CS 686 Project Proposal – Predicting Concussion Risk in NHL Players

Ahmed Abbas

10/20/24

Application Domain

The application domain lies in assessing athlete health. This is interesting to me because I am a big NHL fan (go Flames go!). I recall one of my favourite players, Sidney Crosby, having concussion issues early on in his career, sidelining him for nearly two seasons. This project can be a preventative measure in saving athlete careers. This domain will be slightly difficult because injuries in sports, particularly hockey, have so many factors that influence them. This will make it difficult to generate an accurate risk assessment based on data alone. Thus, some challenges that may arise from this domain are understanding non-obvious correlations between player features and injury risk, as well as ensuring that any AI models used don't make any over generalized predictions. It will be challenging to weigh several player features against each other, to create a well informed and nuanced prediction.

The Problem

The problem I want to tackle is developing an AI system that analyzes historical and present NHL player data to identify any trends associated with concussion occurrences and then predict which current NHL players are vulnerable to future concussions.

Currently, Zone7 is one of the leading AI systems for injury risk forecasting and is working with top soccer clubs to help them make better informed decisions. On the other hand, there is no de facto AI tool used by hockey teams for injury risk prediction to my knowledge. However there has been a study done showing that XGBoost outperformed logistic regression in predicting next-season NHL player injury. It might be a better project idea to try to replicate the results of this study, but it is not specific to concussions. Nevertheless, it will be interesting to see how my results compare to the study's.

The major challenge associated with this problem is that a large number of features that can lead to a concussion such as concussion history, player physical stats (height, weight, age), player position, average time on ice, and many more will need to be accounted for. I will also need to be wary of creating an AI system that can stigmatize players marked as high risk, thus reducing their playing opportunity.

In terms of data, there are datasets on Kaggle that can provide me with the following information: NHL players with concussions from 2012-2023, NHL player injury history from 2000-2023, NHL player information (body size, position, season stats), and NHL game data (including areas of the ice where specific plays were made). These datasets can possibly be combined to create another dataset that contains all the risk factors I want to include.

Artificial Intelligence Techniques

The main technique will be the usage of machine learning classification models, particularly logistic regression, Random Forest, K Nearest Neighbors, and finally XGBoost. These models can allow the system to account for a large number of features that can contribute to concussion risk. They can help predict the likelihood of future concussions based on historical and biometric data. The accuracies of each model can also be compared against each other.

Literature Review

In the references listed, the first three describe how machine learning methods can be used to predict injury in sports in general, and review the current methods. The next two references attempt to find a correlation between general NHL injuries and the many factors that can contribute to injuries. Both references have interesting results where they find which factors do and do not have correlation with NHL injury rates. The next four references get deeper into the meat of my specific problem which is concussions in the NHL. They all aim to analyze the risk factors associated with concussions in the NHL as well as bridging the gap in scientific literature on concussions in ice hockey. All of these references provide interesting statistics on when the most concussions were sustained. Finally, the tenth reference on the list is the most important, as it utilizes machine learning classification models to predict next-season NHL injuries. This is very similar to the problem I am tackling; I will be paying close attention to this study, and compare my results with Luu *et al*.

Planned References

Van Eetvelde, Han,s *et al.* Machine learning methods in sport injury prediction and prevention: a systematic review. *J EXP ORTOP* 8, 27 (2021). https://doi.org/10.1186/s40634-021-00346-x

Amendolara, Alfred, *et al*. An Overview of Machine Learning Applications in Sports Injury Prediction. *Cureus* (2023). https://doi.org/10.7759/cureus.46170

Claudino, Joao Gustavo, et al. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Med Open* (2019). https://doi.org/10.1186/s40798-019-0202-3

McKay, Carly D., et al. "The epidemiology of professional ice hockey injuries: a prospective report of six NHL seasons." *British journal of sports medicine* 48.1 (2014): 57-62. https://bjsm.bmj.com/content/48/1/57

Kneeland, Aaron T. "Man Games Lost in the NHL: A Correlation between Travel, Rest Periods and Injuries in the National Hockey League." *Sport Management Undergraduate* (2014). https://fisherpub.sjf.edu/sport_undergrad/4/

Hutchison, Michael G., et al. A systematic video analysis of National Hockey League (NHL) concussions, part I: who, when, where and what? *British Journal of Sports Medicine* (2015). https://doi.org/10.1136/bjsports-2013-092234

Izraelski, Jason. Concussions in the NHL: A narrative review of the literature. *J Can Chiropr Assoc* (2014). https://pmc.ncbi.nlm.nih.gov/articles/PMC4262814/

Navarro, Sergio M., et al. Short-Term Impact of Concussion in the NHL: An Analysis of Player Longevity, Performance, and Financial Loss. *Journal of Neurotrauma* (2018). https://doi.org/10.1089/neu.2017.5611

Benson, Brian W., et al. A prospective study of concussions among National Hockey League players during regular season games: the NHL-NHLPA Concussion Program. *CMAJ* (2011). https://doi.org/10.1503/cmaj.092190

Luu, Bryan C., *et al*. Machine Learning Outperforms Logistic Regression Analysis to Predict Next-Season NHL Player Injury: An Analysis of 2322 Players From 2007 to 2017. *Orthop J Sports Med* (2020). https://doi.org/10.1177/2325967120953404