Developing a Novel Machine Learning Framework for Season-Ahead Prediction of Injured Reserve Placement in Professional Football Wide Receivers and Tight Ends

Anonymous Author(s)
Affiliation
Address
email

Abstract

Severe injuries that force National Football League (NFL) athletes onto injured reserve (IR) can reshape team strategy and player careers, yet no literature exists for next-season forecasting of these high-impact events. We construct a position-specific, season-ahead IR risk model for wide receivers and tight ends by merging six years (2018-2023) of public tracking, workload, demographic, travel, surface, combine, and injury-history data into 746 player-season observations. We evaluate logistic regression, random forest, multilayer perceptron, and Extreme Gradient Boosting (XGBoost) pipelines using balanced accuracy. XGBoost attains the best score (58.3%), outperforming neural and linear baselines. SHapley Additive exPlanations reveal 40-yard dash time, pre-snap cushion, and total snaps as the most influential predictors, while static anthropometrics and prior IR counts contribute marginally. The framework delivers the first interpretable, position-aware IR forecasting tool and lays a strong foundation for future efforts.

Sports analytics, Machine learning, Injury prediction, NFL, Feature importance, Athlete health

1 Introduction

1.1 Background and Motivation

Injuries are an unavoidable component of professional sports and carry enormous competitive and financial consequences for athletes and teams. In the National Football League (NFL), movements of athletes to the injured reserve (IR) are typically used in the case of a severe injury in which the athlete is required to miss numerous weeks. IR designations, therefore, provide a convenient proxy for high-impact injuries that require long recovery windows and often influence roster construction, salary-cap management, and player health trajectories. Despite incremental advances in sports medicine and strength-and-conditioning practices, the overall incidence of impactful injuries in the NFL has remained persistently high, even post-COVID (10). Publicly available injury reports show that wide receivers (WRs) and tight ends (TEs) consistently rank among the most frequently injured positions. In one recent positional analysis covering the opening four weeks of the 2016–2020 seasons, wide receivers sustained 432 reportable injuries and tight ends 226, together accounting for roughly one quarter of all injuries in that time span (2). These "receiver" positions are exposed to high-velocity sprinting and abrupt cutting while often incurring mid-air or open-field collisions, biomechanics that elevate the risk of lower-extremity trauma, concussions, and soft-tissue strains (7; 11). Players placed on IR can miss the remainder of the season, lose future earnings, and face shortened careers, while teams absorb performance and monetary lossesc(13).

Researchers have tried to quantify and identify critical predictors of injury risk, as understanding key drivers of athlete injuries may enable the proactive use of preventative treatment strategies. Although traditional epidemiological studies have identified specific correlates of injury, their scope is often limited to small sample sizes, linear assumptions, and variability amongst different studies (6). Modern machine learning (ML) methods can overcome these constraints by ingesting high-dimensional data from a myriad of sources, consequently learning nonlinear interactions that may precede an injury (9).

This study aims to develop and introduce an end-to-end ML pipeline that predicts next-season injured reserve occurrence in NFL receivers using carefully curated player-season data, produceing a practically useful and scientifically rigorous decision-support tool while advancing methodological transparency in sports injury analytics.

1.2 Key Contributions

This work expands upon previous efforts by introducing the first study to design and deploy a machine learning-driven prediction model for injured reserve risk in NFL athletes. Furthermore, by limiting analysis to receivers, this study enables the discovery and incorporation of position-specific risk patterns and signals. This study also employs a wide and diverse range of feature types and adds an emphasis on interpretability within the statistical and ML analyses, an underexplored sector in the field.

1.3 Paper Organization

The remainder of this paper includes: (1) a comprehensive review of related work in sports injury prediction, with an emphasis on statistical and machine learning applications to NFL injury risk; (2) the dataset construction process, including demographic, workload, kinematic, and injury history sources, as well as procedures for data cleaning, record linkage, and feature engineering; (3) the predictive modeling pipeline, detailing preprocessing, model selection, evaluation metrics, and interpretability methods; (4) model training and evaluation results, including performance comparisons and key feature importances; (5) a discussion of the findings in the context of injury prevention and decision-making in professional football, along with the study's limitations; and (6) a summary of contributions and directions for future research.

2 Related Works

Only a handful of recent studies have begun to leverage machine learning on NFL athletes. Angileri et al. (2023) examined five seasons of NFL schedule and travel data and found that cumulative round-trip mileage was not a significant predictor of overall injury counts (1). Furthermore, concussion-specific deep learning was among the first supervised approaches applied to reconstructed NFL injury cases. Cai et al. (2017) trained both support-vector machines and random forests alongside convolutional neural networks on whole-brain white-matter fiber-strain features to classify concussions (4). Despite these advances, inconsistent documentation of methods and the predominant focus on next-game or weekly injury events limit comparability across studies and practical utility for season-ahead forecasting.

Despite increasing interest in injury prediction in professional sports, no existing study models next-season injured reserve placement, particularly using public data or focusing on severe outcomes. Prior NFL research largely targets short-term injury events, relies on proprietary biomechanical inputs, and treats injury as a general binary rather than forecasting meaningful, high-impact designations like IR. Furthermore, no prior work builds position-specific models for wide receivers and tight ends, nor integrates the diverse feature types (demographics, injury history, workload, travel, Combine metrics, and surface conditions) into a unified predictive framework. Finally, most efforts treat model interpretability as secondary; to our knowledge, no study has combined ensemble learning with SHAP-based explanations in this domain. This work, therefore, fills a clear and critical gap by delivering a reproducible, long-range, position-specific, and interpretable public-data model for NFL IR risk.

3 Dataset

3.1 Cohort Definition

The unit of analysis for this study is the player-season, defined as a single NFL athlete's participation in a given regular season as either a wide receiver (WR) or tight end (TE). Each row in the base dataset corresponds to a unique (player, season, team) tuple, with workload features aggregated across the entire regular season. We computed a correlation analysis of the core demographic features (age, height, weight, and 40-yard dash time), The strongest linear association is between height and weight (r = 0.83), followed by those between weight and 40-yard dash (0.55) as well as height and 40-yard dash (0.49), informing decisions about imputation strategies in Section 3.5. Players are included if they appear in the Next Gen Stats receiving dataset for that season. This, therefore, results in the implementation of Next Gen Stats' minimum target requirement (43 from 2018-2020, 45 from 2021 onward). Players with mid-season team changes (e.g., via trade or release) have their statistics merged under one combined record and the team they began the season with. The final cohort includes all player-seasons from 2018 through 2023 meeting these criteria, yielding a sample of 746 WR/TE player-seasons.

3.2 Feature Definitions

The following features were used in the injury risk prediction model: **targets** (number of times the player was targeted with a pass), **yards** (total receiving yards), **avgCushion** (average pre-snap distance between the player and nearest defender), **avgIntendedAirYards** (mean depth of intended targets), **avgSeparation** (average distance from defender at catch or incompletion), **avgYAC** (average yards after catch), **percentShareOfIntendedAirYards** (player's share of team air yards), **turf_percentage** (proportion of games played on artificial turf), **total_round_trip_miles** (total estimated travel mileage), **total_snaps** (count of offensive snaps), **avg_opponent_pass_epa** (mean EPA allowed by opposing defenses), **sos_win_pct** (opponent win percentage as strength of schedule), **age_season_start** (player's age on September 1 of the season), **height** (player height in inches), **weight** (player weight in pounds), **forty** (40-yard dash time in seconds), and **past_ir_count** (number of prior seasons on injured reserve).

Information on data sourcing for each feature is provided in Appendix **??**.

3.3 Label Definition

The binary outcome variable, ir_next_season, serves as the target label for predictive modeling. It is defined as follows: for a given player-season observation in year n, ir_next_season equals 1 if that player is placed on injured reserve (IR) at any time between March 14 of year n + 1 and March 13 of year n + 2. Otherwise, it is assigned a value of 0. Of the n = 746 examples, 152 were labeled with *ir_next_season*=1, indicating that approximately 20.4% of player-season observations resulted in an injured reserve placement during the subsequent league year.

This March 14 cutoff aligns with the typical official start of the NFL league year, ensuring that transactions are cleanly attributed to the upcoming season and avoiding ambiguity surrounding late-season injuries or postseason roster changes.

4 Methodology

To develop a robust framework for predicting next-season injured reserve placement among NFL players, we implemented a comprehensive methodology encompassing data preprocessing, statistical analysis, model training, hyperparameter tuning, and interpretability.

4.1 Data Processing

Continuous features destined for logistic regression and multilayer perceptron are standardized to zero mean and unit variance. Tree models operate on raw values.

During training, we apply Synthetic Minority Over-sampling Technique (SMOTE), a common strategy to reduce imbalance by artificially generating additional training examples within the minority class, inside each cross-validation fold for the multilayer perceptron and logistic regression models, and adjust class-weight where supported (Random Forest) or scale_pos_weight in XGBoost.

The final dataset contained no missing values for any feature except for the 40-yard dash time, a Combine metric that is frequently unavailable for undrafted players or those who did not participate in pre-draft testing. Out of 746 total player-season records, 25 were missing a 40-yard dash entry (3.35%). Missing values were imputed using K-nearest neighbors (KNN) imputation, with height and weight used as the predictor features. This approach preserves underlying physiological relationships and enables the retention of all samples for modeling.

4.2 Modeling Approaches

4.2.1 Logistic Regression

Logistic regression is a linear classification model that estimates the probability of a binary outcome by applying the sigmoid function to a weighted sum of input features. Its simplicity and convex optimization landscape make it an effective and interpretable baseline in many binary classification tasks.

4.2.2 Tree-Based Models

Tree-based models operate by recursively splitting the dataset on feature thresholds to form decision paths, ultimately leading to leaf nodes that represent class predictions. We adopt two ensemble methods, which combine multiple decision trees to form a more robust and accurate model.

- **Extreme Gradient Boosting (XGBoost)** is an efficient implementation of gradient-boosted decision trees. It builds models sequentially, where each new tree attempts to correct the errors made by previous ones.
- **Random Forest** constructs an ensemble of decision trees in parallel, using bootstrap samples of the data and random feature subsets at each split. Each tree votes independently, and the final prediction is made by majority vote.

4.2.3 Multilayer Perceptron (MLP)

Neural networks consist of layers of interconnected nodes (neurons) that learn hierarchical representations of input data by optimizing weights through backpropagation. These models excel at capturing complex, non-linear relationships, making them a strong candidate for structured prediction tasks. In this study, we employed a fully connected multilayer perceptron (MLP), a feedforward architecture composed of an input layer, one hidden layer with ReLU activations, and a final sigmoid output layer.

4.3 Hyperparameter Optimization

We employ an exhaustive grid search with stratified five-fold cross-validation, selecting the best hyperparameters based on mean F1 score across folds. Specific hyperparameters and tuning grids can be found in Appendix **??**.

4.4 Model Evaluation and Interpretability

Model performance is assessed on the test set with the *balanced accuracy* score, a metric that reflects class-aware model performance on imbalanced datasets. SHapley Additive exPlanations (SHAP) are computed on the test set for the top-performing model, enhancing model transparency by highlighting which features most consistently influence predictions across the entire test set.

4.5 Statistical Significance Testing

Prior to modeling, we test each continuous feature for normality with the Shapiro-Wilk test. Normality was established for avgCushion and failed to be established for all other features. Due to widespread non-normality, the Mann-Whitney U test was used to assess statistical significance between the features and the binary target variable. Significant associations were identified between next-season injury and both average separation and average opponent pass EPA. Complete results of the normality and significance testing, including calculated p-values for each feature evaluated at the $\alpha = 0.05$ threshold, are provided in Appendix **??**

5 Results

5.1 Model Performance

Table II summarizes the balanced accuracy of the tested machine learning models. XGBoost achieved the highest performance at 58.3%, followed closely by the multilayer perceptron (57.0%) and random forest (55.9%). Logistic regression lagged behind with a balanced accuracy of 48.1%, reflecting its limited effectiveness on this imbalanced classification task. These results suggest that non-linear models, particularly tree-based and neural approaches, are better suited for capturing the complex patterns associated with IR risk.

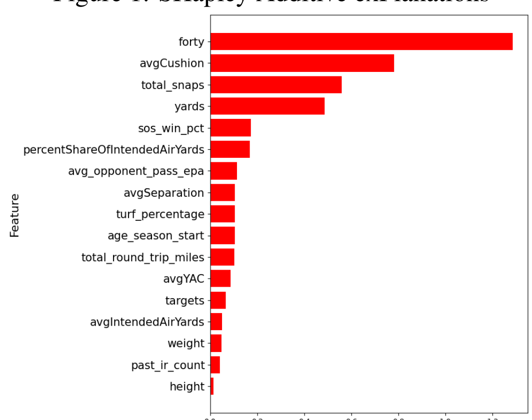
Table 1: Model Types and Their Balanced Accuracies

Model Type	**Balanced Accuracy (%)**
XGBoost	58.3
Multilayer Perceptron (MLP)	57.0
Random Forest (RF)	55.9
Logistic Regression (LR)	48.1

5.2 Feature Importance

Figure 1 displays SHapley Additive exPlanations (SHAP) applied to the top-performing XGBoost model, ranking features by their average contribution to predictions of next-season IR placement. The 40-yard dash time (forty) was the most impactful, suggesting that reduced straight-line speed may correlate with underlying biomechanical or conditioning-related injury risks. avgCushion also ranked highly, potentially reflecting usage patterns or physical playstyles that influence exposure to contact. In contrast, total_snaps and yards likely reflect cumulative workload, consistent with prior findings linking overuse to injury. Static attributes such as height, weight, and past IR count contributed minimally to the model's predictive behavior.

Figure 1: SHapley Additive exPlanations



6 Discussion

6.1 Key Insights

The XGBoost classifier reached a balanced accuracy of 58.3%, modest in absolute terms yet superior to all baselines examined. Feature attribution revealed that straight-line speed (*forty*) and pre-snap cushion (*avgCushion*) carry the greatest marginal contributions, while static anthropometrics, prior IR history, and travel- or surface-based variables offered limited incremental signal.

These results suggest that publicly available tracking metrics can serve as an early warning system for severe, season-disrupting injuries. Declines in straight-line speed or increasing reliance on tight separation may alert performance staff to underlying biomechanical fatigue long before clinical symptoms appear. From a roster-management perspective, probabilistic IR forecasts could inform depth-chart planning and contract structuring, especially for receivers whose market value is closely tied to availability.

Model performance may improve through several extensions: (1) incorporating time-series representations of weekly workload to capture acute and chronic load interactions, (2) expanding labels to include duration on IR or injury type to enable survival analysis, (3) integrating domain-informed engineered features to better capture nonlinearity, and (4) improving model performance and generalizability by expanding data collection.

6.2 Limitations and Modeling Challenges

The study is constrained by the small positive class (20.4% IR incidence) and the limited six-season window, which restrict statistical power and encourages overfitting. IR designations are an imperfect proxy for injury severity; off-season surgeries and voluntary short-term injured lists remain uncaptured. Publicly scraped data introduce measurement error in both workload and injury histories, and fuzzy-matching of names may retain residual linkage noise. Finally, SMOTE oversampling and class-weighting mitigate but do not fully resolve the inherent trade-off between recall and precision, a key consideration when deploying risk scores in a high-stakes sports environment.

7 Conclusion

This study presents the first position-specific model for forecasting season-ahead injured-reserve placements among NFL wide receivers and tight ends. Leveraging six seasons of tracking, workload, and injury-history data, the proposed XGBoost pipeline achieved a balanced accuracy of 58.3%, outperforming neural, random-forest, and logistic baselines. SHAP analyses highlighted straight-line speed, pre-snap cushion, and cumulative workload as primary drivers of risk, offering actionable levers for medical and coaching staffs. Although modest in absolute terms, these results demonstrate that even noisy, open-source signals carry predictive value when synthesized through non-linear learning. The framework lays the foundation for richer time-series features, survival objectives, and biomechanical covariates. Future investigations should test the framework across longer observation windows and incorporate player-worn sensor data to sharpen individual risk trajectories. Ultimately, integrating models of this kind into athlete management workflows could shorten rehabilitation timelines, inform contract strategies, and, most importantly, protect player health.

A Appendix

This appendix provides supplementary methodological details, statistical testing results, data source descriptions, record linkage protocols, and additional exploratory analyses referenced in the main paper. These materials are intended to support reproducibility and transparency while keeping the main body concise.

A.1 A. Data Sources

Player-season records were assembled from the following repositories:

- **Demographics and Snap Counts**: Extracted from the `nfl-data-py` roster files. Missing values were back-filled from ESPN player profiles.
- **Workload and Kinematic Metrics**: Advanced tracking statistics (targets, yards, avgCushion, avgSeparation, etc.) were scraped from the NFL Next Gen Stats (NGS) public portal (2018–present).
- **Injury History**: Collected using a Selenium pipeline on Pro Sports Transactions (PST), restricted to injured reserve (IR) movements (2010–2025). These logs provide transaction dates and descriptive notes.

- **Combine Metrics**: Forty-yard dash times obtained from the `nfl-data-py` Combine file.
- **Travel and Surface**: NFL schedule files from `nfl-data-py` were linked to a static stadium reference table (manual validation of stadium latitude–longitude and surface type). This enabled computation of round-trip travel mileage and turf exposure.
- **Opponent Strength**: Opponent win percentage and average opponent pass EPA calculated from the NFL schedule file.

A.2 B. Record Linkage and Deduplication

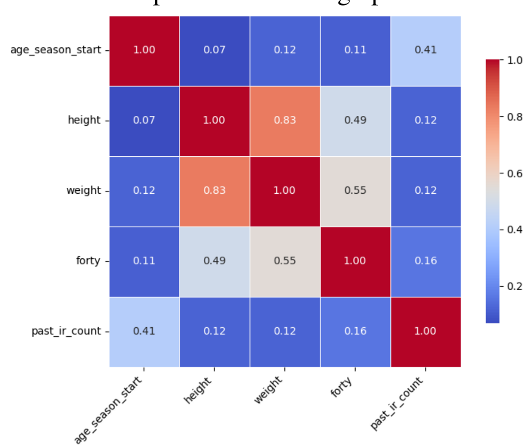
To unify datasets:

- Names were matched using `fuzzywuzzy.token_set_ratio` with a threshold of 95%.
- Ambiguous cases were manually reviewed to resolve false positives and negatives.
- Exact duplicates were removed.
- Multi-row duplicates from mid-season trades were merged by summing snap counts and recalculating averages across the full season.

A.3 C. Correlation Analysis of Demographic Features

Correlation analysis was conducted on the primary demographic variables (age, height, weight, and 40-yard dash). The strongest linear relationship was observed between height and weight ($r = 0.83$), followed by weight and 40-yard dash ($r = 0.55$), and height and 40-yard dash ($r = 0.49$). These relationships guided the KNN imputation strategy for missing 40-yard dash times.

Figure 2: Correlation Heatmap Between Demographic and Performance Features



A.4 D. Statistical Testing Results

We assessed normality of each continuous feature using the Shapiro–Wilk test. Except for `avgCushion`, features rejected the null hypothesis of normality at $\alpha = 0.05$. Consequently, we used the Mann–Whitney U test to compare feature distributions by next-season IR status.

A.5 E. Hyperparameter Search Details

Hyperparameters for each model were tuned via stratified five-fold cross-validation, maximizing mean F1 score. The search grids were:

- **Logistic Regression**: Regularization strength $C \in \{0.01, 0.1, 1, 10\}$, penalty $\in \{\ell_1, \ell_2\}$.
- **Random Forest**: Number of trees $\in \{100, 300, 500\}$, maximum depth $\in \{5, 10, 20, \text{None}\}$, class weights balanced.

Table 2: Shapiro–Wilk normality test p-values for continuous features ($\alpha = 0.05$). Bold indicates failure to reject normality.

Feature	Shapiro–Wilk p
Targets	0.0000
Yards	0.0000
Avg Cushion	**0.7970**
Avg Intended Air Yards	0.0006
Avg Separation	0.0000
Avg YAC	0.0000
% Share of Intended Air Yards	0.0000
Turf Percentage	0.0000
Total Round Trip Miles	0.0000
Total Snaps	0.0000
Avg Opponent Pass EPA	0.0000
Strength of Schedule Win%	0.0049
Age at Season Start	0.0000
Height	0.0000
Weight	0.0000
40-Yard Dash (s)	0.0000
Past IR Count	0.0000

Table 3: Mann–Whitney U test p-values comparing distributions by next-season IR status ($\alpha = 0.05$). Bold indicates statistical significance.

Feature	Mann–Whitney U p
Targets	0.9983
Yards	0.7244
Avg Cushion	0.2255
Avg Intended Air Yards	0.1703
Avg Separation	**0.0294**
Avg YAC	0.3271
% Share of Intended Air Yards	0.2578
Turf Percentage	0.3466
Total Round Trip Miles	0.1806
Total Snaps	0.7086
Avg Opponent Pass EPA	**0.0402**
Strength of Schedule Win%	0.3945
Age at Season Start	0.6283
Height	0.7508
Weight	0.7126
40-Yard Dash (s)	0.2783
Past IR Count	0.1524

- **XGBoost**: Learning rate $\in \{0.01, 0.1, 0.3\}$, maximum depth $\in \{3, 5, 7\}$, subsample $\in \{0.5, 0.8, 1.0\}$, scale_pos_weight $\in \{1, 3, 5\}$.
- **Multilayer Perceptron**: Hidden units $\in \{32, 64, 128\}$, dropout $\in \{0.2, 0.5\}$, learning rate $\in \{0.001, 0.01\}$, batch size $\in \{32, 64\}$.

—

References

[1] Angileri H S, McLoughlin D E, Owen M M, May J M, Terry M A, Tjong V K. Association of injury rates among players in the National Football League with playoff qualification, travel distance, game timing, and the addition of another game: Data from the 2017–2022 seasons. *Orthop J Sports Med*. 2023;11(8):23259671231177633. doi:10.1177/23259671231177633.

[2] Baker H P, Pirkle S, Cahill M, Reddy M, Portney D, Athiviraham A. The injury rate in National Football League players increased following cancellation of preseason games because of COVID-19. *Arthrosc Sports Med Rehabil*. 2021;3(4):e1147–e1154. doi:10.1016/j.asmr.2021.05.002.

[3] Binney Z O, Hammond K E, Klein M, Goodman M, Janssens A C J W. NFL injuries before and after the 2011 Collective Bargaining Agreement (CBA). *arXiv* preprint arXiv:1805.01271. 2018.

[4] Cai Y, Wu S, Zhao W, Li Z, Wu Z, Ji S. Concussion classification via deep learning using whole-brain white-matter fiber strains. *PLOS ONE*. 2018;13(5):e0197992. https://doi.org/10.1371/journal.pone.0197992.

[5] Hershman E B, Anderson R, Bergfeld J A, Bradley J P, Coughlin M J, Johnson R J, et al. An analysis of specific lower-extremity injury rates on grass and FieldTurf playing surfaces in National Football League games: 2000–2009 seasons. *Am J Sports Med*. 2012;40(10):2200–2205. doi:10.1177/0363546512458888.

[6] Maak T G, et al. Sports performance and injury research: Methodologic limitations and recommendations for future improvements. *Arthroscopy*. 36(11):2938–2941.

[7] Manoharan A, Barton D, Khwaja A, Latt L D. Return-to-play rates in NFL wide receivers and running backs after ACL reconstruction: An updated analysis. *Orthop J Sports Med*. 2021;9(1):2325967120974743. doi:10.1177/2325967120974743.

[8] McCormick W F, Lomis M J, Yeager M T, Tsavaris N J, Rogers C D. Field surface type and season-ending lower-extremity injury in NFL players. *Transl Sports Med*. 2024;2024:6832213. doi:10.1155/2024/6832213.

[9] Musat C L, Mereuta C, Nechita A, Tutunaru D, Voipan A E, Voipan D, et al. Diagnostic applications of AI in sports: A comprehensive review of injury risk prediction methods. *Diagnostics*. 2024;14(22):2516. doi:10.3390/diagnostics14222516.

[10] Platt B, Abed V, Khalily C, Sullivan B, Skinner M, Jacobs C, Johnson D, Stone A V. Injury rates remained elevated in the second National Football League season after the onset of the COVID-19 pandemic. *Arthroscopy, Sports Medicine, and Rehabilitation*. 2023;5(2):e325–e329. https://doi.org/10.1016/j.asmr.2022.11.026.

[11] Sanchez E, Weiss L, Williams T, Ward P, Peterson B, Wellman A, Crandall J. Positional movement demands during NFL football games: A three-year review. *Applied Sciences*. 2023;13(16):9278. https://doi.org/10.3390/app13169278.

[12] Sclafani S, Frane N, Humphrey T J, Miceli J, Trasolini R. Increased regular-season soft-tissue injury rates in National Football League players may be associated with the canceled 2020 NFL preseason due to COVID-19. *Cureus*. 2022;14(5):e24674. doi:10.7759/cureus.24674.

[13] Turnbull M R, Gallo T F, Carter H E, Drew M, Toohey L A, Waddington G. Estimating the cost of sports injuries: A scoping review. *J Sci Med Sport*. 2024;27(5):307–313. https://doi.org/10.1016/j.jsams.2024.03.001.

[14] Venishetty N, Xiao A X, Ghanta R, Reddy R, Pandya N K, Feeley B T. Lower-extremity injury rates on artificial turf versus natural grass surfaces in the National Football League during the 2021 and 2022 seasons. *Orthop J Sports Med*. 2024;12(8). doi:10.1177/23259671241265378