

Artificial Intelligence in BIM Clash Management: Assessment of Current Techniques, Automation Levels, and Deployment Readiness

Behzad Abbasnejad ^{1,*}, Alireza Ahankoob ^{1,*}, Guillermo Aranda-Mena ¹ and Amirhossein Karamoozian ²

¹ School of Property, Construction and Project Management, RMIT University, La Trobe Street, Melbourne, VIC 3000, Australia; guillermo.aranda-mena@rmit.edu.au

² Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen 518052, China; karamoozian@szu.edu.cn

* Correspondence: behzad.abbasnejad@rmit.edu.au (B.A.); alireza.ahankoob@rmit.edu.au (A.A.)

Abstract

Automated clash detection in Building Information Modelling (BIM) has reached operational maturity; however, subsequent stages of clash management, including filtering, prioritisation, resolution, and prevention, remain largely dependent on manual expert judgement. Artificial intelligence (AI) has been increasingly applied to address these limitations, yet focused synthesis remains limited on how AI techniques support different stages of BIM-based clash management, what levels of automation they achieve, and how ready they are for professional deployment. This review follows PRISMA 2020 and synthesises 21 empirical studies published between 2019 and 2026, providing one of the first focused syntheses of AI applications across the BIM-based clash management lifecycle. Five families of AI techniques are identified, including rule-based systems, supervised machine learning, deep learning, graph neural networks, and reinforcement and generative AI. The results show that research is concentrated on filtering, while resolution and prevention remain underexplored despite their greater coordination significance. A structured automation taxonomy is developed, revealing that systems described as autonomous are evaluated only in simulated environments. The synthesis further indicates that progress is constrained primarily by structural rather than algorithmic limitations, as most studies rely on single-project validation, report performance without class-level metrics, and produce black-box outputs that cannot meet the accountability requirements of professional coordination practice. The primary barrier to deployment is therefore not model capability but the absence of shared benchmark datasets, rigorous cross-project validation, interpretable outputs, and sociotechnical integration with professional practice.

Keywords: artificial intelligence; automation level; building information modelling; clash management; deployment readiness; machine learning; systematic literature review



Academic Editor: Ahmed Senouci

Received: 4 June 2026

Revised: 25 June 2026

Accepted: 26 June 2026

Published: 1 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Construction projects require continuous coordination across multiple design disciplines operating within constrained spatial and regulatory conditions. The integration of architectural, structural, and building services models in Building Information Modelling (BIM) environments produces a large volume of geometric conflicts, or clashes, that must be identified and resolved prior to construction. The implications of inadequate clash management are well established in the literature. In complex building types with high

service density, clash volumes can reach tens of thousands per coordination cycle, placing substantial pressure on manual review processes [1]. Unresolved clashes propagate into construction as field conflicts, where resolution incurs substantially higher costs and schedule disruption [2,3]. Rework associated with coordination failures accounts for a significant proportion of project overruns, commonly estimated between four and twelve percent of total cost [4,5].

BIM has enabled automated clash detection through tools such as Autodesk Navisworks and Solibri, and this capability has been widely adopted across different contexts [1,6]. However, detection alone has not resolved the coordination problem. Two persistent limitations constrain its effectiveness. First, existing detection systems generate a disproportionately high volume of irrelevant clashes, with empirical studies showing that the majority do not require any coordination action [7]. This false positive burden increases cognitive load and reduces decision efficiency. Second, the management of relevant clashes across filtering, prioritisation, and resolution remains entirely dependent on human expertise. Decisions regarding clash relevance, escalation priority, element modification, sequencing, and trade-offs across disciplines are complex, contextually embedded, and unsupported by automated decision frameworks in current practice [8,9]. As a result, while detection has been automated, filtering, prioritisation, and resolution remain labour intensive, inconsistent, and dependent on coordinator availability and experience.

Recent advances in artificial intelligence (AI) and machine learning (ML) have prompted a growing body of research seeking to support or automate key stages of BIM-based clash management, including relevance classification, prioritisation, resolution recommendation, and prevention. Despite increasing technical sophistication, focused synthesis remains limited on how these techniques support different stages of the clash management lifecycle, what levels of automation they achieve, and how ready they are for deployment in professional coordination practice. Existing reviews of AI in BIM [10,11] treat clash management as one application among many; no prior review has examined it as a primary focus. To address this gap, this systematic literature review (SLR) synthesises empirical studies published between 2019 and 2026. To the best of the authors' knowledge, it is among the first focused reviews of AI applications in BIM-based clash management. The review examines the techniques employed, the degree of automation achieved, the validation conditions reported, and the requirements for operational deployment across the clash management lifecycle.

Three research questions guide the review. RQ1: What AI techniques have been applied to BIM-based clash management, across which lifecycle stages, and with what reported performance? RQ2: What methodological characteristics, dataset properties, validation approaches, and automation levels are reported in existing studies? RQ3: What evidence is available on the deployment readiness of AI-enabled clash management systems, and what technical, methodological, and organisational conditions are required for operational use?

This review makes three contributions. First, it provides one of the earliest focused syntheses of AI applications across the full clash management lifecycle in BIM. Second, it introduces a four-tier automation taxonomy, grounded in human factors theory, to classify AI systems based on decision authority, action initiation, and human oversight. Third, it reframes deployment readiness as a sociotechnical and evidential problem rather than a purely algorithmic one, showing that current barriers arise from limited cross-context validation, opaque outputs, weak integration with professional coordination workflows, poorly specified mechanisms of oversight and accountability, and insufficient causal reasoning in prevention-focused studies.

The remainder of the paper is organised as follows. Section 2 provides background on AI techniques applicable to clash management. Section 3 describes the systematic review methodology. Section 4 presents the results of the bibliometric and thematic analysis. Section 5 discusses the findings and draws theoretical and practical implications. Section 6 concludes the paper and integrates the review's limitations and future research directions.

2. Background

AI Techniques Applicable to Clash Management: Capabilities, Data Requirements, and Limitations

Several families of AI and machine learning techniques have been explored for clash management in BIM, each differing in capability, data requirements, and limitations in supporting generalisable decision-making. Supervised machine learning methods, including decision tree ensembles, probabilistic classifiers, and shallow neural networks, are the most widely applied for clash filtering and relevance classification. These models learn from labelled historical clashes using structured attributes extracted from BIM clash reports, such as element type, discipline pair, penetration geometry, spatial coordinates, and system priority [8,12]. They typically require datasets ranging from several hundred to several thousand labelled instances and are effective in identifying patterns in clash occurrence and prioritisation. However, their explanatory capacity is limited to statistical relationships within the training data, and their generalisation is highly sensitive to dataset diversity. Models trained on single-project datasets often capture project-specific modelling conventions, resulting in reduced performance when applied to other contexts [13]. This limitation is further affected by class imbalance, as most BIM-detected clashes are irrelevant, requiring resampling or cost-sensitive approaches to maintain classification reliability [14,15].

Deep learning and computer vision approaches extend these capabilities by enabling analysis without structured BIM attributes. Convolutional neural networks (CNNs) and object detection models such as YOLO interpret clash report screenshots as image inputs, allowing platform-independent processing [16,17]. These methods require substantially larger datasets, often tens of thousands of labelled images, and their representations may omit precise geometric and semantic information necessary for detailed coordination assessment. Graph-based approaches, particularly graph neural networks (GNNs), address relational complexity by representing BIM models as networks of elements and their spatial or functional connections, supporting reasoning about interdependencies and coordination sequences [18,19]. These models require structured graph representations and sufficient relational data, and their performance depends on the consistency of modelling practices across datasets.

Reinforcement learning and generative AI methods are primarily applied to clash resolution, where design modifications are learned through simulated environments guided by constraint-based reward functions [20–22]. These approaches require extensive computational resources and a large number of simulation iterations to achieve stable outcomes. In contrast, multi-criteria decision analysis (MCDA) hybrids support clash prioritisation by combining expert-defined weighting schemes with BIM-based automation, producing interpretable ranking outputs without requiring labelled datasets [23]. However, their reliance on predefined criteria limits scalability and transferability across projects.

Across these technique families, the literature consistently indicates that model generalisation is more sensitive to dataset diversity than to dataset size, because clash patterns are strongly shaped by project-specific design conventions, regulatory conditions, and coordination practices [9,14,24]. A model trained within one coordination context learns the labelling conventions of that context as much as it learns about clashes in any transferable sense, and performance degradation across project types is therefore not a feature of any individual algorithm but a structural property of how training data are produced and used.

A second cross-cutting limitation concerns interpretability: with the exception of rule-based and shallow supervised methods, none of the technique families described above produce outputs that practitioners can interrogate, audit, or contest, a condition that matters more as automation levels increase and AI outputs move closer to consequential coordination decisions. These two properties, contextual dependency of training data and the absence of interpretable outputs, are not incidental weaknesses of particular studies. They define the evidential terrain on which the review's corpus sits, and they structure the analysis that follows.

3. Methodology

This review was conducted as a systematic literature review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [25] and drawing on the evidence-informed methodology of Tranfield et al. [26]. The protocol was designed to ensure reproducibility, transparency, and systematic coverage of empirical evidence on AI applications across the BIM clash management lifecycle. Five sequential phases were applied: scope, database search, screening and eligibility assessment, data extraction and quality assessment, and thematic synthesis.

3.1. Scope

The review was scoped to empirical studies applying AI and machine learning-based computational techniques to BIM-based clash management. Clash management was defined broadly to include filtering and classification, prioritisation, resolution, correction, dependency modelling, prediction, and prevention in building and civil infrastructure projects. Studies addressing AI in BIM workflows broadly, but without specific engagement with clash management, were excluded.

The research questions were operationalised using the PICO framework adapted for engineering research [27]. In this review, the population refers to BIM-based design and model coordination workflows. The intervention concerns the application of AI and machine learning techniques to clash management. The comparator represents conventional manual or software-assisted coordination practices. The outcomes include reported measures of classification accuracy, prediction performance, filtering effectiveness, resolution support, coordination efficiency, validation quality, and the level of automation achieved.

3.2. Database Selection and Search Strategy

The literature search was conducted in Scopus in May 2026. The search combined three concept blocks: BIM, clash management, and AI/machine learning techniques. The keywords and Boolean structure used for the search are presented in Table 1. The search was limited to publications from 2019 to May 2026, written in English, and published as peer-reviewed journal articles or indexed conference proceedings.

Scopus was selected as the sole database because comparative bibliometric studies identify it as a broad, multidisciplinary index with strong coverage of peer-reviewed journals and conference literature, including recent publications in applied engineering and built environment research [28]. This coverage is appropriate for AI-enabled BIM clash management, which is an emerging and rapidly developing field spanning construction informatics, engineering, computer science, and digital construction. The use of Scopus also supports a transparent and reproducible search strategy by defining a clearly bounded indexed evidence base for systematic analysis.

Table 1. Search string structure used in Scopus.

Search Block	Concept	Search Terms
Block 1	BIM context	"building information model*" OR BIM
Block 2	Clash-management terms	"clash detection" OR "clash management" OR "clash resolution" OR "clash relevance" OR "clash filtering" OR "clash classification" OR "clash correction" OR "clash correction sequence" OR "clash dependency network" OR "component change list" OR "design clash*" OR "design conflict*" OR "spatial conflict*" OR "interference detection" OR "model coordination" OR "design coordination" OR "resolution option*" OR "clash triage"
Block 3	AI and computational techniques	"artificial intelligence" OR AI OR "machine learning" OR "deep learning" OR "neural network*" OR ANN OR "artificial neural network" OR "support vector machine" OR SVM OR "random forest" OR "decision tree" OR XGBoost OR "gradient boosting" OR "naive Bayes" OR "Bayesian network" OR "reinforcement learning" OR "Q-learning" OR "multi-agent reinforcement learning" OR MARL OR "graph neural network" OR GNN OR CNN OR "convolutional neural network" OR RNN OR LSTM OR "recurrent neural network" OR "generative adversarial network" OR GAN OR cGAN OR "deep Q-network" OR DQN OR YOLO OR "computer vision" OR "object detection" OR SHAP OR "explainable AI" OR XAI OR "rule-based reasoning" OR "knowledge-based" OR "knowledge representation" OR "graph theory" OR "dependency network" OR "simulated annealing" OR "linear programming"
Boolean structure	Combined query	TITLE-ABS-KEY (Block 1 AND Block 2 AND Block 3)
Limits	Search filters	Publication period: 2019–May 2026; language: English; source types: peer-reviewed journal articles and indexed conference proceedings

3.3. Inclusion and Exclusion Criteria

Explicit inclusion and exclusion criteria were established prior to screening. Table 2 presents the full criteria applied to this systematic review.

Table 2. Inclusion and exclusion criteria applied at each screening stage.

Criterion	Inclusion	Exclusion
Topic focus	AI/ML applied to BIM-based clash detection, filtering, classification, prioritisation, resolution, prediction, or prevention	BIM or AI studies without direct engagement with BIM-based clash management
AI/ML requirement	At least one AI/ML technique/method empirically applied and evaluated	Conventional BIM clash detection only; conceptual AI proposals without implementation or testing
BIM integration	Uses BIM models, clash reports, BIM software outputs, IFC, Navisworks, Revit, or equivalent BIM-based data	Studies based only on 2D CAD, non-BIM spatial data, or general project data
Empirical content	Reports empirical application, simulation, case study, model validation, expert validation, or practitioner evaluation	Reviews, conceptual papers, opinion papers, editorials, or descriptive studies with no empirical evaluation
Publication type	Peer-reviewed journal articles and indexed conference proceedings	Grey literature, industry reports, theses, dissertations, non-indexed abstracts, editorials, and non-research publications
Language	English-language publications	Non-English publications
Full text	Full text accessible for data extraction	Abstract-only records or unavailable full text

3.4. Screening Process and PRISMA Flow

The Scopus search returned 56 records after applying the publication-year and language filters. No duplicate records were identified. Screening followed PRISMA 2020 in two stages. In Stage 1, titles and abstracts were screened against the inclusion and exclusion criteria. Twenty-five records were excluded because they were review or conceptual papers, addressed general BIM/AI, digital twin, or Industry 4.0 topics outside clash management, focused on education, cost, safety, evacuation, or project management, or were outside the scope of BIM design/model clash management.

In Stage 2, 31 full-text reports were assessed for eligibility. Ten reports were excluded at full-text assessment because they were outside the BIM design/model clash-management scope, lacked empirical AI or computational clash-management evaluation, focused on broader BIM coordination or AI interaction rather than clash-management evaluation, or duplicated a more complete journal article. The final corpus included 21 studies for qualitative synthesis. Figure 1 presents the PRISMA 2020 flow diagram.

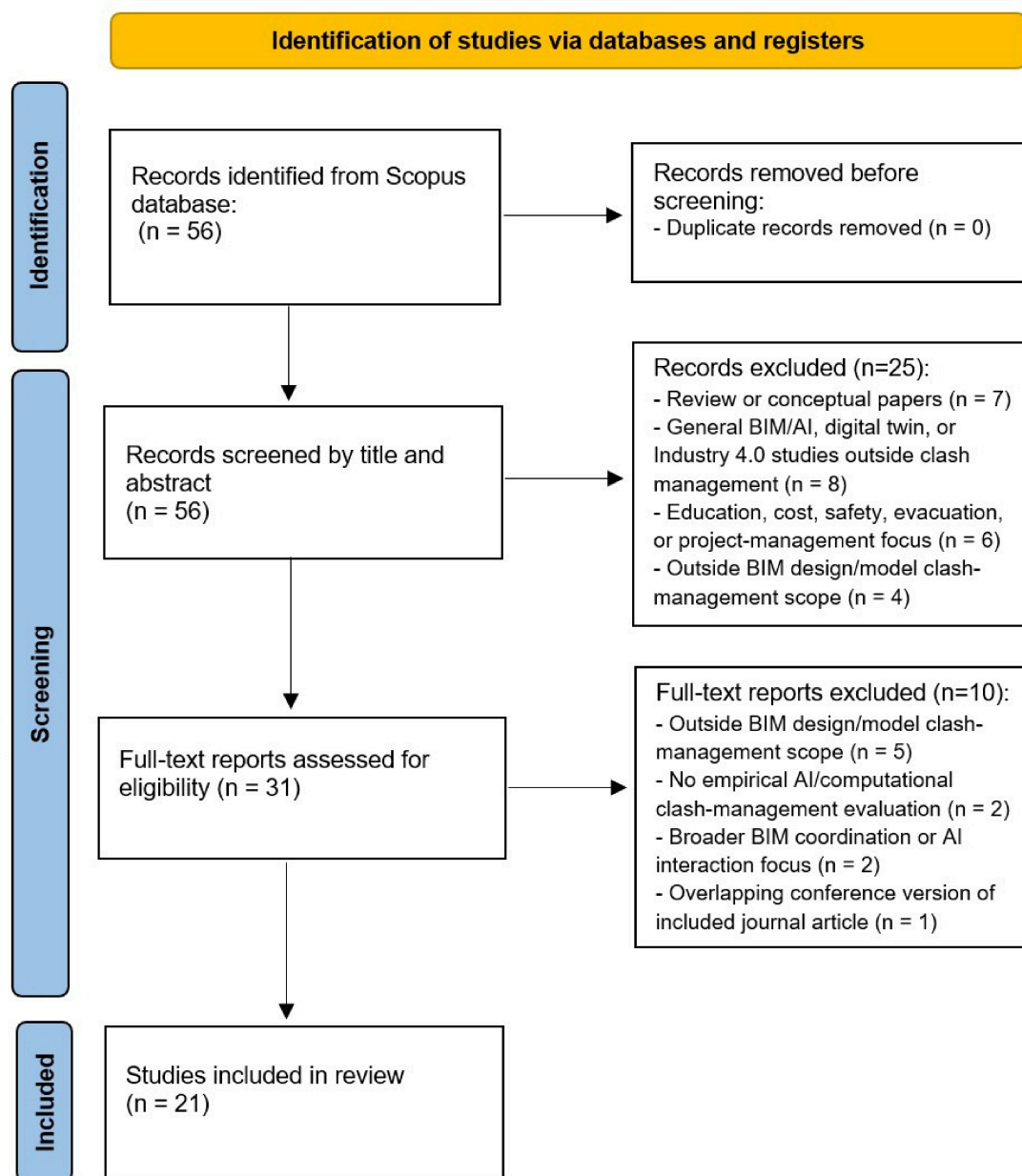


Figure 1. Study selection process following PRISMA 2020 guidelines.

3.5. Data Extraction

A structured coding matrix was applied across all 21 studies. Nine fields were extracted: AI technique, pipeline stage, application domain, input data type, dataset characteristics, performance metrics, validation approach, class imbalance treatment, and automation level with human oversight, based on the four-tier taxonomy. Data extraction was completed by the primary researcher and verified by the second author through a second-pass review of the source documents. Any differences were resolved through discussion until agreement was reached.

3.6. Quality Assessment

Methodological quality was assessed using a framework adapted from the Mixed Methods Appraisal Tool (MMAT) [29], tailored to AI and ML research in construction informatics. Five dimensions were examined: (1) clarity of research design; (2) adequacy of datasets; (3) rigour of validation procedures; (4) completeness of performance reporting; and (5) discussion of generalisability. Each dimension was assessed using three descriptive categories: adequate, partial, or limited. These categories were used to characterise the strength and transparency of the evidence, rather than to classify studies as acceptable or unacceptable. “Adequate” indicates that the dimension was clearly reported and well supported by the study design, data, or validation approach. “Partial” indicates that the dimension was addressed and provides useful evidence, although its support is more context-specific or narrower in scope. “Limited” indicates that the dimension was reported only briefly or provided restricted evidence for interpretation in that specific area. The first two authors independently reviewed the included studies, and any differences in interpretation were resolved through discussion and consensus. Quality appraisal was not applied as an exclusion criterion but was used to inform the interpretation of methodological patterns and evidence characteristics across the corpus. The study-level appraisal results are reported in Supplementary Tables S1 and S2.

3.7. Bibliometric Analysis

In addition to qualitative thematic synthesis, a bibliometric keyword co-occurrence analysis was performed using VOSviewer (Version 1.6.20) on the 21 included studies. Author keywords were extracted from each study and entered into VOSviewer to generate network and temporal overlay visualisations. The network visualisation maps co-occurrence clusters among keywords, identifying thematically coherent research groupings. The temporal overlay illustrates the average publication year of each keyword, highlighting the chronological trajectory of methodological development and the emergence of distinct research paradigms across the corpus. Results are presented in Section 4 alongside the thematic findings.

3.8. Synthesis Approach

Given the heterogeneity of AI techniques, problem framing, performance metrics, and dataset characteristics across the 21 studies, quantitative meta-analysis was not feasible. The synthesis approach is qualitative thematic analysis following the three-stage iterative process of Braun and Clarke [30]: open coding of each study’s contributions, methods, and limitations; axial coding to identify cross-study relationships and patterns; and selective coding to define the final thematic structure. Seven themes emerged from this process, including the clash management pipeline, AI technique taxonomy, input data modalities, the irrelevant clash problem, application domains, dataset characteristics and validation approach, and automation level. A research gap analysis and deployment readiness assessment conclude the findings.

4. Results

4.1. Bibliometric Analysis

4.1.1. Co-Occurrence Network Analysis

Keyword co-occurrence analysis was performed in VOSviewer on the 21 included studies. The network visualisation (Figure 2) reveals four thematically coherent clusters. The first, centred on BIM and clash detection, connects to deep learning, Navisworks, and GNN, reflecting the centrality of detection-oriented research and the growing presence of graph-based approaches. The second cluster, centred on machine learning, brings together the classical supervised methods used for clash filtering, including random forest, naïve Bayes, gradient boosting, and artificial neural networks (ANNs), alongside studies on clash relevance. This demonstrates that relevance classification constitutes a methodologically coherent subfield. The strongest co-occurrence link in the map, between BIM and machine learning, indicates that machine learning has become as definitionally central to this literature as clash detection itself. The third cluster bridges the detection and resolution sides of the map through mechanical, electrical, and plumbing (MEP) and rule-based reasoning, while the fourth groups clash resolution, path planning, rebar design, and GAN as a peripheral resolution sub-domain. The weak connectivity between the machine learning and resolution clusters provides bibliometric confirmation that these two research threads have developed largely in parallel.

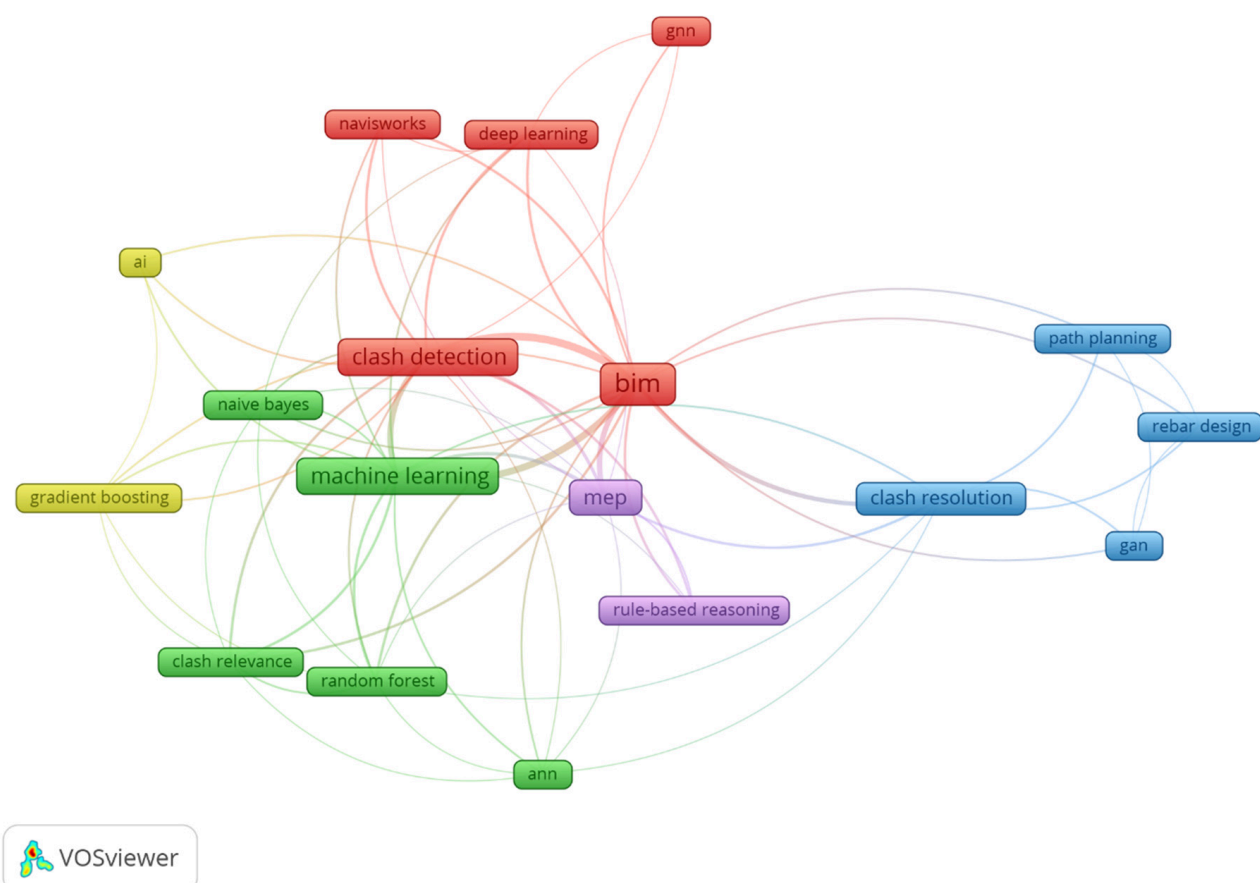


Figure 2. Keyword co-occurrence network.

4.1.2. Research Paradigm

The temporal overlay (Figure 3) maps the average publication year per keyword, revealing a coherent methodological progression across three phases. In the earliest phase (approximately 2022–2023), keywords such as rule-based reasoning, path planning, and

rebar design reflect a dominant focus on deterministic and logic-based coordination approaches. The intermediate period (2023–2024) shows a shift toward supervised machine learning, with keywords including machine learning, random forest, and naïve Bayes clustering here, corresponding to the relevance classification literature. The most recent phase (2024–2026) is marked by deep learning, gradient boosting, GNN, and GAN keywords, reflecting the move toward relational and generative architectures. The positioning of GNNs as a late-emerging cluster within the bibliometric structure suggests that graph-based approaches mark the most recent phase of methodological evolution in the field, with adoption and system-level integration lagging behind their technical development.

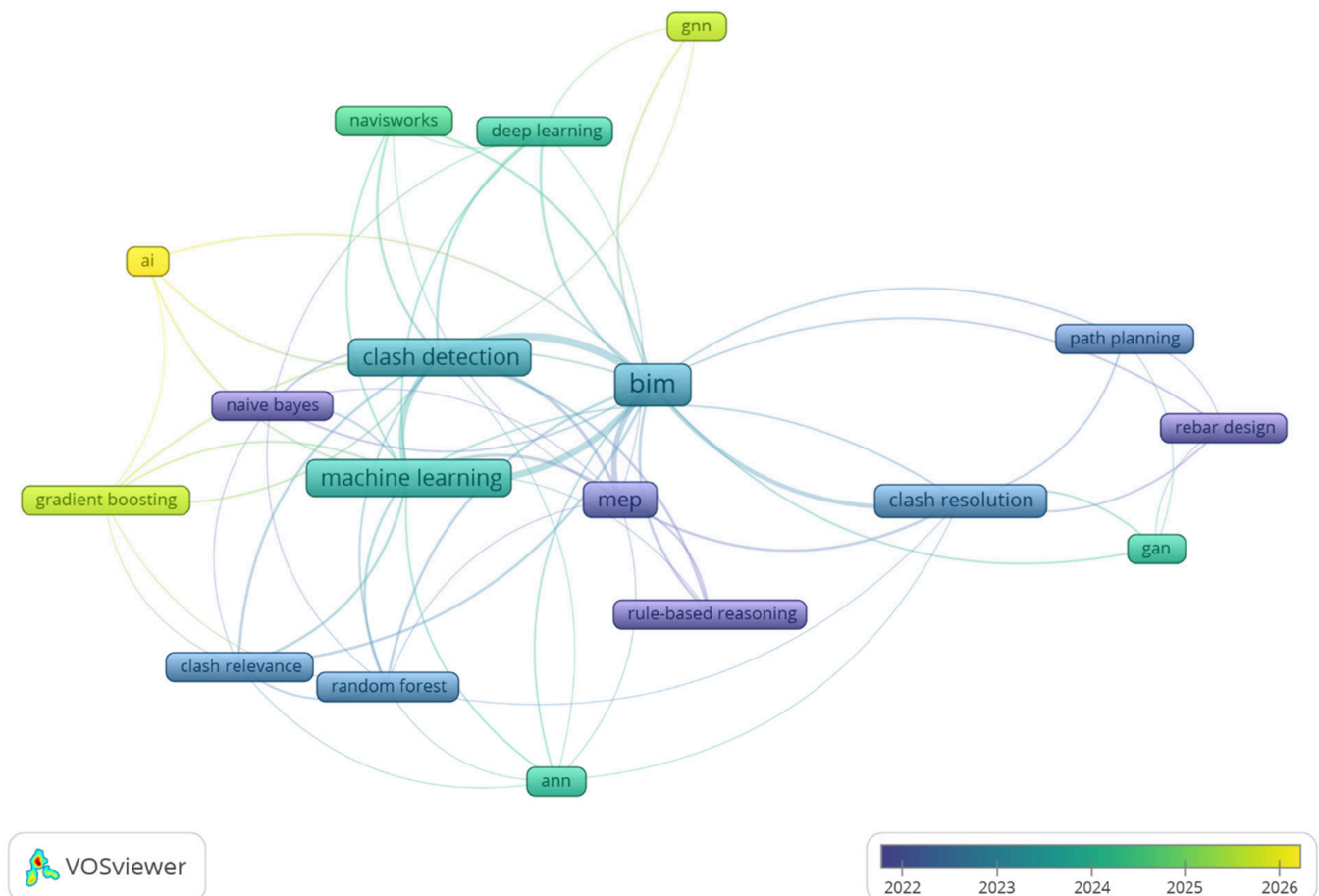


Figure 3. Temporal overlay of keyword co-occurrence.

4.2. Thematic Analysis

The final 21 studies were subjected to qualitative thematic analysis conducted in NVivo, following the three-stage iterative process of Braun and Clarke [30]. Open coding of each study's contributions, methods, limitations, and reported outcomes produced an initial code set, which was progressively consolidated through axial and selective coding into a final thematic structure. Seven themes emerged from this process: the clash management pipeline, AI technique taxonomy, input data modalities, the irrelevant clash problem, application domains, dataset characteristics and validation rigour, and automation level. Table 3 provides the full corpus overview across these dimensions.

Table 3. Corpus overview of reviewed studies (n = 21).

Reference	AI Methods	Scope	Domain	Dataset	Evaluation/Validation Approach
Hu & Castro-Lacouture [8]	Jrip, RF, J48, NB, LR, BN	Filtering	MEP-Structural	204 clashes (1 project)	3:1 train-test split; 10-fold CV; 100 repeated samples; Kruskal-Wallis test
Huang & Lin [31]	Rule-based; DT, k-NN, SVM; ensemble	Classification	MEP-Structural	326 clashes (single shopping mall, Taiwan)	70/30 split; 5-fold CV; testing on 98 cases using confusion matrix, precision, recall, and F1-score
Lin & Huang [32]	Hybrid rule-based + ML (DT, SVM, k-NN; ensemble)	Filtering	MEP-Structural	326 clashes (single shopping mall, Taiwan)	70/30 split; 5-fold CV; 30 repeated tests; precision, recall, and F1-score
Hsu et al. [33]	Backpropagation Neural Network (BPNN) + Simulated Annealing	Resolution	MEP-Structural	133 clashes from a student residence project	10 experimental runs; all runs reached zero clashes; average runtime 49.61 min
Hu et al. [19]	Graph theory + combinatorial optimisation (Greedy, Linear Programming (LP), local search)	Resolution	MEP	191 clashes (5-storey residence)	Algorithm performance testing
Liu et al. [21]	Multi-Agent Reinforcement Learning (MARL)	Resolution	RC Rebar	Simulated one-storey RC frame	40 simulation runs; success-rate analysis and BIM clash-checking of generated rebar layout
Luo et al. [15]	Rule-based (ER + GR rules)	Detection & Filtering	Underground Pipelines	~11,190 clashes (1 pipeline project)	Contribution rate evaluation
Ahmadpanah et al. [16]	YOLOv5 object detection	Filtering	Structural vs. MEP	563 clash report images from multiple projects	F1-confidence curve + confusion matrix
Liu et al. [22]	pix2pix conditional GAN + Deep Q-Network	Resolution	Structural (rebar, precast)	100 CAD drawings from practical precast concrete wall panel projects; regular and irregular panel cases	Comparison with manual design time; BIM-based clash-free rebar generation; reported time savings up to 80%
Bitaraf et al. [23]	Knowledge-based; Best-Worst Method (BWM); Navisworks API	Prioritisation	MEP-Structural	1 demo project	Comparative cost/time analysis; plugin-based demonstration in Navisworks; prioritisation using CI formula;
Harode et al. [12]	Modified Delphi (feature/label validation)	Resolution (feature design)	MEP (duct-pipe)	23 factors, 5 options; 15 interviews	2-round expert consensus
Harode et al. [9]	ANN (best); kNN, DT, ET, RF; MLSOL oversampling	Resolution	MEP (duct-pipe)	506 clashes; 6 BIM models; 2 facility types	70/30 split; 5-fold CV; GridSearchCV

Table 3. Cont.

Reference	AI Methods	Scope	Domain	Dataset	Evaluation/Validation Approach
Koo & Guerra [34]	ANN (multilayer perceptron)	Filtering	Multi-discipline	600 clashes (1 building)	Train/test split (70.4%/29.6%)
Koo et al. [35]	ANN, kNN, SVM, RF, GB	Filtering + Risk Prediction	Multi-discipline	1000 clashes from 2 real-world building projects	70/30 train-test split; 5-fold CV; expert validation of relevance and risk labels
Shehadeh & Alshboul [36]	CNN, GNN, RNN-LSTM; VR interface; SHAP attribution	Prevention and Prediction	Multi-discipline	Not reported; 28 projects	Area Under the Receiver Operating Characteristic Curve (AUC-ROC): CNN = 0.83; GNN = 1.00; LSTM = 1.00
Chong et al. [37]	Fine-tuned cGAN + A* algorithm	Resolution	Generic BIM path planning	2000 synthetic images; 5 test scenarios	Pixel-similarity vs. expert ideal
Oduro et al. [38]	RF, ANN, KNN, SVM, LR (relevance); GB, DT, AdaBoost (risk)	Filtering + Risk Prediction	Multi-discipline	1000 clashes; 2 buildings	70/30 split; AUC-ROC; Friedman test
Alshboul & Shehadeh [13]	XGBoost (adaptive, real-time)	Predictive clash detection	Multi-discipline	30 BIM projects; 5 element-pair scenarios	Performance comparison before and after adaptive XGBoost integration; accuracy 0.92, recall 0.93
Ailem & Boton [39]	Ensemble RF + XGBoost + MLP; SMOTE; Navisworks plug-in	Filtering	Multi-discipline	36,562 conflicts; 2 public projects	Technical testing in Navisworks, practitioner assessment, and criteria-based evaluation; F1-score = 0.95, recall = 0.96.
Kumar et al. [40]	BPNN with bagging ensemble and voting classifier	Resolution	MEP	450 expert-derived clash instances; single study	Per-attribute accuracy reporting; overall accuracy 93.57%; no cross-validation reported
Chen [41]	GNN with SMEN and TASCE	Detection and Risk Prediction	Multi-discipline	4 benchmark datasets	Benchmark testing

4.2.1. The Clash Management Pipeline

Analysis of the corpus reveals four discrete and sequential stages at which AI has been applied within the clash management lifecycle: filtering and classification, prioritisation, resolution, and prevention and prediction. These stages form a logical pipeline from raw clash detection output to proactive design intervention; however, few studies address more than one stage, and none develop an integrated end-to-end clash management system spanning filtering, prioritisation, resolution, and prevention. This indicates that the field has evolved through isolated task-specific interventions rather than a fully integrated coordination framework.

Filtering and classification constitute the dominant focus of the corpus, with ten studies seeking to identify, classify, or remove irrelevant clashes from raw BIM detection reports. This group includes Hu and Castro-Lacouture [8], Huang and Lin [31], Lin and

Huang [32], Luo et al. [15], Ahmadpanah et al. [16], Koo and Guerra [34], Koo et al. [35], Oduro et al. [38], Ailem and Boton [39], and Chen [41]. Resolution is addressed by eight studies, including Hsu et al. [33], Hu et al. [19], Liu et al. [21], Liu et al. [22], Harode et al. [9], Harode et al. [12], Chong et al. [37], and Kumar et al. [40], and refers to the identification or execution of corrective design actions. Prioritisation appears as the primary focus in only one study, Bitaraf et al. [23]. Prediction and prevention are similarly limited, with Shehadeh and Alshboul [36] distinctive in framing clash prediction as a temporal design evolution problem, rather than as a post-detection filtering or management task. Alshboul and Shehadeh [13] combine prediction and resolution and are therefore counted within the resolution group. Koo et al. [35] is positioned between filtering and prioritisation because it first distinguishes true and false positive clashes and then predicts the risk level of true positive clashes.

This distribution shows that the literature has concentrated on the most immediately tractable task, reducing information overload after clash detection, while giving less attention to the more consequential coordination functions of resolution, prioritisation, and prevention. The field has therefore made greater progress in deciding which clashes matter than in determining how clashes should be resolved or prevented before they emerge.

4.2.2. AI Techniques: Taxonomy, Application and Performance

The distribution of AI techniques across clash management stages reveals how research effort is allocated within the field. Figure 4 illustrates the mapping of technique families to different stages of the BIM clash management process. Classical supervised machine learning dominates the corpus, with strong contributions to filtering [8,13,32,34,35,38,39] and additional applications in resolution [9,33], classification [31,35], and prioritisation [23]. Rule-based and knowledge-based approaches are primarily applied to filtering and resolution [15,31–33], with limited extension to detection tasks through the rule contribution rate framework of Luo et al. [15]. Deep learning and computer vision methods contribute mainly to filtering [16,35], classification [35], and resolution [22,37], reflecting their application across both structured and image-based data contexts. Graph-based and relational methods are concentrated in resolution and detection [19,21,22,41], indicating their role in modelling element topology, inter-clash dependencies, and the relational structure of BIM coordination environments that tabular and image-based approaches cannot represent. The following subsections analyse the AI techniques employed in the literature, with attention to study scope, reported performance, and methodological contributions across different stages of the clash management pipeline.

Rule-Based and Knowledge-Based Systems

Rule-based and knowledge-based systems encode domain expertise explicitly rather than learning it from labelled data, representing the most interpretable but least scalable AI paradigm in the corpus. This category covers four studies spanning filtering and resolution stages, and their collective contribution lies in demonstrating that codified professional knowledge, when structured as logical conditions, can discriminate valid from invalid clashes with a precision that purely statistical approaches cannot replicate on small or domain-specific datasets.

In the filtering domain, Luo et al. [15] apply two rule tiers to urban underground pipeline networks: Elementary Rules targeting geometrically obvious invalid clashes, and Geometric Rules enforcing pipeline-type-specific tolerance thresholds. A contribution rate formula measures each rule's discriminative value objectively, enabling systematic rule pruning. The result exceeds 95 percent of raw clashes removed as invalid, the highest filtering rate in the corpus, and illustrates the power of standards-grounded logic when

the design code environment is stable and well-defined. The fundamental constraint is transferability: rules must be re-engineered for each code jurisdiction, building typology, and design convention, creating a maintenance burden that scales poorly across diverse project portfolios.

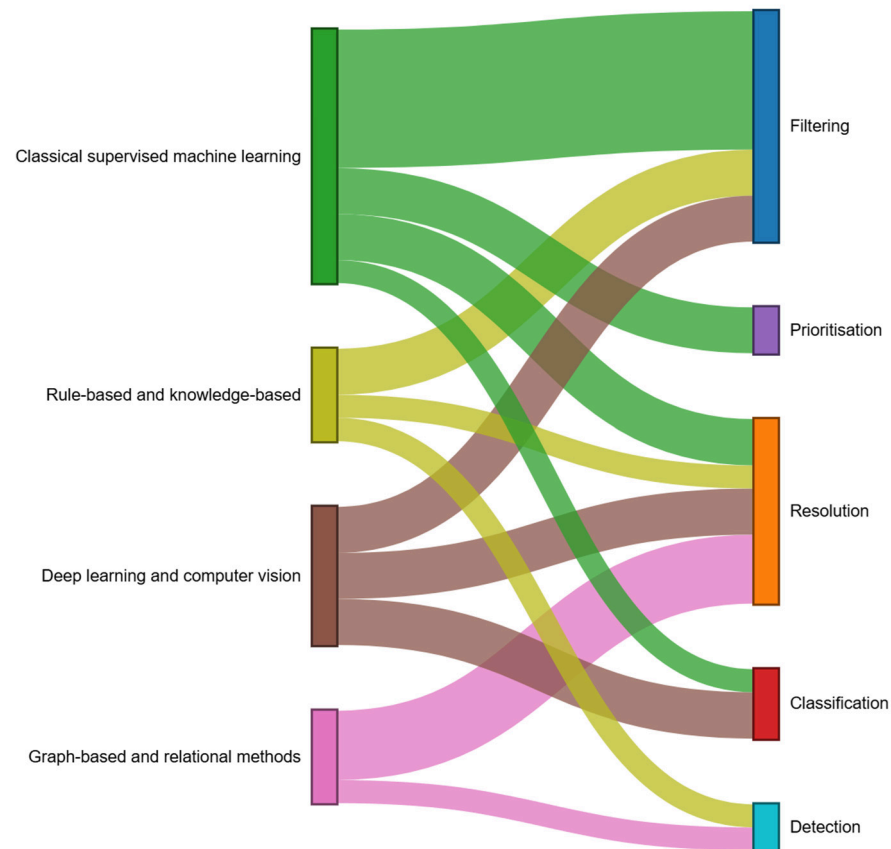


Figure 4. Sankey diagram mapping AI technique categories to clash management stages across the reviewed corpus (note: studies addressing multiple stages are counted in each applicable flow).

Hybrid rule-ML architectures in Huang and Lin [31] and Lin and Huang [32] address this limitation by using expert rules as a pre-classifier that separates obviously permissible clashes before ML training, reducing the class imbalance that degrades ML performance on small datasets. The rules absorb straightforward cases; the ML model learns the ambiguous ones. Lin and Huang [32] report an average filtering performance of 0.96, among the highest in the corpus, though the single-project Taiwanese shopping mall dataset limits independent verification of this figure across broader coordination contexts. The hybrid architecture is conceptually significant because it demonstrates that rule knowledge and statistical learning are complementary rather than competing; each addresses the failure mode of the other.

In the resolution domain, Hsu et al. [33] developed a knowledge-based resolution system combining parallel BPNN learning with simulated annealing and Revit API-based model revision. Constructor knowledge was elicited through questionnaires and used to generate resolution guidelines, while simulated annealing searched for clash-reducing revisions to MEP systems and elements. The system was evaluated on 133 design clashes in the basement MEP systems of a student residence, and all ten experimental runs reached zero clashes. However, the validation remains limited to a single case and does not demonstrate deployment in live multidisciplinary coordination workflows.

Classical Supervised Machine Learning

Classical supervised ML is the dominant AI paradigm in the corpus, spanning nine studies across filtering, resolution, and risk prediction tasks. Its prevalence reflects the strong fit between BIM clash reports and tabular classification methods, as clash data are commonly structured as numerical and categorical attributes. Two main applications emerge: relevance classification for filtering and action prediction for resolution.

In filtering studies, Hu and Castro-Lacouture [8] established the basic feature engineering template by showing that structured clash attributes can support relevance classification, with Jrip producing auditable decision rules from 204 MEP structural clashes. Subsequent studies refined this approach. Koo and Guerra [34] extend relevance prediction beyond MEP-focused datasets by applying an ANN multilayer perceptron model to architectural, structural, and MEP clashes. Their results show that combinations of discipline, element type, element detail, and volume-related attributes can support clash relevance prediction, achieving over 80% testing accuracy. Oduro et al. [38] extended filtering to joint relevance and risk prediction across 1000 labelled clashes, showing that Random Forest performs better for binary relevance classification, while Gradient Boosting is stronger for multi-class risk prediction. Koo et al. [35] extend relevance filtering toward risk-informed clash management. Using 1000 clashes from two real-world projects, the study predicts both clash relevance and risk level, showing how ML can support the removal of false positives and the prioritisation of medium- and high-risk clashes.

The strongest movement toward deployment is found in Ailem and Boton [39], who developed a Navisworks plug-in combining structured metadata extraction, supervised classification and user-facing filtering. Trained on 36,562 conflicts from two public building projects, the tool was evaluated through technical testing, practitioner assessment and criteria-based evaluation, and was explicitly aligned with ISO 19650 information-management principles. Alshboul and Shehadeh [13] further advanced deployment readiness through an XGBoost model connected to the Navisworks API and tested across 30 projects and five element pair scenarios, providing the broadest multi-project validation design in the filtering subset.

In resolution studies, Harode et al. [12] and Harode et al. [9] provide the most rigorous supervised ML contribution. Harode et al. [12] first used a two-round modified Delphi study with 15 professionals across five firms to validate 23 clash factors and five resolution action categories. Harode et al. [9] then framed resolution as a multilabel classification task, using 506 hard clashes from six healthcare and academic BIM models. Five algorithms were trained with MLSOL to address class imbalance, with ANN achieving the strongest performance at approximately 80 percent mean accuracy.

Kumar et al. [40] extend resolution recommendation through a BPNN-based bagging ensemble with a voting classifier, trained on 450 expert-derived MEP clash cases. The model predicts eight resolution labels, including the element to revise and feasible directional modifications, and reports 93.57 percent voting classifier accuracy. However, expert knowledge is elicited without structured consensus validation, and implementation remains dependent on an experienced designer.

Deep Learning and Computer Vision

Deep learning and computer vision studies introduce image-based clash representation as an alternative to structured BIM attributes, covering four studies in the corpus. This modality is attractive because it is comparatively schema-independent. Once trained, a model can process visual clash reports generated from different BIM environments without requiring a bespoke attribute extraction pipeline. However, the evidence indicates a clear performance trade-off. Image-based models reduce dependence on structured data

preparation, but they may also discard much of the geometric and semantic information that coordinators use to assess detected clashes.

YOLO-based object detection also shows potential for image-based clash analysis. Ahmadpanah et al. [16], for example, apply YOLOv5 to 563 Navisworks clash report images to recognise selected structural objects, particularly beams and trusses. However, the study remains limited to object recognition as an initial step toward image-based clash filtering, rather than full relevance or risk classification.

Shehadeh and Alshboul [36] provide the most architecturally ambitious deep learning contribution. Their framework combines CNNs for spatial feature extraction, GNNs for modelling element interdependencies, and RNN-LSTM models for capturing temporal design evolution across 28 projects, with deployment through a VR interface integrated with Revit Live. The study reports AUC-ROC values of 1.00 for both the GNN and RNN-LSTM models. Such perfect discrimination is highly unusual in real-world multi-class prediction and should be interpreted cautiously, as it may indicate leakage between training and test partitions or insufficiently independent validation. Nevertheless, the conceptual contribution remains important. The study reframes clash prevention as a temporal prediction problem across design evolution, which is a promising but still underdeveloped direction in the field.

Chong et al. [37] extend image-based approaches into generative clash coordination by fine-tuning a conditional GAN on 2000 synthetic BIM clash images to generate clash-free MEP routing configurations. The model is benchmarked against an A* algorithm with a dynamic cost function across five obstacle-density scenarios. The cGAN performs substantially worse than A* in similarity to the expert-defined optimal solution, producing unclear and impractical routing outputs. This result reflects an architectural mismatch. GANs learn statistical image distributions, whereas clash resolution is a constrained spatial optimisation problem. The study therefore establishes A* as the stronger baseline for constrained MEP pathfinding at the current maturity of generative image-based methods.

Graph-Based and Relational Methods

Graph-based and relational methods address a structural limitation shared by tabular and image-based techniques: the treatment of clashes as isolated events. In practice, building systems are spatially and functionally interdependent, and a correction applied to one element changes the spatial conditions of all clashes involving that element, potentially introducing new conflicts or restricting later adjustments. Four studies in the corpus model this interdependence, forming the most conceptually advanced group in the review.

Hu et al. [19] advance resolution research by treating clash correction as a sequencing problem rather than as a set of isolated clash decisions. The study represents 191 MEP clashes in a five-storey student residence as a clash dependency network, where nodes denote clashes and directed relations capture correction dependencies. Using greedy algorithms, linear programming, and local search, the method identifies a correction sequence that reduces feedback dependencies and groups interrelated clashes. Its contribution is therefore not automated geometric resolution, but structured support for managing ripple effects during human-led clash correction.

Chen [41] advances the relational paradigm substantially by representing the entire BIM model as a heterogeneous attributed graph in which building elements serve as nodes and their spatial and functional relations form typed edges. A dual-module graph neural network comprising a variational multi-semantic encoder (SMEN) and a temporal-aware semantic coordination engine (TASCE) encodes geometric, taxonomic, and compliance features and aligns them with evolving design states across four benchmark datasets. This

is the most expressive model in the corpus and achieves leading performance on conflict detection accuracy and risk prediction.

Reinforcement learning extends the relational paradigm from detection and sequencing to automated resolution generation. Liu et al. [21] deploy a multi-agent reinforcement learning (MARL) system on a simulated one-storey reinforced concrete (RC) frame comprising 63 beams, 23 columns, and 1120 rebars, where each reinforcing bar acts as an autonomous agent navigating toward a clash-free configuration. The success rate is evaluated across 40 runs with visual inspection of agent path convergence. Liu et al. [22] apply a related approach to irregular precast concrete wall panels, coupling a pix2pix conditional GAN that generates feasible rebar layout candidates with a Deep Q-Network (DQN) that refines them toward clash-free configurations. Both studies mark a substantive shift from reactive coordination assistance to proactive design generation and represent the furthest point on the resolution automation spectrum reached by any empirical study in the corpus.

4.2.3. Input Data Modalities

Three primary input data modalities are employed across the 21 studies, each differing in information content, BIM integration requirements, and performance implications. Tabular clash attributes, consisting of structured numerical and categorical data extracted from clash reports, represent the dominant modality, used in 14 studies. Common features include discipline pair, element type, penetration volume and depth, spatial coordinates, floor level, and system priority. Across studies, material type is consistently found to be non-predictive of clash relevance, whereas element type, discipline combination, and penetration geometry emerge as the most discriminative attributes [8,34].

Image-based data, derived from visual screenshots of clash reports in platforms such as Navisworks, are used in three studies [16,35,37]. This modality enables computer vision approaches that are independent of BIM data schemas, supporting portability across software environments. However, the reduced information fidelity of image-based representations remains a limitation, as pixel-based inputs do not retain the full numerical precision, semantic metadata, and contextual attributes available in structured BIM clash reports. Graph and geometric representations are employed in two studies, offering the highest level of expressiveness [19,41]. These approaches capture element interdependencies that other modalities cannot represent but require Industry Foundation Classes (IFCs)-compliant data and graph-based infrastructure, which constrains their applicability in practice.

4.2.4. The Irrelevant Clash Problem

A consistent finding across the corpus is the substantial volume of false positives, low-risk issues, and otherwise low-priority clashes produced by raw BIM detection. Koo et al. [35] show that among 1000 randomly selected clashes from two real-world projects, only approximately half were true positive clashes with medium or high risk that warranted focused design-phase attention. Luo et al. [15] report that more than nineteen-twentieths of clashes in an underground pipeline project are invalid. Lin and Huang [32] describe most software-flagged clashes as pseudo-conflicts that are geometrically identified but structurally acceptable. Ailem and Boton [39] argue that the burden created by irrelevant alerts undermines data quality management required under ISO standards.

Across project types and detection settings, the reviewed studies consistently show that raw clash reports contain a substantial number of clashes that do not require the same level of human coordination. This establishes the main practical motivation for AI in clash management. The core challenge is not sensitivity, which current software already provides at high levels, but specificity, meaning the ability to distinguish clashes

that require human coordination from those that can be filtered, deprioritised, or resolved with minimal intervention. Studies that present only aggregate accuracy without reporting class-level precision, recall, or risk-level performance therefore risk overstating model capability, particularly where false positives or low-priority clashes dominate the dataset.

4.2.5. Application Domains

MEP coordination is the dominant application domain in the corpus, with eleven studies addressing MEP or MEP-structural contexts explicitly, including Hu and Castro-Lacouture [8], Huang and Lin [31], Lin and Huang [32], Hsu et al. [33], Ahmadpanah et al. [16], Bitaraf et al. [23], Harode et al. [9,12], Koo and Guerra [34], and Kumar et al. [40]. A further six studies operate across multi-disciplinary configurations spanning architectural, structural, and MEP systems [13,35,36,39,41]. Underground pipeline infrastructure constitutes the sole non-building application [15]. This concentration in MEP contexts reflects the volume and spatial density of building services elements installed within confined zones, such as suspended ceilings, wall cavities, service shafts, and plant rooms, where multiple discipline systems must coexist within tightly constrained cross-sections, generating the highest volumes of geometric conflicts and the highest false positive rates in automated detection.

A separate sub-domain concerns reinforcement detailing conflicts in reinforced concrete and precast structural elements, addressed by two studies [21,22]. This sub-domain differs from service coordination because reinforcement layout is governed by strict structural design requirements, including cover limits, spacing rules, and bend geometry, and errors carry direct structural safety implications not present in routine service coordination. The problem is formulated as constrained layout generation rather than conflict classification, implying fundamentally different performance criteria, liability frameworks, and adoption pathways from the MEP coordination literature.

4.2.6. Dataset Characteristics and Validation Approach

The datasets underpinning the corpus reveal empirical foundations substantially narrower than reported performance claims imply. Dataset sizes range from 133 clashes in Hsu et al. [33] to 36,562 conflicts in Ailem and Boton [39], with the majority of studies concentrated at the lower end of this range. Three studies did not report clash counts at all: Liu et al. [22], Shehadeh and Alshboul [36], and Chen [41] reported omissions that render independent assessment of their evidential basis impossible. The rebar studies of Liu et al. [21,22] relied entirely on simulated or synthetic environments with no real project coordination data involved.

The most prevalent validation strategy is a train-test split or cross-validation within a single project. This design estimates generalisation to unseen clashes from the same production context; it does not estimate performance on data generated by a different project team, building type, or contractual regime, and the two are not equivalent. Only Harode et al. [9], training across six BIM models spanning two facility types, and Ailem and Boton [39], drawing from two projects, approached cross-context validation, and neither reported performance disaggregated by context. No study conducted a systematic transfer experiment in which training and test contexts were explicitly varied and performance degradation measured as a function of contextual difference. This absence is not a minor limitation; it is the condition that separates proof-of-concept evidence from evidence of generalisable performance.

Class imbalance received explicit methodological treatment in only two studies. Harode et al. [9] applied MLSOL oversampling, and Ailem and Boton [39] applied SMOTE. Given that genuine clashes constitute a structural minority of raw detected conflicts on most

projects, studies reporting aggregate accuracy without class-disaggregated precision and recall cannot rule out inflated metrics driven by majority-class dominance. At the upper end of reported performance, the AUC-ROC values of 1.00 for the GNN and RNN-LSTM models in Shehadeh and Alshboul [36], obtained on an undisclosed dataset, are inconsistent with typical real-world multi-class prediction outcomes and are examined critically in Section 5. The corpus as a whole has produced internally coherent proof-of-concept evidence across a range of tasks and architectures. The cross-context transfer evidence that professional deployment would require remains entirely absent.

4.2.7. Automation Level

A four-tier automation taxonomy was developed for this review to classify the role of AI systems in BIM-based clash management. The taxonomy is informed by two complementary reference points, including the human factors framework of Parasuraman et al. [42], which conceptualises automation as a continuum from full human control to full autonomous execution, and the risk-tier logic of the EU AI Act [43], which foregrounds the governance implications of increasingly consequential AI outputs. Rather than applying the full ten-level scale proposed by Parasuraman et al. [42], this review consolidates it into four functionally distinct levels that better reflect the human-machine oversight configurations observable in the reviewed clash management studies: L1 decision support, L2 recommendation, L3 automated output with human review, and L4 autonomous execution. Table 4 presents the resulting classification of the included studies, including each study's scope, assigned automation level, and basis for assignment.

Table 4. Automation level of AI systems in BIM-based clash management studies.

Study	Scope	Automation Level	Basis for Assignment
Hu & Castro-Lacouture [8]	Filtering	L3—Automated output, human review	ML used to automatically classify clashes as relevant or irrelevant but requires BIM coordinators to review outputs before any clashes are discarded.
Huang & Lin [31]	Filtering/Classification	L3—Automated output, human review	Rule-based reasoning and ML classifiers automate clash classification/filtering, with human inspection still required.
Lin & Huang [32]	Filtering	L3—Automated output, human review	Hybrid rule-based and supervised ML method automatically filters irrelevant clashes, with serious clash misclassification still requiring human oversight.
Hsu et al. [33]	Resolution	L3—Automated output, human review	BPNN + SA generates resolved sequence; engineer sign-off required.
Hu et al. [19]	Resolution	L3—Automated output, human review	Graph algorithm outputs optimal resolution order; no automatic update.
Liu et al. [21]	Resolution	L4—Autonomous execution	MARL agents autonomously resolve rebar clashes in simulation.
Luo et al. [15]	Detection/Filtering	L3—Automated output, human review	The rule engine automatically applies Elementary and Geometric Rules to flag and remove invalid clashes; contribution rate metrics evaluate rule performance at system level, not individual clash decisions, confirming automated output reviewed by human.

Table 4. Cont.

Study	Scope	Automation Level	Basis for Assignment
Ahmadpanah et al. [16]	Filtering	L3—Automated output, human review	YOLOv5 classifies objects in clash report images; coordinator review is required before use.
Liu et al. [22]	Resolution	L4—Autonomous execution	cGAN + DQN autonomously redesigns rebar layout to clash-free state.
Bitaraf et al. [23]	Prioritisation	L1—Decision support	BWM weights and scores clashes; all priority decisions remain with coordinator.
Harode et al. [12]	Resolution (feature design)	L1—Decision support	Delphi expert consensus defines feature/label space; no automated output.
Harode et al. [9]	Resolution	L2—Ranked recommendation	ANN recommends one of five resolution options; human engineer selects and implements.
Koo & Guerra [34]	Filtering	L3—Automated output, human review	ANN automatically scores clash relevance across the full dataset; coordinators review the scored output to prioritise their queue, confirming the system acts first and the human reviews the resulting ranked list.
Koo et al. [35]	Filtering + Risk Prediction	L3—Automated output, human review	ANN, kNN, SVM, RF and GB models predict clash relevance and risk level from BIM clash attributes; outputs support filtering and prioritisation, while review and resolution remain with BIM/VDC experts.
Shehadeh & Alshboul [36]	Prevention/Prediction	L3—Automated output, human review	CNN/GNN/RNN-LSTM predict clash likelihood; VR interface surfaces predictions; human designer acts.
Chong et al. [37]	Resolution	L3—Automated output, human review	cGAN and A* generate clash-free routing paths from synthetic image scenarios; outputs require human validation before implementation.
Oduro et al. [38]	Filtering/Risk	L3—Automated output, human review	Two-stage pipeline automatically flags irrelevant clashes and assigns risk levels across the full dataset; coordinators retain final judgement over clash disposition, confirming human review of automated output.
Alshboul & Shehadeh [13]	Prediction/Resolution	L3—Automated output, human review	XGBoost predicts and advises resolution in real-time; human oversight retained.
Ailem & Boton [39]	Filtering	L3—Automated output, human review	Navisworks plug-in filters automatically; coordinator reviews output before action.
Kumar et al. [40]	Resolution	L2—Ranked recommendation	BPNN ensemble generates resolution guidelines, including the element to revise and feasible modification directions; however, implementation remains with an experienced designer.
Chen [41]	Detection/Risk	L3—Automated output, human review	GNN detects conflicts and predicts risk; output is a prediction score, not an executed action.

Note: A* refers to the A-star pathfinding algorithm.

This consolidation is methodologically justified for two reasons. First, most studies in the corpus do not describe their operational workflows with sufficient detail to support reliable distinctions between adjacent levels in a ten-point automation scale. Second, a four-tier structure provides a clearer basis for linking technical functionality to decision authority, action initiation, human oversight, and governance risk. The taxonomy, therefore, does not rank studies by quality, effectiveness, or desirability. Instead, it classifies where automation stops, where human intervention begins, and how responsibility is distributed between the system and the human user.

At L1, the system provides decision support, such as scoring, feature validation, analytical information, or conceptual guidance, while the human remains responsible for interpretation, judgment, and action. At L2, the system generates recommendations, such as predefined resolution options, but the human selects from these options and initiates the coordination response. At L3, the system produces a more complete automated output, such as a filtered clash list, classified report, resolved sequence, prediction result, or routing proposal, which is then reviewed, accepted, modified, or rejected by a human before downstream use. The boundary between L2 and L3, therefore, lies at the point where automation ends. L2 provides recommended options for humans to select and act upon, whereas L3 produces an output that is ready for workflow use and requires human review to ensure reliability, completeness, and contextual appropriateness. At L4, the system autonomously executes clash resolution or design modification without human involvement in the immediate decision-action loop.

Applying this taxonomy, two studies operate at L1: Bitaraf et al. [23] and Harode et al. [12]. Both support human judgment through scoring frameworks or feature validation processes, respectively, but neither automates coordination decisions nor generates outputs for direct workflow use. Two studies are classified as L2: Harode et al. [9] and Kumar et al. [40]. In both cases, the system generates recommendations, specifically predefined resolution options, which the human engineer selects and implements the preferred option.

The largest group of studies operates at L3. Fifteen studies fall into this category because they generate automated outputs that require human review before consequential use. This includes filtering and classification studies by Hu and Castro-Lacouture [8], Huang and Lin [31], Lin and Huang [32], Luo et al. [15], Ahmadpanah et al. [16], Koo and Guerra [34], Koo et al. [35], Oduro et al. [38], and Ailem and Boton [39]; non-autonomous resolution studies by Hsu et al. [33], Hu et al. [19], and Chong et al. [37]; and prediction or detection studies by Shehadeh and Alshboul [36], Alshboul and Shehadeh [13], and Chen [41]. The dominance of L3 indicates that the field has moved beyond passive decision support, yet still largely relies on human review of AI-generated coordination outputs. Two studies are classified as L4: Liu et al. [21] and Liu et al. [22]. Both operate exclusively in simulated environments, and neither reports deployment in live BIM coordination workflows.

This distribution reveals a significant gap. The literature has advanced automated output generation, particularly at L3, but has not developed an equivalent account of the human-machine oversight configuration required to govern such outputs. Most studies assume that human reviewers will detect errors, manage uncertainty, and prevent inappropriate downstream use, yet they rarely specify review criteria, workload implications, expertise requirements, accountability, or vigilance risks. The central concern is therefore not only technical error, but the under-specified nature of human review itself.

4.2.8. Deployment Readiness and Practical Integration

To assess whether reported model performance translates into deployment readiness, this review evaluates the literature against three observable criteria adapted from the Technology Readiness Level framework described by Mankins [44], which distinguishes

validation in controlled settings from demonstration in operational environments. These criteria include validation across multiple projects to test robustness in varied coordination contexts, practitioner evaluation to determine whether outputs are interpretable and actionable within real workflows, and integration with live BIM platforms to confirm correct operation within actual software environments and data formats.

When assessed against these criteria, most studies show limited deployment readiness, as the evidence reported in Tables 3 and 4 indicates that evaluation and validation approaches are typically confined to controlled experimental settings, with little or no integration into active coordination workflows. Exceptions include Ailem and Boton [39], who deployed their ensemble model as a Navisworks plug-in and evaluated it against manual coordinator judgements across nine test cases; Alshboul and Shehadeh [13], who described a phased real-world deployment scenario across 30 projects; Koo et al. [35], who validated relevance and risk labels with VDC experts and framed ML outputs as decision support for filtering and prioritisation; and Harode et al. [9,12], whose two-paper series progressed from expert validation of the feature space through to ML model development on real project clash data.

The practitioner-facing evaluation by Ailem and Boton [39], which included a five-responder questionnaire assessing the plug-in's usability and alignment with coordinator practice, represents the closest approximation to end-user validation in the corpus. Shehadeh and Alshboul [36] incorporated VR usability testing via Oculus Quest 2 and Unity integration with Revit Live, introducing a novel interface dimension that no other study in the corpus addressed. Across the corpus, however, model transparency remains under-addressed relative to predictive performance, and no study explicitly aligned its AI system with risk classification frameworks such as those proposed under the EU AI Act [43], despite the safety implications of automated resolution advice in construction projects.

5. Discussion

This review shows that AI research for BIM-based clash management has advanced methodologically but remains weakly positioned for professional deployment. Four issues explain this gap: research effort is concentrated on technically tractable tasks; evidence remains too project-specific to support generalisable claims; predictive performance has outpaced interpretability and accountability; and most systems are developed as isolated technical artefacts rather than as components of professional coordination workflows.

The distribution of studies across the clash management lifecycle reveals a strong concentration on clash filtering, which represents one of the most structured and lowest-consequence tasks. In contrast, clash resolution and prevention, which directly influence coordination quality, project delays, and rework costs, remain comparatively underexplored. This reflects a broader tendency in applied AI research to prioritise problems that are measurable and computationally manageable over tasks requiring contextual judgement and multidisciplinary reasoning [45,46].

This concentration is partly explained by the alignment between classical supervised learning methods and the data structure of filtering tasks, where BIM clash reports provide labelled, tabular, and comparatively well-bounded classification inputs. Classical supervised machine learning and rule-based systems are therefore well suited to filtering, where features such as element type, discipline pair, penetration geometry, and spatial location can be extracted and modelled as structured attributes. They are less suited to the spatial, sequential, and constraint-intensive reasoning required for resolution and prevention. More expressive approaches, including graph neural networks, reinforcement learning, and generative methods, are better aligned with these higher-consequence tasks, but require

richer datasets, stronger validation designs, and greater computational and integration infrastructure than the field currently provides.

The evidential foundation of the field also remains limited. Most reported performance metrics are derived from single-project datasets and evaluated under in-sample or project-specific conditions. Such findings do not constitute transferable evidence. Without benchmark datasets, standardised labelling protocols, and cross-project validation, reported accuracies remain context-dependent and difficult to compare. This limitation reflects the epistemic character of construction data itself. Clash labels are not objective ground truths, but outcomes of project-specific coordination practices, regulatory conditions, and professional judgement. Larger datasets drawn from a single project context may improve in-sample performance, but they do not resolve external validity concerns and may reinforce project-specific modelling and labelling biases.

Methodological advances further expose a tension between predictive performance and practical legitimacy. Ensemble learning, deep learning, and graph-based models improve pattern recognition and relational analysis, yet these gains are not matched by equivalent progress in interpretability, accountability, or workflow integration. This matters because clash coordination decisions can carry contractual, financial, and safety consequences. As Rudin [47] argues, high-stakes decision environments require inherently interpretable systems or robust explanation mechanisms, since post-hoc explanations cannot guarantee faithful representation of model reasoning. Across the reviewed corpus, explainability is largely absent. Only Shehadeh and Alshboul [36] incorporated SHapley Additive exPlanations (SHAPs) to interpret model outputs. A system that cannot justify its recommendations cannot support accountable coordination practices, regardless of predictive accuracy.

The automation-level analysis reinforces this concern. Fifteen of the twenty-one studies are classified at L3, where AI systems generate coordination outputs that humans are expected to review before downstream action. This indicates that the field has converged on a configuration in which the system acts first, and the coordinator verifies. Yet studies rarely specify how verification should occur, what criteria should guide acceptance or rejection, how much review effort is required, or what level of model uncertainty should trigger independent professional judgement. Under the framework proposed by Parasuraman et al. [42], such configurations are vulnerable to automation bias and weakened oversight, particularly when systems report high apparent accuracy. Almost none of the reviewed studies empirically evaluate reviewer behaviour, human-system interaction, or the allocation of responsibility between coordinators and AI systems.

From a sociotechnical perspective, the persistent gap between experimental validation and operational deployment reflects a broader failure to jointly design technical systems and organisational workflows [48]. Existing AI clash management systems are typically developed in isolation from the coordination practices, contractual structures, communication protocols, and accountability mechanisms that shape real project environments. This weak integration limits the likelihood that technically effective systems will be consistently adopted or produce comparable outcomes in practice.

A related limitation concerns the instability of both the underlying data and the problem definition itself. Clash reports function as boundary objects [49], interpreted differently by designers, coordinators, and construction teams. Consequently, labels vary across projects and organisational contexts. Models trained on labels from one coordination environment may become misaligned when deployed elsewhere. This issue is compounded by contextual data friction [50], where assumptions embedded in data generation do not transfer between settings. The problem is especially significant for resolution and

prevention, which depend heavily on sequencing, constructability, design intent, and professional judgement.

The literature on clash prevention exposes a deeper epistemological issue. Predicting clashes and preventing clashes are not equivalent tasks [51], yet several studies risk treating predictive performance as a proxy for preventive capability [13,36]. Correlational models trained on historical project data can identify patterns associated with clashes, but they cannot establish causal mechanisms without causal modelling or controlled experimentation [52,53]. High predictive accuracy may therefore reflect recognition of project complexity signatures rather than identification of the upstream conditions that generate clashes. Future research should therefore shift from simply predicting where clashes may occur to identifying the design, modelling, and coordination conditions that repeatedly lead to them. This shift is supported by adjacent AEC AI research, such as Mirhosseini et al. [54], which demonstrates how predictive modelling can be reframed around actionable upstream conditions rather than outcome prediction alone. When combined with methods such as SHAP-based attribution, this approach could improve explanatory transparency, strengthen accountability, and provide a more credible basis for causal investigation in clash prevention research [47,52].

These structural limitations also carry increasing regulatory significance. The European Union AI Act [43] provides a useful reference point for assessing the governance conditions required for AI systems deployed in regulated environments. Systems that autonomously modify design geometry or influence safety-critical coordination decisions may fall within high-risk governance expectations, particularly where their outputs affect structural safety, regulatory compliance, or occupational risk. Yet the reviewed corpus provides limited evidence that current systems would satisfy comparable requirements for risk management, data governance, technical documentation, and meaningful human oversight. First, risk management provisions requiring continuous monitoring, hazard detection, and human override mechanisms are largely absent. Second, data governance requirements mandating representative, high-quality, and unbiased datasets are not met by single-project corpora characterised by contextual bias and unresolved class imbalance. Third, technical documentation requirements necessary for conformity assessment are unsupported because most studies provide insufficient dataset transparency and replication detail. The limitations identified in this review are therefore not only methodological. They are preconditions for deployment that the field has not yet addressed.

Beyond formal regulatory compliance, deployment readiness also depends on trust in the digital information environment in which AI-enabled coordination occurs. BIM-based clash management relies on shared models, clash records, issue histories, and automated outputs that move across disciplinary and organisational boundaries. In this context, unclear model update procedures, insecure information exchange, opaque algorithmic recommendations, or weakly governed data flows can undermine practitioner trust and create organisational resistance to AI adoption. Broader research on AI-enabled digital ecosystems and smart infrastructure similarly emphasises that technical automation must be supported by legal accountability, cybersecurity, data protection, and transparent governance arrangements [55]. For AI-enabled clash management, reliable deployment therefore requires not only accurate models, but also auditable data provenance, secure model exchange, transparent update logs, and clearly assigned responsibility for accepting, modifying, or rejecting AI-generated coordination outputs. These requirements are particularly important because collaborative BIM environments involve the continuous exchange of sensitive project data, creating potential vulnerabilities to unauthorised access, data manipulation, and model corruption.

Algorithmic transparency and human oversight are equally central to responsible AI deployment in clash management contexts. Transparency, in this sense, refers not merely to access to source code or model architecture, but to the practical ability of practitioners, including project managers, BIM coordinators, engineers, and contract administrators, to understand, interrogate, and critically evaluate AI-generated outputs. This aligns with emerging regulatory frameworks, most notably the EU AI Act [43], which classifies AI systems used in safety-critical infrastructure applications as high-risk and imposes obligations relating to technical documentation, conformity assessment, and human oversight mechanisms.

Research Gaps

The four arguments developed above converge on two categories of research gaps. Methodologically, the most critical gap is the absence of systematic cross-context validation. Most studies rely on single-project or internally validated datasets, and even those drawing on multiple projects rarely test transferability by training and evaluating across clearly separated project contexts. Shared benchmark datasets remain limited, formal explainability mechanisms are rare, and human-machine function allocation is seldom specified. These absences collectively constrain cumulative knowledge building, meaningful cross-study comparison, practitioner trust, and safe deployment at higher automation levels. The field remains fragmented, with studies addressing separate parts of the clash management pipeline. Although some studies cover adjacent functions, such as detection and filtering or filtering and risk prediction, no study develops a validated end-to-end workflow that connects filtering, prioritisation, resolution, and prevention.

Technically, three gaps reflect the static and consequence-blind nature of current approaches. First, all models operate on fixed BIM snapshots; none accommodate dynamically evolving design states, which is the normal condition of a live coordination workflow. Second, inter-clash dependencies are largely ignored: resolution of one element displaces spatial conditions for others, yet no filtering or resolution model accounts for this propagation. Third, no study applies asymmetric cost-sensitive evaluation, despite the coordination reality that missing a structurally significant clash carries substantially greater consequences than a false positive. Addressing these gaps requires investment in shared data infrastructure, sociotechnical co-design with practitioners, and evaluation frameworks calibrated to the consequence profiles that determine real coordination value.

6. Conclusions

This review systematically examined 21 empirical studies published between 2019 and 2026 on AI applications for BIM-based construction clash management. The findings show that the field has progressed from simple binary filtering models toward more advanced ensemble, graph-based, reinforcement learning, and generative approaches. However, this progress remains uneven across the clash management lifecycle. Research is concentrated on filtering and classification, while prioritisation, resolution, and prevention remain comparatively underdeveloped. No study develops an integrated end-to-end system spanning filtering, prioritisation, resolution, and prevention. These findings demonstrate clear technical advancement in AI-enabled BIM coordination but also reveal a fragmented evidence base.

The review further shows that the principal barriers to deployment are structural rather than purely algorithmic. Most studies rely on single-project datasets, limited validation environments, and context-specific labelling practices, restricting the generalisability and operational reliability of reported performance. In parallel, limited attention has been given to interpretability, human oversight, workflow integration, and practitioner

accountability, despite their importance in high-stakes coordination environments. The automation-level analysis further supports this concern. Most studies operate at L3, where AI generates outputs that still require human review, and the studies classified as L4 are limited to simulated environments. As a result, current research demonstrates technical feasibility under controlled conditions but provides limited evidence of readiness for real project deployment. Operational use will require shared benchmark datasets, cross-project validation, interpretable and auditable outputs, clearly defined human oversight mechanisms, integration with live BIM coordination workflows, and stronger governance of data provenance, cybersecurity, and accountability.

This review makes three contributions to the existing literature. First, it provides one of the first focused syntheses of AI applications across the BIM-based clash management lifecycle. Second, it introduces a four-tier automation taxonomy grounded in human factors theory, showing that studies classified as autonomous remain confined to simulation-based validation. Third, it reframes deployment readiness as a sociotechnical and evidential problem rather than a purely algorithmic one, showing that current barriers arise from limited cross-context validation, opaque outputs, weak integration with professional coordination workflows, poorly specified mechanisms of oversight and accountability, and insufficient causal reasoning in prevention-focused studies.

The review concludes that the field faces a validity challenge rather than a modelling challenge. Future progress depends less on increasing architectural complexity alone and more on developing shared benchmark datasets, practice-aligned evaluation frameworks, interpretable systems, practitioner co-design, and robust integration between AI systems and professional coordination workflows. More broadly, the findings highlight that AI deployment in construction cannot be treated solely as a technical optimisation problem. In coordination and risk-management contexts, tacit knowledge, professional judgement, and human intuition remain central to project delivery and decision-making, requiring AI systems to support rather than displace human expertise.

This review is subject to several limitations that frame the interpretation of its findings. First, the corpus comprises 21 studies, which provides a focused but still emerging evidence base. The findings, therefore, support analytical synthesis rather than statistical generalisation. Second, the search was limited to Scopus, English-language publications, journal articles, and indexed conference proceedings. Relevant studies indexed only in Web of Science, IEEE Xplore, ASCE Library, Engineering Village, ScienceDirect, non-English outlets, doctoral theses, industry reports, or software documentation may therefore have been missed. Future review papers could extend the evidence base by searching multiple databases and applying backward and forward citation chasing. Third, the heterogeneity of the included studies in terms of AI methods, datasets, validation procedures, evaluation metrics, and problem formulations precluded quantitative meta-analysis.

Future research should prioritise three directions. First, the field requires publicly accessible, multi-project and multi-disciplinary benchmark datasets with standardised labelling protocols. Without shared benchmarks, reported performance will remain context-dependent and difficult to compare across studies. Second, fragmented AI capabilities should be integrated into end-to-end clash management systems that connect filtering, prioritisation, resolution, and prevention while retaining clear human oversight at high-consequence decision points. Explainability should be treated as a core design requirement, since practitioners must be able to interpret, contest, and justify AI-generated coordination outputs. Third, future studies should expand the empirical and technological scope of the field by examining large language models, multimodal AI, digital twins, and natural language processing for unstructured coordination data such as clash comments, resolution notes, and RFI records. Progress will also require closer alignment between researchers,

practitioners, software vendors, and standards frameworks, particularly ISO information-management standards and IFC-based OpenBIM interoperability [56].

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/buildings16132623/s1>, Supplementary Table S1: Quality appraisal criteria and rating basis. Supplementary Table S2: Study-level quality appraisal of included studies. PRISMA checklist [25].

Author Contributions: Conceptualisation, B.A.; methodology, B.A. and A.A.; formal analysis, B.A. and A.K.; investigation, B.A.; data curation, B.A.; writing—original draft preparation, B.A.; writing—review and editing, A.A., G.A.-M. and A.K.; supervision, B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Acknowledgments: During manuscript preparation, the authors used ChatGPT (Version 5.5) and Microsoft Copilot (Version 134.0.3124.51) to support language editing, proofreading, and refinement of academic expression. These tools were not used to conduct reviews, select studies, extract data, appraise quality, code themes, generate findings, or determine the interpretation of results. All intellectual content was developed and verified by the authors, who take full responsibility for the final manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation/Term	Full Term
A*	A-star algorithm
AdaBoost	Adaptive Boosting
AEC	Architecture, Engineering and Construction
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BIM	Building Information Modelling
BN	Bayesian Network
BPNN	Backpropagation Neural Network
BWM	Best-Worst Method
CAD	Computer-Aided Design
cGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
CV	Cross-Validation
DQN	Deep Q-Network
DT	Decision Tree
ET	Extra Trees
F1-score	Harmonic mean of precision and recall
GAN	Generative Adversarial Network
GB	Gradient Boosting
GNN	Graph Neural Network
IFC	Industry Foundation Classes
ISO	International Organisation for Standardisation
J48	C4.5 Decision Tree Algorithm (Weka implementation)

Abbreviation/Term	Full Term
Jrip	Repeated Incremental Pruning to Produce Error Reduction
k-NN/KNN	k-Nearest Neighbours
LP	Linear Programming
LR	Logistic Regression
MARL	Multi-Agent Reinforcement Learning
MCDA	Multi-Criteria Decision Analysis
MEP	Mechanical, Electrical and Plumbing
ML	Machine Learning
MLP	Multilayer Perceptron
MLSOL	Multilabel Synthetic Oversampling
MMAT	Mixed Methods Appraisal Tool
NB	Naïve Bayes
OpenBIM	Open Building Information Modelling
PICO	Population, Intervention, Comparator, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RC	Reinforced Concrete
RF	Random Forest
RFI	Request for Information
RNN-LSTM	Recurrent Neural Network–Long Short-Term Memory
SA	Simulated Annealing
SHAP	SHapley Additive exPlanations
SLR	Systematic Literature Review
SMEN	Variational Multi-Semantic Encoder
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
TASCE	Temporal-Aware Semantic Coordination Engine
VDC	Virtual Design and Construction
VR	Virtual Reality
XGBoost	Extreme Gradient Boosting
YOLO	You Only Look Once

References

1. Pärn, E.A.; Edwards, D.J.; Sing, M.C.P. Origins and probabilities of MEP and structural design clashes within a federated BIM model. *Autom. Constr.* **2018**, *85*, 209–219. [\[CrossRef\]](#)
2. Mehrbod, S.; Staub-French, S.; Mahyar, N.; Tory, M. Beyond the clash: Investigating BIM-based building design coordination issue representation and resolution. *J. Inf. Technol. Constr.* **2019**, *24*, 33–57.
3. Chahrour, R.; Hafeez, M.A.; Ahmad, A.M.; Sulieman, H.I.; Dawood, H.; Rodriguez-Trejo, S.; Kassem, M.; Naji, K.K.; Dawood, N. Cost-benefit analysis of BIM-enabled design clash detection and resolution. *Constr. Manag. Econ.* **2021**, *39*, 55–72. [\[CrossRef\]](#)
4. Hwang, B.-G.; Thomas, S.R.; Haas, C.T.; Caldas, C.H. Measuring the Impact of Rework on Construction Cost Performance. *J. Constr. Eng. Manag.* **2009**, *135*, 187–198. [\[CrossRef\]](#)
5. Josephson, P.E.; Hammarlund, Y. The causes and costs of defects in construction: A study of seven building projects. *Autom. Constr.* **1999**, *8*, 681–687. [\[CrossRef\]](#)
6. Teo, Y.H.; Yap, J.H.; An, H.; Yu, S.C.M.; Zhang, L.; Chang, J.; Cheong, K.H. Enhancing the MEP Coordination Process with BIM Technology and Management Strategies. *Sensors* **2022**, *22*, 4936. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Malsane, S.; Merin, M.J.; Gokul, G.J.; Nithin Dev, P.; Babu, S.T. Management of High-Volume Clashes During Clash Detection. In *Mathematical, Computational Intelligence and Engineering Approaches for Tourism, Agriculture and Healthcare*; Springer: Singapore, 2022; pp. 285–295.
8. Hu, Y.; Castro-Lacouture, D. Clash Relevance Prediction Based on Machine Learning. *J. Comput. Civ. Eng.* **2019**, *33*, 04018060. [\[CrossRef\]](#)
9. Harode, A.; Thabet, W.; Gao, X. Developing a Machine-Learning Model to Predict Clash Resolution Options. *J. Comput. Civ. Eng.* **2024**, *38*, 04024005. [\[CrossRef\]](#)
10. Zhang, Y.; Wang, J.; Ahmad, R.; Li, X. Integrating lean production strategies, virtual reality technique and building information modeling method for mass customization in cabinet manufacturing. *Eng. Constr. Archit. Manag.* **2021**, *29*, 3970–3996. [\[CrossRef\]](#)

11. Sacks, R.; Girolami, M.; Brilakis, I. Building Information Modelling, Artificial Intelligence and Construction Tech. *Dev. Built Environ.* **2020**, *4*, 100011. [\[CrossRef\]](#)
12. Harode, A.; Thabet, W.; Leite, F. Formulation of Feature and Label Space Using Modified Delphi in Support of Developing a Machine-Learning Algorithm to Automate Clash Resolution. *J. Constr. Eng. Manag.* **2024**, *150*, 04023173. [\[CrossRef\]](#)
13. Alshboul, O.; Shehadeh, A. Adaptive Integration of BIM and Navisworks for Real-Time Clash Detection Using the XGBoost Algorithm. *J. Constr. Eng. Manag.* **2026**, *152*, 04025232. [\[CrossRef\]](#)
14. Adegun, I. BIM Clash Report Analysis Using Machine Learning Algorithms. Master's Thesis, University of Alberta, Edmonton, AB, Canada, 2024.
15. Luo, S.; Yao, J.; Wang, S.; Wang, Y.; Lu, G. A sustainable BIM-based multidisciplinary framework for underground pipeline clash detection and analysis. *J. Clean. Prod.* **2022**, *374*, 133900. [\[CrossRef\]](#)
16. Ahmadpanah, H.; Haidar, A.; Latifi, S.M. BIM and machine learning (ML) integration in design coordination: Using ML to automate object classification for clash detection. *Ecaade Proc.* **2023**, *2*, 619–628.
17. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
18. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Hu, Y.; Castro-Lacouture, D.; Eastman, C.M.; Navathe, S.B. Automatic clash correction sequence optimization using a clash dependency network. *Autom. Constr.* **2020**, *115*, 103205. [\[CrossRef\]](#)
20. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 1998; Volume 1.
21. Liu, J.; Liu, P.; Feng, L.; Wu, W.; Li, D.; Chen, F. Towards automated clash resolution of reinforcing steel design in reinforced concrete frames via Q-learning and building information modeling. *Autom. Constr.* **2020**, *112*, 103062.
22. Liu, P.; Qi, H.; Liu, J.; Feng, L.; Li, D.; Guo, J. Automated clash resolution for reinforcement steel design in precast concrete wall panels via generative adversarial network and reinforcement learning. *Adv. Eng. Inform.* **2023**, *58*, 102131. [\[CrossRef\]](#)
23. Bitaraf, I.; Salimpour, A.; Elmi, P.; Shirzadi Javid, A.A. Improved Building Information Modeling Based Method for Prioritizing Clash Detection in the Building Construction Design Phase. *Buildings* **2024**, *14*, 3611. [\[CrossRef\]](#)
24. Sain, S.R. The Nature of Statistical Learning Theory. *Technometrics* **1996**, *38*, 409. [\[CrossRef\]](#)
25. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Tranfield, D.; Denyer, D.; Smart, P. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *Br. J. Manag.* **2003**, *14*, 207–222. [\[CrossRef\]](#)
27. Papaioannou, D.; Sutton, A.; Booth, A. *Systematic Approaches to a Successful Literature Review*; SAGE Publications Ltd.: Thousand Oaks, CA, USA, 2016; pp. 1–336.
28. Aghaei Chadegani, A.; Salehi, H.; Yunus, M.M.; Farhadi, H.; Fooladi, M.; Farhadi, M.; Ale Ebrahim, N. A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. *Asian Soc. Sci.* **2013**, *9*, 18. [\[CrossRef\]](#)
29. Hong, Q.N.; Fàbregues, S.; Bartlett, G.; Boardman, F.; Cargo, M.; Dagenais, P.; Gagnon, M.-P.; Griffiths, F.; Nicolau, B.; O'Cathain, A. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ. Inf.* **2018**, *34*, 285–291. [\[CrossRef\]](#)
30. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [\[CrossRef\]](#)
31. Huang, Y.; Lin, W.Y. Automatic Classification of Design Conflicts Using Rule-based Reasoning and Machine Learning—An Example of Structural Clashes Against the MEP Model. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*; IAARC Publications: Edinburgh, UK, 2019; Volume 36, pp. 324–331. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Lin, W.Y.; Huang, Y.-H. Filtering of Irrelevant Clashes Detected by BIM Software Using a Hybrid Method of Rule-Based Reasoning and Supervised Machine Learning. *Appl. Sci.* **2019**, *9*, 5324. [\[CrossRef\]](#)
33. Hsu, H.-C.; Chang, S.; Chen, C.-C.; Wu, I.C. Knowledge-based system for resolving design clashes in building information models. *Autom. Constr.* **2020**, *110*, 103001. [\[CrossRef\]](#)
34. Koo, H.J.; Guerra, B.C.; Saka, S.; Leite, F. Clash Relevance Prediction in BIM Model Coordination Using Artificial Neural Network. In *Construction Research Congress 2024*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2024; pp. 127–136.
35. Koo, H.J.; Guerra, B.C.; Saka, S.; Leite, F. Streamlining BIM Coordination and Clash Resolution through Risk and Relevance Analysis. *J. Constr. Eng. Manag.* **2026**, *152*, 04026092. [\[CrossRef\]](#)
36. Shehadeh, A.; Alshboul, O. Enhancing Engineering and Architectural Design Through Virtual Reality and Machine Learning Integration. *Buildings* **2025**, *15*, 328. [\[CrossRef\]](#)
37. Chong, Y.C.; Wong, R.T.K.; Jasser, M.B.; Issa, B.; Chua, H.N. A Comparative Study Between a Finetuned Conditional Generative Adversarial Network (cGAN) and the A* Algorithm for BIM Clash Coordination. In *Proceedings of the 2025 IEEE 8th International Conference on Electrical, Control and Computer Engineering (InECCE)*, Kuantan, Malaysia, 27–28 August 2025; pp. 671–676.

38. Oduro, S.; Koo, H.J.; Guerra, H.J.B.C. Comparative Analysis of Machine Learning Algorithms for Clash Relevance and Risk Level Prediction in BIM Design Coordination. In *Computing in Civil Engineering 2025*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2025; pp. 638–648.
39. Ailem, R.; Boton, C. Automating clash relevance filtering in BIM-based multidisciplinary coordination using machine learning. *Autom. Constr.* **2026**, *181*, 106644. [[CrossRef](#)]
40. Kumar, A.; Kumar, D.; Swara, H.; Gupta, A.; Gupta, T. Resolving Design Clashes Using Machine Learning. In Proceedings of the 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 24–28 June 2024; pp. 1–7.
41. Chen, Y. Automatic Detection of Construction Conflicts and Risk Prediction Based on the Integration of Building Information Modeling and Graph Neural Networks. *Int. J. Image Graph.* **2026**, 2750091. [[CrossRef](#)]
42. Parasuraman, R.; Sheridan, T.B.; Wickens, C.D. A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern.—Part A Syst. Hum.* **2000**, *30*, 286–297. [[CrossRef](#)] [[PubMed](#)]
43. EUAI. Regulation (EU) 2024/1689 on Artificial Intelligence. Official Journal of the European Union, L Series. Available online: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng> (accessed on 1 May 2026).
44. Mankins, J.C. Technology Readiness Levels: A White Paper. 1995. Available online: http://www.artemisinnovation.com/images/TRL_White_Paper_2004-Edited.pdf (accessed on 1 May 2026).
45. Koschmann, T.D. *Mind Over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*; Dreyfus, H.L., Dreyfus, S.E., Eds.; Basil Blackwell: Oxford, UK, 1986; 223p.
46. Pasquale, F. *New Laws of Robotics: Defending Human Expertise in the Age of AI*; Harvard University Press: Cambridge, MA, USA, 2020. [[CrossRef](#)]
47. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [[CrossRef](#)] [[PubMed](#)]
48. Trist, E.L.; Bamforth, K.W. Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Hum. Relat.* **1951**, *4*, 3–38.
49. Star, S.L.; Griesemer, J.R. Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–1939. *Soc. Stud. Sci.* **1989**, *19*, 387–420.
50. Leonelli, S. *Data-Centric Biology: A Philosophical Study*; University of Chicago Press: Chicago, IL, USA, 2019.
51. Pearl, J. *The Book of Why: The New Science of Cause and Effect*; Basic Books: New York, NY, USA, 2018.
52. Kitchin, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*; Sage: Thousand Oaks, CA, USA, 2014.
53. Hernán, M.A.; Hsu, J.; Healy, B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. *CHANCE* **2019**, *32*, 42–49. [[CrossRef](#)]
54. Mirhosseini, S.A.; Chang, R.; Omrany, H.; Zou, Y.; Liu, Y. Application of machine learning technologies in predicting severity of heat-related illnesses among construction workers. *Smart Sustain. Built Environ.* **2026**, *12*, 1–30. [[CrossRef](#)]
55. Peráček, T.; Kaššaj, M. The Impact of Effective Public Administration and European Digital Connectivity on Building Smart Cities: An Analysis of Legal Frameworks, Policy Initiatives and Transformative Impact in the Emerging Era of Artificial Intelligence. *Jurid. Trib.-Rev. Comp. Int. Law* **2025**, *15*, 720–745.
56. Aranda-Mena, G.; Wakefield, R. Interoperability of building information—Myth of reality? In *eWork and eBusiness in Architecture, Engineering and Construction. ECPPM 2006: European Conference on Product and Process Modelling 2006 (ECPPM 2006)*, Valencia, Spain, 13–15 September 2006; CRC Press: Boca Raton, FL, USA, 2006; p. 127.

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.