Journal of Scientometric Research, 2026; 15(3):1032-1048.
https://www.jscires.org

Research Article

AI in Crime Prediction Research: Scientometric Evidence of Conceptual Change, Method Evolution, and Evaluation Reporting Gaps

Cem Eroglu[1,*], Hüseyin Cakir[2]

[1]Department of Digital Forensics, Graduate School of Informatics, Gazi University, Ankara, TÜRKİYE.
[2]Department of Computer Education and Instructional Technology, Faculty of Education, Gazi University, Ankara, TÜRKİYE.

ABSTRACT

This study examines the role of Artificial Intelligence (AI) in crime prediction research using scientometric and bibliometric methods, with a targeted analysis of evaluation reporting practices in influential AI-based crime prediction studies. The study introduces the Conceptual Shift Index (CSI), a novel scientometric indicator for quantifying conceptual change, proposes the Method Evolution Network (MEN) to examine the structural dynamics of methodological change, and develops the Evaluation Reporting Framework (ERF) to assess the completeness of evaluation reporting. A total of 2,821 crime prediction publications from the Web of Science and Scopus databases were analysed using Biblioshiny. In addition, six of the most highly cited recent AI-based crime prediction studies (2022-2026) were selected to provide a targeted assessment of contemporary evaluation reporting practices. The findings show that the share of AI-based studies increased from 8.8% in 2016 to 62.2% by 2026, indicating that machine learning and deep learning approaches have become increasingly prominent within crime prediction research since 2017. This conceptual change is quantitatively reflected by a CSI value of 0.737. Furthermore, MEN analysis reveals increased methodological interaction, with network density rising from 0.196 to 0.592. Application of the ERF produced an average Coverage Score of 0.78 and showed that class imbalance and temporal resolution are not systematically reported. The findings point to structural deficiencies in evaluation processes in crime prediction research. In response, this study proposes an evaluation framework to improve the interpretability and comparability of performance metrics across studies.

Keywords: Conceptual change, Crime prediction, Evaluation framework, Model interpretability, Performance metrics, Scientometrics.

Correspondence:
Cem Eroglu
Department of Digital Forensics, Graduate School of Informatics, Gazi University, Ankara, TÜRKİYE.
Email: cem.eroglu1@gazi.edu.tr
ORCID: 0000-0001-8491-6398

Received: 15-05-2026;
Revised: 09-07-2026;
Accepted: 21-09-2026.

INTRODUCTION

Crime prediction research has evolved rapidly in the past two decades, especially with predictive policing, spatial analysis (Perry *et al*., 2013), and Artificial Intelligence (AI)-based modeling (Eroglu & Cakir, 2024; Wheeler & Steenbeek, 2021). Earlier research was based on statistical, hotspot, and environmental risk models, while more recent research tends to use AI-based approaches to analyse complex spatial, temporal, and contextual crime patterns (Kaur & Saini, 2024). With the rapidly increasing volume and methodological diversity, there is an increasing need to systematically map the conceptual evolution, methodological development, and emerging trends in crime prediction research.

DOI: 10.5530/jscires.20260212

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

Recent advances in AI have accelerated the development of crime prediction models and extended the scope of prediction tasks such as crime occurrence, crime counts, hotspot prediction, and risk estimation (Eroglu & Cakir, 2026; Kaur & Saini, 2024; Liang *et al*., 2022). While many studies evaluate the performance of individual prediction models, little attention has been paid to the evolution of the research field, to changes in methodological relationships over time, and to whether evaluation practices have evolved in parallel with methodological advances.

In addition, the increasing relevance of AI-based crime prediction suggests an additional conceptual change for computational social science beyond methodological pluralism, which is reflected in the growing use of probabilistic modelling, algorithmic optimization, and validation approaches informed by performance metrics that are typical for Machine Learning (ML) and Deep Learning (DL) (Eroglu & Cakir, 2024). In this paper, the conceptual change is understood as measurable changes in the relative prominence of concepts and methodological terminology over time and not a broader epistemological change. These methods have

improved the analysis of complex data and have shown promising predictive performance across a range of crime prediction tasks (Travaini *et al*., 2022). However, few evidence-synthesis and research-mapping studies have discussed the role of AI in crime prediction research in a broader sense.

In this context, this study aims to systematically chart the major changes and identify recent trends in AI-based crime prediction research using scientometric and bibliometric methods. Historical trends from 1970 to 2026 were reviewed, current research patterns were described, and possible future directions have been suggested.

To quantify conceptual change in the literature, the current study proposes the Conceptual Shift Index (CSI), a scientometric indicator based on normalized keyword distributions. Mathematically, CSI corresponds to the total variation (L1, also known as Manhattan) distance between normalized keyword distributions. However, unlike general-purpose distributional distance measures, CSI is introduced here as a domain-specific scientometric indicator to quantify conceptual change in crime prediction research on an interpretable 0-1 scale. Moreover, the Method Evolution Network (MEN) is introduced to describe the structural dynamics of methodological changes. CSI measures the degree of conceptual change, while MEN analyses the interactions among methods and their restructuring over time at a structural level.

Apart from documenting methodological developments, this paper also examines the implications of such changes for model evaluation. The primary limitation in crime prediction studies is not algorithm performance but the evaluation and interpretation of model output. Model performance metrics are often reported without sufficient contextual information (e.g., target definition, spatial and temporal resolutions, data imbalance) and thus constrain cross-study comparability. Hence, an Evaluation Reporting Framework (ERF) is developed to evaluate model performance.

METHODOLOGY

This study adopts a scientometric and bibliometric approach (Table 1) to analyse the development of crime prediction research. Previous studies (Alomran *et al*., 2025; Mohamad, 2025; Duplock *et al*., 2026) have widely used scientometric and bibliometric mapping approaches to examine thematic evolution and research trends across various disciplines. Bibliometrics and scientometrics are closely related analytical approaches that enable the quantitative mapping of scientific production, thematic evolution, and collaboration structures within a research field (Dutta *et al*., 2026). In addition, to examine evaluation practices, a further analysis was conducted using a targeted subsample of high-impact studies published in recent years.

Under this framework, the analysis first focused on the overall body of crime prediction publications without explicitly considering the involvement of AI. AI-based terms were then introduced to construct a separate dataset for the subsequent analysis phase (Appendix 1).

The publications selected for the study were journal articles and conference proceedings. Journal articles were prioritized given the peer-review process and stronger citation impact. Conference papers were included because they often contain new AI methods, especially in computer science and ML/DL. For consistency and reliability, other types of documents such as editorials, books, notes, errata, and retracted publications were excluded (Appendix 1). Furthermore, no language restriction was explicitly applied during the database retrieval. However, since the search strategy was based on English-language search terms in the Title, Abstract, and Author Keywords fields, the retrieved dataset primarily comprised English-language publications.

The datasets of the Web of Science (WoS) and Scopus were combined using the Merge Collections module in Biblioshiny (version 5.0). During the merging process, the metadata fields were automatically aligned across databases, and duplicate records were removed using the integrated multistage matching process based primarily on DOI matching and, in the absence of DOI information, normalized title and publication year matching. In addition, Biblioshiny standardized the author names and harmonized the metadata fields, thus producing a single bibliographic dataset for further analysis and ensuring consistency and comparability. No additional manual data cleaning or record exclusion was performed after the automated harmonization and deduplication process, apart from the predefined document-type filters described above.

Following the merging process, metadata completeness and consistency were assessed within Biblioshiny before conducting the bibliometric and scientometric analyses. The metadata completeness reports for the overall crime prediction dataset and the AI-based subset are provided in Appendix 2A and Appendix 2B, respectively.

The temporal periods used in the scientometric analyses were defined empirically based on the publication trends observed in the bibliometric analysis. As illustrated in Figure 1, the share of AI-based publications showed the first clear structural increase in 2017, indicating the beginning of a distinct shift in the methodological orientation of the literature.

Accordingly, 2017 was selected as the boundary between the pre-AI (1980-2016) and post-AI (2017-2026) periods. Although the early years after 2017 may be regarded as reflecting the initial diffusion of AI-based approaches, all publications from 2017 onwards were analysed as a single post-AI period to provide a consistent basis for the comparative scientometric analyses. As different Biblioshiny modules rely on different metadata fields

(e.g., author keywords, country information, and collaboration metadata), the effective sample size varies across analyses. Therefore, the number of publications reported for individual analyses may differ from that of the overall merged dataset.

RESEARCH FINDINGS

This section presents the findings from the bibliometric and scientometric analyses, as well as a targeted analysis of evaluation practices.

Bibliometric analysis: mapping trends in crime prediction research

This section analyses publication and citation trends over time, focusing on the temporal evolution and academic impact of crime prediction research. Analysis of the dataset (1970-2026; WoS & Scopus; 2,821 publications) shows that the average document age is 7.26 years, with 18.8 citations per publication. In total, the studies are distributed across nearly 1,600 sources, indicating a broad and diverse publication structure. Figures 1 and 2 present these trends and illustrate the growing role of AI in crime prediction research. The figures begin in 2003 because AI-based crime prediction publications before that year were sporadic and insufficient to show a meaningful annual trend.

The results obtained from WoS and Scopus show that the first study on AI-based crime prediction was published in 1997, followed by a second study in 2003. However, before 2013, AI-based studies remained limited both in number and in their proportion within the overall crime prediction literature.

A clear shift emerges after 2013, with a marked increase in AI-based publications and a steady rise in their share within the overall literature. This growth accelerates particularly after 2017 and becomes more pronounced after 2020.

Between 2023 and 2025, the share of AI-based studies exceeded 55%, indicating a substantial increase in the prominence of AI-based approaches within crime prediction research. AI-based approaches are no longer marginal and have become a prominent component of the field. This trend reflects the increasing use of ML and DL methods in recent studies (Figure 1).

Citation patterns are broadly consistent with these publication trends. Until 2013, the unstable citation trend for AI-based studies was apparent due to the small number of papers published. However, since 2013, the relative citation impact of AI studies has been significant and above the average overall literature during 2013-2018 (Figure 2).

After 2019, this dominance lessened, and both groups converged at similar levels. Thus, AI initially showed high citation impact but has gradually become more widely integrated into the field. These findings are consistent with an increasing methodological emphasis on ML- and DL-based approaches in crime prediction research. However, the apparent downward trend in recent years should be viewed with caution, as it results mainly from citation lag.

Scientometric analysis: life cycle analysis of crime prediction research

This section describes the life cycle model of the temporal evolution of crime prediction research through the Life Cycle of Scientific Production module in Biblioshiny. The module fits a logistic growth model to annual publication data and estimates key parameters, including the theoretical saturation level, duration of growth, and peak publication year, to characterize the developmental stage of a research field (Aria & Cuccurullo, 2017; Aria *et al*., 2026). Accordingly, the resulting estimates should be interpreted as model-based projections of publication dynamics rather than precise forecasts.

The Biblioshiny life-cycle module reports point estimates for the model parameters but does not provide confidence intervals or formal measures of uncertainty. Thus, the predicted saturation levels and the milestone years should be interpreted as descriptive outputs of the fitted logistic model, and not as exact predictions of the future publication trends.

The analysis compares the growth dynamics of crime prediction research with AI-based crime prediction studies. In this way, the position of AI-based crime prediction studies within the entire evolution of the field and their relative impact can be assessed. Estimated growth times for the total crime prediction research and its AI-based subset are, respectively, about 14.7 years and 8.2 years. Thus, crime prediction research based on AI has developed and diffused in much faster time than the whole development process of the total literature. The peak production year is estimated to be 2023 for the total literature and 2024 for AI-based studies, indicating the growing relative importance of AI-based studies within the total crime prediction research over recent years.

In terms of the current level of progress, the total crime prediction research has already reached approximately 81.5% of its projected saturation level, and the AI-based literature has already reached approximately 84.8%. Both results show a maturity but, AI-based studies have achieved this stage in a shorter period of time. Moreover, the projection that AI-based literature will reach 90% saturation by the end of 2027 and total literature in 2030 indicates that AI-based publications account for an increasing share of recent scientific output within crime prediction research.

These saturation estimates and milestone years are obtained from the fitted logistic growth model and thus should be interpreted as model-based predictions rather than exact predictions, as future publication trajectories may be affected by scientific, technological, or societal developments. These results suggest that AI-based approaches have become a more and more important methodological part of crime prediction research. Overall, crime

Table 1: Summary of the scientometric and bibliometric analysis steps.

Method	Objective	Analysis Type	Data Source	Time Frame	Tool
Bibliometric analysis	Quantify performance; surface influential works/ sources; map thematic/content structure.	Journal/source productivity and impact.	Crime prediction studies (overall; 2,821 publications, WoS & Scopus); AI-based crime prediction studies (1,035 publications, WoS & Scopus).	Up to May 1, 2026	Biblioshiny (Version 5.0)
		Publication and citation counts			
		Most-cited papers.			
Scientometric analysis	Map social and intellectual structure; identify hubs/clusters; detect seminal years.	Thematic trends	Crime prediction studies (overall; 2,631 publications, WoS & Scopus).	From 2008 to May 1, 2026	Biblioshiny (Version 5.0)
			Crime prediction studies (overall; 2,821 publications, WoS & Scopus).	Up to May 1, 2026	
		Country collaboration network	Crime prediction studies (overall; 2,036 publications, Scopus); AI-based crime prediction studies (overall; 962 publications, Scopus).	From 2017 to May 1, 2026	
			Crime prediction studies (overall; 631 publications, Scopus).	Pre 2016	

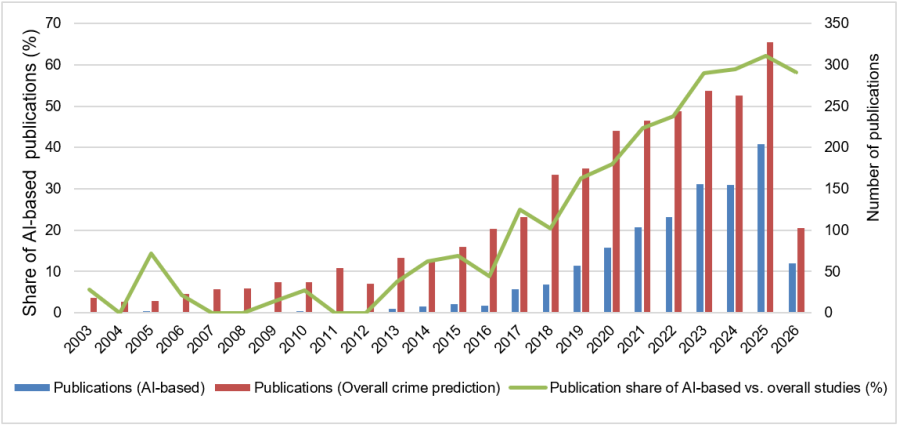


Figure 1: Publication trends in AI-based and overall crime prediction research (2003-2026; WoS & Scopus - 1,035 publications; Biblioshiny output).

prediction research is evolving in a broader, longer-term horizon, but AI-based studies have grown at a much faster pace in the literature, gained more visibility, and have become a key element in shaping the dynamics of recent developments in the field.

Bibliometric Analysis of Influential Studies and Targeted Analysis of Evaluation Practices

This section analyses the most highly cited studies in crime prediction at both local and global levels. The local analysis maps citation relationships within crime prediction research, while the global analysis examines how these studies are cited across different disciplines. As shown in Table 2, Chainey *et al.* (2008) received 158 Local Citations (LC) and 412 Global Citations (GC), followed by Mohler *et al.* (2011) with 131 and 558 respectively. These results suggest that they occupy a central role within the citation network of crime prediction research.

Table 2 shows that the early period of research in crime prediction was dominated by approaches based on hotspot analysis and spatial risk modelling. These studies provided the methodological basis of the field, but they were from a time when methods based on AI were still limited.

In contrast, the higher global citation rates of more recent studies suggest that the field mirrors the increasing relevance of data-driven and algorithmic models in the literature. However,

despite this change, there is still a large lack of standardization in the way assessments are conducted.

For the targeted evaluation analysis (Table 3), only publications from 2022 to 2026 were included from the WoS records obtained with the AI-based search strategy outlined in Appendix 1, to reflect current evaluation reporting practices, and the publications were ranked by citation count. The most cited records were then screened for eligibility and studies that performed AI-based crime prediction and reported model evaluation were included. This process resulted in a purposive sample of six studies selected to exemplify recent practice in evaluation reporting, rather than to constitute a comprehensive systematic review.

Similar limitations are found in recent AI-based crime prediction studies as shown in Table 3. Overall, target variables vary from binary occurrence to continuous counts and risk scores, but their definitions, class imbalance, interpretability, temporal and spatial resolution are often underreported.

The same limitations may also be seen in the earliest and most highly cited studies in the field. In the six key studies reviewed in Table 2 (Bowers *et al*., 2004; Caplan *et al*., 2011; Chainey *et al*., 2008; Gerber, 2014; Mohler *et al*., 2011; Mohler *et al*., 2015), crime prediction was performed, but the target variable was mostly undefined and non-standardized. Prediction performance in these studies was usually assessed by operational metrics like the hotspot capture rate, hit rate, or PAI, but important evaluation elements like class imbalance and distribution of the target variable were not systematically reported. This suggests that the shortcomings in evaluation noticed in recent AI-based studies are not novel but have been structural limitations within crime prediction research for some time.

Scientometric analysis: thematic trends in crime prediction research

In this section, topic trend analysis, CSI, and MEN were used to demonstrate the presence of AI in the field, and the changes in fundamental concepts over time were analysed.

Trend topics in overall crime prediction research

Trend topic analysis recognizes the most used terms in the literature and gives a visual summary of how they evolve over time. Therefore, instead of 1980-2026, the period 2008-2026 was selected to represent changes in key topics, as well as changes in the literature, and the increasing presence of AI in crime prediction research.

Trend topics were created using the Biblioshiny Trend Topics module based on the Author Keywords (DE). Only terms with a minimum frequency of seven occurrences were kept and a maximum of three terms per year was displayed. There was no additional manual preprocessing or synonym standardization before the trend topic analysis. The size of the bubbles indicates term frequency and the horizontal line indicates the period when the term was present in the selected trend topics (Figure 3).

In the earlier years of 2008-2016, the literature was marked by a greater prominence of concepts related to traditional criminological and statistical approaches such as geographic profiling, routine activities, time series, multivariate and statistical analysis. However, after 2017, the literature showed increasing prominence of data-driven analytical themes, including data mining, patterns, early warning systems, prediction, and computer crime.

A significant methodological change occurred after 2020. Concepts based on AI and ML, such as prediction models, DL, and ML, showed a marked increase in frequency. Larger bubble sizes for the period 2020-2026 indicate that these methods have become increasingly prominent within the literature.

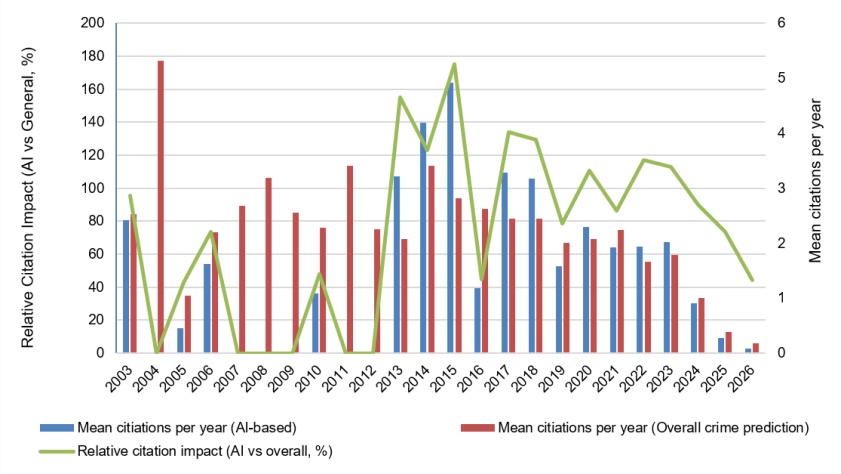


Figure 2: Citation trends in AI-based and overall crime prediction research (2003-2026; WoS & Scopus - 1,035 publications; Biblioshiny output).

Table 2: Top cited overall crime prediction research (1980-2026, Biblioshiny output).

Study	LC	GC	Citations per Year (LC/GC)
Chainey *et al.* (2008)	158	412	8.32 / 21.68
Mohler *et al.* (2011)	131	558	8.19 / 34.88
Gerber (2014)	102	387	7.85 / 29.77
Bowers *et al.* (2004)	81	201	3.52 / 8.74
Caplan *et al.* (2011)	73	197	4.56/12.31
Mohler *et al.* (2015)	73	229	6.08/19.08

These findings suggest that AI-related methods increasingly complement established criminological approaches while becoming more prominent within the literature. Accordingly, the thematic trends indicate an increasing prominence of AI-related concepts and methods within crime prediction research.

As shown in the bibliometric analysis section, the adoption of AI has increased since 2017. The trend topic analysis indicates conceptual changes in the terminology used within crime prediction research. To examine whether these conceptual developments are accompanied by changes in collaboration patterns, the following section analyses international co-authorship networks by comparing overall crime prediction studies with AI-based studies.

Scientometric analysis: conceptual shift index

A term-level change measure was used to more objectively assess the conceptual shift over time in crime prediction research. In this context, p-pre and p-post represent the normalized relative usage rates of the relevant keyword in the pre-AI (1980-2016) and post-AI (2017-2026) periods. To measure term-level change, Δp (temporal term change) was calculated for each keyword as follows:

$$\Delta p = p\text{-post} - p\text{-pre} \quad (1)$$

In addition, the CSI was used to measure the overall level of conceptual change in the literature. CSI is defined as half the sum of the absolute differences of the normalized keyword distributions between two periods:

$$CSI = \tfrac{1}{2} \sum |p_i(post) - p_i(pre)| \quad (2)$$

Where:

Δp: Temporal change in the normalized relative frequency of a keyword between the pre-AI and post-AI periods.

p-pre: Normalized relative frequency of a keyword in the pre-AI period (1980-2016).

p-post: Normalized relative frequency of a keyword in the post-AI period (2017-2026).

p_i(pre): Normalized relative frequency of keyword *i* in the pre-AI period.

p_i(post): Normalized relative frequency of keyword *i* in the post-AI period.

i: Individual keyword.

Σ: Summation over all keywords.

½: Normalization factor ensuring that CSI ranges from 0 to 1.

CSI ranges from 0 (identical keyword distributions) to 1 (completely different keyword distributions), thereby providing an intuitive measure of conceptual redistribution between two periods. The calculated Δp values show a significant increase in method-oriented concepts in the post-AI era. For example, crime prediction (Δp=+0.0278) and machine learning (Δp=+0.0269) are among the terms showing the highest increase, while concepts such as deep learning (Δp=+0.0119) and artificial intelligence (Δp=+0.0088) are also rapidly increasing in the literature. In contrast, a significant decrease is observed in the relative weight of classic criminological concepts such as victimization (Δp=−0.0223) and fear of crime (Δp=−0.0105).

These results quantitatively demonstrate that new methods have been introduced in crime prediction research, but the research focus has increasingly included data-driven and computational methods in parallel with established criminological approaches. The calculated CSI of 0.737 means that the normalized keyword distributions differ by 0.737 on a scale of 0-1, which suggests a substantial conceptual redistribution between the two periods. This value therefore implies a large reallocation of the conceptual lexicon used in the literature. Since CSI is introduced here as a descriptive scientometric indicator, the value should be interpreted in the context of this dataset and not as a universal benchmark.

A sensitivity analysis was conducted with three different minimum keyword-frequency thresholds (≥2, ≥3, and ≥5) to test the robustness of the proposed CSI. The CSI values were 0.663 to 0.668 in the case of excluding low-frequency keywords, while the value was 0.737 in the case of including all keywords (Table 4). The pruning of infrequent terms led to a slight decrease of the index, but the CSI values remained consistently high, implying that the detected conceptual shift is not merely an effect of rare keywords. Moreover, the CSI value did not change (0.737 before

and after harmonization), indicating that the results were robust to the term-standardization procedure.

Scientometric analysis: method evolution network

The present study developed a MEN to examine the methodological transition in crime prediction research at the conceptual and structural level. A method-oriented glossary was created (Appendix 1) using the title, abstract and author keyword fields. Each publication was screened for the existence of glossary terms and methods identified within the same publication were linked by co-occurrence links. That is, nodes are methodological terms and edges are their co-occurrence in single publications. An edge was formed when two methodological terms were found in the same publication (title, abstract or keywords). The weight of the edges corresponds to the number of co-occurrences in the publication set.

The textual data were converted to lowercase, punctuation and special characters were removed and whitespace was standardized before method detection. We aligned methodological variants and common abbreviations (e.g. ML/machine learning, ANN/artificial neural network, SVM/support vector machine) using the predefined method glossary, which is publicly available in the author's GitHub repository (see Code Availability).

The analysis showed that one or more method terms were identified in 2,238 (79.33%) publications in the literature. The temporal analysis of the structure of the network shows a great increase in the methodological diversity and interaction levels. The network structure had 28 nodes and 74 connections in the pre-AI period (1980-2016), which increased to 41 nodes and 187 connections in the transition period (2017-2019). The network structure after the AI period (2020-2026) reached 47 nodes and 640 links.

More importantly, the network density increased from 0.196 in the pre-AI period to 0.592 in the post-AI period. Network density is normalized by the maximum number of possible connections, making it less sensitive to differences in network size than raw node or edge counts. Higher density indicates that multiple analytical methods are more often combined in single studies, as network density is the ratio of actual connections to all possible method pairs.

These results suggest that the methodological network evolved from a relatively isolated methodological approach to a dense and more interconnected structure, mirroring the increase of methodological hybridization within the literature. In the pre-AI period, the main part of the network was occupied by classical spatial approaches, like hotspot mapping, while ML methods have dominated the transition and post-AI periods.

To further examine whether the increase in network density was mainly attributable to the larger post-AI corpus, an additional robustness analysis based on equal-size random subsampling was performed. 1,000 random subsamples were drawn from the post-AI corpus with the same number of publications containing detected methods as in the pre-AI corpus. Mean network density was 0.380 (SD=0.023; 95% interval 0.338-0.425), which was significantly greater than the original pre-AI density (0.196), but less than the full post-AI density (0.592). These results point to an increased methodological integration that cannot be explained solely by the greater number of publications in the post-AI era.

Table 3: Evaluation reporting practices in recent AI-based influential crime prediction studies in WoS (2022-2026; targeted analysis).

Study	Prediction Type	Target variable (Y_label)	Class imbalance (no crime proportion)	Temporal / Spatial resolution	Metrics used
Zhang *et al.* (2022)	Crime occurrence prediction	Binary (0/1)	93%	Biweekly / 150 m grid	Accuracy=0.89; ROC=0.586
Tang *et al.* (2023)	Regression	Not clearly defined	Not reported	Daily / 3 km × 3 km grid	MAE=0.9095; MAPE=0.5154 (NYC Crime)
Deng *et al.* (2023)	Regression	Continuous (crime risk intensity)	Not reported	Not reported / 200×200 ft grid	PAI, PEI, RRI, ROC
Wu *et al.* (2022)	Crime occurrence classification	Binary (0/1)	Not reported	Daily / Census tract	F1-score, relative improvement (2% to 7%)
Saraiva *et al.* (2022)	Mixed (Regression + Classification)	Classification / Binary (0/1) & crime count	Not reported	Not reported / 500 m grid	Accuracy=0.83; Recall=0.99; Precision=0.79; F1-score=0.89
Liang *et al.* (2022)	Regression	Crime count (continuous)	Not reported	Hour-level / region (grid-like)	MSE, Accuracy

Table 4: Sensitivity analysis of the CSI.

Analysis	CSI
All keywords	0.737
Minimum keyword frequency ≥2	0.666
Minimum keyword frequency ≥3	0.668
Minimum keyword frequency ≥5	0.663
Raw keywords (no harmonization)	0.737

However, it is important to note that the classical criminological approaches have not completely disappeared; on the contrary, they are still present within the network structure by integrating with new methods. The network shows the bridging role of boosting and ensemble-based methods. The results show that AI has brought new approaches to studying crime prediction and that AI-based approaches are associated with a more connected methodological structure in the literature. In this context, the proposed MEN approach allows the researchers to analyse the methodological transformation in the literature not only in terms of conceptual trends, but also in terms of structural relations between methods.

Scientometric analysis: mapping collaboration networks across countries

The country collaboration network highlighted key contributors, strong collaborative relationships, and regions that need greater scientific collaboration. The network map covered countries with at least one publication, regardless of their citation levels. The results suggest a clear pattern of global collaboration in crime prediction research.

Collaboration in overall crime prediction research from 1980 to 2016 was largely concentrated in North America and Western Europe, where the U.S., the U.K., China, and Canada were key contributors, while other European countries contributed relatively little. Although no separate figure is presented for 1980-2016, the conclusions are based on co-authorship and country-level collaboration analyses. The limited number of studies in the pre-AI period also suggests that the field was still in its early development phase.

From 2017 to 2026, AI-based publications point to a change in the structure of collaboration networks; however, the U.S. (71 co-authorships), China (70), India (37), and the U.K. (37) remained key hubs. Several emerging regions became more prominent, particularly Malaysia (33), Australia (26), Pakistan (22), Saudi Arabia (22), Italy (21), Hong Kong (20), and the United Arab Emirates (17), while the Netherlands, Germany, Sweden, and France became less prominent within the collaboration network (Figure 4). Country-level collaboration indicators for both the pre-AI and post-AI periods are summarized in Appendix 3. Overall, AI-based publications exhibit broader international collaboration patterns, with several additional regions becoming more visible during the study period.

For comparison, the non-AI crime prediction literature, consisting of crime prediction publications that were not classified as AI-based, showed a different collaboration pattern during the same period (2017-2026). The U.S. (115) and the U.K. (75) continued to lead, while China (53) emerged as an important contributor. Additionally, countries such as Australia (35), Canada (31), the Netherlands (20), Denmark (17), Germany (17), Spain (17), and Malaysia (16) also expanded their presence. However, certain nations, such as Uganda, Austria, and Mexico, reported just one collaboration, indicating uneven scientific integration.

Overall, these findings indicate that AI-based publications are associated with broader international collaboration patterns and the participation of additional regions in the research network. However, these descriptive findings do not by themselves demonstrate that AI caused the observed changes in international collaboration.

DISCUSSION

This section will examine changes in methods used for crime prediction and shifts in knowledge production.

Potential of AI algorithms in crime prediction

Although crime prediction research uses a variety of algorithms, there is no evidence that any one algorithm performs consistently across different contexts, regardless of crime type or setting (Travaini *et al*., 2022; White *et al*., 2025). Indeed, the results from the NIJ Recidivism Forecasting Challenge suggest that complex ML and DL models do not necessarily outperform simpler, more interpretable methods, such as logistic regression (White *et al*., 2025). Accordingly, there is a need to view algorithmic success beyond accuracy by considering interpretability and computational costs toward practical applicability.

Generally, hybrid approaches are employed in crime prediction. The pursuit of improved predictive performance through hybrid approaches may impose operational constraints due to processing costs and latency in real-time and at large scales. Thus, in crime prediction, algorithmic success should not be pursued solely through the lens of enhanced performance but should also be viewed as a cohesive component of system integrity.

On the other hand, these findings indicate that DL models primarily show considerable potential in dealing with high-dimensional data concerning complex interrelations (Eroglu & Cakir, 2024). Traditional ML methods have already demonstrated that high accuracy can be achieved using similar data sources (White *et al*., 2025; Travaini *et al*., 2022). Hence, the relationship between precision and complexity does not allow the

latter to indicate the superiority of a methodological approach (Eroglu & Cakir, 2026).

This assessment is also consistent with the scientometric findings of the current study. The CSI (0.737) and MEN results (network density increased from 0.196 to 0.592) indicate an increase in method diversity rather than the superiority of any particular algorithm, and that methods are increasingly being used together.

The findings above highlight that the leading issue for crime prediction does not lie so much in choosing algorithms as in how different models are compared. As algorithms increase the variety in metrics and evaluation methodologies, the comparability and interpretation of model performance become more difficult. What appears to be lacking in the literature on crime prediction is not complex models but the necessary consistency within evaluation methodologies.

Contributions to and gaps in the literature

As a result, introducing AI-based approaches into crime prediction research has diversified methods and research dynamics. The increase in concepts such as "machine learning" and "deep learning" attests to a shift toward a data-driven and algorithmic structure. Trend topic analyses indicate that concepts such as "machine learning" and "deep learning" have become increasingly prominent, reflecting a shift toward a more data-driven and algorithmic structure.

However, this also generates several methodological and structural issues. The use of different prediction tasks, such as classification, regression, and risk score prediction, does not allow a direct comparison of performance metrics. The same metric values can have vastly different implications under different problem definitions and data structures. This causes the interpretation of model performance to become problematic in a context-independent manner. The development of international collaboration further increases the diversity of data and model contexts and thereby increases the demand for a shared, interpretable language of evaluation.

The structure of the data sources also directly influences the evaluation. Model performance metrics are less reliable when crime data are imbalanced, incomplete, or intervention-based. Model outputs should be interpreted carefully when using historical data for prediction, as this may reinforce existing patterns of law enforcement.

Interpretability and ethical dimensions are also part of this process. Since many AI-based models are "black boxes," how their results are derived is difficult to understand; hence, using model outputs in decision-making processes is risky. Therefore, model performance should be evaluated not only by accuracy but also by transparency, fairness, and societal impacts.

In addition, this research suggests that CSI is a novel scientometric indicator that quantifies the conceptual shift in the literature. The high CSI value indicates a substantial shift in the conceptual vocabulary and methodological emphasis of the literature, reflecting the increasing prominence of data-driven and computational approaches alongside established criminological perspectives. In this respect, CSI extends traditional trend analyses by providing a quantitative indicator of conceptual change.

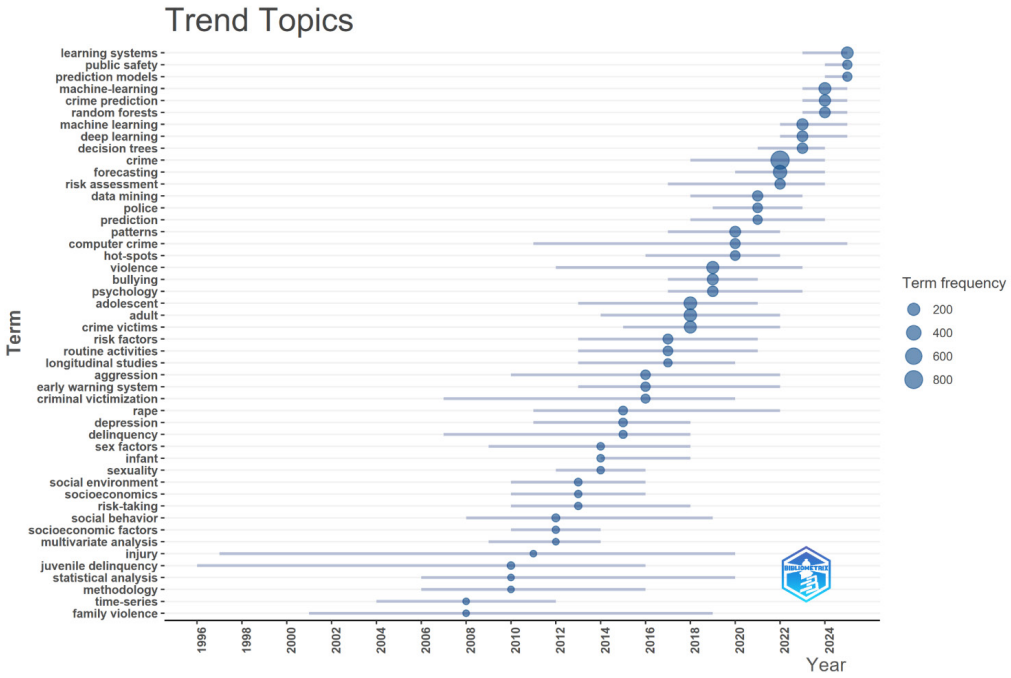


Figure 3: Trend topics in crime prediction research (2008-2026; min freq=7; 3 terms/year; Biblioshiny output).

Table 5: Equal-size subsampling robustness analysis of the MEN (1,000 random subsamples).

Network	Density
Pre-AI (1980-2016)	0.196
Post-AI (2020-2026)	0.592
Post-AI equal-size subsampling (mean, n=1,000)	0.380
Standard deviation	0.023
95% interval	0.338-0.425

This conceptual and methodological shift is also reflected in the structural relationships among methods. The MEN analysis proposed in this paper reveals that the methods applied within the literature have become denser and more integrated over time. The increased network density implies that crime prediction research shifted from a methodological approach characterized by isolation to a hybrid and interactive methodological structure. These findings suggest that AI-based methods have expanded the methodological structure of the literature.

Together, the core problem in crime prediction research lies not in the data or models but in how those models are evaluated and interpreted. In other words, the problem in the literature is not model performance per se but the contextual and systemic inadequacy of the evaluation processes. This limits the comparability and interpretability of crime prediction model results and underlines the demand for more systematic, transparent, and standardized approaches to evaluation.

Although the present study focuses on crime prediction research, the proposed CSI and MEN are not domain-specific approaches. Because both methods are based on changes in keyword distributions and methodological co-occurrence structures, they can be applied to other scientific fields undergoing conceptual or methodological transformation. Potential application areas include healthcare, education, environmental science, digital forensics, and computational social science, where data-driven methods and AI have substantially influenced research practices. Nevertheless, the interpretation of CSI and MEN should always consider the conceptual characteristics, publication patterns, and methodological context of the specific research field under investigation.

Evaluation Reporting Framework for Crime Prediction Models

The proliferation of AI-based methods within crime prediction research has made models increasingly diverse. However, this has complicated model performance evaluation. As shown in previous sections, performance metrics are commonly presented independently of contextual information across model types, data structures, and units of analysis. This represents a significant limitation for correctly interpreting model results and comparing studies.

However, the most fundamental issues in evaluation in crime prediction research concern the considerable heterogeneity in the target variable (Y_label), temporal and spatial resolution, and data imbalance across studies. Indeed, some studies only deal with the presence/absence of crime prediction (binary), while others output crime counts or risk scores. Likewise, predictions can be made on different time scales, from hours to days and weeks, and on various spatial scales, from points to grids and regions.

Under such heterogeneity, the direct comparability of metrics used in model performance evaluation, such as ROC-AUC, F1, and MAE, becomes impractical. Specifically, performance metrics reported without referencing the distribution of the target variable (Y_label) and the time-space unit under analysis limit the accurate interpretation of model performance.

Accordingly, this paper advances the ERF, insisting on systematic reporting of metrics and several key contextual elements, i.e., the definition of the target variable, data distribution, and temporal and spatial resolution. Thus, it outlines a first step toward providing more meaningful, comparable, and interpretable assessment results in crime prediction research.

Rather, this study shows that the real problem in crime prediction research is not the performance of algorithms but the lack of an interpretable and systematic structure of evaluation processes. The same metric values can have completely different meanings for different problem definitions. For instance, a high accuracy value is misleading under severe class imbalance; similarly, low error metrics, such as MAE, may be misleading when interpreted without considering the underlying crime distribution. Thus, performance metrics must be understood contextually rather than separately.

Therefore, a multi-layered evaluation framework is proposed to improve the interpretability of crime prediction models. This framework proposes that the following key elements be presented together to make the performance metrics meaningful:

Clear definition of the target variable (e.g., crime probability, crime rate, or crime type).

- Data distribution and class imbalance,
- Spatial resolution,
- Temporal resolution,
- Type of performance metrics,
- Validation setup used.

These elements provide the context in which performance metrics can be interpreted and, thus, allow for more reliable inter-study comparisons. Otherwise, models with the same metric values can sometimes denote different levels of performance.

Within the proposed ERF, a Coverage Score (CS) is introduced to quantify the completeness of evaluation reporting across studies. The CS is calculated as follows:

$$CS = \frac{Number\ of\ reported\ components}{Total\ number\ of\ components} \quad (3)$$

The proposed CS was intentionally designed as a simple binary indicator to provide a transparent and reproducible assessment of reporting completeness across studies. Accordingly, each reporting component was scored as either present (1) or absent (0). Although this approach does not distinguish between partial and complete reporting, it offers a straightforward basis for comparison. Future studies may further refine the framework by introducing weighted or multi-level scoring schemes that capture different degrees of reporting completeness.

A CS close to 1 indicates that the study reports the evaluation components more comprehensively, whereas a CS close to 0 indicates substantial shortcomings in evaluation reporting. Each element contextualizes the model output evaluation and shows the potential for misinterpretation if the elements are missing. This will minimize the excessive use of context-free metrics in the literature and make the evaluation process systematic, transparent, and comparable.

Based on Table 6, which summarizes the most cited and hence most influential papers in the domain, the application of the proposed evaluation framework yields an average CS of 4.67/6, or 0.78.

The targeted analysis indicates that reporting completeness varies across evaluation components rather than being uniformly poor. Among the six studies, class imbalance was the most consistently underreported element, followed by the temporal resolution, whereas performance metrics, spatial resolution, and validation setup were generally reported. In crime prediction studies, model performance interpretation is limited because the aspects defining performance, such as the target variable, spatial unit, and temporal resolution, are not defined.

Thus, performance metrics are difficult to interpret and compare reliably without explicitly stating whether the predictions are hourly, daily, or yearly and whether the spatial unit is street segments, grids, or aggregates at the city level. Target definitions should be provided along with crime occurrence rates or class imbalance within the prediction window. The validation setup should also be transparently reported since it is an important factor in determining the model's trustworthiness.

Within this targeted sample, the findings suggest that incomplete and decontextualized evaluation reporting remains an important methodological issue. Accordingly, the proposed ERF should be regarded as an exploratory diagnostic framework for identifying reporting gaps rather than as a definitive assessment standard. Broader validation using larger and systematically selected samples is warranted. Accordingly, the proposed framework provides an exploratory basis for assessing evaluation reporting practices and may serve as a methodological guide for future studies.

Limitations

This study, despite its contributions, has several limitations. First, the combined WoS and Scopus datasets have broad coverage of the literature, but relevant publications indexed in other databases may not have been captured. Second, the search strategy was

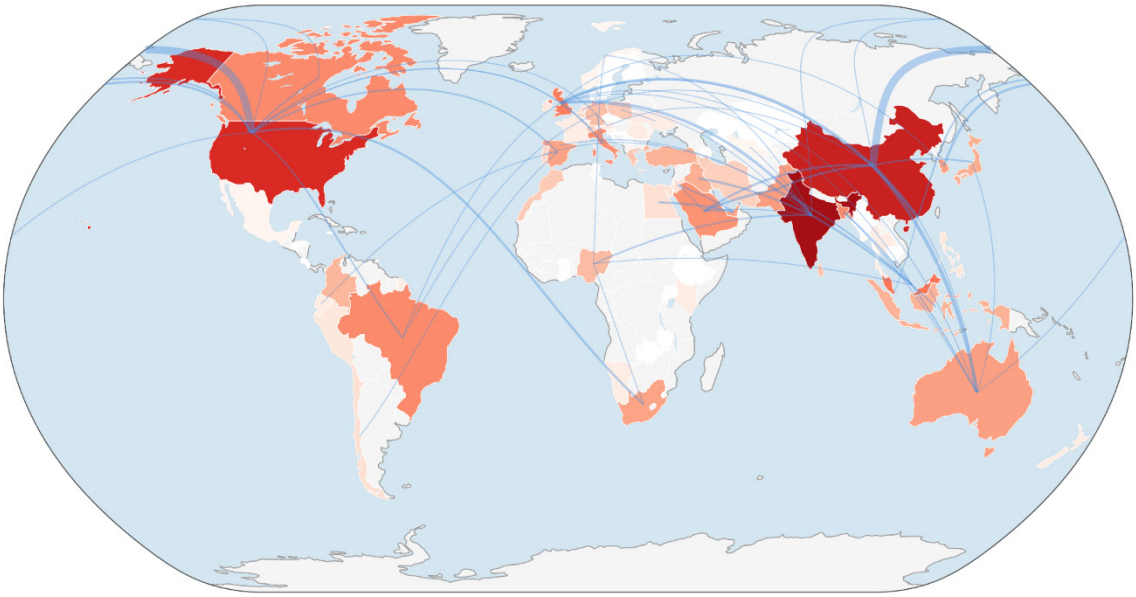


Figure 4: Mapping international collaboration in AI-based crime prediction research (2017-2026, 962 publications in Scopus, min. edges: 2, edge size: 5, Biblioshiny output).

Table 6: Evaluation matrix and CS of selected AI-based crime prediction studies.

Study	Target variable	Spatial resolution	Temporal resolution	Imbalance reported	Metrics reported	Validation setup	CS
Zhang *et al.* (2022)	1	1	1	1	1	1	6/6=1.00
Tang *et al.* (2023)	0	1	1	0	1	1	4/6=0.67
Deng *et al.* (2023)	1	1	0	0	1	1	4/6=0.67
Wu *et al.* (2022)	1	1	1	0	1	1	5/6=0.83
Saraiva *et al.* (2022)	1	1	0	0	1	1	4/6=0.67
Liang *et al.* (2022)	1	1	1	0	1	1	5/6=0.83

based on pre-defined keywords in English and we may have missed studies with different terminology or published in other languages. The proposed ERF was ultimately tested on a small, targeted sample with a simple binary scoring system and should therefore be viewed as an exploratory diagnostic framework that needs further validation with larger, more systematically selected samples. Hence, the results should be interpreted in the context of the selected databases, the retrieval strategy, and the exploratory application of the proposed ERF.

CONCLUSION

This study identifies a conceptual and methodological shift in AI-based crime prediction research through scientometric analyses. The results indicate a growing integration of data- and algorithm-based approaches with traditional criminological perspectives in the literature, as well as changes in methodological foci, validation processes, and research applications.

However, the results seem to suggest that the main problem in the literature is not with the algorithmic performance but with the evaluation and interpretation of the results of models. Performance metrics are often reported over data structures, model types, and units of analysis without the requisite context, making cross-study comparisons difficult.

More specifically, this study makes three complementary contributions to the literature. First, the proposed CSI provides an interpretable scientometric indicator to measure the conceptual change in crime prediction literature. Second, the MEN provides a transparent view of the methodological integration in terms of method co-occurrence patterns. Third, the proposed ERF offers an exploratory diagnosis framework for improving the transparency and comparability of evaluation reporting in AI-based crime prediction studies.

Taken together, these contributions offer a better, more comparable, more meaningful, and more interpretable basis for the evaluation of AI-based models of crime prediction. However, further development of AI-based crime prediction will be helped by the development of more advanced algorithms and by improving the interpretation and standardization of their outputs. In this respect, the study suggests the need for a methodological reframing of crime prediction studies.

Future research should validate and refine the proposed CSI, MEN, and ERF further with larger datasets, more bibliographic sources, and wider application domains. The use of standard evaluation reporting practices and the application of these methods to other research domains enabled by AI could further improve the transparency, comparability, and reproducibility of research using predictive modeling. The proposed CSI and MEN could be a useful framework for exploring conceptual and methodological changes in other fast-moving research areas beyond the research of crime prediction. Their application to different scientific fields can help to identify common as well as discipline-specific patterns of research evolution and thus support comparative scientometric analyses across disciplines.

DATA AVAILABILITY

The data used in this study were obtained from the WoS and Scopus databases. The data selection process is described in Appendix 1 to ensure reproducibility. Further information can be obtained from the corresponding author on request.

CODE AVAILABILITY

The scripts used to compute the Conceptual Shift Index (CSI) and to construct the Method Evolution Network (MEN) are publicly available at the author's GitHub repository (https://github.com/cem5113/crime_prediction_scientometrics).

ABBREVIATIONS

LDA: Latent Dirichlet allocation; **OECD:** Organisation for Economic Co-operation and Development; **GDP:** Gross Domestic Product; **UNEP:** United Nations Environment Programme; **SDG:** Sustainable Development Goal; **COVID-19:** Coronavirus Disease 2019; **DTM:** Document-Term Matrix; **h-index:** Hirsch Index; **g-index:** g-index; **R:** R Statistical Programming Language.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

Alomran, A., Basha, I., Hussain, A., Nasir, M., & Shaikh, M. (2025). Mapping of global aging and poverty research: a scientometric and visualization approach. Journal of Scientometric Research, 14(3), 821-837. https://doi.org/10.5530/jscires.20251955

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007

Aria, M., Cuccurullo, C., D'Aniello, L., & Spano, M. (2026). Biblioshiny and the SAAS workflow: an integrated framework for transparent and reproducible science mapping. A demonstration through the replication of a study. Journal of Informetrics, 20(3), 101837. https://doi.org/10.1016/j.joi.2026.101837

Bowers, K., Johnson, S., & Pease, K. (2004). Prospective hot-spotting - the future of crime mapping? British Journal of Criminology, 44(5), 641-658. https://doi.org/10.1093/bjc/azh036

Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. Justice Quarterly, 28(2), 360-381, Article 922514474. https://doi.org/10.1080/07418825.2010.486037

Chainey, S., Tompson, L., & Uhlig, S. (2008). The utility of hotspot mapping for predicting spatial patterns of crime. Security Journal, 21(1-2), 4-28. https://doi.org/10.1057/palgrave.sj.8350066

Deng, Y., He, R., & Liu, Y. (2023). Crime risk prediction incorporating geographical spatiotemporal dependency into machine learning models. Information Sciences, 646, 13, Article 119414. https://doi.org/10.1016/j.ins.2023.119414

Duplock, R., Casali, G. L., & McLennan, C. l. (2026). Beyond the narrative: structural mapping and thematic evolution in entrepreneurial intentions research. Journal of Scientometric Research, 15(1), 24-37. https://doi.org/10.5530/jscires.20260019

Dutta, E., Singh, S., Rath, J. P., Soodan, V., & Jassal, T. (2026). Mapping research trends and collaborations in sustainable development goals: a bibliometric analysis. Journal of Scientometric Research, 15(1), 62-77. https://doi.org/10.5530/jscires.20260237

Eroglu, C., & Cakir, H. (2024). The usability of stacking-based ensemble learning model in crime prediction: a systematic review. Crime Prevention & Community Safety, 26(4), 440-489. https://doi.org/10.1057/s41300-024-00222-7

Eroglu, C., & Cakir, H. (2026). Time-aware crime prediction: a multi-output ai framework for operational risk prioritization. IEEE Access, 14, 88405-88417. https://doi.org/10.1109/ACCESS.2026.3701892

Gerber, M. S. (2014). Predicting crime using Twitter and kernel density estimation. Decision Support Systems, 61, 115-125. https://doi.org/10.1016/j.dss.2014.02.003

Kaur, M., & Saini, M. (2024). Role of Artificial Intelligence in the crime prediction and pattern analysis studies published over the last decade: a scientometric analysis. Artificial Intelligence Review, 57(8), Article 202. https://doi.org/10.1007/s10462-024-10823-1

Liang, W., Wang, Y., Tao, H., & Cao, J. (2022). Towards hour-level crime prediction: A neural attentive framework with spatial-temporal-categorical fusion. Neurocomputing, 486, 286-297. https://doi.org/10.1016/j.neucom.2021.11.052

Mohamad, A. (2025). Mapping the intellectual landscape of big data in accounting and finance: a decade of bibliometric analysis (2013-2023). Journal of Scientometric Research, 14(1), 201-220. https://doi.org/10.5530/jscires.20251109

Mohler, G., Short, M., Brantingham, P., Schoenberg, F., & Tita, G. (2011). Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493), 100-108. https://doi.org/10.1198/jasa.2011.ap09546

Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. Journal of the American Statistical Association, 110(512), 1399-1411. https://doi.org/10.1080/01621459.2015.1077710

Perry, W. L. (2013). Predictive policing: the role of crime forecasting in law enforcement operations. Rand Corporation.

Saraiva, M., Matijosaitiene, I., Mishra, S., & Amante, A. (2022). Crime prediction and monitoring in porto, portugal, using machine learning, spatial and text analytics. ISPRS International Journal of Geo-Information, 11(7), 18, Article 400. https://doi.org/10.3390/ijgi11070400

Tang, J., Xia, L., & Huang, C. (2023). Explainable spatio-temporal graph neural networks. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (pp. 2432-2441). https://doi.org/10.1145/3583780.3614871

Travaini, G. V., Pacchioni, F., Bellumore, S., Bosia, M., & De Micco, F. (2022). Machine Learning and criminal justice: a systematic review of advanced methodology for recidivism risk prediction. International Journal of Environmental Research and Public Health, 19(17), Article 10594. https://doi.org/10.3390/ijerph191710594

Wheeler, A. P., & Steenbeek, W. (2021). Mapping the risk terrain for crime using machine learning. Journal of Quantitative Criminology, 37(2), 445-480. https://doi.org/10.1007/s10940-020-09457-7

Wu, J., Abrar, S., Awasthi, N., Frias-Martinez, E., & Frias-Martinez, V. (2022). Enhancing short-term crime prediction with human mobility flows and deep learning architectures. EPJ Data Science, 11(1), Article 53. https://doi.org/10.1140/epjds/s13688-022-00366-2

White, V., Rief, R., Hudgins, C. c., Pimsler, M. L., & Hunt, J. (2025). A synthesis of the 2021 NIJ forecasting challenge winning reports. https://www.ojp.gov/pdffiles1/nij/309826.pdf

Zhang, X., Liu, L., Lan, M., Song, G., Xiao, L., & Chen, J. (2022). Interpretable machine learning models for crime prediction. Computers Environment and Urban Systems, 94, 9, Article 101789. https://doi.org/10.1016/j.compenvurbsys.2022.101789

Cite this article: Eroglu C, Cakir H. AI in Crime Prediction Research: Scientometric Evidence of Conceptual Change, Method Evolution, and Evaluation Reporting Gaps. J Scientometric Res. 2026;15(3):1032-48.

Appendix 1: Study selection process: identification, screening, and eligibility.

		Overall crime prediction studies	AI-based crime prediction studies
Identification	Search Criteria	Time frame: no lower limit Search date: 1 May 2026 Study type: Journal articles and conference papers	Time frame: no lower limit Search date: 1 May 2026 Study type: Journal articles and conference papers
	Keywords	("crime predict*" OR "crime forecast*" OR "crime estimat*" OR "predict* crime" OR "forecast* crime" OR "estimat* crime" OR "predictive policing" OR "predictive police*" OR "hotspot forecast*" OR "hotspot predict*" OR "crime hotspot*" OR "risk terrain modelling" OR "RTM" OR "recidivism forecast*" OR "recidivism predict*" OR "repeat offender prediction" OR "early warning system*" OR "early warning model*" OR "victimization risk" OR "victim risk modelling" OR "risk of victimization" OR "offender risk prediction" OR "criminal career forecast*") AND crime	… AND ("artificial intelligence*" OR "machine learning*" OR "deep learning*" OR "artificial neural networks*" OR "neural networks*" OR "random forest*" OR "XGBoost*" OR "recurrent neural network*" OR "convolutional neural network*" OR "long short*term memory" OR "LSTM*" OR "reinforcement learning*" OR "transformer model*" OR "support vector machine*" OR "gradient boosting*" OR "naive bayes*" OR "bayesian network*" OR "autoencoder*" OR "generative adversarial networks*" OR "natural language processing*" OR "computer vision*" OR "predictive analytics*" OR "recommendation system*" OR "feature engineering*" OR "hyperparameter optimization*" OR "tensorFlow*" OR "PyTorch*" OR "keras*" OR "scikit*learn*" OR "LightGBM*")
	Identified	WoS (*n*=1,722) and Scopus (*n*=15,227)	WoS (*n*=489) and Scopus (*n*=6600)
Screening		Keywords were searched again in the 'Abstract, Keywords and Title' section in WoS (*n*=1,423) and Scopus (*n*=2,667)	Keywords were searched again in the 'Abstract, Keywords and Title' section in WoS (*n*=364) and Scopus (*n*=999)
Eligibility	Selected Study	Deleted 1,269 duplicate records, (n=2,821)	Deleted 328 duplicate records, (n=1,035)

Completeness of metadata -- **2821** docs merged from 2 DBs

Original size **4090** docs -- Deleted **1269** duplicated docs

Metadata	Description	Missing Counts	Missing %	Trend	Status
DT	Document Type	0	0.00	1970 2026	Excellent
SO	Journal	0	0.00	1970 2026	Excellent
LA	Language	0	0.00	1970 2026	Excellent
PY	Publication Year	0	0.00	1970 2026	Excellent
TI	Title	0	0.00	1970 2026	Excellent
TC	Total Citation	0	0.00	1970 2026	Excellent
AU	Author	1	0.04	1970 2026	Good
AB	Abstract	8	0.28	1970 2026	Good
C1	Affiliation	32	1.13	1970 2026	Good
DI	DOI	264	9.36	1970 2026	Good
DE	Keywords	376	13.33	1970 2026	Acceptable
RP	Corresponding Author	471	16.70	1970 2026	Acceptable
ID	Keywords Plus	833	29.53	1970 2026	Poor
WC	Science Categories	1400	49.63	1970 2026	Poor
CR	Cited References	2821	100.00	1970 2026	Completely missing

Showing 1-15 of 15

Appendix 2A: Metadata completeness assessment of the merged bibliographic dataset: Overall crime prediction dataset (*n*=2,821).

Completeness of metadata -- **1035** docs merged from 2 DBs

Original size **1363** docs -- Deleted **328** duplicated docs

Metadata	Description	Missing Counts	Missing %	Trend	Status
DT	Document Type	0	0.00	1997 2026	Excellent
SO	Journal	0	0.00	1997 2026	Excellent
LA	Language	0	0.00	1997 2026	Excellent
PY	Publication Year	0	0.00	1997 2026	Excellent
TI	Title	0	0.00	1997 2026	Excellent
TC	Total Citation	0	0.00	1997 2026	Excellent
AB	Abstract	1	0.10	1997 2026	Good
AU	Author	1	0.10	1997 2026	Good
C1	Affiliation	5	0.48	1997 2026	Good
DI	DOI	72	6.96	1997 2026	Good
DE	Keywords	80	7.73	1997 2026	Good
RP	Corresponding Author	251	24.25	1997 2026	Poor
ID	Keywords Plus	280	27.05	1997 2026	Poor
WC	Science Categories	672	64.93	1997 2026	Critical
CR	Cited References	1035	100.00	1997 2026	Completely missing

Showing 1-15 of 15

Appendix 2B: Metadata completeness assessment of the merged bibliographic dataset: AI-based crime prediction dataset (*n*=1,035).

Appendix 3: Country-level collaboration indicators before and after the AI transition (Top 25).

Rank	Pre-AI period (1980-2016) - Overall crime prediction studies (631 publications)	Collaboration links	Rank	AI period (2017-2026) - AI-based crime prediction studies (962 publications)	Collaboration links
1	USA	56	1	USA	71
2	United Kingdom	52	2	China	70
3	Netherlands	21	3	India	37
4	Australia	19	4	United Kingdom	37
5	China	19	5	Malaysia	33
6	Germany	16	6	Australia	26
7	Canada	14	7	Pakistan	22
8	Spain	12	8	Saudi Arabia	22
9	Sweden	10	9	Italy	21
10	Ireland	10	10	Hong Kong	20
11	Slovenia	10	11	Spain	17
12	France	9	12	United Arab Emirates	17
13	Israel	7	13	Nigeria	15
14	Italy	7	14	Canada	14
15	Denmark	7	15	Brazil	13
16	Norway	6	16	South Africa	12
17	Portugal	6	17	Korea	12
18	Brazil	5	18	Bangladesh	10
19	Saudi Arabia	4	19	Japan	9
20	Belgium	4	20	Iraq	9
21	Hong Kong	3	21	Poland	9
22	Turkey	3	22	Slovenia	9
23	Pakistan	3	23	Germany	9
24	India	3	24	Austria	8
25	Korea	2	25	Jordan	8