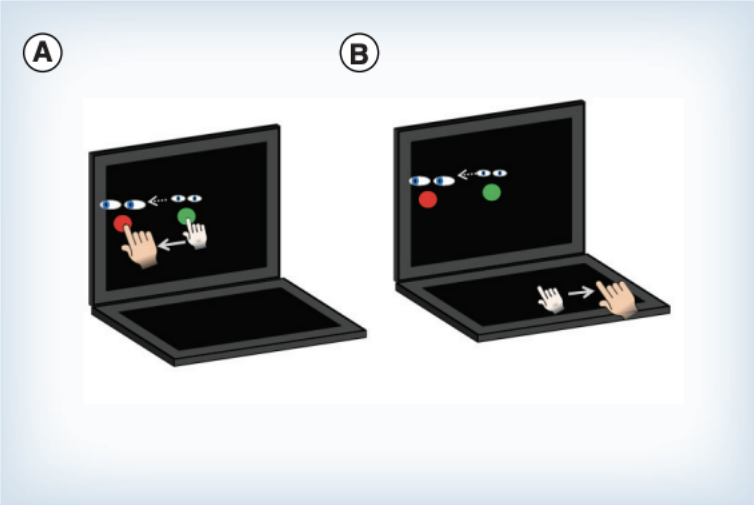

Fig. 1. Experimental setup used in the BrDi test study by Hurtubise et al. (2016). Reprinted with permission from [11].

B. Data Preparation

The original dataset is 150 rows each with 57 features. Features listed include identification and personal data such as initials, id, age, date of birth, height, weight, concussion history (Yes / No), and number of prior concussions. Lastly, 41 different metrics were collected during the neuroscience test. Many columns contained missing values and while the usual practice would be to impute missing, we reason this is not appropriate for our data as it is collected from humans. Thus, any metric and rows with missing values were dropped. Lastly, nonpredictive columns such as a player's initials, position, and shooting dexterity were removed leaving a dataset containing 117 rows and 39 features. We framed the prediction task as binary classification and built a machine learning pipeline incorporating:

- Preprocessing: Applied common feature scaling (Standard, MinMax, Robust) and feature selection using linear SVM and Extra Trees importance scores.
- Class Balancing: Due to moderate imbalance (60% played, 40% did not) within a small dataset, we used class weighting instead of oversampling to avoid synthetic data generation or further reduction via undersampling. Class weighting adjusts model training by penalizing misclassification of the minority class without altering the dataset.
- Training and Tuning: Hyperparameter search is conducted on 80% of the data using 10-fold cross-validation; the remaining 20% is held out for final validation. All splits are stratified to preserve class balance. After comparing multiple train/validation ratios (70/30, 80/20, 90/10), we find that 80/20 offers the best trade-off between training size and validation reliability. Hyperparameters were optimized via Random Search and Tree-Structured Parzen Estimator (TPE). Each setup was repeated 5 times using stratified 80/20 train/validation splits with 10-fold CV.

TABLE I. Average Generalization Results from XAI and Black-Box Models over 5 trials of tuning using Random Search.

	Train			Generalize		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Model						
Weighted-DecisionTreeClassifier	0.725	0.716	0.713	0.611	0.608	**0.603**
Weighted-LinearTreeClassifier	0.645	0.637	0.639	0.609	0.592	0.593
Weighted-LinearBoostClassifier	0.727	0.718	0.720	0.550	0.533	0.533
Weighted-NeuralNetClassifier	0.621	0.622	0.572	0.549	0.575	0.526
Weighted-SVC	0.669	0.658	0.656	0.530	0.525	0.522
Weighted-LogisticRegression	0.636	0.611	0.613	0.535	0.517	0.514
Weighted-RandomForestClassifier	0.740	0.731	**0.732**	0.519	0.517	0.513
Weighted-ElasticNet	0.644	0.624	0.626	0.534	0.508	0.501
Weighted-LGBMClassifier	0.732	0.714	0.716	0.491	0.475	0.473
Weighted-XGBClassifier	0.350	0.591	0.440	0.340	0.583	0.430
Mean	**0.649**	**0.662**	**0.643**	**0.527**	**0.543**	**0.521**

TABLE II. Average ensemble model results over 5 trials of tuning using Random Search

	Train			Generalize		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Model						
Weighted-BaggedLogisticRegression	0.593	0.617	0.587	0.578	0.600	**0.565**
Weighted-BaggedElasticNet	0.657	0.615	0.610	0.609	0.558	0.550
Weighted-AdaBoostLogisticRegression	0.573	0.630	0.577	0.513	0.567	0.517
Weighted-AdaBoostElasticNet	0.634	0.606	0.608	0.540	0.508	0.504
Weighted-BaggedDecisionTree	0.715	0.710	0.704	0.497	0.500	0.486
Weighted-BaggedSVC	0.659	0.641	0.642	0.479	0.467	0.467
Weighted-AdaBoostSVC	0.350	0.591	0.440	0.340	0.583	0.430
Weighted-AdaBoostDecisionTree	0.913	0.895	**0.895**	0.440	0.425	0.423
Mean	**0.637**	**0.663**	**0.633**	**0.500**	**0.526**	**0.493**

III. Results

As this is preliminary research using BrDi data, we evaluated a broad set of ML models. We implemented hyperparameter tuning for six explainable models—Decision Tree, Random Forest, Linear Tree, Logistic Regression, Elastic Net, and Support Vector Machine (SVC)—and four black-box models—Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), Linear Boosting Tree, and a Neural Network. In addition, we assessed ensemble methods using bagging and AdaBoost. Model performance was assessed using precision, recall, and weighted F1 score—metrics suited for imbalanced data.

We report average generalization performance across five trials using Random Search in Table I containing results from black-box models and XAI models. The XAI models such as Weighted Decision Tree classifier, Weighted Linear Tree, Weighted Support Vector Classifier (SVC), Weighted Logistic Regression, Weighted Random Forest and Weighted Elastic Net classifiers have outperformed the XAI models. Across most models, a clear gap is observed between training and validation scores, indicating overfitting. This is likely due to high feature dimensionality, limited sample size, and inherent dataset noise. Ideally, high-performing models would show consistent accuracy across both training and validation sets, with predictions aligning along the diagonal of the confusion matrix. Among all models, the Weighted Decision Tree achieved the highest average generalization F1 score (0.603), suggesting its potential to detect meaningful patterns in BrDi-based cognitive-motor data relevant to NHL participation. While this finding is promising, we interpret the overall results as preliminary due to signs of overtraining. To mitigate variability and improve reliability, all experiments were repeated using five random seeds with stratified train/validation/test splits. This approach maximizes data utilization while supporting consistent evaluation.

To further enhance generalization performance, we evaluated ensemble approaches in Table II using bagging and boosting strategies. Specifically, we applied ensemble learning to weak classifiers (e.g., Decision Trees and Logistic Regression) using techniques such as Random Forests and AdaBoost. While ensemble models typically improve robustness and reduce variance, results in our case were mixed. As shown in Table II, ensemble methods like Random Forest and LGBM achieved high training F1 scores but often failed to generalize well, likely due to overfitting on the limited dataset. In contrast, the weighted Decision Tree, while simpler, achieved the most consistent balance between training and generalization performance (**F1 = 0.627**), suggesting that increased model complexity does not necessarily yield better predictive outcomes under small-sample constraints. These findings reinforce the importance of model simplicity and interpretability in small-data regimes and suggest that, despite their theoretical advantages, ensemble methods may not offer practical gains in early-stage, data-constrained predictive modeling for concussion-related NHL outcomes.

Confusion matrices provide insight into classification errors and the distribution of predictions across actual classes. Fig. 2 illustrates the performance of the Weighted Decision Tree classifier on the best single trial in which the classifier achieved strong generalization, correctly classifying 75% of instances with minimal false positives and false negatives. In contrast, Fig. 3 represents average results over five random splits and reveals a drop in performance, with more misclassifications occurring in both positive and negative classes. The classifier achieved F1 score ≈ 0.675. From Fig. 3, we observe that while the classifier identifies a substantial number of true positives (49), it also produces a relatively high number of false positives (26), lowering the precision for the NHL class. This indicates the model sometimes incorrectly predicts NHL participation for players who did not reach that level. This means that on average, the classifier maintains moderate recall—identifying most NHL players—but at the cost of reduced precision, as it more frequently misclassifies non-NHL

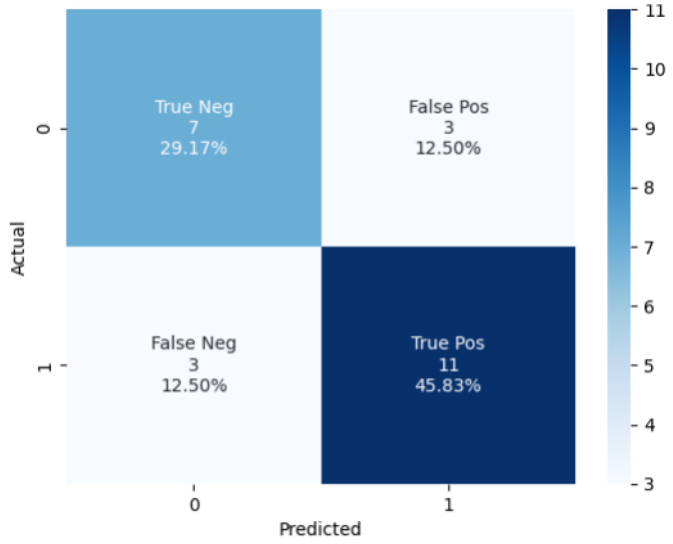

Fig. 2. Confusion matrix for Weighted Decision Tree which produced the best generalization score for a *single trial*.

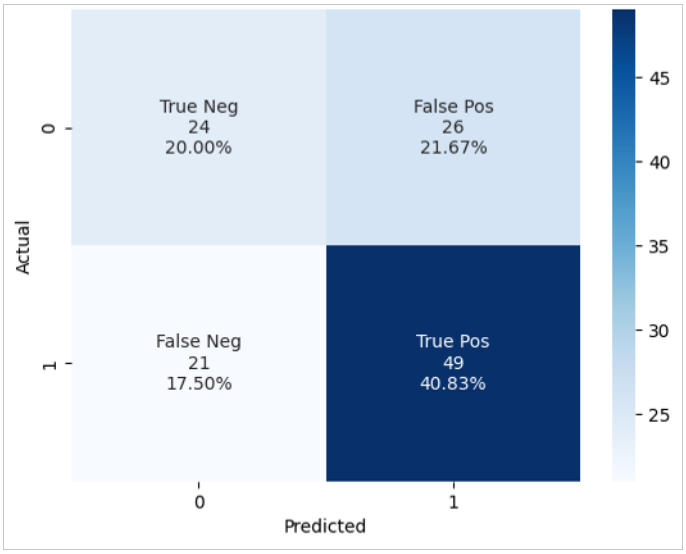

Fig. 3. Confusion matrix for the Weighted Decision Tree, the best XAI model over 5 trials.

players as NHL players. This contrast highlights the variability in performance due to data splits and the importance of averaging across multiple trials. While some individual runs yield near-optimal results, reporting average performance offers a more conservative and reliable estimate of model generalizability. Conversely, the number of false negatives (21) affects the recall, meaning the model misses several actual NHL players. In the best-performing single trial (Fig. 2), both false positives and false negatives are significantly reduced, leading to a higher F1 score ≈ 0.786. This trial achieves balanced and high precision and recall, meaning it correctly identifies most NHL players while keeping false alarms low. Overall, these matrices reinforce the trade-off between sensitivity and specificity and highlight the importance of evaluating classifier performance across multiple data splits.

A portion of the trained weighted decision tree is visualized in Fig.4, highlighting its interpretability. Fig. 4 shows the features used in making classifications and the order of importance in which they were chosen. In a situation where two models have similar accuracy, we prioritize using the explainable model as it gives a way to reason about the decision. This builds trust in model results through transparency in decisions, aids in future decisions, mitigate potential biases, and aids developers and users of the model in future work. For example, we see in Fig. 4 that the feature *Delta OffAxis* is more important than *DR Errors HR* as the tree splits on the former feature first. As the top-performing model - the Weighted Decision Tree is also interpretable, we recommend it for this task. However, in cases where black-box models outperform, practitioners must weigh performance against interpretability, depending on the value placed on transparency, trust, and fairness. We report the average across five runs to ensure robustness.

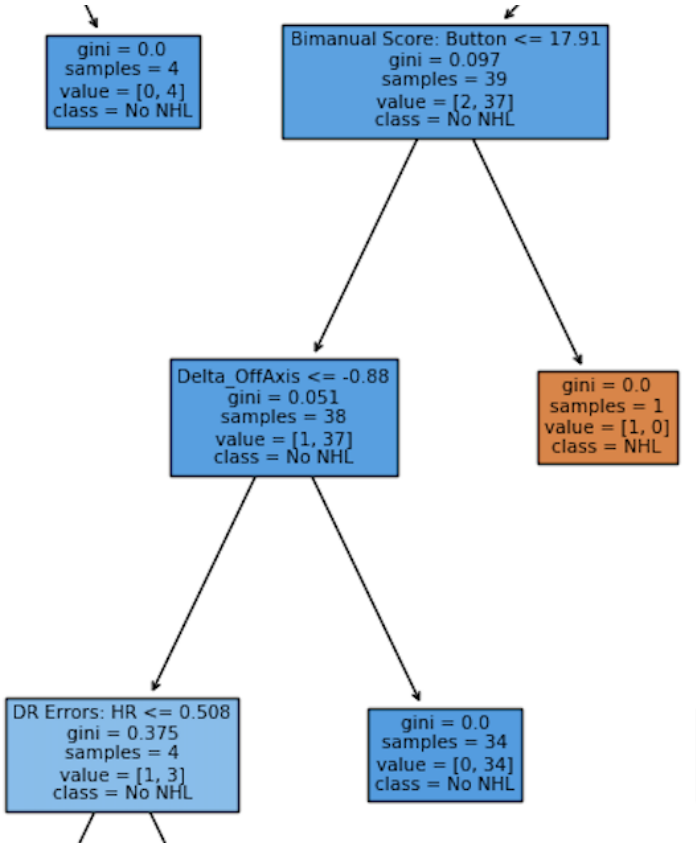

Fig. 4. A zoomed in section of a larger representation of the proposed trained Weighted Decision Tree model.

Hypothesis Result: We evaluated our hypothesis by analyzing model performance in predicting NHL participation based on BrDi-derived cognitive-motor metrics. Across five randomized trials, the weighted Decision Tree model achieved the highest generalization F1 score (0.603), indicating a moderate ability to distinguish between athletes who did and did not reach the NHL. This suggests that CMI impairments post-concussion—quantified via BrDi performance—carry predictive value in determining elite-level outcomes.

While results are preliminary due to dataset limitations, the observed trends provide supportive evidence for our hypothesis: athletes with lower CMI performance were, on average, less likely to appear in the NHL. Notably, explainable models performed comparably or better than black-box models, reinforcing the utility of interpretable cognitive features in predictive assessments.

IV. Limitations: Threats to Validity

This section describes threats to the validity of this study including dataset size, training hardware configurations, and noise. These are explained in detail as follows:

A. Dataset Size and Overfitting

A primary limitation is the small dataset of 93 samples, which restricts model capacity to learn complex relationships between visuomotor features and NHL outcomes. While regularized models such as Logistic Regression (with L1, L2, and Elastic Net penalties) and Support Vector Machines (SVM) are generally robust on small datasets, overfitting remains a concern. We mitigate this by using regularization techniques to penalize complexity and improve generalization. We also tested multiple SVM kernels (linear, radial basis function, sigmoid) and regression variants to control for feature sparsity and correlation. However, as shown in Table I many models showed stronger performance on training data than on the independent holdout set, indicating overfitting. While this reduces confidence in the generalizability of results, it is a known ML challenge in small-sample studies. Nonetheless, our best-performing models still exceeded the NHL participation baseline of 49%, suggesting potential signal in BrDi features despite the limited data. Ideally, model performance on training and validation sets would be within a narrow margin ($\pm 5\%$), supporting greater confidence in real-world applicability. Future studies will benefit from larger datasets to validate the predictive power of cognitive-motor metrics.

B. Dataset Noise

Player outcomes are influenced by numerous factors beyond cognitive-motor performance, introducing noise into the dataset and complicating model training. Players may fail to reach the NHL due to non-concussive injuries, underdevelopment, competitive depth, or external issues such as contract disputes. For instance, although not included in our dataset, Jesse Niinimäki—drafted 15th overall—never played in the NHL despite early promise, illustrating the unpredictability of player trajectories. This highlights the inherent difficulty in modeling success using limited variables. To reduce this noise in future work, additional features such as injury history, performance statistics, and scouting evaluations should be incorporated to provide a more holistic view of each athlete's potential.

C. Neurocognitive Efficiency

Although this work centers on XAI, insights from neuroscience offer context for the challenge of building accurate classifiers. Neurocognitive efficiency—defined as the ratio of motor output to neuromotor effort—reflects how effectively the brain translates cognitive processes into physical actions [12]. Studies across sports such as golf, baseball, and volleyball have shown that expert athletes often exhibit superior performance with reduced cortical activation, suggesting more efficient visuomotor processing [13]–[16]. This phenomenon implies that even athletes with concussion-related cognitive deficits may compensate through higher neurocognitive efficiency. Such compensatory mechanisms likely introduce additional variability in our visuomotor data, contributing to dataset noise and complicating classifier accuracy.

D. Ethical Considerations

Although not a direct limitation, ethical use of AI in sports warrants attention. Predictive models, especially those involving health-related data must uphold principles of fairness, transparency, and accountability. We emphasize that any deployment of our models should align with human-centered values—namely, nonmaleficence (do no harm), beneficence (act in the athlete's best interest), fairness (avoid bias), and transparency. These principles ensure AI supports, rather than undermines, the integrity and equity of sports.

V. Conclusion and Future Work

This work investigates whether post-concussion cognitive-motor integration (CMI), measured via BrDi Test performance, can predict NHL participation using explainable machine learning models. By framing the task as a binary classification problem and evaluating a variety of models, we demonstrate that cognitive-motor metrics derived from neuroscience-informed tasks can carry predictive value in elite sports analytics. The findings provide preliminary but supportive evidence for our hypothesis: that athletes with greater post-concussion CMI impairments, as measured by BrDi performance, are less likely to play in the NHL. The best-performing model—a weighted Decision Tree—achieved a generalization F1 score of 0.603, indicating it could meaningfully differentiate athletes based on cognitive-motor function. While constrained by sample size and overfitting risk, these results suggest that cognitive metrics may offer predictive insight into elite athletic outcomes post-injury.

Our study also highlights the importance of explainability in sports AI applications. Interpretable models offer transparency

and trust—critical for practical use in scouting, return-to-play decisions, and athlete development. Future work should explore the integration of additional features such as historical performance, draft rank, and injury history, as well as regression tasks to predict draft position, games played, or concussion severity. With larger datasets and expanded features, this line of research has the potential to meaningfully augment talent assessment and health management in elite athletics. Future work should address overfitting by incorporating additional features known to influence NHL outcomes—such as draft rank, scoring rate, and nationality—and by expanding the dataset. We also recommend exploring regression tasks, such as predicting draft position, number of games played, or concussion count, which may provide more granular insights into athlete trajectories. This work represents a step toward integrating neuroscience-based performance metrics with predictive analytics in elite sports. With continued data collection and model refinement, such approaches may offer valuable tools for athlete evaluation, injury risk assessment, and long-term performance forecasting.

REFERENCES

[1] M. Schuckers, "Draft by numbers: using data and analytics to improve national hockey league (nhl) player selection," in *MIT Sloan Sports Analytics Conference*, 2016, pp. 1–8.

[2] P. McCrory, W. Meeuwisse, J. Dvorak, M. Aubry, J. Bailes, S. Broglio, R. C. Cantu, D. Cassidy, R. J. Echemendia, R. J. Castellani *et al.*, "Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in berlin, october 2016," *British journal of sports medicine*, vol. 51, no. 11, pp. 838–847, 2017.

[3] M. Chaudhary, M. S. Adams, S. Mukhopadhyay, M. Litoiu, and L. E. Sergio, "Sabotage Detection Using DL Models on EEG Data from a Cognitive-Motor Integration Task," *Frontiers in Human Neuroscience*, vol. 15, p. 662875, 2021.

[4] J. K. Register-Mihalik, A. C. Littleton, and K. M. Guskiewicz, "Are divided attention tasks useful in the assessment and management of sport-related concussion?" *Neuropsychology review*, vol. 23, pp. 300–313, 2013.

[5] G. Manley, A. J. Gardner, K. J. Schneider, K. M. Guskiewicz, J. Bailes, R. C. Cantu, R. J. Castellani, M. Turner, B. D. Jordan, C. Randolph, and et al., "A systematic review of potential long-term effects of sport-related concussion," *British Journal of Sports Medicine*, vol. 51, no. 12, p. 969–977, 2017.

[6] J. A. Brown, M. Dalecki, C. Hughes, A. K. Macpherson, and L. E. Sergio, "Cognitive-motor integration deficits in young adult athletes following concussion," *BMC sports science, medicine and rehabilitation*, vol. 7, no. 1, pp. 1–12, 2015.

[7] M. Dalecki, D. Albines, A. Macpherson, and L. E. Sergio, "Prolonged cognitive–motor impairments in children and adolescents with a history of concussion," *Concussion*, vol. 1, no. 3, p. CNC14, 2016.

[8] Y. Liu, O. Schulte, and C. Li, "Model trees for identifying exceptional players in the nhl and nba drafts," *Machine Learning and Data Mining for Sports Analytics*, p. 93–105, Apr 2019.

[9] J. N. Cohen, K. M. Thompson, V. K. Jamnik, N. Gledhill, and J. F. Burr, "Relationship of fitness combine results and national hockey league performance: A 25-year analysis," *International Journal of Sports Physiology and Performance*, vol. 17, no. 6, pp. 908–916, 2022.

[10] L. Farah, "Evaluating the efficacy of talent identification and development in the national hockey league entry draft," 2022.

[11] J. Hurtubise, D. Gorbet, Y. Hamandi, A. Macpherson, and L. Sergio, "The effect of concussion history on cognitive-motor integration in elite hockey players," *Concussion*, vol. 1, no. 3, p. CNC17, 2016.

[12] B. D. Hatfield, K. J. Jaquess, L. Lo, and H. Oh, "The cognitive and affective neuroscience of superior athletic performance," *Handbook of Sport Psychology*, p. 487–512, 2020.

[13] G. Gallicchio, A. Cooke, and C. Ring, "Practice makes efficient: Cortical alpha oscillations are associated with improved golf putting performance." *Sport, exercise, and performance psychology*, vol. 6, no. 1, p. 89, 2017.

[14] B. Tomasino, M. Maieron, E. Guatto, F. Fabbro, and R. I. Rumiati, "How are the motor system activity and functional connectivity between the cognitive and sensorimotor systems modulated by athletic expertise?" *Brain research*, vol. 1540, pp. 21–41, 2013.

[15] H. Nakamoto and S. Mori, "Experts in fast-ball sports reduce anticipation timing cost by developing inhibitory control," *Brain and cognition*, vol. 80, no. 1, pp. 23–32, 2012.

[16] C. Del Percio, C. Babiloni, M. Bertollo, N. Marzano, M. Iacoboni, F. Infarinato, R. Lizio, M. Stocchi, C. Robazza, G. Cibelli *et al.*, "Visuo-attentional and sensorimotor alpha rhythms are related to visuo-motor performance in athletes," *Human brain mapping*, vol. 30, no. 11, pp. 3527–3540, 2009.