



A football player wears a vest holding a GPS sensor. The data captured feed into an algorithm.

Could an algorithm predict an injury?

Machine learning can tell athletes when to train and when they risk an injury that could jeopardize their career. **By Andrada Fiscutean**

In 2005, 17-year-old aspiring footballer Alessio Rossi tore two ligaments in his right ankle during training for lower league Italian football club USD Olginatese. The injury ended his dream of playing at the highest level. Today, Rossi is a postdoctoral researcher at the University of Pisa, Italy, where he collects and analyses reams of data to help prevent players at top teams getting injuries of their own.

When Rossi was playing, his coaches' instincts and experiences were all they had to predict whether he might receive an injury. Now, a footballer training with top-level teams, such as those in the English Premier League, will wear a tight top under their jersey outfitted with GPS, an accelerometer, a gyroscope and a digital compass. While they run drills, the sensors track their heart rate,

speed and distance covered.

"We follow a team for an entire season, recording GPS data during training and matches," Rossi explains. He then uses machine learning to try to detect patterns. "This gives us the probability that a player will get injured in the next days or next weeks."

These data reveal an athlete's workload – how often they train and how intensely. Just enough training can pave the way to medals, but too much puts pressure on the body and can lead to injuries. Coaches have always taken into account the condition of players when scheduling training sessions. Now they can calculate more precisely the probability that individual athletes will get injured during the next match, the next week or the next month.

Professional footballers experience between 2.5 and 9.4 injuries per 1,000 hours

of exertion (D. Pfirrmann *et al.* *J. Athl. Train.* **51**, 410–424; 2016), about one-third of which are from overuse and therefore potentially predictable. Most injuries last about a week, but recurrent ones – about 15% of the total – often require more rest. During this time, the player's physical and mental conditioning declines, and their careers might suffer as a result of damage to their reputation or concerns over their ability to fully recover. Their time off might also increase their teammates' workload, increasing the likelihood of injuries throughout the squad.

The problem is even worse outside elite athletics. Youth sport, for instance, is experiencing "a pandemic of injuries", says Dhruv Seshadri, a biomedical engineer at Case Western Reserve University in Cleveland, Ohio. Athletes as young as 10 are pushing themselves harder, trying to propel themselves towards a career in professional sports. The rate of injury for youth football players, for example, can be as high as 19.4 injuries per 1,000 hours of exertion.

Sports scientists started using data analytics such as those Rossi employs only in the past decade, but the hope is that the approach could save careers and money, as well as improve results. Researchers and coaches are working to develop methods of collecting and analysing the quantity of data required to make predictions, not only for team sports such as football or basketball, but also for events such as figure skating or tennis, in which people often compete as individuals.

A multivariate approach

Assessing workload and predicting injuries the old-fashioned way – through experience – has some value, but also limitations. Rossi's aim is to help coaches combine what their gut is telling them with what the data reveals. To do that, the first step is to collect as much information as possible.

In a trial, his team used a sensor package inserted into clothing to monitor 26 elite football players, recording 931 individual training sessions over the course of 23 weeks (A. Rossi *et al.* *PLoS ONE* **13**, e0201264; 2018). From this, the researchers extracted 12 variables, including total distance run, distance run faster than 5.5 metres per second, and the number of high-intensity accelerations and decelerations, which put extra stress on the body.

In other sports, the most useful data to gather might differ. Baseball players, for instance, can wear smart sleeves with accelerometers that measure joint angles, velocity and stress, whereas figure skaters can attach accelerometers and gyroscopes to their hips to record jumps. Earrings, bodysuits, vests

or bands can measure heart rate and oxygen saturation, and wristbands can record sleep quality and body temperature. Some sports scientists also incorporate contextual data, such as a player's mood, body mass index and previous injuries, as well as how much water players have drunk in a given period and how far they have travelled recently by bus or by plane.

This multivariate approach is key to predicting injuries, says Derek McHugh, a data scientist at Kitman Labs in Dublin. The sports analytics company partners with professional football, basketball and ice hockey teams. "In the last five or ten years, teams have been capturing huge volumes of data," McHugh says. But having that information is one thing; making sense of it is another. "There's a gap between how much data we've been capturing and what's actually been done with it," he says. "We're trying to fill that gap using machine learning."

The value of reasoning

In Rossi's trial, he used decision-tree classifiers – a supervised machine-learning technique that involves asking a series of questions based on different variables to reach a conclusion. The variables in Rossi's model include an athlete's previous health issues, the total distance they have covered in a training session and the distance covered at high speed. By asking a series of questions of the data, the system is able to predict 80% of injuries – although for some recurrent health issues, such as specific sprains and strains, the system can spot the warning signs almost every time.

Other researchers use variations on Rossi's decision-tree-based method, such as 'random forest' or 'gradient boosting' techniques, which use multiple decision trees to incrementally improve forecasts. Another machine-learning technology, known as deep neural networks, could yield even greater accuracy. The technique still relies on parameters such as previous injuries or total distance run, but in this case, the exact rules used to make predictions are not known by the data scientist. However, although neural networks are popular in many areas of science, Rossi thinks that the approach is not currently workable in sport. These algorithms are a black box – the reasoning behind their findings is not easily interpretable. Coaches want to know why an athlete might get hurt, not just that they will, he explains. "We prefer to reduce the performance of our algorithm, but to provide interpretable results."

McHugh agrees, stressing that the ability to explain how an algorithm reached the conclusion it did is paramount in team sports. "If a coach can't easily understand why the



GPS data from sensors carried by players can help coaches to decide who is at risk of injury.

algorithm has arrived at its injury prediction, they can't begin to take action to mitigate an athlete's elevated risk," he says. "Similarly, when an athlete gets injured, they want to know what were the things that were under their control that contributed to that injury, and what they can do to avoid those in the future."

In individual sports such as figure skating, however, the machine-learning systems used by analysts such as Rossi are still out of reach. The relatively small amount of information that can be gathered from an individual means that analysis is often done manually, on a week-by-week basis, says Lindsay Slater, a sports scientist at the University of Illinois at Chicago. "The ultimate goal is to automate everything, but it's just difficult when everything is so unique to each skater," says Slater, who is also sports science manager for the US national figure skating team. Only as more athletes wear sensors will researchers have the chance to collect enough data to automate the process.

A possible solution would be to aggregate data from numerous teams and individuals practising the same sport. McHugh and Rossi both regard this as the ultimate goal of sport analytics – a step that could markedly improve injury predictions. But doing so will be a challenge. First, teams play at different levels and in different leagues. Second, sportspeople wear different devices, so manufacturers would need to agree to follow the same standards. And third, competitors would need incentives to share information that could improve their rivals' fitness as well as theirs.

The AI teammate

Sport is gradually entering a new era, in which artificial intelligence might act as an assistant coach. Algorithms could enable a teenager to train smarter and avoid a career-ending injury,

or help a professional athlete to compete for a few years longer. But the technology's success depends, in part, on the ability of data scientists to convince coaches to include data in their decision process.

The teams that McHugh has worked with have seen a reduction in injuries of between 5% and 40%, he says. Yet, not every coach is happy to join forces with AI. "Coaches sometimes don't feel good, because it seems like trying to substitute the human element," Rossi says. But in reality, data is only a tool. "The interpretation of the results, the change of the training load, is done by coaches," he says.

McHugh agrees that people have to make the final call. "Once the injury probability for an athlete on a given day is output from an injury model, the athlete or coach must then decide whether the predicted risk is acceptable or not, usually depending on the context," he says. There might be a big game that day, and the player might be especially important to the team. "Even though the predicted injury probability may be as high as 70%, the coach may be willing to take that chance," he says.

The pitch to coaches should be that marrying their years of experience with the machine's analytical insights offers an advantage over relying entirely on intuition. As sensors become more accurate and wearable, and algorithms become more powerful, the accuracy of AI-powered predictions of injury will improve. "What wakes me up in the morning is the opportunity to help the athlete in ways that we couldn't do yesterday," says Seshadri. Still, in sports there will always be an element of surprise. "We control what we can, and know to expect the unexpected," says Slater.

Andrada Fiscutean is a freelance writer in Bucharest, Romania.