

## ORIGINAL ARTICLE



Converging Healthcare &amp; Technology

## INTERNATIONAL JOURNAL OF CONVERGENCE IN HEALTHCARE

Published by  
IJCIH & Pratyaksh Medical Care LLP

www.ijcih.com

## Exploratory Data Analysis for Predicting Injury in Soccer Players

Durgansh Sharma<sup>1</sup>, Akhilesh Tiwari<sup>2</sup>, Vaibhava Sharma<sup>3</sup>

<sup>1</sup>Associate Professor, School of Business and Management, Christ (Deemed to be University), Delhi-NCR, Ghaziabad, U.P.-201003, <sup>2</sup>Associate Professor, School of Business and Management, Christ (Deemed to be University), Delhi-NCR, Ghaziabad, U.P.-201003, <sup>3</sup>Senior Executive, NLB Services, Noida, U.P.-201301

## Abstract

This research work proposes the model for analysing the outcome for sports injury amongst soccer players. The most important part is to keep the player motivated towards the game while keeping the player away from the game for that instance. In this work, we are keeping injury as a dependent variable along with multiple independent variables affecting the injury. The decision-making model using R is adapted for generating the model for predicting sports injury. The inclusion of predictive analytics transforms the soccer player's moves, which after processed using the proposed framework predicts the injury in advance depending upon the life styles of the respective soccer player. The majority of the injuries like Cruciate Ligament Rupture, Hamstring Injury, Ankle Injury are taken under consideration towards this research work.

**Keywords:** Sports Injury, Lifestyle injuries, Soccer player injuries, Data Scraping.

## Introduction

The injured player never would stop playing and it may lead to further worsen the injury and time towards its respective healing. The entire future of the player may get affected if proper guidance won't be provided during the time, the player is injury prone. To reduce the related sports injuries, it requires exploration of the reasons behind the injury and its related features before the implementation of solutions towards the prevention of injuries. The analysis of sports injury leads towards the correct ways for

prevention which are required during the execution of the game and preparatory exercises specifically for structural measures to alert the sports persons towards injuries that may occur and spoil the entire career.

As per the research problem data, mapping and augmentation were major concerns. We face major challenges towards the availability of the dataset for the related work. We need to create a dataset by mapping the related variables from the following websites "transfermarket.com" and "lequipe.fr". The transfermarket.com provided various attributes about soccer players like player's name, sponsored club, current age, and injury history, while from lequipe.fr we extracted the weight of the players. The software libraries "requests" and "BeautifulSoup" were used as a web scraping tool to extract the needful data from the related web presence. The most crucial phase was to merge the two datasets in one dataset, to merge them, it was needed to find common attributes between the two datasets like player's name and

## Corresponding Author:

**Dr. Durgansh Sharma**

Associate Professor, School of Business and Management, Christ (Deemed to be University), Delhi-NCR, Ghaziabad, U.P.-201003  
e-mail: durgansh.sharma@christuniversity.in

club. But the club in the two datasets may be different depending upon the dataset requirements. Thus the usage of “lib” and “difflib” for computing deltas was needed to be applied that allow us to compare sets of data.

This research work has provided an eye-opener towards the life changing threats for the soccer player in the form of injuries which affects their entire career prospects. Furthermore, we tried to analyse all the perspectives and finally proposed a model to help the players to get recovered fast and remain motivated and focussed to their natural game. We have analysed sportsperson wise most frequent injuries like Cruciate Ligament Rupture, Hamstring Injury, Ankle Injury.

**Review of Literature:** Prediction of a soccer game mainly depends upon the physical and psychological fitness of the soccer player. This research work has worked upon the establishment of prediction towards the winning of the soccer game using various deep learning techniques and compared them to find out the best among them for the prediction of the game. The data used towards prediction was comprised of 592 FIFA world cup soccer matches. The suggested framework of this research article shall be helpful in future research work once extended towards the physical fitness of the player as well. <sup>[1]</sup>

This research work has proposed an image analysis model to avoid the injuries to the soccer players keeping the reflected outcome that the 1/5<sup>th</sup> of the accidental injuries during the game of soccer have existed due the wrong posture during real time playing of the game. The research work extracted the wrong postures of the players from the images captured during the game played. The wrong stance could lead to a major injury, which can be avoided once informed to the player to improve the posture from region of interest of the player, which is measured and reflected out of 3 million annotations of the players. <sup>[2]</sup>

The referred research work found helpful towards predicting sports injury in the game of soccer or football, it elucidated that the medical perspective is referred as the most important perspective for the analysis of the game. The evidence from Medical Journal of Australia, Victoria were also considered towards the sports injury in various games including soccer or football. <sup>[3]</sup> The data analysis of soccer game and respective prediction for the dynamic sports like soccer or football is a topic for the research

towards time-to-event analysis, it was satisfactorily provided in the research work from British Journal of Sports Medicine. <sup>[4]</sup>

The proposed framework was considering multiple independent factors helpful to predicts the results of soccer matches. This research work focused on time-series approach to generate the knowledge base for that respective match which helped on sports match prediction and enhanced the approach to predict sports matches. The framework delivered was used to reflect the tides/flows of a sports match which helps prediction more accurate. They have implemented a football results predictor called FRES (Football Result Expert System) based on their framework to provide reasonable and stable predictions for the soccer matches. <sup>[5]</sup>

This research work endeavoured to forecast the outcome of several sporting events. They have tried to automate ratings of the football or basketball gameantedate the upcoming powerful computers and related networks which have amplified to the current internet services. They visualized the randomness amongst the outcomes of football and basketball games and proposed a framework towards deciding the ratings of the game played according to the independent factors and probabilistic calculations as respective outcomes. <sup>[6]</sup>

## Methodology

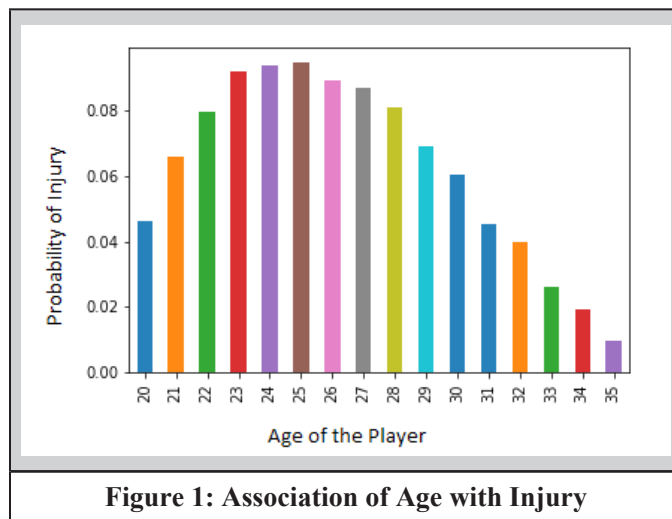
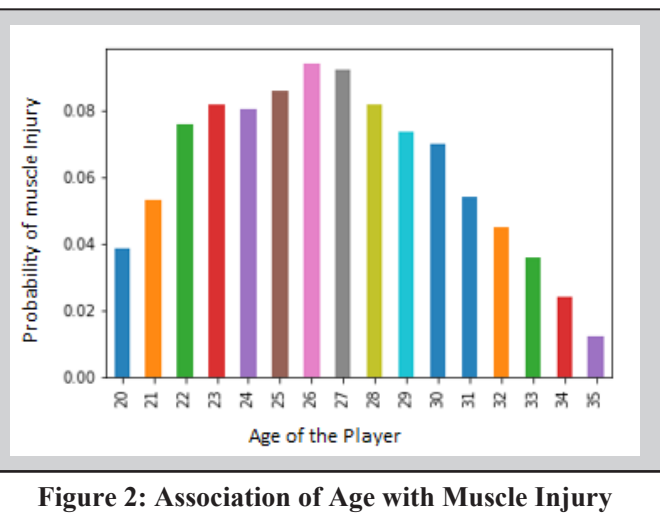
After associating the two different datasets and culminating into one source we got the data of 2580 players, out of which we removed 59 players whose date of birth was not mentioned in the available dataset. The injuries extracted from the dataset comprised of (Hamstring Injury, Muscular problems, Muscle Injury, Ankle Injury, Unknown Injury, Knee Injury, Thigh Problems, Knee Problems, Torn Muscle Fibre, Knock, Minor Knock, Fitness Adductor problems, Shoulder Injury, Cruciate Ligament Rupture, Back trouble, Influenza, Calf Injury, Thigh Muscle Strain, Calf Problems, Tear in the abductor muscle, Groin Injury, Rest Foot Injury, Achilles tendon problems, Bruise, groin strain, Muscle Fatigue, Ankle problems, Cold, Strain, Abdominal Influenza, Flu, Biceps femoris muscle injury, Concussion, Torn muscle bundle, Hip Injury, Leg Injury). While analysing the frequency of different types of injuries for soccer players, we found the details mentioned in Table-1.

**Table 1: Frequency of Injuries**

| Injury Type             | Frequency |
|-------------------------|-----------|
| Hamstring Injury        | 825       |
| Muscular Problem        | 666       |
| Muscle Injury           | 583       |
| Ankle Injury            | 516       |
| Unknown Injury          | 514       |
| Achilles Tendon Rupture | 314       |

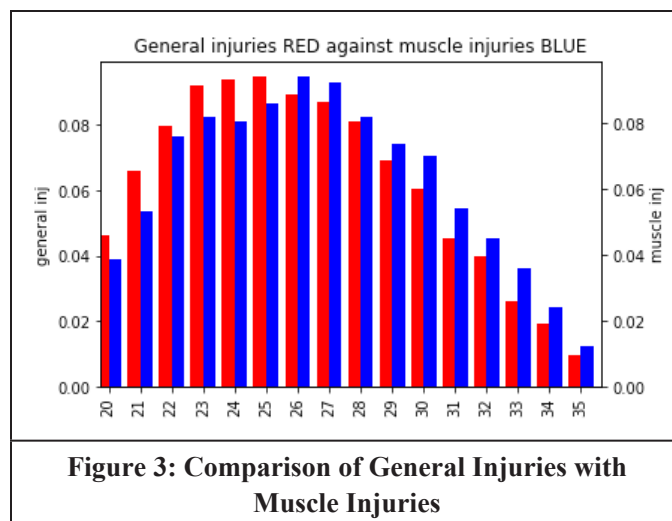
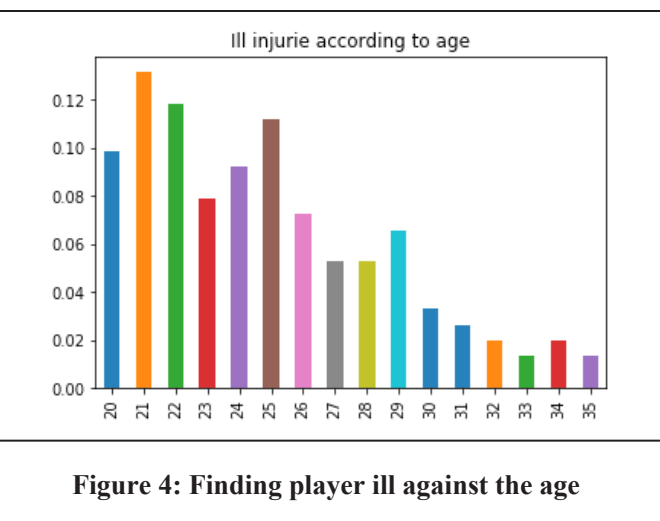
Then we selected the set of muscle injuries majorly labelled for the soccer players like “Hamstring Injury”, “Torn Muscle Fibre”, “Thigh Muscle Strain”, “Achilles Tendon Problem” etc. and try to associate them with the age of the player.

There were a set of questions directly or indirectly associate with the injury of the player, initially first parameter was age of the player and relevance with injury. Figure-1 reflects that the soccer players within the range of 23 to 27 years of age are more prone towards getting injured.

**Figure 1: Association of Age with Injury****Figure 2: Association of Age with Muscle Injury**

The Figure-2 reflects the association amongst the muscle injury and age. But, no major differences were shown. Then we mixed both the outcomes of injury with age of the player. The general injuries were marked in “Red” whereas the muscle injuries were marked in “Blue”. The outcomes in Figure 3 reflects that the muscle

injuries were bound to happen later in the sports career of the player when the player get older. Then we plotted a histogram towards the injury being ill against the age of the player as the outcomes of Figure 4 reflects that the younger players are falling ill more than the older ones, depending upon their lifestyles.

**Figure 3: Comparison of General Injuries with Muscle Injuries****Figure 4: Finding player ill against the age**

Next move was to analyse the recovery time from an injury according to player's age. The data which reflects more famous injuries like Cruciate Ligament Rupture, Hamstring Injury, Ankle Injury etc. were taken under consideration. The entire dataset comprising of injuries was applied with group by with the specific injury and calculated the average of recovery time. But, it further

needed data cleansing by removing all the rows of sports person that don't have recovery time. As reflected in Figure 5 we found that for the three injuries that we choose namely "Cruciate Ligament Rupture", "Hamstring Injury", "Ankle Injury" we can't lead to any conclusion, but when we observed all injuries together, it inferred that the graph decline with the age.

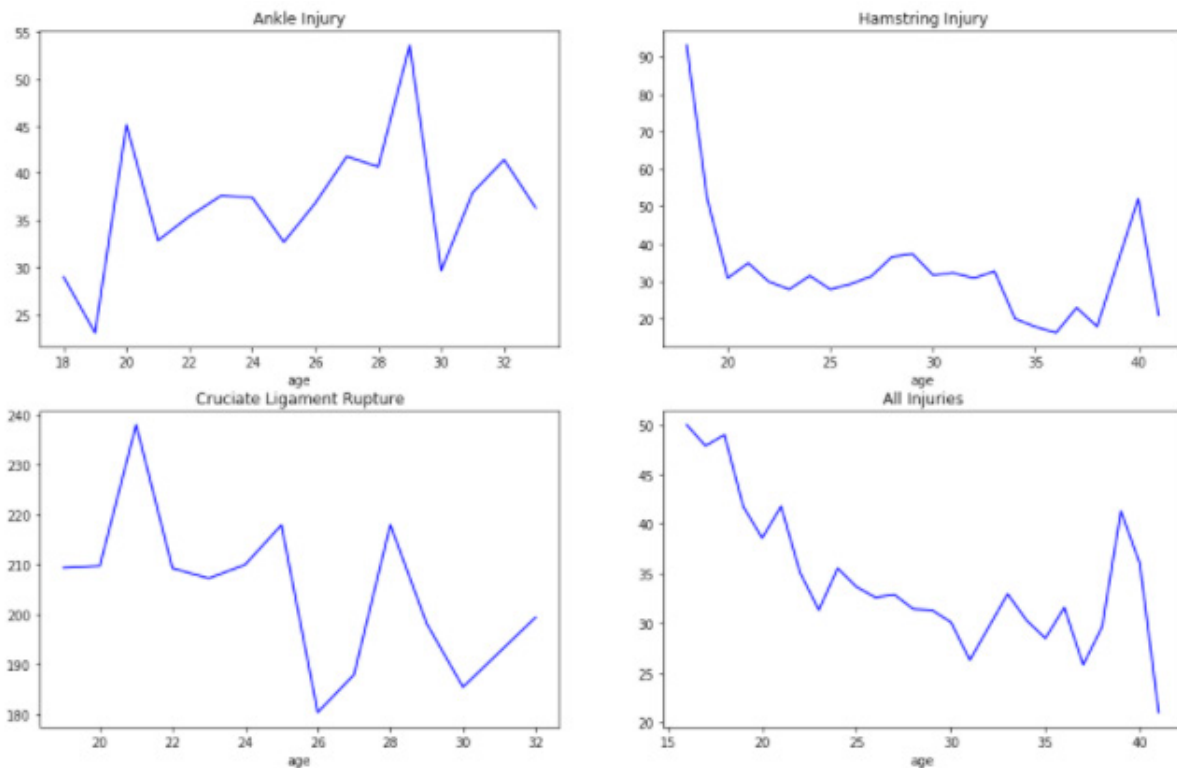


Figure 5: Time taken for Injury Recovery according to age

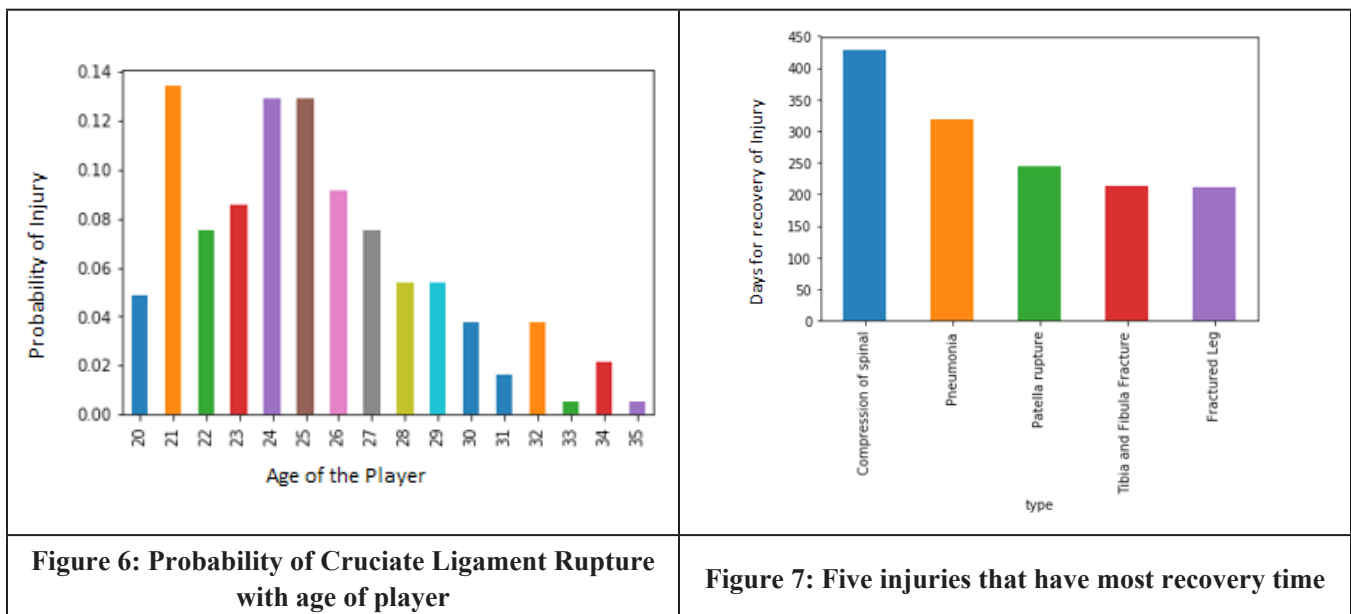
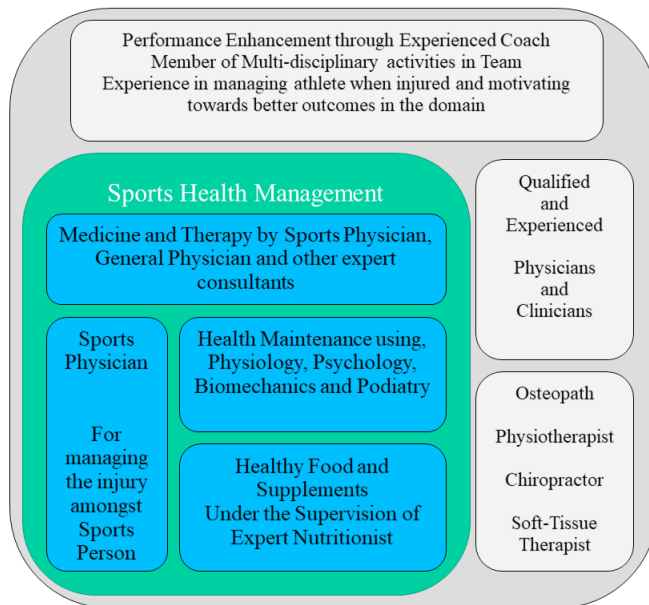


Figure 6: Probability of Cruciate Ligament Rupture with age of player

Figure 7: Five injuries that have most recovery time



**Figure 8: Medical Support-Injured Soccer Player**

Furthermore, looking forward to the famous injury ‘Cruciate Ligament Rupture’ that occurs in soccer players with older age, we analysed and found that it is affecting more of the younger players as reflected in Figure 6. This again reflects the incorrect lifestyle and playing styles of the younger players and not the older ones.

While expediting towards the recovery time of injuries occurred to the soccer player, we analysed the top-five injuries taking most of the time towards recovery from the injury as reflected in Figure 7 and in turn affecting the player’s career prospects as well.

The method adopted to enhance the capturing of a player’s injury well in time and provide the healing shall be the best way for prospects of the player. The proposed model reflected in Figure-8 is showcasing all the parameters feasible towards the medical revival of the player. As a subset of Sports Health Management, comprised of the converged and timely availability of medicines, physicians, and right supplements to heal up the injury. Chiropractors and Soft-tissue experts would be a blessing in disguise for the player. Thus, it is suggested that the right physiotherapist or occupational therapist with the working knowledge as soft-tissue healer and chiropractor shall be given priority appointment in the team. Above all, an experienced coach with a member of multi-disciplinary activities shall revive the player in very little time and never let the player be disheartened.

## Conclusion and Future Scope

This research work finally proposed a model in Figure 8, which shall be further expedited with the help of IoT based artificially intelligent and related systematic automation towards better healthcare for the bright soccer player in future.

**Source of Funding:** Self

**Conflict of Interest:** Nil

## References

1. Sharma D, Sharma V. Optimal Predictive Model for FIFA Worldcup. International Journal of Recent Technology and Engineering. 2019 Apr; 7(6c). <https://www.ijrte.org/wp-content/uploads/papers/v7i6C/F90410476C19.pdf>
2. Sharma D, Sharma V. Novel Architecture for Reducing Sports Injuries in Football. International journal of convergence in healthcare. 2021 Jan 16;1(1):10-6. <https://www.ijcih.com/index.php/ijcih/article/view/5>
3. Tharanga Fernando D, Berecki-Gisolf J, Finch CF. Sports injuries in Victoria, 2012-13 to 2014-15: evidence from emergency department records. The Medical Journal of Australia [Internet]. 2018;208(6):255–60. Available from: <https://www.mja.com.au/journal/2018/208/6/sports-injuries-victoria-2012-13-2014-15-evidence-emergency-department-records>
4. Nielsen RO, Bertelsen ML, Ramskov D, Møller M, Hulme A, Theisen D, et al. Time-to-event analysis for sports injury research part 2: time-varying outcomes. British Journal of Sports Medicine. 2019;53(1):70–8.
5. Min B, Kim J, Choe C, Eom H, McKay RI. A compound framework for sports prediction: The case study of football. Knowledge-Based Systems. 2008;21(7):551-62.
6. Stefani RT. Football and basketball predictions using least squares. IEEE Transactions on systems, man, and cybernetics. 1977 Feb;7(2):117-21.