Interconectando Saberes

ISSN: 2448-8704
Vol. 11, Núm. 22
julio-diciembre 2026
DOI: https://doi.org/10.25009/is.v11i22.3042

Universidad Veracruzana – Instituto de Investigaciones y Estudios Superiores Económicos y Sociales (IIESES) http://is.uv.mx

CÓMO CITAR / HOW TO CITE:

Manzano Silva, O. (2026). Inteligencia artificial para el análisis predictivo del delito: Estado del arte y desafíos para la seguridad pública. *Interconectando Saberes*, *11*(22), 299-312. https://doi.org/10.25009/is.v11i22.3042

RECIBIDO:
12 de junio de 2026

ACEPTADO:
23 de octubre de 2026

PUBLICADO:
20 de noviembre de 2026

LICENCIA

Esta obra está bajo una licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional (CC BY-NC-ND 4.0)

Inteligencia artificial para el análisis predictivo del delito:
Estado del arte y desafíos para la seguridad pública

Omar Manzano Silva
Consultor de Seguridad Pública, México

Resumen

El artículo revisa estudios recientes sobre el uso de inteligencia artificial (IA) en la predicción del delito y sus aportes potenciales a la seguridad pública. Se analizaron investigaciones empíricas publicadas entre 2019 y 2025 que aplican IA, aprendizaje automático o aprendizaje profundo al análisis delictivo. Los estudios, realizados en países de América y Asia, emplean métodos supervisados y no supervisados, redes neuronales, análisis espacio-temporal y visión por computadora. La literatura reporta aplicaciones para identificar patrones y zonas de riesgo, estimar tendencias delictivas y apoyar la asignación de recursos. Sin embargo, los enfoques y resultados son heterogéneos, y la mayoría de los modelos requiere validación en contextos reales. Se concluye que su incorporación a la seguridad pública debe considerar la calidad y representatividad de los datos, la interpretabilidad, los sesgos y la privacidad, y evaluar su impacto en las políticas públicas.

Palabras clave: Inteligencia artificial, Aprendizaje automático, Aprendizaje profundo, Predicción del delito, Seguridad pública.

Abstract

This article reviews recent studies on the use of artificial intelligence (AI) for crime prediction and its potential contributions to public security. It analyzes empirical research published between 2019 and 2025 that applies AI, machine learning, or deep learning to crime analysis. Conducted in countries across the Americas and Asia, the studies use supervised and unsupervised methods, neural networks, spatiotemporal analysis, and computer vision. The literature reports applications for identifying patterns and risk areas, estimating crime trends, and supporting resource allocation. However, approaches and findings are heterogeneous, and most models require validation in real-world settings. The review concludes that integrating these tools into public security requires attention to data quality and representativeness, interpretability, bias, and privacy, as well as evaluation of their impact on public policies.

Keywords: Artificial intelligence, Machine learning, Deep learning, Crime prediction, Public security.

INTRODUCCIÓN

El objetivo de esta etapa de la investigación en curso es identificar, clasificar y analizar estudios recientes sobre análisis delictivo con IA para distinguir los modelos y técnicas empleados y señalar de entre ellos los avances, limitaciones y brechas existentes actualmente en la implementación de este tipo de tecnología como parte del diseño de políticas públicas de seguridad.

La amplia utilización de la Inteligencia Artificial (IA) y el aprendizaje automático (*Machine Leraning* ML) por sus innegables ventajas en precisión, rapidez y economía de análisis de datos ha sido ampliamente empleada en el ámbito empresarial primero y en el sector público después, en lo referente a este último, su uso en estudios de criminología y seguridad pública ha ido cobrando importancia dada la necesidad de tener eficaces modelos predictivos que soporten la toma de decisiones.

El presente documento presenta parte de una investigación en curso que tiene como objetivo diseñar y proponer un modelo de análisis predictivo que, basado en inteligencia artificial, esté orientado a la prevención de incidencias de robo de vehículos, con el propósito de fortalecer el proceso de toma de decisiones de la alta dirección en las instituciones estatales de seguridad pública. Al respecto este documento presenta la realización de una búsqueda sistemática para la generación del estado del arte correspondiente en la aplicación de IA/ML/DL en análisis delictivo.

MARCO TEÓRICO

Para el presente trabajo se ha realizado y se presenta una breve revisión de enfoques clásicos de criminología predictiva (criminología ambiental, teorías de oportunidad, análisis espacio-temporal), así como una introducción de la IA y el aprendizaje automático como herramientas disruptivas que pueden mejorar la operatividad del combate a la delincuencia.

Teorías clásicas de criminología

La criminología ambiental se centra en cómo el entorno físico y social influye en la ocurrencia del delito, por tanto, realiza un análisis del mismo. La llamada Escuela de Chicago y la teoría de la desorganización social son antecedentes clave. En este sentido se sugiere que aquellos barrios con instituciones débiles y cohesión social baja tienden a generar más criminalidad. Con esta teoría como base pueden analizarse zonas específicas de las ciudades, con desorganización social percibida, para analizar en ellas patrones de comportamiento delictivo. Asociado a esto se encuentran los preceptos de criminología ambiental que analiza factores de diseño y mobiliario urbano como iluminación, movilidad y concentración de actividades para identificar áreas de riesgo.

Las denominadas teorías de oportunidad plantean que la existencia del delito ocurre cuando se presentan circunstancias favorables para el infractor, esto es cuando confluyen en el tiempo ciertas variables percibidas como favorables a los ojos del infractor, ya que según la de la elección racional los individuos evalúan los costos y los beneficios antes de delinquir.

Con estos antecedentes surgen los modelos como el propuesto por (Cohen y Felson) de actividades rutinarias. En el modelo de actividades rutinarias los autores explican que el crimen surge cuando coinciden tres elementos: a) ausencia de guardianes eficaces (instituciones débiles/circunstancias favorables), b) un delincuente motivado (cuya evaluación resulta en mayores beneficios que costos) y, c) una víctima u objetivo adecuado.

Finalmente, en las teorías de criminología predictiva, las relativas al análisis espacio-temporal examinan la distribución geográfica y temporal de los delitos, en ellas las principales herramientas de análisis son los mapas delictivos, estadísticas y técnicas predictivas para identificar patrones de concentración (hot spots), estos análisis permiten, según este modelo anticipar dónde y cuándo es más probable que ocurran delitos, facilitando la prevención policial y comunitaria.

Los enfoques clásicos presentados coinciden en la idea de que el crimen es producto de características individuales y, de la interacción entre personas, oportunidades y contextos sociales/espaciales. Tomando en consideración que la criminología predictiva usa estas teorías en el diseño de políticas públicas más efectivas, debe evaluarse la elaboración de estas tomando en consideración la prevención comunitaria, el urbanismo seguro y las estrategias policiales focalizadas según resulte conveniente, se pretende además que esto pueda definirse en cada caso particular.

Con todo esto sobre la mesa, el uso de la Inteligencia Artificial y el denominado aprendizaje automático, muy ampliamente explotado en entornos comerciales y de negocios por sus resultados altamente eficientes, se vislumbran como poderosas herramientas que permitan la el análisis y definición de patrones delictivos precisos, y a su vez, con ellos la elaboración de estrategias de combate, mejor aún de previsión del delito; todo lo anterior a partir de la explotación de bases de datos existentes.

Bases de datos y minería de datos

Los sistemas de bases de datos se utilizan para administrar colecciones de datos: a) muy valiosos, b) relativamente grandes y c) accedidos por múltiples usuarios y aplicaciones, a menudo al mismo tiempo.

Las bases de datos tradicionales cumplen la función de almacenar datos que se consideran valiosos para la gestión operativa, en general son grandes en tamaño y accesibles desde aplicaciones creadas para su alimentación, procesamiento y consulta; soportan la operación diaria de los procesos de cómputo y permiten la extracción de sus datos para el análisis de los mismos.

Los modelos básicos de bases de datos son: bases de datos relacionales (RDBMS) para datos normalizados y no relacionales (NoSQL) para permitir la escalabilidad, flexibilidad en datos heterogéneos. Actualmente se han desarrollados programas para el procesamiento de masivas cantidades de datos no estructurados, mediante mecanismos de distribución del trabajo entre procesadores.

Las RDBMS son tradicionales, tienen un buen soporte transaccional, SQL, integridad comprobada y modelos normalizados; actualmente siguen sosteniendo la operación principal en muchas empresas e instituciones del sector público. Las bases de datos NoSQL tienen ventajas en velocidad, escalado horizontal, tolerancia a fallos y formatos flexibles y son cada vez más utilizadas debido a las ventajas listadas previamente.

Las propuestas actuales de explotación de bases de datos son arquitecturas modernas que combinan tanto RDBMS como NoSQL para determinar los datos existentes y su ubicación como un aspecto crítico para proyectos de Big Data y para alimentar modelos de

aprendizaje automático (*Machine Learning*) y de aprendizaje profundo (*Deep Learning*).

En este orden de ideas la minería de datos y el Descubrimiento de Conocimiento en Bases de Datos. (KDD por sus siglas en inglés -*Knowledge Discovery in Databases*-) como proceso para el descubrimiento de patrones, asociaciones e identificación de estructuras significativas así somo anomalías en los conjuntos de datos y en los conjuntos de datos masivos (*big data*) se ha desarrollado para la explotación de conjuntos de datos (donde y como estén almacenados).

Las etapas de la minería de datos implican: a) determinación de objetivos (qué se busca y para qué), b) el llamado preprocesamiento (comprende procesos de selección, limpieza, enriquecimiento y reducción de datos), c) definición de un modelo de análisis y, finalmente, d) el análisis de resultados que implica la validación, e interpretación de lo obtenido.

La relación práctica entre estos conceptos de Inteligencia Artificial se describe a continuación

- **Infraestructura (Bases de datos)**: Proporcionan datos normalizados, es decir limpios, escalados y accesibles que son necesarios para el entrenamiento de modelos.
- **Data Mining / KDD**: Es un proceso interdisciplinario que integra herramientas estadísticas, bases de datos y de aprendizaje automático (ML por sus siglas en inglés) para la generación de conocimiento; como un mediador entre los datos crudos e el conocimiento accionable.
- **Machine Learning**: Sirve en las tareas de clasificación, regresión y segmentación sobre datos preparados
- **Deep Learning**: este tipo de ML es aplicable a grandes volúmenes de datos o a datos complejos (imágenes, texto, señales) que implican el uso de redes profundas; requiere de infraestructuras para entrenamiento distribuido y acelerado (GPU/TPU).

Las bases de datos (ya sean relacionales o NoSQL) son una infraestructura indispensable en proyectos de análisis de datos, sobre esa infraestructura la minería de datos y los procesos KDD convierten los datos en modelos que gestionados con técnicas de ML pueden apoyar la toma de decisiones

Redes neuronales artificiales

Las redes neuronales artificiales son una arquitectura informática de reciente creación cuyo desarrollo se volvió posible derivado del avance del hardware que permite el procesamiento paralelo de información a bajo costo. El esfuerzo de la comunidad tecnológica y su investigación permanente ha consolidado teorías de cálculo colectivo y ha permitido la delimitación de problemas más adecuados para este enfoque, superando en resultados de procesamiento masivo y automático los métodos simbólicos tradicionales.

Las redes neuronales artificiales complementan la denominada inteligencia artificial (IA por sus siglas en inglés) cuando posibilitan soluciones en tareas como el reconocimiento de patrones, clasificación, búsquedas asociativas y optimización de datos incompletos o imprevisibles (Gómez Quesada et al., 1994).

Entre los rasgos distintivos de estas tecnologías frente al procesamiento tradicional clásico (previo al desarrollo de estas tecnologías disruptivas) destacan:

a) **Paralelismo masivo**: se refiere a como miles de unidades procesan información de manera simultánea.

b) **Interconexión elevada**: son nodos que se organizan en capas o niveles con diferentes grados de conexión según el modelo y la aplicación.

c) **Representación distribuida**: el almacenamiento de la información ocurre en patrones de conexiones que abarcan toda la red.

d) **Tolerancia a fallos**: el sistema configurado es capaz de mantener con precisión su funcionamiento aun cuando algunas neuronas artificiales presenten errores.

e) **Cálculo colectivo**: los equipos trabajan de conjuntamente en la resolución de problemas, sin que se requieran instrucciones aisladas.

f) **Autoorganización**: la red de equipos de cómputo puede modificar su estructura lógica para descubrir y aprender nuevos patrones.

La capacidad de procesamiento paralelo distingue a las redes neuronales artificiales se lo cual otorga resiliencia y aprendizaje adaptativo. Lo anterior, las convierte en herramientas idóneas para problemas complejos donde los métodos convencionales ya resultan insuficientes.

Aprendizaje automático

Los sistemas aprenden automáticamente a partir de datos sin ser programados explícitamente para cada tarea, a esta capacidad se refiere el denominado aprendizaje automático (ML por sus siglas en inglés –*Machine learning*-).

Todos los beneficios explicados para una red neuronal pueden aprovecharse si dicha red neuronal es "entrenada" para el aprendizaje. Así, el funcionamiento adecuado de una red neuronal depende de un proceso de entrenamiento, en el cual se definen los resultados esperados, se ajustan los "pesos de sus neuronas" (asignando un valor de cálculo) para obtener los resultados deseados. Este ajuste es iterado, lo cual resulta en la introducción de datos de entrenamiento, obteniendo resultados, analizándolos y corrigiendo los "pesos de sus neuronas" según el error cometido y la contribución de cada neurona al resultado.

De manera clásica y fundamental puede decirse que existen dos grandes categorías de aprendizaje: aprendizaje supervisado y aprendizaje no supervisado. En ambos casos se refiere a redes neuronales que "aprenderán" mediante la introducción de datos de entrenamiento y el posterior análisis de los resultados que las redes neuronales produzcan, si se definen los resultados esperados la red buscará aproximarse a dicho resultado, se dice entonces que el aprendizaje es supervisado; si no se definen resultados esperados, entonces la red consolidará de entradas similares, resultados similares mediante patrones de similitud, se dice por lo tanto que el aprendizaje es no supervisado.

El Aprendizaje supervisado utiliza "pares de datos" de entrada y de salida esperada, en este caso la red neuronal compara sus resultados con la salida deseada y procede a la modificación de pesos en la búsqueda de reducción del error. Este proceso se repite hasta encontrar una diferencia mínima en los resultados obtenidos en comparación con los definidos como deseados.

Algunas de las técnicas y modelos de aprendizaje supervisado mayormente utilizados en la actualidad, según los estudios recientes revisados, se presentan en la tabla 1.

Por su parte, el Aprendizaje no supervisado, emplea solo valores de entrada, sin proporcionar valores de salida esperada, en este caso los pesos de las neuronas de la red se ajustan para que entradas similares generen salidas consistentes. Este tipo de aprendizaje de la red abstrae propiedades colectivas y agrupa patrones de similitud. Se dice entonces que el aprendizaje no supervisado se asemeja más a los sistemas biológicos naturales de aprendizaje, pues la red aprende de manera autónoma.

Tabla 1

Técnicas y modelos recientes de aprendizaje supervisado

Modelo / técnica	Breve descripción
Multilayer Perceptron (MLP)	Red neuronal multicapa; base de arquitecturas profundas.
Gradient Boosting Decision Tree (GBDT)	Algoritmo de ensamble basado en boosting; muy usado en Kaggle/XGBoost.
Support Vector Machine (SVM)	Algoritmo de clasificación/regresión.
Extreme Gradient Boosting (XGBoost)	Algoritmo de boosting altamente eficiente y preciso; muy usado en competencias de datos.
Support Vector Machine (SVM)	Algoritmo de clasificación/regresión basada en hiperplanos.
K-Nearest Neighbors (KNN)	Algoritmo de clasificación/regresión basado en vecinos más cercanos.
Árboles de decisión / Decision Tree	Algoritmo clásico de clasificación y regresión; fácil de interpretar.
Máquinas de Soporte Vectorial (SVR)	Algoritmo de regresión basado en SVM.
Random Forest	Ensamble de árboles de decisión para clasificación/regresión.
Regresión lineal	Modelo estadístico para clasificación y regresión.
Logistic Regression	Modelo estadístico para clasificación binaria/multiclase.
Multiclass Logistic Regression	Modelo estadístico para clasificación multiclase.
Naive Bayes	Clasificador probabilístico basado en teorema de Bayes.

Tabla 2

Técnicas y modelos recientes de aprendizaje no supervisado

Modelo / técnica	Breve descripción
Análisis de series temporales[1]	Incluye modelos clásicos (ARIMA) y DL (LSTM).
Análisis de Componentes Principales (ACP/PCA)	Técnica de reducción de dimensionalidad que simplifica conjuntos de datos complejos, facilita la visualización y el análisis y mejora el rendimiento de otros algoritmos de ML al eliminar redundancia
Clustering (K-means, etc.)	Agrupación de datos sin etiquetas.

El entrenamiento de redes neuronales consiste en la modificación iterada de conexiones para mejorar el desempeño resultante. El aprendizaje supervisado se basa en ejemplos con salidas conocidas, mientras que el no supervisado permite que la red descubra por sí misma patrones y agrupaciones en los datos.

Actualmente esta división clásica del aprendizaje automático, en la práctica, el campo se ha ampliado a más categorías: a) aprendizaje semi-supervisado que combina datos etiquetados y no etiquetados, b) aprendizaje por refuerzo en el que el modelo aprende mediante recompensas y penalizaciones en un entorno dinámico y c) aprendizaje profundo (*deep learning*) este no es una categoría aparte, sino un subconjunto de ML que puede ser supervisado, no supervisado o por refuerzo, pero usando redes neuronales profundas.

Aprendizaje profundo (Deep Learning)

El aprendizaje profundo (DL por las siglas en inglés de *Deep Learning*) es un subcampo del Aprendizaje Automático (ML por las siglas en inglés de *machine learning)* ha demostrado gran capacidad para el manejo de datos complejos y de gran escala, se considera que la aplicación en la investigación (de casi cualquier tema que posea big data) redundará en grandes avances de análisis.

El DL se generó a partir de los avances en algoritmos de optimización librerías especializadas y hardware, y constituye una evolución del aprendizaje de las redes neuronales artificiales que posibilita que el procesamiento paralelo de grandes volúmenes de datos. El desarrollo del DL ha hecho posible trabajar con redes más extensas y flexibles, entrenadas sobre conjuntos masivos de información (Martínez, 2020).

Este enfoque, busca que los sistemas informáticos aprendan de manera autónoma y ejecuten tareas utilizando múltiples capas de procesamiento. Las redes se organizan en capas de entrada, ocultas y salida, en las que la información se transforma progresivamente hasta producir una respuesta final (Martínez, 2020).

[1] El estudio citado utiliza *Machine Learning* semi supervisado pues usa tanto aprendizaje supervisado como no supervisado

La visión artificial, el reconocimiento de voz, los asistentes virtuales, los famosos chatbots, la robótica, publicidad y música son aplicaciones prácticas del aprendizaje profundo. En el ámbito de la seguridad pública, el análisis predictivo y la capacidad para el procesamiento de datos tanto históricos como contextuales convierten al DL en una herramienta para la generación de estrategias anticipatorias de patrones criminales, para la identificación de zonas potenciales de riesgo y, no menos importante, para la optimización del uso de recursos; todo ello fortalece la toma de decisiones y, por tanto, también la prevención del delito (Iberdrola, 2023).

Los beneficios mas destacados del aprendizaje profundo no supervisado son: automatización de procesos, capacidad para variar datos sin premisas establecidas, eficacia en la corrección y disminución de errores, alto rendimiento en el análisis de volúmenes masivos de datos; adicionalmente al valor agregado que se genera cuando se hace factible clasificar datos con eficiencia, elaborar predicciones altamente precisas y posibilitar la construcción de modelos explicativos basados en el análisis predictivo producido (Martínez, 2020; Iberdrola, 2023).

Algunas de las arquitecturas o métodos que pueden ser adecuados por resultar más prometedores según Raghu y Schmidt (2020) y autores de estudios similares consultados se muestran en la tabla 3.

Los fundamentos teóricos del DL, así como la factibilidad de su implementación han permitido el uso de sus técnicas en casos prácticos para la resolución de resolver problemas reales, en estos casos se realiza una evaluación del desempeño del DL en *datasets* específicos. Las aplicaciones interdisciplinarias en las que se han realizado estos análisis son variadas (medicina, bioinformática, ciencias sociales, ingeniería y, por supuesto, criminología) y, el enfoque de estas investigaciones es definir la eficiencia del funcionamiento (del modelo que se aplica) en un entorno real, así como identificar aquellas limitantes (que pueden ser interpretabilidad, ética, entre otras) a las que dicho modelo se enfrenta. Los modelos de DL recientemente utilizados en estos estudios se presentan en la tabla 4.

Tabla 3

Técnicas y modelos recientes de aprendizaje profundo

Modelo / técnica	Breve descripción
CNN (Convolutional Neural Networks)	Arquitectura para imágenes y datos visuales, arquitectura especializada en datos visuales (imágenes, video), muy usada en visión por computadora
RNN (Recurrent Neural Network)	Arquitectura para secuencias y series temporales. Diseñadas para procesar secuencias (texto, series temporales, audio)
LSTM (Long Short-Term Memory)	Variante de RNN que resuelve el problema de dependencias largas en secuencias
Graph Neural Networks (GNN)	Arquitectura diseñada para trabajar con datos estructurados en forma de grafos (nodos y aristas).
Redes neuronales profundas (Deep Neural Networks)	Arquitectura con múltiples capas ocultas; base de DL.
Perceptrón multicapa (MLP)	Arquitectura de red neuronal con capas ocultas; base de DL.
Aprendizaje por transferencia (Transfer Learning)	Reutilización de modelos preentrenados en nuevas tareas.
Backpropagation	Algoritmo de entrenamiento para redes neuronales (ajuste de pesos).
Redes Neuronales (NN)	Modelos inspirados en el cerebro humano; base de DL.
Técnicas de aumento de datos (Data Augmentation)	Estrategias para ampliar datasets y mejorar generalización.

Tabla 4

Técnicas y modelos recientes de aprendizaje profundo

Modelo / técnica	Breve descripción
Multiclass Neural Network (Azure)	Red neuronal multicapa para clasificación multiclase; implementada en plataforma ML de Microsoft Azure.
YOLOv8n (Google Colab Pro)	Arquitectura de detección de objetos en tiempo real.
Keras / TensorFlow	Bibliotecas para construir y entrenar redes neuronales profundas.

Como puede apreciarse la relación conceptual (según Egger et al., 2021) entre los diferentes conceptos técnicos presentados se explica de la siguiente manera: la Inteligencia artificial (IA) es el marco general que busca replicar funciones cognitivas humanas mediante Tecnologías de la Información, el aprendizaje automático (o Machine learning, ML) es un conjunto de técnicas que permiten a los sistemas informáticos identificar y aprender patrones a partir de datos, el aprendizaje profundo (o Deep learning, DL) es un subcampo del ML que se instrumenta mediante redes neuronales profundas, cuya característica es su capacidad para extraer representaciones jerárquicas de los datos, la minería de datos (data mining) son los procesos tecnológicos de preparación de datos que alimenta los modelos ML/DL y de identificación o descubrimiento de tendencias y patrones.

METODOLOGÍA DE REVISIÓN DEL ESTADO DEL ARTE

Para la realización del estado del arte en el estudio en curso referente al uso de Inteligencia Artificial (IA) y aprendizaje automático (*Machine Learning*) para apoyar la toma de decisiones en instituciones de seguridad pública, en particular el objetivo específico de diseñar una arquitectura computacional basada en algoritmos de *machine learning* y procesamiento de lenguaje natural para la construcción de un modelo predictivo automatizado enfocado en el robo vehicular, se requirió la compilación y análisis de estudios similares para la elaboración del estado del arte, al respecto se realizó el análisis documental que se describe.

1. Detección de estudios empíricos contemporáneos
2. Selección de estudios recientes. Identificación de estudios empíricos contemporáneos publicados en el periodo de entre 2019 y 2025
3. Identificación de herramientas de IA más utilizadas en los estudios actuales mencionados en el punto inmediato anterior.
4. Definición de criterios de inclusión: uso explícito de modelos y técnicas de IA/ML/DL en predicción delictiva.
5. Selección y análisis de estudios empíricos desde fuentes primarias gratuitas de artículos indexados, tesis de posgrado, conferencias internacionales disponibles en motores de búsqueda
6. Ordenamiento de los hallazgos de búsqueda en subapartados por región o enfoque.

El resultado de esta búsqueda, selección y ordenamiento razonado se presenta en siguiente sección.

RESULTADOS

Como se especificó, la construcción del estado del arte relativo al uso de IA y ML en la toma de decisiones de instituciones de seguridad pública, se realizó mediante un proceso sistemático de búsqueda y análisis documental. El proceso implicó identificar y seleccionar estudios empíricos recientes (2019–2025), que especificaran el uso de las herramientas de IA/ML/DL más empleadas en la predicción delictiva. Los criterios de inclusión consideraron la aplicación explícita de modelos predictivos.

Los hallazgos organizados por región y enfoque, permitieron estructurar una clara comparativa de los aportes relevantes hallados. La síntesis de este proceso razonado dio origen a la construcción de la síntesis que se presenta en siguiente tabla, que presenta en orden cronológico los estudios seleccionados, sus principales características metodológicas, así como las técnicas empleadas en su elaboración.

Tabla 5

Estudios recientes sobre Análisis delictivo con IA

Autor	Año	Objetivo	Aspectos metodológicos	Conclusiones
Ricardo Francisco Reier Forradellas, Sergio Luis Náñez Alonso, Javier Jorge-Vázquez, Marcela Laura Rodríguez	2020 (publicado en 2021)	Desarrollar un modelo predictivo de delitos Lugar: Buenos Aires Tipo de delito: homicidios, robos, lesiones, hurtos por comuna.	Técnica / modelo: Uso de redes neuronales del modelo SEMMA (Sample, Explore, Modify, Model, Assess). Tipo de datos: Datos abiertos de 2016–2019. Clustering con K-means (K=4) y red neuronal multicapa en Python.	Construcción de un modelo de predicción de delitos por comuna. Se consideró que la metodología SEMMA si permitió la simplificación de del proceso, además la red neuronal demostró capacidad para la identificación de patrones temporales y espaciales de criminalidad.
Santiago Giraldo Alegría, Luis Eduardo Ordoñez Palacios, Víctor Bucheli Guerrero, Hugo Ordoñez Erazo	2020	Predecir la tendencia de víctimas de secuestro Lugar: Colombia Tipo de delito: secuestro.	Técnica / modelo: Entrenamiento de Redes Neuronales (Keras/Tensorflow, MLPRegressor) con 6 capas ocultas y 505 nodos. Comparación con Máquinas de Soporte Vectorial (SVR). Tipo de datos: Dataset de 39,863 registros de la Fiscalía (1982–2020).	La Red Neuronal obtuvo menor RMSE que la SVR, mostrando mejor desempeño. Se predijo la disminución progresiva de secuestros en los próximos 10 años, correlacionada con acuerdos de paz y políticas públicas.
Hugo Ordóñez, Carlos Cobos, Víctor Bucheli	2019–2020	Predecir la tendencia de hurtos Lugar: Colombia y en tres ciudades principales (Bogotá, Medellín, Cali). Tipo de delito: hurto	Técnica / modelo: Regresión con máquinas de Soporte Vectorial (SVR), ajustando parámetros (ε=0.1) para mejorar RMSE. Tipo de datos: Dataset de 2,662,402 registros de la Fiscalía (1960–2019). Se filtraron 350,320 registros de hurto	El modelo SVR ajustado mostró mejor precisión que la regresión lineal y SVR estándar. Se concluyó que los hurtos tienden a disminuir en Colombia y en las principales ciudades, reflejando el impacto de políticas públicas y estrategias de seguridad.
Kyoung-Sook Kim & Yeong-Hoon Jeong (Corea)	2021	Predecir tipos de crimen y herramientas utilizadas para reducir la tasa de criminalidad mediante IA.	Técnica / modelo: Algoritmos de Multiclass Logistic Regression y Multiclass Neural Network en Microsoft Azure; limpieza y división de datos (70% entrenamiento, 30% prueba). Tipo de datos: Uso de datos de Baltimore (2011–2016)	La regresión logística mostró mayor precisión y estabilidad que la red neuronal Se elaboró predicción de alta probabilidad el uso de "HANDS" y el crimen "COMMON ASSAULT"; Como resultado se ha propuesto la extenciòn del modelo para predecir reincidencia y desplegarlo como servicio web.
Hugo Calderón Vilca, Oscar Abad Nauto y col. (Perú)	2021	Diseñar una arquitectura de red neuronal capaz de predecir si una persona podría cometer un delito según sus condiciones familiares y sociales. Tipo de delito: N/A	Técnica:/ modelo: Tres arquitecturas de perceptrón multicapa con algoritmo backpropagation; normalización de datos y pruebas con diferentes capas ocultas. Tipo de datos: Datos del Censo Penitenciario Nacional 2016 (14,000 registros)	La arquitectura con 24 entradas, dos capas ocultas (21 y 21 neuronas) y 19 salidas obtuvo mejor desempeño (error cuadrático medio 0.278, precisión 96.03%); La implementación de la aplicación web para instituciones de justicia; confirmó la relación entre condiciones sociales/familiares y probabilidad de delinquir.
Ana Laura Lira Cortés & Carlos Fuentes Silva (México, Querétaro)	2021	Revisar y probar modelos de IA basados en redes neuronales para predecir delitos en espacios urbanos. Lugar: ciudades Atlanta, Baltimore, Chicago y Portland	Técnica / modelo: Análisis de datos históricos y espacio-temporales; pruebas con algoritmos como Naive Bayes, SVM, Random Forest, Redes Neuronales, RNN y LSTM; comparación de	Precisión de hasta 81% de precisión en redes neuronales; Percisiónentre 75–90% de predictibilidad de LSTM según tipo de robo. Identificación de patrones espacio-temporales de incidencia delictiva;

Autor	Año	Objetivo	Aspectos metodológicos	Conclusiones
		Tipo de delitos: delitos en espacios urbanos	precisión en ciudades de EUA.	Se encontró necesidad de modelos locales Se plantean consideraciones éticas sobre privacidad y sesgos en IA.
Wajiha Safat, Sohail Asghar, Saira Andleeb Gillani	2021	Analizar y comparar múltiples algoritmos de ML y DL para predecir y pronosticar delitos (incluyendo tendencias futuras) Lugar: Chicago y Los Ángeles,.	Técnica / modelo: Algoritmos de ML (Logistic Regression, SVM, Naïve Bayes, KNN, Decision Tree, MLP, Random Forest, XGBoost) y modelos de series temporales (LSTM y ARIMA). Tipo de datos: Uso de datasets masivos (Chicago: 7M registros; Los Ángeles: 2.6M).	Los modelos LSTM y ARIMA mostraron buen desempeño en predicción temporal. Identificación de patrones estacionales y áreas de alta densidad delictiva. Conclusión de los autores relativa a la mejora de la precisión ante la combinación de ML y DL además de permitir pronósticos útiles en planificación policial.
Dr. M. Bheemalingaiah, E. Nageswara Rao, A. Srirama Kanaka Ratnam, Sudhakar Vecha	2022	Proponer un sistema avanzado de predicción de crímenes cibernéticos mediante algoritmos de IA y visión por computadora para apoyar a las agencias de seguridad. Lugar: Chicago, Vancouver Tipo de delito: crímenes cinbernéticos	Técnica /modelo: Regresión lineal, KNN, árboles de decisión, SVM, redes neuronales profundas Integración de visión por computadora y análisis estadístico. Tipo de datos: grandes bases de datos de delitos	La mejora de la detección y prevención de delitos puede lograrse mediante la combinación de ML, DL y visión por computadora. El modelo DNN alcanzó precisión mayor a 84.2%, en comparación con otros métodos, por tanteo se le considera potencial para sistemas de vigilancia continua.
Muzammil Khan, Azmat Ali, Yasser Alharbi	2022	Desarrollar un modelo de predicción de delitos en San Francisco utilizando técnicas de clasificación para optimizar recursos policiales. Lugar: San Francisco	Técnica / modelo Comparación de tres algoritmos: Naive Bayes, Random Forest y Gradient Boosting Decision Tree. Tipo de datos: Dataset de Kaggle (2003–2015, 878,049 registros). Evaluación con métricas de precisión, recall y análisis exploratorio de datos.	El modelo Gradient Boosting Decision Tree obtuvo la mayor precisión de 98.5%, superando a Naive Bayes y Random Forest. La conclusión elaborada por los autores es que el modelo Gradient Boosting Decision Tree es el más eficaz para la predicción de delitos y para el apoyo en la distribución eficiente de recursos policiales.
Nohemí Abilene Díaz Linares	2023	Analizar la relación entre el robo de vehículo y la calidad de vida urbana en la Ciudad de México (2018–2020) desde la perspectiva de la criminología ambiental. Lugar: Ciudad de México Tipo de delito: Robo de vehículo	Técnica /modelo: Construcción de un Índice de Calidad de Vida Urbana (ICVU) Uso de análisis de componentes principales (ACP), estratificación óptima de Dalenius y Hodges, y autocorrelación espacial (Índice de Moran). Tipo de datos: datos de 2018 a 2020	Existencia de asociación espacial positiva baja entre robo de vehículo y calidad de vida urbana. Consolidación de segmentos de alta calidad de vida y alto robo de vehículo en delegaciones centrales de la CDMX.
William Andrés Villavicencio Benalcázar	2023	Levantar un estado del arte sobre técnicas de aprendizaje automático aplicadas a la ciberseguridad.	Técnica / modelo Revisión bibliográfica de algoritmos de ML (Naïve Bayes, Árboles de decisión, Random Forest, SVM, KNN). Análisis de aplicaciones en detección de intrusos, malware, ingeniería social y seguridad en redes.	Las técnicas de ML permiten la detección más eficiente de amenazas que los métodos tradicionales. Combinar clasificadores (boosting, híbridos) es sugerido para mayor robustez en ciberseguridad.

Autor	Año	Objetivo	Aspectos metodológicos	Conclusiones
Espíndola, J. N.	2024	Analizar la distribución espacio-temporal de actividades delictivas en Morelia mediante aprendizaje automático no supervisado. Lugar: Morelia, México	Técnica / modelo: Clustering y técnicas de minería de datos Uso de mapas de calor y análisis georreferenciado para identificar patrones. Tipos de datos: registros oficiales de criminalidad.	Identificación de zonas críticas y patrones recurrentes de delitos El enfoque no supervisado identificó dinámicas delictivas, es posible proponer estrategias de prevención basadas en evidencia.
Simón Ricardo Burgos Alarcón	2025	Analizar la criminalidad en el departamento de Córdoba (2021–2024) y generar alertas tempranas para mejorar la gestión de la seguridad pública. Lugar: Córdoba	Técnica / modelo: Modelos supervisados y no supervisados, análisis de series temporales y análisis espacial. Visualización mediante mapas de calor y análisis de sentimiento con NLP (NLTK, TextBlob, VADER). Tipo de datos: Datos históricos de fuentes oficiales y medios regionales.	Identificación de patrones espaciales y temporales de criminalidad. Anticipación de incidentes mediante modelos predictivos y, por tanto, proponer estrategias preventivas para la región.
Raquel Victoria Pacheco Olivo, Víctor Geovanny Vallejo Villavicencio, Diego Armando Enríquez Gutiérrez, Alexander Vladimir Rojas Cevallos, José Octavio Bonilla Peñaherrera	2025	Diseñar un modelo de aprendizaje automático para videovigilancia inteligente que detecte armas de fuego en restaurantes del cantón Quito. Lugar: Quito	Técnica / modelo: Enfoque cuantitativo, aplicativo y descriptivo. Entrenamiento con CNN y arquitectura YOLOv8n en Google Colab Pro. Técnicas de aumento de datos y aprendizaje por transferencia. Tipo de datos: Dataset de 1,009 imágenes de armas (Roboflow).	Precisión del 99.5%, recall del 98% y mAP@0.5 de 0.987 Demostración de la eficacia del modelo para la detección automática de armas en tiempo real. Lo que permite el fortaleciendo la seguridad en restaurantes y promoviendo confianza ciudadana.

A continuación, se presentan los hallazgos, en un resumen cuya organización obedece a subapartados por región o enfoque:

En América se encontraron estudios en México, Ecuador, Colombia, Perú y Argentina. Los estudios en México se realizaron estudios aplicando: análisis espacio-temporal, técnicas espaciales y *clustering* no supervisado (Lira Cortés & Fuentes Silva, 2021; Díaz Linares, 2023; Espíndola, 2024), en distintas arquitecturas de redes neuronales para el mapeo de patrones delictivos y la relación entre entornos urbanos y distintos tipos de delitos. En Ecuador la aplicación de ML y tecnologías de videovigilancia inteligente con IA para detección de: a) amenazas digitales y b) detección automática de armas (Villavicencio, 2023; Pacheco Olivo et al., 2025). Colombia fue el país con mayor frecuencia de este tipo de estudios, se utilizaron comparaciones entre redes neuronales, SVR, SVR ajustado, modelos predictivos para el análisis anticipatorio de delitos (Giraldo Alegría et al., 2020; Ordóñez et al., 2020; Burgos Alarcón, 2025). En Perú y Argentina igualmente se realizaron estudios con SEMMA y redes neuronales multicapa (Reier Forradellas et al., 2021) para anticipar conductas delictivas (Calderón Vilca et al., 2021) con redes neuronales en distintos tipos de delitos

En otros países encontrados la predicción del crimen aborda los siguientes ángulos: con enfoque en visión por computadora y redes neuronales profundas, realizado en China e India (Bheemalingaiah et al., 2022), otro en clasificación comparativa con datos de San Francisco con predicción temporal con LSTM/ARIMA (Safat et al., 2021; Khan et al., 2022), otro en predicción y pronóstico

temporal con conjuntos masivos de datos (Pakistán – Safat), uno final en Corea que analiza la tipología del crimen con datos históricos (Kim & Jeong, 2021).

CONCLUSIONES

Con esta revisión sistemática se ha encontrado una diversidad regional en los estudios realizados en América, en países como México, Ecuador, Colombia, Perú y Argentina han desarrollado investigaciones aplicando distintas técnicas de IA y ML para la predicción del crimen, mientras en Asia se destaca lo encontrado en China, India, Pakistán y Corea, con enfoques en visión por computadora, redes neuronales profundas y análisis temporal.

Se identifica a Colombia como un referente en estudios de este tipo pues es el país con mayor frecuencia de estudios empíricos en la región, allí se han aplicado comparaciones entre técnicas y modelos para la predicción de delitos, lo que muestra un interés sostenido en metodologías comparativas y en la validación de modelos predictivos.

También se halló variedad metodológica dado el empleo de técnicas diversas tanto de aprendizaje supervisado como no supervisado, así como de arquitecturas de redes neuronales multicapa. Esta diversidad refleja la inexistencia de un único enfoque dominante, en contraste se realizan análisis constantes de diferentes modelos en la búsqueda de adaptabilidad y precisión.

El uso de IA/ML/DL en análisis criminológico es una tendencia internacional que se configura como una tendencia global, sus con aplicaciones que van desde la vigilancia inteligente hasta la predicción temporal de delitos; por lo que se estima que la predicción del crimen mediante IA es un campo emergente y multidisciplinario, con aportes relevantes en distintas regiones.

Finalmente, para mejorar la calidad de una evaluación crítica de la validez de los estudios sería pertinente establecer si los datos son representativos e identificar si contienen sesgos asì como la naturaleza de los mismos

Para la medición del impacto real en políticas públicas se requiere la implementación de los modelos desarrollados, así como su evaluación en campo. En tanto no suceda así, no puede realizarse una generalización de resultados pues la aplicación práctica es limitada, excepto en casos como el de videovigilancia en Quito, la mayoría de modelos no refieren que hayan sido implementados como políticas públicas.

Los estudios presentados muestran poca integración interdisciplinaria, en ellos hay predominancia de enfoques técnicos, con poca articulación con teorías criminológicas, por tantos se considera necesaria la incorporación de perspectivas que aborden el tema desde aproximaciones de criminología, sociología, derecho, psicología, etc.

Se observó que los estudios presentados utilizan conjuntos de datos históricos; se considera que hace falta comprobar su eficacia, en posibilidad de hacerlo, en escenarios reales con datos en tiempo real y la realización de estudios longitudinales que permitan identificar la eficacia de los modelos diseñados a través del tiempo.

En tanto a consideraciones éticas deben tenerse en cuanta la privacidad en el uso de datos sensibles y, dada la utilización de estos datos se considera la posibilidad de sesgos en datos oficiales que potencian el riesgo de reproducir desigualdades sociales.

Los estudios revisados muestran que la IA ha permitido avances significativos en la predicción del delito, desde modelos espacio-temporales hasta aplicaciones en ciberseguridad y videovigilancia; se confirma que las redes neuronales y modelos híbridos ofrecen mejor desempeño que métodos tradicionales, y que la integración de análisis espacial y NLP abre nuevas posibilidades para la gestión de la seguridad pública. Sin embargo, la validez de los resultados depende de la calidad de los datos y de la capacidad de los modelos para adaptarse a contextos locales. La investigación futura debe enfocarse en la interpretabilidad de los modelos, la reducción de sesgos y la evaluación de impacto en políticas públicas reales.

REFERENCIAS

Bheemalingaiah, M., Nageswara Rao, E., Srirama Kanaka Ratnam, A., & Vecha, S. (2022). Advanced cyber crime rate prediction using deep learning & AI based algorithm. *Journal of Harbin Institute of Technology, 54*(3), 140–146. https://doi.org/10.11720/JHIT.54032022.17

Burgos Alarcón, S. R. (2025). *Análisis predictivo en seguridad pública: Generación de alertas con datos históricos del departamento de Córdoba* [Proyecto aplicado, Especialización en Ciencia de Datos y Analítica, Universidad Nacional Abierta y a Distancia]. Repositorio Institucional UNAD. https://repository.unad.edu.co/handle/10596/73242

Calderon-Vilca, H. D., Carhuaricra Rivera, L., Abad Nauto, O., Carrillo Estrada, J., Calderon-Vilca, E. F., & Ibarra, M. (2021, septiembre). Review of crimes in Peru and proposal of a neural network architecture to predict if a person could commit a crime. En *Proceedings of the 29th Conference of the Open Innovations Association FRUCT* (pp. 416–422). Open Innovations Association FRUCT. https://doi.org/10.5281/zenodo.4770802

Díaz Linares, N. A. (2023). *Entorno urbano y delincuencia: Robo de vehículo y calidad de vida urbana en la Ciudad de México, 2018–2020* [Tesis de maestría, Universidad Autónoma del Estado de México].

Egger, J., Pepe, A., Gsaxner, C., Jin, Y., Li, J., & Kern, R. (2021). Deep learning—a first meta-survey of selected reviews across scientific disciplines, their commonalities, challenges and research impact. *PeerJ Computer Science, 7*, e773. https://doi.org/10.7717/peerj-cs.773

Giraldo-Alegría, S., Ordoñez-Palacios, L., Bucheli-Guerrero, V., & Ordoñez-Erazo, H. (2020). Modelo de redes neuronales para predecir la tendencia de víctimas de secuestro en Colombia. *Investigación e Innovación en Ingenierías, 8*(3), 38–49. https://doi.org/10.17081/invinno.8.3.4702

Gupta, M. K., & Chandra, P. (2020). A comprehensive survey of data mining. *International Journal of Information Technology, 12*, 1243–1257. https://doi.org/10.1007/s41870-020-00427-7

Iberdrola. (2023). *Aplicaciones del deep learning*. https://www.iberdrola.com/conocenos/nuestro-modelo-innovacion/deep-learning

Khan, M., Ali, A., & Alharbi, Y. (2022). Predicting and preventing crime: A crime prediction model using San Francisco crime data by classification techniques. *Complexity, 2022*, Article 4830411. https://doi.org/10.1155/2022/4830411

Kim, K.-S., & Jeong, Y.-H. (2021). A study on crime prediction to reduce crime rate based on artificial intelligence. *Korean Journal of Artificial Intelligence, 9*(1), 15–20. https://doi.org/10.24225/kjai.2021.9.1.15

Lira Cortés, A. L., & Fuentes Silva, C. (2021). Artificial intelligence models for crime prediction in urban spaces. *Machine Learning and Applications: An International Journal, 8*(1), 1–15. https://doi.org/10.5121/mlaij.2021.8101

Martínez, J. (2020). *Deep learning: Fundamentos y aplicaciones*.

Navarro Espindola, J. (2024). *Análisis espacio-temporal de actividades delictivas en Morelia con un enfoque de aprendizaje automático no supervisado* [Tesis de licenciatura, Universidad Nacional Autónoma de México]. Repositorio Institucional de la UNAM. https://repositorio.unam.mx/contenidos/3713132

Ordóñez, H., Cobos, C., & Bucheli, V. (2020). Modelo de machine learning para la predicción de las tendencias de hurto en Colombia. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E29), 494–506.

Pacheco Olivo, R. V., Vallejo Villavicencio, V. G., Enríquez Gutiérrez, D. A., Rojas Cevallos, A. V., & Bonilla Peñaherrera, J. O. (2025). Diseño de un modelo de aprendizaje automático para videovigilancia inteligente para la detección automática de armas de fuego enfocado a restaurantes ubicados en el cantón Quito. *Ciencia Latina Revista Científica Multidisciplinar, 9*(5), 11182–11205. https://doi.org/10.37811/cl_rcm.v9i5.20463

Raghu, M., & Schmidt, E. (2020). *A survey of deep learning for scientific discovery* [Preprint]. arXiv. https://arxiv.org/abs/2003.11755

Reier Forradellas, R. F., Náñez Alonso, S. L., Jorge-Vázquez, J., & Rodríguez, M. L. (2021). Applied machine learning in social sciences: Neural networks and crime prediction. *Social Sciences, 10*(1), Article 4. https://doi.org/10.3390/socsci10010004

Safat, W., Asghar, S., & Gillani, S. A. (2021). Empirical analysis for crime prediction and forecasting using machine learning and deep learning techniques. *IEEE Access, 9*, 70080–70094. https://doi.org/10.1109/ACCESS.2021.3078117

Villavicencio Benalcázar, W. A. (2023). *Estudio de técnicas de aprendizaje automático utilizadas en ciberseguridad* [Tesis de maestría, Universidad Politécnica Salesiana]. Repositorio Institucional de la Universidad Politécnica Salesiana. https://dspace.ups.edu.ec/bitstream/123456789/26498/1/UPS-CT011003.pdf