

Una revisión sistemática sobre la prevención espacial de los delitos

Ourania Kounadi, Alina Ristea, Adelson Araujo Jr. and Michael Leitner

Antecedentes: La vigilancia policial predictiva y el análisis del delito con un enfoque espaciotemporal reciben cada vez más atención entre una variedad de comunidades científicas y ya se están implementando como herramientas policiales eficaces. El objetivo de este artículo es proporcionar una visión general y una evaluación del estado del arte en la predicción espacial de los delitos centrándose en el diseño del estudio y los aspectos técnicos.

Métodos: Seguimos las pautas PRISMA para informar esta revisión sistemática de la literatura y analizamos 32 artículos, entre el año 2000 al 2018, que fueron seleccionados de 786 artículos que ingresaron a la fase de selección, y un total de 193 artículos que pasaron por la fase de elegibilidad. La fase de elegibilidad incluyó varios criterios que se agruparon en:

- a) el tipo de publicación;
- b) relevancia para el alcance de la investigación;
- c) características del estudio.

Resultados: El tipo más predominante de inferencia de pronóstico es el método de puntos críticos (es decir, clasificación binaria). Se utilizaron principalmente métodos tradicionales de aprendizaje automático, pero también enfoques basados en la estimación de la densidad del kernel y, con menos frecuencia, enfoques de proceso puntual y aprendizaje profundo. Las principales medidas del rendimiento de la evaluación son la precisión de la predicción, seguida del índice de precisión de la predicción y la puntuación. F1. Finalmente, el enfoque de validación más común fue la división tren-prueba, mientras que otros enfoques incluyen la validación cruzada, dejar uno fuera y el horizonte móvil.

Limitaciones: Los estudios actuales a menudo carecen de informes claros de los experimentos de estudio, procedimientos de ingeniería de características y utilizan terminología inconsistente para abordar problemas similares.

Conclusiones: Existe un notable crecimiento en los estudios de predicción espacial del delito como resultado del trabajo técnico interdisciplinario realizado por académicos de diversos orígenes. Estos estudios abordan la necesidad social de comprender y combatir el crimen, así como el interés de las fuerzas del orden en la predicción casi en tiempo real.

Implicaciones: Aunque identificamos varias oportunidades y fortalezas, también hay algunas debilidades y amenazas para las cuales brindamos sugerencias. Los estudios futuros no deberían descuidar la yuxtaposición de algoritmos (existentes), cuyo número aumenta constantemente. Para permitir la comparación y reproducibilidad de los estudios, describimos la necesidad de un protocolo o estandarización de los enfoques de pronóstico espacial y sugerimos el informe de los elementos de datos clave de un estudio.

FONDO

La criminología ambiental proporciona una base teórica importante para explorar y comprender la distribución espacial del crimen (Bruinsma y Johnson 2018). La ocurrencia de delitos dentro de un área fluctúa de un lugar a otro. Además, la ocurrencia de delitos se desprende de una multitud de factores y muestra una mayor importancia estratégica, complejidad e interacción con otras redes, como las institucionales o comunitarias. En la investigación criminológica, estos factores se denominan principalmente atractores y generadores de delitos (Kinney et al. 2008). Las fluctuaciones espaciales y las dependencias de atractores y generadores sugieren que el crimen no es aleatorio en el tiempo y el espacio. Una base sólida para el análisis del crimen predictivo espacial es la teoría del patrón criminal (Brantingham y Brantingham 1984). Se utiliza para explicar por qué ocurren delitos en áreas específicas, sugiere que el crimen no es aleatorio y que puede ser organizado u oportunista. En particular, muestra que cuando el espacio de actividad de una víctima se cruza con el espacio de actividad de un delincuente, hay mayores posibilidades de que ocurra un delito. El perímetro de actividad de una persona está espacialmente limitado por los lugares a los que asiste (nodos). Por ejemplo, si uno de los nodos personales está en un vecindario con alta criminalidad, los delincuentes encontrarán nuevas oportunidades para ofender.

Si el delito no es aleatorio, se puede estudiar y, como tal, se pueden modelar sus patrones, incluido el componente espacial. Como consecuencia, las teorías de la criminología ambiental han sido probadas científicamente y en la última década varios campos de investigación han logrado grandes avances en el desarrollo de métodos para la predicción y evaluación (espaciotemporal) del delito (Caplan et al. 2011; Mohler et al. 2011, 2015; Perry 2013; Wang y Brown 2011; Yu et al. 2011).

La mayoría de las técnicas de predicción se utilizan para pronósticos retrospectivos, es decir, predecir el futuro a través de datos históricos. Los datos históricos sobre delitos se utilizan solos o junto con atractores y generadores de delitos (que pueden ser demográficos, ambientales, etc.) en diversos tipos de modelos de predicción (Mohler et al. 2011; Ohyama y Amemiya 2018; Yu et al. 2011). Además de los datos estáticos, como las variables demográficas o socioeconómicas, como predictores, los investigadores han incluido recientemente características dinámicas de espacio y tiempo, dando así un impulso a la predicción de la ocurrencia de delitos. Estos modelos consisten en datos de redes sociales (Al Boni y Gerber 2016; Gerber 2014; Kadar et al. 2017; Wang y cols. 2012; Williams y Burnap

2015), y datos de recogida y devolución de taxis (Kadar y Pletikosa 2018; Wang et al. 2016; Zhao y Tang 2017).

Aunque los modelos actuales de predicción del delito muestran una precisión cada vez mayor, se ha puesto poco énfasis en dibujar el panorama empírico y técnico para delinear las fortalezas y oportunidades para investigaciones futuras, pero también para identificar debilidades y amenazas. En este artículo, nos centramos en la prevención espacial del delito, que es la prevención espacial de información relacionada con el delito. Tiene muchas aplicaciones, como la prevención espacial del número de delitos, el tipo de actividad delictiva, la siguiente ubicación de un delito en una serie u otra información relacionada con el delito. En este punto, Debemos tener en cuenta que nos encontramos con artículos que afirman realizar pronósticos de delitos espaciales o pronósticos de delitos mientras extrapolan en el espacio o detectan conglomerados espaciales.

En general, los artículos en el campo del análisis espacial del crimen utilizan el término predicción como sinónimo de pronóstico y tienen preferencia por el término predicción (Perry 2013). Sin embargo, existen varios tipos de predicción espacial con aplicaciones de interpolación o extrapolación. La prevención es una predicción que extrapola una variable estimada a un tiempo futuro. Si bien la predicción puede ser sinónimo de prevención, a menudo también se utiliza para inferir valores desconocidos independientemente de la dimensión temporal (por ejemplo, predecir el delito en el área A utilizando un modelo derivado del área B). Cressie (1993, págs. 105-106) se refiere a la predicción espacial como un proceso de inferencia para predecir valores en ubicaciones desconocidas a partir de datos observados en ubicaciones conocidas. Su terminología incluye las nociones temporales de suavizado o interpolación, filtrado y predicción, que tradicionalmente utilizan unidades de tiempo en lugar de ubicaciones.

Como resultado, al buscar literatura sobre pronósticos es necesario agregar el término “predicción”, que deriva de un conjunto de artículos mucho mayor que los que en realidad hacen “sólo” pronósticos. En este artículo, definimos el término “Pronóstico espacial del crimen” como un enfoque de inferencia sobre el crimen tanto en el tiempo como en el espacio. En el cuadro siguiente, agregamos límites de definición describiendo variaciones de los enfoques de pronóstico que consideramos en nuestro estudio.

Nos impulsa la necesidad de armonizar los conceptos y metodologías existentes dentro y entre las comunidades de criminología, sociología, geografía e informática. El objetivo de este artículo es realizar una revisión sistemática de la literatura sobre análisis espacial de predicción del delito, con un enfoque en la predicción del delito, para comprender y evaluar el estado del arte en materia de conceptos y métodos, dado el ritmo sin precedentes de los estudios empíricos publicados. A continuación, enumeramos las preguntas de investigación de este estudio.

1. ¿Cuáles son los tipos de información pronosticada para los cuales el espacio juega un papel importante? (Sección "Resumen de publicaciones seleccionadas sobre prevención espacial de delitos").
2. ¿Cuáles son los métodos de pronóstico comúnmente utilizados? (Sección "Métodos de prevención de delitos espaciales").
3. ¿Cuáles son las similitudes y diferencias técnicas entre los modelos espaciales de prevención de delitos? (Sección "Métodos de prevención de delitos espaciales").
4. ¿Cómo se mide el desempeño predictivo en el pronóstico espacial de delitos? (Sección "Consideraciones al analizar el desempeño de los pronósticos").
5. ¿Cuáles son las estrategias de validación de modelos comúnmente utilizadas? (Sección "Consideraciones al analizar el desempeño de los pronósticos").
6. ¿Cuáles son las principales dependencias y limitaciones del desempeño de la prevención del delito? (Sección "Consideraciones al analizar el desempeño de los pronósticos").

Antes de presentar los resultados (sección "Resultados") y discutirlos en forma de análisis FODA (sección "Discusión"), resumimos el trabajo bibliográfico previo sobre predicción y análisis del delito (sección "Trabajo relacionado") y luego presentamos los resultados. metodología para seleccionar los trabajos y asegurar la calidad del estudio (sección "Métodos"). Por último, en la sección "Conclusión" concluimos con los principales hallazgos de cada pregunta de investigación. Con nuestro trabajo, pretendemos arrojar luz sobre futuras direcciones de investigación e indicar los obstáculos a considerar al realizar pronósticos espaciales de delitos.

TRABAJO RELACIONADO

Los artículos identificados como estudios de revisión o de trabajos relacionados (un total de 13) se remontan a 2003 y están conectados con la estrategia de palabras clave que utilizamos (encuentre más detalles en la sección "Selección de estudios"). Además, para revisar los artículos (un total de 9), también incluimos dos editoriales, un capítulo de libro y un artículo de investigación, porque contienen una extensa revisión de la literatura en el campo del análisis predictivo del crimen.

Cinco artículos se centran en la minería de datos con un alcance mucho más amplio que nuestros temas de interés, es decir, predicción, pronóstico o análisis espacial. El más antiguo propone un marco para la extracción de datos sobre delitos (Chen et al. 2004). Agrupa las técnicas de minería en ocho categorías, que incluyen (a) extracción de entidades (ejemplo de uso: para identificar personas), (b) agrupación (ejemplo de uso: para distinguir entre grupos que pertenecen a diferentes pandillas), (c) minería de reglas de asociación (ejemplo de uso: para detectar ataques a la red), (d) minería de patrones secuenciales (ejemplo de uso: igual que antes), (e) detección de desviación (ejemplo de uso: para identificar fraude), (f) clasificación (ejemplo de uso: para identificar spam de correo electrónico), (g) comparador de cadenas

(ejemplo de uso: para detectar información engañosa) y (h) análisis de redes sociales (ejemplo de uso: construir el papel del criminal en una red). La regla de asociación, la agrupación y la clasificación son los que se han discutido en otras revisiones de minería de datos sobre delitos, como para la identificación de delincuentes (es decir, elaboración de perfiles) (Chauhan y Sehgal 2017), aplicaciones para resolver delitos (Tongsatapornwatana 2016), y aplicaciones del análisis de carreras criminales, perfiles de investigación y análisis de patrones (con respecto al comportamiento criminal) (Tongtae y Srisuk 2008). Además, Hassani et al. (2016) realizaron una revisión en profundidad reciente que analizó más de 100 aplicaciones de extracción de datos sobre delitos. Su taxonomía de aplicaciones identifica cinco tipos que incluyen los descritos previamente por Chen et al. (2004) con la excepción de la minería de patrones secuenciales, la detección de desviaciones y el comparador de cadenas. En cuanto a algoritmos específicos, el énfasis se pone en tres tipos: árboles de decisión, redes neuronales y máquinas de vectores de soporte. Chen et al. (2004) cubre un amplio espectro de análisis e investigación de delitos y, como tal, identifica un par de estudios relacionados con la detección y prevención de puntos críticos bajo las categorías mineras de agrupación y clasificación. Estos estudios de revisión técnica nos dieron ejemplos de los elementos de datos que necesitamos extraer y analizar, cómo las técnicas que se utilizan y las tareas que se realizan (Tongsatapornwatana 2016), así como el propósito del estudio y la región (Hassani et al. .2016).

El artículo más antiguo, pero todavía relevante para nuestro trabajo, es un editorial de seis estudios de prevención del delito (Gorr y Harries 2003). Los autores se refieren al pronóstico de delitos como un nuevo dominio de aplicación, que incluye el uso de sistemas de información geográfica (SIG), realiza predicciones a largo y corto plazo con métodos univariados y multivariados, y unidades de área de límites fijos versus unidades de área ad hoc para el espacio. y datos de series temporales. Más de 15 años después, este dominio de aplicación no es nuevo pero aún presenta las mismas características descritas anteriormente. Otro editorial de Kennedy y Dugato (2018) presenta un número especial sobre la prevención de delitos espaciales utilizando el enfoque de modelado de terreno de riesgo (RTM). El enfoque de la mayoría de los artículos es analizar los factores que conducen a pronósticos precisos porque la base del enfoque RTM se basa en la Teoría de los lugares riesgosos de Kennedy y Caplan (2012).

Esta teoría parte de la proposición de que los lugares varían en términos de riesgo debido a la influencia espacial de factores criminógenos. Por último, un estudio de revisión reciente resume estudios anteriores de predicción de delitos de cuatro métodos, a saber, máquinas de vectores de soporte, redes neuronales artificiales, teoría difusa y series temporales multivariadas (Shamsuddin et al. 2017) . Los autores sugieren que los investigadores propongan métodos híbridos para producir mejores resultados. En nuestro estudio agrupamos y discutimos un número mucho más amplio de métodos (una lista de 66 en el archivo adicional 1 C) y también identificamos enfoques híbridos (por ejemplo, métodos de conjunto), uno de los cuales se remonta a 2011.

Además, identificamos dos artículos que describen métodos espaciales para la predicción espacial del crimen per se. El artículo de Bernasco y Elfers (2010) analiza métodos estadísticos y espaciales para analizar el crimen. Es interesante que distinguen dos tipos de resultados espaciales para el modelado, incluida la distribución espacial y el movimiento. Cuando se trata de distribución espacial, que es relevante para el alcance de nuestro artículo, los autores describen los siguientes métodos espaciales, incluidos modelos de regresión espacial, filtrado espacial, regresión ponderada geográficamente y regresión multinivel con dependencia espacial. El artículo de Chainey et al. (2008) se centra en el mapeo de puntos críticos como un enfoque básico para la predicción del delito. Las técnicas que describen y se examinan empíricamente son elipses espaciales, mapeo temático de áreas geográficas, mapeo temático de cuadrícula y estimación de densidad del núcleo (KDE). Entre estos, KDE obtuvo la puntuación más alta del índice de precisión de predicción (PAI). Sorprendentemente, la mayoría de los métodos espaciales (con la excepción de KDE y RTM) no han sido utilizados por los autores de nuestros artículos seleccionados (consulte los métodos discutidos en la sección “Métodos de pronóstico de delitos espaciales”).

Con respecto a la vigilancia policial predictiva, una revisión reciente explica su definición, cómo funciona, cómo evaluar su efectividad y también proporciona una descripción general de las aplicaciones existentes (en su mayoría comerciales) (Hardyns y Rummens 2018). Uno de los aspectos innovadores de esta revisión es la sección sobre la evaluación de la vigilancia policial predictiva utilizando tres criterios, a saber, la exactitud de la predicción, el efecto de las implementaciones de la vigilancia policial predictiva en las tasas de criminalidad reales y los costos relativos a los métodos que se reemplazan. Los autores de este artículo apoyan la definición de vigilancia policial predictiva que se origina en Ratcliffe (2015, p. 4), que dice: “el uso de datos históricos para crear un pronóstico espaciotemporal de áreas de criminalidad o puntos críticos de criminalidad que serán la base para las decisiones de asignación de recursos policiales con la expectativa de que tener agentes en el lugar y momento propuesto disuadirá o detectará la actividad criminal”. En general, la prevención espacial de la delincuencia tiene un alcance más amplio y no es sinónimo de vigilancia policial predictiva. Además, los artículos que examinamos no pretenden ayudar en las decisiones policiales (aunque esto puede ser una consecuencia indirecta), sino que tienen un enfoque académico y explicativo. Sin embargo, la efectividad del primer criterio del análisis predictivo, tal como lo enmarcan Hardyns y Rummens (2018), está altamente relacionada con nuestro alcance y, por lo tanto, se analiza en mayor detalle, desde una perspectiva técnica, en sección “Consideraciones al analizar el desempeño de las predicciones”.

Una segunda revisión sistemática de vigilancia policial predictiva realizada por Seele (2017) examina el potencial de los macrodatos para promover la sostenibilidad y reducir el daño y también analiza aspectos éticos y legales vinculados a los algoritmos predictivos. Similarmente, Ozkan (2018) también revisa los big data para la investigación criminal. Este artículo ofrece una discusión crítica sobre los beneficios y limitaciones de la investigación basada en datos y llama la atención sobre la amenaza inminente de eliminar las pruebas de

hipótesis convencionales, que tradicionalmente han sido una herramienta científica integral para los científicos sociales y criminólogos.

A excepción de Seele (2017), ningún otro estudio de trabajos relacionados sigue un procedimiento sistemático con respecto a los métodos para identificar y seleccionar investigaciones relevantes y, posteriormente, recopilar y analizar datos de ellas. Además, nuestro trabajo se centra únicamente en el pronóstico espacial del delito, que es más limitado que la extracción de datos sobre el delito y más amplio que la vigilancia policial predictiva, como se analizó anteriormente. Por último, pretendemos contribuir con referencia científica para cuestiones técnicas en futuros estudios. Para lograr esto, seguimos un protocolo de revisión (sección “Métodos”), para responder seis preguntas de investigación (mencionadas en “Antecedentes”) que no han sido abordadas sistemáticamente en estudios anteriores.

MÉTODOS

Este estudio sigue la guía de presentación de informes “PRISMA” (Elementos de informes preferidos para revisiones sistemáticas y metanálisis) (Liberati et al. 2009). PRISMA sugiere una lista de verificación de 27 elementos relacionados con las secciones de una revisión sistemática de la literatura y su contenido, así como un diagrama de flujo de cuatro fases para la selección de artículos.

Adoptamos y modificamos la guía PRISMA de acuerdo con las necesidades de nuestro estudio. Nuestro diagrama de flujo contiene tres fases para la selección de artículos. La primera fase es la “identificación” e implica la selección de fuentes de información y una estrategia de búsqueda que arroja un conjunto de posibles artículos. La segunda fase consiste en “seleccionar” los artículos seleccionados de la primera fase y eliminar aquellos que no son relevantes para el alcance de la investigación. La tercera fase es la “elegibilidad”, en la que aplicamos una lectura más exhaustiva de los artículos y seleccionamos aquellos que son relevantes para nuestras preguntas de investigación. El recuento de artículos en cada fase y sus pasos posteriores se ilustran en la Fig. 1.

El número de artículos seleccionados en la fase de Identificación se basa en once palabras clave relacionadas con la predicción del delito (es decir, predecir el delito, predicción del delito, vigilancia policial predictiva, predicción del delito, predicción del delito, prevención del delito, extracción de datos sobre el delito, extracción de datos sobre el delito, estimación del delito), aprendizaje automático sobre delitos, big data sobre delitos). Además, agregamos siete términos más espacialmente explícitos (es decir, punto crítico del crimen, predicción espacial del crimen, modelado del terreno con riesgo de crimen, análisis espacial del crimen, modelado espacio-temporal del crimen, modelado espaciotemporal del crimen, crimen casi repetido). En un análisis posterior, hemos visualizado la frecuencia de palabras de los títulos de los artículos seleccionados como evidencia de la relevancia de las palabras clave.

Esto se puede encontrar en el archivo adicional 1 B: Una nube de palabras de alta frecuencia extraídas de los títulos de artículos seleccionados.

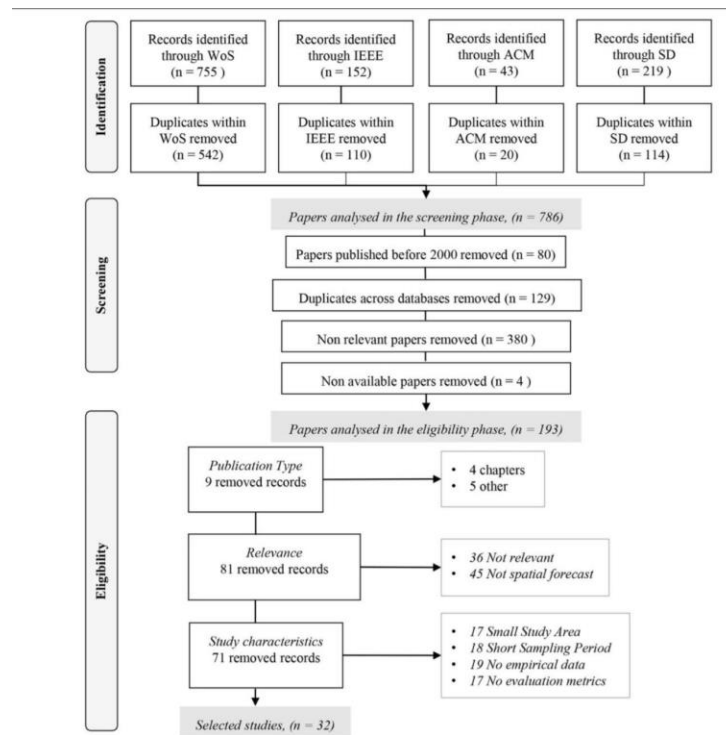


Fig. 1 Las tres fases del proceso de selección de estudios: identificación, cribado y elegibilidad.

A continuación, seleccionamos fuentes de información para realizar búsquedas bibliográficas. Aunque existe un número notable de motores de búsqueda y bases de datos académicas, nos centramos en bases de datos de investigación académicas e integrales que incluyen campos donde la predicción espacial de delitos es un tema representativo. Consideramos las siguientes bases de datos, incluida Web of Science de Clarivate Analytics (WoS), Instituto de Ingenieros Eléctricos y Electrónicos (IEEE).Xplore, ScienceDirect de Elsevier (SD) y la biblioteca digital de la Association for Computing Machinery (ACM). Consideramos que un proceso de búsqueda óptimo debe incluir múltiples bases de datos de búsqueda académica, realizándose las búsquedas con el mejor nivel de detalle posible. Además, como también discuten Bramer et al. (2017) en un estudio exploratorio de combinaciones de bases de datos, si la pregunta de investigación es más interdisciplinaria, es probable que una base de datos científica más amplia, como Web of Science, agregue valor.

Con respecto a Google Scholar (GS) existen opiniones divergentes entre los investigadores sobre si GS aporta información relevante para una revisión interdisciplinaria o no. Holone (2016) analiza que algunas búsquedas en motores, específicamente GS, tienen una tendencia a exponer información de forma selectiva mediante el uso de algoritmos que personalizan la información para los usuarios, llamándolo efecto burbuja de filtro. Haddaway et al. (2015) encontraron que cuando se buscaron artículos específicos, la mayoría de la literatura identificada usando Web of Science también se encontró usando GS. Sin embargo, sus hallazgos mostraron una superposición de moderada a pobre en los resultados cuando se

utilizaron cadenas de búsqueda similares en Web of Science y GS (10–67%), y que GS omitió alguna literatura importante en cinco de seis estudios de caso.

En cada base de datos, utilizamos palabras clave en versiones de palabras en singular y plural (por ejemplo, puntos críticos de delincuencia). Para WoS, utilizamos la opción de búsqueda avanzada, buscando artículos escritos en inglés y haciendo coincidir nuestras palabras clave con el tema o título. Para IEEE, buscamos nuestras palabras clave en los metadatos o en los títulos de los artículos. En SD y ACM, utilizamos la opción de búsqueda avanzada con funciones booleanas que buscaban nuestras palabras clave en el título, resumen o palabras clave del artículo. Los artículos identificados se integraron directamente en el gestor de referencias gratuito Mendeley. Por último, eliminamos duplicados dentro de cada base de datos, lo que resultó en 786 artículos para la segunda fase, la fase de selección. La última búsqueda en fase de Identificación se realizó el 5 de febrero de 2019.

Si bien el uso de análisis estadísticos y geoestadísticos para el pronóstico del delito se ha considerado durante bastante tiempo, durante las últimas dos décadas ha habido un interés creciente en desarrollar herramientas que utilicen grandes conjuntos de datos para hacer predicciones del delito (Perry 2013) . Por lo tanto, el análisis predictivo se ha incluido en las prácticas de aplicación de la ley (Brayne 2017). Esta es la razón principal por la que durante la fase de selección, primero excluimos los artículos publicados antes de 2000. En segundo lugar, eliminamos los duplicados en las cuatro bases de datos seleccionadas (WoS, IEEE, SD y ACM). En tercer lugar, revisamos todos los artículos para identificar los “no relevantes”. Esta decisión se tomó definiendo los artículos “relevantes” de modo que contuvieran los siguientes tres elementos.

El primer elemento es que un artículo aborde eventos delictivos con límites geográficos explícitos. Ejemplos comunes de artículos excluidos son los que tratan sobre el miedo al delito, las características de los delincuentes, las características de los delincuentes, las características de las víctimas, el perfil geográfico, el camino hacia el delito y los delitos cibernéticos o financieros. El segundo elemento para que un artículo sea “relevante” es que emplee un algoritmo de pronóstico y no se limite a un análisis exploratorio o de agrupamiento. El tercer elemento es que existe alguna forma de predicción espacial. Esto significa que existen unidades espaciales de análisis predefinidas, como la inferencia para cada bloque censal del área de estudio. Para los elementos de relevancia, nuestra estrategia fue la siguiente: (a) leer el título y las figuras y/o mapas de la pantalla, (b) si no está seguro de la relevancia, leer el resumen, (c) si aún no está seguro de la relevancia, buscar palabras relevantes (p. ej., geo*, ubicación, espacial) dentro del documento. El último paso de la fase de selección fue eliminar artículos relevantes a los que los autores no tuvieron acceso debido a restricciones de suscripción. La fase de selección resultó en 193 artículos relevantes que serán considerados para la tercera y última fase.

Durante esta fase final, la fase de Elegibilidad, leemos el resumen y el cuerpo principal de los 193 artículos (por ejemplo, área de estudio, datos, métodos y resultados). El objetivo fue extraer elementos de datos que componen los criterios de elegibilidad del artículo. Estos se agrupan en tres categorías, a saber: (a) tipo de publicación, que es el primer elemento de datos, (b) relevancia: consta de elementos de datos, relevancia y propósito, y (c) características del estudio: consta de elementos de datos, área de estudio, período de muestreo, datos empíricos, métricas de evaluación. A continuación, analizamos cada categoría y los elementos de datos que implica. El primer dato es el tipo de publicación. Las revisiones de la literatura a veces excluyen los artículos de conferencias porque su calidad no se evalúa como los artículos del International Scientific Indexing (ISI). Sin embargo, para algunas disciplinas, como la informática, muchas conferencias se consideran medios de publicación de gran reputación. En la fase de selección, encontramos una gran cantidad de artículos de académicos de informática o ciencias de la información, por lo que en esta etapa decidimos no excluir los artículos de conferencias ($n = 65$), pero también los artículos que no pertenecen al ISI ($n = 19$). En total, excluimos nueve artículos que son capítulos de libros o pertenecen a otras categorías (p. ej., editorial).

Los siguientes dos criterios de “relevancia” (es decir, relevancia y propósito) abordan la adecuación del contenido de los artículos a nuestro alcance de investigación. Durante esta fase se volvió a comprobar la relevancia del artículo. Por ejemplo, se descubrió que algunos artículos que parecían ser relevantes en la fase de selección (es decir, un artículo sobre eventos delictivos, espacio y pronóstico) no lo eran después de leer la parte central del artículo. Por ejemplo, la predicción se mencionó en el resumen, pero lo que los autores dieron a entender fue que la predicción es una perspectiva de investigación futura del análisis que realmente se realizó en el artículo. Además, agregamos el elemento de datos “propósito” para separar los métodos que modelan y exploran las relaciones entre las variables dependientes e independientes (por ejemplo, atractores de delitos para robos) de aquellos que realizan un pronóstico espacial. El número de trabajos que fueron excluidos por estos criterios ascendió a 81.

Por último, hay cuatro criterios más de “características del estudio” relevantes para la calidad y homogeneidad de los artículos seleccionados. Primero, el área de estudio debe ser igual o mayor que una ciudad. Las ciudades son menos propensas a sufrir efectos de borde en comparación con unidades administrativas más pequeñas dentro de una ciudad que comparten límites con otras unidades (por ejemplo, distritos). Además, cuanto más pequeña sea el área de estudio, más probable es que las conclusiones se adapten a las características del estudio y no sean escalables. En segundo lugar, el período de tiempo de la muestra de delitos debe ser igual o mayor a un año para garantizar que se capturen los patrones de estacionalidad. Estos dos ítems también aumentan la homogeneidad de los estudios seleccionados. Sin embargo, existen diferencias significativas entre los estudios que se analizan con más detalle, en la sección Resultados. Los dos últimos criterios son la restricción al análisis de datos empíricos (por ejemplo, se excluyeron pruebas de conceptos o artículos puramente metodológicos) y el uso de medidas que evalúen el desempeño de los modelos (por ejemplo, error cuadrático medio). Los

dos últimos criterios garantizan que sólo analicemos estudios que sean útiles para abordar nuestras preguntas de investigación. El número de artículos que se excluyeron debido al tipo de publicación, la relevancia o las características del estudio fue 71. Además, la Fig. 1 muestra el número de artículos excluidos para cada elemento de datos (por ejemplo, 17 artículos se excluyeron debido a insuficiencia de datos para el tamaño del área de estudio). Finalmente, todo el proceso de selección arrojó 32 trabajos.

Tabla 1 Elementos de datos analizados en diferentes etapas del estudio

Etapa de estudio	Elementos de datos
Identificación	Autores; año; título; fuente de datos
Poner en pantalla	Relevancia_1; disponibilidad
Elegibilidad	Tipo de publicación; datos empíricos; actuación evaluación; tamaño espacial; tamaño temporal; propósito, relevancia_2
Resultados "Características del estudio"	Año; título; disciplina; revista/conferencia; estudiar área país, institución
Resultados "Resumen de publicaciones seleccionadas sobre previsión espacial de delitos"	Área de estudio; escala; periodo de muestreo; meses; tipo; muestra; inferencia; tarea; unidad espacial; unidad temporal
Resultados "Métodos de previsión de delitos espaciales"	Método propuesto; mejor método propuesto; método de referencia; tipo de algoritmo propuesto; método propuesto entrada, variables
Resultados "Consideraciones al analizar el desempeño del pronóstico"	Métrica de evaluación; estrategia de validación

CALIDAD DE ESTUDIO

Dos de los cuatro autores de esta investigación realizaron la selección de los artículos a analizar. Antes de cada fase, estos dos autores discutieron y diseñaron el proceso, probaron muestras y dividieron la carga de trabajo.

Diez, los resultados se fusionaron, analizaron y discutieron hasta que ambos autores llegaron a un consenso para la siguiente fase. Los mismos dos autores verificaron varios de los resultados para asegurar la coherencia metodológica entre ellos. La lectura de los artículos durante la fase final (es decir, elegibilidad) se realizó dos veces, alternando las muestras de los artículos entre los dos autores, para garantizar que se incluyeran todos los artículos elegibles. Además, en caso de que alguna información sobre el contenido del documento no fuera clara para los dos autores, se comunicaron por correo electrónico con los autores correspondientes para obtener aclaraciones.

Con respecto a las subsecciones de resultados que constituyen cuatro etapas de estudio (secciones “Características del estudio”, “Resumen de publicaciones seleccionadas sobre pronóstico de delitos espaciales”, “Métodos de pronóstico de delitos espaciales” y “Consideraciones al analizar el desempeño de los pronósticos”), uno o dos autores realizaron todos y cada uno de los autores contribuyeron a extraer información y revisarla. Para extraer información estructurada como elementos de datos seguimos un procedimiento de tres pasos que se repitió en cada etapa.

Primero, los autores leyeron los artículos para extraer manualmente los elementos de datos y sus valores (1—extraer). Luego, los autores discutieron y verificaron dos veces los elementos de datos y sus valores (2: discusión/consenso). En caso de que la información aún no estuviera clara, nos comunicamos por correo electrónico con los autores correspondientes para obtener aclaraciones (3: consulta). Esta información se estructuró como una matriz donde las filas representan los artículos y las columnas son varios elementos de información procesada (por ejemplo, un elemento de datos es el año de publicación). La Tabla 1 muestra los elementos de datos en la etapa en la que fueron explotados. Los atributos (valores) de los ítems se analizan en la sección “Resultados”.

El riesgo de sesgo en los estudios individuales se evaluó mediante la escala del estudio. Se establecieron restricciones espaciales y temporales (ya definidas en la fase de elegibilidad) para garantizar que analicemos estudios de mediana y gran escala y que los hallazgos no estén vinculados a características específicas de localidad o estacionalidad. Además, no identificamos publicaciones duplicadas (es decir, dos o más artículos con las mismas muestras y experimentos) y no identificamos peculiaridades de los estudios, como características especiales y poco comunes o temas de investigación.

Por último, el riesgo de sesgo entre los estudios se evaluó mediante una encuesta en línea. Nos pusimos en contacto con los autores de las publicaciones (en algunos casos no pudimos identificar los datos de contacto) y les pedimos que respondieran a una breve encuesta sobre los resultados de su artículo. El correo electrónico introductorio definía el sesgo entre estudios como “El sesgo entre estudios puede resultar de la no publicación de estudios completos (sesgo de publicación) y de la publicación selectiva de resultados (es decir, informes selectivos dentro de los estudios) y es un riesgo importante de sesgo para una revisión sistemática y metanálisis”. Diez, explicamos el contenido de la encuesta que es para identificar, si hay resultados no reportados que sean considerablemente diferentes a los de sus trabajos. Esta información se evaluó mediante dos preguntas (para más detalles agregamos el cuestionario como archivo adicional 1 de este documento). De los 32 artículos, recibimos respuestas para 11 artículos (n = 12, con dos autores respondiendo al mismo artículo). Un factor que explica la baja tasa de respuesta es que muchos autores han cambiado de posición (los artículos se remontan a 2001) y para algunos no pudimos identificar sus nuevos datos de contacto, mientras que para otros recibimos varias respuestas por correo electrónico que no se entregaron.

En cuanto a los resultados de las respuestas, siete autores respondieron que nunca realizaron un estudio similar al que fueron contactados y cinco respondieron que habían realizado un estudio similar al que fueron contactados. Un estudio similar se definió como un estudio en el que: (a) el diseño del estudio, la selección de variables/predictores independientes, la selección de los métodos y la parametrización de los métodos son los mismos, y (b) los datos pueden ser diferentes. De aquellos que realizaron un estudio similar, cuatro respondieron que sus resultados no eran diferentes y uno respondió que sus resultados eran considerablemente diferentes. Sin embargo, en una respuesta explicativa de seguimiento, este autor respondió que cambiar la parametrización produjo resultados diferentes con respecto a la clasificación de desempeño de tres algoritmos y que los datos y el área de estudio eran los mismos. Según esta encuesta a pequeña escala, no hay indicios de que exista riesgo de sesgo entre los estudios. Sin embargo, es necesario seguir investigando este asunto.

Tabla 2 Un resumen de las características generales de los artículos como revista o congreso, país de área de estudio, institución y disciplina científica del primer autor.

Top 5 de revistas o congresos (nº de artículos)	Top 3 países (recuento)	
Revista internacional de previsión (3)	Estados Unidos (23)	
Geografía Aplicada (2)	Brasil (2)	
Revista europea sobre política e investigación criminal (2)	Reino Unido (2)	
Ciencia de datos de HCE (2)		
Conferencia Internacional sobre Sistemas, Hombre y Cibernética (2)		
Top 4 instituciones (nº de artículos)	4 disciplinas principales (n.º de artículos)	
Universidad de Massachusetts Boston, EE. UU. (3)	Informática, Sistemas de información (12)	Administración pública (3)
Universidad Carnegie Mellon, EE. UU. (2)	Criminología y Penología (5)	Geociencias, Multidisciplinario (2)
Universidad Federal de Rio Grande do Norte, Brasil (2)		
Universidad de Indiana: Universidad Purdue de Indianápolis, EE. UU. (2)		

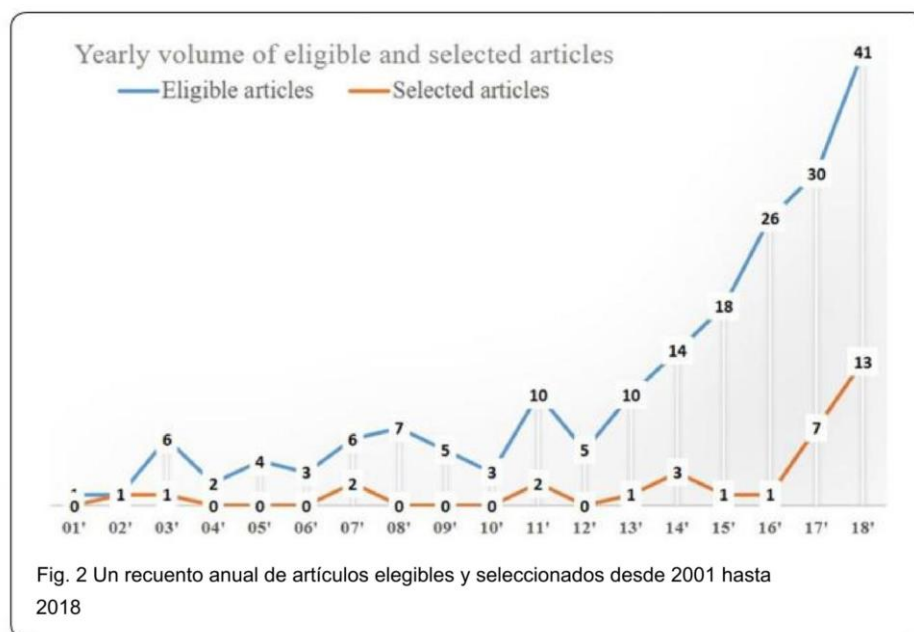
RESULTADOS

Características del estudio

En esta sección discutimos las características genéricas de los artículos seleccionados. Para empezar, el tipo de publicación es ligeramente mayor para los artículos de revistas ISI (n=18) que para los artículos de congresos (n=14). Los 32 artículos se publicaron en una variedad de revistas y congresos y no se observó preferencia por un medio de publicación en particular. En concreto, cuatro revistas y un congreso publicaron dos o tres de los artículos seleccionados cada uno (Tabla 2) y todos los demás artículos se publicaron en diferentes revistas y congresos. Por otro lado, existe poca variación respecto al país del área de estudio. La mayoría de los estudios se realizaron en los EE. UU., lo que probablemente sea una estadística algo sesgada, considerando el gran tamaño de la población estadounidense, así como el idioma utilizado (p. ej., inglés) en el proceso de selección de estudios. De manera similar, las instituciones que han publicado más de un artículo sobre pronóstico espacial de delitos tienen su sede en Estados Unidos, con excepción de la Universidad Federal de Rio Grande do Norte, Brasil, que tiene publicaciones recientes en este campo.

También recogimos la disciplina asociada a cada artículo. Para hacerlo utilizamos la afiliación del primer autor y extrajimos la disciplina hasta el nivel de departamento, si esto fuera posible. Diez utilizamos como referencia las 236 categorías/disciplinas utilizadas en Journal Citation Reports (JCR)¹ por el Web of Science Group. Luego, cada afiliación de autores se relacionó con una de las categorías. En la Tabla 2, vemos las disciplinas que aparecieron más de una vez (es decir, ciencias de la computación, criminología, administración pública y geociencias). Aunque recopilamos una variedad de disciplinas, estas son las que encontramos más de una vez y representan la mayoría de los artículos (n = 22). Por lo tanto, los estudiosos de estas disciplinas parecen tener un mayor interés en la prevención espacial del delito.

La Figura 2 muestra el número de artículos elegibles y seleccionados por año durante el período de selección del estudio. Incluimos los artículos elegibles además de los seleccionados por dos razones. En primer lugar, muchos de los artículos elegibles analizaron la prevención espacial de la delincuencia, pero no cumplieron con los criterios definidos para este estudio. En segundo lugar, es posible que otros artículos no sean relevantes para la predicción, pero sí lo son para temas más amplios de predicción o modelización. El gráfico de la Fig. 2 muestra una tendencia de rápido aumento en los últimos años. Para los artículos elegibles, el número de artículos aumentó sustancialmente desde 2013, mientras que para los artículos seleccionados, una tendencia similar es evidente en los últimos 2 años.



Resumen de publicaciones seleccionadas sobre prevención de delitos espaciales

¹ JCR: <https://clarivate.com/products/scientific-and-academic-research/research-analytics-evaluation-and-management-solutions/journal-citation-reports/>

En la Tabla 3 incluimos cada artículo seleccionado junto con información relacionada con el espacio (es decir, área de estudio y escala espacial), tiempo (es decir, período de muestreo y período en meses), datos sobre el delito (es decir, tipo de delito y tamaño de la muestra del delito). y predicción (es decir, información predicha, tarea, unidad espacial y unidad temporal).

En esta sección, consideramos estos 10 elementos de datos como información inicial y básica al momento de informar un estudio espacial de pronóstico de delitos. Un lector que quiera replicar o modificar el enfoque metodológico presentado en la investigación de seguimiento necesitará los mismos 10 elementos de datos para evaluar si dicho enfoque es adecuado para el estudio de seguimiento y los objetivos de la investigación del autor.

Más importante aún, cuando falta alguno de estos elementos de datos, la evaluación de la generalización (o sesgo) de las conclusiones y la interpretación de los resultados es limitada. Desafortunadamente, la mayoría de los 32 artículos seleccionados (n=21) tenían al menos un ítem con información indefinida o poco clara para cinco de los 10 ítems de datos (Fig. 3). De estos, el 52% (n=11) fueron artículos de congresos y el 48% (n=10) fueron artículos de ISI. Por otro lado, el 73% (n=8) de los artículos sin información indefinida o poco clara fueron artículos ISI y el 27% (n=3) fueron conferidos papeles de referencia.

La mayoría de los estudios se realizaron a nivel de ciudad. En dos estudios, el área de pronóstico cubrió un condado, que es la unidad administrativa estadounidense que generalmente se expande a través de los límites de una ciudad. En un artículo, las predicciones cubrían un país entero (EE.UU.). La ciudad de Nueva York (EE.UU.) fue examinada en cuatro estudios, Pittsburgh (EE.UU.) fue examinada en tres estudios y Portland (EE.UU.), Natal (Brasil) y Chicago (EE.UU.) fueron examinadas en dos estudios. Todas las demás publicaciones se basaron en áreas de estudio individuales.

Los datos sobre delincuencia más antiguos que se utilizaron en los 32 artículos son del año 1960 y los datos sobre delincuencia más recientes son del año 2018. El período de muestreo varía desde 1 año hasta 50 años. Hay un artículo con un período de muestreo indefinido. Sin embargo, del artículo se puede inferir que la duración del período de muestreo es de al menos 1 año. En cuanto al tamaño de la muestra de los datos sobre delitos, oscila entre aproximadamente 1.400 incidentes y 6,6 millones, lo que es relevante tanto para el número de tipos de delitos como para la duración del período de muestreo. En cuanto al número de tipos de delitos, hay cuatro estudios que agregaron y analizaron todos los tipos de delitos juntos, doce estudios que se centraron en un tipo de delito en particular, catorce estudios que analizaron desde dos hasta 34 tipos de delitos diferentes, y tres estudios con información indefinida sobre el tipo de delito analizado. El robo con allanamiento de morada fue el tipo de delito que se examinó con mayor frecuencia en estudios que analizaron sólo un tipo de delito.

Los últimos cuatro elementos de datos de la Tabla 3 describen detalles de la información pronosticada, a la que nos referimos como “inferencia”. La unidad temporal es la granularidad

temporal de la inferencia y varía desde una granularidad fina de 3 h hasta 1 año. La unidad temporal más frecuente en todos los artículos es 1 mes (utilizada en 12 artículos). Además, el día y la semana se han utilizado en ocho estudios y el año en siete estudios.

Otras unidades temporales menos frecuentes son 3 h, día durante 1 mes, noche durante 1 mes, 2 semanas, 2 meses, 3 meses y 6 meses. De manera similar, la unidad espacial es la granularidad espacial de la inferencia y varía desde un área pequeña con un tamaño de celda de cuadrícula de $75\text{ m} \times 75\text{ m}$ hasta un área grande, como comisarías de policía o incluso países. La unidad espacial es difícil de analizar y comparar por dos razones. En primer lugar, las unidades espaciales no tienen un formato estándar como el tiempo y son bidimensionales. Así, pueden variar en tamaño y forma. En segundo lugar, en aproximadamente un tercio de los artículos esta información se informó de manera deficiente (Fig. 3). En el caso de unidades administrativas (por ejemplo, bloques o distritos censales), la forma y el tamaño suelen variar, pero si alguien busca más detalles o las unidades mismas, en la mayoría de los casos las autoridades administrativas pueden recuperarlos. Sin embargo, los conglomerados espaciales también pueden variar en forma y tamaño, pero si no se informan los detalles (por ejemplo, la dirección de las elipses, el rango de tamaño de los conglomerados, el tamaño promedio de los conglomerados) es difícil cuantificar y replicar dichos conglomerados. También encontramos casos en los que los autores informan las dimensiones del tamaño de una celda de una cuadrícula sin mencionar las unidades de medida. Sin embargo, la celda de la cuadrícula parece ser el tipo preferible de unidad espacial y se utiliza en la mayoría de los artículos ($n=20$).

Los elementos de datos “inferencia” y “tarea” se refieren a los tipos de información pronosticada y tarea predictiva, respectivamente. La inferencia y la tarea se definen según la información que los autores evaluaron y no según el resultado de un algoritmo de predicción. Por ejemplo, un algoritmo puede derivar la intensidad del crimen en el espacio (es decir, la salida del algoritmo), que los autores utilizaron para extraer puntos críticos (es decir, salida procesada para ser evaluada) y luego evaluar sus resultados usando una medida de clasificación como precisión, exactitud u otras. Algunos métodos predictivos, como el bosque aleatorio, se pueden utilizar tanto para tareas de clasificación como desregresión. No está claro por qué algunos autores optan por aplicar una aplicación de regresión de un método y luego procesar, derivar y evaluar un resultado de clasificación, aunque podrían hacerlo aplicando directamente una aplicación de clasificación del mismo método. Además, la inferencia “puntos críticos” en la Tabla 3 incluye las siguientes cuatro categorías:

1. Los puntos críticos y los no críticos se definen utilizando un enfoque estadístico que separa el espacio entre áreas de alta y baja criminalidad (2020) 9:7
2. Los puntos críticos y los no críticos se definen utilizando un umbral arbitrario que separa el espacio entre áreas de alta y baja criminalidad.
3. El delito y la no delincuencia se definen utilizando un enfoque estadístico que separa el espacio entre las áreas donde es probable que ocurra el delito y las áreas donde es poco probable que ocurra el delito.

4. El delito y la no delincuencia se definen utilizando un enfoque estadístico que separa el espacio entre áreas donde hay al menos un delito estimado y áreas donde no hay ningún delito estimado.

Table 3 An overview of the 32 selected papers with information about the space and time of the research, the crime data, and forecasting details

No*	Authors and date	Space		Time		Crime Data			Forecasting			
		Study area	Scale	Sampling period	Months	Type	Sample	Inference	Task	Spatial unit	Temporal unit	
1	Araujo Junior et al. (2017)	Natal, Brazil	City	2006–2016	132	<i>U</i>	<i>U</i>	# of crimes	Regression	Rectangular grid (<i>U</i>), districts	Week	
2	Araujo et al. (2018)	Natal, Brazil	City	2006–2016	132	<i>U</i>	<i>U</i>	Hotspots	Binary classification	k-means cells of varying size (<i>U</i>)	Week	
3	Bowen et al. (2018)	DeKalb, USA	County	2011–2014	48	Violent crime	<i>U</i>	Hotspots	Binary classification	Census block groups	Month	
4	Brown and Oxford (2001)	Richmond, USA	City	1994–1999	72	Break and enter	≈ over 24,000	# of crimes	Regression	Grid cells of 1.66 km ² , precincts	Week, month	
5	Cohen et al. (2007)	Pittsburgh, USA	City	1991–1998	96	2 crime types	1.3 million	# of crimes	Regression	1219 m × 1219 m grid cells	Month	
6	Dash et al. (2018)	Chicago, USA	City	2011–2015	60	34 crime types	6.6 million	# of crimes	regression	Communities	Month, year	
7	Drawve et al. (2016)	Little Rock, USA	City	2008–2009	18	Gun crime	1429	Hotspots	Binary classification	91 m × 91 m grid cells	6 months	
8	Dugato et al. (2018)	Milan, Italy	City	2012–2014	36	Residential burglary	20,921	Hotspots	Binary classification	Grid cells of 2500 m ²	Year	
9	Gimenez-Santana et al. (2018)	Bogota, Colombia	city	2012–2013	24	3 crime types	<i>U</i>	Hotspots	Binary classification	75 m × 75 m grid cells	Year	
10	Gorr et al. (2003)	Pittsburgh, USA	City	1991–1998	96	5 crime types	≈ 1 million	# of crimes	Regression	Police precincts	Month	
11	Hart and Zandbergen (2014)	Arlington, USA	City	2007–2008	24	4 crime types	6295	Hotspots	Binary classification	Grid cells of 3 different sizes (<i>U</i>)	Year	
12	Hu et al. (2018)	Baton Rouge, USA	City		2011	12	Residential burglary	3706	Hotspots	Binary classification	100 m × 100 m grid cells	Week
13	Huang et al. (2018)	New York, USA	City		2014	12	4 crime types	103,913	Category of crime	Binary classification	Districts	Day, month
14	Ivaha et al. (2007)	Cardiff, UK	City	2001–2003	26	Criminal damage	<i>U</i>	Percent of crime in clusters	Regression	Clusters of varying size (<i>U</i>)	Day	
15	Johansson et al. (2015)	Sweden three cities: Stockholm, Gothenburg, and Malmo	Cities	2013–2014	12	Residential burglary	5681	Hotspots	binary Classification	Grid cells (<i>U</i>)	3 months	
16	Kadar and Pletikosa (2018)	New York, USA	City	2014–2015	24	All crime and 5 crime types	174,682	# of crimes	Regression	Census tract	Year	
17	Liesenfeld et al. (2017)	Pittsburgh, USA	City	2008–2013	72	All crime	9936	# of crimes	Regression	Census tracts	Month, year	
18	Lin et al. (2018)	Taoyuan City, Taiwan	City	2015–2018	39	Motor vehicle thefts	≈ 8580	Hotspots	Binary classification	5 to 100 × 5 to 100 grid cells (<i>U</i>)	Month	

Table 3 (continued)

No*	Authors and date	Space		Time		Crime Data			Forecasting		
		Study area	Scale	Sampling period	Months	Type	Sample	Inference	Task	Spatial unit	Temporal unit
19	Malik et al. (2014)	Tippecanoe, USA	County	2004–2014	120	all crime	≈ 310,000	Hotspots	Binary classification	Grid cells (<i>U</i>), law beats, census blocks	Week
20	Mohler (2014)	Chicago, USA	City	2007–2012	72	2 crime types	78,852	Hotspots	Binary classification	75 m × 75 m, 150 m × 150 m grid cells	Day
21	Mohler and Porter (2018)	Portland, USA	City	2012–2017	60	4 crime types	<i>U</i>	Hotspots	Binary Classification	Grid cells of 5806 m ²	Week, 2 weeks, month, 2 months, 3 months
22	Mohler et al. (2018)	Indianapolis, USA	City	2012–2013	24	4 crime types	<i>U</i>	Hotspots	Binary classification	300 m × 300 m grid cells	Day
23	Mu et al. (2011)	Boston, USA	City	2006–2007	24	Residential burglary	<i>U</i>	Hotspots	Binary classification	20 × 20 grid cells (<i>U</i>)	Month
24	Rodriguez et al. (2017)	San Francisco, USA	City	2003–2013	120	Burglary	<i>U</i>	Properties of clusters	Regression	Clusters (<i>U</i>)	Day
25	Rosser et al. (2017)	"Major city", UK (<i>U</i>)	City	2013–2014	13	Residential burglary	5862	Hotspots	Binary classification	Street segments (<i>U</i>)	Day
26	Rumi et al. (2018)	Brisbane, Australia; New York City, USA	Cities	2012–2013 (AUS); 2012–2013 (USA)	9 and 12	6 crime types	<i>U</i>	Hotspots	Binary classification	Census regions	3 h
27	Rummens et al. (2017)	"Large city", Belgium (<i>U</i>)	city	2011–2014	48	3 crime types	163,800	Hotspots	Binary classification	200 m by 200 m grid cells	2 weeks, daytime month, night time month
28	Shoesmith (2013)	USA	Country	1960–2009	600	2 crime types	<i>U</i>	Crime rate	Regression	USA regions	Year
29	Yang et al. (2018)	New York, USA	city	January 2014–April 2015	16	7 crime types	<i>U</i>	Hotspots	Binary classification	0.01 latitude × 0.01 longitude grid cell size	Day, week, month
30	Yu et al. (2011)	"City in the Northeast", USA (<i>U</i>)	City	<i>U</i>	<i>U</i>	Residential burglary	<i>U</i>	Hotspots	binary Classification	grid cells (<i>U</i>)	Month
31	Zhao and Tang (2017)	New York, USA	City	2012–2013	12	<i>U</i>	<i>U</i>	# of crimes	Regression	2 km × 2 km grid cells	Day, week
32	Zhuang et al. (2017)	Portland, USA	City	March 2012–December 2016	58	All crime	<i>U</i>	Hotspots	Binary classification	183 m × 183 m grid cells	2 weeks

U = undefined or unclear information

* No: Reference number of the paper that are used in Fig. 5

En cuanto a las categorías tres y cuatro, algunos autores se refieren a estas áreas como puntos críticos y otros no. Agrupamos las cuatro categorías y las definimos como puntos críticos y no puntos críticos, lo que representa el resultado de una clasificación binaria que separa el

espacio en áreas para el patrullaje policial que son alarmantes y no alarmantes. Reconocemos que en el campo del análisis espacial de la delincuencia, los puntos críticos son áreas de alta intensidad delictiva. Sin embargo, en nuestros artículos seleccionados no parece haber una definición clara del término "punto de acceso".

La mayoría de los artículos (n=20) infirieron los puntos críticos como resultado de una clasificación binaria. Nueve estudios infirieron el número de delitos o la tasa de criminalidad en cada unidad espacial. Sin embargo, tres estudios parecen ser de alguna manera diferentes y únicos del resto. Huang et al. (2018) evaluaron la categoría de delito pronosticada como el resultado de un problema de clasificación binaria (por ejemplo, si la categoría A está presente en el área X; sí o no) Ivaha et al. (2007) dedujeron el número total de delitos en una región, los grupos espaciales (o puntos críticos) y la proporción de delitos dentro de cada grupo. Por último, Rodríguez et al. (2017) evaluaron las propiedades (es decir, ubicación y tamaño) de los grupos inferidos.

Métodos de prevención de delitos espaciales

Los primeros tres elementos de datos que se extrajeron para ser analizados en esta sección son el método de pronóstico propuesto, el mejor método de pronóstico propuesto y el método de pronóstico de referencia. Este último es el método utilizado como medida de comparación del método propuesto. Analizamos la frecuencia de los métodos para cada uno de los tres tipos de pronóstico. El mejor método de pronóstico propuesto es el que tiene mejor desempeño a lo largo de los experimentos realizados. Por ejemplo, si un experimento se evalúa por separado en múltiples tipos de delitos, solo contamos el método con el mejor desempeño en la mayoría de los casos. En caso de que dos métodos funcionen de manera similar (como lo demuestran los resultados estadísticos y lo informado por los autores de los artículos), se consideran ambos métodos. Esto fue necesario porque algunos artículos propusieron más de un método para compararlo con un método de referencia, pero al final, estos artículos proponen el método con mejor desempeño. Además, esto reduce los recuentos de frecuencia sesgados de los métodos propuestos. Por otro lado, consideramos como línea de base el método con el que los autores querían comparar los métodos propuestos. Por ejemplo, Zhuang et al. (2018) propusieron tres redes neuronales profundas y utilizaron seis algoritmos de aprendizaje automático adicionales como métodos de referencia para evaluar en qué medida eran mejores los métodos propuestos en comparación con los métodos de referencia. En este caso, contamos los seis algoritmos de aprendizaje automático como métodos de referencia.

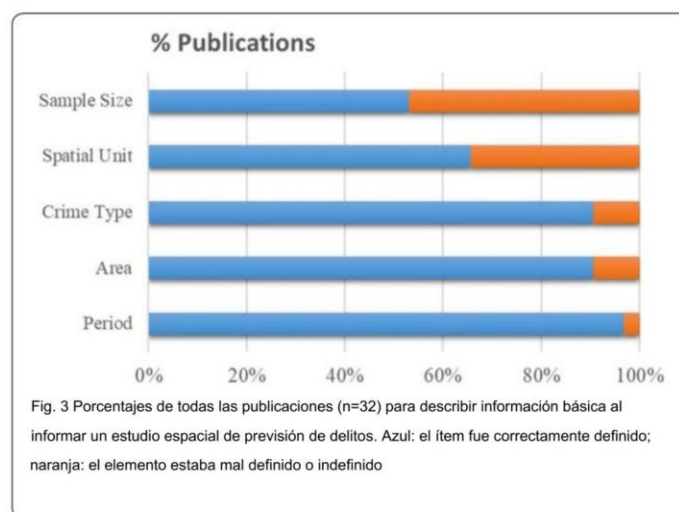
En la Tabla 4, mostramos los métodos "principales" (es decir, contados con frecuencia dentro de los 32 artículos seleccionados) por cada elemento. El bosque aleatorio (RF) es el método propuesto más utilizado. El perceptrón multicapa (MLP) aparece como el método principal en los tres elementos (es decir, propuesto, mejor, línea de base).

Otros métodos mejores propuestos son los basados en la estimación de densidad del núcleo (KDE) y el modelado de terreno de riesgo (RTM).

Curiosamente, las máquinas de vectores de soporte (SVM) se han propuesto en cinco artículos, pero no se encuentran entre los cuatro métodos mejor propuestos. Por otro lado, se prefieren muchos modelos estadísticos bien conocidos como métodos de referencia, como el modelo autorregresivo (AR), la regresión logística, el modelo autorregresivo integrado de media móvil (ARIMA) y la regresión lineal, así como KDE. -basado y K vecinos más cercanos. En el archivo adicional 1 C agregamos tablas detalladas que muestran para cada artículo los elementos de datos del método propuesto, el mejor método propuesto y el método de referencia.

En las siguientes secciones, clasificamos los métodos de pronóstico propuestos por tipo de algoritmo (sección "Tipo de algoritmo de los métodos de pronóstico propuestos") y por tipo de entradas que toman (sección "Entradas del método propuesto").

Esta tarea fue un desafío porque no existe un consenso científico sobre una taxonomía única o una categorización de técnicas y métodos analíticos. Papamitsiou y Economides (2014) revisaron artículos sobre análisis educativo, categorizando los métodos de minería de datos en clasificación, agrupamiento, regresión, minería de textos, minería de reglas de asociación, análisis de redes sociales, “descubrimiento con modelos”, visualización y estadísticas. Otros investigadores resumirían todas estas categorías, por ejemplo, como aprendizaje supervisado, aprendizaje no supervisado y análisis exploratorio de datos. Vlahogianni et al. (2014) utilizan diferentes categorizaciones para los artículos revisados en pronóstico de tráfico, incluidos aspectos relacionados con el enfoque del modelo para tratar las entradas y otras propiedades relevantes para dividir las metodologías propuestas. La granularidad adecuada de las propiedades para definir una categorización útil puede ser problemática y particular para cada campo.



Tipo de algoritmo de los métodos de pronóstico propuestos

Otra característica adecuada para clasificar los métodos de predicción son las similitudes entre los algoritmos.

Dividimos todos los algoritmos utilizados en los artículos informados en (i) basado en kernel (ii) proceso puntual, (iii) aprendizaje automático tradicional y (iv) aprendizaje profundo, de acuerdo con los siguientes criterios. Los algoritmos basados en kernel se preocupan particularmente por encontrar una curva de tasa de criminalidad para cada lugar g que se ajuste a un subconjunto de puntos de datos dentro de los límites de un kernel dado (ver Ec. 1).

Observamos que la principal diferencia entre los algoritmos basados en kernel es el uso de diferentes tipos de kernel. Hart y Zandbergen (2014) experimentaron con diferentes tipos de núcleos y arrojaron algunas conclusiones útiles. En nuestros artículos seleccionados, seis de ellos han utilizado algoritmos basados en kernel y el más utilizado es la estimación de densidad de kernel (KDE) bidimensional simple ($n = 2$). Sin embargo, observamos que algunos métodos son una variación del modelo KDE simple, en la forma del KDE espaciotemporal (STKDE) utilizado en el artículo de Hu et al. (2018), el KDE Network-Time (NTKDE) propuesto por Rosser et al. (2017), o el KDE de covarianza dinámica (DCKDE) propuesto por Malik et al. (2014). También hemos considerado el método de suavizado exponencial utilizado en el artículo de Gorr et al. (2003) como un algoritmo basado en kernel, ya que trabaja con una función de ventana (kernel) en la agregación de series de tiempo.

$$\lambda_g(t) = \sum_{t < t_i} \kappa_g(t - t_i)$$

Los procesos puntuales se pueden distinguir de los algoritmos basados en kernel en la medida en que existe un factor de velocidad de fondo μ que se puede calcular estocásticamente, como con un proceso de Poisson. El factor de tasa de fondo incluye el modelado de covariables o características del lugar g , como variables demográficas, económicas, geográficas, etc. (ver Ec. 2). A partir de la explicación de Mohler (2014), suponemos que la introducción de la tasa de fondo hace que el proceso puntual sea más adecuado para el modelado multivariado en comparación con los algoritmos basados en kernel. En los artículos revisados, los algoritmos se pueden distinguir entre sí basándose en sus formulaciones matemáticas de κ y μ , pero también en su selección de parámetros internos, principalmente basados en la maximización de probabilidad. Solo tres artículos propusieron dicho algoritmo, incluido el proceso de punto marcado de Mohler (2014), el muestreo de importancia eficiente de máxima verosimilitud (ML-EIS) de Liesenfeld et al. (2017), y el Proceso Hawkes de Mohler et al. (2018).

En el caso de los algoritmos de aprendizaje automático, su formulación suele estar asociada con la búsqueda de una función f que asigna vectores de características X a una salida dada Y .

Estos algoritmos se distinguen entre sí en la forma en que se estima esta función, siendo algunos más precisos y complejos que otros. Incluimos en esta categoría todos los algoritmos que están explícitamente asociados con la regresión o clasificación. Se diferencian de los algoritmos de categorías anteriores porque f se construye sólo después de un proceso de entrenamiento. Este paso de entrenamiento tiene como objetivo encontrar un modelo que minimice el error de estimación entre la salida predicha y la salida original. La mayoría de los artículos publicados ($n=20$) se incluyeron en esta clase de algoritmos. Los algoritmos tradicionales de aprendizaje automático más propuestos fueron RF y MLP (empatados en $n=6$), seguidos de SVM junto con la regresión logística ($n=4$) y la regresión binomial negativa utilizada en estudios de RTM junto con ($n=3$).

Aunque los algoritmos de aprendizaje profundo tienen la misma formulación que los algoritmos tradicionales de aprendizaje automático, presentan una estructura interna mucho más compleja que afecta su uso. La estructura de capas profundas hace que el presupuesto computacional sea necesario principalmente durante el entrenamiento.

Además, la necesidad de muestras también es mayor que para los otros enfoques. Entre los artículos publicados, los tres que han utilizado este tipo de algoritmo sostienen que tiene el mejor rendimiento general. Esto incluye las redes neuronales profundas (DNN) instaladas por Lin et al. (2018), el marco DeepCrime de Huang et al. (2018) y la arquitectura de memoria a largo plazo (LSTM) propuesta por Zhuang et al. (2017). El artículo de Huang et al. (2018) incluso presenta una arquitectura neuronal dedicada a un enfoque independiente de las características, con una capa recurrente dedicada a codificar las dependencias temporales directamente de los sucesos delictivos. Aún así, ninguno de estos artículos ha discutido el rendimiento del tiempo computacional frente a otros algoritmos, ni tamaños de muestra suficientes para obtener modelos precisos. Al momento de escribir este artículo, sostenemos que no existe una guía clara sobre cuándo se debe llevar a cabo un enfoque de redes neuronales profundas, aunque en los últimos años ha comenzado a surgir evidencia de su efectividad.

Tabla 4 Los cuatro métodos principales propuestos, mejores propuestos y de referencia aplicados en los 32 artículos seleccionados

Los 4 principales métodos propuestos (n.º de artículos)	Los 4 mejores métodos propuestos (n.º de artículos)	Los 4 principales métodos de referencia (n.º de artículos)
Bosque aleatorio (7)	Bosque aleatorio (5)	Basado en modelos autorregresivos (5)
Perceptrón multicapa (6)	Perceptrón multicapa (5)	Regresión logística (3)
Basado en la estimación de la densidad del grano (5)	Basado en la estimación de la densidad del grano (5)	Media móvil integrada autorregresiva, perceptrón multicapa, regresión lineal, basado en KDE,
Máquinas de vectores de soporte (5)	Modelado del terreno de riesgo (3)	KNN: (2)

Entrada del método propuesto

Otro factor de división son los insumos de los métodos de pronóstico, es decir, las variables independientes. Hay algunos métodos de pronóstico que aceptan como entrada la latitud, la longitud y la marca de tiempo de los eventos delictivos (datos sin procesar), mientras que otros necesitan aplicar agregaciones o transformaciones explícitas antes de alimentar sus modelos. En este artículo, nos referimos como ingeniería de características al proceso de elaboración, escalamiento y selección de covariables o características para explicar mejor una variable de

predicción que a menudo requiere un conocimiento considerable de un dominio específico. Un ejemplo es la agregación de acontecimientos delictivos en series espaciotemporales, que pueden descomponerse en rezagos autorregresivos y utilizarse como características. Esta ingeniería de características también se puede aplicar a metodologías de datos auxiliares que no están directamente relacionadas con el crimen. Por ejemplo, Lin et al. (2018) cuentan los puntos de referencia en la cuadrícula contando el número de elementos en cada celda (agregación espacial) y crean una nueva característica para cada tipo de punto de referencia, mientras que Huang et al. (2018) definen que una parte de su algoritmo es una capa de incrustación de región para procesar solo la ubicación sin procesar de los puntos de referencia de la ciudad. Creemos que el factor de división por entradas del método puede ser información útil para un investigador potencial que desee realizar pronósticos espaciales y consulte esta sección de nuestro artículo. El procesamiento de datos requiere conocimiento del dominio y es una tarea costosa (en términos de tiempo), especialmente cuando se trata de grandes conjuntos de datos espaciotemporales. Por lo tanto, puede ser preferible evitar el proceso de ingeniería de características de algunos investigadores. Por otro lado, es posible que se prefiera utilizar datos para derivar sus variables con patrones particulares.

A los métodos que tienen un enfoque interno para agregar eventos delictivos en variables espaciotemporales los llamamos “independientes de la ingeniería de características” y “dependientes de la ingeniería de características”. En otras palabras, estos métodos necesitan explícitamente agregaciones para derivar variables espaciotemporales a partir de los datos sin procesar, independientemente del algoritmo de pronóstico. La mayoría (n=24) de los artículos reportados tienen una forma explícita de transformar sus eventos delictivos, así como datos auxiliares, en características para alimentar sus algoritmos (es decir, dependientes de la ingeniería de características). Aunque hemos encontrado muchas formas diferentes de agregación de datos en características, tanto espacial como temporalmente, el procedimiento de asignación de características a menudo no se informa lo suficiente, lo que dificulta la reproducción de la metodología propuesta.

Aun así, en Malik et al. se detallan flujos de trabajo o marcos de trabajo bien definidos seguidos de métodos dependientes de la ingeniería de características. (2014) y Araújo et al. (2018). Sintetizaron sus métodos de pronóstico en (1) agregar datos brutos espacialmente, siguiendo una metodología de mapeo del crimen (por ejemplo, contando eventos dentro de celdas de la cuadrícula), (2) generar series de tiempo y sus características, (3) usar un modelo de pronóstico usando un algoritmo, y (4) visualizar los resultados. En los métodos dependientes de ingeniería de características, la agregación y la generación de series de tiempo se realizan por separado como pasos de procesamiento antes de ajustar un modelo, mientras que esto no es necesario para los métodos independientes de ingeniería de características.

Consideraciones al analizar el desempeño del pronóstico

En esta sección, analizamos las medidas de desempeño de pronóstico (sección "Descripción general de las métricas de evaluación") y analizamos cuáles se utilizan para cada tarea de

pronóstico, incluidas las de clasificación y regresión (sección "Métricas por tarea de pronóstico"). Décimo, exploramos estrategias de validación por tipos de algoritmos (sección "Algoritmos y estrategias de validación"). Finalmente, resumimos y discutimos las principales dependencias y limitaciones de las subsecciones anteriores (sección "Dependencias y limitaciones").

Descripción general de las métricas de evaluación

Como se menciona en la sección "Métodos de pronóstico de delitos espaciales" , los artículos seleccionados incluyen datos de referencia para el pronóstico, modelos, modelos novedosos o modelos de conjunto propuestos por los respectivos autores. Las métricas de evaluación de tales modelos son, en general, bien conocidas en criminología, ciencia SIG, matemáticas o estadística. Sin embargo, es importante mencionar que pocos autores destacan la necesidad de combinar o utilizar diversas métricas de evaluación.

No podemos hacer una comparación de todos los resultados de la evaluación de los 32 artículos debido a diversas razones, como diferentes unidades espaciales y temporales, áreas de estudio o métodos de pronóstico aplicados. Sin embargo, podemos discutir ciertas similitudes entre ellos. La elección de una métrica de evaluación depende en gran medida del resultado principal de la predicción, como los recuentos (por ejemplo, para una distribución de Poisson), valores o tasas normalizados (por ejemplo, para una distribución gaussiana continua) o clasificación binaria (crimen). está ausente o presente). Las métricas de evaluación más frecuentes utilizadas en los artículos seleccionados son la precisión de predicción (PA, n=9), seguida del índice de precisión de predicción (PAI, n=7), la puntuación F1 (n=7), la precisión y la recuperación (n=5), el error cuadrático medio (MSE, n=4), el error cuadrático medio (RMSE, n=3), R-cuadrado (n=3), el índice de tasa de recaptura (RRI, n=3) , la tasa de aciertos (n=2), el área bajo la curva (AUC, n=2) y el error de pronóstico absoluto medio (MAFE, n=2). Algunas métricas adicionales se utilizan solo una vez, a saber, la estimación cuadrática media espacio-temporal (STMRSE), el RMSE promedio (RMSE), la puntuación de precisión deregresión (RAS), la puntuación de precisión de regresión (RPS), el LjungBox prueba, el error absoluto medio (MAE), el error porcentual absolutomedio (MAPE), macro-F1, micro-F1, el error de pronóstico medio (cuadrado) (M(S)FE), el coeficiente de correlación de Pearson y el coeficiente de Nash. Generalmente, las métricas derivadas de la matriz de confusión, a saber, exactitud, precisión, recuperación y puntuación F1, se utilizan juntas para evaluar clasificaciones binarias.

Analizamos las tres métricas de evaluación principales (PA, PAI, F1-Score) en relación con su distribución entre los elementos de datos de la disciplina, el tipo de algoritmo de pronóstico propuesto, la inferencia de pronóstico, la tarea de pronóstico, la unidad espacial y la unidad temporal. Curiosamente, encontramos que los informáticos utilizan exclusivamente la PA, mientras que los criminólogos prefieren aplicar la PAI. Además, mientras que el PA y el F1-Score se han probado preferentemente para predicciones a corto plazo (es decir, menos o igual a 3 meses), el PAI se ha utilizado para predicciones tanto a corto como a largo plazo. No se detectó ningún otro patrón obvio entre los demás elementos de información

sobre el uso y preferencia de estas métricas de evaluación.

Métricas por tarea de prevención

La tarea de pronóstico más común es la clasificación binaria ($n=21$) para los puntos críticos de delincuencia ($n=20$) y la categoría de delincuencia ($n=1$). Si bien la tarea de clasificación es frecuentemente discutida al inicio de los experimentos, algunos artículos consideran en la evaluación del desempeño un ítem diferente al de la salida del algoritmo, transformando así los productos de regresión en valores binarios. Los ejemplos más destacados incluyen los modelos RTM (Drawve et al. 2016; Dugato et al. 2018; Gimenez-Santana et al. 2018), donde el resultado del algoritmo es una puntuación de riesgo. Esta puntuación se reclasifica posteriormente en un resultado binario (una puntuación de riesgo positiva o negativa) a los efectos de la evaluación.

Además, Rummens et al. (2017) proponen un modelo de conjunto combinado que consta de LR y MLP que infiere valores de riesgo, similar a RTM, donde los autores consideran como punto crítico de delincuencia los valores con un riesgo superior al 20%.

La tarea de regresión ($n=11$) se utiliza en gran medida para predecir el número de delitos ($n=8$) y el rendimiento se mide mediante varias medidas de error, como MSE ($n=4$) o RMSE ($n=3$). Araujo et al. (2017) proponen dos nuevas métricas de evaluación, a saber, la puntuación de precisión de regresión (RAS), que representa el porcentaje de éxito en la predicción de una muestra, y la puntuación de precisión de regresión (RPS), que define la precisión de la RAS. El RPS mide el MSE de muestras de éxito normalizadas por la varianza de la muestra de entrenamiento (Araujo et al. (2017). Rodríguez et al. (2017) presentan la Eficiencia de Nash-Sutcliffe (NSE), que derivan de la predicción de modelos hidrológicos, como una estadística normalizada que determina la magnitud relativa de la varianza residual en comparación con la varianza de los datos medidos.

Sin embargo, el número de delitos no es la única inferencia considerada en los modelos de regresión. Por ejemplo, Ivaha et al. (2007) predicen el porcentaje de criminalidad en conglomerados, utilizando elipses espaciales como unidades espaciales, Rodríguez et al. (2017) investigan las propiedades de los conglomerados, mientras que Shoesmith (2013) infiere las tasas de criminalidad a partir de datos históricos sobre la criminalidad.

Además de las métricas de evaluación mencionadas anteriormente, tres artículos analizan los gráficos de vigilancia para la evaluación de predicciones. Mohler (2014) utiliza una métrica gráfica de vigilancia que muestra la fracción de delitos preventivos durante un período de tiempo frente al número de celdas de la cuadrícula con delitos reales para cada día (Fig. 4a). El mismo autor menciona que esta métrica es similar a la curva característica operativa del receptor, o curva ROC, aplicada por Gorr (2009), pero difiere porque no utiliza una tasa de falsos positivos asociada en el eje x. De manera similar, Hu et al. (2018) aplican la curva PAI, también un gráfico de vigilancia que muestra el porcentaje de área en el eje x y el PAI o el valor de la tasa de aciertos en el eje y (Fig. 4b, c). De manera similar, Rosser et al. (2017)

utilizan gráficos de vigilancia de la tasa de aciertos, que representan la tasa de aciertos media frente a la cobertura para la red y los enfoques de predicción basados en cuadrículas (Fig. 4c). Estos gráficos son muy útiles para visualizar los valores de las métricas en el territorio encuestado.

Algoritmos y estrategias de validación

Como se menciona en la sección "Métodos de prevención de delitos espaciales" , en muchos de los artículos, el método de prevención propuesto no incluye un algoritmo novedoso, sino que aplica principalmente nuevas variables que no se han utilizado antes. Al recordar los cuatro tipos de algoritmos, a saber (i) basados en kernel, (ii) procesos puntuales, (iii) aprendizaje automático tradicional y (iv) aprendizaje profundo, notamos una diversidad entre el pronóstico propuesto y el métodos de referencia. Además, las estrategias de validación también son diversas. La mitad de los estudios (n=16) consideran dividir los datos en subconjuntos de entrenamiento y prueba.

La mayoría de estos estudios incluyen un 70% de entrenamiento (actual) y prueba aplicada a la tarea de clasificación, y un 30% de series de prueba (futuras). Johansson et al. (2015) utilizan un enfoque combinado, que incluye un horizonte móvil, que produce diez veces el tamaño de una muestra para el algoritmo KDE, que contiene el 70% del conjunto de datos sobre delitos original (manteniendo la proporción 70/30). El resultado final se calcula como la media de las diez mediciones. La Figura 5 ofrece una buena descripción general de todos los algoritmos y sus estrategias de validación.

Esta visualización del árbol de decisiones incluye cinco elementos de datos centrales: tarea de predicción, método de pronóstico de entrada propuesto, tipo de algoritmo de pronóstico propuesto, estrategia de validación y métricas de evaluación. Clasificación m se refiere a aquellas métricas de evaluación que se utilizan particularmente para tareas de clasificación (por ejemplo, PA, puntuación F1). La regresión m es una composición de métricas de error para el análisis de regresión (por ejemplo, MSE, RMSE, MAE) mientras la criminología m incluye métricas de análisis de delitos (p. ej., PAI, RRI).

Los algoritmos basados en kernel se utilizan preferentemente para predecir los puntos críticos (n=5) y el número de delitos (n=1).

Curiosamente, Malik et al. (2014) ponen en discusión el hecho de que regiones con demografía similar tienden a ilustrar tendencias similares para ciertos tipos de delitos. Esta observación se incluye en su modelo de predicción "Método de estimación de densidad de núcleo de covarianza dinámica (DSKDE)" y se compara con el modelo de referencia "Descomposición de tendencias estacionales basada en Loess (STL)". Hart y Zandbergen (2014) y Johansson et al. (2015) utilizan un enfoque de KDE basado en el kernel sin compararlo con un método de referencia, y ambos consideran el PAI como una de las métricas de evaluación. Solo dos de los estudios basados en kernel consideran datos auxiliares (Gorr et al. 2003; Rosser et al. 2017),

sin embargo, ambos utilizan diferentes estrategias de validación (horizonte móvil y división de prueba de tren, respectivamente) y métricas de evaluación (MAE), MSE, MAPE en la primera publicación y Hit Rate en la segunda publicación). Por lo tanto, vale la pena señalar que, si bien se utiliza el mismo algoritmo base, como KDE, otros componentes del proceso de predicción pueden ser diferentes.

Dos de cada tres algoritmos de proceso de puntos no explican la estrategia de validación seguida en los estudios (Liesenfeld et al. 2017; Mohler 2014). Möhler (2014) muestra un interesante enfoque de proceso puntual que utiliza únicamente datos históricos sobre delitos y captura patrones de riesgo de delincuencia tanto a corto como a largo plazo. Este artículo incluye la evaluación de la trama de vigilancia (ver la subsección “Métricas por tarea de pronóstico”), comparando los componentes crónicos y dinámicos de los puntos críticos para homicidios y todos los tipos de delitos.

La tercera categoría de algoritmos de pronóstico, el ML tradicional, se divide casi por igual entre tareas de clasificación y regresión. Solo tres artículos que analizan algoritmos de ML tradicionales no mencionan información sobre la comparación inicial (Araújo et al. 2018; Rodríguez et al. 2017; Rummens et al. 2017). La mayoría de los algoritmos de ML ($n = 11$) utilizan la estrategia de validación dividida entre entrenamiento y prueba aplicada a la tarea de clasificación.

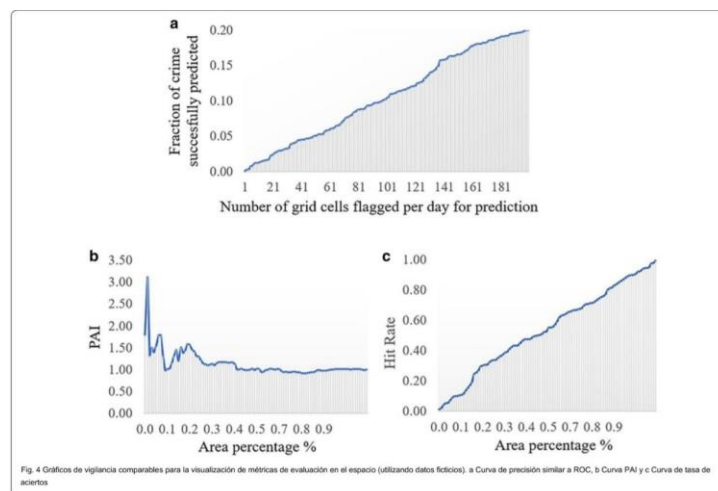
Curiosamente, uno de los artículos (Yu et al. 2011) analiza un enfoque de validación diferente, el “Leave-One-Month-Out” (LOMO), donde en lugar de ejecutar la clasificación solo una vez en los conjuntos de datos de entrenamiento y prueba, se ejecuta en $S-1$ conjuntos (S =número de conjuntos/meses).

Un conjunto cada vez mayor de técnicas de pronóstico se basan en DL; sin embargo, para esta revisión, incluimos solo tres artículos, todos ellos de predicción a corto plazo y provenientes de la disciplina de las ciencias de la computación (Huang et al. 2018; Lin, Yen, y Yu 2018; Zhuang et al. 2017).

Dos de los tres artículos consideran variables geográficas auxiliares y aplican la estrategia de validación del horizonte móvil, mientras que el tercer artículo aborda únicamente los rezagos de la delincuencia siguiendo un enfoque de validación cruzada de 10 factores. Los tres artículos consideran una clasificación binaria evaluada mediante métricas como la puntuación PA y F1. Zhuang et al. (2017) proponen una red neuronal espacio-temporal (STNN) para pronosticar la ocurrencia de delitos, al tiempo que incorpora información espacial. Luego comparan el STNN con tres métodos de última generación, incluida la red neuronal recurrente (RNN), la memoria a corto plazo (LSTM) y la unidad recurrente cerrada (GRU). Dado que el modelo está diseñado para todos los tipos de datos sobre delitos, cada tipo de delito puede conducir a diferentes desempeños del STNN debido a su variabilidad en el tiempo y el espacio. Presumiblemente, aparecerán desafíos para los tipos de delitos con bajos volúmenes de datos, porque las redes neuronales requieren una cantidad suficiente de datos para su entrenamiento.

Dependencias y limitaciones

Aunque la mayoría de los artículos utilizan métricas de evaluación estándar, como PA para un resultado binario, generalmente no incluyen métricas complementarias para garantizar que se cubran todos los aspectos del desempeño de la predicción. A menudo, el PA se utiliza por sí solo para medir el rendimiento del modelo (Araújo et al. 2018; Malik et al. 2014; Mu et al. 2011). Se necesitan métricas complementarias, porque si bien un método puede tener una puntuación de evaluación más alta que otros, pueden proporcionar información adicional.



Por ejemplo, si bien muestra un PAI alto, el valor del Índice de eficiencia de predicción (PEI) (Hunt 2016) puede reducirse. PEI es otra métrica de evaluación que se calcula mediante la relación entre el PAI y el PAI máximo posible que un modelo puede alcanzar. La diferencia entre el PAI y el PEI se puede explicar porque ambas métricas tienen diferentes dependencias con el tamaño de la celda. Las métricas complementarias también superan las limitaciones de algunas medidas de evaluación. Por ejemplo, el PA es la suma de verdaderos positivos y verdaderos negativos dividida por el número total de instancias, lo que representa el porcentaje que se clasifica correctamente. Sin embargo, esta información puede no ser suficiente para juzgar el desempeño de un modelo, porque omite información sobre la precisión.

El Hit rate y el PAI se obtienen mediante una división, así, cuando el denominador es pequeño, ambas métricas son altas. En consecuencia, cuando la incidencia de delitos es baja, los resultados se ven muy afectados. Además, las métricas tradicionales son de naturaleza global, pero en la predicción o pronóstico espacial, también nos interesa la distribución espacial de la predicción. Puede haber áreas locales de buen y mal desempeño de predicción, lo que resulta en un valor global promedio. Una métrica complementaria para un resultado de regresión podría ser calcular el I de Moran del error de predicción y explorar la variación del desempeño predictivo en toda el área de estudio. Idealmente, el error de predicción debería seguir una distribución espacial aleatoria. En general, encontramos un interés bajo o nulo en desarrollar métricas de evaluación (locales) espaciales, temporales o espaciotemporales.

La relevancia de las métricas de evaluación puede estar sesgada por varias razones. Un ejemplo puede ser el desequilibrio de clases. Un modelo puede tener una gran precisión y al mismo tiempo predecir muy bien las ubicaciones sin delitos. En cambio, las localizaciones con delitos no están bien pronosticadas. Algunos autores intentan mejorar la relación negativa-positiva entre células con delito y sin delito, ajustando la ponderación de los puntos críticos y los puntos fríos (Yu et al. 2011), o cambiando el conjunto de entrenamiento, mientras que el conjunto de prueba mantiene sus datos reales originales (Rumi et al.2018).

Otra dependencia son los diferentes tipos de agregación que tienen lugar durante el modelado por atributos de tiempo, espacio o tipos de delito. Por ejemplo, si bien la mayoría de los artículos informan que trabajan con tipos de delitos desagregados, algunos de ellos consideran agregar tipos de delitos a, por ejemplo, “delitos violentos”, sin especificar qué tipos se incluyen. Además, los efectos que tienen las agregaciones espaciotemporales.

El desempeño de los pronósticos generalmente no se analiza, pero podría realizarse fácilmente con un análisis de sensibilidad.

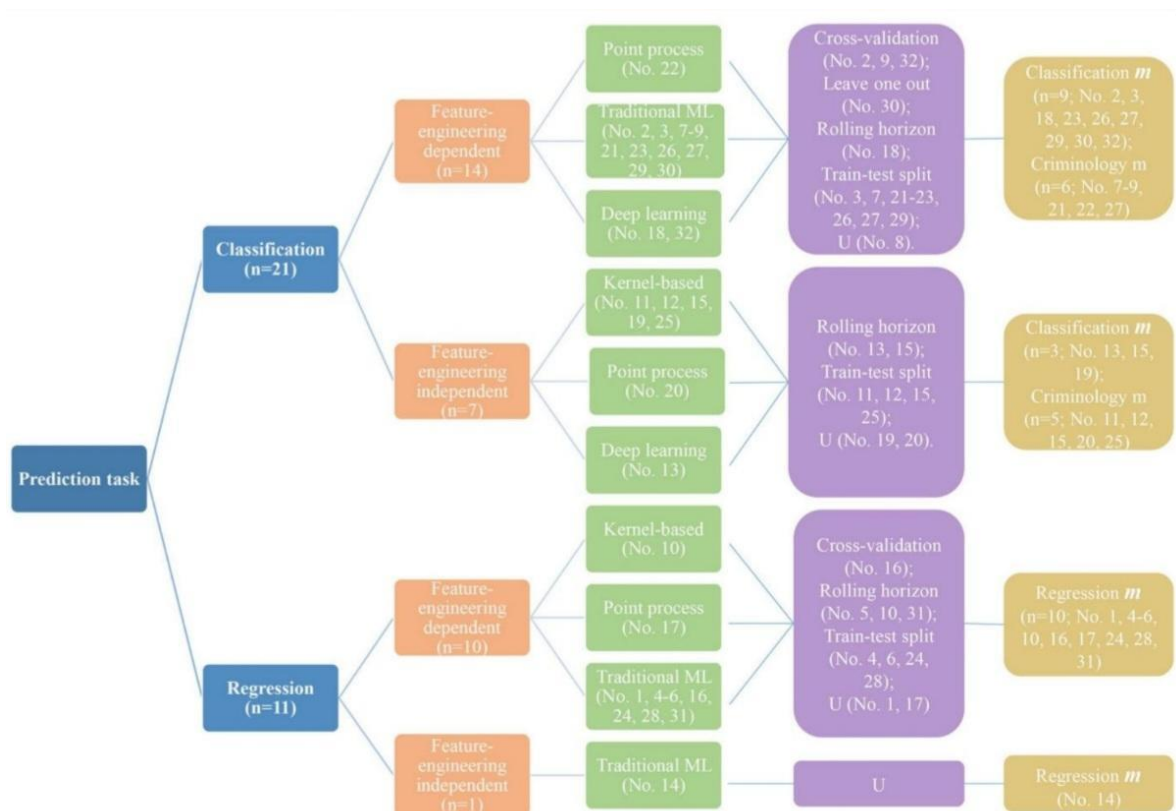


Fig. 5 Descripción general de los métodos de pronóstico (consulte la sección "Métodos de pronóstico de delitos espaciales") y su evaluación de desempeño (consulte la sección "Consideraciones al analizar el desempeño de los pronósticos") vinculados a los 32 artículos seleccionados. Las referencias de los artículos vinculadas a su número se muestran en la Tabla 3. La letra m denota una métrica de evaluación. La letra "U" denota un elemento indefinido.

DISCUSIÓN

En esta sección realizamos un análisis FODA de los hallazgos más significativos.

FORTALEZAS

Uno de los elementos más fuertes de los esfuerzos de investigación actuales es la incorporación de información espacial o espaciotemporal en los algoritmos de predicción tradicionales. Los ejemplos de este enfoque son STAR y STKDE (Shoemith 2013; Rosser et al. 2017).

Además, KDE, un método tradicional en el campo, se ha adaptado para considerar problemas de muestreo, como datos dispersos (DCKDE) y formato de cuadrícula (NTKDE) (Malik et al. 2014; Rosser et al. 2017). Además, el interés de la comunidad científica por la incorporación y efecto del big data en la predicción queda patente en el apartado de trabajos relacionados. Este interés también se ve respaldado por la tendencia a introducir variables dinámicas en el proceso de modelado, como el cálculo de la entropía de los visitantes de Foursquare o la población ambiental de las redes sociales y el transporte. En cuanto a la evaluación del desempeño, los gráficos de vigilancia (Fig. 4) brindan una imagen más detallada de la precisión de la información pronosticada. Dado que incluyen la cobertura del área en el eje x, la policía puede utilizarlos como herramienta de decisión para identificar el umbral que equilibra la precisión de la predicción con el tamaño de las áreas de patrullaje.

DEBILIDADES

En general, no siempre se informan detalles significativos de los experimentos del estudio y los elementos comúnmente no definidos son la unidad espacial de análisis y el tamaño de la muestra. De manera similar, para los métodos que dependen de la ingeniería de características, los procedimientos de elaboración no se describen suficientemente. Los elementos anteriores hacen que un estudio sea difícil de reproducir o comparar sus resultados con un posible estudio futuro. Además, no encontramos ninguna herramienta de código abierto que implemente pronósticos espaciales de delitos utilizando los métodos mejor propuestos reportados. Una herramienta de este tipo podría mejorar la posibilidad de reproducir los resultados de un estudio de prevención existente.

Sugerimos que siempre se informen todos los elementos de datos analizados en la sección “Resumen de publicaciones seleccionadas sobre pronóstico espacial de delitos” (para obtener una descripción general, consulte la Tabla 3). Sin embargo, se podría desarrollar un “protocolo de pronóstico espacial” detallado de manera similar a los protocolos para otros enfoques de modelado, como el protocolo ODD (Grimm et al. 2010). Además, la unidad espacial más común es la celda de la cuadrícula, que puede no necesariamente alinearse con los lugares donde normalmente se despliegan los recursos policiales. Hasta el momento, no hemos encontrado ningún estudio que aborde suficientemente este tema. En cuanto a la evaluación del desempeño, la mayoría de los autores utilizan métricas estándar.

Una métrica estándar “global”, como MAE, no puede describir la distribución del error de predicción en el espacio, que puede variar mucho. Por lo tanto, proponemos desarrollar nuevas métricas de evaluación espacial o espaciotemporal locales. Finalmente, otras cuestiones de modelización apenas se discuten, si es que se discuten, como el exceso de información, la multicolinealidad, el sesgo de muestreo y la escasez de datos.

OPORTUNIDADES

Hay un tremendo aumento en los estudios de prevención espacial de la delincuencia. Del conjunto de 32 artículos seleccionados, se publicaron 7 y 11 artículos en 2017 y 2018, respectivamente, en comparación con aproximadamente un artículo por año entre 2000 y 2016 (Fig. 2). Esto muestra el creciente interés de académicos de diversas disciplinas (compárese con la Tabla 2) en este tipo de investigación. El tipo de delito que más se ha estudiado es el robo con allanamiento de morada. No está claro por qué es más probable que se estudien este tipo de delito en particular y los delitos contra la propiedad en general. Una oportunidad futura podría ser probar sistemáticamente si existe un patrón de delitos contra la propiedad que superan consistentemente a otros tipos de delitos y por qué.

Además, a excepción de RTM y KDE, los artículos seleccionados no utilizaron otros métodos espaciales mencionados en la sección de trabajo relacionado (sección “Trabajo relacionado”). La razón puede ser que los autores tienen diferentes antecedentes, como ciencias de la computación y criminología, y es posible que no estén familiarizados con dichos métodos. Esto abre una oportunidad de investigación para explorar y comparar métodos espaciales menos utilizados con enfoques tradicionales, como RTM o KDE. Otra oportunidad sería comparar el rendimiento de pronóstico de los métodos entre sí. En esta revisión presentamos tendencias metodológicas, pero no fue posible realizar una comparación justa entre los métodos espaciales. En primer lugar, algunos métodos no se compararon con un método de referencia. Otros autores compararon el mismo método con un conjunto diferente de características.

Incluso si hubiera artículos con un conjunto similar de características, una comparación entre ellos estaría sesgada debido a variaciones en los datos de muestra, áreas de estudio, períodos de muestreo, etc. Los futuros estudios empíricos deberían centrarse en la comparación de algoritmos, cuyo número aumenta constantemente. Fusionamos los artículos seleccionados en cuatro categorías de algoritmos de pronóstico, incluidos los procesos puntuales basados en kernel, el aprendizaje automático tradicional y el aprendizaje profundo. Los algoritmos tradicionales de aprendizaje automático estaban presentes en la mayoría de los métodos propuestos, siendo MLP y RF los más comunes, mientras que los modelos AR fueron los métodos de referencia más utilizados. Una sugerencia es comparar algoritmos nuevos o desarrollados recientemente con los propuestos con más frecuencia, en lugar de continuar realizando comparaciones adicionales con métodos tradicionales o más simples que repetidamente han demostrado tener un rendimiento inferior.

AMENAZAS

Destacamos que los estudios espaciales de prevención del delito carecen de una terminología coherente, especialmente para términos como “predicción”, “pronóstico” y “puntos críticos”. La tarea predictiva predominante es la clasificación binaria (n=21) y la inferencia de pronóstico predominante son los puntos críticos (n=20).

Es importante entender la razón detrás de esta tendencia. ¿Es el análisis de regresión menos útil o más difícil de predecir? Aunque notamos un aumento constante en el desarrollo de algoritmos de clasificación o características para incorporar en la tarea de clasificación, reconocemos la importancia de ambas tareas de predicción. Además, para visualizar el panorama delictivo de un área, es importante examinar tanto los puntos críticos como los fríos o una clasificación multiclase hacia los puntos más críticos de criminalidad. Sin embargo, ninguno de estos fue el foco de los artículos examinados. Reconocemos que pronosticar los puntos críticos es importante para que la policía asigne recursos. Sin embargo, ¿qué pasa con la información que puede derivarse de otros tipos de agrupaciones espaciales, como puntos fríos, valores atípicos de puntos fríos o valores atípicos de puntos calientes, comúnmente denominados LL, LH, HL (bajo-bajo, bajo-alto, alto-bajo, respectivamente) y calculados por la estadística local de Moran (Anselin 2005)? La ciencia necesita progresar en el conocimiento, lo que requiere comprender y examinar todos los aspectos de un fenómeno. Finalmente, sólo un tercio de todos los artículos realizaron predicciones a largo plazo. Si bien esta tendencia es positiva porque las fuerzas del orden tienen interés en la predicción casi en tiempo real, no se debe pasar por alto que la predicción a largo plazo desempeña un papel importante en la comprensión del riesgo de delincuencia y proporciona un panorama amplio para la planificación estratégica.

CONCLUSIÓN

En este artículo, nos centramos en la “Pronóstico espacial de la delincuencia”, que es un enfoque de inferencia sobre la delincuencia tanto en el tiempo como en el espacio. Realizamos una revisión sistemática de la literatura que sigue la guía de informes "PRISMA" (Liberati et al. 2009) para comprender y evaluar el estado del arte relacionado con conceptos y métodos en estudios empíricos sobre delincuencia con muchas aplicaciones y especial atención a la delincuencia. Abordamos varias preguntas de investigación que tratan del papel del espacio en el procedimiento de pronóstico, los métodos utilizados, el rendimiento predictivo y, finalmente, las estrategias de validación del modelo.

Identificamos cinco tipos de inferencia, a saber (1) puntos críticos (la mayoría de los artículos), (2) número de delitos, (3) tasa de delitos, (4) categoría de delitos, (5) porcentaje de delitos en grupos. y (6), propiedades de los conglomerados. Con respecto a los métodos de pronóstico,

los autores propusieron principalmente métodos tradicionales de aprendizaje automático, pero también enfoques basados en la estimación de la densidad del núcleo y, con menos frecuencia, enfoques de proceso puntual y aprendizaje profundo. Cuando se trata de medir el rendimiento, se utilizó una gran cantidad de métricas, siendo las tres principales la precisión de la predicción, seguida del índice de precisión de la predicción y la puntuación F1. Finalmente, el enfoque de validación más común fue la división tren-prueba, mientras que otros enfoques incluyen la validación cruzada, dejar uno fuera y el horizonte móvil.

Este estudio fue impulsado por la creciente publicación de estudios espaciales de prevención del delito y (análisis predictivos del delito en general). Más de la mitad de los artículos seleccionados (n=32) fueron publicados en los últimos 2 años. En concreto, se publicó alrededor de un artículo por año entre 2000 y 2016, mientras que en 2017 y 2018 se publicaron 7 y 11 artículos, respectivamente. Al mismo tiempo, hay un crecimiento global de la producción de publicaciones científicas. Born-mann y Mutz (2015) adaptaron un modelo exponencial a este crecimiento y calcularon una tasa creciente de producción de alrededor del 3% anual, mientras que se estima que el volumen se duplicará en aproximadamente 24 años. Sin embargo, los patrones anuales de los artículos seleccionados muestran un aumento mucho mayor que indica la importancia y el potencial futuro de los estudios relacionados con la prevención espacial del delito.

Además, nos gustaría esbozar las principales limitaciones que pueden impedir la reproducibilidad y, por tanto, el avance de este tema a largo plazo. En primer lugar, la terminología utilizada no es coherente, posiblemente debido al hecho de que los científicos que trabajan en este tema tienen diversos antecedentes (por ejemplo, criminología, informática, geociencias, políticas públicas, etc.). En segundo lugar, los detalles significativos de los experimentos del estudio se informan vagamente o no se informan en absoluto.

Con respecto al último punto, sugerimos reportar los siguientes ítems de datos: área de estudio, escala, período de muestreo, meses, tipo, muestra, inferencia, tarea, unidad espacial y unidad temporal (en total 10 ítems). Los elementos adicionales que se deben informar son el método propuesto, el mejor método propuesto, el método de referencia, las métricas de evaluación y la estrategia de validación (en total, 5 elementos).