Universitat Oberta de Catalunya

OIKONOMICS
Revista de economía, empresa y sociedad

DIGITALIZACIÓN Y PRL

De la prevención reactiva a la prevención predictiva: cómo los datos y la IA están transformando la seguridad y salud laboral

Xavier Baraza
Director de los Estudios de Economía y Empresa (UOC)

Joan Torrent-Sellens
Director del Centro de Investigación para la Transformación Digital y la Governanza (UOC)

Josep-Anton Mir-Tutusaus
Profesor de los Estudios de Economía y Empresa (UOC)

RESUMEN La creciente disponibilidad de datos y el desarrollo de técnicas de *machine learning* están ampliando las posibilidades de la prevención de riesgos laborales, lo que permite avanzar desde un enfoque fundamentalmente reactivo hacia una prevención anticipatoria. Este artículo analiza el papel de los modelos predictivos en esta transformación, explicando su funcionamiento y sus principales aplicaciones para anticipar accidentes, estimar su gravedad, identificar periodos de mayor riesgo y predecir otros resultados relacionados con la salud laboral. Se aborda especialmente el paso de la predicción a la acción preventiva, destacando que el valor de estos modelos no reside únicamente en su capacidad predictiva, sino en su utilización para establecer prioridades y apoyar la toma de decisiones. Asimismo, se analizan sus principales limitaciones, relacionadas con la calidad y representatividad de los datos, los sesgos, la explicabilidad y la privacidad. Se concluye que la prevención predictiva debe entenderse como una herramienta complementaria al conocimiento y criterio profesional, orientada a utilizar mejor la información disponible para actuar antes de que se produzca el daño.

PALABRAS CLAVE prevención predictiva; aprendizaje automático; analítica predictiva; predicción de accidentes; evaluación del riesgo; seguridad y salud laboral

DIGITALIZATION AND OCCUPATIONAL RISK PREVENTION

From reactive to predictive prevention: how data and AI are transforming occupational safety and health

ABSTRACT *The growing availability of data and advances in machine learning techniques are expanding the possibilities for occupational safety and health (OSH), enabling a shift from a predominantly reactive approach towards more anticipatory prevention. This article examines the role of predictive models in this transformation, explaining how they work and their main applications for anticipating accidents, estimating their severity, identifying periods of increased risk, and predicting other occupational health outcomes. Particular attention is paid to the transition from prediction to preventive action, emphasizing that the value of these models lies not only in their predictive capacity but also in their*

use for setting priorities and supporting decision-making. The article also discusses their main limitations, including data quality and representativeness, bias, explainability, and privacy. It concludes that predictive prevention should be understood as a tool that complements professional knowledge and judgement, aimed at making better use of available information to act before harm occurs.

KEYWORDS *predictive prevention; machine learning; predictive analytics; accident prediction; risk assessment; occupational safety and health*

Introducción. ¿Podemos anticiparnos al próximo accidente?

La prevención de riesgos laborales (PRL) se ha apoyado tradicionalmente en el análisis de lo que ya ha ocurrido. Los accidentes de trabajo, los incidentes, las enfermedades profesionales y otros daños relacionados con el trabajo proporcionan información para identificar sus causas y adoptar medidas destinadas a evitar que vuelvan a producirse. Este enfoque retrospectivo continúa siendo esencial, pero presenta una limitación evidente: buena parte del conocimiento preventivo se obtiene después de que el daño se haya producido.

Al mismo tiempo, las organizaciones disponen de una cantidad creciente de información sobre las condiciones de trabajo: registros de accidentes e incidentes, características de los puestos, evaluaciones de riesgos, exposiciones laborales, organización de los turnos o duración de las bajas son algunos ejemplos. El desarrollo de técnicas estadísticas avanzadas y de aprendizaje automático (*machine learning*) está ampliando la capacidad para analizar estos datos e identificar patrones complejos (Armenteros-Cosme *et al*., 2025), que pueden pasar inadvertidos mediante aproximaciones convencionales.

Surge así una pregunta especialmente relevante para la PRL: ¿podemos utilizar la información acumulada sobre el pasado para anticipar dónde, cuándo o bajo qué circunstancias es más probable que se produzca un daño? Esta es la lógica de los modelos predictivos. Diferentes investigaciones han mostrado su aplicación en la predicción de la ocurrencia y gravedad de los accidentes (Ait Lamkademe y Naddami, 2026), así como en otros resultados relacionados con la seguridad y salud laboral (Luo *et al*., 2023; Obasi y Benson, 2025).

Estas herramientas abren la posibilidad de avanzar desde una prevención fundamentalmente reactiva hacia otra más proactiva y predictiva. Sin embargo, predecir no significa adivinar. Los modelos proporcionan estimaciones probabilísticas cuya fiabilidad depende, entre otros factores, de la calidad y representatividad de los datos utilizados. Por tanto, su verdadero interés preventivo reside no solo en su capacidad para anticipar determinados resultados, sino en convertir esas predicciones en información útil para decidir dónde y cómo actuar.

Este artículo analiza qué son los modelos predictivos, sus principales aplicaciones en PRL y las oportunidades y limitaciones que plantea su utilización como herramienta de apoyo a la toma de decisiones preventivas.

1. De mirar al pasado a anticipar el riesgo

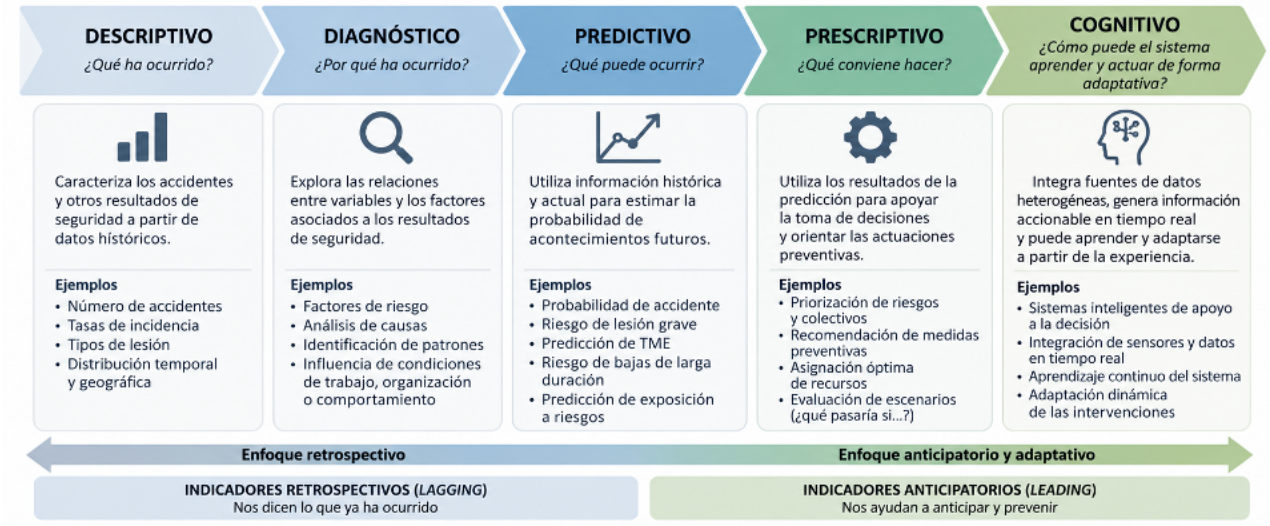
Históricamente, la PRL ha generado una gran cantidad de información. Accidentes e incidentes, índices de siniestralidad, investigaciones de accidentes, mediciones higiénicas, resultados de la vigilancia de la salud, absentismo o características de los puestos de trabajo constituyen fuentes de datos habituales en las organizaciones. Sin embargo, buena parte de esta información se ha utilizado tradicionalmente con una finalidad descriptiva o explicativa. Watkins *et al*. (2025) señalan que la evaluación del desempeño en seguridad se ha sustentado tradicionalmente en indicadores retrospectivos (*lagging indicators*), vinculados a acontecimientos ya ocurridos y, por tanto, de naturaleza esencialmente reactiva.

El desarrollo de nuevas capacidades de tratamiento de datos permite ampliar este enfoque. La revisión sistemática realizada por Maheronnaghsh *et al*. (2023) muestra que las técnicas de *machine learning* (ML) ya se utilizan en seguridad y salud laboral para diferentes tareas de clasificación, predicción y evaluación del riesgo. En la misma dirección, Armenteros-Cosme *et al*. (2025) constatan una expansión reciente de los modelos predictivos aplicados a la prevención

y destacan su capacidad para trabajar con múltiples variables e identificar patrones que pueden resultar difíciles de reconocer mediante aproximaciones analíticas convencionales.

Esta evolución, recogida en la figura 1, puede interpretarse a partir de diferentes niveles de utilización de los datos: descriptivo, diagnóstico, predictivo, prescriptivo y cognitivo (Ezerins *et al*., 2022). El análisis descriptivo responde fundamentalmente a la pregunta *¿qué ha ocurrido?* y permite caracterizar los accidentes y otros resultados de seguridad a partir de datos históricos. El análisis diagnóstico trata de avanzar hacia *¿por qué ha ocurrido?*, explorando las relaciones entre las variables y los factores asociados a esos resultados. El análisis predictivo modifica la perspectiva temporal y plantea: *¿qué puede ocurrir?*, utilizando información histórica y actual para estimar la probabilidad de acontecimientos futuros. El análisis prescriptivo traslada la pregunta hacia *¿qué conviene hacer?*, utilizando los resultados de la predicción para apoyar la toma de decisiones. Finalmente, el análisis cognitivo representa un nivel adicional de sofisticación, en el que sistemas capaces de integrar fuentes de datos heterogéneas generan información accionable en tiempo real y pueden aprender y adaptarse a partir de la experiencia.

Figura 1. Evolución del uso de los datos en prevención de riesgos laborales



Fuente: elaboración propia a partir de Ezerins et al. (2022)

En PRL, el salto entre los dos últimos niveles es especialmente relevante. Un modelo puede, por ejemplo, identificar qué situaciones presentan una mayor probabilidad de accidente o cuáles podrían asociarse con consecuencias más graves. Pero una predicción, por precisa que sea, no constituye por sí misma una intervención preventiva. Su valor aparece cuando permite priorizar recursos, intensificar controles o dirigir las actuaciones hacia aquellos escenarios en los que pueden resultar más necesarias (Huang *et al*., 2018). El planteamiento de estos autores resulta especialmente pertinente porque sitúa la predicción dentro de una secuencia que conduce desde el análisis de los datos hasta la decisión y, finalmente, la acción en seguridad.

Esta transición no implica sustituir los instrumentos tradicionales de la prevención. La investigación de accidentes, la evaluación de riesgos o el conocimiento experto siguen siendo necesarios para interpretar adecuadamente los resultados. Los modelos predictivos añaden una capacidad diferente: utilizar la experiencia acumulada en los datos para anticipar escenarios y apoyar la decisión sobre dónde y cuándo intervenir. Por tanto, la cuestión no es elegir entre prevención tradicional y prevención predictiva, sino integrar ambas aproximaciones en un proceso de toma de decisiones mejor informado.

2. ¿Cómo funcionan los modelos predictivos?

Los modelos predictivos utilizan datos disponibles para identificar patrones y estimar la probabilidad de que se produzca un determinado resultado. Su construcción requiere definir, en primer lugar, qué se quiere predecir (variable resultado) y qué información puede contribuir a esa predicción (variables predictoras). En PRL, estas últimas pueden proceder de fuentes muy diversas, como registros de accidentes e incidentes, características de los puestos y las tareas, condiciones ambientales o información generada por inspecciones y otras actividades preventivas. Las recientes revisiones muestran que las técnicas de *machine learning* permiten analizar conjuntamente múltiples variables y abordar diferentes problemas de clasificación y predicción en seguridad y salud laboral (Maheronnaghsh *et al*., 2023; Armenteros-Cosme *et al*., 2025).

La construcción de un modelo comprende diferentes etapas: preparación de los datos, entrenamiento, validación y predicción. En primer lugar, los registros disponibles deben transformarse en variables analizables y seleccionarse aquellas que pueden aportar información relevante. Posteriormente, durante el entrenamiento, el algoritmo aprende los patrones que relacionan esas variables con el resultado que se pretende anticipar. Finalmente, el modelo debe evaluarse utilizando datos que no hayan intervenido en su construcción para comprobar si mantiene su capacidad predictiva ante nuevas observaciones. La figura 2 sintetiza este proceso, desde los datos inicialmente disponibles hasta la obtención de una estimación de riesgo. En una aplicación sobre registros de accidentes de una planta siderúrgica, Sarkar *et al*. (2019) siguieron precisamente una secuencia de preprocesamiento, generación y selección de variables, entrenamiento, optimización y validación de diferentes modelos.

Figura 2. Proceso simplificado de construcción de un modelo predictivo en PRL



Fuente: elaboración propia a partir de Sarkar et al. (2019)

Como se aprecia en la columna «Datos» de la figura 2, a menudo, la fuente más rica de información en PRL no se encuentra solamente en bases de datos estructuradas, como los registros de accidentes e incidentes, sino también en texto libre, como las descripciones narrativas de los accidentes, las evaluaciones cualitativas o las notas de los inspectores. En este sentido, la irrupción del procesamiento de lenguaje natural (NLP, por sus siglas en inglés) y, más recientemente, de los grandes modelos de lenguaje (LLM, por sus siglas en inglés), permite extraer automáticamente variables predictoras a partir de estos textos o datos no estructurados. Al transformar orígenes de datos complejos en variables analizables, estas tecnologías enriquecen sustancialmente la capacidad del modelo para capturar matices que escaparían a una recolección de datos puramente tabular.

Además, no existe un único algoritmo adecuado para realizar predicciones. Pueden emplearse desde modelos paramétricos tradicionales, como la regresión logística o árboles de decisión, hasta técnicas de *ensemble learning* como Random Forest o Gradient Boosting (por ejemplo, XGBoost, LightGBM, CatBoost), así como arquitecturas de redes neuronales artificiales. Yapi *et al*. (2026) compararon diferentes familias de modelos para predecir accidentes laborales a

corto plazo a partir, entre otras variables, de información procedente de inspecciones de seguridad. Su análisis incluyó modelos estadísticos, árboles, métodos de *boosting* y diferentes arquitecturas de redes neuronales. Un resultado especialmente relevante es que una mayor complejidad no garantiza necesariamente una mejor capacidad predictiva, porque el rendimiento de los distintos algoritmos depende de las características de los datos y del problema analizado.

Por ello, la evaluación del modelo constituye una etapa fundamental. La exactitud global (*global precision*) puede resultar engañosa cuando el acontecimiento que se pretende anticipar es poco frecuente, como sucede habitualmente con los accidentes laborales (Topolska *et al*., 2026). Un modelo podría clasificar correctamente la mayoría de las situaciones simplemente porque en estas no ocurre ningún accidente y, sin embargo, detectar mal precisamente los casos de mayor interés preventivo. Por esta razón, el rendimiento se evalúa utilizando métricas robustas ante clases desbalanceadas, como la sensibilidad o *recall*, la precisión (*precision*), el valor F1 (*F1-score*) o la exactitud balanceada (*balanced accuracy*). Yapi *et al*. (2026) recurren a este tipo de métricas y a procedimientos de validación temporal para evaluar el comportamiento de los modelos ante datos posteriores a los utilizados durante su entrenamiento.

Finalmente, es importante precisar qué significa realmente predecir en este contexto. Un modelo predictivo no determina qué accidente ocurrirá ni cuándo sucederá exactamente. Habitualmente, estima una probabilidad o un nivel de riesgo que permite identificar situaciones o periodos que merecen mayor atención. Yapi *et al*. (2026) señalan, de hecho, que predecir la fecha exacta de un accidente rara vez resulta factible y plantean como objetivo más realista identificar intervalos temporales con un riesgo elevado. Por tanto, la predicción debe entenderse como una estimación probabilística basada en patrones presentes en los datos, no como una certeza sobre acontecimientos futuros.

3. ¿Qué podemos predecir en prevención de riesgos laborales?

La prevención predictiva puede orientarse a diferentes resultados dependiendo del problema que se quiera abordar. El más inmediato es la ocurrencia de accidentes e incidentes, estimando la probabilidad de que se produzcan en determinadas situaciones, actividades o periodos. En este ámbito, también los *near misses* pueden aportar información relevante al actuar como señales precursoras de acontecimientos posteriores (Sarkar *et al*., 2019; Topolska *et al*., 2026; Yapi *et al*., 2026). Un segundo ámbito es la gravedad de los accidentes. Anticipar sus posibles consecuencias permite distinguir situaciones que no necesariamente presentan una mayor probabilidad de ocurrencia, pero que podrían generar daños especialmente importantes (Luo *et al*., 2023; Obasi y Benson, 2025; Li *et al*., 2025). Desde el punto de vista preventivo, esta distinción es relevante, porque probabilidad y gravedad representan dimensiones diferentes del riesgo.

La predicción puede incorporar también una dimensión temporal. Más que intentar determinar el momento exacto en el que se producirá un accidente, el objetivo puede ser identificar periodos en los que su probabilidad aumenta. Este planteamiento permite detectar variaciones del riesgo a corto plazo y resulta especialmente interesante para actividades en las que las condiciones de trabajo cambian con frecuencia (Yapi *et al*., 2026). La predicción se convierte así en una herramienta para reconocer ventanas temporales en las que puede ser necesario intensificar la atención preventiva.

El campo de aplicación es todavía más amplio cuando se consideran otros resultados relacionados con la salud laboral. Los modelos predictivos pueden orientarse, por ejemplo, hacia trastornos musculoesqueléticos y riesgos ergonómicos (Chan *et al*., 2022), absentismo y bajas potencialmente prolongadas (Llamas Blázquez, 2025), así como hacia otros riesgos relacionados con la salud, como la fatiga o la exposición a agentes peligrosos (Armenteros-Cosme *et al*., 2025). La literatura recoge aplicaciones en contextos muy diversos, como la construcción (Luo *et al*., 2023; Topolska *et al*., 2026; Umar y Haque, 2026), las industrias extractivas (Li *et al*., 2025), la industria manufacturera (Sarkar *et al*., 2019), la agricultura (Mermer *et al*., 2026), el sector sanitario (Alcaide *et al*., 2026a) y la gestión de residuos (Yapi *et al*., 2026), entre otros.

Esta diversidad refuerza el carácter transversal de la prevención predictiva: los resultados que interesa anticipar y las variables disponibles cambian entre actividades, pero la lógica predictiva puede aplicarse a problemas preventivos muy diferentes. En conjunto, permite anticipar si puede producirse un accidente, estimar sus posibles consecuencias, detectar cuándo aumenta el riesgo o identificar la probabilidad de determinados daños para la salud. Sin embargo, conocer anticipadamente el riesgo solo adquiere verdadero valor preventivo cuando esa información se traduce en una decisión. La cuestión pasa entonces de qué podemos predecir a qué hacemos con esa predicción.

4. De la predicción a la prevención: convertir los datos en decisiones

La capacidad de anticipar un resultado no constituye por sí misma una actuación preventiva. Un modelo puede identificar situaciones con mayor probabilidad de accidente, periodos de riesgo elevado o escenarios asociados con consecuencias más graves, pero la predicción solo adquiere valor cuando conduce a una decisión y, finalmente, a una intervención. El potencial de los datos reside precisamente en su capacidad para generar información útil para decidir y actuar, y no únicamente en mejorar el conocimiento sobre los riesgos (Huang *et al*., 2018).

En la práctica, las predicciones pueden utilizarse para establecer prioridades preventivas. Identificar determinadas tareas, condiciones o periodos con un nivel de riesgo superior permite concentrar en estos la atención y los recursos disponibles. Esto puede traducirse en priorizar inspecciones, revisar condiciones de trabajo, reforzar controles o intensificar la supervisión cuando resulte necesario. De este modo, la predicción introduce una dimensión anticipatoria en la toma de decisiones: no se trata de esperar a que se produzca el daño para actuar, sino de utilizar la información disponible para orientar la intervención antes de que ocurra (Yapi *et al*., 2026).

Sin embargo, identificar dónde o cuándo existe un mayor riesgo no determina automáticamente qué medida preventiva debe adoptarse. Por tanto, conviene distinguir dos decisiones complementarias: la predicción puede ayudar a determinar dónde y cuándo actuar, mientras que la evidencia científica puede contribuir a decidir cómo actuar. La elección entre medidas técnicas, organizativas, formativas, ergonómicas o de otro tipo requiere considerar el conocimiento disponible, las características específicas del contexto y el criterio de los profesionales de la prevención. Los modelos predictivos deben entenderse como herramientas de apoyo a la decisión, integradas en el proceso preventivo y no como mecanismos autónomos de decisión.

Este paso de la predicción a la intervención constituye todavía uno de los principales retos de la prevención predictiva. La literatura muestra que el desarrollo de herramientas para identificar y anticipar riesgos ha avanzado más rápidamente que la evidencia disponible sobre su implementación efectiva en los procesos preventivos (Chan *et al*., 2022; Jetha *et al*., 2025). Por ello, una elevada capacidad predictiva no puede confundirse con una elevada efectividad preventiva. Un modelo puede identificar correctamente situaciones de mayor riesgo y, sin embargo, no producir ninguna mejora si esa información no genera una actuación adecuada o si las medidas adoptadas no resultan efectivas.

El proceso tampoco debería finalizar con la intervención. Las medidas adoptadas deben ser posteriormente evaluadas, para comprobar si han reducido el riesgo o mejorado los resultados de seguridad y salud (Jetha *et al*., 2025). La prevención predictiva requiere integrarse en un ciclo de vida continuo (habitualmente gestionado mediante prácticas de MLOps): ingesta de datos →entrenamiento del modelo →predicción →priorización →intervención →evaluación →monitorización del modelo y reentrenamiento frente a variaciones en el entorno laboral (*concept drift*). Además, desde un punto de vista técnico, este ciclo no debe considerarse estático. Los entornos laborales son dinámicos: la introducción de nueva maquinaria, la rotación de plantillas o los cambios organizativos alteran progresivamente las condiciones de trabajo subyacentes. Cuando esto ocurre, los algoritmos sufren lo que se conoce como *deriva de concepto* (*concept drift*), provocando que sus predicciones pierdan fiabilidad con el paso del tiempo. Por ello, la implementación real de estos sistemas exige adoptar un enfoque de operaciones de *machine learning* (MLOps), que garantice la monitorización continua del rendimiento del modelo en producción y su reentrenamiento periódico para adaptar sus inferencias a la realidad cambiante de la empresa.

Los resultados obtenidos por este ciclo no estático generan, a su vez, nueva información que puede contribuir a mejorar futuras decisiones y favorecer el aprendizaje preventivo de la organización. El objetivo último no es disponer de algoritmos cada vez más sofisticados *per se*, sino utilizar la información que generan para intervenir mejor y hacerlo antes de que se produzca el daño.

5. Las limitaciones de predecir: datos, sesgos y explicabilidad

La capacidad de un modelo predictivo depende, en gran medida, de la calidad de los datos con los que ha sido construido. Registros incompletos, errores de clasificación, diferencias en los criterios de notificación o información poco representativa pueden afectar directamente a las predicciones. Además, en PRL, muchos de los datos disponibles

proceden de accidentes, incidentes, inspecciones o evaluaciones recogidos originalmente con otras finalidades. Por ello, disponer de un gran volumen de información no garantiza necesariamente disponer de datos adecuados para predecir: su calidad, consistencia y representatividad resultan determinantes (Ezerins *et al*., 2022; Baraza y Torrent-Sellens, 2026).

Los modelos también pueden reproducir o amplificar los sesgos presentes en los datos históricos. Determinados accidentes, trabajadores o situaciones pueden estar infrarrepresentados, porque los acontecimientos graves suelen ser mucho menos frecuentes que los leves. Este desequilibrio resulta especialmente relevante en PRL: un modelo puede presentar una elevada exactitud global y, al mismo tiempo, tener dificultades precisamente para identificar los acontecimientos menos frecuentes y potencialmente más críticos (Luo *et al*., 2023; Li *et al*., 2025). Para mitigar este problema matemático de clases desbalanceadas, la ingeniería aplica estrategias específicas durante la fase de entrenamiento. Entre estas destacan las aproximaciones algorítmicas, como el aprendizaje sensible al coste (*cost-sensitive learning*), que penaliza drásticamente al algoritmo cuando falla al predecir un accidente grave, o las técnicas de remuestreo de datos, como la generación de datos sintéticos de la clase minoritaria (por ejemplo, mediante el algoritmo SMOTE). La aplicación de estas técnicas es un paso ineludible para garantizar que el modelo no normalice el riesgo ignorando los escenarios de mayor criticidad. Por tanto, la evaluación debe considerar no solo el rendimiento general del modelo, sino también qué casos identifica correctamente y cuáles deja sin detectar.

Una tercera cuestión es la explicabilidad. Una predicción resulta preventivamente más útil cuando permite comprender qué variables han contribuido a identificar una situación como de mayor riesgo. Para resolver este problema, las técnicas de inteligencia artificial explicable (*Explainable Artificial Intelligence*, XAI) abren la «caja negra» de los algoritmos. El uso de métodos matemáticos de atribución local, como SHAP (*Shapley Additive exPlanations*) o LIME, permite a un técnico de PRL entender por qué el modelo ha tomado una decisión concreta. Con estas herramientas, el sistema no se limita a emitir una alerta ciega (por ejemplo, «Riesgo de incidente del 80 %»), sino que detalla su justificación probabilística: «Riesgo alto debido a: variable A (acumulación de horas extras en la última semana) y variable B (turno nocturno actual)». Esta transparencia algorítmica es la que verdaderamente facilita la interpretación clínica del dato y favorece la confianza y utilización de los resultados por parte de los profesionales (Armenteros-Cosme *et al*., 2025; Obasi y Benson, 2025). La transparencia y la explicabilidad adquieren especial importancia cuando los modelos incorporan información individual, porque aparecen también cuestiones relacionadas con la privacidad, la equidad y el posible uso inadecuado de los datos (Baraza y Torrent-Sellens, 2026; EU-OSHA, 2022).

Estas limitaciones no invalidan el potencial de la prevención predictiva, pero sí condicionan su utilización. La calidad de los datos, el control de los sesgos, la explicabilidad y la protección de la información deben considerarse juntamente con la capacidad predictiva del modelo. Sobre todo, una estimación algorítmica no debería interpretarse como una decisión automática: requiere conocimiento del contexto, valoración preventiva y supervisión humana (Baraza y Torrent-Sellens, 2026; EU-OSHA, 2022). Por tanto, el reto no consiste en sustituir el criterio profesional por un algoritmo, sino en determinar cómo pueden combinarse ambos para mejorar la toma de decisiones. De hecho, la investigación sobre la eficiencia predictiva ya ha señalado cuál es la mejor combinación de datos y conocimientos para optimizar esta relación. Esta nos sugiere que apliquemos la complementariedad de inteligencias (entre inteligencia artificial, IA, e inteligencia humana, IH) en función de las distintas categorías de la predicción y los componentes de la toma de decisiones (Torrent, 2025). En la tabla 1 se reproduce este contexto. Es decir, se analizan los puntos fuertes y débiles de la IA y la IH en función de las características de los datos y los conocimientos en la toma de decisiones sobre predicción en una organización económica. En cuanto a la predicción, la IA optimiza sus capacidades en contextos de conocimientos conocidos, es decir, en situaciones de predicción con fenómenos conocidos y abundantes datos. Por su parte, la IH es mucho más verosímil que la IA en contextos de desconocimientos conocidos, es decir, en situaciones de predicción de fenómenos poco o nada conocidos y con pocos datos. En estos contextos, algunas de las habilidades de la IH, como nuestras capacidades innatas e intuitivas para reconocernos, compararnos o adaptarnos a nuevas circunstancias, son de gran utilidad para la previsión.

Tabla 1. La matriz de confusión para la toma de decisiones en IA + IH

		Conocimientos	
		Conocidos	**Desconocidos**
Datos	**Abundantes**	**Conocimientos conocidos** (muchos datos de un tema conocido) **IA**: fraude, diagnósticos médicos, sentencias judiciales	**Conocimientos desconocidos** (muchos datos de un tema desconocido) **IA falla**: causalidad inversa, variables omitidas **IA + IH**: análisis contrafactual y explicabilidad de IA
	Escasos	**Desconocimientos conocidos** (pocos datos sobre un tema conocido; predicciones con pocos datos) **IH**: reconocimientos, analogías, intuición	**Desconocimientos desconocidos** (pocos datos sobre un acontecimiento que no ha sucedido nunca) **Ni IA, ni IH**: creación creativa, innovaciones radicales inesperadas

Fuente: Torrent-Sellens (2025)

6. El futuro de la PRL: algoritmos para apoyar, no para sustituir

La incorporación de modelos predictivos a la PRL no debería plantearse como una sustitución de los profesionales, sino como una ampliación de su capacidad para interpretar información y anticipar riesgos. Los algoritmos pueden procesar grandes volúmenes de datos, identificar patrones difíciles de reconocer mediante aproximaciones convencionales y generar estimaciones que ayuden a establecer prioridades. Sin embargo, transformar esas estimaciones en decisiones preventivas requiere comprender las condiciones reales de trabajo, valorar sus implicaciones y seleccionar las actuaciones más adecuadas. Por tanto, el valor de estas herramientas reside en combinar la capacidad analítica de los modelos con el conocimiento y el criterio profesional (Baraza y Torrent-Sellens, 2026).

Esta complementariedad puede favorecer una prevención más dinámica y anticipatoria. Frente a indicadores que describen acontecimientos ya ocurridos, la disponibilidad creciente de datos permite incorporar señales capaces de advertir cambios en las condiciones de riesgo (Watkins *et al*., 2025). La progresiva aplicación del *machine learning* en seguridad y salud laboral muestra las posibilidades de esta evolución (Maheronnaghsh *et al*., 2023). El objetivo no debería ser aspirar a predecir todos los accidentes, sino mejorar progresivamente la capacidad para reconocer situaciones que requieren atención y actuar antes de que el daño se materialice.

Este escenario también implica una evolución de las competencias de los profesionales de la prevención (Alcaide *et al*., 2026b). No será necesario que todos se conviertan en especialistas en *machine learning*, pero sí que puedan comprender qué información utiliza un modelo, interpretar adecuadamente sus resultados y reconocer sus limitaciones. La alfabetización en datos y una comprensión básica de los modelos predictivos serán necesarias para utilizarlos de manera crítica y convertir sus resultados en conocimiento preventivo útil (Alcaide *et al*., 2026b). De este modo, la tecnología modifica no solo las herramientas disponibles, sino también las capacidades necesarias para utilizarlas adecuadamente.

En definitiva, el futuro de la prevención predictiva dependerá menos de disponer de algoritmos cada vez más complejos que de conseguir integrarlos adecuadamente en la gestión preventiva. La tecnología puede contribuir a detectar patrones, anticipar escenarios y orientar prioridades; el profesional aporta el conocimiento del contexto necesario para interpretar esa información y transformarla en acción. La prevención predictiva no supone sustituir una forma de hacer prevención por otra, sino incorporar una nueva capacidad para anticiparse al riesgo y tomar decisiones mejor informadas.

Conclusiones

Los modelos predictivos amplían las posibilidades de la PRL al permitir utilizar los datos disponibles no solo para comprender lo ocurrido, sino también para anticipar accidentes, estimar sus consecuencias e identificar situaciones que

requieren mayor atención preventiva. Su aplicación en diferentes resultados y sectores muestra el potencial de avanzar hacia una prevención más anticipatoria.

Sin embargo, predecir no equivale a prevenir. El valor de estos modelos depende de que sus resultados se traduzcan en prioridades, decisiones e intervenciones efectivas, y de que posteriormente se evalúe su impacto. Al mismo tiempo, la calidad de los datos, los sesgos, la explicabilidad y la privacidad condicionan su utilización responsable.

El futuro de la prevención predictiva no pasa por sustituir el criterio profesional, sino por combinar la capacidad analítica de los algoritmos con el conocimiento preventivo. El objetivo final no es predecir más, sino utilizar mejor la información disponible para actuar antes de que se produzca el daño. Tener en cuenta las características de los datos y los acontecimientos a predecir es un buen punto de partida para ejercer la complementariedad de inteligencias en PRL.

Referencias bibliográficas

AIT LAMKADEME, Hajar; NADDAMI, Ahmed (2026). «Predictive analysis of causal factors influencing occupational accidents in construction workplaces using a machine learning unified data model». *International Journal of System Assurance Engineering and Management*, vol. 17, págs. 2140-2159. DOI: https://doi.org/10.1007/s13198-025-03064-0

ALCAIDE, Neus; BARAZA, Xavier; COMBE, Gemma; CORRONS, August; FICAPAL-CUSÍ, Pilar; NOS, Isabel. (2026a). «Predicting musculoskeletal sick leave in tertiary hospital staff using classification trees for targeted ergonomic intervention». *Work: A Journal of Prevention, Assessment and Rehabilitation*. DOI: https://doi.org/10.1177/10519815261489557

ALCAIDE, Neus; BARAZA, Xavier; FICAPAL-CUSÍ, Pilar (2026b). «From technical skills to future competencies: training and development of occupational health and safety professionals». *International Journal of Training and Development*. DOI: https://doi.org/10.1111/ijtd.70037

ARMENTEROS-COSME, Pablo; ARIAS-GONZÁLEZ, Marcos; ALONSO-ROLLÁN, Sergio; MÁRQUEZ-SÁNCHEZ, Sergio; CARRERA, Albano (2025). «Advancements in artificial intelligence and machine learning for occupational risk prevention: A systematic review on predictive risk modeling and prevention strategies». *Sensors*, vol. 25, n.º 17, art. 5419. DOI: https://doi.org/10.3390/s25175419

BARAZA, Xavier; TORRENT-SELLENS, Joan (2026). «Artificial intelligence and emerging risks in occupational safety and health». *Encyclopedia*, vol. 6, n.º 1, art. 25. DOI: https://doi.org/10.3390/encyclopedia6010025

CHAN, Victor C. H.; ROSS, Gwyneth B.; CLOUTHIER, Allison L.; FISCHER, Steven L.; GRAHAM, Ryan B. (2022). «The role of machine learning in the primary prevention of work-related musculoskeletal disorders: A scoping review». *Applied Ergonomics*, vol. 98, art. 103574. DOI: https://doi.org/10.1016/j.apergo.2021.103574

EUROPEAN AGENCY FOR SAFETY AND HEALTH AT WORK (EU-OSHA) (2022). *Artificial intelligence for worker management: implications for occupational safety and health*. Luxemburgo: Publications Office of the European Union. DOI: https://doi.org/10.2802/76354

EZERINS, Maira E.; LUDWIG, Timothy D.; O'NEIL, Tara; FOREMAN, Anne M.; AÇIKGÖZ, Yalçin (2022). «Advancing safety analytics: A diagnostic framework for assessing system readiness within occupational safety and health». *Safety Science*, vol. 146, art. 105569. DOI: https://doi.org/10.1016/j.ssci.2021.105569

HUANG, Lang; WU, Chao; WANG, Bing; OUYANG, Qiumei (2018). «Big-data-driven safety decision-making: A conceptual framework and its influencing factors». *Safety Science*, vol. 109, págs. 46-56. DOI: https://doi.org/10.1016/j.ssci.2018.05.012

JETHA, Arif; BAKHTARI, Hela; IRVIN, Emma; BISWAS, Aviroop; SMITH, Maxwell J.; MUSTARD, Cameron; ARRANDALE, Victoria H.; DENNERLEIN, Jack T.; SMITH, Peter M. (2025). «Do occupational health and safety tools that utilize artificial intelligence have a measurable impact on worker injury or illness? Findings from a systematic review». *Systematic Reviews*, vol. 14, art. 146. DOI: https://doi.org/10.1186/s13643-025-02869-1

LI, Yang; SANMIQUEL, Lluis; ZHANG, Zhengxin; ZHAO, Guoyan; BASCOMPTA, Marc (2025). «Discovering the underground coal mining accident patterns in Spain from 2003 to 2021: Insights through machine learning techniques». *Safety Science*, vol. 181, art. 106677. DOI: https://doi.org/10.1016/j.ssci.2024.106677

LLAMAS BLÁZQUEZ, Pablo (2025). «Predicting workplace absenteeism using machine learning: A pilot study in occupational health». *Journal of Occupational Medicine and Toxicology*, vol. 20, art. 38. DOI: https://doi.org/10.1186/s12995-025-00482-5

LUO, Xixi; LI, Xinchun; GOH, Yang Miang; SONG, Xuefeng; LIU, Quanlong (2023). «Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents». *Safety Science*, vol. 163, art. 106138. DOI: https://doi.org/10.1016/j.ssci.2023.106138

MAHERONNAGHSH, Sara; ZOLFAGHARNASAB, Hooshiar; GORGICH, Malihe; DUARTE, J. (2023). «Machine learning in Occupational Safety and Health – A systematic review». *International Journal of Occupational and Environmental Safety*, vol. 7, n.º 1, págs. 14-32. DOI: https://doi.org/10.24840/2184-0954_007-001_001586

MERMER, Omer; ZHANG, Eddie; DEMIR, Ibrahim (2026). «Explainable ensemble machine learning for predicting injury severity in agricultural accidents». *Journal of Agromedicine*, págs. 1-15. DOI: https://doi.org/10.1080/1059924X.2026.2658048

OBASI, Izuchukwu Chukwuma; BENSON, Chizubem (2025). «An explainable machine learning framework for predicting injury severity in extractive industry accidents». *Results in Engineering*, vol. 28, art. 107552. DOI: https://doi.org/10.1016/j.rineng.2025.107552

SARKAR, Sobhan; VINAY, Sammangi; RAJ, Rahul; MAITI, J.; MITRA, Pabitra (2019). «Application of optimized machine learning techniques for prediction of occupational accidents». *Computers & Operations Research*, vol. 106, págs. 210-224. DOI: https://doi.org/10.1016/j.cor.2018.02.021

TOPOLSKA, Katarzyna; WOŹNIAK, Zuzanna; HOŁA, Bożena (2026). «Application of machine learning and deep learning methods for the prediction of near misses and occupational accidents in the construction industry». *Archives of Civil and Mechanical Engineering*, vol. 26, art. 244. DOI: https://doi.org/10.1007/s43452-026-01606-2

TORRENT-SELLENS, Joan (2025). *Inteligencia artificial y trabajo: hacia la complementariedad de inteligencias*. Universitat Oberta de Catalunya.

UMAR, Mubarak; HAQUE, Faizanul (2026). «Health and safety in the construction industry: A scoping review of challenges and emerging technologies». *Discover Civil Engineering*, vol. 3, art. 175. DOI: https://doi.org/10.1007/s44290-026-00563-2

WATKINS, Deborah; BISHOP, Emma; NAYLOR, Steven; FRANKISH, Ruth; STAVES, Malcolm; DONY, John; MILLER, Paul; MCCOOL, Rachael; FERRANTE DI RUFFANO, Lavinia (2025). «A scoping review of the evidence base for the performance of leading indicators for improving safety outcomes: Available evidence, implications for practice and future directions». *Journal of Safety Research*, vol. 95, págs. 530-544. DOI: https://doi.org/10.1016/j.jsr.2025.10.024

YAPI, Aho; LATOUCHE, Pierre; GUILLIN, Arnaud; BAILLY, Yan (2026). «A new machine learning framework for occupational accidents forecasting with safety inspections integration». *Journal of Safety Research*, vol. 97, págs. 244-260. DOI: https://doi.org/10.1016/j.jsr.2026.03.004

Cita recomendada: BARAZA, Xavier; TORRENT-SELLENS, Joan; MIR-TUTUSAUS, Josep-Anton. «De la prevención reactiva a la prevención predictiva: cómo los datos y la IA están transformando la seguridad y salud laboral». *Oikonomics* [en línea]. Noviembre 2026, n.º 27. ISSN 2339-9546. DOI: https://doi.org/10.7238/o.n27.2611



Xavier Baraza
jbaraza@uoc.edu
Director de los Estudios de Economía y Empresa (UOC)

Profesor agregado de Seguridad y Salud Laboral y director de los Estudios de Economía y Empresa de la Universitat Oberta de Catalunya. Es investigador del grupo de investigación interdisciplinaria i2TIC-IA Lab (UOC-DIGIT) y centra su investigación en el análisis de la macroaccidentabilidad en diversos sectores de actividad, la evolución de las competencias de los profesionales de la seguridad y salud laboral y la modelización de accidentes químicos industriales. Ingeniero Químico (UB, 1999), doctor en Ingeniería Química (UB, 2002) y máster en Química Experimental (UB, 2000), Prevención de Riesgos Laborales (UPC, 2003) y Dirección y Administración de Empresas (EEN, 2007).



Joan Torrent-Sellens
jtorrent@uoc.edu
Director del Centro de Investigación para la Transformación Digital y la Gobernanza (UOC)

Catedrático de Economía, director del Centro de Investigación en Transformación Digital y Gobernanza (UOC-DIGIT) y coinvestigador principal del grupo de investigación interdisciplinaria en TIC y Laboratori de IA (i2TIC-IA Lab). Su investigación se centra en la economía política de la inteligencia artificial y sus efectos económicos, empresariales, laborales, de bienestar y salud y de sostenibilidad de los procesos de transformación y transición digital. Licenciado en Ciencias Económicas y Empresariales (UAB, 1994), máster en Análisis Económicos (UAB, 2001) y doctor en Economía (UOC, 2002).



Josep-Anton Mir-Tutusaus
jmirtu@uoc.edu
Profesor de los Estudios de Economía y Empresa (UOC)

Profesor lector de *Big Data* y *Business Intelligence* de los Estudios de Economía y Empresa de la Universitat Oberta de Catalunya. Es investigador del grupo TURBA Lab (UOC-TRÀNSIC) y centra su investigación en la intersección entre las ciencias ambientales, las ciencias sociales y la inteligencia artificial mediante el análisis cuantitativo y el *machine learning*. Su trabajo busca desmontar el discurso «inmaterial» de la tecnología, analizando el coste físico (huella hídrica, huella de carbono y extracción de materiales), así como el coste social y laboral de la transformación que produce la inteligencia artificial. Doctor en Ciencia y Tecnología Ambientales (ICTA-UAB, 2017), licenciado en Biotecnología (UAB, 2012) y máster en Bioinformática y Bioestadística (UOC-UB, 2023).

Los textos publicados en esta revista están sujetos –si no se indica lo contrario– a una licencia de Reconocimiento 4.0 Internacional de Creative Commons. Puede copiarlos, distribuirlos, comunicarlos públicamente, hacer obras derivadas siempre que reconozca los créditos de las obras (autoría, nombre de la revista, institución editora) de la manera especificada por los autores o por la revista. La licencia completa se puede consultar en https://creativecommons.org/licenses/by/4.0/deed.es_ES.



ODS

