UWL REPOSITORY

repository.uwl.ac.uk

Intelligent Eyes on Buildings: A Scientometric Mapping and Systematic Review of AI-Based Crack Detection and Predictive Diagnostics of Building Structures

Mohagheghi, Mehdi, Bahadori-Jahromi, Ali ORCID logoORCID: https://orcid.org/0000-0003-0405-7146 and Room, Shah (2026) Intelligent Eyes on Buildings: A Scientometric Mapping and Systematic Review of AI-Based Crack Detection and Predictive Diagnostics of Building Structures. Encyclopedia, 6 (4).

https://doi.org/10.3390/encyclopedia6040075

This is the Published Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/14812/

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

encyclopedia

MDPI

Systematic Review

Intelligent Eyes on Buildings: A Scientometric Mapping and Systematic Review of AI-Based Crack Detection and Predictive Diagnostics of Building Structures

Mehdi Mohagheghi *, Ali Bahadori-Jahromi and Shah Room

Department of Civil Engineering and Built Environment, School of Computing and Engineering, University of West London, St Mary's Road, Ealing, London W5 5RF, UK; ali.jahromi@uwl.ac.uk (A.B.-J.); shah.room@uwl.ac.uk (S.R.)
* Correspondence: mehdi.mohagheghi@research.uwl.ac.uk

Abstract

Artificial Intelligence (AI)-based crack detection in buildings uses computer vision and deep learning to automatically identify structural cracks from inspection images. In recent years, many studies have explored this topic, but the overall development of the field, its methodological practices, and the remaining challenges are still not fully clear. Unlike most previous reviews that focus mainly on technical methods, this study combines a large-scale scientometric mapping of the research field with a focused technical analysis of recent AI-based crack detection methods specifically applied to building structures. This study therefore provides a dual-layer review covering research published between 2015 and 2025. A total of 146 Scopus-indexed publications were analysed using Visualization of Similarities viewer (VOSviewer) to examine publication growth, thematic evolution, collaboration patterns, and citation structures. In addition, a focused technical review of 36 highly relevant studies was carried out to analyse task formulations, model families, datasets, evaluation protocols, and methodological practices. The results show a rapid increase in research activity after 2020, largely driven by advances in deep-learning and Unmanned Aerial Vehicle (UAV)-based inspections. At the same time, collaboration networks remain uneven, and citation influence is concentrated in a limited number of research communities. The technical review further shows that most studies focus on detection-level tasks, particularly You Only Look Once (YOLO)-based models, while predictive diagnostics, automated inspection reporting, and decision-oriented Structural Health Monitoring (SHM) are still rarely addressed. Current datasets and evaluation protocols also remain mostly perception-oriented, which makes it difficult to assess robustness, generalisability and long-term predictive capability.

Keywords: AI; computer vision; scientometric analysis; crack detection; building inspection; predictive diagnostics; deep learning

Check for updates

Academic Editor: Antonio Formisano

Received: 1 February 2026
Revised: 15 March 2026
Accepted: 20 March 2026
Published: 27 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Ensuring the long-term safety and performance of buildings has become increasingly essential as urban regions continue to expand and building components deteriorate due to ageing, environmental exposure, and material fatigue. Among various types of surface deterioration, cracks are widely recognised as the earliest and most informative indicators of structural distress in buildings. Continuous monitoring of crack formation and growth is therefore essential to help prevent progressive damage and to maintain structural reliability

https://doi.org/10.3390/encyclopedia6040075

over time [1]. Previous studies on building condition assessments report that traditional inspection practices continue to rely primarily on manual visual assessments and handheld measurement tools. Such approaches have been associated with high labour demands and observer-dependent variability, which can limit their suitability for large-scale or frequent inspection tasks [2]. Recent work in AI, computer vision (CV), and deep learning (DL) has increasingly focused on automated methods for identifying and analysing building cracks. Within this body of research, convolutional neural networks (CNNs) are frequently reported to provide reliable recognition results across challenging surface textures, variable lighting conditions, and noisy image data [3]. Also, several recent studies have applied object detection frameworks from the YOLO family to real-time crack localisation, particularly in inspection scenarios using drone imagery or handheld cameras Figure 1. These studies report improved practicality compared with manual inspection workflows [4].

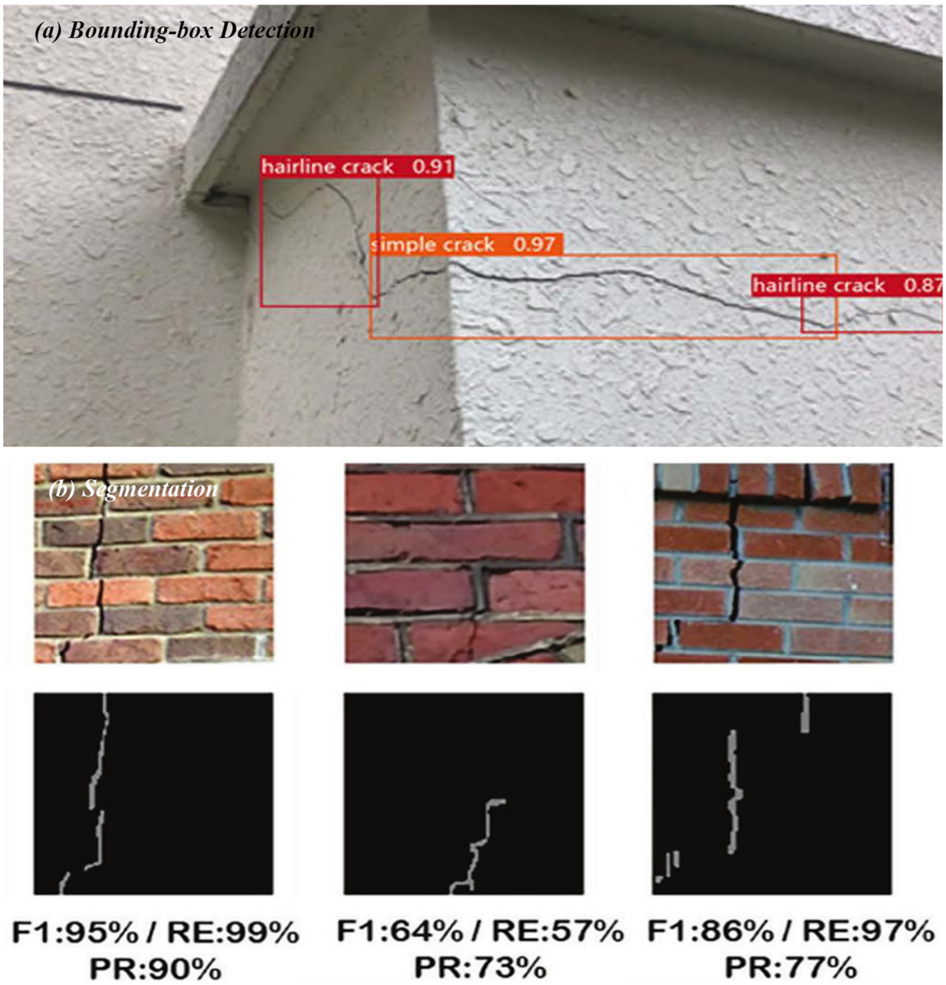

Figure 1. Examples of AI-based building crack detection: (**a**) bounding-box detection (adapted from ref. [5]); (**b**) segmentation (adapted from ref. [6]).

In addition to object detection-based approaches, several recent studies have investigated the use of transfer learning to enhance detection accuracy in building defect inspection tasks, particularly when available datasets are small or highly specialised. This strategy has been reported as a practical option in contexts where access to large, fully annotated datasets remains limited [7]. Some recent studies have also explored Generative Adversarial Networks (GANs) as a way to mitigate data imbalance and limited training samples in damage and defect detection tasks, including in related engineering and industrial inspection contexts. However, these approaches are still relatively limited in building-specific crack detection studies [8–10].

Alongside these methodological approaches, recent work has also considered how inspection data are collected in practice. Drones, smartphones, and high-resolution cameras are widely discussed as practical options in building inspection studies. In some cases, UAV-mounted cameras are able to capture façade images with enough clarity, which can then be used for automated crack detection tasks [11]. Recent studies have also examined the use of lighter deep-learning models and edge-AI devices for real-time crack recognition on mobile platforms. These approaches are reported to support on-site building inspection workflows, especially where real-time processing is required [12]. Semantic segmentation models, including the U-Net architecture and DeepLab families, have also helped improve pixel-level outlining of fine cracks. This has made more detailed defect quantification and modelling possible [13].

As these methods and technologies developed, more researchers started to focus on AI-based crack detection in buildings, with many related studies published between 2020 and 2024 [3]. Multiple studies have focused on topics such as CNN architectures, segmentation accuracy, drone-assisted inspection, and mobile-based detection systems, and they are reflecting increasing interest in automated approaches to building assessments [14]. Despite the growing number of studies, the existing literature remains fragmented. While many studies focused on algorithmic performance, image processing techniques, or specific façade conditions, broader research patterns, thematic clusters, and collaboration structures are still unclear [15]. Moreover, existing review studies have mainly examined isolated technical or application-specific aspects of crack detection, and as a result, they do not provide a unified, building-oriented view of AI-based crack analysis [1,3,16–25].

Whitin this broader context, automated crack detection has become a core element of modern SHM. As buildings age and urban environments grow denser, the need for scalable, objective, and automated inspection technologies becomes increasingly clear. Based on prior research, computer-vision and machine-learning methods are widely used for infrastructure monitoring and SHM, offering efficient strategies for defect detection and structural assessments of bridges, pavements, and other civil systems [26–29]. In the case of buildings, unique characteristics, such as diverse materials, localised damage patterns, and operational environments, introduce challenges that are not fully addressed by broader infrastructure-oriented reviews. Because of this gap, several aspects of AI-based building crack detection are still discussed separately in the literature, which makes a focused and systematic review necessary. Such a review can help connect scientometric trends with technical methods and provide a clearer direction for future research and practical applications.

Existing review papers usually provide narrative or technical summaries of the literature and can be helpful for understanding specific methods or applications. Also, many of these reviews are based on a limited number of studies, which means they do not always capture how the research field is structured at a wider level. Aspects such as patterns of influence, collaboration among researchers, and the gradual development of knowledge are often only briefly addressed. Scientometric analysis looks at the literature from a different angle by relying on quantitative, data-driven methods to study scientific activity, making it possible to identify relationships and structures that may not be obvious in narrative reviews [30]. This approach uses quantitative data taken from publications, citation records, and research networks to examine how the field has changed and developed over time [31].

In this study, we examine research on AI-based crack detection in buildings using two main perspectives. First, we use scientometric analysis to look at overall research activity, including how studies are distributed over time, how researchers collaborate, and which topics receive the most attention. We then review the technical aspects of existing work more closely, focusing on methodological choices, evaluation settings, and modelling strategies. By combining these two views, the study provides both a general overview of the field and a more detailed understanding of how technical approaches have evolved [32,33]. We organise this study around two complementary objectives. Using this dual approach, we aim to link broader scientific patterns with detailed technical evidence in building structural health monitoring.

As described by Ellegaard and Wallin (2015), scientometric analysis can be used to study research output, identify influential researchers, and examine how knowledge develops and spreads within a field. They also note that this type of analysis can reveal patterns that are often not clear in traditional narrative reviews [34]. While scientometric techniques have been applied to broader research areas, including artificial intelligence in civil engineering, crack analysis in concrete structures and pavements, and automated building inspections, fewer studies have focused specifically on AI-based building crack detection [35].

Structural health monitoring systems can rely on different types of data, including vibration signals, acoustic measurements, strain sensing, and visual inspection data [36]. However, in recent years, vision-based SHM approaches have become increasingly prominent due to advances in computer vision and deep learning. These approaches use image data collected through cameras, drones, or inspection platforms to automatically detect structural defects such as cracks. The present review primarily focuses on this vision-based branch of SHM, where AI and deep-learning models are applied to image data for crack detection in building structures.

So far, AI-based crack detection in buildings has not been the subject of a dedicated scientometric analysis. Despite the increasing number of related studies, the literature is still dispersed across multiple research areas. A targeted scientometric review may help organise this work, provide a clearer view of global research patterns, and identify directions that merit further investigation. Such an analysis can reveal how topics such as real-time detection, drone-based imaging, segmentation modelling, and lightweight architectures have evolved, while also identifying underexplored areas that warrant future research. Given this dispersed research landscape and the absence of a comprehensive and building-focused scientometric understanding of the domain, it becomes important to examine how studies on AI-based crack detection in building structures have collectively developed over time. The present study attempts to address the following central question: how has research on AI-based crack detection in building structures evolved over the past decade in terms of its scientific growth, thematic organisation, and patterns of collaboration? Therefore, the first part of this study aims to conduct a comprehensive scientometric analysis of AI-based crack detection research in building structures from 2015 to 2025. The objectives of this part are to:

(1) Analyse publication growth and citation dynamics.
(2) Identify leading authors, institutions, and countries in the field.
(3) Map collaboration networks at the author, institutional, and national levels.
(4) Examine keyword co-occurrence patterns and thematic clusters.
(5) Identify emerging trends, methodological developments, and research hotspots.
(6) highlight research gaps and propose future research directions.

While these objectives address the large-scale scientometric structure of the field based on the full corpus of 146 Scopus-indexed publications, they do not provide sufficient insight into the detailed technical choices adopted in state-of-the-art AI-based crack detection and predictive diagnostics studies. And, in a second layer of analysis, this study conducts a focused systematic technical review of a core subset of the most recent and thematically relevant publications. This technical review is guided by the following research questions (TRQs).

TRQ1—Application domains and problem formulations.

What types of crack detection, defect characterisation, and predictive diagnostic tasks have been formulated in recent AI-based studies for buildings?

TRQ2—AI and machine-learning model families.

Which families of machine-learning and deep-learning models have been employed, and how do their reported performances compare?

TRQ3—Datasets and data acquisition settings.

What are the main characteristics of the datasets used in these studies in terms of data modality, size, annotation strategy, and public availability?

TRQ4—Evaluation protocols and performance metrics.

How are training and evaluation protocols designed, and which performance metrics are used to assess model effectiveness?

TRQ5—Hyperparameter optimisation and methodological rigour.

To what extent are systematic hyperparameter optimisation and robust methodological practices reported in the core studies?

Structure of the Paper

The rest of this paper is organised as follows: Section 2 outlines the data collection process and scientometric methodology. Section 3 presents the bibliometric and network analysis results. Section 4 reports the systematic technical review results. Section 5 provides an in-depth discussion of the thematic clusters, methodological insights, and emerging trends. Finally, Section 6 concludes the paper by summarising the main contributions and offering directions for future research.

2. Materials and Methods

This study adopts a dual-method framework, consisting of (1) a scientometric analysis of the entire publication landscape and (2) a systematic technical review of a core subset of recent high-relevance studies. This integrated approach allows the study to capture both macro-level scientific trends and micro-level methodological practices. The methodological framework of this study follows established scientometric and bibliometric protocols [33] and adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines for transparent reporting of literature selection [37], with the completed PRISMA 2020 checklist provided in Appendix A. A detailed PRISMA flow diagram is provided in Appendix B [38].

The workflow consists of four major phases: (i) database selection and search strategy, (ii) data collection, (iii) screening and eligibility assessment, and (iv) bibliometric and scientometric analysis. Figure 2 presents the flowchart for the scientometric study.

2.1. Database Selection

Scopus was selected as the primary source database due to its comprehensive coverage of engineering, its computer science and built environment publications, its structured metadata, and its wide adoption in scientometric studies within civil engineering and AI research [39,40]. In addition, Scopus provides strong coverage of interdisciplinary studies and conference publications in these areas. Searches in other databases (e.g., Web of Science) showed a high level of overlap with Scopus-indexed records. Therefore, restricting the final dataset to Scopus helped maintain consistent metadata, which is important for reliable scientometric analysis. Similarly, several previous scientometric reviews in construction engineering used Scopus as the exclusive or primary database [41,42].

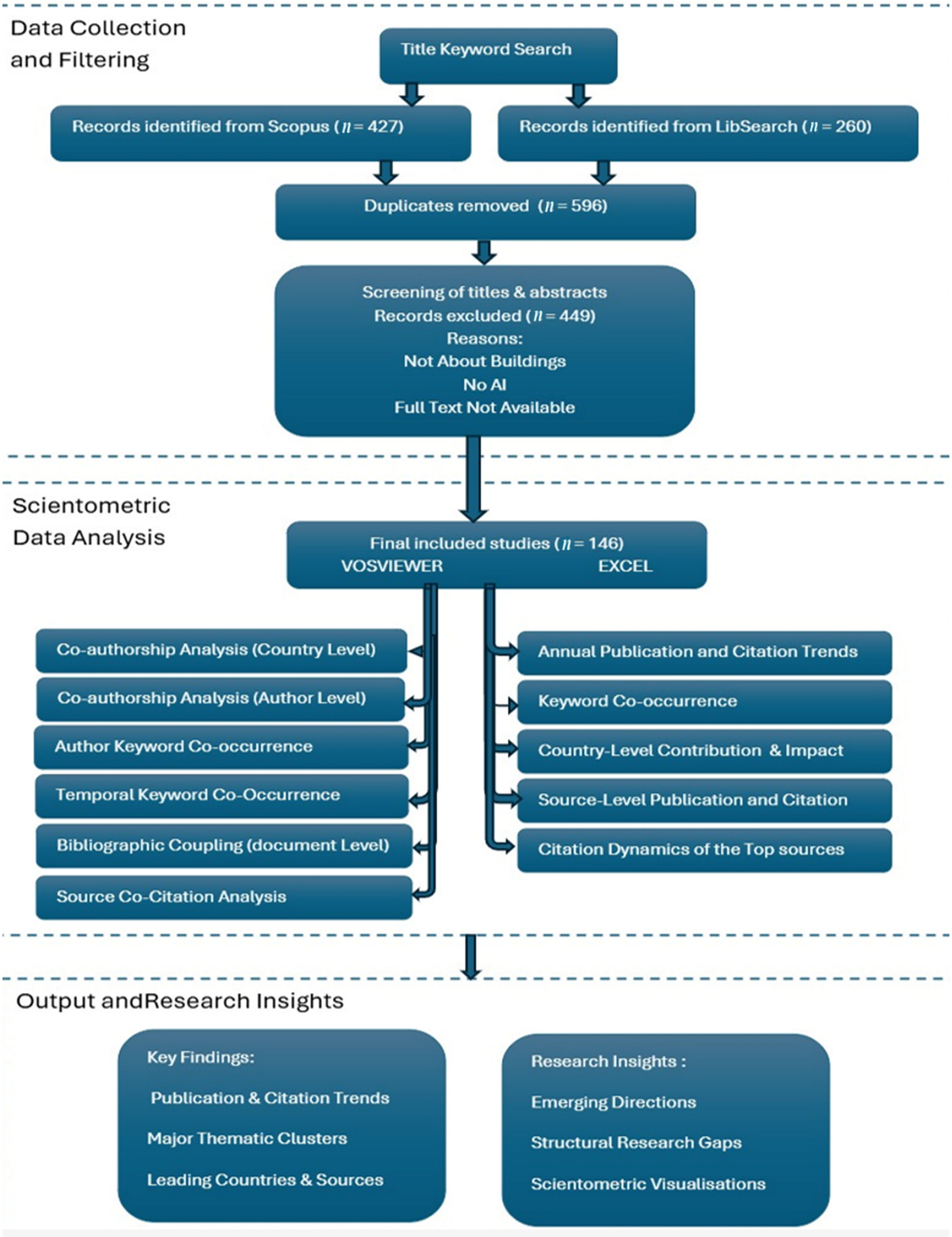

Figure 2. Flowchart for the scientometric study.

To supplement Scopus and ensure that no relevant studies were missed due to variations in keyword indexing, an additional search was conducted through the University of West London's LibSearch system, an EBSCO-powered unified academic search platform. LibSearch aggregates multiple databases (e.g., EBSCOhost Engineering Collection, ProQuest Engineering, Wiley, and SpringerLink) and therefore increases sensitivity for interdisciplinary topics.

Only studies retrievable in Scopus (via Digital Object Identifier (DOI) or Electronic Identifier (EID) lookups) were retained for the final dataset to maintain metadata consistency for bibliometric mapping, a standard requirement in scientometric methodology [30].

2.2. Search Strategy

All searches were restricted to:

(1) Years: 2015–2025.
(2) Language: English.
(3) Document types: Articles, reviews, conference papers, and conference reviews.

According to previous studies, restricting the search to title-only terms significantly increases precision and reduces false positives in engineering-focused scientometric reviews [32]. Therefore, the search was performed in the TITLE fields using a Boolean operator. The full Boolean string is provided in Appendix C. A parallel search using identical conceptual keywords was conducted in LibSearch, yielding 260 records. All records were exported, including DOI, title, authors, year, and source information.

2.3. Data Cleaning and Deduplication

All records from Scopus and LibSearch were merged in Microsoft Excel. Duplicate publications were identified through DOIs, title similarity, or Scopus EID codes, following recommendations by [33]. After removing duplicates, 596 unique documents remained.

Some LibSearch records were not detected in Scopus through the keyword search due to variations in indexing or different terminology. To ensure dataset completeness, these items were manually checked in Scopus using a DOI lookup, which is recommended when conducting multi-database integration in scientometric studies [39].

2.4. Screening and Eligibility Assessment

Title and abstract screening were performed manually following the PRISMA 2020 guidelines [37]. The exclusion criteria were:

(1) Topics not related to buildings (e.g., bridges, tunnels, pavements, and pipelines).
(2) Studies not applying AI, Machin Learning (ML), DL, or computer vision.
(3) Studies unrelated to structural health monitoring, inspection, or defect detection or defect progression prediction.
(4) Full text not accessible through University of West London (UWL) library subscriptions.

A total of 449 records were excluded. The final dataset consisted of 146 studies, forming the basis for bibliometric and scientometric analysis (Table 1).

Table 1. Summary of the search results and screening process.

Stage	**Count**
Scopus Search Results (A)	427
LibSearch Search Results (B)	260
After Merging A and B	595
After Screening	146

2.5. Bibliometric and Scientometric Analysis

Scientometric mapping and network visualisation were conducted using VOSviewer (v1.6.20), a widely used tool for co-authorship, co-citation and co-occurrence analysis. The following analyses were performed:

(1) Co-authorship networks (authors, institutions, countries).
(2) Co-citation networks.

(3) Bibliographic coupling.
(4) Keyword co-occurrence mapping.
(5) Temporal keyword evolution.
(6) Publication trend analysis (2015–2025).
(7) Source and publisher impact analysis.

Microsoft Excel was used to support descriptive analytics, trend plots, year-wise distributions, and data preprocessing. Descriptive statistical analyses, including publication trends, citation distributions, and source impact charts, were generated using Microsoft Excel, which is commonly employed in scientometric studies for basic performance analysis. This mixed-methods scientometric approach aligns with best practices established in the prior literature [32].

2.6. Systematic Technical Review Method

In addition to the scientometric mapping, this study conducted a systematic technical review of a core subset of publications to examine methodological practices in depth and answer TRQ1–TRQ5. From the initial corpus of 146 scientometric publications, a structured scoring-based selection process was applied. Each article was manually screened by reviewing its title, abstract, and, when necessary, the introduction, methodology, and conclusion sections. Articles were assigned a relevance and methodological detail score on a 1–5 scale, where:

(1) A score of 1–2 = low relevance (e.g., non-building domain or insufficient methodological reporting);
(2) A score of 3 = moderate relevance;
(3) A score of 4–5 = high relevance (directly addressing AI-based crack detection in buildings with substantial methodological detail).

Only studies scoring 4 or 5 were retained, yielding 36 core publications, published primarily between 2023 and 2025, with several highly relevant earlier works (2020–2022) preserved for their methodological significance.

We conducted the technical review by following the systematic review principles proposed by Kitchenham and Charters [43] and used the PRISMA 2020 guidelines to improve transparency and reproducibility [37]. We defined the inclusion criteria to focus on studies that address crack detection, segmentation, or defect prediction in building structures and that apply AI-, ML-, or DL-based methods. We also required the selected studies to report enough methodological detail to allow for data extraction, including information on datasets, model architectures, evaluation metrics, validation strategies, and hyperparameter optimisation. We excluded studies that focused on cracks in bridges, pavements, tunnels, or laboratory-only specimens, as well as those that did not use AI-based approaches or failed to provide sufficient methodological detail. We also applied a detailed data extraction protocol to the 36 selected papers. For each study, we collected information on the application domain, preprocessing steps, model architectures, comparative models, and model families. We also extracted dataset-related details, including dataset type, source, size, and annotation strategy, along with performance metrics, validation protocols, and approaches to hyperparameter optimisation. Where available, deployment-related details, including hardware specifications and access to open data or code, were noted. In addition to these quantitative variables, we recorded qualitative information, including reported limitations, failure cases, and directions for future work. This approach supports a structured synthesis of the literature in line with common practices in systematic mapping studies [44,45].

For clarity and consistent referencing throughout the technical review, each of the 36 included studies was assigned a unique identifier (D1–D36). The full list of identifiers and corresponding bibliographic details is provided in Table 2.

Table 2. Identifier list of the 36 studies included in the technical review (D1–D36).

No.	Paper	Ref	Year	No.	Paper	Ref	Year	No.	Paper	Ref	Year
D1	Shin, c.	[46]	2025	D13	Mariniuc, a.m.	[47]	2024	D25	Garrido, i.	[48]	2022
D2	Dinh, n.n.h.	[49]	2024	D14	Wang, j.	[50]	2025	D26	Ren, w.	[5]	2025
D3	Wang, j.;	[51]	2025	D15	Ding, w.	[52]	2023	D27	He, y.	[53]	2024
D4	Wang, j.	[54]	2024	D16	Ribeiro, w.s.	[55]	2024	D28	Busheska, a.	[56]	2023
D5	Kottari, p.	[57]	2024	D17	Huo, z.	[58]	2024	D29	Tan, y.	[59]	2022
D6	Wei, g.	[60]	2023	D18	Shi, t.	[61]	2025	D30	Akgül, i.	[62]	2023
D7	Yadav, d.p.	[63]	2024	D19	Zhou, x.	[64]	2025	D31	Lee, k.	[65]	2020
D8	Ren, w.	[66]	2025	D20	Ramkumar, g.	[67]	2024	D32	Asif, k.m.s.	[68]	2025
D9	Keerthana, b	[69]	2024	D21	Interlando, m.	[70]	2024	D33	Bian, x.	[71]	2024
D10	Tan, y.	[72]	2024	D22	Liu, j.; mao, p.	[73]	2025	D34	Wang, h.	[74]	2023
D11	Xue, s.	[75]	2024	D23	Angan, r.b.	[76]	2023	D35	Tang, s.	[77]	2024
D12	Chen, y.	[78]	2023	D24	Munawar, h.s.	[79]	2022	D36	Dissanayake, d.m.k.i	[80]	2024

2.7. Quality Assessment

To check the overall quality of the selected studies, a basic quality appraisal was used. The appraisal focused on three aspects: risk of bias, reporting completeness, and reproducibility. Studies were considered to have a lower risk of bias when they clearly described their datasets, model architectures, and evaluation procedures. Reporting completeness was checked by looking at whether key methodological details, such as preprocessing steps, validation strategies, performance metrics, and hyperparameter settings, were clearly described. Reproducibility was examined by considering access to implementation details and, where available, open datasets or code repositories. This approach is in line with common practice in systematic reviews [37,43,81].

For data synthesis, we combined tabular summaries with a narrative comparison of the selected studies. The tables were used to summarise model families, datasets, evaluation metrics, and optimisation procedures, while the narrative comparison helped highlight common methodological patterns, recurring challenges, and remaining gaps across the literature. This combined strategy supports a clear and consistent way of addressing the technical review questions (TRQ1–TRQ5) and also links the technical findings to the broader results of the scientometric analysis. To organise the analysis, this section is split into two parts. The first part summarises the scientometric results based on the full set of 146 publications. The second part focuses on methodological insights drawn from the technical review of 36 selected studies. By looking at both aspects together, the analysis covers the broader research landscape while also examining recent technical developments in more detail.

3. Scientometric Results

This section reports the results of the scientometric analysis based on 146 publications related to AI-based crack detection and predictive diagnostics in building structures. The results are discussed in two parts. First, basic bibliometric indicators are used to describe publication trends, country-level contributions, and the distribution of publication sources. The second part focuses on science mapping results obtained using VOSviewer, which help illustrate key research themes, cluster structures, and changes in research focus over time.

3.1. Annual Publication and Citation Trends (2015–2025)

Figure 3 shows the annual number of publications, the total citations received by papers published each year, and the average citations per document between 2015 and 2025. Overall, the figure indicates a clear increase in research activity related to AI-based crack detection in buildings. Publication output remained relatively low from 2015 to 2020, with between 2 and 8 papers published per year. A clear change can be observed from 2022 onwards, when the number of publications rose to 19, corresponding to a 137% increase compared with 2021. This upward trajectory continued in 2023, reaching 23 publications, and ultimately peaked in 2024 with 44 documents, the highest productivity recorded in the dataset. While the publication count appears to decline in 2025 (30 documents), this reduction is most likely due to incomplete indexing for the current year rather than a true decrease in research activity.

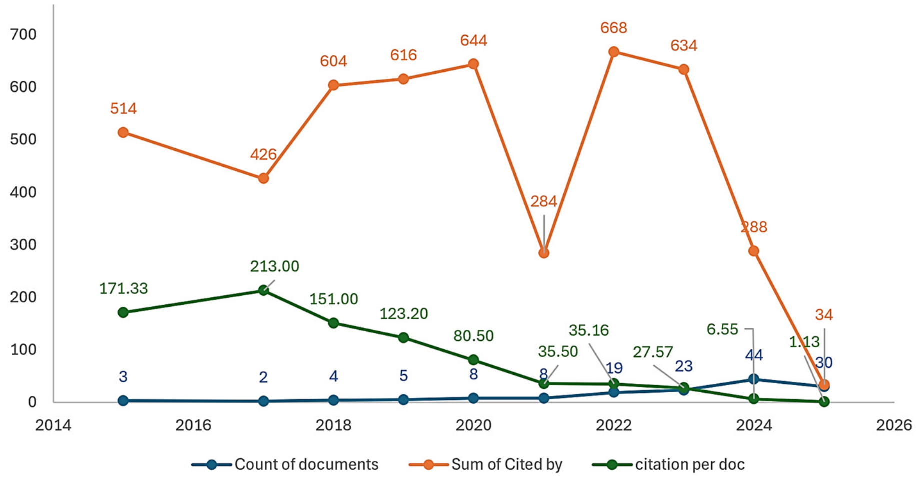

Figure 3. Annual publication output, total citations, and average citations per document (2015–2025).

To correct for temporal variation in exposure time, the average citation per document offers a more balanced indicator of scientific visibility and influence. As expected in a time-dependent citation process, early publications display significantly higher averages, such as 171 citations per document in 2015 and 213 in 2017, because they have been available for substantially longer periods and have served as foundational references in the field. From 2018 onward, the average citation per document gradually declines, not necessarily due to reduced research impact but primarily because (i) the number of publications increases, and (ii) newer studies have had far less time to accumulate citations. For example, the average falls to 35 citations per document in 2021, 6.55 in 2024, and 1.13 in 2025, reflecting the normal citation lag typical of rapidly expanding scientific domains. Therefore, the downward trend in average citation values should be interpreted as a function of a shorter exposure time and higher publication volume, rather than diminished relevance or quality of recent outputs.

3.2. *Keyword Co-Occurrence Analysis (2015–2020 vs. 2021–2025)*

To capture changes in research priorities, keyword frequencies were examined separately for 2015–2020 and 2021–2025. The results show a clear thematic shift from early methodological exploration toward specialised, application-driven crack detection research.

3.2.1. Early Research Themes (2015–2020) (Figure 4)

In the first period, research output was relatively limited, and the most common keywords reflect broad conceptual experimentation. "Deep learning" ($n = 4$) was the most frequent term, followed by "structural health monitoring" ($n = 3$) and "monitoring" ($n = 3$). Some core methodological terms appear less frequently in the dataset, including "damage detection," "neural networks," "digital image processing," and "transfer learning" (each $n = 2$). Less frequent terms such as "concrete deterioration," "Class Activation Mapping

(CAM),” and “ambient temperature” (each n = 1) show that these topics have received relatively limited attention so far.

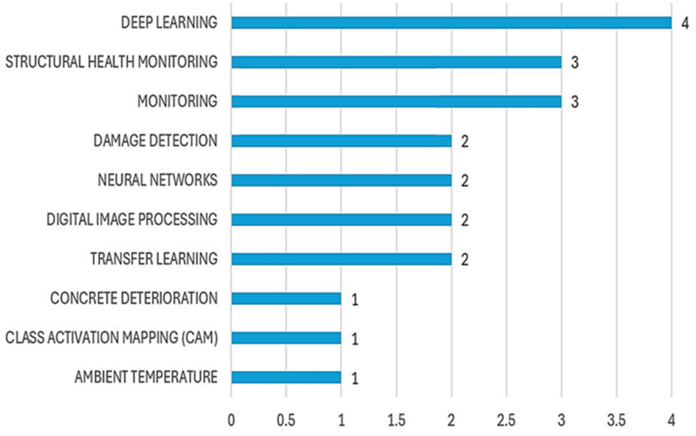

Figure 4. Top 10 keywords in early-stage research (2015–2020).

The results indicate that research activity between 2015 and 2020 was still at an early stage, with the field mainly focusing on:

(1) The general adoption of deep-learning techniques;
(2) A limited crack-specific focus;
(3) Early conceptual work on defect detection within broader SHM contexts.

3.2.2. Expansion and Application-Oriented Themes (2021–2025) (Figure 5)

In the second period, both the number of publications and the level of thematic focus increase noticeably. “Deep learning” becomes the most frequent term, appearing 34 times, which shows its growing use in crack and defect detection studies. At the same time, domain-specific keywords appear more often, particularly “crack detection” (n = 12) and “damage detection” (n = 9).

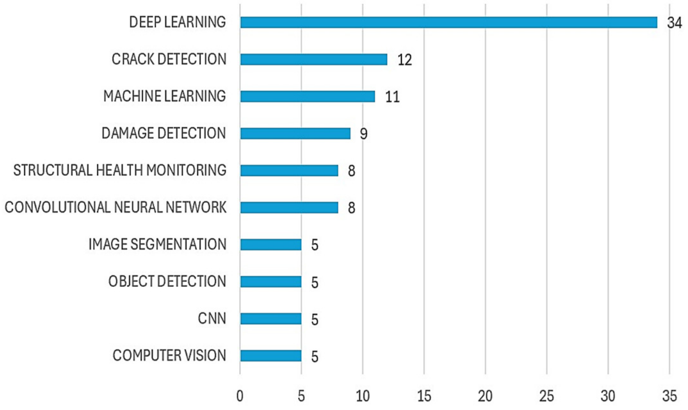

Figure 5. Top 10 keywords in recent research (2021–2025).

3.2.3. Thematic Evolution Summary

A comparison of both periods reveals a clear trajectory. The evolution of keywords across the two periods is summarised in (Table 3).

Overall, the field has transitioned from general AI adoption to highly specialised crack detection pipelines, driven by the rapid rise in deep-learning architectures and real-world deployment via computer-vision and UAV-based SHM systems.

Table 3. Keyword evolution by period.

Period	Dominant Keywords	Interpretation
2015–2020	DL (4), SHM (3), Monitoring (3), and NN/Transfer Learning/DIP (2)	Early conceptual and methodological development
2021–2025	DL (34), Crack Detection (12), ML (11), Damage Detection (9), CNN (8), and Segmentation/Object Detection/CV (5)	Mature, application-driven, and model-centric research

3.3. Country-Level Contribution and Scientific Impact

A cross-country comparison reveals substantial variation in both research productivity and citation influence in AI-based crack detection (Figure 6). China remains the dominant contributor, producing 54 documents, far more than any other country. However, its total citations (1054) and modest average impact (19.52 citations per article) place it below the global mean (32.27), indicating high output but only moderate scientific visibility. South Korea and the United States show similar publication output, with 12 papers each. However, despite producing the same number of publications, South Korea has a higher citation count overall (851 citations, averaging 70.92 per paper) than the United States (752 citations, averaging 62.67 per paper), which points to a stronger citation impact at the article level. India, by contrast, contributes 10 publications but receives comparatively few citations, with only 34 citations in total and an average of 3.40 citations per paper, indicating relatively low impact compared with its publication output.

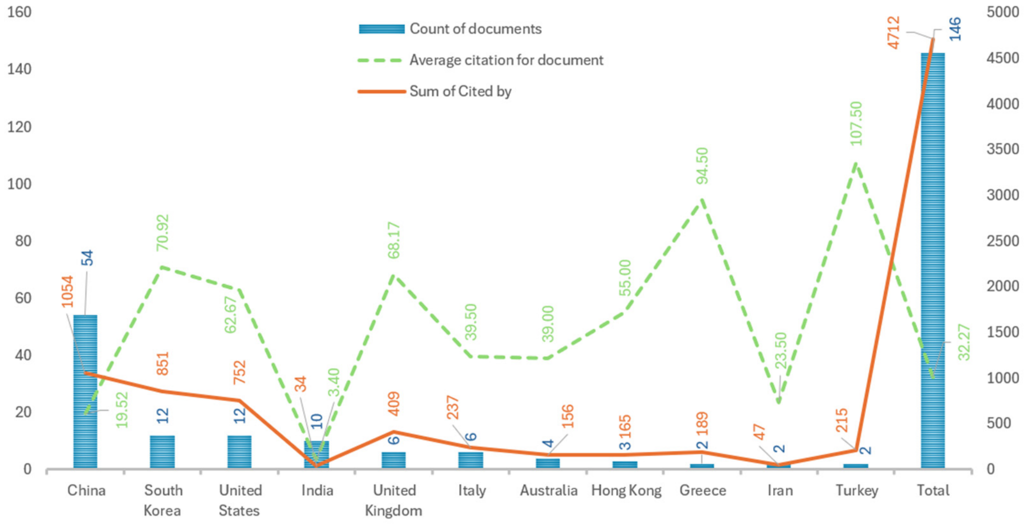

Figure 6. Country-level documents, total citations, and average citation impact.

Several low volume countries produce disproportionately high scientific impact. Notably, Greece (2 documents; 94.50 citations per article) and Turkey (2 documents; 107.50 per article) significantly exceed global citation norms, indicating exceptionally influential contributions despite limited output. Mid-range contributors show more nuanced patterns. The United Kingdom (6 documents; 68.17 citations per article) clearly outperforms the global average, while Italy and Australia, though publishing relatively few papers (6 and 4, respectively), maintain moderate citation influence (39.50 and 39.00 per paper). Hong Kong also demonstrates respectable impact (55 citations per article), whereas Iran (2 documents; 23.50 citations per article) performs noticeably below other low-output countries.

Taken together, the country-level patterns show a clear distinction between nations that contribute primarily through a high publication volume and those that exert influence through high-impact, selectively published work. While a few countries dominate output (e.g., China), the most influential contributions often originate from nations with smaller but more impactful research portfolios (e.g., South Korea, the United Kingdom, Greece, and Turkey). These contrasting profiles highlight the dual structure of the field (Table 4), where both productivity-driven and quality-driven pathways play central roles in shaping scientific visibility and the diffusion of global knowledge.

Table 4. Country categories based on publication output and citation impact.

Category	**Countries**	**Characteristics**
High output and moderate impact	China	Largest publication volume but only moderate citation influence.
Moderate output and high impact	United States and South Korea	Reasonable productivity with strong citation performance; influential contributions.
Moderate output and very low impact	India	Comparable output but very limited global visibility and low citation impact.
Low output and high impact	United Kingdom Greece, and Turkey	Few publications but exceptionally high impact; citation averages far above global benchmark.
Low output and moderate impact	Italy, Australia, and Hong Kong	
Low output and low impact	Iran	Minimal contribution and comparatively low citation influence.

3.4. Source-Level Publication and Citation Analysis

The analysis shows that research in this area is spread across journals, with clear differences in publication output and citation impact (Figure 7). Sensors and Automation in Construction appear as the most productive outlets, each publishing nine papers. However, their citation profiles are quite different: Automation in Construction records the highest influence in the dataset (770 citations; 85.56 citations per paper), whereas Sensors accumulates a more moderate 403 citations (44.78 per paper). Mid-range journals show a similar variation in citation impact. Construction and Building Materials publishes five papers but achieves one of the field's strongest citation rates (474 citations; 94.80 citations per paper); on the other hand, Buildings, despite releasing six papers, receives only 50 citations (8.33 per paper). The Journal of Structural Control and Health Monitoring and Engineering Structures further illustrate this pattern: although each publishes only three to five papers, they achieve high visibility, with Engineering Structures reaching 107 citations per article.

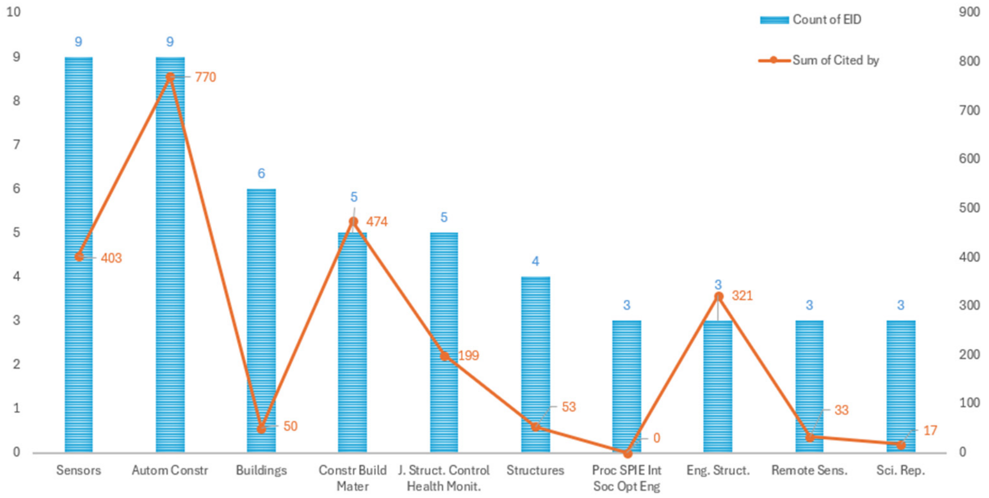

Figure 7. Leading publication sources ranked by document output and citation impact.

Lower-output sources such as Remote Sensing and Scientific Reports contribute modestly to the field, while the Proceedings of SPIE show no measurable citation uptake despite multiple publications.

Overall, these findings underscore that journal productivity does not directly predict scientific influence (Table 5). A small group of outlets, particularly Automation in Construction and Construction and Building Materials, serve as key hubs for high-impact dissemination, while others primarily support publication volume rather than citation visibility.

Table 5. Classification of journals by citation impact.

Source Category	Representative Journals	Interpretation
High-impact sources	Automation in Construction (85.56 cpp), Construction and Building Materials (94.80 cpp), and Engineering Structures (107 cpp)	Journals with strong visibility and influential publications; high citation-per-paper ratios regardless of output level.
Moderate-impact sources	Sensors (44.78 cpp), Journal of Structural Control and Health Monitoring (39.80 cpp), and Remote Sensing (33 cpp)	Outlets with solid, consistent impact; contribute meaningfully but not dominantly to the field.
Low-impact sources	Structures (13.25 cpp), Buildings (8.33 cpp), Scientific Reports (5.67 cpp), and Proceedings of SPIE (0 cpp)	Journals with limited citation influence; contribute mainly to publication volume rather than scholarly visibility.

3.5. Citation Dynamics of the Top 10 Most-Cited Publications (2015–2025)

Figure 8 illustrates the citation trajectories of the ten most influential papers, revealing two clear waves of high-impact contributions. The earliest papers, particularly the 2015 computer-vision study (487 citations) and the 2017 ML-based damage detection paper (339 citations), show steady, long-term citation growth. These works established the conceptual and methodological basis for later research and consistently attract 50–80 citations per year, maintaining dominance across the entire period.

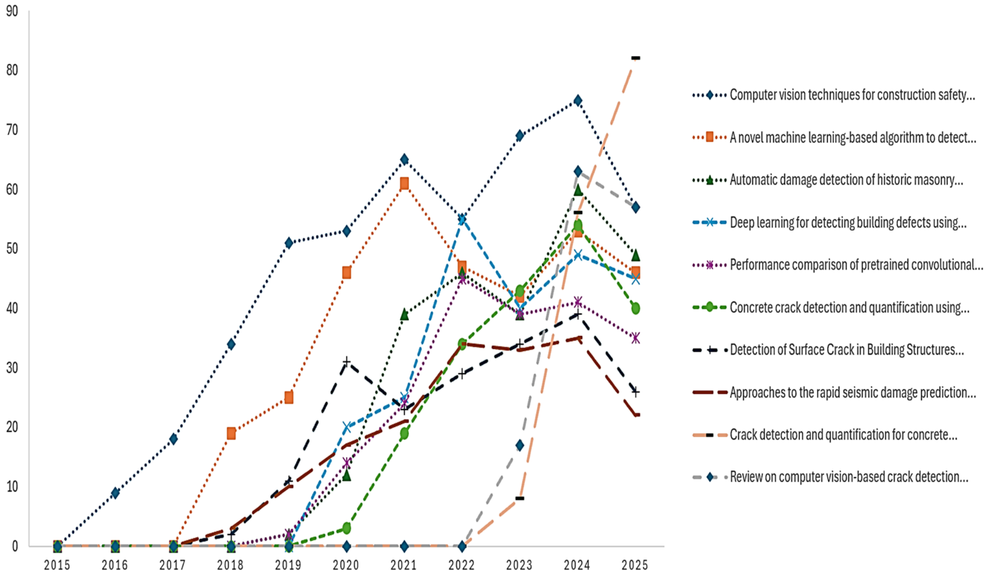

Figure 8. Citation trends of the top 10 most-cited papers (2015–2025).

Articles published between 2018 and 2020 display accelerated citation growth beginning around 2020, reflecting the shift toward CNNs, deep feature extraction, and UAV-based imaging. Their citation trajectories indicate strong methodological influence during the field's transition to deep learning. Several recent papers show exceptionally rapid uptake. For example, the 2023 transformer/UAV-based crack detection study has already reached 149 citations, including 82 citations in 2024 alone, making it one of the fastest-growing contributions. Similarly, the 2022 review has accumulated 143 citations in a short period. These patterns indicate immediate and substantial impact despite limited time in circulation.

Annual citation totals show that these ten papers collectively account for a large share of all citations in many years, particularly after 2022. Older works dominate citations before 2021, while more recent deep-learning and transformer-based studies increasingly shape citation activity thereafter. A citation pattern analysis shows that the field has moved from early ML- and CV-based studies toward deep learning, with recent work focusing more on practical applications. The growing attention given to newer papers suggests that this remains an active and fast-developing research area.

After examining basic bibliometric indicators such as publication trends, citation patterns, country contributions, and source impact, the analysis moves to science mapping using VOSviewer. Instead of focusing only on numerical summaries, these visualisations help show how different research elements are connected. They highlight collaboration networks, group related topics into clusters, and track how key concepts change over time. The following subsections present the results of this science mapping analysis and provide a clearer, network-based view of how authors, countries, keywords, documents, and sources relate to one another in the literature on AI-based crack detection.

3.6. Co-Authorship Analysis (Country Level)

We analysed country-level co-authorship to explore patterns of international research collaboration. To focus on countries with a meaningful level of contribution, we included only those with at least two publications, which also helped reduce visual noise from single-paper cases. We used fractional counting to distribute the publication credit more evenly among collaborating countries and to avoid overrepresenting papers with many international authors. Countries were weighted by their number of publications to reflect overall research participation. A network visualisation was then used to show collaboration links and clusters, as it provides a clear view of the underlying structure without splitting the map into too many small groups.

Figure 9 shows the global collaboration network, with China playing a central role. China produces the highest number of publications and collaborates closely with several major contributors, including the United States, the United Kingdom, South Korea, and Australia. The large node size and strong connections associated with China reflect both its high research output and its active involvement in international collaboration. The United States and South Korea also collaborate closely and appear as a secondary group within the network, highlighting their strong bilateral links and active contributions to this research area. The United Kingdom plays a connecting role by linking research collaborations between China and several European countries.

Countries such as Italy, Turkey, Ukraine, and Indonesia show fewer collaboration links and appear only weakly connected within the network. Although these countries contribute to the field, their research activity is often limited to a small number of publications or to collaborations concentrated within specific regions. A regional collaboration group can also be seen among India, Iran, and Saudi Arabia, with India acting as the main connection point. Within this group, collaboration between the countries is relatively strong, but their links to the main global research contributors are still limited.

Findings show that the network indicates that global research on AI-based crack detection is concentrated in a small group of countries, with China playing a central role in international collaboration. Other countries also contribute to the field, but their connections to the main collaboration network remain relatively limited.

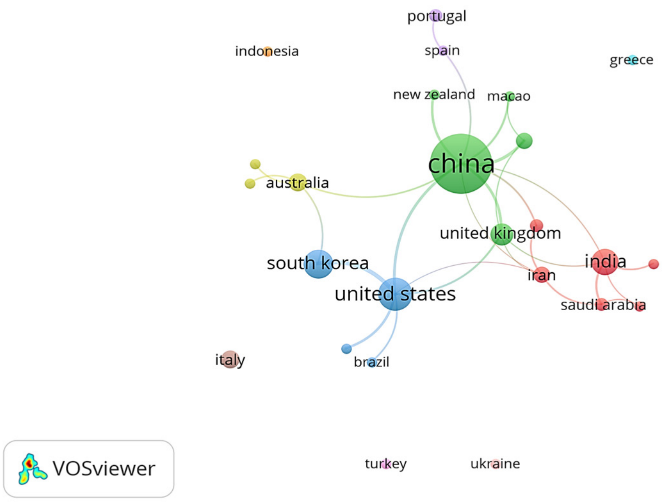

Figure 9. Country-level co-authorship network (minimum: 2 documents).

3.7. Co-Authorship Analysis (Author Level)

We also analysed co-authorship at the author level to better understand collaboration patterns within the research community and to identify influential researchers. Unlike the country-level analysis, this approach focuses on individual authors and shows who collaborates most frequently, as well as how research influence is distributed across the author network. To create a clear author-level collaboration map, we included only authors with at least two publications, which helped focus the analysis on more active contributors. Fractional counting was used again to distribute credit more evenly across multi-author papers, in line with the earlier analyses. We weighted authors using normalised citation counts to better reflect their research influence without bias from publication year. In addition, an overlay based on publication year was applied to distinguish earlier contributors from more recent authors. An overlay visualisation was chosen because it shows collaboration patterns while also indicating how author activity and influence change over time.

Figure 10 shows the overlay co-authorship network, which includes several collaboration clusters formed by both well-established and more recent researchers. Authors such as Wang J., Wang P., Li Y., and Ueda T. appear as the largest nodes in the network, reflecting their higher number of publications, frequent collaborations, and relatively high normalised citation counts. These features place them among the more influential contributors in this research area.

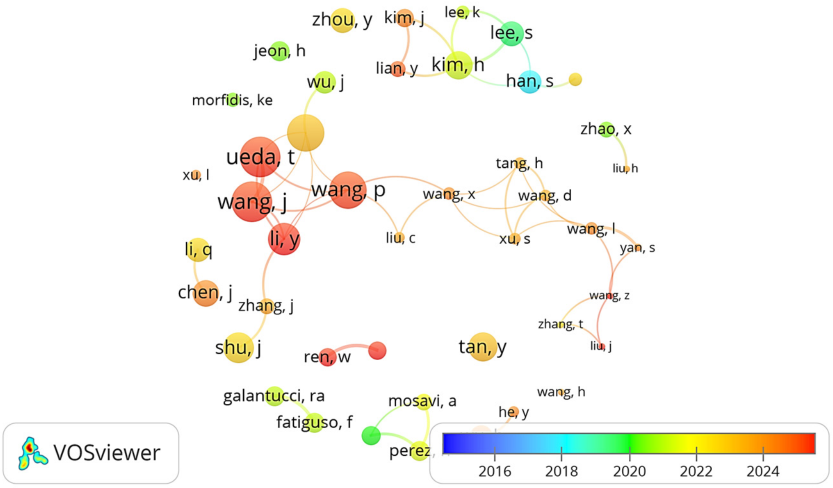

Figure 10. Author co-authorship overlay map.

Several structurally distinct collaboration clusters are visible. The largest cluster includes authors such as Wang J., Wang P., Li Y., Chen J., Xu S., and Tang H (predominantly based in China/East Asia), reflecting dense intra-group cooperation. A second major cluster contains authors like Lee S., Kim J., Kim H., and Han S. (largely Korean), characterised by strong internal connectivity and relatively recent publication activity, as indicated by their yellow-to-green colouring. Another smaller cluster includes authors such as Galantucci R.A., Fatiguso F., Perez A., and Mosavi A. (associated with European institutions), forming a more loosely connected collaboration group. These clusters reflect collaboration patterns that, while structurally defined by VOSviewer, tend to align with regional research communities (Table 6).

Table 6. Summary of author collaboration clusters.

Cluster	Representative Authors	Characteristics	Temporal Pattern (Overlay Colour)
Central High-Influence Cluster	Wang J., Wang P., Li Y., and Ueda T.	Highest connectivity; multiple co-authored papers; high normalised citation impact.	Orange–red (established and sustained influence)
Cluster A (China/East Asia–Structural Grouping)	Chen J., Xu S., Tang H., and Wang D.	Densely interconnected; strong internal collaboration; method-driven contributions.	Yellow–orange (recent but influential work)
Cluster B (Korea–Structural Grouping)	Lee S., Kim J., Kim H., and Han S.	Coherent regional subgroup; frequent intra-cluster collaboration; emerging influence.	Green–yellow (active mainly after 2020)
Cluster C (Europe–Structural Grouping)	Galantucci RA, Fatiguso F., Perez A., and Mosavi A.	Moderately connected; project-based collaboration; niche methodological contributions.	Green–yellow (mid-recent research activity)
Peripheral Authors	Zhou X., Liu H., Wang H., and Ren W.	Few collaborations; low normalised impact; peripheral position in the network.	Blue–green (older or less active contributions)

The overlay colouring also highlights the temporal evolution of influential contributors. Authors shaded in yellow and green (e.g., Kim H., Wu J., Han S., and Perez A.) represent those whose most impactful work has emerged largely after 2020, demonstrating the recent expansion of the field and the entry of new high-impact researchers. In contrast, orange–red nodes indicate earlier influential authors with sustained citation impact over time. Smaller, peripheral nodes, such as Zhou X., Liu H., and Wang H., reflect authors with limited collaborations or lower normalised citation impact. These authors contribute to the field but remain less integrated into its core research networks.

The overlay map depicts a decentralised and evolving research landscape, driven by several moderately connected regional clusters, a small set of highly influential central authors, and a growing influx of emerging contributors since 2020.

Based on the collaboration clusters summarised in Table 6 and the author network shown in Figure 10, the field appears to be organised around several regionally concentrated research groups, particularly in East Asia, Korea, and parts of Europe. A small number of highly connected authors link these clusters, suggesting that collaboration within these networks plays an important role in shaping the development and dissemination of research in AI-based crack detection for building structures.

3.8. Author Keyword Co-Occurrence Analysis (Overlay Visualisation)

This VOSviewer overlay map visualises how key research topics co-occur across publications, revealing the conceptual structure of the field and showing how research themes have evolved. Unlike frequency tables, the co-occurrence network highlights how ideas are interconnected and which concepts jointly shape the scientific discourse in AI-based crack detection.

A minimum threshold of three keyword occurrences was selected to exclude noise terms while preserving the main conceptual structure of the field. Author keywords were used because they directly reflect the authors' intended thematic focus, and full counting ensured that all co-occurrence relationships contributed meaningfully to theme formation. Occurrence was chosen as the weight, so node size reflects topic frequency, and the overlay (average publication year) was applied to visualise the temporal evolution of concepts. The chosen layout settings (attraction = 4; repulsion = 1) yielded a clear and interpretable structure with well-separated clusters and minimal label overlap. The co-occurrence overlay map, Figure 11**,** shows a highly interconnected thematic structure dominated by deep learning, which appears as the largest and most central keyword in the network. Its strong links to crack detection, computer vision, CNN, machine learning, and semantic segmentation highlight its role as the methodological core of AI-based crack detection research.

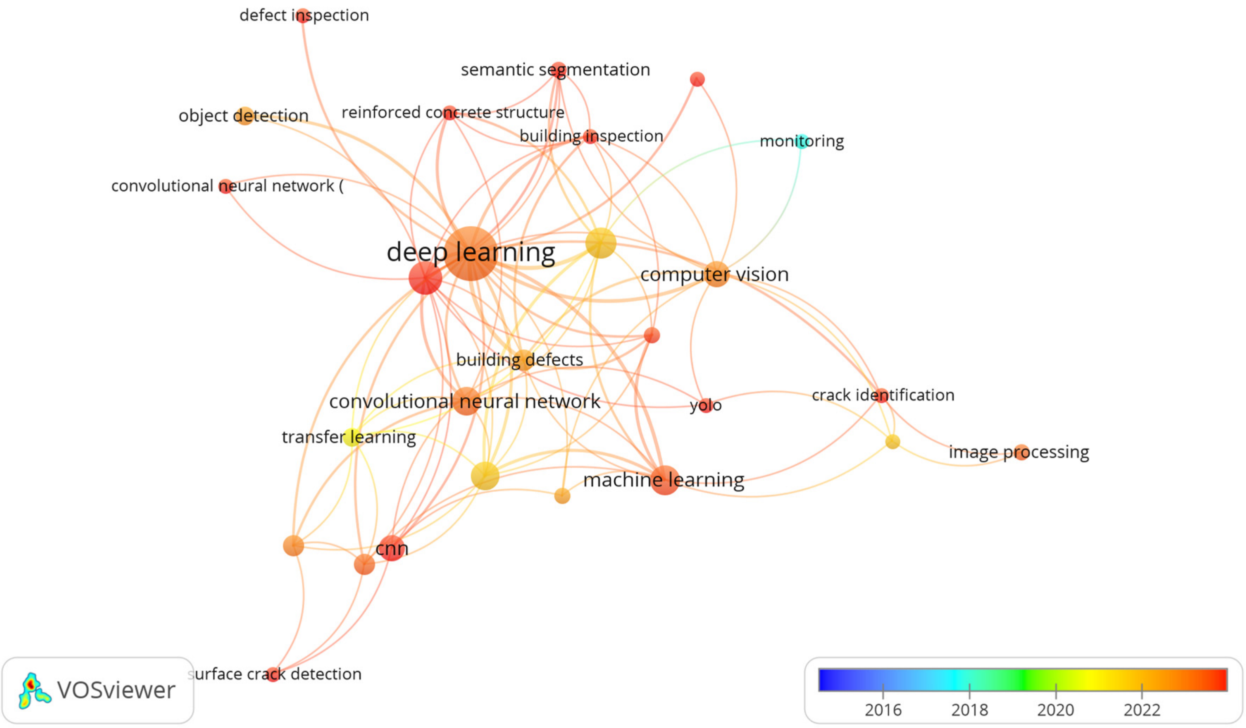

Figure 11. Overlay map of keyword co-occurrence showing core themes and temporal evolution.

Three major thematic clusters emerge:

(1) A deep-learning and crack-detection cluster, including CNN, YOLO, semantic segmentation, and object detection, representing the technical backbone of model development.

(2) A structural health monitoring cluster, linking terms such as structural health monitoring, monitoring, building defects, and concrete structures, capturing the application domain.

(3) A computer-vision and image-processing cluster, featuring machine learning, image processing, defect detection, and artificial intelligence, representing the broader computational tools supporting the field.

The overlay colours show a gradual change in research focus over time. Early studies from 2016 to 2019 mainly relied on classical computer-vision and monitoring methods. Around 2020, CNN-based approaches became more common, and in more recent work (2021–2024), deep-learning methods such as YOLO and real-time detection have appeared more frequently. Taken together, the map suggests that the field has become more consistent in its methods, with increasing emphasis on advanced deep-learning models and application-oriented detection systems.

3.9. Temporal Keyword Co-Occurrence Analysis (2015–2025)

A temporal keyword co-occurrence analysis was performed to trace how the conceptual focus of AI-based crack detection has evolved. This approach complements the overall keyword map by showing when specific themes emerged, matured, or declined. All three maps use identical VOSviewer settings (fractional counting, minimum of 2 occurrences, overlay visualisation, and attraction set to 4/repulsion set to 0) to ensure comparability.

3.9.1. 2015–2021: Early Formation Stage (Figure 12)

The early period shows a small but coherent thematic structure centred on deep learning, damage detection, CNN, SHM, and digital image processing. Colours indicate that most foundational concepts appeared between 2018 and 2020, marking the introduction of deep learning and the transition away from classical image-processing techniques. Connectivity is limited, reflecting the still-emerging nature of the field.

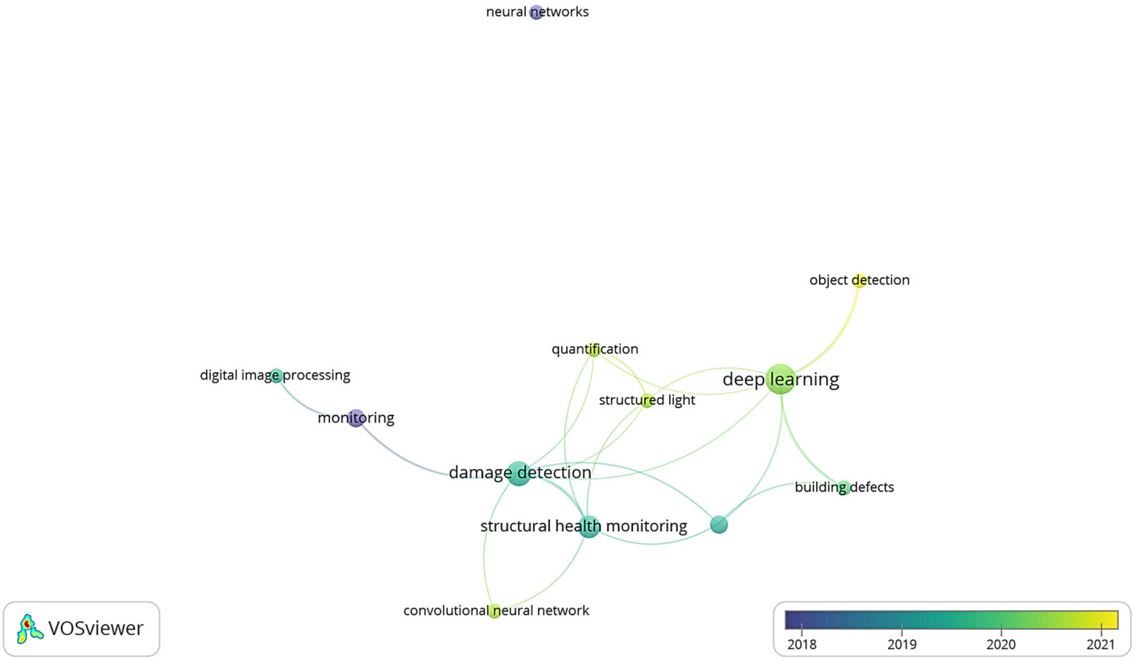

Figure 12. Keyword co-occurrence overlay for 2015–2021.

3.9.2. 2022–2023: Expansion and Consolidation (Figure 13)

This stage shows a clear expansion in both the number and diversity of keywords. New themes, such as transfer learning, UAV/drone inspection, CAM, and automatic damage detection, indicate a move toward operational and field-ready applications. Stronger link density reveals that research became more integrated, with 2021–2022 marking a rapid growth in segmentation-based and data-collection-driven methods.

3.9.3. 2024–2025: Frontier Innovation Stage (Figure 14)

The most recent period shows the emergence of cutting-edge topics, including YOLOv8, vision transformers, point cloud analysis, instance segmentation, GANs, and 3D reconstruction. These concepts cluster tightly around deep learning, indicating a shift toward high-performance, multimodal, and real-time detection systems. The yellow hues show that most of these innovations have appeared only very recently (2024–2025) (Table 7).

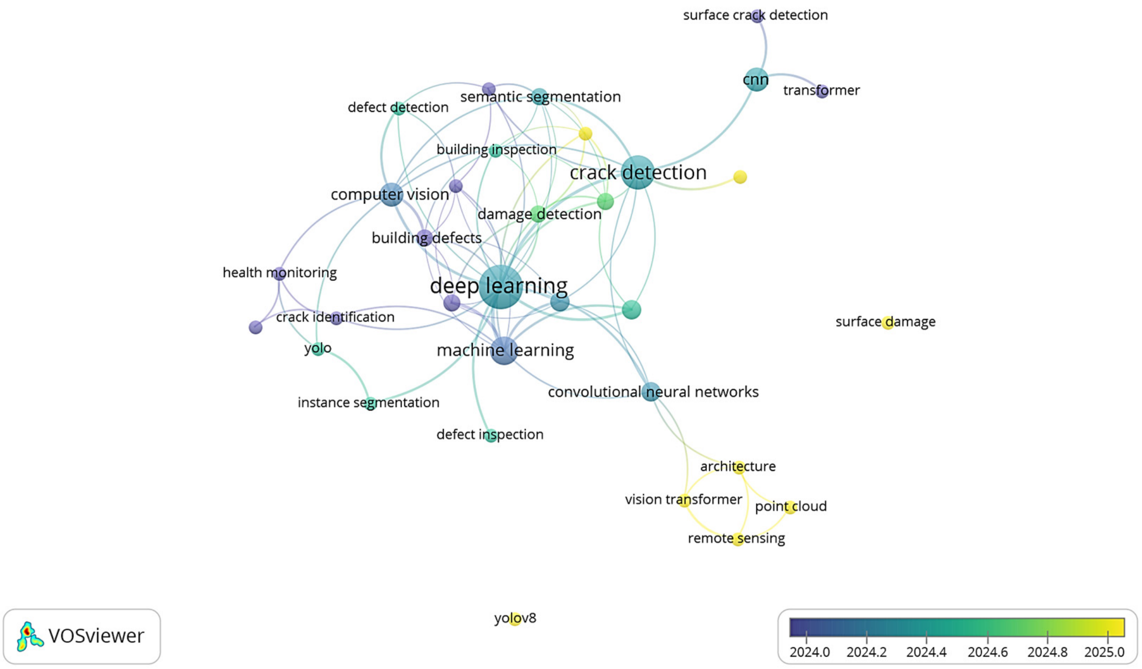

Figure 13. Keyword co-occurrence overlay for 2022–2023.

Table 7. Summary table—temporal evolution of research themes (2015–2025).

Period	Dominant Themes	Research Characteristics
2015–2021	Deep learning, CNN, SHM, and digital image processing	Early formation stage; emergence of DL-based crack detection; limited network connectivity; shift away from classical image processing.
2022–2023	Transfer learning, UAV inspection, segmentation, and automatic damage detection	Rapid expansion; increased methodological variety; stronger integration of detection and inspection workflows.
2024–2025	YOLOv8, transformers, point clouds, GANs, and instance segmentation	Frontier innovation; adoption of state-of-the-art architectures; movement toward real-time, multimodal, high-precision crack detection.

Together, the three maps reveal a clear research trajectory:

(1) Early emergence of CNN-based detection → (2) expansion into segmentation and UAV-based inspection → (3) adoption of state-of-the-art architectures, such as the YOLO series and transformers.

(2) This demonstrates the field's rapid evolution toward automated, multi-sensor, and high-precision crack detection.

3.10. Bibliographic Coupling Analysis (Document Level)

This visualisation identifies how influential documents in the field are intellectually connected based on shared references. Unlike co-authorship or keyword analyses, bibliographic coupling reveals the structural backbone of the literature, showing which papers build on similar foundations and therefore form conceptual research streams.

A 20-citation threshold filters out low-impact papers, ensuring the map represents only influential, field-shaping documents. Total Link Strength (TLS) highlights how strongly each document is connected to others via shared references, making conceptual clusters easy to detect. Selecting the largest connected cluster (23 docs) ensures the figure remains readable while capturing the dominant intellectual structure of the field. The network layout was chosen because it most clearly shows cluster boundaries and relational density.

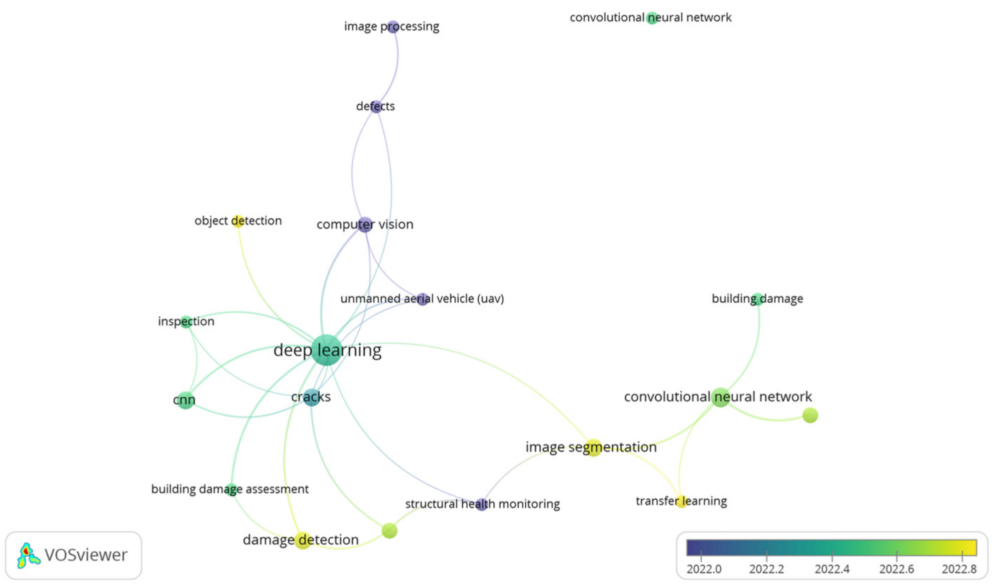

Figure 14. Keyword co-occurrence overlay for 2024–2025.

The bibliographic coupling map reveals four coherent clusters, each reflecting a distinct line of research grounded in shared reference structures (Figure 15). The thematic focus of each cluster was identified by examining the titles and author-provided keywords of the documents (not merely from the visual layout).

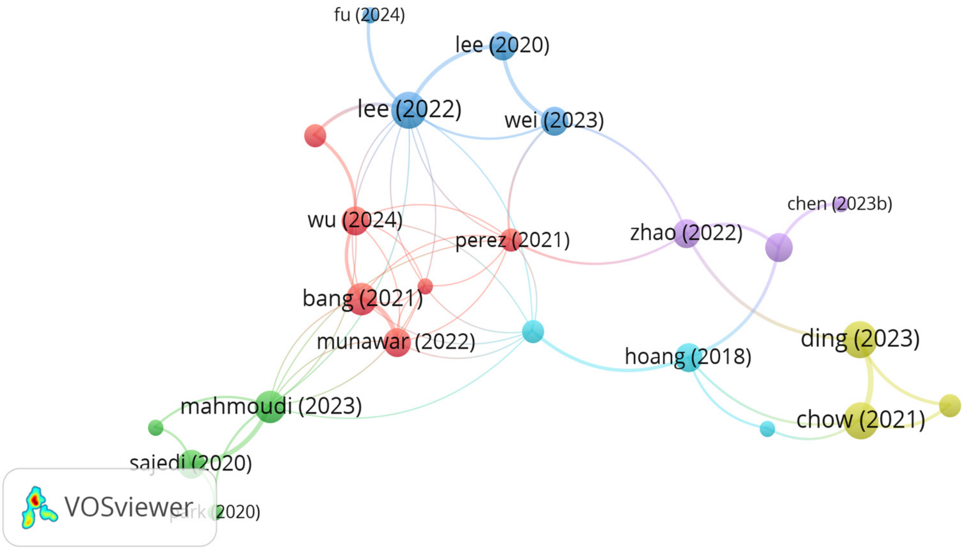

Figure 15. Bibliographic coupling network of highly cited documents (min. 20 citations).

(1) Core Conceptual Cluster including representative studies such as Lee (2020–2022), Wei (2023), and Fu (2024) (Lee 2020–2022; Wei 2023; Fu 2024): This is the densest and most interconnected group. Based on document titles/keywords, these studies centre on deep-learning architectures and methodological innovation. They form the methodological backbone of recent AI-driven crack detection research.

(2) Applied Structural Monitoring Cluster including representative studies such as Hoang (2018), Ding (2023), and Chow (2021): Titles and keywords show a focus on structural health monitoring, vibration-based assessments, and concrete damage evaluations. This cluster bridges earlier SHM research with more recent computer-vision-based inspection approaches.

(3) Emerging Deep-Learning Implementation Cluster including representative studies such as Wu (2024), Bang (2021), Munawar (2022), and Perez (2021): These papers share strong coupling and emphasise practical DL implementation, including real-time defect detection, CNN pipelines, and accelerated processing. Their thematic alignment is confirmed through their reported methods and keywords.

(4) Peripheral But Growing Cluster including representative studies such as Mahmoudi (2023), Sajedi (2020), and Park (2020): Although less connected to the central cluster, titles/keywords indicate a consistent focus on building inspection, hybrid sensing, thermal imaging, and nondestructive testing. This reflects the field's expansion toward multimodal inspection technologies.

These clusters collectively demonstrate that the field is structured around four dominant methodological and application-driven research streams.

3.11. Source Co-Citation Analysis (Largest Connected Cluster)

A source-level co-citation analysis was conducted to reveal the intellectual structure of the research field and identify the journals that form its conceptual foundation. In co-citation mapping, two sources are considered intellectually related when they are frequently cited together in the literature; therefore, this technique provides insights into the underlying knowledge base that researchers draw upon. To focus the analysis on meaningful scholarly relationships, a minimum threshold of two citations per source was applied. Higher thresholds reduced the network to only a handful of journals, while lower thresholds introduced noise from marginal or weakly referenced outlets. After thresholding, only the largest connected component, consisting of twenty journals, was visualised. This approach is standard in science mapping, as it captures the core structure of the domain without diluting interpretability. The network was generated using full counting, Total Link Strength (TLS) as the weighting measure, and a default Association Strength normalisation layout with attraction set to 4 and repulsion set to 0, producing a clear and analytically coherent map.

The co-citation network of sources reveals how journals collectively shape the intellectual foundations of AI-based crack detection and structural health monitoring. Three major clusters appear in Figure 16. The first cluster, dominated by Construction and Building Materials, Structures, Science, and Computer Methods in Applied Mechanics and Engineering, represents the engineering and materials-science backbone of the field, providing theoretical frameworks in applied mechanics, structural modelling, and construction materials. The second cluster centres on Automation in Construction and includes technology-oriented journals and conference outlets; this group reflects the digital construction and computer-vision lineage of the research, connecting work on automated inspection, sensing technologies, and robotics. The third cluster, anchored by Structural Health Monitoring and Mechanical Systems and Signal Processing, captures the diagnostic and vibration-based monitoring tradition, highlighting the influence of SHM and signal-processing approaches. The spatial separation but visible linking paths among clusters indicate that the field is strongly interdisciplinary: classical engineering science, automation technologies, and SHM techniques jointly underpin the development of modern deep-learning-based crack detection systems. Overall, the map demonstrates that the literature draws from both traditional civil/mechanical engineering sources and emerging digital-construction journals, creating a tightly integrated conceptual ecosystem. A summary of key co-citation clusters and their thematic role is shown in Table 8.

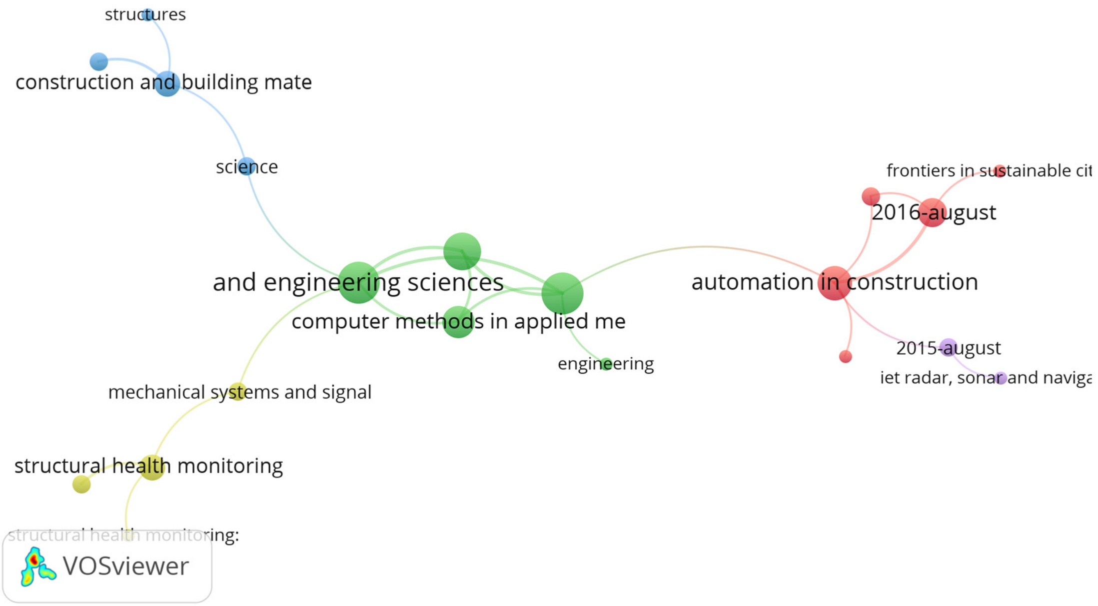

Figure 16. Co-citation network of sources showing major foundational journal clusters.

Table 8. Summary of key co-citation clusters and their thematic role.

Cluster	Representative Journals	Role in the Field
Engineering & Materials Cluster	Construction and Building Materials, Structures, Science, and Computer Methods in Applied Mechanics and Engineering (CMAME).	Provides foundational mechanics, materials science, structural modelling, and engineering theory.
Automation & Digital Construction Cluster	Automation in Construction, Institution of Engineering and Technology (IET) Radar/Sonar, and Frontiers in Sustainable Cities	Supplies computer vision, sensing, UAV, and robotics perspectives for automated inspection.
SHM & Signal Processing Cluster	Structural Health Monitoring, and Mechanical Systems and Signal Processing	Contributes diagnostic, vibration-based, and monitoring methodologies.

4. Systematic Technical Review Results

Recent systematic reviews have explored advances in computer-vision-based crack detection and deep-learning techniques for structural health monitoring and infrastructure assessments [3,82]. These studies highlight the evolution of AI-based methods for crack identification, segmentation, and quantification across various civil engineering applications. However, they do not offer a detailed, building-focused technical synthesis of methodological practices in recent AI-based crack detection studies.

Building on this foundation, the present systematic technical review examines methodological trends across 36 high-relevance studies to address the five Technical Research Questions (TRQ1–TRQ5) outlined in the Introduction.

The data items extracted from each study were structured according to the schema summarised in Table 9.

4.1. Overview of the Selected Studies

4.1.1. Publication Year Distribution

The 36 core studies selected for the systematic technical review are strongly concentrated in the most recent years, reflecting the rapid development of AI-based crack detection methods for buildings. As shown in Figure 17, the majority of studies were published between 2023 and 2025, with 16 papers (44.4%) appearing in 2024 and 9 papers (25.0%) in 2025. Earlier contributions are relatively sparse, with only 1 study published in 2020, 3 in 2022, and 7 in 2023. This temporal pattern indicates that methodological research on

AI-based crack detection in buildings is an emerging and fast-growing topic, and that most technical practices analysed in this review correspond to very recent work.

Table 9. Data extraction schema and its alignment with TRQ1–TRQ5.

Field	Description	TRQ
Study ID/Reference	Identifier or citation of each included study	—
Application Domain & Problem Formulation	Crack detection, segmentation, defect characterisation, severity estimation, and predictive diagnostics	TRQ1
AI/ML/DL Model Family	Main model category used (CNNs, YOLO variants, U-Net, transformers, and hybrid models)	TRQ2
Dataset Characteristics	Dataset modality (image/video), dataset size, annotation type (bbox/mask/severity labels), and resolution	TRQ3
Evaluation Metrics	mean Average Precision (mAP), Intersection over Union (IoU), precision, recall, F1-score, dice, etc.	TRQ4
Training & Validation Protocol	Train/validation/test split, cross-validation strategy, and evaluation scenarios	TRQ4
Hyperparameter Settings & Optimisation Strategy	Learning rate, batch size, optimiser, and epochs; presence/type of Hyperparameter Optimization (HPO) (grid, random, Bayesian, and Optuna)	TRQ5

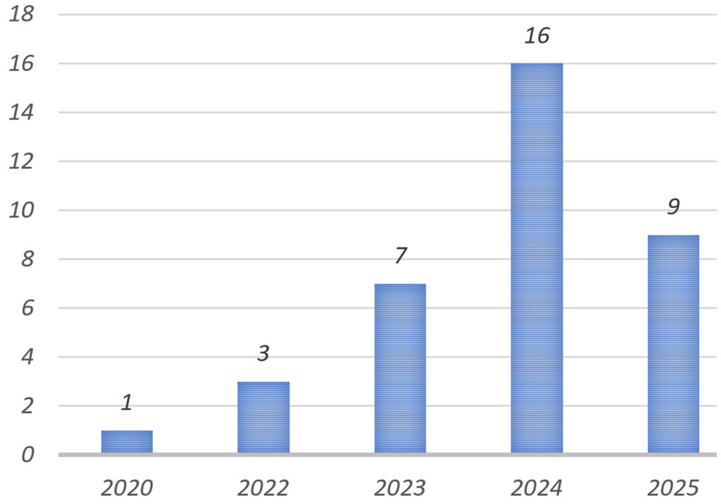

Figure 17. Annual distribution of the 36 studies (2020–2025).

4.1.2. Application Domain Distribution

The application focus of the reviewed studies is predominantly building-oriented. As summarised in Table 10, 21 out of 36 papers (58.3%) explicitly target generic building scenarios, while an additional 5 studies focus on “concrete buildings” and 5 on “building façades”. Only a small number of works address more specialised contexts, such as concrete structural components (2 studies), apartment buildings (1 study), ceramic tile façades (1 study), or dedicated building defect datasets (1 study). Overall, this distribution confirms that the core set of studies is strongly aligned with the building-centric scope of this review, rather than broader infrastructure domains, such as bridges, pavements, or tunnels.

Table 10. Application domain distribution.

Application Domain	Building	Concrete Buildings	Building Façade	Concrete Structure	Building Defect Dataset	Apartment Buildings	Ceramic Tile Building Façade
No.	21	5	5	2	1	1	1

4.1.3. Citation Influence of the Selected Studies

The citation profile of the 36 core studies is highly skewed, with a small group of influential papers and a larger number of recent works that have not yet accumulated substantial citations. As illustrated in Figure 18, a few studies have already attracted significant attention, including D15 (148 citations), D29 (95), D24 (77), D31 (46), and D25 (35), indicating that certain methodological contributions have become reference points in the field. In contrast, a considerable proportion of the selected papers, particularly those published in 2023–2025, currently have few or no citations, which is expected given their recency. This observation suggests that the technical review mainly reflects recent methodological developments, while their longer-term impact has not yet become fully clear.

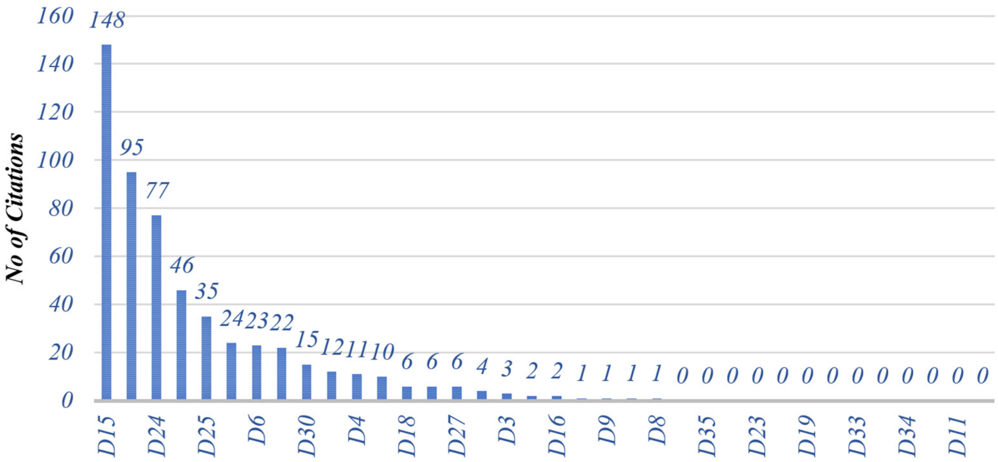

Figure 18. Citation metrics for selected studies.

4.2. Task Formulations (TRQ1)

In this subsection, we examine how the 36 studies describe crack- and defect-related tasks, as this provides a useful context for the methodological trends discussed later. Task types were identified and grouped into broader themes based on their main focus.

4.2.1. Distribution of Task Formulations

Task types identified in the reviewed studies are summarised in Table 11 and visualised in Figure 19. Binary classification, object detection, and semantic segmentation dominate the task landscape (13, 14, and 10 studies), showing that current research is primarily focused on recognising and localising cracks. SHM-oriented workflows (Task 10) are equally prevalent (14 studies), indicating a strong interest in integrating detection models into practical inspection pipelines. More specialised tasks, such as quantification, severity estimation, type classification, and defect segmentation, are far less common, occurring in only 1–5 studies. Multi-defect tasks (Tasks 7–8) appear in six studies each, showing an emerging trend toward broader defect coverage. High-level analytical tasks (predictive prognosis, Large Language Model (LLM)-based reporting, and maintenance decision support) are almost absent, highlighting a major research gap. Overall, the field remains detection-centric, with significant opportunities for expanded diagnostic and predictive modelling.

Structural health monitoring (SHM) is defined in this study as a system-level application objective rather than a stand-alone computer-vision task.

In the reviewed literature, SHM-oriented studies do not introduce new perception-level outputs but instead integrate the outputs of one or more lower-level tasks (e.g., crack detection, segmentation, or quantification) into inspection workflows, condition assessment processes, or maintenance-oriented decision contexts.

Table 11. Task distribution.

Task1	Binary Crack Classification
Task 2	Crack Object/Instance Detection (Bounding Box)
Task 3	Crack Semantic Segmentation (Pixel-wise)
Task 4	Crack Quantification (Width/Length/Density Measurement)
Task 5	Crack Severity Estimation
Task 6	Crack Type Classification (Longitudinal, Transverse, Shear, etc.)
Task 7	Non-crack Defect Classification
Task 8	Multi-defect Detection (Crack + Other Surface Defects)
Task 9	Non-crack Defect Segmentation
Task 10	Visual Inspection & Structural Health Monitoring (SHM)
Task 11	Predictive Damage Assessment/Prognosis
Task 12	Image Captioning for Crack/Defect Description
Task 13	LLM-based Automated Structural Report Generation
Task 14	Maintenance & Repair Decision Support

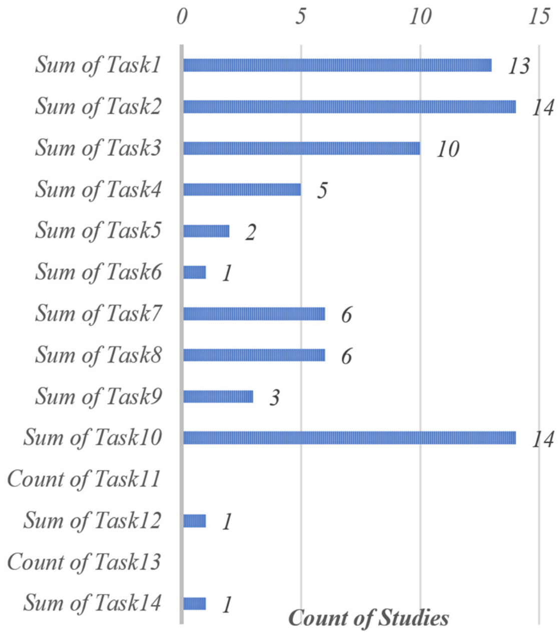

Figure 19. Task frequency.

Accordingly, SHM is treated as a distinct application-layer task, characterised by its functional role and operational intent, rather than by a unique computational formulation or output representation.

4.2.2. Co-Occurrence of Tasks Within Studies

The 36 reviewed studies exhibit a diverse distribution of task combinations, indicating that AI-based crack assessments in buildings are increasingly moving from single-purpose models toward multi-functional inspection pipelines. Most papers (≈60%) address two tasks, commonly pairing crack classification (Task 1) or object detection (Task 2) with SHM-oriented visual assessment (Task 10). More advanced combinations, such as detection + segmentation + SHM, appear in studies D15, D22, D24, D27, D29, D33, and D34, reflecting a growing trend toward integrated workflows.

Only four studies (D23, D32, D35, and D36) address four different tasks, making them the most comprehensive multi-step approaches in the dataset. By comparison, eleven studies focus on a single task, most often binary crack classification (Task 1) or SHM-level assessment (Task 10). This suggests that many existing studies still concentrate on individual capabilities rather than covering a complete inspection workflow. Looking at how tasks appear together across studies, some clear groupings can be observed. Crack segmentation (Task 3) is most often combined with SHM-level assessment (Task 10) and crack quantification (Task 4), which suggests that segmentation is commonly used as a starting point for more detailed façade condition analysis. Similarly, multi-defect detection (Task 8) frequently appears alongside non-crack defect classification (Task 7), showing a tendency to move toward broader façade analysis rather than focusing on cracks alone.

In general, the results suggest that research is slowly moving beyond single-task models toward more integrated solutions. However, this development is not uniform, and many studies still stop short of supporting full, real-world building inspection workflows.

The full task co-occurrence matrix is presented in Table 12, summarising how the 36 studies are distributed across Tasks 1–14 and the total number of tasks addressed in each study.

Table 12. Task co-occurrence matrix. Note: * indicates that the corresponding task is addressed in the study.

	Task1	**Task2**	**Task3**	**Task4**	**Task5**	**Task6**	**Task7**	**Task8**	**Task9**	**Task10**	**Task11**	**Task12**	**Task13**	**Task14**	**Total Tasks**
D1	*														1
D2												*			1
D3			*						*						2
D4	*	*													2
D5		*					*								2
D6							*								1
D7	*						*								2
D8		*				*									2
D9	*						*								2
D10				*				*							2
D11								*							1
D12	*	*													2
D13							*			*					2
D14			*												1
D15			*	*						*					3
D16		*													1
D17	*			*											2
D18					*										1
D19									*	*					2
D20	*									*					2
D21								*							1
D22		*	*							*					3
D23		*						*		*				*	4
D24	*	*	*							*					3
D25									*						1
D26		*								*					2
D27		*	*							*					3
D28	*				*										2
D29		*	*							*					3
D30	*														1
D31								*							1
D32	*						*	*		*					4
D33	*	*	*												3
D34			*	*						*					3
D35		*	*	*						*					4
D36	*									*					2
Total	13	13	10	5	2	1	6	6	3	14	0	1	0	1	

4.2.3. Interpretation of Task Trends

Looking across the 36 studies, a few common patterns can be seen in task formulation. Many studies focus on visual inspection and SHM-level tasks (Task 10), rather than only detecting cracks, which points to a growing interest in more integrated building assessment approaches [25]. Object detection (Task 2) and semantic segmentation (Task 3) also appear often, which shows that locating cracks in space is still a central focus in many studies, especially those related to UAV-based or façade-level inspections [25]. Findings show more advanced tasks, such as predictive diagnostics (Task 11), LLM-based report generation (Task 13), and maintenance decision support (Task 14), are still rarely addressed. This suggests that most studies have not yet moved toward higher-level reasoning or fully automated decision-making [83]. A similar lack of attention can be seen for severity estimation (Task 5) and crack type classification (Task 6). Most existing models still focus mainly on detecting cracks, rather than interpreting their structural significance [25].

A small number of studies that cover multiple tasks (for example, D23, D32, and D35) point to a growing interest in more flexible deep-learning approaches. These studies aim to handle tasks such as detection, segmentation, and SHM-related analysis within a single workflow, rather than treating them separately [84]. In general, most studies still rely on basic tasks like binary classification and object detection. However, there are early signs that research is starting to move toward more integrated, multi-task problems, including SHM-related tasks and some initial work on vision–language and automated applications [83].

4.3. Model Families Used in Recent Studies (TRQ2)

This section looks at the types of model families used across the selected studies and how their use has changed over time. The analysis focuses on which modelling approaches are most common and how different model families tend to be used for specific crack-related tasks.

4.3.1. Distribution of AI Model Families

The model-family classification adopted in this review is summarised in Appendix D. Across the selected studies, CNN-based architectures are used far more often than other model families (Table 13; Figure 20). They remain the main choice in AI-based crack detection for buildings, largely because they work well for visual feature extraction and can be trained or fine-tuned using existing pre-trained models, even when datasets are relatively small or diverse. Detection-oriented frameworks from the YOLO and Region-based Convolutional Neural Networks(R-CNN) families form the next most common group. Their use reflects a growing interest in object-level crack localisation and inspection scenarios where faster or near real-time performance is required. By comparison, transformer-based models and CNN–transformer hybrids appear much less frequently. This suggests that their use in building-focused crack detection is still at an early stage, possibly due to higher computational demands and a stronger dependence on large training datasets.

Table 13. Distribution of model families across the reviewed studies.

		Studies
Model Family	F1	D1, D4, D9, D11, D12, D13, D17, D18, D19, D24, D28, D29, D30, D31, D32, D33, D34, D36
	F2	D1, D15, D21
	F3	D7, D27
	F4	D6, D8, D10, D16, D22, D23, D26
	F5	D20, D25, D35
	F6	D2

Classical or shallow machine-learning approaches are rarely reported, likely due to their limited capacity to model the complex, irregular patterns of cracks compared to deep-feature-learning methods. Vision–language models are almost absent from the reviewed studies. This limited presence may reflect the relative novelty of such models in the structural inspection domain, the methodological complexity involved in integrating visual and semantic representations, and the fact that their potential applications for building crack analysis are still largely unexplored. Overall, these findings suggest a research landscape that is still grounded in established deep-learning paradigms, while gradually beginning to investigate more advanced and conceptually complex modelling approaches.

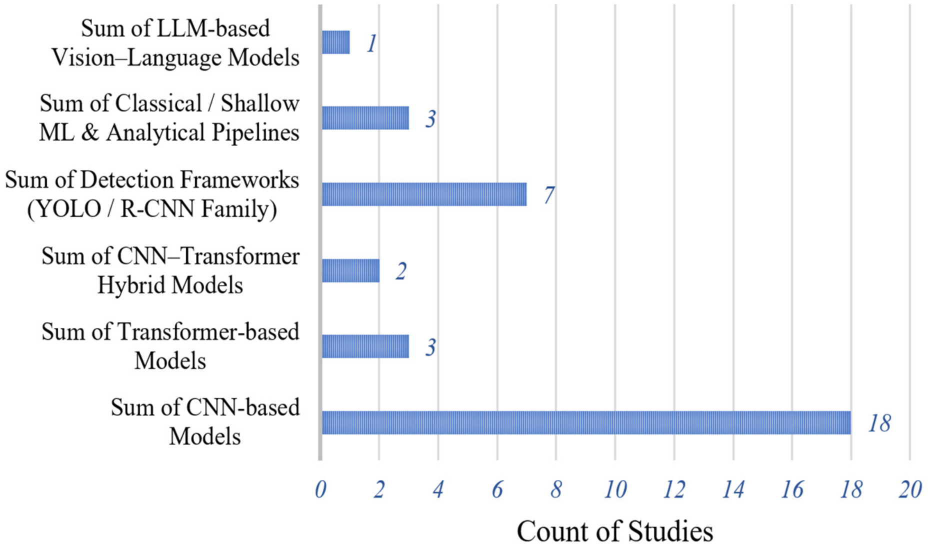

Figure 20. Frequency of model families in the selected studies.

4.3.2. Temporal Evolution of Model Families (2020–2025)

The temporal evolution of model families adopted in the selected studies was analysed to identify recent methodological shifts in AI-based crack detection research. Figure 21 illustrates the temporal evolution of model families used in the selected studies between 2020 and 2025. Due to the selection strategy of this review, which prioritises recent high-relevance publications, only a limited number of studies are represented before 2023. Consequently, early years mainly reflect isolated CNN-based approaches rather than a comprehensive methodological trend.

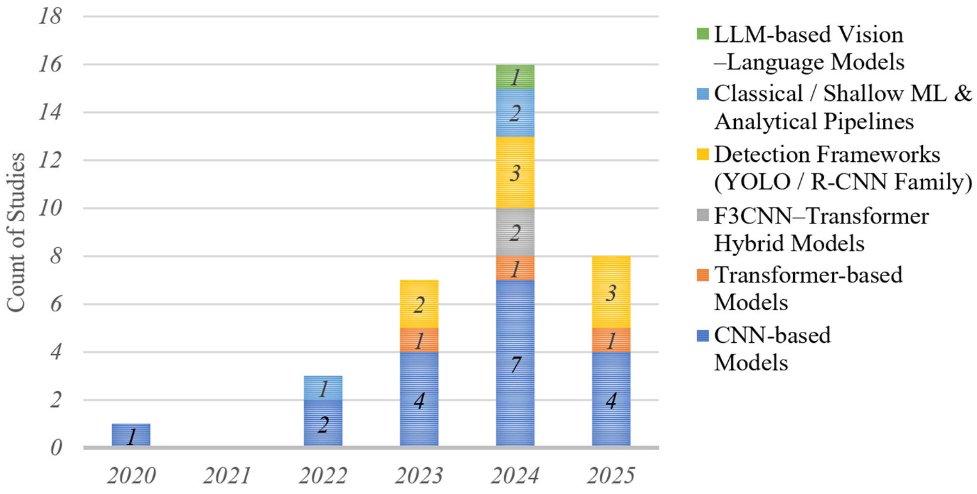

Figure 21. Temporal evolution of model families (2020–2025).

From 2023 onward, a clear diversification of model families is observed. Transformer-based and CNN–transformer hybrid models begin to appear alongside conventional CNN architectures, indicating an increasing interest in capturing both local and global visual features. The year 2024 marks the highest level of methodological diversity, with all major model families being actively explored. In addition, YOLO-based detection frameworks show stronger representation in 2024 and 2025, suggesting a growing emphasis on real-time and application-oriented crack detection solutions.

4.3.3. Alignment Between Tasks and Model Families

This subsection examines how different model families are aligned with specific task formulations in the selected studies. Figure 22 presents the distribution of model families across the defined task categories, revealing clear preferences between modelling approaches and problem types.

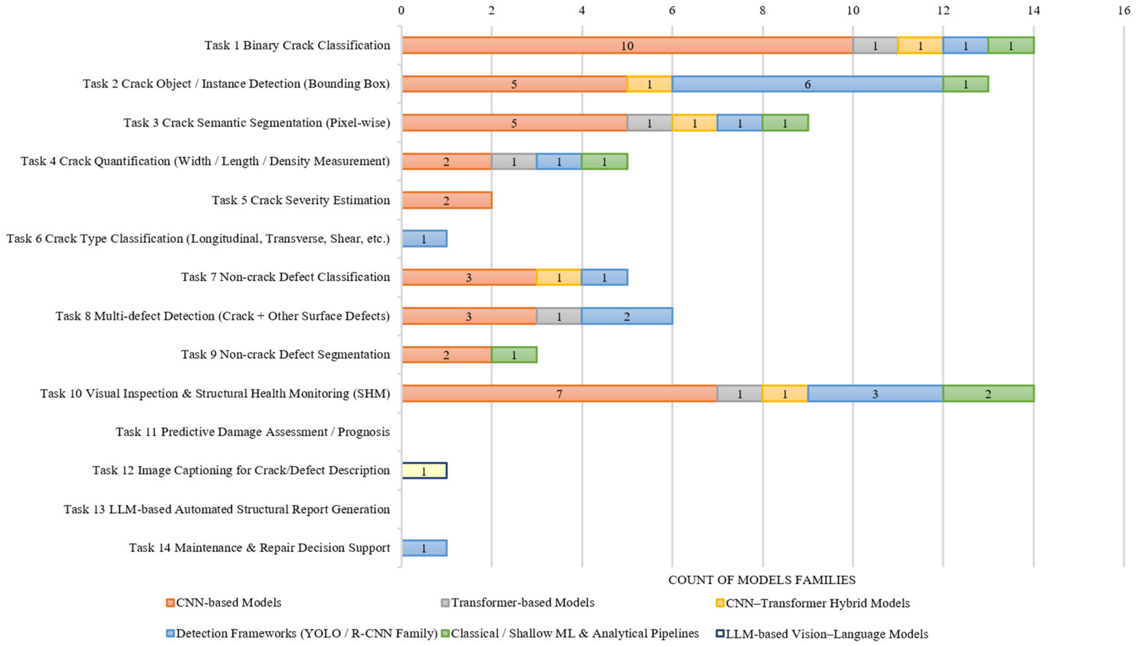

Figure 22. Task–model family alignment across the selected studies.

The following observations are based on patterns identified in the analysed literature dataset included in this review.

CNN-based models demonstrate the broadest applicability, being employed across nearly all task categories, including classification, detection, segmentation, and structural health monitoring. This confirms their role as general-purpose backbones for building-related crack analysis. Detection frameworks from the YOLO/R-CNN family show a strong concentration in object-level tasks, particularly crack instance detection and multi-defect detection, reflecting their design focus on localisation and real-time performance.

Transformer-based and CNN–transformer hybrid models are more selectively adopted, appearing primarily in semantic segmentation and SHM-oriented tasks where global contextual reasoning is beneficial. Classical or shallow machine-learning pipelines are limited to a small number of low-level or rule-driven tasks. Finally, vision–language models appear only in image-captioning applications, indicating an early-stage exploration rather than a mature methodological trend.

4.3.4. Interpretation of and Insights into Model Selection Trends

The continued dominance of CNN-based models in building crack detection can be attributed to their architectural inductive biases, such as locality and translation equivariance, which make them relatively data-efficient and robust when trained on limited or heterogeneous datasets [85]. These properties align well with typical building inspection scenarios, where large-scale, uniformly annotated datasets are often unavailable.

In contrast, transformer-based vision models generally require substantially larger datasets to compensate for the lack of strong inductive biases and to achieve stable generalisation performance [86]. This partially explains their more selective adoption in the reviewed studies, where transformers are mainly applied to tasks such as semantic segmentation or high-level structural health monitoring rather than basic classification.

YOLO-based detection frameworks have gained increasing popularity in recent years, particularly for crack and multi-defect detection tasks. Their appeal lies in their real-time inference capability, end-to-end training pipeline, and strong performance in detecting small objects, features that are especially relevant for on-site building inspections and UAV-based imaging applications [87,88].

Hybrid CNN–transformer architectures represent an emerging trend, aiming to combine the local feature extraction strength of CNNs with the global contextual modelling ability of transformers. These models are still relatively limited in number but show growing potential, particularly in segmentation-oriented tasks. Finally, vision–language and LLM-based models appear only marginally in the current literature, reflecting their novelty and the unresolved challenges associated with multimodal data requirements and task formulation in building-focused crack detection.

Taken together, these model adoption patterns indicate that architectural innovation in AI-based crack detection has progressed faster than the development of task-specific reasoning and decision-oriented modelling capabilities.

4.4. Dataset Characteristics (TRQ3)

An overview of dataset modalities and annotation practices across the selected studies is presented in this section.

4.4.1. Data Modality Distribution

As illustrated in Figure 23, the datasets used across the reviewed studies are dominated by ground-based Red, Green and Blue (RGB) imagery. UAV-based RGB data and combined ground–UAV datasets form a secondary group, while thermal, depth/3D, synthetic, and multimodal inputs are only sparsely represented.

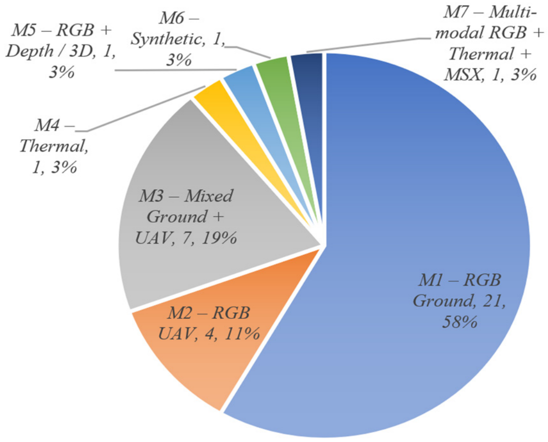

Figure 23. Dataset modality distribution.

This distribution reflects a strong reliance on conventional visual sensing approaches, with more complex or specialised modalities remaining relatively underexplored within current building-focused crack detection studies.

4.4.2. Annotation Strategy Distribution

The annotation strategies adopted in the reviewed literature are summarised in Figure 24. Image-level labels and bounding box annotations account for the largest share, whereas pixel-level semantic masks, instance segmentation, and geometry-based annotations appear far less frequently. Automatic or annotation-free approaches are observed only in isolated cases.

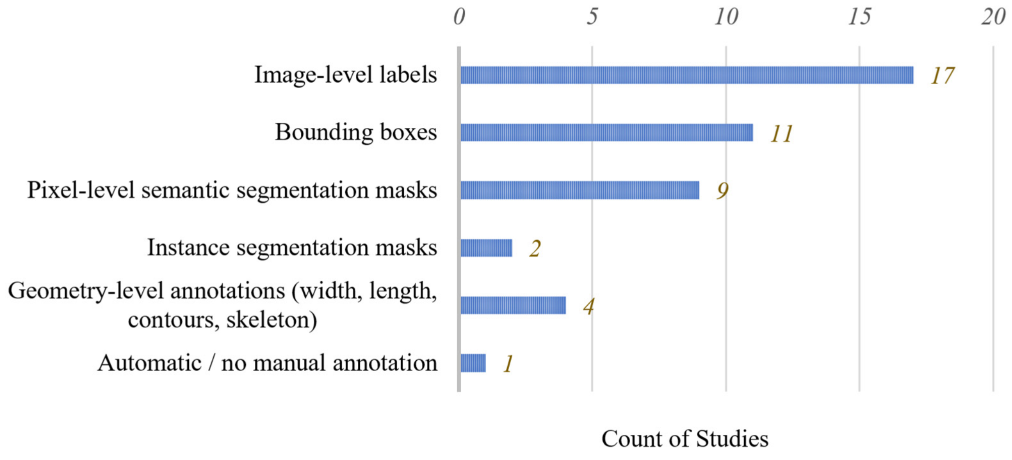

Figure 24. Annotation type distribution.

Overall, the prevalence of coarse annotation schemes suggests a preference for methods that balance detection capability with manageable annotation effort, while fine-grained annotations are typically reserved for segmentation-oriented investigations, where pixel-level accuracy is essential despite the higher labelling cost [89,90].

4.4.3. Interpretation of Dataset Trends

The observed dominance of RGB-based datasets is supported by recent comprehensive reviews of the crack detection literature, which highlight that most studies rely on standard image sources due to their accessibility and established use in training deep-learning models [91]. Moreover, surveys of crack detection technologies note that the availability and utilisation of diverse public datasets remain limited, and that researchers often draw from a combination of proprietary and benchmark collections rather than specialised modalities [1]. Thermal, depth, synthetic, and multimodal datasets appear only in a small number of studies. This is likely related to the difficulty of collecting such data and the fact that multimodal approaches are still relatively new in automated crack detection. As a result, there remains clear room for expanding available datasets and for exploring more detailed annotation strategies in future work [92]. These dataset characteristics show that most existing benchmarks are mainly designed for detection-related tasks. In contrast, they offer much less support for more detailed structural interpretations, temporal analyses, or predictive assessments.

4.5. Results for Evaluation Protocols and Performance Metrics (TRQ4)

Beyond model architectures and datasets, the reliability of reported results strongly depends on how models are evaluated, as evaluation protocols and dataset splits can significantly influence the perceived performance and generalisability of learning-based vision systems [93], with the metric families and validation strategy categories used in this review summarised in Appendices E and F, respectively. This section reviews the evaluation protocols and performance metrics adopted in the selected studies, highlighting prevailing practices, methodological gaps, and implications for comparability of results across the literature.

4.5.1. Evaluation Metrics and Validation Strategies

Stacked Bar Chart Analysis (Metric Families vs. Validation Strategies):

As illustrated in Figure 25, evaluation practices in AI-based crack and building-defect analyses are heavily concentrated around a small set of metric families and validation strategies. Classification (M1), detection (M2), and segmentation metrics (M3) dominate performance reporting, while other metric families, such as severity estimation, quantification, and captioning, appear far less frequently. Across all metric families, fixed hold-out validation (V1) is the most commonly adopted strategy, indicating a strong preference for simple train–test splits over more robust evaluation designs.

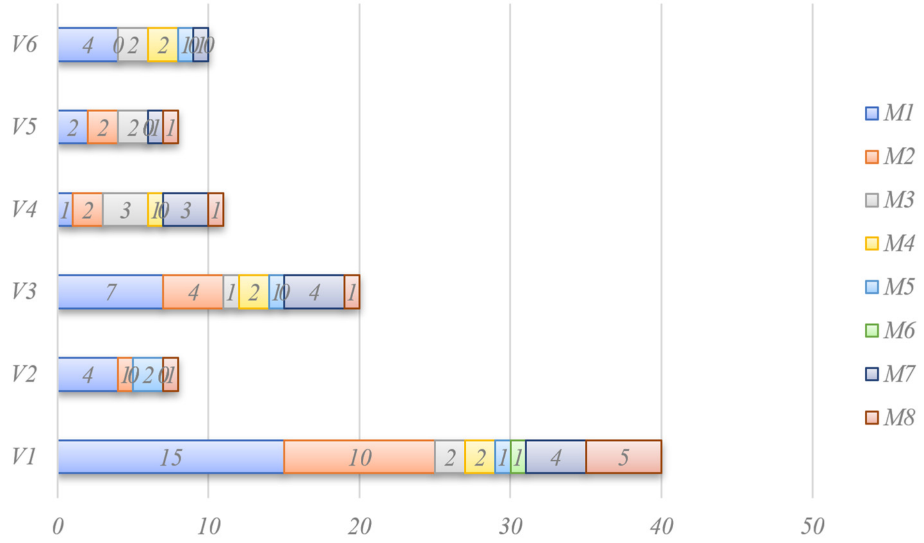

Figure 25. Distribution of evaluation metric families across validation strategies.

More rigorous validation approaches, including external test sets (V3), field testing (V4), and cross-dataset evaluation (V5), are present but remain secondary. Their use is more noticeable in segmentation- and geometry-related studies, suggesting that tasks requiring finer spatial reasoning tend to motivate stronger evaluation protocols. Overall, the stacked distribution highlights a methodological imbalance between the growing technical complexity of models and the relatively conservative evaluation strategies used to assess them.

4.5.2. Sankey Diagram Analysis (Metric–Validation Relationships)

The Sankey diagram (Figure 26) shows how different metric families are linked to validation strategies, and it largely follows the same patterns seen in the stacked chart. Most of the stronger links come from M1 and M2 toward V1, which shows that many studies still rely on conventional metrics evaluated using fixed dataset splits. This pattern suggests that, even though model architectures have advanced, evaluation is often limited when it comes to testing generalisations under new or unseen building conditions.

Weaker but still relevant links connect segmentation (M3) and quantification metrics (M4) with external testing (V3), field validation (V4), and cross-dataset evaluation (V5). This points to a growing awareness of robustness issues, especially for pixel-level analyses and geometric measurement tasks. By comparison, higher-level metric groups, including severity estimation (M5), SHM-related metrics (M7), and vision–language metrics (M6), are rarely paired with more demanding validation strategies. This highlights a mismatch between the complexity of these tasks and the way they are currently evaluated.

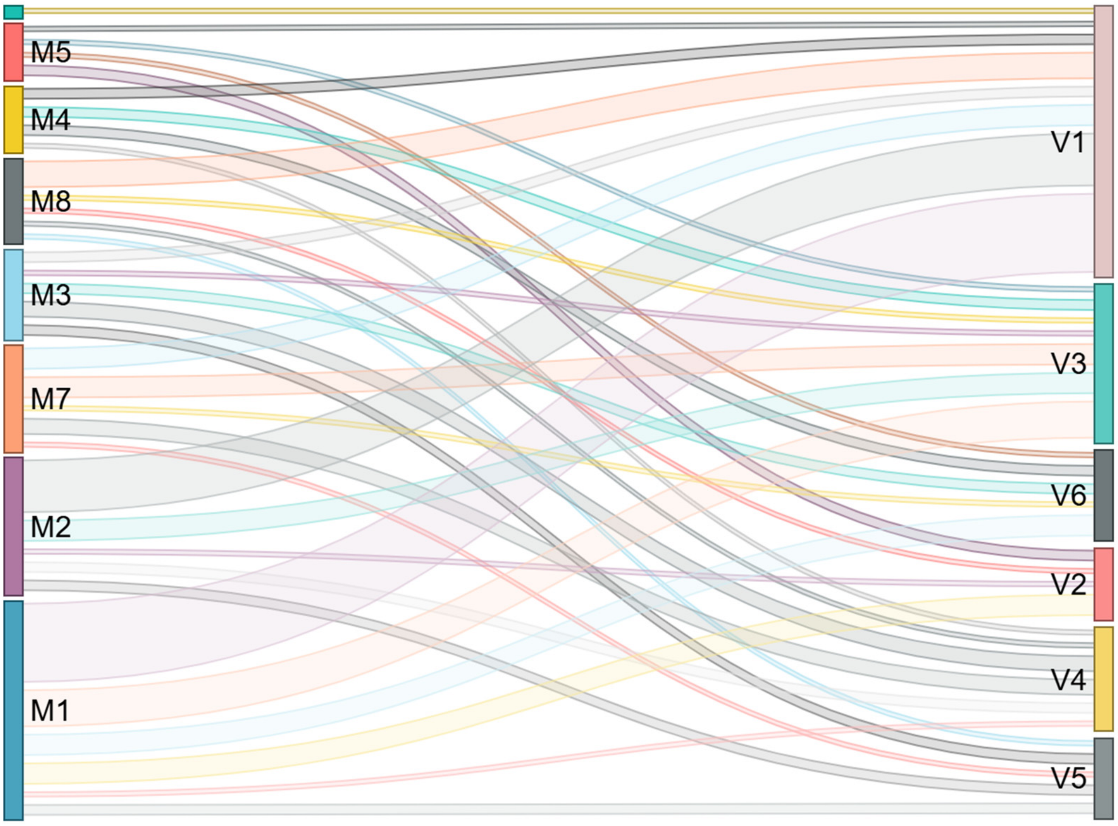

Figure 26. Relationships between metric families and validation protocols.

4.5.3. Evaluation Protocols and Metric Usage Patterns

Looking at the evaluation protocols, most studies rely on fixed hold-out validation strategies. As shown in Table 14, more than half of the reviewed studies employ a single train–validation–test split. At the same time, cross-validation and cross-dataset evaluation remain relatively rare. This gap means that even though models are becoming more complex, robustness and generalisation are still not checked consistently across different data conditions.

Table 14. Distribution of validation strategies across studies.

Validation Family	Doc IDs	Count
V1—Fixed Holdout	D1, D2, D4, D5, D6, D8, D10, D21, D22, D23, D24, D26, D27, D28, D29, D30, D31, D32, D36	19
V2—Cross-Validation	D3, D7, D18, D21	4
V3—External Test Set	D13, D16, D17, D19, D20, D26, D28, D29, D30	9
V4—Field Testing	D15, D19, D23, D25, D35	5
V5—Cross-Dataset Evaluation	D11, D27	2
V6—Weak/Unreported	D9, D12, D14, D33, D34	5

In terms of performance reporting, classification and detection metrics dominate the literature, as evidenced by Table 15. Metrics such as accuracy, precision, recall, F1-score, and mAP are widely used, reflecting the frequency of classification- and detection-oriented tasks. In contrast, segmentation, quantification, severity estimation, and SHM-specific metrics appear far less frequently, indicating that fine-grained performance assessments and structural interpretation are still secondary concerns in many studies.

Table 15. Frequency of metric families used in the reviewed studies.

Metric Family	**Documents (Doc IDs)**
M1—Classification Metrics	D1, D3, D4, D5, D7, D9, D11, D12, D13, D16, D17, D18, D20, D21, D24, D26, D27, D28, D30, D32, D33, D36
M2—Detection Metrics	D4, D6, D8, D10, D11, D16, D17, D19, D21, D22, D23, D26, D27, D29, D31
M3—Segmentation Metrics	D11, D14, D15, D19, D22, D24, D27, D33
M4—Quantification Metrics	D8, D10, D15, D17, D29, D34, D35
M5—Severity Estimation Metrics	D7, D18, D28, D33
M6—Captioning/Natural Language Processing (NLP) Metrics	D2
M7—SHM/Geometry Metrics	D10, D15, D17, D19, D20, D24, D27, D29, D33, D35
M8—Runtime/Computational Metrics	D6, D11, D17, D21, D22, D23, D31

Overall, these patterns point to a methodological gap: advanced tasks are often evaluated using limited validation schemes and a narrow set of metrics, which may overestimate real-world performance. Strengthening evaluation protocols, through external test sets, field validation, and task-appropriate metrics, remains a critical requirement for improving the reliability and deployability of AI-based crack detection systems.

4.5.4. Methodological Implications for Robust Evaluations

Regarding the robustness and generalisability of reported performance results, the evaluation protocols adopted in the reviewed studies play a decisive role. A substantial proportion of the analysed works rely on single fixed train–test or train–validation–test splits, which are known to produce optimistic and potentially unstable performance estimates, as results can be highly sensitive to the specific data partition used [94]. Such evaluation schemes may therefore fail to reflect model behaviour under dataset variability or real-world deployment conditions.

Many studies consider cross-validation and external testing to be more reliable than simple hold-out validation, as these approaches are less affected by data partitioning and better reflect generalisation performance, particularly in challenging domains like building inspections [95,96]. Using these evaluation approaches can lead to more reliable performance estimates, particularly when data are limited or when models are tested on new environments. However, they remain relatively uncommon in the reviewed literature. This indicates that validation practices often lag behind recent advances in model design and task complexity, pointing to the need for more careful validation in future studies, especially for real-world applications. As a result, many reported performance improvements mainly reflect controlled test settings, while robustness, generalisation, and real-world applicability are less well demonstrated.

4.6. Hyperparameter Optimisation and Methodological Rigour (TRQ5)

Hyperparameter optimisation (HPO) is a critical component of methodological rigour in machine-learning studies, as it directly affects model performance stability, fairness of comparison, and generalisation ability. Recent methodological reviews emphasise that insufficient or ad hoc tuning can lead to misleading performance estimates, particularly in deep-learning models with high sensitivity to training configurations [97].

4.6.1. Distribution of HPO Practices in the Selected Studies

The distribution of hyperparameter optimisation practices across the 36 reviewed studies is summarised in Table 16 and visualised in Figure 27. The results show that the majority of studies fall into either the no HPO category (H1) or rely on basic manual tuning strategies (H2), such as fixed learning rates or manually selected batch sizes. Structured manual tuning approaches (H3) are relatively rare and are typically observed in studies addressing more complex tasks or architectures. Notably, fully automated hyperparameter optimisation (H4), including grid search, random search, or Bayesian optimisation, appears in only a single study.

Table 16. HPO categories and associated studies.

HPO_Group	Types_of_HPO	Doc IDs	Count
H1_No_HPO	No tuning; fixed LR/BS/epochs; default settings; transfer learning without tuning; no LR/BS reported; rule-based pipelines without training	D1, D5, D9, D12, D13, D14, D16, D17, D18, D20, D25, D34	12
H2_Basic_ Manual_ Tuning	Manual LR setting; manual batch size; manual optimiser choice; fixed schedules; simple heuristics; no systematic exploration	D2, D3, D4, D6, D7, D10, D21, D22, D23, D24, D26, D27, D28, D30, D31, D32, D33, D36	18
H3_Structured_ Manual_ Tuning	Structured tuning strategies: LR decay, step decay, staged training, multi-stage optimisation, model-specific calibration	D11, D15, D19, D29, D35	5
H4_Automated_HPO	Grid search; random search; Bayesian optimisation; AutoML-style systematic tuning	D8	1

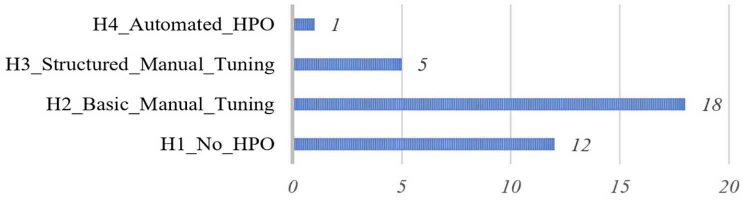

Figure 27. Distribution of HPO practices (H1–H4).

This distribution indicates that systematic exploration of hyperparameter spaces remains uncommon in AI-based crack detection research for buildings, despite its recognised importance in modern machine-learning workflows [98].

4.6.2. Interpretation: Implications for Methodological Rigour

The dominance of H1 and H2 practices suggests that many reported results may be sensitive to default settings or subjective design choices, rather than reflecting well-optimised model behaviour. Prior studies have shown that reliance on fixed or weakly tuned hyperparameters can introduce optimistic bias and hinder reproducibility, especially when models are evaluated on limited or homogeneous datasets [99].

The limited use of structured tuning (H3) in segmentation, quantification, and SHM-oriented studies implies that task complexity often necessitates more careful optimisation, even if such practices are not yet widespread. Automated hyperparameter optimisation (HPO) is rarely used in the reviewed studies, which points to a clear gap in current practice. This is especially noticeable given the growing complexity of YOLO-based detectors, transformer models, and CNN–transformer hybrids. Using automated optimisation methods could help make performance results more robust and easier to compare across different studies [100]. In general, hyperparameter optimisation has not yet become a regular part of crack detection research in building applications. Paying more attention to this step could help improve the quality and consistency of future studies.

5. Discussion

5.1. Insights from Scientometric Analysis

This scientometric analysis offers a broad look at how research on AI-based crack detection in building structures has changed over roughly the past decade. Over time, the number of studies has increased, and the field has slowly started to cluster around a few recurring themes, although approaches still vary quite widely. In many cases, this progress seems to follow advances in deep learning and computer vision, together with the more frequent use of UAV-based imaging and lightweight deployment solutions. Similar developments have also been mentioned in wider scientometric work on deep learning for crack detection and structural health monitoring, especially as modern AI tools became more common, although the pace and focus differ across application areas [13]. At the same time, development across the field is uneven, with parts of the research still fragmented and affected by ongoing methodological constraints.

After 2020, both publications and citations become more common across the dataset. China produces the largest share of studies, but countries like South Korea, the UK, and Greece often show higher citation impact. This kind of mismatch between output and impact has also been reported in earlier scientometric work on SHM [18]. Co-authorship patterns point to a similar picture, where only a few authors connect different groups, while most collaborations remain concentrated within regional clusters. Comparable fragmentation has been documented in scientometric reviews of construction materials and infrastructure technologies, where research groups tend to operate independently with limited cross-regional collaboration [23].

Thematic and temporal analyses reveal a logical methodological progression, from early CNN-based detection toward segmentation, UAV-based workflows, and, more recently, transformer architectures and emerging multimodal sensing. This reflects trends noted in several contemporary crack detection reviews, which emphasise the shift from classical image processing toward more adaptive and real-time deep-learning pipelines [1]. However, the rapid adoption of advanced models also raises concerns regarding validation depth; as highlighted in prior survey studies, many state-of-the-art architectures are trained on small or domain-specific datasets with limited attention to environmental generalisations and long-term robustness [3].

Bibliographic coupling results further show four coherent research streams: methodological innovation, SHM-oriented inspections, applied deep learning, and emerging multimodal sensing. While these clusters reflect conceptual richness, their weak interconnections suggest that the field's intellectual core is still forming. Strengthening these connections could support more consistent research directions and help limit unnecessary repetition in model development.

These observations carry a number of important implications. A strong reliance on deep-learning-based vision models may gradually narrow the range of methods being explored. At the same time, the limited availability of benchmark datasets continues to restrict replicability and makes meaningful comparisons between studies more difficult, a concern that has been raised repeatedly in recent review work [1,3]. UAV-based workflows are becoming more common, but many still fall short of fully integrated operational frameworks that account for environmental uncertainty and track how cracks develop over time.

Looking forward, there are a few areas where current work could be pushed further. One clear issue is the lack of open and shared datasets, which still makes it hard to compare results across studies. There is also growing interest in using more than one type of sensor data, for example by combining RGB images with thermal or Light Detection and Ranging (LiDAR) information. At the same time, many studies would benefit from closer links to basic SHM concepts, such as material behaviour and how cracks change over time. More recently, approaches based on transformer models, simple ways of handling uncertainty, and running models directly on edge devices have started to attract attention, reflecting trends seen across the recent literature [3]. Stronger collaboration between countries, together with better sharing of datasets, could still help the field move forward and make research results easier to apply more broadly.

In general, there has been clear progress, but the field is still affected by fragmented work, limited data resources, and an imbalanced focus across methods. Dealing with these issues will be important if current algorithms are to move beyond strong experimental results and become reliable, scalable tools for real-world building inspections.

These scientometric findings help clarify how the field has developed in terms of research output, collaboration, and thematic emphasis. They do not, however, show how specific methodological decisions have influenced the technical maturity of current AI-based crack detection systems. This gap is addressed in the systematic technical review that follows.

5.2. Insights from the Systematic Technical Review

5.2.1. Technical Implications from Task Formulation and Model Selection

The technical review shows that recent studies usually choose models that fit fairly closely with the task being addressed. Most of the literature focuses on detection-oriented problems and visual inspections within SHM, with an emphasis on methods that can be deployed at scale. Higher-level reasoning and decision support receive much less attention. Similar trends have been reported in earlier reviews, where deep learning is largely applied to perception-level tasks such as classification, detection, and segmentation [23,25].

CNN-based models are still the most commonly used option, largely because they work reliably, run efficiently, and can handle limited or varied datasets reasonably well [23]. At the same time, YOLO-based detection models are being used more often in real-time building inspection tasks, especially in UAV- and façade-level applications where speed and practical efficiency matter most [61]. Higher-level tasks, such as predictive diagnostics, automated maintenance support, and vision–language-based reporting, are still relatively uncommon. This suggests that the field has yet to move beyond perception-focused models toward approaches that support reasoning and decision-making [13].

These technical imbalances are in line with the fragmented collaboration patterns and uneven thematic focus seen in the scientometric analysis. In practice, this means that research areas are not yet well connected, and much of the work follows relatively low-risk, incremental paths [101,102].

5.2.2. Methodological Maturity and Evaluation Limitations

From a methodological perspective, the technical review highlights several areas where the field is still not fully mature. Most studies continue to rely on RGB images and relatively simple annotation schemes, usually limited to image-level labels or bounding boxes. This largely reflects practical constraints related to data collection and labelling effort, but it also limits the ability to support more detailed structural interpretations, accurate crack measurements, and longer-term analyses [23,25]. As a result, most datasets are better suited for detection tasks than for severity assessments or long-term damage modelling.

Model evaluation is another area where limitations repeatedly appear. A large number of studies rely on a single train–test split, with much less use of cross-validation or external testing. As a result, it is not always clear how stable or transferable the reported results are. Earlier reviews in deep-learning-based SHM and crack detection have raised similar points, suggesting that evaluation often does not keep pace with model design and may increase the risk of overfitting [13,61]. Based on these observations, recent work shows faster technical progress in crack detection, but more consistent methods and evaluations are still needed before these approaches can be trusted in real-world use.

5.3. Research Gaps and Future Research Directions

Despite the fact that research on AI-based crack detection for buildings has grown quickly, the combined scientometric and technical review highlights several gaps that still limit practical use. In many cases, these gaps make it difficult to move beyond visual crack recognition toward more decision-oriented structural health monitoring. Based on their relevance, the identified gaps are discussed as primary and secondary research directions.

5.3.1. Primary Research Gaps (PGs)

PG1. Lack of predictive and prognostic modelling:

Most of the reviewed studies concentrate on detecting existing cracks using classification, detection, or segmentation approaches, while much less attention is given to modelling how damage develops over time. As a result, current methods contribute only in a limited way to long-term structural health monitoring and proactive maintenance. Future work would benefit from placing more emphasis on time-aware and predictive learning approaches that can support prognosis and condition-based decision-making.

PG2. Absence of vision–language- and LLM-based inspection reporting:

The review shows that very few studies make use of vision–language models or large language models to turn visual crack or defect detection results into structured inspection reports. This gap creates a clear disconnect between what algorithms produce and how inspections are actually carried out in practice, where interpretation and reporting play a central role. Combining computer-vision methods with language models could help bridge this gap by supporting automated reporting, more transparent assessments, and inspection documentation that is easier to scale.

PG3. Material and regional bias toward concrete-dominated building stocks:

Most of the reviewed studies focus on reinforced concrete elements, with far less attention given to buildings made from a wider range of façade materials, such as brick masonry, stone, or composite systems. This narrow material focus limits how well current approaches can be applied in regions with more diverse building stocks, including the United Kingdom. Paying closer attention to material-specific visual characteristics is therefore important for developing inspection systems that can be transferred across different regional contexts.

5.3.2. Secondary Research Gaps (SGs)

SG1. Limited multimodal and temporal data integration:

Most studies still rely on RGB images and single, one-off inspections, with far less use of thermal, depth, LiDAR, or time-series data. This narrow data focus makes systems less reliable under changing environmental conditions and limits progress toward time-aware monitoring and predictive modelling.

SG2. Limited explainability and interpretability of AI models:

Even as models become more complex, only a small number of studies make use of explainable or interpretable AI methods to support transparency, trust, and engineering validation. Without this level of explainability, wider adoption remains difficult, particularly in safety critical areas such as building inspections and structural health monitoring.

SG3. Insufficient evaluation rigour, reproducibility, and optimisation practices:

In many cases, evaluation still relies on fixed hold-out validation, with little use of cross-dataset testing or real-world evaluations. Practices around hyperparameter tuning and reproducibility are also reported inconsistently, which makes it harder to judge how well results generalise, to compare studies fairly, and to maintain transparency in the research.

The research gaps identified in this review have direct implications for building maintenance and structural health monitoring. Currently, most AI-based systems are primarily designed to detect cracks after they appear, while predictive diagnostic capabilities remain limited. If these gaps are addressed, AI systems could support earlier identification of crack progression and allow maintenance actions to be planned before damage becomes more severe. In addition, improving dataset diversity and integrating multimodal inspection data (such as UAV imagery and other sensing sources) could increase the reliability of automated inspections under real building conditions. Without such developments, many current approaches may remain confined to experimental detection tasks and may offer limited support for practical maintenance planning and long-term structural management.

Beyond the maintenance aspects discussed above, these developments are also related to broader sustainability considerations in the built environment. When AI-assisted crack detection becomes more reliable, it can help identify deterioration at earlier stages and allow for intervention before damage becomes more severe. In practice, this may contribute to extending the service life of buildings, reducing unnecessary material replacement, and improving how inspection and maintenance activities are organised over the lifecycle of a building. In this sense, progress in AI-based crack detection may also support wider sustainability efforts in building management and urban development, which are reflected in international frameworks such as the United Nations Sustainable Development Goals, particularly Sustainable Development Goal (SDG) 9 (Industry, Innovation and Infrastructure) and SDG 11 (Sustainable Cities and Communities) [103].

6. Conclusions

This scientometric assessment helps clarify how research on AI-based crack detection in building structures has developed into a more recognisable and increasingly organised field. When looking at publication trends, themes, and citation links together, it seems that the field has slowly moved away from its early, scattered exploratory stage. The literature is not fully coherent yet, but over time some clearer groupings around methods and approaches have started to form. Recent work tends to focus more on deep learning and segmentation and, in a few cases, on multimodal ideas as well. This shift appears to be linked more to inspection-related needs than to purely algorithmic experimentation. At the same time, citation patterns show that only some research groups and directions have become more central as the field has developed. Others are still weakly connected or appear only occasionally and may still be in an early phase.

While the scientometric analysis focuses on the structure and growth of the field, the technical review looks more closely at the methods used in individual studies and how these choices influence the reported outcomes. Although AI-based crack detection in building structures has progressed in terms of model complexity and application scope, the technical findings point to a gap between innovation and overall methodological maturity. Much of the existing literature still focuses on detection-oriented tasks and visually driven SHM applications, often using CNN- or YOLO-based models that are geared toward practical deployment. In comparison, higher-level tasks, such as predictive diagnostics, decision support, or vision–language integration, receive much less attention and are still relatively underdeveloped.

Methodologically, a large part of the literature continues to depend on RGB data and relatively simple annotations, while more robust validation and systematic hyperparameter tuning remain uncommon. This gap makes it difficult to judge how reliable or transferable reported performance results are, especially when comparing studies that use different datasets or settings.

Looking at the scientometric and technical results side by side, the field appears to be moving past its early phase of rapid exploratory growth and into a more settled stage.

Within this setting, the review highlights a number of unresolved gaps, including limited attention to predictive and prognostic modelling, automated inspection reporting, material diversity, and the integration of multimodal and temporal data. Gaps also remain around explainable AI and overall methodological rigour. Addressing these issues will be important if recent algorithmic advances are to translate into reliable, scalable, and decision-oriented systems for real-world building inspections and structural health monitoring.

Author Contributions: Conceptualization, M.M., A.B.-J. and S.R.; methodology, M.M. and S.R.; software, M.M.; validation, A.B.-J.; formal analysis, M.M.; investigation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M., A.B.-J. and S.R. visualization, M.M.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analysed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence	LR	Learning rate
BS	Batch size	mAP	Mean average precision
CAM		ML	Machine learning
CNNs	Convolutional neural networks	PRISMA	Preferred reporting items for systematic review and meta-analyses
CV	Computer vision	R-CNNs	Region-based CNNs
DIP	Digital image processing	RGB	Red, green and blue
DL	Deep learning	SDG	Sustainable Development Goal
DOI	Digital object identifier	SHM	Structural health monitoring
EID	Electronic identifier	TLS	Total link strength
EBSCO	Elton B. Stephens Company,	U-Net	U-shaped Network
GANs	Generative adversarial networks	UWL	University of West London
HPO	Hyperparameter optimization	UAV	Unmanned Aerial Vehicle
IoU	Intersection over unions	VOSViewr	Visualization of Similarities viewer
LiDAR	Light detection and range	YOLO	You only look once
LLMs	Large language models		

Appendix A

Table A1. PRISMA 2020 checklist for the present systematic review.

Section and Topic	Item #	Checklist Item	Location Where Item is Reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 1

Table A1. *Cont.*

Section and Topic	Item #	Checklist Item	Location Where Item is Reported
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 3
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 4
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 6–7
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 5
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 5—Appendix A
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 6
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 7
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g., for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 7
	10b	List and define all other variables for which data were sought (e.g., participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 7
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 8
Effect measures	12	Specify for each outcome the effect measure(s) (e.g., risk ratio, mean difference) used in the synthesis or presentation of results.	N/A
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g., tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 7
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	N/A
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 7
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 7
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g., subgroup analysis, meta-regression).	N/A
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A

Table A1. *Cont.*

Section and Topic	Item #	Checklist Item	Location Where Item is Reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Appendix C
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Appendix C
Study characteristics	17	Cite each included study and present its characteristics.	N/A
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	N/A
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g., confidence/credible interval), ideally using structured tables or plots.	N/A
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	N/A
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g., confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	N/A
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	N/A
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 36–37
	23b	Discuss any limitations of the evidence included in the review.	Page 37
	23c	Discuss any limitations of the review processes used.	N/A
	23d	Discuss implications of the results for practice, policy, and future research.	Page 38–39
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	N/A
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	N/A
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

Note: N/A indicates not applicable.

Appendix B

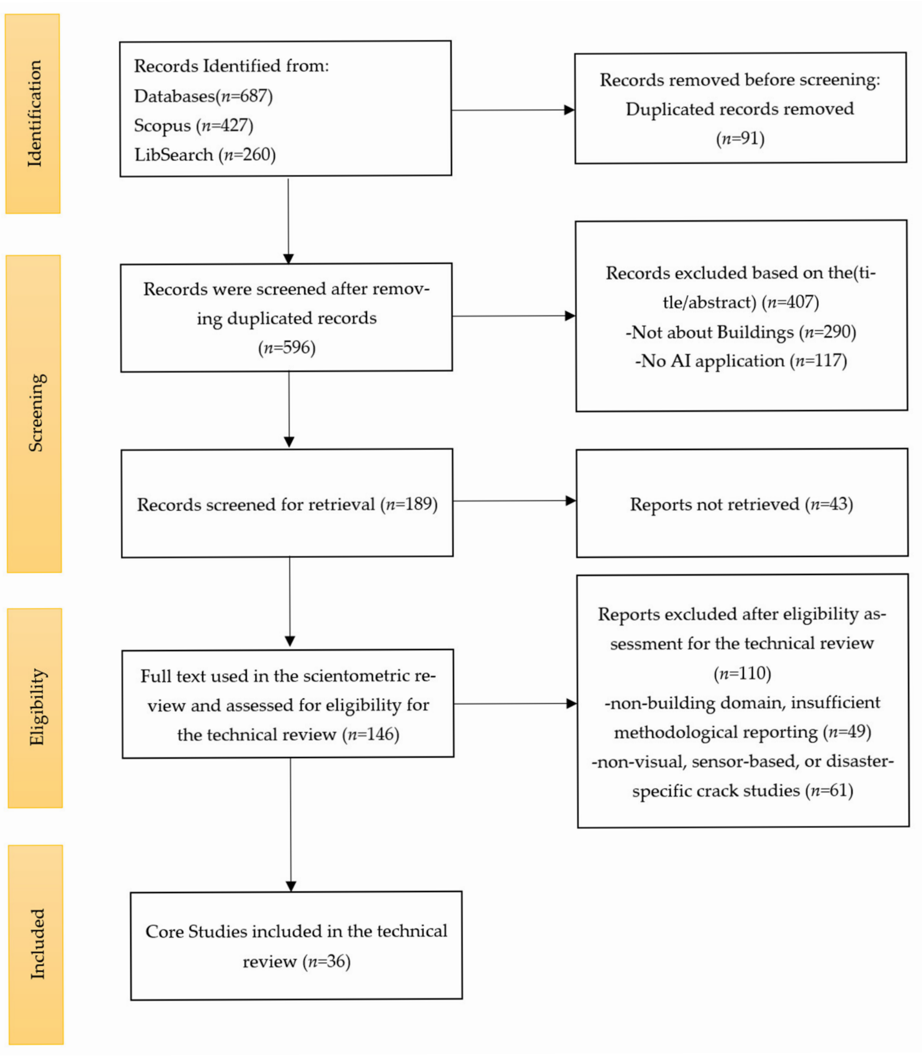

Figure A1. PRISMA flow diagram of the study selection process.

Appendix C

Scopus Boolean Search String

(TITLE (structure+ OR building OR construction) AND TITLE (crack OR defect OR damage OR anomaly OR fracture OR break OR opening OR gap OR microcrack OR health OR fault OR rupture OR Discontinuity OR flaw OR condition) AND TITLE (monitor* OR inspect* OR detect* OR evaluate* OR assess* OR Diagnosis OR Surveillance OR prognosis OR Examin* OR Survey OR Localize* OR predict* OR inspect* OR detect* OR evaluat* OR progress OR propagation OR forecast* OR proactive OR anticipat* OR project*) AND TITLE (AI OR “artificial intelligence” OR “deep learning” OR “machine learning” OR “computer vision” OR “image processing” OR “object detection” OR YOLO OR “convolutional neural network” OR CNN OR ANN OR “neural network” OR “artificial neural network” OR LSTM OR “Long Short-Term Memory” OR “Temporal Convolutional Network” OR transformer OR “vision transformer” OR “attention model” OR ensemble OR autoencoder OR “generative adversarial network” OR GAN OR “transfer learning” OR “feature extraction” OR U-Net OR ResNet OR EfficientNet OR “semantic segmentation” OR “instance segmentation” OR “semi_supervised” OR reinforcement OR autonomous) AND TITLE (building OR house OR residential OR facade)) AND PUBYEAR > 2014 AND PUBYEAR

< 2027 AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO (DOCTYPE,"re")) AND (LIMIT-TO (LANGUAGE,"English")).

Appendix D

Table A2. Classification of AI model families used in the reviewed studies.

	Model Family 1: CNN-Based Models	**Model Family 2: Transformer-Based Models**	**Model Family 3: CNN–transformer Hybrid Models**	**Model Family 4: Detection Frameworks (YOLO/R-CNN Family)**	**Model Family 5: Classical/Shallow ML & Analytical Pipelines**	**Model Family 6: LLM-based Vision–Language Models**
Models	Inception-V3, Xception, ResNet50, ResNet101, ResNet34, ResNet-50 backbone, EfficientUNet++ (EfficientNet-B5 backbone), UNet, U-Net++, FCN/Improved FCN, FastSCNN, K-Net, EfficientNet-B5, MobileNetV2, DenseNet169, GhostNet, Custom CNNs, Oxford DCNN, CCNN, DLCD (ResNet-50 + attention), CycleGAN + UNet + CRF + Guided Filter, U-Net detailed version	Vision Transformer (ViT-B16), Swin-Transformer (Swin-Crack), Swin-B, Swin-T backbone, IBR-Former, Vmamba	CCTNet, KAN-MobileViT, KAN-ViT, KAN-Hybrid, cLGFAF-Net (ResNet34 + VMamba)	YOLOv8 variants, YOLOv7 (BFD-YOLOv7), YOLOv8n (BCCD-YOLO), YOLOv8n-seg (ESE-YOLO), YOLOv4-tiny, YOLOv5, YOLOv5 + DeepSORT, Mask R-CNN, Faster R-CNN (MultiDefectNet), RPN + FPN + RoIAlign, Multi-feature fusion networks (ESE-YOLO, LBA-YOLO)	CrackNet-Hybrid (CNN + SVM), Thermal Gaussian separation pipeline, Morphology + Skeletonization pipeline, Adaptive Image Noise Reduction Pipeline	Encoders: Inception-V3, Xception, ResNet50, Decoders: LSTM, GRU, Bahdanau Attention

Appendix E

Table A3. Classification of evaluation metric families used in the reviewed studies.

Metric Family	Title	Metrics Included
M1	Classification Metrics	Accuracy, Precision, Recall, F1-score, Specificity, ROC-AUC, Confusion Matrix
M2	Detection Metrics	AP, mAP@0.5, mAP@0.5:0.95, PR curves, FPS, Segmentation-based mAP
M3	Segmentation Metrics	IoU, Dice, mIoU, BF-score, Pixel Accuracy
M4	Quantification Metrics	MAE (width/length), Absolute Error, Relative Error (%), Pixel-level geometric error, cm-level localization error
M5	Severity Estimation Metrics	Kappa coefficient, Severity Accuracy
M6	Captioning/NLP Metrics	BLEU-1/2/3/4, ROUGE-L, CIDEr, METEOR, SPICE
M7	SHM/Geometry Metrics	Crack geometry extraction accuracy, Length/width vs. GT, Structural diagnostic KPIs
M8	Runtime/Computational Metrics	FPS, FLOPs, Parameters (M), Inference time

Appendix F

Table A4. Classification of validation strategy families used in the reviewed studies.

Validation Family	Description	Examples
V1—Fixed Hold-Out Split	Any fixed train/val/test split without cross-validation	80/20, 75/25, 70/20/10, 60/20/20, 7:2:1, 8:1:1, 90/10, simple Train/Test
V2—Cross-Validation	k-fold or repeated/random split validation	5-fold CV; repeated runs; multiple random splits
V3—External/Independent Test Set	Evaluation on a completely separate dataset not used in training	Internet images; unseen UAV dataset; new field images
V4—Real-World Field Testing	Validation on real measured cracks, physical experiments, UAV field missions	UAV tests; physical crack measurement; thermal lab testing
V5—Cross-Dataset Evaluation	Training on one dataset and testing on another dataset (domain generalization)	DeepCrack → CrackSeg9k, SDNET2018 evaluation, Diverse Set A/B/C
V6—Unreported/Weak Evaluation	No clear split; metrics vague; qualitative only	No ratio reported; no held-out test set; qualitative comparison only

References

1. Kaveh, H.; Alhajj, R. Recent Advances in Crack Detection Technologies for Structures: A Survey of 2022–2023 Literature. *Front. Built Environ.* **2024**, *10*, 1321634. [CrossRef]
2. Shah Mansouri, T.; Lubarsky, G.; Finlay, D.; McLaughlin, J. Machine Learning-Based Structural Health Monitoring Technique for Crack Detection and Localisation Using Bluetooth Strain Gauge Sensor Network. *J. Sens. Actuator Netw.* **2024**, *13*, 79. [CrossRef]
3. Yuan, Q.; Shi, Y.; Li, M. A Review of Computer Vision-Based Crack Detection Methods in Civil Infrastructure: Progress and Challenges. *Remote Sens.* **2024**, *16*, 2910. [CrossRef]
4. Sohaib, M.; Arif, M.; Kim, J.-M. Evaluating YOLO Models for Efficient Crack Detection in Concrete Structures Using Transfer Learning. *Buildings* **2024**, *14*, 3928. [CrossRef]
5. Ren, W.; Zhong, Z. LBA-YOLO: A Novel Lightweight Approach for Detecting Micro-Cracks in Building Structures. *PLoS ONE* **2025**, *20*, e0321640. [CrossRef] [PubMed]
6. Loverdos, D.; Sarhosis, V. Automatic Image-Based Brick Segmentation and Crack Detection of Masonry Walls Using Machine Learning. *Autom. Constr.* **2022**, *140*, 104389. [CrossRef]
7. Nguyen, C.L.; Nguyen, A.; Brown, J.; Byrne, T.; Ngo, B.T.; Luong, C.X. Optimising Concrete Crack Detection: A Study of Transfer Learning with Application on Nvidia Jetson Nano. *Sensors* **2024**, *24*, 7818. [CrossRef]
8. Gao, Y.; Zhai, P.; Mosalam, K.M. Balanced Semisupervised Generative Adversarial Network for Damage Assessment from Low-data Imbalanced-class Regime. *Comput. Aided Civ. Eng.* **2021**, *36*, 1094–1113. [CrossRef]
9. Gao, S.; Dai, Y.; Xu, Y.; Chen, J.; Liu, Y. Generative Adversarial Network–Assisted Image Classification for Imbalanced Tire X-Ray Defect Detection. *Trans. Inst. Meas. Control* **2023**, *45*, 1492–1504. [CrossRef]
10. Ai, D.; Zhang, R. Deep Learning of Electromechanical Admittance Data Augmented by Generative Adversarial Networks for Flexural Performance Evaluation of RC Beam Structure. *Eng. Struct.* **2023**, *296*, 116891. [CrossRef]
11. Choi, D.; Bell, W.; Kim, D.; Kim, J. UAV-Driven Structural Crack Detection and Location Determination Using Convolutional Neural Networks. *Sensors* **2021**, *21*, 2650. [CrossRef] [PubMed]
12. Chang, S.; Zheng, B. A Lightweight Convolutional Neural Network for Automated Crack Inspection. *Constr. Build. Mater.* **2024**, *416*, 135151. [CrossRef]
13. Wu, Y.; Li, S.; Li, J.; Yu, Y.; Li, J.; Li, Y. Deep Learning in Crack Detection: A Comprehensive Scientometric Review. *J. Infrastruct. Intell. Resil.* **2025**, *4*, 100144. [CrossRef]
14. Ai, D.; Jiang, G.; Lam, S.-K.; He, P.; Li, C. Computer Vision Framework for Crack Detection of Civil Infrastructure—A Review. *Eng. Appl. Artif. Intell.* **2023**, *117*, 105478. [CrossRef]
15. Dorafshan, S.; Thomas, R.J.; Maguire, M. Comparison of Deep Convolutional Neural Networks and Edge Detectors for Image-Based Crack Detection in Concrete. *Constr. Build. Mater.* **2018**, *186*, 1031–1045. [CrossRef]
16. Bai, J.; Wu, D.; Shelley, T.; Schubel, P.; Twine, D.; Russell, J.; Zeng, X.; Zhang, J. A Comprehensive Survey on Machine Learning Driven Material Defect Detection. *ACM Comput. Surv.* **2025**, *57*, 3730576. [CrossRef]
17. Gong, H.; Liu, L.; Liang, H.; Zhou, Y.; Cong, L. A State-of-the-Art Survey of Deep Learning Models for Automated Pavement Crack Segmentation. *Int. J. Transp. Sci. Technol.* **2024**, *13*, 44–57. [CrossRef]
18. Ali, L.; Alnajjar, F.; Khan, W.; Serhani, M.A.; Al Jassmi, H. Bibliometric Analysis and Review of Deep Learning-Based Crack Detection Literature Published between 2010 and 2022. *Buildings* **2022**, *12*, 432. [CrossRef]
19. Gupta, P.; Dixit, M. Image-Based Crack Detection Approaches: A Comprehensive Survey. *Multimed. Tools Appl.* **2022**, *81*, 40181–40229. [CrossRef]
20. Munawar, H.S.; Hammad, A.W.A.; Haddad, A.; Soares, C.A.P.; Waller, S.T. Image-Based Crack Detection Methods: A Review. *Infrastructures* **2021**, *6*, 115. [CrossRef]
21. Hsieh, Y.-A.; Tsai, Y.J. Machine Learning for Crack Detection: Review and Model Performance Comparison. *J. Comput. Civ. Eng.* **2020**, *34*, 04020038. [CrossRef]
22. Ashraf, A.; Sophian, A.; Shafie, A.A.; Gunawan, T.S.; Ismail, N.N. Machine Learning-Based Pavement Crack Detection, Classification, and Characterization: A Review. *Bull. Electr. Eng. Inform.* **2023**, *12*, 3601–3619. [CrossRef]
23. Hamishebahar, Y.; Guan, H.; So, S.; Jo, J. A Comprehensive Review of Deep Learning-Based Crack Detection Approaches. *Appl. Sci.* **2022**, *12*, 1374. [CrossRef]
24. Khan, S.; Jan, A.; Seo, S. Structural Crack Detection Using Deep Learning: An In-Depth Review. *Korean J. Remote Sens.* **2023**, *39*, 371–393. [CrossRef]
25. König, J.; Jenkins, M.; Mannion, M.; Barrie, P.; Morison, G. What's Cracking? A Review and Analysis of Deep Learning Methods for Structural Crack Segmentation, Detection and Quantification 2022. *arXiv* **2022**, arXiv:2202.03714.
26. Dong, C.-Z.; Catbas, F.N. A Review of Computer Vision–Based Structural Health Monitoring at Local and Global Levels. *Struct. Health Monit.* **2021**, *20*, 692–743. [CrossRef]
27. Payawal, J.M.G.; Kim, D.-K. Image-Based Structural Health Monitoring: A Systematic Review. *Appl. Sci.* **2023**, *13*, 968. [CrossRef]

28. Yeum, C.; Dyke, S. Vision-Based Automated Crack Detection for Bridge Inspection. *Comput.-Aided Civ. Infrastruct. Eng.* **2015**, *30*, 759–770. [CrossRef]
29. Cha, Y.-J.; Ali, R.; Lewis, J.; Büyüköztürk, O. Deep Learning-Based Structural Health Monitoring. *Autom. Constr.* **2024**, *161*, 105328. [CrossRef]
30. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef]
31. Perianes-Rodriguez, A.; Waltman, L.; Eck, N.J. van Constructing Bibliometric Networks: A Comparison between Full and Fractional Counting 2016. *J. Informetr.* **2016**, *10*, 1178–1195. [CrossRef]
32. Zupic, I.; Čater, T. Bibliometric Methods in Management and Organization. *Organ. Res. Methods* **2015**, *18*, 429–472. [CrossRef]
33. Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W.M. How to Conduct a Bibliometric Analysis: An Overview and Guidelines. *J. Bus. Res.* **2021**, *133*, 285–296. [CrossRef]
34. Ellegaard, O.; Wallin, J.A. The Bibliometric Analysis of Scholarly Production: How Great Is the Impact? *Scientometrics* **2015**, *105*, 1809–1831. [CrossRef]
35. Plevris, V.; Papazafeiropoulos, G. AI in Structural Health Monitoring for Infrastructure Maintenance and Safety. *Infrastructures* **2024**, *9*, 225. [CrossRef]
36. Kralovec, C.; Schagerl, M. Review of Structural Health Monitoring Methods Regarding a Multi-Sensor Approach for Damage Assessment of Metal and Composite Structures. *Sensors* **2020**, *20*, 826. [CrossRef]
37. Page, M.J.; Moher, D.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. PRISMA 2020 Explanation and Elaboration: Updated Guidance and Exemplars for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n160. [CrossRef]
38. Rabbani, M.G.; Alam, A.; Prybutok, V.R. Digital Health Transformation Through Telemedicine (2020–2025): Barriers, Facilitators, and Clinical Outcomes—A Systematic Review and Meta-Analysis. *Encyclopedia* **2025**, *5*, 206. [CrossRef]
39. Martín-Martín, A.; Orduna-Malea, E.; Thelwall, M.; Delgado López-Cózar, E. Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories. *J. Informetr.* **2018**, *12*, 1160–1177. [CrossRef]
40. Mongeon, P.; Paul-Hus, A. The Journal Coverage of Web of Science and Scopus: A Comparative Analysis. *ResearchGate* **2025**, *106*, 213–228. [CrossRef]
41. Room, S.; Bahadori-Jahromi, A. Biochar-Enhanced Carbon-Negative and Sustainable Cement Composites: A Scientometric Review. *Sustainability* **2024**, *16*, 10162. [CrossRef]
42. Smol, M. Circular Economy in Wastewater Treatment Plant—Water, Energy and Raw Materials Recovery. *Energies* **2023**, *16*, 3911. [CrossRef]
43. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report*; Scientific Research Publishing: Irvine, CA, USA, 2007; Available online: https://www.scirp.org/reference/ReferencesPapers?ReferenceID=1555797 (accessed on 13 December 2025).
44. Petersen, K.; Feldt, R.; Mujtaba, S.; Mattsson, M. Systematic Mapping Studies in Software Engineering. In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, Bari, Italy, 26–27 June 2008; Volume 17.
45. Popay, J.; Roberts, H.; Sowden, A.; Petticrew, M.; Arai, L.; Rodgers, M.; Britten, N.; Roen, K.; Duffy, S. *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme*; University of Lancaster: Lancaster, UK, 2006.
46. Shin, C. A Study on the Artificial Intelligence Model Using Various Pre-Processing Techniques for Crack Detection in Buildings. In *Proceedings of the IEEE Integrated STEM Education Conference, ISEC*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2025.
47. Mariniuc, A.M.; Dorian, D.; Abagiu, M.M. Building Surface Defect Detection Using Machine Learning and 3D Scanning Techniques in the Construction Domain. *Buildings* **2024**, *14*, 669. [CrossRef]
48. Garrido, I.; Barreira, E.; Almeida, R.; Lagüela, S. Introduction of Active Thermography and Automatic Defect Segmentation in the Thermographic Inspection of Specimens of Ceramic Tiling for Building Façades. *Infrared. Phys. Technol.* **2022**, *121*, 104012. [CrossRef]
49. Dinh, N.N.H.; Shin, H.; Ahn, Y.; Oo, B.L.; Lim, B.T.H. Attention-Based Image Captioning for Structural Health Assessment of Apartment Buildings. *Autom. Constr.* **2024**, *167*, 105677. [CrossRef]
50. Wang, J.; Li, Y.; Ueda, T. Comparative Analysis of Deep Learning Segmentation Algorithms in Concrete Damage Detection for Building Inspection Applications. In *Proceedings of the Lecture Notes in Civil Engineering.*; Tangtermsirikul, S., Panuwatwanich, K., Tanapornraweekit, G., Warnitchai, P., Eds.; Springer Science and Business Media Deutschland GmbH: Berlin/Heidelberg, Germany, 2025; Volume 726 LNCE, pp. 465–471.
51. Wang, J.; Ueda, T. Automatic Damage Detection and Segmentation Using Deep Learning Algorithms in Reinforced Concrete Structure Inspections. *Struct. Concr.* **2025**, *26*, 5511–5534. [CrossRef]

52. Ding, W.; Yang, H.; Yu, K.; Shu, J. Crack Detection and Quantification for Concrete Structures Using UAV and Transformer. *Autom. Constr.* **2023**, *152*, 104929. [CrossRef]
53. He, Y.; Yuan, Z.; Xia, X.; Yang, B.; Wu, H.; Fu, W.; Yao, W. Local–Global Feature Adaptive Fusion Network for Building Crack Detection. *Sensors* **2024**, *24*, 7076. [CrossRef]
54. Wang, J.; Wang, P.; Qu, L.; Pei, Z.; Ueda, T. Automatic Detection of Building Surface Cracks Using UAV and Deep Learning-Combined Approach. *Struct. Concr.* **2024**, *25*, 2302–2322. [CrossRef]
55. Ribeiro, W.S.; Zanetti, J.; Totola, L.B.; Junqueira, S.A.C.; Lauff, P.H.P. Crack Detection in Buildings Using the YOLO v8 Network. *Rev. Alconpat.* **2024**, *14*, 288–298. [CrossRef]
56. Busheska, A.; Sabella, N.; Almeida, N.; Rocha, E. Machine Learning and Thermography Applied to the Detection and Classification of Cracks in Buildings. In *Proceedings of the IEEE Conference on Technologies for Sustainability, SusTech*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023; pp. 247–251.
57. Kottari, P.; Arjunan, P. BD3: Building Defects Detection Dataset for Benchmarking Computer Vision Techniques for Automated Defect Identification. In *Proceedings of the BuildSys—ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*; Association for Computing Machinery, Inc.: New York, NY, USA, 2024; pp. 297–301.
58. Huo, Z.; Wu, X.; Cheng, T. Crack Recognition and Defect Detection of Assembly Building Constructions for Intelligent Construction. *J. Meas. Eng.* **2024**, *12*, 485–501. [CrossRef]
59. Tan, Y.; Li, G.; Cai, R.; Ma, J.; Wang, M. Mapping and Modelling Defect Data from UAV Captured Images to BIM for Building External Wall Inspection. *Autom. Constr.* **2022**, *139*, 104284. [CrossRef]
60. Wei, G.; Wan, F.; Zhou, W.; Xu, C.; Ye, Z.; Liu, W.; Lei, G.; Xu, L. BFD-YOLO: A YOLOv7-Based Detection Method for Building Façade Defects. *Electronics* **2023**, *12*, 3612. [CrossRef]
61. Shi, T.; Luo, H. Deep Learning for Automated Detection and Classification of Crack Severity Level in Concrete Structures. *Constr. Build. Mater.* **2025**, *472*, 140793. [CrossRef]
62. Akgül, İ. Mobile-DenseNet: Detection of Building Concrete Surface Cracks Using a New Fusion Technique Based on Deep Learning. *Heliyon* **2023**, *9*, e21097. [CrossRef]
63. Yadav, D.P.; Sharma, B.; Chauhan, S.; Ben Dhaou, I.B. Bridging Convolutional Neural Networks and Transformers for Efficient Crack Detection in Concrete Building Structures. *Sensors* **2024**, *24*, 4257. [CrossRef]
64. Zhou, X.; Tiong, R.L.K. Defects Inspection System for Building Facades Using Drones and Deep Learning Method. *Expert Sys. Appl.* **2026**, *298*, 129715. [CrossRef]
65. Lee, K.; Hong, G.; Sael, L.; Lee, S.; Kim, H.Y. Multidefectnet: Multi-Class Defect Detection of Building Façade Based on Deep Convolutional Neural Network. *Sustainability* **2020**, *12*, 9785. [CrossRef]
66. Ren, W.; Zhong, Z. Building Construction Crack Detection with BCCD YOLO Enhanced Feature Fusion and Attention Mechanisms. *Sci. Rep.* **2025**, *15*, 23167. [CrossRef]
67. Ramkumar, G.; Jency, S. Empirical Evaluation of Building Crack Identification and Classification Using Deep Learning Methodology. In *Proceedings of the International Conference on Integration of Emerging Technologies for the Digital World, ICIETDW*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2024.
68. Asif, K.M.S.; Isty, M.N.; Islam, R.U.; Gani, R.; Islam, T.; Khan, M.S.H.; Ahmed, J.; Rashid, M.R.A. Real-Time UAV-Based Building Surface Defect Detection: A Dataset-Driven Lightweight CNN Framework with Grad-CAM Explainability. In *Proceedings of the IEEE International Conference on Quantum Photonics, Artificial Intelligence, and Networking, QPAIN*; Shah, M.S., Ed.; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2025.
69. Keerthana, B.; Devi, J.A.; Madhuri, S.; Rao, M.S.; Rao, K.V.; Geethika, P.S.S. Building Crack Detection Using CNN and Class Activation Mapping. In *Proceedings of the International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application, ICAIQSA*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2024.
70. Interlando, M.; Pacifico, M.G.; Novellino, A.; Pastore, V.P. Ensembles of Deep Neural Networks for the Automatic Detection of Building Facade Defects From Images. *IEEE Access* **2024**, *12*, 164953–164965. [CrossRef]
71. Bian, X.; Zhao, G.; Li, S.; Wang, Q. Research on Building Crack Detection and Extraction Methods Based on Deep Learning and Image Processing Techniques. In *Proceedings of the International Conference on Information Systems and Computer Aided Education, ICISCAE*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2024; pp. 1302–1306.
72. Tan, Y.; Xu, W.; Chen, P.; Zhang, S. Building Defect Inspection and Data Management Using Computer Vision, Augmented Reality, and BIM Technology. *Autom. Constr.* **2024**, *160*, 105318. [CrossRef]
73. Liu, J.; Mao, P.; Ma, M.; Miao, Z. ESE-YOLO: A Lightweight Instance Segmentation Model for Building Façade Crack Detection. In *Proceedings of the International Conference on Artificial Intelligence and Electromechanical Automation, AIEA*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2025; pp. 237–243.
74. Wang, H.; Xiao, S. Research on Building Crack Width Monitoring Method Based on Computer Vision. In *Proceedings of the International Signal Processing, Communications and Engineering Management Conference, ISPCEM*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023; pp. 314–319.

75. Xue, S. Building Material Defect Detection and Diagnosis Method Based on Big Data and Deep Learning. *Informatica* **2024**, *48*, 37–48. [CrossRef]
76. Angan, R.B.; Hossain, M.S.; Nobel, S.S.; Hasan, M.M. Health Monitoring of Old Buildings in Bangladesh: Detection of Cracks and Dampness Using Image Processing. In *Proceedings of the International Conference on Next-Generation Computing, IoT and Machine Learning, NCIM*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2023.
77. Tang, S.; Wang, H. Research on Health Monitoring and Damage Recognition Algorithm of Building Structures Based on Image Processing. *J. Intell. Fuzzy Syst.* **2024**, *46*, 10305–10314. [CrossRef]
78. Chen, Y.; Zhu, Z.; Lin, Z.; Zhou, Y. Building Surface Crack Detection Using Deep Learning Technology. *Buildings* **2023**, *13*, 1814. [CrossRef]
79. Munawar, H.S.; Ullah, F.; Heravi, A.; Thaheem, M.J.; Maqsoom, A. Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages. *Drones* **2022**, *6*, 5. [CrossRef]
80. Dissanayake, D.M.K.I.; Rathnayake, R.S.M.P.W.; Sathsara, N.T.A.; Silva, C.S. UAV-Based Building Crack Detection Using Convolution Neural Networks. In *Proceedings of the International Conference on Information Technology Research*; Institute of Electrical and Electronics Engineers Inc.: New York, NY, USA, 2024.
81. Higgins, J.P.T.; Green, S. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration*; Scientific Research Publishing: Irvine, CA, USA, 2011; Available online: https://www.scirp.org/reference/referencespapers?referenceid=2225527 (accessed on 13 December 2025).
82. Zhuang, H.; Cheng, Y.; Zhou, M.; Yang, Z. Deep Learning for Surface Crack Detection in Civil Engineering: A Comprehensive Review. *Measurement* **2025**, *248*, 116908. [CrossRef]
83. Khan, M.Z.; Shahzadi, M.; Khan, A.; Ali, U.; Hassan, M.A.S.; Hussain, M. Review on Crack Detection in Civil Infrastructure Using Structural Health Monitoring and Machine Learning Techniques. *Innov. Infrastruct. Solut.* **2025**, *10*, 348. [CrossRef]
84. Chen, Y.-C.; Wu, R.-T.; Puranam, A. Multi-Task Deep Learning for Crack Segmentation and Quantification in RC Structures. *Autom. Constr.* **2024**, *166*, 105599. [CrossRef]
85. Raghu, M.; Unterthiner, T.; Kornblith, S.; Zhang, C.; Dosovitskiy, A. Do Vision Transformers See Like Convolutional Neural Networks? *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12116–12128.
86. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
87. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767. [CrossRef]
88. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
89. Che, L.; Liu, X.; Wen, Y.; Li, Z. Identification of Cascading Failure Initiated by Hidden Multiple-Branch Contingency. *IEEE Trans. Reliab.* **2019**, *68*, 149–160. [CrossRef]
90. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiyama, T.; Omata, H. Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 1127–1141. [CrossRef]
91. Zhang, X.; Wang, H.; Hsieh, Y.-A.; Yang, Z.; Yezzi, A.; Tsai, Y.-C. Deep Learning for Crack Detection: A Review of Learning Paradigms, Generalizability, and Datasets. *arXiv* **2025**, arXiv:2508.10256. [CrossRef]
92. Alexander, Q.G.; Hoskere, V.; Narazaki, Y.; Maxwell, A.; Spencer, B.F. Fusion of Thermal and RGB Images for Automated Deep Learning Based Crack Detection in Civil Infrastructure. *AI Civ. Eng.* **2022**, *1*, 3. [CrossRef]
93. Torralba, A.; Efros, A.A. Unbiased Look at Dataset Bias. In *Proceedings of the CVPR 2011*; IEEE: Colorado Springs, CO, USA, 2011; pp. 1521–1528.
94. Kohavi, R. A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the International Joint Conferences on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995.
95. Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Statist. Surv.* **2010**, *4*, 40–79. [CrossRef]
96. Wilimitis, D.; Walsh, C.G. Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial. *JMIR AI* **2023**, *2*, e49023. [CrossRef] [PubMed]
97. Probst, P.; Boulesteix, A.-L.; Bischl, B. Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *J. Mach. Learn. Res.* **2019**, *20*, 1–32.
98. Ilemobayo, J.; Durodola, O.; Alade, O.; Awotunde, O.; Adewumi, T.; Falana, O.; Ogungbire, A.; Osinuga, A.; Ogunbiyi, D.; Odezuligbo, I.; et al. Hyperparameter Tuning in Machine Learning: A Comprehensive Review. *J. Eng. Res. Rep.* **2024**, *26*, 388–395. [CrossRef]
99. Yang, L.; Shami, A. On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. *Neurocomputing* **2020**, *415*, 295–316. [CrossRef]
100. Hanifi, S.; Cammarono, A.; Zare-Behtash, H. Advanced Hyperparameter Optimization of Deep Learning Models for Wind Power Prediction. *Renew. Energy* **2024**, *221*, 119700. [CrossRef]

101. The Structure of Scientific Collaboration Networks. Available online: https://www.pnas.org/doi/10.1073/pnas.98.2.404 (accessed on 25 December 2025).
102. Wu, L.; Wang, D.; Evans, J.A. Large Teams Develop and Small Teams Disrupt Science and Technology. *Nature* **2019**, *566*, 378–382. [CrossRef] [PubMed]
103. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. Available online: https://www.un.org/ohrlls/sites/www.un.org.ohrlls/files/2030_agenda_for_sustainable_development_web.pdf (accessed on 13 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.