

La integración de los datos en el fútbol de élite. Un nuevo paradigma de investigación

The Integration of Data in Elite Football. A New Paradigm of Research

José Luis Felipe

Grupo de Investigación IGOID. Universidad de Castilla-La Mancha.
Área de Control del Rendimiento. Unión Deportiva Las Palmas

Antonio Alonso-Callejo

Grupo de Investigación IGOID. Universidad de Castilla-La Mancha

Resumen

La tecnología se ha consolidado en el deporte de élite en los últimos años y se utiliza de forma rutinaria, especialmente en el fútbol de élite. Es preciso destacar que, mientras que los procesos subyacentes a las tácticas en el fútbol de élite han mejorado a lo largo de los años, los enfoques científicos no han evolucionado con la misma rapidez. La solución ante este problema es la integración de las nuevas tecnologías y el big data en el día a día de los cuerpos técnicos del fútbol de élite. De este modo, el mundo del fútbol debe aprender a registrar, almacenar, analizar y aplicar toda la variedad y volumen de datos disponibles en aras de la mejora del juego y del espectáculo. Así, los tres grandes retos que se perfilan en el mundo del fútbol en los próximos años son: la prevención de lesiones, la orientación de las tareas de entrenamiento y el desarrollo técnico-táctico. Los nuevos sistemas y técnicas de big data aplicadas al mundo del fútbol permiten implementar el ciclo PDCA (Plan, Do, Check, y Act), considerándose una herramienta válida y fiable para implementar un modelo de resolución de problemas en el contexto del fútbol de élite.

Palabras clave: Big data, Fútbol de élite, Prevención de lesiones, Tareas de entrenamiento, Desarrollo técnico-táctico.

Clasificación JEL: C32, D83, L83

Abstract

Technology has become established in elite sports in recent years and is used routinely, especially in elite football. It should be noted that, while the processes underlying tactics in elite football have improved over the years, scientific approaches have not evolved at the same pace. The solution to this problem is the integration of new technologies and big data into the day-to-day operations of elite football coaching staffs. In this way, the world of football must learn to record, store, analyze, and apply the wide variety and volume of data available for the sake of improving the game and the

spectacle. Thus, the three major challenges that loom in the world of football in the coming years are: injury prevention, orientation of training tasks and technical-tactical development. The new big data systems and techniques applied to the world of football allow for the implementation of the PDCA (Plan, Do, Check, and Act) cycle, considered a reliable and valid tool for implementing a problem-solving model in the context of elite football.

Keywords: *Big Data, Elite Football, Injury Prevention, Training Tasks, Technical-Tactical Development.*

1. Introducción

La tecnología se ha consolidado en el deporte de élite en los últimos años y se utiliza de forma rutinaria, especialmente en los deportes de equipo. El seguimiento de los patrones de movimiento revela resultados importantes en relación con el rendimiento deportivo. En el ámbito del fútbol semiprofesional o de élite, el mejor sistema para cuantificar estos resultados y el perfil cinemático de los jugadores son los Sistemas de Posicionamiento Global, GPS, (Buchheit *et al.*, 2015). Estos dispositivos se han definido como una herramienta válida para la evaluación de la carga externa en deportes intermitentes, con capacidad para registrar en tiempo real datos de tiempo, velocidad, distancia, posición, altitud y dirección (Folgado *et al.*, 2014), lo que los hace muy habituales en el análisis de los deportes de equipo (Cunniffe *et al.*, 2009; Higgins *et al.*, 2009).

Además, es preciso mencionar que se ha incorporado la tecnología de *tracking* (sistema de análisis de partidos con múltiples cámaras), implementada en casi todas las ligas profesionales europeas y competiciones internacionales. Estos sistemas han demostrado ser tan fiables como los sistemas GPS y proporcionan resultados con menos del 5% de error, por lo que han demostrado ser válidos para investigación (Felipe *et al.*, 2019). De este modo, es habitual utilizar este tipo de sistemas para estudios de rendimiento físico y táctico en diferentes competiciones de élite (Di Salvo *et al.*, 2007; Memmert *et al.*, 2017).

Es preciso destacar que, mientras que los procesos subyacentes a las tácticas en el fútbol de élite han mejorado a lo largo de los años, los enfoques científicos no han evolucionado con la misma rapidez (Felipe *et al.*, 2021). La solución ante este problema es la integración de las nuevas tecnologías y del *big data* en el día a día de los cuerpos técnicos del fútbol de élite (Rein y Memmert, 2016). En este sentido, debemos definir el *big data* a través de las denominadas tres V: Volumen, Variedad y Velocidad (Noor *et al.*, 2015). El *big data* tiene como objetivo proporcionar una forma estandarizada para que los investigadores accedan a algoritmos de procesamiento complejos y permitir que usuarios no expertos apliquen tecnologías de análisis de vanguardia a sus datos (Goeks *et al.*, 2010).

De este modo, el progreso de la ciencia y la tecnología ha hecho posible obtener datos de posicionamiento sobre los movimientos de los jugadores (Baca *et al.*,

2009). Los sistemas de seguimiento de posicionamiento establecen la posición del jugador y del balón en las coordenadas X (paralelas a las líneas laterales) y las coordenadas Y a 25 Hz, lo que equivale a unas 135.000 posiciones por sujeto y partido, y un total de unas 3.100.000 posiciones entre todos los jugadores y el balón. La recopilación de todos estos datos y su posterior análisis facilitarán la simplificación de la teoría y la práctica en el deporte (Low *et al.*, 2020). En este sentido, el análisis y la aplicación práctica en los cuerpos técnicos de los datos, tanto de parámetros físicos como técnico-tácticos, ha ido incrementándose en los últimos años.

Así, el uso e integración diaria de estos datos y el desarrollo de modelos predictivos es uno de los temas más apremiantes para la investigación en Ciencias del Deporte (Rein y Memmert, 2016). En concreto, investigaciones previas han incluido variables de rendimiento físico (velocidad máxima y velocidad media en intervalos de 15 minutos; distancias recorridas a diferentes velocidades; distancia media recorrida; velocidad máxima en sprints; aceleración pico; y número de aceleraciones de jugadores en diferentes rangos) y rendimiento técnico (número de pases, control del balón, entradas, remates de cabeza, tiros a puerta, córners y tiros libres por áreas: área defensiva [tercio del campo más cercano a la zona defensiva], área central [tercio central del terreno de juego], y área ofensiva [tercio del terreno de juego más cercano a la portería rival]; número de pases de corta distancia [<10 m], promedio [>10 m], y distancia mínima [<2 m]; número de tiros a puerta; porcentaje de posesión del balón; y entradas e intercepciones a intervalos de 15 minutos, Rein y Memmert, 2016).

Para determinar el comportamiento táctico de los jugadores o equipos, se requiere una serie de variables que demuestren la dinámica de comportamiento de estos agentes. Los centroides, la profundidad y el ancho, y las tasas de estiramiento del equipo parecen proporcionar una base sólida para analizar el comportamiento colectivo en la dinámica del equipo de ataque y defensa (Memmert *et al.*, 2017). Dentro de estos conceptos, el centroide del equipo se define como la posición media de todos los jugadores de un equipo; presenta baja variabilidad cuando se trata de medir la coordinación entre jugadores en un partido de fútbol. Sin embargo, la variación entre cada jugador y su posición específica (centroide individual) ha sido identificada como una variable potencial para determinar con mayor precisión el comportamiento dinámico del jugador (Sampaio y Maças, 2012). Desde una perspectiva de sistema dinámico, se espera que la alta variabilidad en la distancia entre equipos (distancia entre el centroide de dos equipos) refleje perturbaciones en el equilibrio entre el comportamiento del equipo, que preceden a eventos críticos del juego, como oportunidades de gol (Memmert *et al.*, 2017).

Así, el mundo del deporte en general, y del fútbol en particular, debe saber registrar, almacenar, analizar y aplicar toda esta variedad y volumen de datos en aras de la mejora del juego y del espectáculo. Para ello, el objetivo de este trabajo es dar a conocer las diferentes tipologías de *big data* aplicables al mundo del fútbol pro-

fesional. En este contexto, los tres grandes retos que se vislumbran en el mundo del fútbol en los próximos años son:

Prevención de lesiones

El control de la carga física de los jugadores es uno de los grandes objetivos que persiguen los departamentos de preparación física de equipos y deportistas profesionales. Conocer la carga diaria interna y externa del deportista permite reajustar los objetivos de entrenamiento y optimizar el rendimiento físico. Los datos obtenidos gracias a la tecnología mencionada anteriormente pueden sincronizarse con la monitorización de la carga interna generada por el estímulo del entrenamiento, así como la carga mecánica a la que se ve sometido el deportista.

En este sentido, la inteligencia artificial, el *big data* y el *machine learning* se han venido utilizando desde hace años en medicina y ciencias de la salud para generar modelos predictivos de enfermedades (Deo, 2015; Kononenko, 2001). Sin embargo, el aprendizaje automático es aún bastante nuevo en las Ciencias del Deporte y la medicina deportiva (Claudino *et al.*, 2019; Edouard *et al.*, 2022), y, en particular, en la ciencia aplicada al fútbol. El uso del aprendizaje automático se basa en la suposición de que los algoritmos aprenderán a medida que se les alimente con más y más datos. Después de la recopilación y limpieza de datos, los algoritmos pueden construir relaciones entre variables, ya sea sin asistencia humana (enfoques de aprendizaje no supervisados) o con ella (supervisada), quien les proporciona los valores de corte para variables específicas. A través de la alimentación repetida de datos, las computadoras y los algoritmos aprenderán a identificar y seleccionar, entre una gran cantidad de variables, aquellas que explican la variable dependiente (Ayala *et al.*, 2019; López-Valenciano *et al.*, 2018).

El problema con los enfoques anteriores es que los datos recopilados durante los meses previos a una lesión no pueden capturar la naturaleza dinámica de las actividades de fútbol y los cambios que pueden ocurrir en los jugadores. Los jugadores están expuestos a una carga de trabajo específica que puede modificar la relación entre un parámetro y el riesgo de lesión (Bahr, 2016). De hecho, se supone que la precisión con el aprendizaje automático se puede mejorar con datos (p. ej., carga de trabajo, preparación física de microciclo a microciclo, datos fisiológicos y contextuales como condiciones del campo de entrenamiento, tipo de calzado) recopilados lo más cerca posible del evento (la lesión). Los datos en tiempo real sobre la fisiología del jugador (p. ej., frecuencia cardíaca, temperatura corporal y asimetrías térmicas), respuestas mecánicas (p. ej., alteraciones inducidas por la fatiga en la mecánica de movimientos) y otros factores contextuales (p. ej., condición del campo, propiedades de la superficie del calzado) podrían añadir valor y mejorar la precisión de las técnicas de aprendizaje automático en el futuro. Este será uno de los

mayores desafíos en el corto plazo con la integración de *big data* y *machine learning* de una manera que aporte una ventaja competitiva en el fútbol de élite.

Hay al menos dos problemas más asociados con el uso del aprendizaje automático en la predicción de lesiones en el fútbol:

- 1) la baja incidencia de lesiones específicas que impiden que los algoritmos alcancen una mayor precisión (Ekstrand *et al.*, 2023). En el fútbol de élite se pueden esperar alrededor de 15 lesiones por equipo durante una temporada, un número muy pequeño para crear modelos de predicción con alta precisión en un solo equipo.
- 2) La incertidumbre al aplicar algoritmos de aprendizaje automático a otro entorno o al utilizar un procedimiento de recopilación de datos ligeramente diferente al original. La situación se vuelve más desafiante cuando se consideran los diferentes tipos de lesiones dentro de un equipo de fútbol. Cabe tener en cuenta que dentro de un mismo equipo de fútbol pueden darse múltiples tipos de lesiones diferentes, lo que hace que el muestreo específico disminuya exponencialmente, produciéndose el llamado desequilibrio en los datos (Rommers *et al.*, 2020).

Por lo tanto, la aplicación del *big data* y el *machine learning* a la predicción de lesiones aún parece estar lejos de ser altamente predictiva. Sin embargo, la aplicación de estas técnicas puede resultar muy interesante para ayudar a los cuerpos técnicos a identificar signos tempranos de riesgo elevado de lesión muscular. Las investigaciones futuras deberían tener en cuenta la naturaleza multifactorial de este tipo de lesiones para que el machine learning arroje resultados más significativos. El desarrollo de estos sistemas se perfila como el santo grial del *big data* aplicado al fútbol de élite, y aunque la resolución de este problema no parece estar cercana, es necesario continuar en esta línea de investigación para generar modelos que ayuden a maximizar el rendimiento físico y así mejorar la espectacularidad del juego.

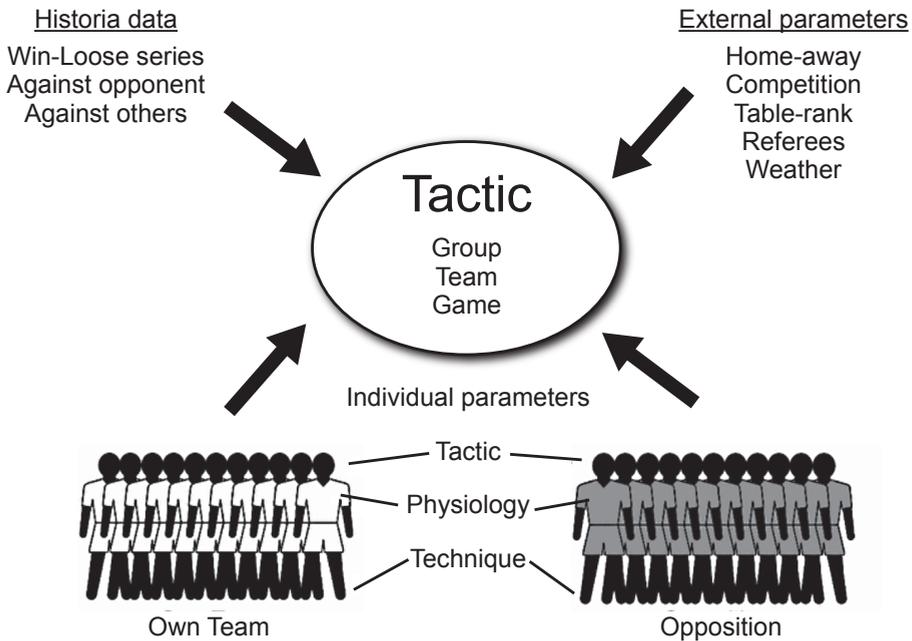
La orientación de las tareas de entrenamiento

La información que proporciona esta tecnología permite identificar los escenarios críticos de competición, es decir, las máximas exigencias a las que se ven sometidos los deportistas al enfrentarse a un partido o competición deportiva en la que su objetivo es conseguir el mejor resultado. La identificación de estas demandas físicas y técnico-tácticas nos permite orientar las tareas de entrenamiento y su planificación para facilitar su proceso adaptativo. (Higgings *et al.*, 2009).

Los datos disponibles actualmente en el fútbol profesional y semiprofesional permiten a un entrenador predecir con precisión la disposición táctica de su

equipo y del rival, y, por tanto, adaptar las tareas de entrenamiento de la semana preparatoria del partido a las demandas físicas y técnico-tácticas del rival. Para ello, debe tener en cuenta el estado del equipo, el estado del rival, así como factores externos como jugar en casa o incluso el clima (Gréhaigne y Godbout, 1995; Lago, 2009; Mackenzie y Cushion, 2013; Sarmiento *et al.*, 2014, Figura 1). Es importante destacar que las tácticas durante un partido no son constantes, sino que deben adaptarse según las interacciones entre y dentro de los dos equipos (Balagué y Torrents, 2005; Gréhaigne y Godbout, 2014). Por ejemplo, la sustitución de un jugador por parte del equipo contrario puede introducir un cambio en las tácticas de juego, al que el entrenador podría tener que responder para ajustar las tácticas de su equipo. Por lo tanto, las tácticas de equipo se rigen por un proceso complejo, que resulta de una red de parámetros interdependientes (Kempe *et al.*, 2014). Así, las tácticas pueden interpretarse como una estructura compleja compuesta por una red de dependencias entrelazadas. En consecuencia, la preparación de los partidos, y las tareas de entrenamiento destinadas a tal fin, deben apoyarse en los datos disponibles para poder adaptarse a esta complejidad.

FIGURA 1
FACTORES QUE INFLUENCIAN
EN LA TÁCTICA EN FÚTBOL



FUENTE: Rein & Memmert, 2016.

Desarrollo técnico-táctico:

La provisión de datos posicionales de los atletas permite identificar sus limitaciones y fortalezas en el comportamiento táctico. En deportes de equipo, la interacción entre los componentes de un mismo equipo, así como la presencia de los contrarios, es fundamental para obtener resultados. Por ello, la información que proporcionan estos sistemas tecnológicos permite establecer redes de interacción con el objetivo de conocer el comportamiento táctico del equipo, con el fin de potenciar las debilidades detectadas y orientar las tareas de entrenamiento en base a información cuantitativa. (Gonçalves *et al.*, 2014).

Tradicionalmente, los análisis técnico-tácticos se basaban en enfoques de análisis notacional, fundamentados en estadísticas y promedios (Hughes y Bartlett, 2002), que incluían, por ejemplo, variables de pase (Liu *et al.*, 2015), posesión del balón (Collet, 2013), recuperación del balón (Vogelbein *et al.*, 2014) o estilo de juego (Tenga *et al.*, 2010). La principal limitación del enfoque notacional tradicional es que casi toda la información contextual se descarta y, por tanto, no se considera muy eficaz para la preparación técnico-táctica en el fútbol de élite actual (Glazier, 2015).

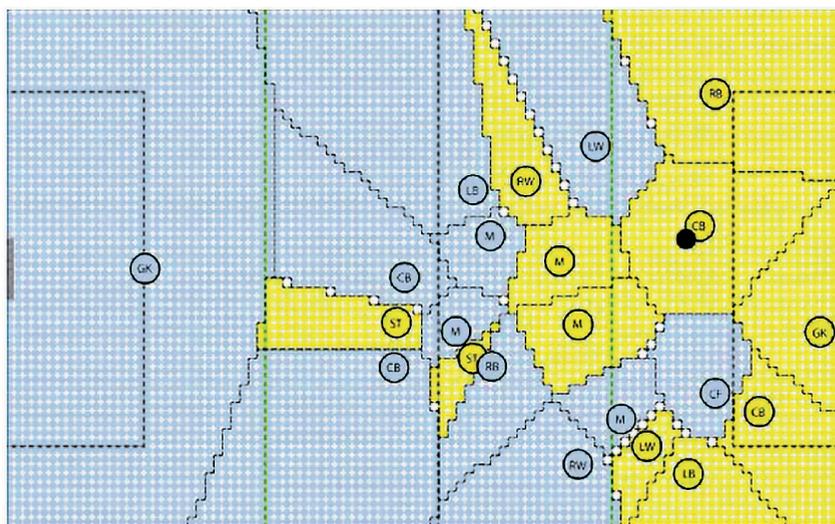
En los últimos años, los desarrollos tecnológicos en deportes de equipo han cambiado radicalmente la forma de trabajar en los aspectos técnico-tácticos día a día. Así, la implementación de sistemas de seguimiento en todos los estadios y campos de entrenamiento reduce drásticamente la incertidumbre en el proceso de toma de decisiones, y los sistemas de análisis notacional tradicionales quedan relegados a un segundo plano.

Un enfoque cada vez más utilizado para estudiar las tácticas de equipo es el método del centroide del equipo (Folgado *et al.*, 2014), que consiste en analizar el centro geométrico de las posiciones de todos los jugadores para entender el comportamiento del equipo completo. A este tipo de análisis se le añadió una mejora, incluyendo el cálculo de la entropía aproximada, ApEn, (Goncalves *et al.*, 2014). Este método sugiere que los movimientos de los jugadores son más regulares respecto al centroide de sus respectivos equipos en comparación con otros equipos. Aunque la aplicación de ApEn se está volviendo más prominente, aún queda por demostrar qué representa realmente esta medida, ya que la regularidad en el comportamiento de los centroides del equipo en sí misma es una descripción muy abstracta del comportamiento del equipo. Sin embargo, las medidas del centroide del equipo se utilizan cada vez más para analizar el comportamiento técnico-táctico del equipo y su integración en las sesiones de entrenamiento previas a la preparación de un determinado partido.

Otro tipo de análisis técnico-táctico reciente se centra en estudiar las tácticas de equipo en función del control del espacio. Con este enfoque, se utiliza, por ejemplo, el área de superficie del equipo calculada a partir del área convexa que englo-

ba a todos los jugadores de un equipo (Moura *et al.*, 2013). Los resultados de esta línea de investigación indican que los equipos atacantes cubren mayores áreas de superficie que los equipos defensivos (Moura *et al.*, 2012). También se aplican comúnmente los diagramas de Voronoi para investigar el control del espacio (Nakanishi *et al.*, 2008). Aquí, el espacio controlado se determina utilizando la ubicación y las distancias entre los jugadores individuales para determinar el espacio controlado (Figura 2).

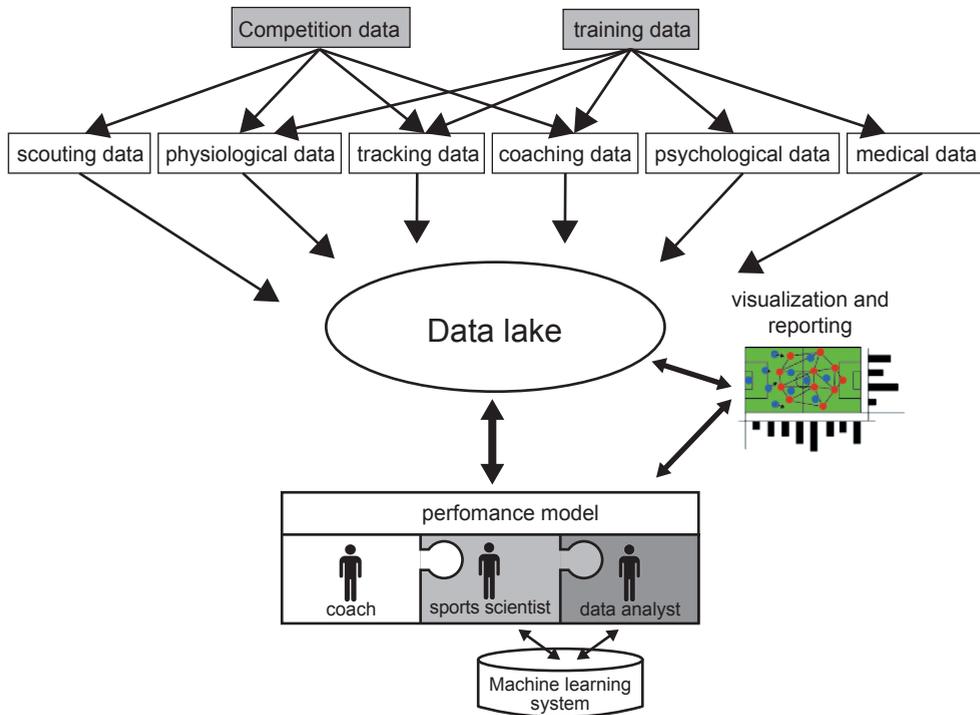
FIGURA 2
APLICACIÓN DE LOS DIAGRAMAS DE VORONOI AL FÚTBOL



FUENTE: Memmert & Rein, 2018.

No obstante, el gran avance en este aspecto se produce con la introducción de algoritmos de aprendizaje automático aplicados al análisis técnico-táctico del fútbol. Este sistema permite estudiar la toma de decisiones tácticas, los movimientos de los jugadores y su interacción tanto entre ellos como con los rivales, facilitando la predicción de comportamientos internos (del equipo) y externos (de los rivales) en función de los datos previamente conocidos (Jordan y Mitchell, 2015). La integración del aprendizaje automático en el mundo del fútbol de élite ofrece como principal ventaja la capacidad de generar modelos que, partiendo de datos conocidos, produzcan resultados precisos sobre los sistemas de juego más apropiados para implementar en función de la tipología del rival. También permite enfocar los entrenamientos de manera más efectiva e incluso generar informes automáticos de lo acontecido en entrenamientos o competiciones (Figura 3).

FIGURA 3
IMPLEMENTACIÓN DE UN MODELO DE MACHINE LEARNING AL FÚTBOL



FUENTE: Rein y Memmert, 2016.

Conclusiones

Los nuevos sistemas y técnicas de *big data* aplicadas al mundo del fútbol permiten implementar el ciclo PDCA (*Plan, Do, Check y Act*) para optimizar el rendimiento físico y técnico-táctico a nivel individual (jugador) y colectivo (equipo). Siguiendo los principios de este modelo, el sistema de seguimiento debe permitir identificar y analizar los principales problemas relacionados con la carga externa, las demandas locomotoras y el rendimiento técnico-táctico de cualquier sesión o partido (*Plan*). Posteriormente, se pueden desarrollar e implementar soluciones holísticas a las situaciones identificadas, en colaboración con el resto del *staff* (*Do*). A continuación, se evalúan los resultados y las metas alcanzadas en relación con el desempeño y las lesiones de los jugadores (*Check*). Finalmente, se estandarizan las estrategias para ayudar en la toma de decisiones respecto a los programas de entrenamiento de los jugadores (*Act*). En conclusión, los sistemas de seguimiento (*tracking*) y *big data* aplicados al mundo del fútbol pueden considerarse herramientas válidas y fiables para implementar un modelo de resolución de problemas en el contexto de los deportes de equipo de élite.

Referencias bibliográficas

- Ayala, F., López-Valenciano, A., Martín, J. A. G., Croix, M. D. S., Vera-Garcia, F. J., del Pilar Garcia-Vaquero, M., Ruiz-Pérez, I., & Myer, G. D. (2019). A preventive model for hamstring injuries in professional soccer: Learning algorithms. *International journal of sports medicine*, 40(05), 344-353.
- Baca, A., Dabnichki, P., Heller, M., & Kornfeind, P. (2009). Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences*, 27(12), 1335-1346.
- Bahr, R. (2016). Why screening tests to predict injury do not work—and probably never will...: a critical review. *British journal of sports medicine*, 50(13), 776-780.
- Balagué, N., & Torrents, C. (2005). Thinking before computing: Changing approaches in sports performance. *International Journal of Computer Science in Sport*, 4(1), 5-13.
- Buchheit, M., Gray, A., & Morin, J. B. (2015). Assessing stride variables and vertical stiffness with GPS-embedded accelerometers: preliminary insights for the monitoring of neuromuscular fatigue on the field. *Journal of Sports Science & Medicine*, 14(4), 698.
- Claudino, J. G., Capanema, D. D. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports medicine-open*, 5, 1-12.
- Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *Journal of sports sciences*, 31(2), 123-136.
- Cunniffe, B., Proctor, W., Baker, J. S., & Davies, B. (2009). An evaluation of the physiological demands of elite rugby union using global positioning system tracking software. *The Journal of Strength & Conditioning Research*, 23(4), 1195-1203.
- Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920-1930.
- Di Salvo, V., Baron, R., Tschann, H., Montero, F. C., Bachl, N., & Pigozzi, F. (2007). Performance characteristics according to playing position in elite soccer. *International journal of sports medicine*, 28(03), 222-227.
- Edouard, P., Verhagen, E., & Navarro, L. (2022). Machine learning analyses can be of interest to estimate the risk of injury in sports injury and rehabilitation. *Annals of physical and rehabilitation medicine*, 65(4), 101431.
- Ekstrand, J., Bengtsson, H., Waldén, M., Davison, M., Khan, K. M., & Häggglund, M. (2022). Hamstring injury rates have increased during recent seasons and now constitute 24% of all injuries in men's professional football: the UEFA Elite Club Injury Study from 2001/02 to 2021/22. *British Journal of Sports Medicine*, 57(5), 292-298.
- Felipe, J. L., Garcia-Unanue, J., Viejo-Romero, D., Navandar, A., & Sánchez-Sánchez, J. (2019). Validation of a video-based performance analysis system (Mediacoach®) to analyze the physical demands during matches in LaLiga. *Sensors*, 19(19), 4113.
- Felipe, J. L., Garcia-Unanue, J., Gallardo, L., & Sanchez-Sanchez, J. (2021). Tracking Systems Used to Monitor the Performance and Activity Profile in Elite Team Sports. *Sensors*, 21(24), 8251.

- Folgado, H., Lemmink, K. A., Frencken, W., & Sampaio, J. (2014). Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science*, 14(sup1), S487-S492.
- Glazier, P. S. (2017). Towards a grand unified theory of sports performance. *Human movement science*, 56, 139-156.
- Goecks, J., Nekrutenko, A., Taylor, J., & Galaxy Team team@ galaxyproject. org. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, 11, 1-13.
- Gonçalves, B. V., Figueira, B. E., Maças, V., & Sampaio, J. (2014). Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of sports sciences*, 32(2), 191-199.
- Gréhaigne, J. F., & Godbout, P. (1995). Tactical knowledge in team sports from a constructivist and cognitivist perspective. *Quest*, 47(4), 490-505.
- Gréhaigne, J. F., & Godbout, P. (2014). Dynamic systems theory and team sport coaching. *Quest*, 66(1), 96-116.
- Higgins, T., Naughton, G. A., & Burgess, D. (2009). Effects of wearing compression garments on physiological and performance measures in a simulated game-specific circuit for netball. *Journal of Science and Medicine in Sport*, 12(1), 223-226.
- Hughes, M. D., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of sports sciences*, 20(10), 739-754.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Kempe, M., Vogelbein, M., Memmert, D., & Nopp, S. (2014). Possession vs. direct play: evaluating tactical behavior in elite soccer. *International Journal of Sports Science*, 4(6A), 35-41.
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1), 89-109.
- Lago, C. (2009). The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *Journal of sports sciences*, 27(13), 1463-1469.
- Liu, H., Gomez, M. Á., Lago-Peñas, C., & Sampaio, J. (2015). Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of sports sciences*, 33(12), 1205-1213.
- López-Valenciano, A., Ayala, F., Puerta, J. M., Croix, M. D. S., Vera-García, F., Hernández-Sánchez, S., ... & Myer, G. (2018). A preventive model for muscle injuries: a novel approach based on learning algorithms. *Medicine and science in sports and exercise*, 50(5), 915.
- Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2020). A systematic review of collective tactical behaviours in football using positional data. *Sports Medicine*, 50, 343-385.
- Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research. *Journal of sports sciences*, 31(6), 639-676.

- Memmert, D., & Rein, R. (2018). Match analysis, big data and tactics: current trends in elite soccer. *German Journal of Sports Medicine*, 69(3), 65-72.
- Memmert, D., Lemmink, K. A., & Sampaio, J. (2017). Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1), 1-10.
- Moura, F. A., Martins, L. E. B., Anido, R. D. O., De Barros, R. M. L., & Cunha, S. A. (2012). Quantitative analysis of Brazilian football players' organisation on the pitch. *Sports biomechanics*, 11(1), 85-96.
- Moura, F. A., Martins, L. E. B., Anido, R. O., Ruffino, P. R. C., Barros, R. M., & Cunha, S. A. (2013). A spectral analysis of team dynamics and tactics in Brazilian football. *Journal of sports sciences*, 31(14), 1568-1577.
- Nakanishi, R., Murakami, K., & Naruse, T. (2008). Dynamic positioning method based on dominant region diagram to realize successful cooperative play. In *RoboCup 2007: Robot Soccer World Cup XI II* (pp. 488-495). Springer Berlin Heidelberg.
- Noor, A. M., Holmberg, L., Gillett, C., & Grigoriadis, A. (2015). Big Data: the challenge for small research groups in the era of cancer genomics. *British Journal of Cancer*, 113(10), 1405-1412.
- Rein, R., & Memmert, D. (2016). Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus*, 5(1), 1-13.
- Rommers, N., Rössler, R., Verhagen, E., Vandecasteele, F., Verstockt, S., Vaeyens, R., Lenoir, M., D'Hondt, E., & Witvrouw, E. (2020). A machine learning approach to assess injury risk in elite youth football players. *Medicine and science in sports and exercise*, 52(8), 1745-1751.
- Sampaio, J., & Maças, V. (2012). Measuring tactical behaviour in football. *International journal of sports medicine*, 33(05), 395-401.
- Sarmiento, H., Marcelino, R., Anguera, M. T., Campaniço, J., Matos, N., & Leitão, J. C. (2014). Match analysis in football: a systematic review. *Journal of sports sciences*, 32(20), 1831-1843.
- Tenga, A., Ronglan, L. T., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *European journal of sport science*, 10(4), 269-277.
- Vogelbein, M., Nopp, S., & Hökelmann, A. (2014). Defensive transition in soccer—are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. *Journal of sports sciences*, 32(11), 1076-1083.