Sports Injury Recovery Period Prediction Based on Injury Type and Diet Plan

Dr. Bhagyashri R Hanji[1], Dhiraj Athreya H[2], Deeksha K B[3], Drushadwathi B Salian[4], Devon Stephen Fernandes[5]

[1,2,3,4,5]*Computer Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bangalore, Karnataka*

Abstract— **Athletic injuries pose a significant challenge, hindering optimal performance and prolonging rehabilitation. To address this issue, we propose a comprehensive approach to predicting injury recovery duration for sports professionals using machine learning techniques. Our primary goal is to develop a predictive model that accurately forecasts recovery periods based on various factors, including injury type, severity, athlete demographics, medical history, and rehabilitation protocols. Utilizing a comprehensive dataset of historical injury records and performance metrics from a diverse range of athletes across various sports, the model aims to capture complex relationships and enhance prediction accuracy.**

Keywords— Scoped Session, Cross-validation, One-Hot Encoding, Support Vector Regressor (SVR), Dynamic HTML Rendering, PostgreSQL.

I. INTRODUCTION

Athletic injuries pose significant challenges, disrupting training, hindering performance, and prolonging rehabilitation. To address this critical issue, we propose a comprehensive approach to predictive modeling for sports-related injuries, leveraging advanced data analytics, machine learning techniques, and web application development. Our system aims to enhance rehabilitation strategies, facilitate timely return to play, and optimize athlete performance.

Effective injury management is crucial for an athlete's long-term success. Conventional methods for estimating recovery periods often lack precision and adaptability, failing to capture the complexities of injury recovery. Our project addresses this gap by developing a sophisticated predictive model capable of accurately forecasting recovery durations for sports-related injuries.

Our system employs the Support Vector Regressor (SVR) machine learning algorithm to predict recovery periods based on a comprehensive set of input variables, including injury type, severity, athlete demographics, medical history, and rehabilitation protocols. The model is trained on a diverse dataset of historical injury records, ensuring its generalizability and applicability across various sports and athlete populations.

Our web application framework, built upon the Flask framework, provides a versatile platform for web application development. Seamless communication with a PostgreSQL database, enabled by SQLAlchemy, facilitates the storage and retrieval of vital athlete and injury data. Advanced features such as session management, dynamic HTML rendering, and an external notification system enhance the user experience, ensuring a seamless and engaging interaction for athletes, coaches, and relevant authorities.

Rigorous data preprocessing, including one-hot encoding for categorical variables and feature engineering, optimizes the model's predictive capabilities. Ensemble methods, decision trees, and fine-tuning through cross-validation further refine the model's performance. To ensure transparency and facilitate informed decision-making, we emphasize model interpretability, providing insights into the factors influencing injury recovery predictions.

The work fills a critical gap in sports management by providing a robust, user-centric platform for predicting sports-related injury recovery periods. Through the integration of cutting-edge technologies and machine learning methodologies, we have created an impactful tool that empowers sports professionals to optimize rehabilitation processes, make informed decisions, and contribute to the sustained success of athletes in the competitive arena.

The rest of the paper is organized as follows. Section 2 details the literature survey. The proposed method is described in section 3. The system architecture is discussed in section 4. The conclusion and future scope are summarized in section 5.

II. LITERATURE SURVEY

The authors of [1] talk about preclinical imaging, radiomics, recuperation, muscle injury, and computed tomography (CT). It is possible to accurately estimate the volume of a skeletal muscle injury using radiomics. The tools used are Pearson's Correlation Coefficient and Least Squares Linear Regression. The paper's authors [2] Determine how accurate the ML techniques currently in use are at predicting injury. To obtain a comprehensive understanding of the subject, statistical techniques such as logistic or linear regression would need to be taken into account. There is usage of algorithms such as data preprocessing techniques (imputation, standardization, and discretization).

Document [3] Write a summary of the most recent research on pediatric genitourinary organ injuries associated with rugby or football. The PRISMA statement is followed when doing a systematic review. A briefing on the use of statistical learning to forecast future injuries in Australian football players is provided in [4]. Complex injury relationships may not be well suited for linear models like logistic regression. The limited number of injuries in the data tended to be over-fit by complex models (RF and SVM). The techniques used in the work include generalized estimating equations (GEE), random forest (RF) models, logistic regression (LR), and training load quantification.

The use of artificial intelligence for performance prediction and injury risk assessment in team sports is covered in paper [5–6]. the work Use artificial neural networks, decision tree classifiers, support vector machines, and Markov processes to predict injury risk and athletic performance in athletes participating in team sports. Anterior Cruciate Ligament injuries, preoperative neurological status, injury mechanism, and electrophysiological studies are all covered in detail in [7-8]. A multifactorial approach is used to account for both internal and external risk factors involved.

The authors of [10] outline the existing data regarding injury prediction in sports and draw attention to the distinctions between association and prediction in this regard. In terms of type, location, prevalence and incidence, recurrence, and severity grade, the injury patterns are described in Paper [11]. The authors of paper [12] provide an overview of sport nutrition, energy requirements, recuperation, decreased muscle mass, and rehabilitation nutrition for athletes recovering from injuries.

III. PROPOSED WORK

Initially the systems that were proposed have two filter techniques that are used, a supervised technique that is Mutual Information and the other one is the unsupervised technique that is Pearson's Correlation Co-efficient, in order to quickly minimize the dimensionality of the dataset and afterwards we proceed with the three wrapper techniques which are Backward Elimination, Forward Selection and Bi-Directional Elimination to find the optimal feature combination for each ML Algorithm. Linear regression being one of the most well known and understood ML algorithm for regression can be used as a bench mark study. A variation in Leave One Out Cross Validation[LOOCV] was the technique used for the evaluation of the model's performance which is also known as the K-Folds validation taken to its logical extreme, with K equal to N, the number of data points in the set. The LOOCV method allows us to use more data on the training of our models than any ither validation method. According to this method, we have two kind of data, one is the training dataset and the other one is the validation dataset. As a final step of the methodology, they implemented ensemble learning techniques on the results of the ML regression algorithms to further improve the predictions of the system.

Data Preprocessing:

Handling Categorical Features: Categorical variables such as injury type and gender are typically one-hot encoded to convert them into numerical form. This transformation allows SVR to work with these variables.

Feature Scaling: Scaling of numerical features might be performed to ensure that all input variables have the same influence on the model. This step can prevent certain features from dominating the prediction.

Outlier Handling: Outliers in the data may be identified and addressed to ensure they do not unduly influence the SVR model's training.

Considerations for Training Data:

Data Quality: Ensuring the quality of historical data is vital. This includes checking for missing values and errors in the dataset.

Data Diversity: The dataset should ideally cover a diverse range of injuries, ages, gender, calorie intakes, and other variables to capture various scenarios.

Model Training:

Kernel Function Selection: Choosing an appropriate kernel function is essential. The choice of the kernel function (e.g., linear, polynomial, RBF) affects how the data is transformed to capture non-linear relationships.

Hyperparameter Tuning: The SVR model has hyperparameters that need to be tuned for optimal performance. This may involve grid search or other techniques to find the best hyperparameters for the specific dataset.

Regularization:

SVR typically includes regularization parameters (C and epsilon) to control the trade-off between model complexity and prediction accuracy. Proper tuning is essential.

Transformation and Prediction:

The SVR model uses the kernel function to transform the input data into a higher- dimensional space. This transformation enables the model to find non-linear patterns in the data, as SVR is particularly effective in capturing non-linear relationships.

After the SVR model is trained, it is used to make predictions. When a user provides their injury-related data as input, SVR takes into account the specific parameters provided by the user and applies the knowledge it has gained during training from the historical dataset. It calculates the estimated recovery time based on this information.

OBJECTIVES

Injury Severity Classification: Create a classification model that can categorize the injuries based on their severity. The severity factors may include values such as mild , moderate and severe, in order to guide appropriate treatment strategies.

Predict Recovery Time: Build a predictive model that estimates the expected recovery time for the sports person with injuries allowing coaches and the medical staff to plan training schedules and rehabilitation programs.

Injury risk assessment: Develop a system that assess an athlete’s risk of sustaining the injury based on the training load and physical conditions including the considerations of calorie intake.

Return to play decision support: Design a model that assists in making decisions about when an injured athlete is ready to return to play safely by taking the considerations of recovery progress and risk factors.

IV. SYSTEM ARCHITECTURE

System Architecture is a roadmap of the system which provides an outline of how different components of the system work together to make the system work efficiently. It is just like a map or a blueprint, guiding the development and ensuring the smooth connections of the system.

According to the architecture depicted in figure 1, the system iterates through the mentioned steps where first step is to receive the data input from the user.

The next step is preprocessing of the data which is done through three steps, namely: Data Normalization, Feature scaling and One-Hot Encoding.

Later, the processed data is used to execute the algorithm, construct the regression lines and compare them with the regression lines of training data.

Through the help of this comparison, the prediction is performed and the recovery period is returned as the outcome of the model.

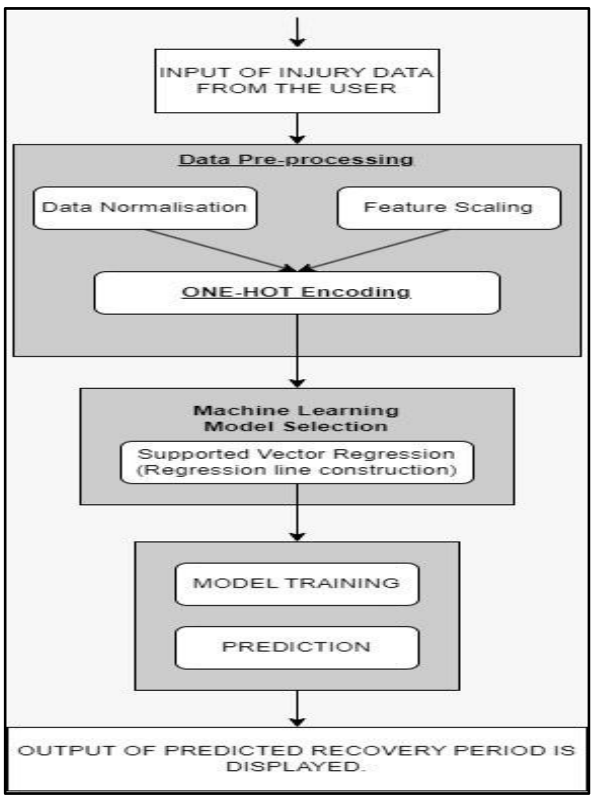

Figure 1. Schematic representation of the Architecture

WORK FLOW DIAGRAM

Work Flow diagram is a road map that iterates the workflow of the system in the stepwise manner. There is a clear understanding about each and every step of the system process and their fore aids in decision making process.

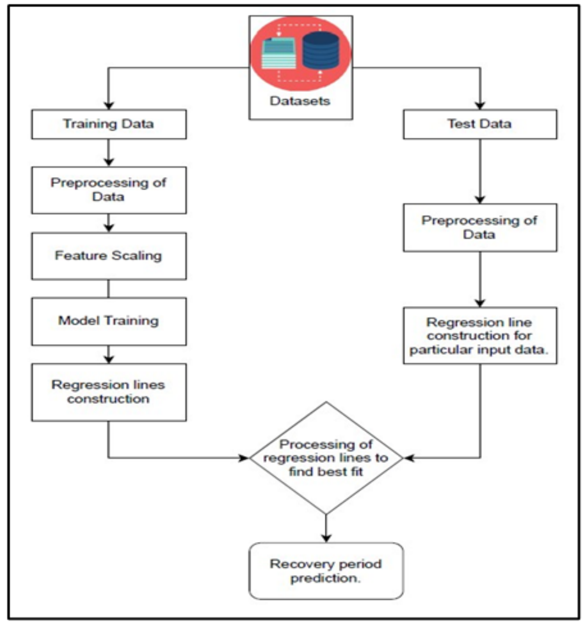

Figure 2. Schematic Representation of Workflow Diagram

1. The initial step is to obtain the inputs from the user regarding the injury occurred to particular athlete.
2. Data preprocessing takes place in the next step which includes three steps namely:
    - Data Normalization: Data normalization is the process of organizing and transforming the data in order to eliminate the redundancy and anomalies that are present in the data.
    - Feature scaling: Feature scaling is the preprocessing methodology in machine learning that standardizes the range of independent variables to ensure equal contribution during the model training.
    - One-Hot Encoding: It is a technique in the machine learning that converts the categorical variables into binary vectors. Each category is represented by a unique binary vector with a 1 at the corresponding category index and 0's in other places.
3. Machine Learning Model constructs the regression lines for the received input data by the user.
4. The regression lines that are constructed according to the input of the users is compared with the priorly constructed regression lines according to the training data too find the best fit and returns the prediction.

SEQUENCE DIAGRAM

Sequence diagram is a detailed time line of the conversations between the various components or objects of the system. It is a type of visualization diagram that gives a clear explanation about the transfer or exchange of data between the components of the system.

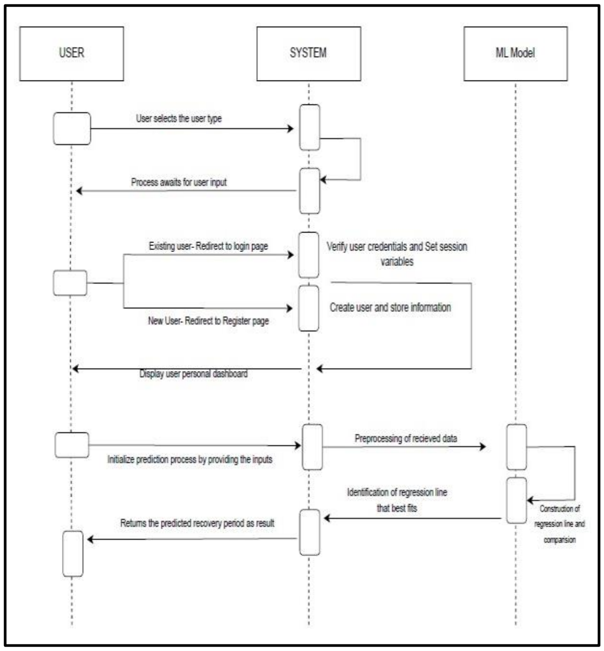

Figure 3. Schematic Representation of Sequence Diagram

MODULES

•User Interface: Design and develop user interfaces and implement user interaction elements like forms, buttons etc.

•Business Logic: Implementation of application logic such as, User authentication, Data validation and External page routing.

•Data access: Access to the various data management system to store and maintain data.

•Machine Learning: Training and Deployment of the machine learning model that aids in the recovery period prediction of the user.

Key aspects of the work include:

- Data Collection and Pre-processing: Meticulously gathering and processing relevant data to extract meaningful information that influences recovery periods.
- Feature Engineering: Identifying and crafting informative features from the data to improve model performance.
- Machine Learning Algorithms: Employing and optimizing advanced machine learning algorithms, such as regression models and decision trees, to accurately predict recovery durations.
- Model Interpretability: Ensuring that the model is interpretable and explainable, enabling insights gained from predictions to contribute to a deeper understanding of injury recovery factors.

By combining diverse datasets and advanced analytics, this project aims to enhance the precision of recovery period predictions, empowering sports professionals, coaches, and medical teams to make informed decisions regarding athlete rehabilitation and return to play.

INPUT GIVEN AND OUTPUT EXPECTED

Input: The input that the system takes is the textual input from the user in the form of either questions or fields where the user is expected to provide all the necessary details in order to perform the prediction. This input provided by the user will be pre-processed and the features are categorized.

Output: The output that is expected from this system is the injury recovery period for the particular injury mentioned by the user in terms of days. The user can also expect some advisory measures to be followed in order to aid his recovery process.

V. CONCLUSION

A. Limitations of the system:

•First and foremost, limitation of this particular system is that, we are focusing only on one particular sports category which helps us use limited dataset to train the model.

•In order to make this system more efficient and to be used in all the sports, it takes a large amount of data in order to train the model which increases the complexity of the system.

•Technical limitations such as dependency on external API's may cause some in-efficiency in the working of the system.

•Other technical frameworks and the libraries used while implementing the system may also create minor problems in the proper working of the proposed system.

•Maintenance of the large amount of data becomes the challenge and requirement for a better database dependency and availability is required.

B. Future Enhancements of the System:

•The major future enhancement that can be done to this particular project is to add the additional sports management features, like expanding the training datasets by collecting the data from various other sports and training the model in an efficient way.

•Enhancing safe and secured way of deployment of system over web networks with usage of stable external dependencies which may include libraries or API's.

•Mobile application connection support can also be included where the features of this system which is available on the web project may also be implemented through mobile application.

•Connection to smart devices such as Smart watches and Fit-bands can be implemented.

•Enhancements of the better notification system can be added to the system where the important information about the athlete and his injury details can be shared with the concerned people.

REFERENCE

[1] Bahr R, Krosshaug T (2005) Understanding injury mechanisms: a key component of preventing injuries in sport. Br J Sports Med 39:324–329 https://doi.org/10.1136/bjsm.2005.018341

[2] Hans, Luciana, Christophe, "Machine learning methods in sport injury prediction and prevention: a systematic review", Journal of Experimental Orthopedics https://jeo-esska.springeropen.com/articles/10.1186/s40634 -021-00346-x

[3] Fonseca ST, Souza TR, Verhagen E, van Emmerik R, Bittencourt NFN, Mendonça LDM, Andrade AGP, Resende RA, Ocarino JM (2020) Sports Injury Forecasting and Complexity: A Synergetic Approach. Sports Med Auckl NZ 50:1757–1770 https://doi.org/10.1007/ s40279-020-01326-4

[4] Bittencourt NFN, Meeuwisse WH, Mendonça LD, Nettel-Aguirre A, Ocarino JM, Fonseca ST (2016) Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition-narrative review and new concept. Br J Sports Med 50:1309–1314

[5] Claudino JG, de Capanema D, O, de Souza TV, Serrão JC, Machado Pereira AC, Nassis GP, (2019) Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. Sports Med - Open 5:28. https://doi.org/10.1186/s40798-019-0202-3

[6] Bittencourt, Meeuwisse, Mendonça LD,Ocarino, Fonseca (2016): "Complex systems approach for sports injuries: moving from risk factor identification to injury pattern recognition- narrative review and new concept", Br J Sports Med 50:1309–1314.

[7] R Bahr, T Krosshaug, "Understanding injury mechanisms: a key component of preventing injuries in sport", Volume 39, Issue 6, British Journal of Sports Medicine

[8] Dosang Cho, Kriangsak Saetia, Sangkook Lee, "Peroneal nerve injury associated with sports-related knee injury", Neurosurg Focus 31 (5):E11

[9] Vangelis, Vasilis, Christos, Dimitris, "A Data Science approach analyzing the Impact of Injuries on Basketball Player and Team Performance", International Hellenic University, Thermi, Greece, Version of Record 4 March 2021

[10] Joshua, Stuart, Cormack, Whiteley, Morgan, Ryan, Timmins, David, "Modeling the Risk of Team Sport Injuries: A Narrative Review of Different Statistical Approaches", School of Behavioral and Health Sciences, Australian Catholic University, Melbourne, VIC, Australia, Volume10-2019

[11] Philip von Rosen, Annette Heijne, Anna Frohm, "High Injury Burden in Elite Adolescent Athletes: A 52- Week Prospective Study"

[12] Sousana K. Papadopoulou, "Rehabilitation Nutrition for Injury Recovery of Athletes: The Role of Macronutrient Intake", International Hellenic University, Published:14 August 2020.