

# **DEVELOPMENT OF MACHINE LEARNING ALGORITHM AND NEURAL NETWORK CAPABLE OF PREDICTING INJURY TO ATHLETES AND PRODUCING POSSIBLE REHABILITATION TIMELINES THROUGH ANALYSIS OF POSITION-BASED INJURY DATA FROM THE NATIONAL FOOTBALL LEAGUE**

Harish Viswanathan and Yayun (Lydia) Hsu, Ph.D. Columbia University  
Columbia University School of Professional Studies

## **Abstract**

Drawing in approximately 400 million fans worldwide, American football is one of the most popular sports in the world. However, the athletes that participate in this sport are always one injury away from having their careers ended. Football has one of the highest rates of injury out of any sport and players are often pushing themselves through the season with multiple injuries. Injuries in American football are especially difficult to avoid because of the physical demand that the sport places on athletes and they are often regarded as being simply part of the sport due to the high frequency of injury experienced amongst players. This study aims to create an algorithm using machine learning and big data analyses to predict potential injuries based on players' positions, field terrain, previous injury history, speed and direction, and the play that is called on the field. Additionally, if the player is predicted to be met with an injury, a rehabilitation timeline is provided based on their injury history and seriousness of the injury.

The research study was conducted using R and RStudio as the software used to create and test methodology. The National Football League provided the data that was used to conduct the study. Initially, the players were classified based on their positions and the history of their injury. An exploratory data analysis was conducted to identify trends in the data and determine correlations between the position of players and their risk to injury. Under similar methodology, other exploratory data analyses were conducted to identify various other trends. Correlations between field terrain and injury history, speed and injury possibility, direction and injury possibility, as well as the play call and injury possibility were all investigated through exploratory data analyses. Once trends were determined, we clustered the data based on the position of the player and the injured body part. A predictive model was then created using nearest neighbor and linear regression to produce a supervised learning algorithm. The model was capable of determining injury possibility to a specific player and also a specific body part. The model was then tested against known player injury records provided by the National Football League. Once the machine learning algorithm had been validated, a single layer Perceptron neural network was employed to create an algorithm that classified whether or not an injury was going to occur. Based on the results of this algorithm, a rehabilitation timeline would be suggested based on the location of the injury and its severity.

The results of the study showed that most injuries were lower body injuries that occurred at the knee and on synthetic field surfaces. The cause of the injuries appeared to stem from the hyperextension or flexion of muscles. Additionally, it was concluded that linebackers contracted the most injuries out of any position group. It was also determined that most injuries occurred during the start of the season and would gradually decline till the end of the season. Also, most injuries were known to occur during passing plays compared to running plays. Furthermore, there are higher rates of injury possibility at the defensive positions compared to offensive positions. The accuracy of the predictive model and the rehabilitation timeline were verified using data from previous player injuries gathered from the National Football League.

The application of the results of this study was tested using data from a local high school football team each game. The predictive model was put to use to determine injury possibility among the high school athletes and the rehabilitation timeline aided in the diagnosis of players' injuries by the school's athletic trainer. The results from the predictive model were also used to work with the team coaches to prevent injury through prevention measures such as mandatory team stretches and yoga sessions.