iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79713

SportSense: An AI-Powered Platform for Predicting Athlete Performance and Injury Risk

Mohammad Afnaan Anwar[1], Fouzan Taha Hussain[2], Syed Abdul Jaleel[3], Dr. K. Kalyani[4]
[1, 2, 3]*Student, Department of Artificial Intelligence & Data Science, Methodist College of Engineering & Technology, Hyderabad, Telangana, India*
[4]*Assistant Professor, Department of Computer Science Engineering, Methodist College of Engineering & Technology, Hyderabad, Telangana, India*

Abstract: Modern sport athlete management practices have evolved over time to be more quantitative, focused on improving athlete performance while considering their health. However, existing methods for analyzing athletes involve biased subjective criteria, lack of anticipation, and incomprehensible data analysis. The present paper presents SportSense, a simplistic solution that predicts injuries and assesses the preparedness of players. SportSense leverages machine learning algorithms like the Random Forest classifier, which evaluates data on workload, physiological conditions, and fatigue from pre-existing datasets. Training and prediction occur offline, followed by embedding the predictions in the web user interface. More specifically, SportSense comprises a web dashboard using the visualization framework Chart.js for visualizing information on risk profile, trends, and workload assessment. Additionally, a personalizable XAI solution using the manually designed TreeSHAP algorithm is presented to determine the crucial features used for making the prediction. In summary, SportSense improves athlete evaluations by detecting critical workload due to fatigue.

Keywords: Artificial Intelligence, Machine Learning, Sports Analytics, Injury Prediction, Explainable AI, Random Forest, Athlete Monitoring, Data Visualization.

I. INTRODUCTION

Given the high level of risk involved in competitive professional sports, one key element of success is ensuring that athletes remain healthy while still maintaining their performance. In recent years, the volume of data surrounding sports has increased significantly to include aspects such as workload, matches played, and performance data collected through analysis of matches. Nevertheless, the effective utilization of all the data that exists presents the biggest challenge faced by numerous sporting teams.

Contemporary problems related to sport science may be considered the problems related to prediction of injuries without physical contact and translation of findings into understandable data. It is certain that machine learning provides powerful tools for predictive modeling; however, at the same time, there are some problems with explainability and good usability of machine learning models. Also, it is necessary to underline that traditional methods do not provide the opportunity to present data in an understandable way for coaches.

The above problems would be addressed through the use of a unique lightweight machine learning algorithm not only for prediction but also easy interpretation of its results. The algorithm utilizes the player's information that is processed into features such as Workload Index and Fatigue Index.

These features will be used by a Random Forest classifier that predicts future injuries. Machine learning algorithms for feature extraction, model training, and prediction generation are implemented off-line using a pipeline approach. The outputs of these operations are embedded into a web dashboard using JSON representation format.

The other significant consideration in connection with SportSense is that it emphasizes its explainability and usability. While it operates based on a machine learning algorithm, the system is also built on a new idea called XAI, which proves to be critical in identifying the factors behind every prediction generated by the system. In terms of forecasting the possibility of an injury, the factors that could be influential include fatigue and lack of adequate rest. Moreover, all predictions are visually clear and comprehensible.

In general, it can be said that SportSense is a perfect example of making use of the machine learning technique, as well as analytics of structured data, to analyze sports science in an easy way that can easily be replicated, without having any kind of server-side component at all. With its use of both offline prediction and interactive visualization, SportSense is a good instrument for preventing injuries among athletes.

II. LITERATURE SURVEY

The implementation of Machine Learning techniques in sports medicine is increasingly common; however, there are certain areas that need further improvement. As discussed by Chen et al., (2023), the application of SVM and Random Forest was effective in predicting injury among the knee joints using biomechanics; however, it did not scale sufficiently enough to be used effectively in real-time management of teams [1]. In addition, Oliver et al., (2020) noted that although decision trees could be useful in predicting injury for young soccer players, class imbalance made it difficult to produce efficient models [3].

However, even though ACWR is considered to be the gold standard in determining whether there is any increase in workloads, based on the recent meta-analysis carried out by Leckey & van Dyk, the ACWR will not be able to predict the outcomes if other factors are not considered [4].

Regarding technology, there is one new development in the field of data security through Blockchain. This has been proven by Tyagi et al. (2025) and Zamare et al. (2024) through their research, which reveals the capacity of the smart contracts in the Ethereum blockchain to provide important securities for DApps [5][6]. SportSense adopts a similar approach where the advantages of decentralized ledger systems are applied in sports science.

Hegde et al. (2024) created a functional crowdsourcing application using Ethereum platform and supporting MetaMask, thus providing transparency in all transactions. The absence of milestones-based payment releases within their model is one of the factors that make it hard to ensure the efficient use of money. Similarly, Narender et al. (2025) tried to improve security and reduce cost via blockchain implementation, however, disregarding user convenience [7].

Moreover, Kumar et al. (2024) implemented DApp for crowdsourcing, which demonstrated the applicability of Web3 technology to facilitate transaction processes while making them more secure [8]. As for governance mechanisms, modern technology and artificial intelligence were explored. For instance, Temsen et al. (2023) provided community-based verification approaches based on DAOs in order to reduce the possibility of fraud [9]. Similarly, Antad et al. (2024) applied artificial intelligence and smart contracts for fraud detection [10].

In conclusion, the current literature review highlights the capability of Machine Learning and Blockchain in improving the efficiency of the sports monitoring system through increased transparency, security, and improved prediction accuracy. However, some limitations exist within the context of this topic. These include the lack of real-time tracking of funds and performance metrics, control mechanisms based on milestones, and the fact that they are complicated to use and understand. The proposed SportSense framework will solve all of these problems, creating an effective sports monitoring system through Random Forest and Web3 technology.

An easy way to comply with IJRASET paper formatting requirements is to use this document as a template and simply type your text into it.

III.PROPOSED ARCHITECTURE

The Architecture of SportSense is a versatile method of organizing the whole application, whereby all processes that come into play, such as taking in sport data in CSV format, processing, performing machine learning, and visualizing in a browser are done in a systematic manner.

A. Data Preprocessing and Feature Engineering

Ingestion is a combination of game history and training data. Data preprocessing occurs using the programming language Python. The following features have been computed:

- Workload Index: It is a composite measure of time, workload intensity, and distance traveled.
- Fatigue Index: It is computed based on the number of games played in a week compared to the number of rest days.

B. The rationale for constructing the engine of this straightforward model comes from the Random Forest technique.

The selected approach was used for the purpose of our task because it helps create nonlinear models and avoid overfitting. To develop our machine learning model, the following variables are necessary: the weekly working hours, breaks, variability in productivity, and injuries to compute the probability of injuries in percent from 0% to 100%.

The learning phase and prediction will occur offline using the Python pipeline. The output of this operation will be a dataset called 'predictions.csv', which contains the output as well as other computations performed for each player separately.

C. Data Integration and Visualization

Static Integration methodology has been implemented in order to not have to implement a backend system. All of the predictions made by the tool will be stored using the JSON format and will be inserted within the HTML pages acting as a frontend system. This specific step is called “the bake step.”

The front end was developed using HTML, CSS, JavaScript, and Chart.js. Graphs were used to visualize information about graphs, injuries, risks, workload, and performance. Graphs are also used to display the athlete’s profile. All actions undertaken by users, such as sorting, are carried out on the client side.

D. Explainability and Decision Support

The use of the TreeSHAP method has been made in the XAI module in order to interpret how the model makes decisions. Through the use of this method, the model identifies those features that have a major impact on causing injury. The identified features are then presented through graphs and texts.

IV. SYSTEM ARCHITECTURE

However, the design of the AI-based system allows for the creation of a true multi-layered architecture, which would be beneficial in terms of ensuring efficient cooperation between the AI algorithm and the blockchain ledger. Specifically, this will include the design of a pipeline with Python programming language, with the results demonstrated in a static dashboard.

In this context, the Agile methodology is applied to allow iteration while developing the system. Iteration will facilitate development of the system in smaller parts that include data preprocessing, developing a machine learning model, and developing the front-end dashboard.

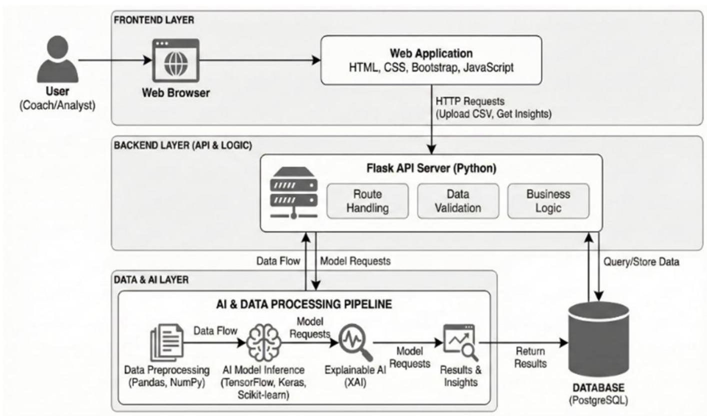

Figure 1. Overall Architecture of the SportSense Platform

SportSense Platform utilizes a scalable and hierarchical architecture framework that provides efficient processing of high-dimensional data pertaining to the athletes, while still maintaining the scalability of the platform through portability and working in an offline mode. Overall architecture of the platform can be separated into three primary layers that comprise of Frontend Layer (Browser Runtime), Offline Data Processing & Machine Learning Pipeline, and Data Integration Layer. By implementing the said architectural framework, SportSense platform is capable of performing heavy computing operations such as training models, predictions and result explanations, while also providing efficient support for frontend layer as well.

The Frontend Layer serves as the interactive layer for coaches and analysts. It has been built using HTML, CSS, and ES6 Javascript, and thus can be used to build an extremely visually-rich interface for tracking squads without making any server-side API calls. While the dynamic interface retrieves data dynamically based on user requests, the frontend layer retrieves data from the statically-loaded JSON objects present within the HTML code itself. The frontend layer is specifically built for processing analytical data and visualizing it using doughnut charts for injury risk assessment, bar charts for workload evaluation, and line graphs for performance prediction. The Data Integration Layer is responsible for connecting the analytics procedure with the user interface layer. In this layer, the output generated by applying machine learning algorithm to the dataset and saved in a file named predictions.csv is transformed to JSON format and added manually to the dashboard HTML files. There is no need for the backend layer and connecting it to any database, which means that there is no use of the “backend,” owing to the presence of the bake function. This makes the application consistent and reproducible, and runs in any modern browser.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

The smart engine for Sportsense runs within the Data & AI layer and uses a Python pipeline off-line. For example, it takes advantage of different data manipulation packages, including Pandas and NumPy to pre-process data and engineer features before feeding into Scikit-learn for prediction. The process entails the conversion of unstructured data about players' performance to structured feature vectors and using them as inputs for training different machine learning models, including Random Forest in predicting injuries and Gradient Boosting predictor for player performance. Furthermore, there is the XAI layer that was manually designed to use the Tree SHAP model to decompose the predictions into feature contributions. It assists in pinpointing the different factors behind each prediction, including players' tiredness and congestion of match calendars.

To summarize, the architecture design philosophy rests on simplicity, modularity, and the client-side implementation of all features. The model calculations' decoupling from visualization at run time and the absence of any backend services allow the solution to be portable and reproducible in an academic environment for evaluation purposes.

V. RESULTS

The evaluation of SportSense technology can be conducted taking into account the level of accuracy of models for the prediction of injuries, the understanding of outcomes from AI results, and the efficiency of the whole procedure of data analysis offline. The use of multidimensional datasets via the Random Forest model allows one to place all athletes into various groups in a systematic way. From the obtained analysis results, it is evident that the SportSense technology acts as an effective medium connecting complicated data with coach-related decision-making processes.

Furthermore, the integration of the explainability feature makes the model more trustworthy because of its capability to offer information on the essential factors considered when predicting. Secondly, with the use of static deployment, there will be consistency in all outputs generated without the involvement of another service. In conclusion, from the performance of the entire system, there is efficient analysis and deployment.

Fig. 2 Player Data

The software interface provides a comprehensive analysis of certain individuals like Ivan Petrov who have a high tendency to get injured. The analysis includes information about previous injuries suffered and how fatigued these players currently are.

Fig. 3 Dashboard

The Squad Overview Dashboard serves as the control tower of the entire team management process, providing coaches with a summary of the current status of their whole squad from a physical fitness perspective.

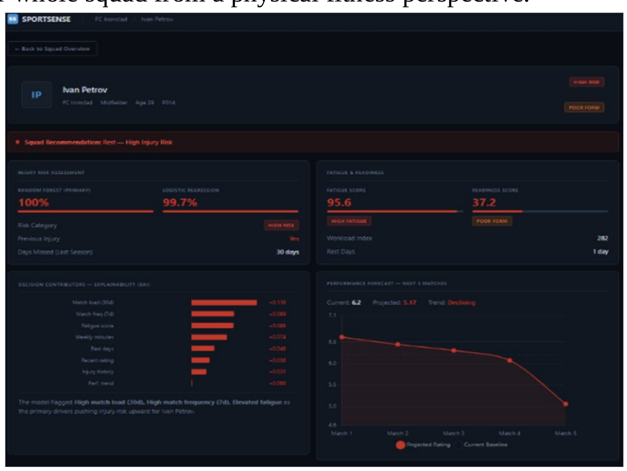

Fig. 4 Player report

Here we can observe where the results of Random Forest and Logistic Regression algorithms converge, where Fatigue and Readiness indicators are represented, and what match loads and frequencies create a high risk situation.

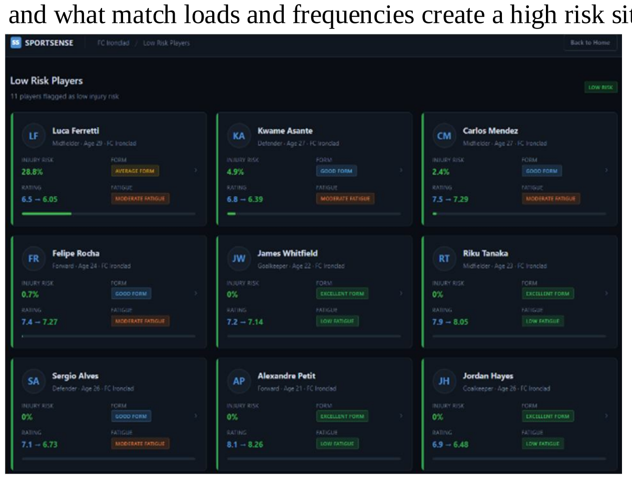

Fig. 5 Low Risk Players

The following figure gives the profile of the athlete that falls under the category of “Low Risk,” which is determined through the use of random forests algorithm. Such type of categorization shows that the athlete has 0% injury risk assessment score, and his/her Readiness Score is very high i.e., 90.0+.

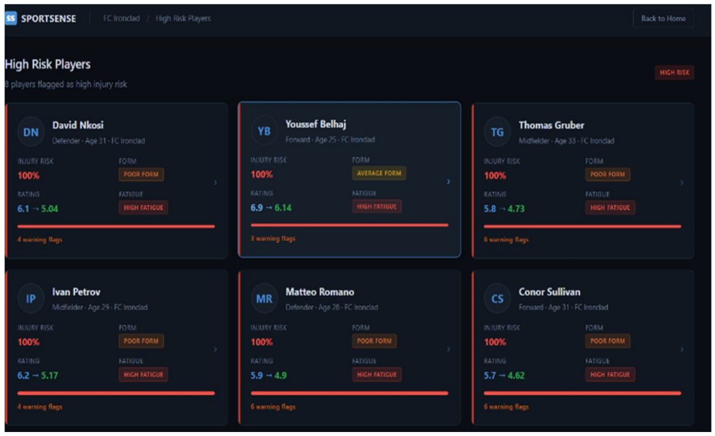

Fig. 6 High Risk Players

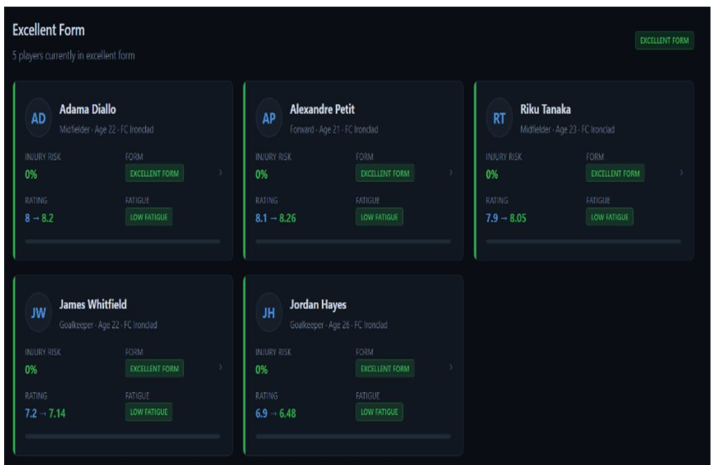

Fig. 7 Excellent form

This is the best form an athlete can ever achieve, and this is evident from the profiles of athletes such as Kai Havertz. The program determines that athletes have attained "Excellent Form" because their readiness levels are highly optimized (for instance, 95.8), while their fatigue levels are highly minimized (for instance, 10.4).

VI. DISCUSSION

From the development and validation of the SportsSense paradigm, it is apparent that there has been a rather drastic paradigm shift in terms of how health and performance are managed within professional athletes. For example, while traditional sports medicine tends to adopt a relatively reactive method for managing injuries in athletes, whereby injuries or any other kind of physical problems are only considered once they begin to manifest themselves, the newly introduced paradigm shifts towards more proactive methods by adopting the use of offline machine learning pipelines that can help predict potential problems using physiological data of the athletes. These problems may include fatigue, workload, and heavy games.

The biggest advantage of this paper is that it has solved the problem known as the "Black Box," where machine learning algorithms cannot be used because of some issues that are faced by the industry. During such conditions, the management requires interpretability rather than probabilistic values. For solving this issue, SportSense has applied an artificial intelligence tool known as the XAI module. An interesting thing about it is that it has used a TreeSHAP algorithm that decomposes all the results in individual features without applying any other library. This enables the determination of the influence of parameters such as fatigue, rest period, or prior injuries on prediction.

In addition, it adopts both probabilistic learning algorithms and deterministic measures such as readiness scores and early warning indicators. This results in an overall view on the well-being of the athlete. There are two stages involved in its analysis, making it more holistic than traditional systems based on thresholding. As opposed to analyzing different variables independently, the model is capable of determining the relationship among different parameters like workload, recovery, performance, and injuries. This implies that two athletes with similar workload requirements can be treated differently based on their recovery and latest performance.

When talking about the usability factor, it becomes apparent that high-level sports analytics is possible to develop in a user-friendly and intuitive way. In comparison to conventional approaches, which make sports analytics dependent on the calculations being done on the server side, the SportSense solution opts for the alternative client-side strategy. The data has already been calculated beforehand and presented in a structured JSON format within static HTML pages. Hence, it becomes apparent that SportSense can work in an offline mode without any integration with the live data through the API. Moreover, the elements of visualizations provided by Chart.js can help users effortlessly switch between teams and players.

In addition, the separation of computation from visualization in terms of design enhances reproducibility and ease of deployment. Offline computation of predictions and their integration into the frontend guarantee that the output will always be reliable without relying on any third-party services at runtime. However, the design flaw with this tool is that, despite being perfect for use in an educational context, it comes with its limitation. This limitation is that the input of real-time data is unnecessary, meaning the tool cannot operate in real time.

Strategically speaking, SportSense could be considered a potential solution for the construction of future systems in sports analytics. Although currently SportSense operates with static datasets, it should be noted without hesitation that it could potentially be scaled up to bigger datasets or could even serve as a basis for creating more complicated systems. The presence of all three processes separately within one pipeline guarantees progress on the part of the system in the future, whether it involves incorporating live streaming data or including explanations through the use of certain libraries.

VII. CONCLUSION

In conclusion, the design and study of the SportSense project have brought to light the enormous transformation that occurs when artificial intelligence is introduced into the sports sciences field. By overcoming the difficulties encountered by the use of conventional methods in evaluating the performance of the athletes, a new technique for predicting injuries and monitoring the athletes' performance levels has been devised. In this particular instance, the major aim was to employ artificial intelligence to create a prediction model using the random forest method.

However, it should be emphasized that the implementation of XAI has been one of the main tools in bridging the gap between the results of machine learning and their application in coaching. This is because XAI allows coaches to learn the reasons for making a prediction, such as extreme fatigue, high loads during matches, or insufficient recovery time. Therefore, it is apparent from the results that coaches are more likely to accept evidence-based decisions when they understand the reasons for making a particular prediction.

At present, SportSense being a static and web-based predictive analysis tool is considered to provide a very stable platform on which the effectiveness of SportSense can be evaluated. It is guaranteed due to the fact that the designed architecture of the tool allows running SportSense offline and visualizing its results in the software. At the same time, such an approach implies inability to adapt and update it in the process of use.

Some of the future areas for research include the implementation of semi-dynamic and real-time systems in this case. The application of the pipeline system, which automatically performs the embedding of the data without the need for manual embedding of the data, is one such example. Other research efforts could also include the application of even more data from other sources. Besides, there could also be an improvement in the efficacy of the models based on real-world data. There could also be improvements in explainability and prediction models.

REFERENCES

[1] Chen, X., Li, Y., & Wang, H. (2023). Healthcare Data-Based Prediction Algorithm for Potential Knee Joint Injury of Football Players. International Journal of Distributed Sensor Networks

[2] C. Barton et al., “Injury Prediction in Competitive Runners With Machine Learning,” Int.

[3] J L. Oliver, F. Ayala, and M. B. A. de Ste Croix, “Using Machine Learning to Improve Injury Risk Prediction in Elite Youth Soccer Players,” J. Sports Sci., 2020.

[4] C. Leckey and N. van Dyk, “Workload Monitoring and Injury Risk in Sports,” Br. J. Sports Med., 2019.

[5] E. Cust et al., “Machine-Learning Models for Injury Prediction in Team Sports,” J. Sports Sci. Med., 2018, doi: 10.1186/s40798-019-0202-3

[6] A. Rossi et al., “Sequence-Based Machine Learning for Injury Prediction,” J. Sci. Med. Sport, 2019.

[7] B. Hulin et al., “Training Load and Injury: Acute :Chronic Workload Ratio,” J. Strength Cond. Res., 2016.

[8] P. Zahariadis et al., “An Overview of Machine Learning Applications in Sports Injury Prediction,” 2023.

[9] Y. Lu, A. Sperry, D. Shapiro, S. Stephan, and T. Pavlovic, “A Deep Learning Approach to Injury Forecasting in NBA Basketball (METIC),” J. Sports Analytics, 2021. DOI: 10.3233/JSA-200529.

[10] S. Jauhiainen et al., “New Machine Learning Approach for Detection of Injury Risk Factors in Young Team Sport Athletes,” 2021.

[11] Y. Lu et al., “Machine Learning for Predicting Lower Extremity Muscle Strain in NBA Athletes,” Orthop. J. Sports Med., 2022.

[12] P. Reche-Soto et al., “Enhancing Sports Injury Risk Assessment in Soccer Using Machine Learning,” 2024..

[13] M. Baladaniya and A. K. Choudhary, “Artificial Intelligence in Sports Science: A Systematic Review and Meta-analysis,” J. Clin. Med. Kazakhstan, vol. 22, no. 3, pp. 64–72, 2025.

[14] C. Vallio et al., “Artificial Intelligence in Sports Biomechanics: A Scoping Review,” Bioengineering, vol. 12, 2025. DOI: 10.1016/j.jbiomech.2020.110106.

[15] P. Plisky et al., “Association of Y Balance Test Reach Asymmetry and Injury Risk,” Clin. Sports Med., 2015.

[16] M. Luu et al., “Machine Learning Outperforms Logistic Regression to Predict Next-Season NHL Player Injury,” Orthop. J. Sports Med., 2020. DOI:10.1177/2325967120953404

[17] K. He and S. Zhang, “Prediction Model of Juvenile Football Players’ Sports Injury Based on Text Mining,” Mobile Information Systems, 2021.

[18] A. Rossi, T. Perri, A. Cicchetti, and M. Marcone, “Predicting Athletic Performance from Physiological Parameters Using Machine Learning: Example of Bocce Ball,” 2019.

[19] J. G. Cladino et al., “Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports,” Sports Med.–Open, 2022. DOI : 10.1186/s40798-019-0202-3

[20] A. Majumdar et al., “Machine Learning for Understanding and Predicting Injuries in Football,” 2022.

[21] M. Choi, K. Lee, and K. Lee, “Injury Prediction in Korean Adult Field Hockey Players Using Machine Learning and SHAP-Based Feature Importance Analysis,” Appl. Sci., vol. 15, p. 8946, Aug. 2025, doi: 10.3390/app15168946