**3. Literature Review**

3.1 - **Introduction to Injury Prediction in Sports**

An extensive amount of research has been conducted in the world of sports injury prediction and prevention none more so than the NFL the sheer amount of data and money involved makes the need for an effective injury prediction system vital for team success. The NFL generated somewhere in the region of $16bn in 2019[11] showing that there is the money available to conduct studies on injury prevention. The effect of a singular players injury can not only effect on field performance but it can also hurt the team financially through lost revenues from negative results[1]. According to a study conducted in 1991 an estimated 1.2 million athletes are injured across all age groups in one entire football season[2], this number is staggering and it shows the need for a comprehensive prediction model that could hopefully be implemented and benefit all throughout the football system. We have seen in recent years in both the NFL and Rugby union the push for the mitigation of head injuries as a result of poor tackle technique, however it remains a massive problem in both sports. Recent studies have found that the NFL concussion rate was on a steady increase in the mid 2010’s reaching as high as 16% in 2017/18,[7] and these effects can cause massive issues in the lives of NFL players as it opens them up to a risk of developing Chronic Traumatic Encephalopathy (CTE) in their later lives as we have seen around 96% of deceased ex-NFL players have had CTE been diagnosed as the ultimate cause of death [7]. The lower body is also of major risk when playing NFL as the 1991 study found that injuries to the lower extremities being the knees and ankles accounted for 50% of all injuries in their study[2], these are often caused due to poor field and weather conditions but also through poor tackling technique. A similar study in 2001 which focused on injuries to Quarterbacks showed that 82.3% of injuries were caused by direct contact while overuse injuries only accounted for 15% of total quarterback injuries, showing us that tackling technique and style is clearly a problem within the NFL and actions must be taken to limit the amount of injuries we see today[5].

3.2 - **Influence of External Factors on Injury Rates**

Various external factors that can influence the likelihood of someone getting injured in the NFL from pitch conditions to team performance which I will examine in this section. Pitch conditions have been a point of contention in recent years with it appearing as though more players are developing serious injuries in their lower extremities such as ankles and knees. A study conducted in 2009 found that artificial surfaces produce greater torque and rotational stiffness[12] which can likely lead to more injuries. Weather is another crucial variable that can influence injury risk, studies have shown that sports that have a winter calendar tend to experience greater injury rates than those of summer calendars or with indoor arenas[13]. Research has also shown us that playing on a wet surface can lead to more ligament injuries[1]. Cold exposure can also lower muscle efficiency leading to more injuries[1]. The amount of travel and rest provided to NFL players between matches is another important variable given the nature of the sport it requires lots of travel and lots of heavy contact rest between games is important. However, according to a study conducted in 2023, the cumulative travel distance of each NFL team had little statistical significance on the amount of injuries they had[6]. The timing of the bye week also had very little statistical significance on injuries[6] implying rest is maybe not as important as we think. One finding of that 2023 study was that teams who made it to the playoffs tended to have less injuries than teams who fell short[6], this is maybe because lower performing teams are forced to play their best players even when hurt in an attempt to make the playoffs, while better teams have the luxury to rest theirs. Rule changes are probably the main external variable that people have control over as the league are able to make rules to make tackling safer to reduce injuries. It has been shown in the past to have a big impact on injuries as we seen between 1975-1984 the rate of cervical spine injury fell dramatically due to new regulations being brought in around tackle technique and safety[2].

3.3 – **Genetic and Physiological factors in injury likelihood**

A player's likelihood of injury is influenced by a variety of factors, including their genetics as well as psychological factors. A player’s mental state can significantly influence their injury risk as it can cause them to make bad decisions leading to injury. Studies have suggested that players who are injured tend to have a higher susceptibility to stress compared to players who aren’t injured[14]. Research also suggests that the ratio of internal workload such as psychological factors and external workload from GPS data to be an important predictor for injuries[8]. The Functional Movement Screen(FMS) is a test of an athletes movement techniques to see if the athletes movements could put them at risk of injury. A study conducted on 46 football players on the FMS found that players who scored lower than 14 on the test were 11 times more likely to get injured in the first 3 weeks of the season[9], this shows the importance of proper movement technique in preventing injuries. Genetics can also play an important role in predicting a players possibility of injury. Studies have shown that a family history of ligament injuries is prominent in athlete injury[15]. These studies show us that in some cases athletes are just genetically presupposed to getting injured.

3.4 – **Injury Distribution**

The most common injuries that occur to NFL players are lower body injuries in the hamstrings, calves, ankles and knees as these account for almost 50% of American football injuries[2], with strains and sprains accounting 40% of injuries and fractures dislocations accounting for 10% and 15% respectively[2]. The current incidence rate per 1000 players for injuries within the NFL is between 23.1and 64.7[11] which does seem quite high, however, in comparison with professional rugby which has an incidence rate of between 83.9 and 90.1[16], although there method for calculation may be different. A study conducted on NFL quarterbacks between 1980 and 2001 found 1534 injuries in this period[5], these quarterbacks missed on average 18.8 days with a median of 6 days[5]. This really shows the importance of being able to predict when a player will get injured as this can prevent players from missing future games. This same study found that 70% of shoulder studies occurred while the quarterback is being tackled[5], showing that finding methods to limit contact to the quarterback while also understanding when to pull the player to prevent injury is vital to maintaining team performance throughout the season. A teams playstyle can also play an important role of predicting how likely a player is to get injured, a team who runs a lot with their quarterback would be more likely to face injuries as it puts them at risk of contact injuries. The study on quarterback injuries found that 12.6% of all quarterback shoulder injuries happen when the quarterback is acting as a ball carrier[5]. This may seem low however, quarterback runs account for roughly 1% of plays within an NFL game[17]. The most common type of shoulder injury that happen to quarterbacks is an AC joint sprain while being tackled, on average quarterbacks missed 22.1 days with a median of 12.5 days[5]. Now we are going to look at how we might be able to predict these injuries and the challenges involved with that task.

3.5 – **Challenges with injury Prediction Modelling**

The idea of injury prediction using machine learning is certainly not a new phenomenon as people have attempted to do this across various sports. A major problem that people have discovered when trying to create these models is the ‘black box’ nature of these machine learning models, this makes it very difficult for implementation as the results can be difficult to interpret and lack any real reporting transparency[1]. Another major problem for these previous studies is data availability, given the specified domain of the problem set it’s very difficult to create a comprehensive model for all sports, meaning all current studies have relied on limited data based on a specific team or specific leagues[1][3]. Data imbalance is massive challenge for all previous studies, the nature of how infrequent injuries occur means that the ratio of instances where injuries do take place to instances where injuries do not take place is massively skewed in favour of the latter category[3][7] [8]. This makes modelling incredibly difficult and emphasises the importance of sampling methods and generating synthetic data to ease modelling. The Injury prediction space remains relatively new and so therefore the consensus on which features are most appropriate and beneficial for the model is not a well investigated question, new studies should focus on encompassing a wider range of features, this will help to discover the most important features and will help to build a comprehensive model that may be more beneficial across multiple sports[1][8].

3.6 – **Preventative Measures and Risk Management**

As mentioned above an intelligent system for injury prevention will be incredibly beneficial for coaches, as it can provide them actionable insights that will help to decrease the likelihood of injury for their players[1]. Preseason and off-season sessions should focus on conditioning and limit physical contact drills so as to prevent injuries while also putting players in a position to be able to take contact without experiencing injuries[2]. It’s also important that an injured player is not rushed back and his return-to-play is managed focusing on conditioning as a preventative measure to ensure there isn’t a risk of reaggravation[2]. Studies have emphasised the importance of a rigorous preseason regime to prepare players for the NFL season[4], the effects of a limited preseason were seen during the covid struck 20/21 season where an increase in injury rate throughout the season[4]. When considering the predictive models, a comprehensive list of internal and external factors should be considered when creating such models, also by leveraging various models and classifiers the system can better adjust to a players specific needs and be able to provide personalised injury prevention strategies.[1]

3.7 – **Conclusion and Future Directions**

As discussed any future work in this topic should focus on expanding the data collection methods currently in place, this will allow more accurate and versatile models to be built[1]. Also by consolidating all this information into one comprehensive source, it will allow for greater research to take place across multiple sports, which will have a knock on effect of improving overall usefulness of these models for coaches. A key challenge for this research is the ‘black box’ nature of the models and so placing a greater emphasis on any future research on creating more explainable AI and having better transparency, this will create more trust in the models and lead to greater levels of implementation. This review examined the various factors be that internal or external that effect the likelihood of injury and so focusing on analysing and optimising the factors so as to build the most efficient and accurate models is vital. This review discussed the problems current research has on data imbalance, measures must be implemented to limit the effects this imbalance will have on the models being created. Ultimately, the implementation of predictive analysis for injuries will have a massive effect on how the sport of NFL is coached by providing greater insights for management while also helping to minimise the influence injuries will have on the revenue for teams and the league.