

Systematic Review

Machine Learning in Architectural Heritage Conservation: A Systematic Review and Thematic Synthesis of Applications, Challenges, and Future Directions

Kaiming Li ¹ , Baitong Du ², Dailuo Li ³, Zhenxiang Liu ¹, Xiaotong Zhang ¹ and HaeYoon Kim ^{1,*}

¹ Department of Design, General Graduate School, Dongseo University, 47 Jurye-ro, Sasang-gu, Busan 47011, Republic of Korea; likaiming612@gmail.com (K.L.); jongsyang@hotmail.com (Z.L.); zxtssdlh@naver.com (X.Z.)

² College of Arts and Media, Shenyang Institute of Technology, Wanghua District, Shenyang 113122, China; dubaitong@126.com

³ Faculty of History, Institute of Archaeology, Liaoning University, Shenbei New District, Shenyang 110136, China; lidailuo2005@163.com

* Correspondence: haeyoonkim@empal.com

Abstract

Machine learning (ML) is increasingly used in architectural cultural heritage conservation, but existing evidence remains fragmented across tasks, data types, and technical workflows. This systematic review synthesizes 33 studies on machine learning applications in architectural heritage conservation and identifies major application domains, methodological patterns, evidence gaps, and future research directions. Following a PRISMA-oriented review process, the literature was searched in Web of Science, Scopus, IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar up to 31 January 2026. Eligible studies were screened according to predefined inclusion and exclusion criteria, extracted using a structured coding form, and synthesized through qualitative thematic analysis. The included studies were grouped into three major domains: semantic understanding of heritage data, multi-scale damage detection and structural health monitoring, and system-level integration through heritage building information models (HBIM), multimodal data fusion, and digital twins. The evidence indicates a transition from isolated data-driven analysis toward predictive and decision-support systems, while persistent limitations remain in model generalization, benchmark datasets, semantic-to-HBIM transformation, multimodal fusion, and real-world deployment. Machine learning has substantial potential to support preventive, interpretable, and context-aware conservation, but future research requires more transparent reporting, shared datasets, external validation, and closer integration with conservation expertise.

Keywords: machine learning; heritage conservation; systematic review; thematic synthesis; structural health monitoring; digital twin; HBIM



Academic Editor: Fabrizio Banfi

Received: 12 June 2026

Revised: 7 July 2026

Accepted: 8 July 2026

Published: 10 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

Architectural cultural heritage is the material carrier of human history and collective memory, containing irreplaceable artistic, scientific, and social value [1]. However, these heritage sites are facing a dual threat: at the natural level, extreme weather caused by climate change and the erosion of coastal historic cities by rising sea levels, as well as the continuous erosion of their material structure by slow-moving processes such as long-term weathering [2]. From a human perspective, rapid urbanization, inappropriate

tourism development, and human-caused damage have further exacerbated the vulnerability of heritage [3]. How to record, monitor, and protect architectural heritage in a scientific, efficient, and sustainable manner has become a core challenge in the global heritage conservation field.

The introduction of digital technology has ushered in a new paradigm for heritage preservation. Photogrammetry, laser scanning and other reality-capture technologies can quickly and accurately acquire three-dimensional geometric and appearance information of architectural heritage and generate high-fidelity point cloud or mesh models, which greatly improve the efficiency and accuracy of information recording [4,5]. However, this wave of digitalization has also brought new challenges: while massive amounts of 3D data are extremely rich in metric data, they are essentially raw and unstructured, lacking semantic information and hierarchical structure [6]. Without effective interpretation, its potential for in-depth analysis and decision support will be difficult to realize. Faced with the dilemma of “abundant data but scarce information,” artificial intelligence technology, with machine learning at its core, provides a key solution. Machine learning can automatically learn patterns from complex, high-dimensional data, enabling semantic segmentation, lesion identification, and behavior prediction of heritage data, and driving the transformation of heritage protection from passive, reactive restoration to proactive, preventative, and even predictive protection models [7,8].

Over the past decade, the application of machine learning in the field of architectural cultural heritage conservation has developed into several relatively mature research directions. In terms of semantic understanding and modeling, researchers are dedicated to developing intelligent algorithms for automatically identifying building components. Early methods projected 3D models into 2D images, classified them using supervised learning, and then mapped them back into 3D space [9]. Subsequent research shifted to directly processing the geometric and textural features of point cloud or mesh data [10]. Algorithms have also evolved from traditional models such as random forests and support vector machines to deep learning models such as PointNet and dynamic graph convolutional neural network (DGCNN) models [11,12], and multi-level, multi-resolution classification strategies have been developed to handle complex building structures [13]. These semantic segmentation results laid the foundation for building heritage building information models (HBIM) [4]. In structural health assessment and predictive protection, image recognition technology based on convolutional neural networks (CNNs) is widely used to automatically detect surface defects such as cracks, spalling, vegetation growth, and material discoloration [14,15]. Furthermore, researchers have constructed a digital twin system integrating IoT sensors, HBIM models, and machine learning algorithms. By continuously monitoring environmental and structural parameters, analyzing time-series data to identify degradation trends, and simulating the long-term effects of different interventions, they achieved data-driven predictive maintenance decisions [16]. In terms of emerging applications, machine learning has also been used to identify architectural style features in specific regions [17] and to infer and repair missing parts of historical buildings using generative adversarial networks (GANs) [18,19], further expanding the application boundaries of this field.

Machine learning has dramatically improved the architectural cultural heritage conservation process, but most research faces multiple deep challenges that hamper the transition from academic exploration to application. Most investigations are conducted in an ad hoc manner: at an individual site or on a single type of data or problem related to conservation. At present, the academic community lacks an integrative perspective. This makes it difficult to form a systematic understanding of the overall development landscape and the core technological paths of this field. The architectural heritage is said to exhibit “non-uniqueness”.

This implies that no two sites are alike. Models that are trained on a particular scene may show a drastic drop in performance when applied to scenes that are drastically different in style or era [20]—for instance, a model trained on an Italian church may be unable to recognize components of an ancient Chinese wooden building [21]. This crisis of generalization blocks the large-scale deployment of such models. Furthermore, constructing datasets for architectural heritage is challenging. The process of acquiring, cleaning, and labeling data is tedious, labor-intensive, and requires deep domain knowledge. This significantly limits the potential of deep learning models. How to automatically convert discrete labels into HBIM or digital twin objects with correct topological relationships, parametric geometric shapes and rich historical attributes is also a problem that has not yet been solved. The current process largely relies on cumbersome manual intervention [4]. The interconnected challenges indicate that this field urgently needs systematic knowledge organization and integration to transcend the limitations of case studies and construct a more macro-level and profound cognitive framework.

To address these challenges, this study aims to conduct a systematic review of existing literature on the application of machine learning in architectural cultural heritage conservation. Rather than simply listing individual studies, this review follows a transparent search, screening, extraction, and synthesis process to identify the main technical pathways, recurring methodological limitations, and practical barriers in the field. The review is guided by five research questions: (1) What machine learning methods have been applied to architectural heritage conservation? (2) Which conservation tasks and heritage data types are most frequently addressed? (3) What datasets, metrics, and validation strategies are used to evaluate these methods? (4) What limitations constrain model generalization, semantic transformation, and practical deployment? (5) What future directions can promote scalable, interpretable, and conservation-oriented ML applications?

The contributions of this systematic review are threefold. First, it provides a structured evidence map of an emerging interdisciplinary field and clarifies the relationships among semantic segmentation, structural health monitoring, multimodal integration, HBIM, and digital twin research. Second, it identifies practical implications for architects, conservation experts, heritage managers, and technical developers by distinguishing mature application areas from unresolved methodological bottlenecks. Third, it strengthens the methodological transparency of knowledge synthesis in this field by combining PRISMA-oriented study selection, structured data extraction, quality-oriented appraisal, and thematic synthesis, thereby offering a reproducible basis for future reviews and empirical studies.

2. Methods

To clarify the methodological role of reality-capture and 3D data in this review, these technologies were treated not only as background tools but also as key data sources that shaped the search strategy, eligibility criteria, and thematic coding. Accordingly, studies involving laser scanning, photogrammetry, point clouds, mesh models, image-based 3D reconstruction, HBIM, and digital twin workflows were considered within the scope of the review when they used machine learning or artificial intelligence for heritage documentation, interpretation, diagnosis, or conservation decision support. This methodological framing ensures that the technical discussion of 3D data acquisition is directly connected with the subsequent literature search and synthesis process.

The review process included protocol definition, database searching, duplicate removal, title and abstract screening, full-text eligibility assessment, structured data extraction, methodological quality appraisal, coding, and thematic synthesis. The reporting logic was aligned with PRISMA 2020 [22] principles, with particular attention to transparent eligibility criteria, information sources, study selection, and synthesis procedures. Studies

were included covering point cloud semantic segmentation, structural health monitoring, surface and material damage detection, multimodal data fusion, HBIM, and digital twin system integration.

2.1. Search Strategy

A systematic literature search was conducted in Web of Science, Scopus, IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar, with the final search completed on 31 January 2026. The search strategy combined three groups of terms related to machine learning and artificial intelligence, architectural or built heritage, and conservation-related applications. The core search string included: (“machine learning” OR “deep learning” OR “artificial intelligence” OR “computer vision” OR “neural network*” OR “semantic segmentation” OR “digital twin” OR “HBIM”) AND (“architectural heritage” OR “built heritage” OR “cultural heritage building*” OR “historic building*” OR “heritage building*” OR monument*) AND (conservation OR preservation OR restoration OR documentation OR monitoring OR “damage detection” OR “structural health monitoring” OR “condition assessment”). The same conceptual string was adapted to the syntax of each database, and the detailed database-specific strings are provided in Supplementary Table S1.

Eligible studies were peer-reviewed English-language journal articles and peer-reviewed full conference or proceedings papers that reported original empirical, technical, or mixed-method findings on the application of ML or AI to architectural or built heritage conservation. Review articles, conference abstracts, editorials, non-peer-reviewed conference materials, short papers without sufficient methodological or empirical information, non-English publications, studies without ML/AI methods, and studies unrelated to architectural heritage were excluded. The study selection process is summarized in Figure 1.

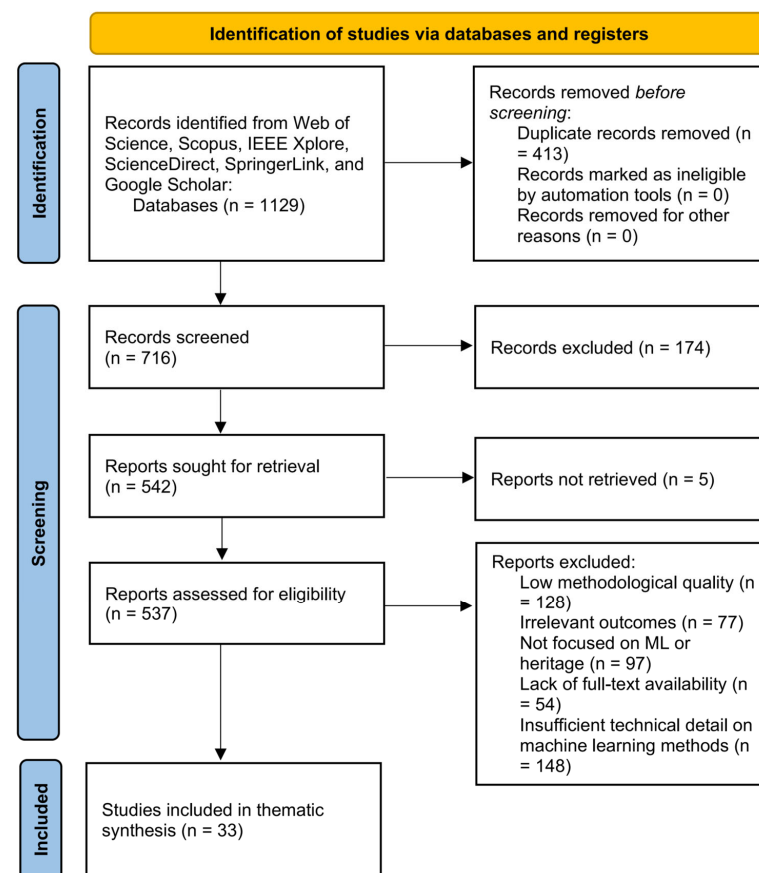


Figure 1. PRISMA flow diagram.

2.2. Study Selection and Data Extraction

After duplicate removal, titles and abstracts were screened against the eligibility criteria. Potentially relevant studies were then assessed in full text. Screening decisions were checked by two researchers, and disagreements were resolved through discussion. For each included study, data were extracted using a structured form covering author, year, country/region, study object/case, conservation task, data type, machine learning method, validation strategy, evaluation metrics, limitations. Extracted evidence was further divided into meaning units. Each meaning unit contained a core item of identifiable information, such as an algorithm application description, technical parameter, performance indicator, data constraint, implementation experience, or stated limitation [23].

2.3. Methodological Quality Appraisal and Coding

Because the included studies were heterogeneous in task type and evaluation design, methodological quality was appraised using criteria tailored to ML-based heritage conservation research. The appraisal considered whether each study clearly reported its data source, sample size or data volume, annotation procedure, model architecture, training-validation-test strategy, evaluation metrics, baseline comparison, external or cross-site validation, code or data availability, and discussion of generalization or practical deployment. These items were used to interpret the strength and transferability of the evidence rather than to exclude studies after eligibility screening. The coding and classification process followed a bottom-up inductive principle to capture the diversity of the original research [24]. To enhance reliability and transparency, the two researchers cross-checked the initial coding and quality appraisal results. Differences in the interpretation of meaning units, quality indicators, or thematic categories were resolved through discussion. The purpose of this process was to establish a coherent link between individual study findings and higher-level themes, while reducing the influence of individual disciplinary backgrounds on classification decisions [25].

To make the appraisal auditable at the study level, a methodological quality matrix was added as Supplementary Table S2. Each included study was assigned scores from 1 to 3 for the five criteria: data source and sample reporting, validation strategy and metrics, baseline comparison, external or cross-site validation, and code or data availability. A score of 3 indicated that the criterion was clearly reported and methodologically strong, 2 indicated partial reporting or adequate reporting with limitations, and 1 indicated weak reporting, non-applicability, or information not available from the reviewed paper. The overall reliability score summarized reporting completeness and evidence transferability, rather than the substantive importance of the study.

2.4. Thematic Synthesis

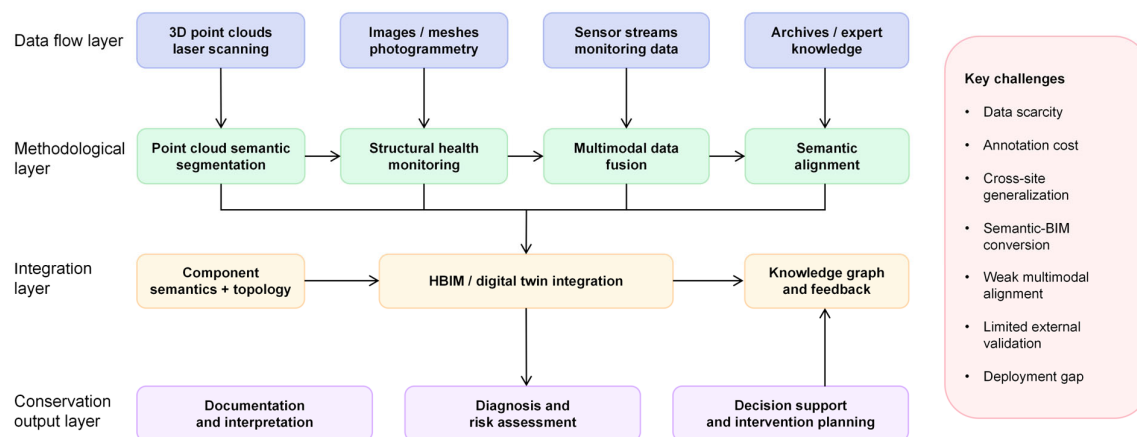
Thematic refinement and pattern identification were undertaken. The synthesis remained close to the original research descriptions and avoided over-interpreting findings beyond the reported evidence. Specific procedures included: (a) horizontal comparison of meaning units within each theme to identify similarities and differences in methods, datasets, metrics, application scenarios, and limitations; (b) cross-study grouping of recurrent patterns, including supervised and deep learning trends in point cloud parsing, unsupervised or weakly supervised strategies under label scarcity, multi-scale damage identification, multimodal integration, and digital twin frameworks; and (c) construction of a logical network linking technical pathways, application levels, and research challenges to present a holistic picture of ML applications in architectural heritage conservation [26,27].

The thematic synthesis was used to answer the research questions and to identify both application trends and evidence gaps. Particular attention was paid to data scarcity, se-

mantic and BIM conversion barriers, model generalization, insufficient external validation, early-stage multimodal fusion, and the limited translation of algorithms into conservation decision-making [28,29]. On this basis, the review summarizes future research priorities, including high-quality multimodal benchmark datasets, hybrid intelligent models that integrate domain knowledge, deeper multimodal fusion, model interpretability, and expert-in-the-loop decision support.

2.5. Analytical Framework

The analytical framework has been revised into a layered schematic in Figure 2. The schematic converts the previous textual description of four interrelated areas into a visual structure that shows how heritage data inputs flow into point cloud semantic segmentation, structural health monitoring, multimodal data fusion, and HBIM/digital twin integration, and how these methodological links are constrained by recurring challenges such as data scarcity, generalization, semantic-BIM conversion, weak multimodal alignment, and practical deployment gaps.



Note. Data flow proceeds from heritage data acquisition to ML interpretation, HBIM/digital twin integration, and conservation decisions. Methodological links connect the four application areas, while recurring challenges constrain transfer from algorithms to field practice.

Figure 2. Layered schematic of data flow, methodological links, and challenges across four interrelated ML application areas in architectural heritage conservation.

This systematic review followed a transparent process of literature retrieval, screening, eligibility assessment, data extraction, methodological appraisal, coding, classification, and thematic synthesis. This process provides the basis for the results reported below and supports a structured understanding of current ML applications, development trends, and unresolved challenges in architectural cultural heritage conservation.

3. Results

3.1. Study Selection and Characteristics of Included Studies

This review included 33 studies. Table 1 shows the characteristics of the studies. These studies were published between 2017 and 2025 and covered countries such as Iran, Canada, Italy, Denmark, Algeria, China, India, Spain, Turkey, Portugal, the Philippines, Australia, and Indonesia. Italy and China had a significant number of related studies, demonstrating the international application trend of machine learning in architectural heritage conservation. The types of objects included in the studies were diverse, including historical buildings, religious buildings, palaces, ancient villages, grottoes, temples, theaters, industrial heritage, architectural components, historical building image datasets, and point cloud data.

Table 1. Basic characteristics of the included studies.

Author (Year)	Country	Heritage Object/Case	ML Method	Evaluation Metrics
Bahrami and Albadvi (2023) [30]	Iran	Multiple heritage buildings across Iran (Semnan, Hamedan, Tehran, etc.)	CNN; MobileNetV2; ResNet152V2; InceptionResNetV2; Grad-CAM	Accuracy; Precision; Recall; F1-score; AUC; Loss
Baradaran Rahimi et al. (2025) [31]	Canada	Escalier St-Joseph, historic architectural interior	Neural Radiance Fields	PSNR; SSIM
Battini et al. (2024) [32]	Italy	Ducal Palace of Urbino (historical vault systems in a Renaissance palace)	PointNet; DGCNN; shape-grammar-based synthetic data generation	Accuracy; Overall Accuracy; Precision; Recall; F1-score; IoU
Boesgaard et al. (2022) [33]	Denmark	Museum Storerooms; Medieval Churches	XGBoost, RandomForest	Precision, Recall, F1-score
Bouchachi et al. (2025) [34]	Algeria	Algerian historical monuments	ChatGPT-4-turbo; LLM/generative AI; prompt-based multimodal analysis	Expert validation; pathology certainty rate 40% in first experiment; >90% confidence/accuracy in selected multi-image + glossary condition; reference veracity analysis
Bounouioua et al. (2025) [35]	Algeria	Algerian National Theatre Mahieddine Bachtarzi	Proprietary AI/ML algorithm in Aurivus; AI-assisted point-cloud recognition and segmentation; Revit + NavVis IVION integration	NR
Buldo et al. (2024) [36]	Italy	Renaissance Palaces	Random Forest; Mean Decrease in Impurity	Overall Accuracy, Precision, Recall, F1-score (including weighted F1), Training Time, Confusion Matrix
Cao et al. (2022) [37]	Italy	Sacromonte Calvario di Domodossola chapels	Random Forest; Dynamic Graph Convolutional Neural Network (DGCNN)	Overall Accuracy; weighted Precision; weighted Recall; weighted F1-score
Casillo et al. (2024) [38]	Italy	University of Salerno Science Library	AutoEncoder, kNN	Accuracy, Precision, Recall, F1-score
Croce et al. (2021) [4]	Italy	Pisa Charterhouse	Random Forest	Precision; Recall; Overall Accuracy; F-measure
Fang et al. (2024) [39]	China	Ancient Historic Buildings	CNN, GAN, FeatureGenNet, AncientNet, ContextAdaptNet, NeuralCD	Accuracy, Ablation Study
Fiorini et al. (2024) [40]	Italy	Palazzo Pitti	K-means Clustering	Accuracy; F1-score; OverSeg; Median IoU; MAD of IoU
Galantucci et al. (2023) [41]	Italy	Monasteries, Underground Arcades, Rural Churches	Hierarchical Clustering; Random Forest	Precision; Sensitivity / Recall; Overall Accuracy; F1-score
Gbran et al. (2025) [42]	Indonesia	Semarang Architectural Heritage	Unsupervised Hierarchical Clustering; Supervised Random Forest	Precision, Recall, F1-score, Overall Accuracy, IoU, classification agreement, intra-cluster variance, silhouette coefficient
Gokak et al. (2025) [43]	India	IHDS cultural heritage building image dataset	MobileNetV3; ResNet; Grad-CAM; CNN; transfer learning; data augmentation; model pruning	Accuracy
Grilli and Remondino (2020) [20]	Italy	Urban Architecture (11th–20th Century), Renaissance-style Lodge, Medieval Squares	Random Forest	Precision, Recall, F1-score
Karadag (2022) [18]	Turkey	Early Ottoman Tombs	Conditional GAN, Pix2PixHD, Deep Learning, Image-to-Image Translation	SSIM, MSE, PSNR, Generator Loss, Discriminator Loss

Table 1. Cont.

Author (Year)	Country	Heritage Object/Case	ML Method	Evaluation Metrics
Karimi et al. (2024) [44]	Portugal	Cultural heritage buildings with patterned azulejo tiles in Northern Portugal	YOLOv7; MobileNetV2; Simple Sequential Model; InceptionV3; Inception-ResNetV2; CNN-based classification; object detection	Precision; Recall; mAP; IoU; Accuracy for binary classification
Kulkarni et al. (2025) [45]	India	Kappachenikeshwara Temple, Veerabhadreshwara Temple, and Hazara Rama Temple, Hampi	U-Net GAN; U-Net generator; PatchGAN discriminator; CNN-based image-to-image reconstruction; adversarial loss; L1 reconstruction loss	SSIM; PSNR
Llamas et al. (2017) [46]	Spain	Architectural Heritage Elements Dataset	AlexNet; Inception V3; ResNet; Inception-ResNet-v2; CNN	Accuracy; Recall@2; Recall@3; loss/cost curves
Mercado et al. (2025) [47]	Philippines	San Sebastian Basilica	Random Forest Regressor; Gradient Boosting Regressor; XGBoost Regressor; physics-informed corrosion model	R ² ; MSE; RMSE; MAE; training time; inference time
Mesanza-Moraza et al. (2020) [48]	Spain	Church/Religious Buildings	Random Forest, Decision Tree, Cross-Validation	Overall reliability, Cohen's kappa, F1-measure, Precision, Recall, RU, RP, Confusion Matrix, Out-of-Bag error
Mishra et al. (2024) [49]	India	Dadi-Poti tombs, Hauz Khas Village, New Delhi	Custom YOLOv5s, Faster R-CNN with ResNet-101	YOLOv5 mAP = 93.7%, Precision = 85.9%, Recall = 91.8%; Faster R-CNN mAP = 85.1%
Pan et al. (2024) [50]	China	Taoping Qiang Village	KP-SG, KP-FCNN, PointNet, PointNet++, DGCNN, RandLA-Net	OA, IoU, mIoU
Patrucco et al. (2025) [51]	Italy	20th Century Industrial Concrete Heritage	CNN (DeepLab V3+); U-Net, FC-DenseNet	Accuracy, Mean IoU, Precision, Recall, F1-score
Pierdicca et al. (2020) [52]	Italy	Churches, Chapels, Cloisters, Porticos, Arcades, etc.	Improved DGCNN (with HSV color and normal features)	Accuracy, Precision, Recall, F1-score, IoU (Intersection over Union), Support, Confusion Matrix
Snehapriya and Umamageswari (2025) [53]	India	Ancient Indian Monuments	SSD, SVM, KNN, CNN, Random Forest	Accuracy, Precision, Recall, F1-score, ROC-AUC, Confusion Matrix, Execution Time
Song et al. (2025) [14]	China	Ming and Qing Dynasty Ancient Building Complex	YOLOv8	mAP (71.97–73.44%), AP per class, Precision, Recall, F1-score, Log-Average Miss Rate (LAMR), Confusion Matrix, Heatmap
Wu et al. (2025) [54]	China	Chinese Historical Architecture	Swin Transformer, Global Channel-Spatial Attention, Transformer	Accuracy, Precision, Recall, F1-score
Xu and Chen (2025) [55]	China	Nanjing Qixia Temple Sheli Pagoda	D3ENet (YOLOv8-based)	Precision, Recall, F1-score, mAP@0.5, mAP@0.5:0.95
Ying et al. (2025) [56]	Australia	Stone decay images from historical buildings in Melbourne; rock heritage site in Western Victoria; Royal Exhibition Building, Melbourne	Fine-tuned SAM, YOLOv8 prompts, U-Net, U-Net++, YOLOv8-seg, FastSAM, MobileSAM, SAM2, COLMAP SfM/MVS	IoU, DSC, inference time, mean reprojection error, visual inspection
Yu et al. (2022) [57]	China	Dunhuang Mogao Grottoes	PConv, EdgeConnect, SSD300, RetinaNet, GHM RetinaNet, FSAF, Faster R-CNN, Libra Faster R-CNN, Libra RetinaNet, YOLO v5-S, Cycle-GAN, CUT, DualAST, AdaIN, WCT, Avatar-Net	PSNR, SSIM, HaarPSI, TV, mAP/AP, deception rate, user study
Yu et al. (2025) [58]	China	Historic villages, Great Wall, traditional architecture	YOLOv11, IODA	Precision, Recall, mAP50, mAP50-95

Conventional ML/DL studies are summarized using measured performance metrics such as accuracy, F1-score, IoU, mAP, RMSE, or MAE where these were reported. By contrast, the ChatGPT-based pathology diagnosis study was classified as LLM-based reasoning with expert validation; its reported confidence or expert agreement was not treated as benchmarked supervised-learning accuracy.

Machine learning methods primarily included convolutional neural networks, random forests, support vector machines, k-nearest neighbors, YOLO series models, PointNet/PointNet++, DGCNN, U-Net, GAN, Transformer, SAM, and NeRF. Based on the research applications, these methods were mainly used for tasks such as architectural heritage image classification, object detection, semantic segmentation, point cloud recognition and segmentation, image restoration, 3D reconstruction, damage identification, structural condition prediction, HBIM-assisted modeling, and multimodal diagnosis. The evaluation metrics mainly include Accuracy, Precision, Recall, F1-score, IoU, mAP, SSIM, PSNR, RMSE, and MAE, indicating that existing research not only focuses on model recognition, classification, and detection performance, but also gradually attaches importance to multi-dimensional evaluation results such as 3D reconstruction quality, image restoration effect, and prediction error.

3.2. Thematic Synthesis of Application Domains

3.2.1. Point Cloud-Based Semantic Segmentation and HBIM Reconstruction

In architectural cultural heritage preservation, supervised semantic segmentation is increasingly becoming a key way to enhance the efficiency of digital analysis. This approach applies a machine learning model to point cloud data to recognize old building elements such as walls and more. As a classic supervised learning method, random forest has been widely applied in classification tasks involving elements such as church domes and stone columns. For instance, a research team conducted segmentation experiments on the point cloud data of the “Sacromonte Calvario di Domodossola” chapel and the ArCH dataset. By extracting geometric covariance features—such as anisotropy and planarity—the RF method achieved satisfactory recognition results: for Chapel 6, the overall accuracy reached 0.976, with a weighted F1-score of 0.977; on Scenes A and B of the ArCH dataset, the overall accuracies were 0.937 and 0.874, respectively [37]. In another study focusing on the stone walls of the San Vicentejo Hermitage in Spain, an RF model utilized 33 attribute variables—encompassing geometric dimensions and contour features—to distinguish between dressed stones and rough stones; during the training phase, the overall reliability reached as high as 99.662%, with an F1-measure of 0.999. However, upon entering the validation phase, due to the problem of class imbalance, the producer’s accuracy for coarse stone identification dropped to 48%, while the overall validation accuracy stood at 94.87% [48]. According to the studies, the performance of the RF method is great in some cases. Nevertheless, this performance highly depends on the quality of the supplied manually annotated samples and the soundness of the feature design.

On the contrary, deep learning methods like dynamic graph convolutional neural networks show better generalization ability in segmentation tasks through feature extraction. The DGCNN model was modified for scene classification by incorporating normal vectors and HSV color encoding. The enhanced model was tested on ArCH dataset that contains 11 scenes of Sacri Monti such as Ghiffa and Varallo. Based on the obtained results, the enhanced model (XYZ + HSV + Norm) got a mean score of 0.918 on the Trompone Church scene. Moreover, this proves to be a significant improvement as the original DGCNN got only 0.897. Furthermore, with also other comparative models that took that scene, the enhanced model clearly outperformed them all. Also, the cross-validation score on unseen scenes got 0.825 and F1-score 0.814. Most noteworthy, the precision and IoU also got better

with reference to the recognition of vaults, columns, and stairs [52]. Despite its advantages, DGCNN is still not robust enough for cross-dataset generalization. For example, in the Sacrimonti dataset, the overall accuracy for Chapels 3, 6, and 7 are merely 0.738, 0.761, and 0.628, respectively, which indicates a strong reliance on the richness and diversity of the training data [37].

In response to the adaptability problems caused by multi-source heterogeneous data, scholars have additionally suggested a novel neural network architecture, namely the KP-SG model, to promote the recognition of historical architectural elements. On the dataset of Taoping Village drone photogrammetry and terrestrial laser scanning, and specifically on the KP-SG model, the 11 classes were able to be accurately segmented thanks to the spatial feature enhancement module and global feature aggregation layer. In particular, the “buildings” and “vegetation” recognition rates were 81% and 83%, respectively. Moreover, the mean Intersection over Union (IoU) increased 2.53 percentage points from the KP-FCNN baseline, yielding an overall accuracy of 84.61%. Experiments showed that the SFE module is critical in capturing local geometric structure (small components). Without it, the mIoU reduced from 53.03% to 51.20% [50]. In addition, the shape of the input block affects the segmentation quality. Spheres, when used as input blocks and having a similar volume, avoid this segmentation bias at edges and corners.

Model performance also varied according to architecture, training data, and acquisition conditions. In the Bologna Porticoes dataset, the RF-based method achieved an average F1-score of 0.93 for wall and window classes during internal testing [20]. However, performance declined when the model was applied to external datasets with greater scene diversity. The weighted F1-score decreased to 0.89 in Vicinanza Maggiore, another Bologna district, and to 0.78 in Trento Square. Differences in architectural character, window morphology, and ornamentation increased class confusion and contributed to this decline.

Feature engineering also played an important role in improving generalization. In a study aimed at segmenting the point cloud of the Palacio de Sástago in Spain [36], an algorithm was trained with different geometric features such as verticality and planarity. The authors evaluated their features using the mean impurity decrease metric, achieving a weighted F1-score of 0.9710 on the ground floor using only 15 features. When the authors used their customized set of six ad hoc features combined with the Z coordinate, the weighted F1-scores for the ground and first floors were 0.9200 and 0.9018, respectively. This clearly shows that feature optimization helps reduce feature engineering complexity. Nonetheless, the issue of data imbalance continues to restrict the generalization performance of the model.

In the preservation of architectural cultural heritage, it was shown that unsupervised and weakly supervised learning methods can be effective approaches to alleviate problems relating to scanty labeled data and environmental interference. To counter the effect of lighting changes on color analysis, HSV clustering was used to transform the RGB color space of point clouds to the HSV space; this separation of chromaticity and lightness led to successful recognition of decay morphologies in different natural lighting conditions [41]. The approach does not necessitate manual annotation or predetermined parameters, it automatically segments the point clouds through hierarchical clustering to accurately identify defects such as damp areas, biofilm, biological colonization and spots/sediments. In various validation cases of interior low-light vaults and exterior high-light facades, the overall accuracy was above 0.8 and the F1-score was more than 0.7. Importantly, even in the most complex cases where the color ranges of biofilms and spots overlapped, the method remained able to discriminate robustly which supervised methods often fail to do so due to the difficulties in annotating the data.

When assessing façade vulnerabilities, the automatic segmentation and classification of stones using the K-means geometric classification algorithm is based on the stone area, aspect ratio, depth variability (median absolute deviation) and variability of surface normal horizontal angles [40]. By calculating the volume of rock mass hanging over the support surface, this method will develop a spalling risk assessment measure based on protrusion volume. In the application to the façade of the Pitti Palace, the algorithm returned 0.89 accuracy, 0.94 F1-score and a median Intersection over Union (IoU) of 71.2%. It overcame the time-consuming and repetitive issue of manual visual inspection and performed stably under the complex courtyard lighting.

The investigation of past vaults has been found to introduce synthetic data-driven methods to combat the problem of shortage of labeled samples [32]. A handful of shape grammars and procedural modeling techniques were used by the researchers to design a synthetic point cloud dataset of six types of vaults; barrel vaults, groin vaults, etc., the dimension of the vaults was modified parametrically ranging from 2 to 5 m with a step size of 0.1 m to enhance the geometric diversity of the dataset. Classification experiments of purely synthetic data resulted in overall accuracy of 99%. Similarly, accuracy of 90% was achieved for semantic segmentation experiments. When testing the models on a real point cloud scene, specifically the Ducal Palace of Urbino, PointNet and DGCNN architectures achieved overall accuracies of 69% and 74% at overlap rates of 5×5 and 2×2 , respectively. This finding suggests that realistic synthetic data, when trained on the data set, can avoid the necessity of real data annotation.

Random sampling methods are a crucial way to alleviate computational problems caused when extracting features from large point clouds, which could provide vital key technical support for the real-time processing of poorly annotated scenarios (Table 2).

Table 2. Performance comparison of machine learning models for semantic segmentation of architectural heritage points clouds.

Study	Method	Dataset/Heritage Site	Key Metric	Value	Limitations
Supervised—Traditional Machine Learning					
Cao et al. (2022) [37]	Random Forest (RF)	Sacromonte Chapel 6	OA/F1	0.976/0.977	Dependent on manual annotation quality
Grilli & Remondino (2020) [20]	Random Forest (RF)	Bologna Porticoes → Trento Sq.	Weighted F1	0.93 → 0.78	Cross-dataset accuracy declines due to style variation
Buldo et al. (2024) [36]	RF + feature selection	Palacio de Sástago (Spain)	Weighted F1	0.971	Requires domain-specific feature engineering
Supervised—Deep Learning					
Pierdicca et al. (2020) [52]	DGCNN (enhanced)	ArCH (Trompone Church)	Accuracy/F1	0.918/0.814	Cross-dataset OA as low as 0.628
Pan et al. (2024) [50]	KP-SG	Taoping Village (drone + TLS)	OA/mIoU	84.6%/53.0%	Sensitive to input block shape (sphere vs. box)
Unsupervised/Weakly Supervised					
Galantucci et al. (2023) [41]	HSV hierarchical clustering	Interior vaults/exterior facades	OA/F1	>0.80/>0.70	Overlapping color ranges for biofilm & spots
Fiorini et al. (2024) [40]	K-means geometric	Palazzo Pitti façade	Accuracy/F1/mIoU	0.89/0.94/71.2%	Limited to geometric features only
Battini et al. (2023) [32]	PointNet/DGCNN (synthetic)	Ducal Palace of Urbino (real)	OA (synth/real)	99%/69–74%	Domain gap between synthetic and real data

Note: OA = overall accuracy; F1 = F1-score; mIoU = mean Intersection over Union; TLS = terrestrial laser scanning. Cross-dataset results are indicated by “→”.

Presently, the mapping of point cloud segmentation results against semantic objects in Historical Building Information Modeling (HBIM), such as those concerned with vault types and material properties, still presents a set of core challenges. The main ones manifest as mismatches of geometric rules, subjects of ontological hierarchies and breaks of semantic-geometric coupling. Victoria Andrea Cotella accentuates a significant gap in knowledge present in existing literature; at present, there are no studies systematically demonstrating that the results of the point cloud segmentation process can be translated into parameterized historical building component models automatically. On the other hand, historical constructions are frequently worn in their infills while losing materials and deforming during their service life. Therefore, they generally have distorted geometric properties. This is in fundamental conflict with the assumptions of orthogonality and verticality (e.g., the Manhattan World assumption) upon which existing automatic reconstruction algorithms are based, resulting in many incompatibilities of geometric rules. The geometric quality of today's AI-generated BIM models is also low. Their Level of Development (LOD) usually does not go beyond LOD 100. Furthermore, relevant case studies are dominated by simple office or residential buildings and do not effectively apply to complex curves or non-standard historical forms [59].

At the ontology and semantic correspondence levels, according to Valeria Croce, previous studies have not yet analyzed the conversion process from semantically annotated data to parametric information models (HBIM). Especially, how to accomplish effective transitions from unstructured point cloud data to parametric geometric reconstruction. The geometric mismatch is also evident when irregular heritage geometry is fitted within conventional BIM-based parametric modelling frameworks, which are mainly designed for planar or cylindrical elements with standardized dimensions and therefore do not easily accommodate the irregular and organic surfaces of heritage buildings. They suggest setting up a semantic bridge between realistic models and parameterized models, by keeping "category names" consistent, ensuring dual coherence in geometric and semantic correspondence, to avoid the problem of the semantic-geometric coupling breakdown. Nonetheless, it should be noted that the existing workflow still lacks a global algorithm to cover all steps. Thus, it requires the use of several different specialized software tools by researchers, including cloud compare, Revit, etc., which gives rise to data interoperability issues [4].

Post-analysis, however, Ferial Bounouioua et al.'s research [35] report points out serious deficiencies. Even though AI algorithms have been trained to identify architectural and structural elements, confusion often persists in sections located within spaces with large surface areas and height. This difficulty arises in accurately identifying the points with the associated attributes of each individual point. More often than not, this leads to the dragging and redirection of points manually. As of now, algorithms have not been sufficiently trained yet to autonomously distinguish various architectural components, especially those with details such as wood carvings or vault structural systems. In addition, these algorithms are currently designed primarily for indoor spaces and are not equipped to deal with the facade of the exterior as of now. While mapping segmentation results to BIM objects, involvement of extensive manual work inside the BIM software is bound to take place, as a full pipeline from point cloud labeling to the parametric modeling of historic components has not yet been basically automated.

3.2.2. Automated Damage Detection and Condition Assessment Across Material Types

It is very important to detect surface defects for the conservation of architectural heritage. For this detection, the exact localization of the defect would be required, such as crack, moss, peeling, stain, etc. As shown in Table 3, various deep learning models

are used to enhance detection accuracy and environmental adaptability by considering various materials (ceramic tiles, sandstone, wood structures) and their respective imaging conditions (drone aerial photography, ground scanning, handheld photography).

Table 3. Performance comparison of deep learning models for surface defect detection in architectural heritage.

Study	Model	Material/Heritage Site	Defect Types	mAP/Accuracy	Dataset Size	Limitations
Mishra et al. (2022) [49]	YOLOv5 (custom)	Dadi-Poti Mausoleum, Delhi	Discoloration, exposed bricks, cracks, spalling	mAP 93.7%	10,291 images	Outperforms Faster R-CNN; crack mAP = 98.9%
Yu et al. (2025) [58]	IODA (improved dense OD)	Wood, brick, stone, tile (multi-material)	High-density, blurred boundary damage	AP reported	Multi-material datasets	Prone to misclassifying real damage as background
Xu & Chen (2025) [55]	D3ENet (YOLOv8s backbone)	Ancient stone pagodas	Cracks, corrosion	mAP 58.3%	—	Performance drops significantly in rain/snow/night lighting
Karimi et al. (2024) [44]	YOLOv7 (custom) + MobileNetV2	Ceramic tiles (historic)	Missing tiles, cracks, pits	mAP 63.9%/ Acc. 96.7%	>5000 images	Tile joints misidentified as cracks
Song et al. (2025) [14]	YOLOv8 (UAV)	Blue-tiled roofs, Jiangnan region	5 damage types (incl. lichen)	mAP 73.4%	UAV (2–5 m altitude)	High false-negative rate for lichen (LAMR = 0.72)
Snehapriya & Umamageswari (2025) [53]	SSD (optimised)	Ancient buildings, India	Cracks, moss, leakage	96.9–97.5%	5624 images	95/2375 normal samples misclassified as defects

Note: mAP = mean Average Precision; LAMR = log average miss rate; OD = object detection; SSD = Single Shot Detector; UAV = unmanned aerial vehicle.

Mayank Mishra et al. [49] developed a real-time object detection model based on a custom YOLOv5 to automatically identify four types of defects in cultural heritage structures: discoloration, exposed bricks, cracks, and spalling. On a dataset of 10,291 images from the Dadi-Poti Mausoleum in Hauz Khas village, Delhi, the model achieved a mAP of 93.7%, with the highest mAP (98.9%) for crack detection. The training time was only 1 h and 53 min, significantly outperforming Faster R-CNN [49]. Yu et al. [58] proposed an improved dense object detection algorithm (IODA) for scenarios with blurred boundaries and high-density damage. This method reduces missed detections by introducing a super pixel-based loss function and utilizes a graph-based loss function to characterize the relationships between defects. Its effectiveness was validated on a dataset covering multiple materials including wood, brick, stone, and tile. However, confusion matrix analysis shows that the model is still prone to misclassifying real damage as background, mainly because it is difficult to distinguish between low-contrast disease and natural texture.

Xu et al. [55] proposed the D3ENet model by improving the backbone network of YOLOv8s. In the task of detecting cracks and corrosion in ancient stone pagodas, this model achieved a mAP of 58.3%, while reducing the number of parameters by approximately 2.5 million and improving the inference speed to 5.8 ms. However, experimental results showed that its detection performance significantly decreased under complex lighting conditions such as rainy and snowy nights, indicating that lighting factors have a significant impact on the identification of low-contrast defects. Karimi et al. [44] constructed a dataset containing over 5000 images (covering various lighting conditions such as shadows, clouds, and rain) and used a custom YOLOv7 model to detect four types of damage to tiles, achieving an overall mAP of 63.9%, with a detection accuracy of 90.7% for missing tiles.

They also applied MobileNet-v2 to a binary classification task, achieving an accuracy of 96.67%. The study found that, against a background of complex patterns, the model is prone to misidentifying tile joints as cracks or misidentifying specific patterns as pits, reflecting that it still faces challenges in texture differentiation.

Song et al. [14] used UAV aerial images (flying at altitudes of 2–5 m) and YOLOv8, and performed fine-grained detection of five types of damage to the blue-tiled roofs of ancient buildings in the Jiangnan region, achieving a mAP of 73.44%. However, the false negative rate was high in the lichen growth recognition task (LAMR of 0.72), mainly due to the high integration of its visual features with the texture of aged tiles and the lack of clear and consistent shape boundaries. Snehapriya et al. [53] validated the performance of the optimized SSD model on high-resolution images of ancient buildings in India. In a dataset of 5624 images, the model achieved detection accuracies of 96.9%, 97.5%, and 97.5% for cracks, moss, and leakage, respectively, with a single execution time of only 25 milliseconds, significantly outperforming traditional methods such as SVM and KNN. However, the confusion matrix shows that 95 out of 2375 normal samples were misclassified as defects, indicating that the model's boundary discrimination ability in complex contexts still has room for improvement.

In recent years, machine learning technology has made a series of exploratory advances in assessing material-level degradation in the protection of architectural cultural heritage [60,61]. For the quantitative characterization of corrosion rates in steel structures, Mercado et al. [47] developed a LoRa-based IoT hardware system that can predict atmospheric corrosion rates using only readily available environmental data such as temperature and relative humidity, effectively avoiding the dependence of traditional dose–response functions on pollutant data, which is often difficult to obtain in real-world scenarios. This study compared the performance of three ensemble regression models: random forest, gradient boosting, and XGBoost. The results showed that random forest performed best in prediction accuracy ($R^2 = 0.9913$, $MAE = 0.89 \mu\text{m}/\text{year}$), while XGBoost maintained high accuracy while exhibiting excellent computational efficiency (training time of only 2.62 s). Furthermore, the research team proposed a physical mechanism corrosion rate model that integrates electrochemical principles. This hybrid model merges a temperature-based Arrhenius relation with a power law dependence on relative humidity further enhancing the physical intuitiveness of the corrosion process with a combination of environmental sensor data and physical mechanism modeling. Based on this, it constructed a dual-output prediction framework combining a high-precision regression output with a classification output. The framework allows for forecasting of corrosion rates and real-time risk alerts. The performance results of the three ensemble regression models and the physical mechanism model considered in the study were summarized in Table 4 in terms of predictive accuracy, computational efficiency, input features, and interpretability.

The existing studies on specific identification of bioerosion in tropical hot humid environments were mainly for detecting disease caused by moss, lichen and mold. Gbran et al. were able to develop a hybrid diagnostic framework that employs an unsupervised hierarchical clustering algorithm, with the point cloud photogrammetric data processed after HSV transformation classified in high resolution. The analysis was able to determine the manifestation of certain diseases, namely, “biopatina” and “biocolonization” [42]. The research was validated in four typical hot and humid locations in Semarang, Indonesia, specifically the Vihara Buddhagaya Watugong temple and the Kota Lama underground structure. The findings revealed that the unsupervised model could exceed 85% accuracy on average across all locations, with an F1-score greater than 0.83, indicating good robustness to variable light and geometric conditions. The results of quantitative analysis indicate that Vihara Buddhagaya Watugong experienced 37.6% biocolonization

under high-humidity conditions (75–94% RH). On the other hand, Kota Lama Semarang experienced a 45.6% biocolonization rate.

Table 4. Comparison of ensemble regression models for atmospheric corrosion rate prediction in steel structures [47].

Model	R ²	MAE (µm/Year)	Training Time (s)	Input Features	Remarks
Random Forest	0.9913	0.89	—	Temperature, relative humidity	Highest prediction accuracy; best overall performance
Gradient Boosting	—	—	—	Temperature, relative humidity	Comparable accuracy; moderate computational cost
XGBoost	High (≈RF)	—	2.62	Temperature, relative humidity	Best computational efficiency; suitable for real-time deployment
Physical mechanism model	—	—	—	Arrhenius + RH power-law	Combines electrochemical principles with sensor data; enhances interpretability

Note: R² = coefficient of determination; MAE = mean absolute error; RH = relative humidity. All models were deployed on LoRa-based IoT hardware. The physical mechanism model was not evaluated under the same regression metrics but serves as an interpretable benchmark. “—” indicates data not explicitly reported in the source.

Figure 3 shows that the supervised random forest model performs better than the unsupervised HSV clustering method in all four sites, but the difference is small and shows the effectiveness of annotation-free methods in resource-poor field environments. Meanwhile, the study also found that while supervised random forest models have high overall accuracy (91.3%), they can confuse visually similar “biofilm” with “surface deposits/spots,” leading to a higher false positive rate. Furthermore, their performance is highly dependent on time-consuming manually labeled data. In addition, current models are mainly based on static snapshots for classification and lack the ability to model the continuous progression of biological growth stages, making it difficult to quantify the evolution of diseases over time [42].

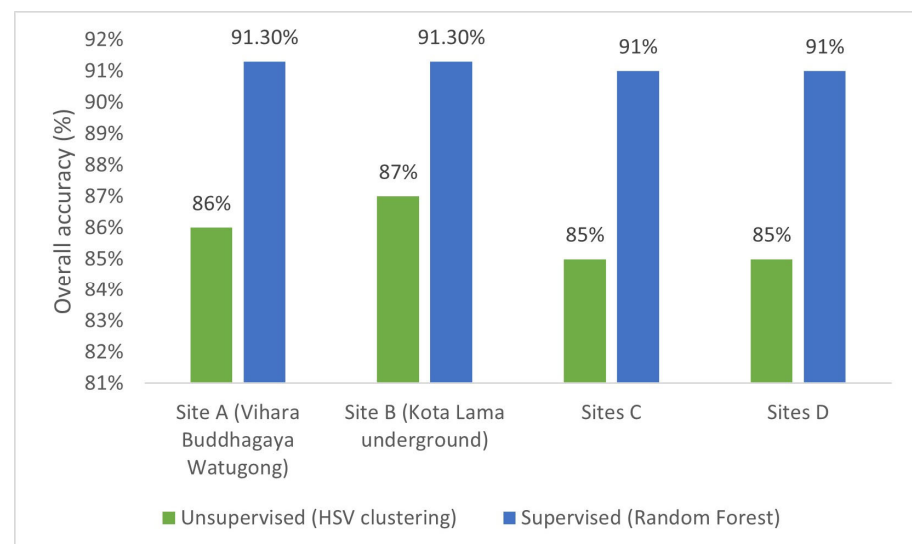


Figure 3. Supervised vs. unsupervised classification accuracy across heritage sites.

3.2.3. Intelligent System Integration: Digital Twins, Multimodal Fusion, and Interpretable Deployment

The core of building a digital twin foundation lies in achieving a three-layer coupling mechanism between the Internet of Things (IoT), Architectural Information Modeling (HBIM), and the digital twin itself. This mechanism integrates the technical paths of

different disciplines into a cohesive system through the collaborative design of data flow, control flow, and interaction flow. Specifically, the data flow follows the path of “real world → cloud → HBIM → virtual reality”; the control flow operates according to the closed-loop logic of “alarm → contingency plan → simulation → decision”; and the interaction flow forms a cyclical process of “expert annotation → model feedback → knowledge accumulation,” supporting the continuous accumulation and updating of knowledge.

At the data flow level, the system first collects multi-source data from the real world, uploads it to the cloud for processing, then integrates it into the HBIM platform, and finally maps it to the virtual reality visualization layer. Taking the Taoping Qiang Village practice as an example, the research team used UAV digital photogrammetry to obtain large-scale terrain and vegetation data (a total of 272 million points), and combined it with ground laser scanning to supplement high-precision data of building facades and hidden spaces (a total of 1.53 billion points). After semantic segmentation through the KP-SG neural network model, the data was imported into Revit to generate a parametric HBIM, and the multi-source datasets and semantic information were integrated based on the Cesium platform, ultimately constructing a digital twin platform that supports real-time spatial analysis and VR roaming [50]. Similarly, in the case of monitoring historical brick and stone towers, the experimental modal parameters (including frequency, mode shape, and damping ratio) obtained from environmental vibration tests were automatically projected onto the finite element model and iteratively updated through a genetic algorithm. Within 2000 iterations, the model kept the relative frequency error within 4%, thus ensuring the accuracy of the digital twin in dynamic response [62]. As illustrated in Figure 4, this three-layer coupling mechanism integrates IoT-based physical sensing, AI/ML-driven cloud processing and HBIM parametric modeling, and VR-enabled expert decision-making into a unified and closed-loop digital twin framework for architectural heritage conservation.

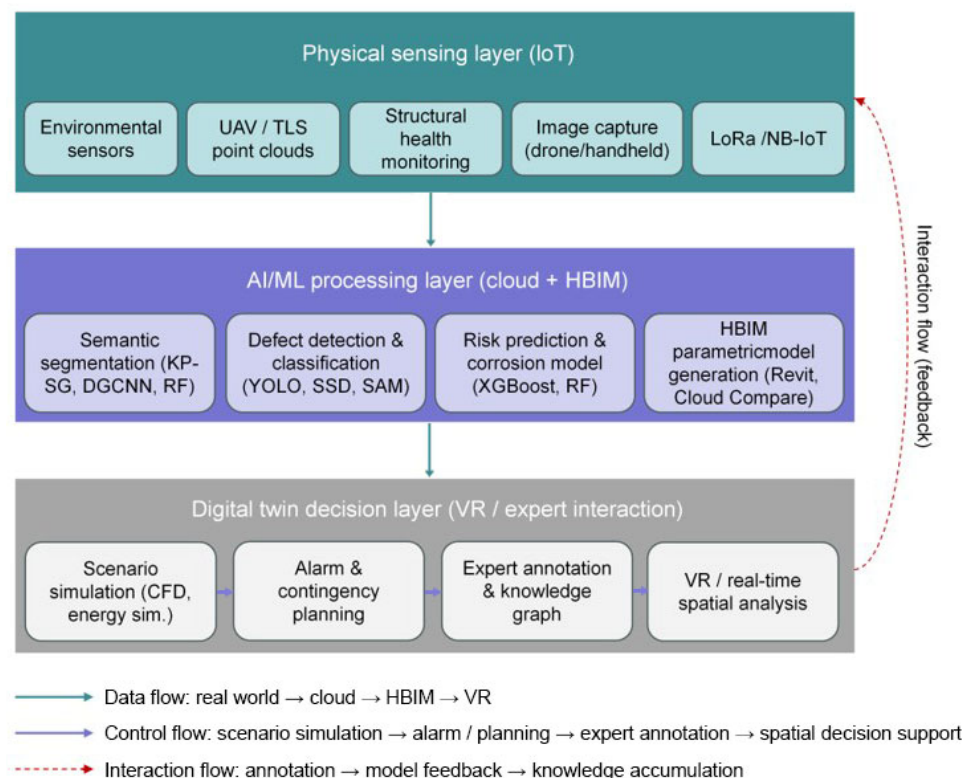


Figure 4. Three-layer coupling mechanism of the digital twin framework for architectural heritage conservation.

The control flow design strictly follows a closed-loop logic of “alarm → contingency plan → simulation → decision”. The system deploys IoT sensor networks such as LoRa (supporting NB-IoT and LoRaWAN protocols) to monitor environmental parameters such as temperature, humidity, and pollutants, as well as structural health indicators in real time. The collected data is transmitted to the cloud in real time and integrated into the HBIM platform. Subsequently, AI/ML models (such as decision trees, XGBoost, and Mask R-CNN) analyze historical and real-time data to identify anomalies and predict potential risks. Based on this, the digital twin platform combines computational fluid dynamics and dynamic energy simulation to perform scenario simulations, providing support for preventative maintenance planning and decision-making. For example, some research has used the Node-RED visualization tool to achieve a three-layer coupling from the physical sensing layer to the digital decision-making layer [63].

The interaction flow constructs a knowledge loop mechanism of “expert annotation → model feedback → knowledge accumulation”. In the digital twin model of heritage buildings, the knowledge graph framework integrates laser scanning data, IoT sensor input, and stakeholder expertise into a unified digital representation, embedding semantic attributes (such as building type, historical description, and path classification) to form a computable parametric component library. This process achieves a closed loop from expert cognition to model feedback and then to knowledge graph visualization. Taking Saudi Arabia as an example, the research team used IBM Watson and TensorFlow to process sensor data and combined it with Autodesk Revit to build a dynamic knowledge accumulation mechanism, providing sustainable knowledge support for the digital conservation of heritage buildings [64].

At present, in architectural heritage conservation, collaborative diagnosis based on multimodal data still remains dependent on unimodal analysis. Most studies rely on a single data source (images, point clouds, etc.) for lesion identification. Various attempts to integrate multimodal information have been carried out, such as merging ChatGPT with image-text analysis. These attempts have revealed bottlenecks including weak cross-modal alignment, large semantic gaps, and a rigid and superficial usage of domain knowledge.

The ChatGPT-based pathology diagnosis study should be interpreted differently from benchmarked supervised-learning experiments. As shown in Figure 5, when only one image was provided, the reported 40% value reflected agreement with expert judgment in a small diagnostic reasoning task, not a supervised model accuracy measured on a labeled test set. The generated solutions were standardized and general, but the model did not reliably retain image-specific details, such as whether the case involved dry-stone masonry. When three multi-view images and the ICOMOS architectural lesion terminology glossary were provided as multimodal input, the reported confidence of pathology interpretation increased to more than 90%. This value indicates increased diagnostic confidence and terminology consistency under expert-guided multimodal prompting, rather than true benchmarked classification accuracy. The comparison therefore suggests that visual diversity and domain-specific vocabulary can support LLM-based reasoning, but such outputs still require expert validation and should not be directly compared with ML/DL metrics such as mAP, F1-score, IoU, or supervised classification accuracy.

In addition, the study proposes a hybrid workflow in which human experts validate and refine AI-generated diagnostic suggestions. This expert-in-the-loop process should be understood as qualitative expert validation and reasoning support, rather than as a direct measurement of supervised-learning accuracy. It provides a practical route for combining automated LLM-based reasoning with domain expertise in multimodal fusion [34].

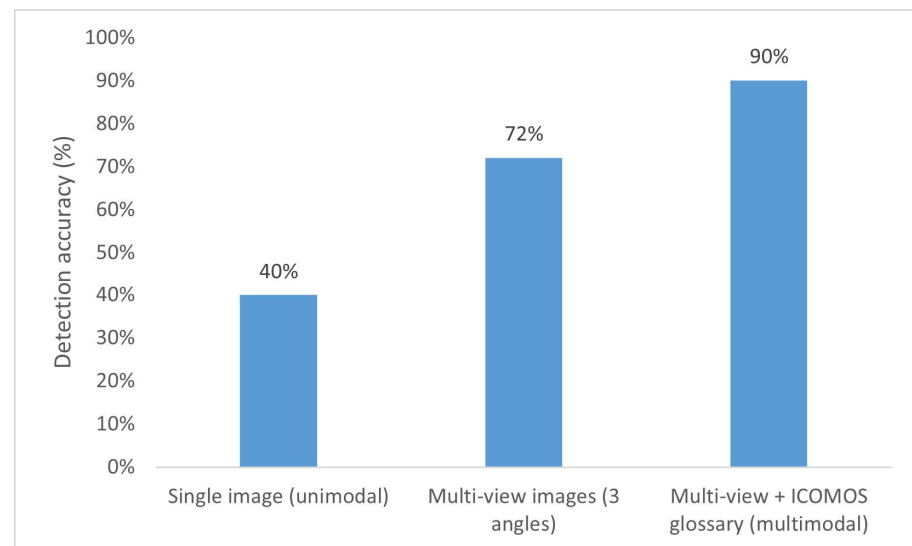


Figure 5. Improvement in ChatGPT-based pathology interpretation confidence through multimodal data integration.

4. Discussion

This systematic review shows that ML applications in architectural cultural heritage conservation have developed into a multi-level and interdisciplinary research landscape. The evidence can be organized into three closely connected levels: data interpretation, condition assessment, and system integration. At the first level, point cloud semantic segmentation and related intelligent parsing technologies are moving from traditional supervised learning toward unsupervised, weakly supervised, and synthetic-data-assisted paradigms that respond to scarce annotation resources. At the second level, multi-scale damage identification has become a central application direction, with deep learning models increasingly used for fine-grained recognition and localization of cracks, spalling, biological erosion, and material degradation. At the third level, HBIM and digital twin frameworks are emerging as integrative platforms that connect multimodal data, monitoring systems, and conservation decision support. Across these levels, however, the review also identifies persistent barriers in model generalization, semantic-BIM transformation, multimodal fusion, and real-world deployment.

The use of machine learning to preserve our architectural cultural heritage has many specificities that make it different from more broadly studied fields like computer vision or remote sensing. In the task of semantic segmentation of point clouds, it has been revealed in the study that deep learning models have provided great success in general datasets [65,66]. However, they fail to generalize well in the particular subdomain of cultural heritage. The geometric shapes, structural systems, and ornamental styles of historical buildings are highly heterogeneous, regional, and distinctive. Thus, models trained on a particular scene or dataset cannot be transferred to buildings that drastically differ in style [67]. This is far removed from the standardized repetitive nature of objects in a city environment or self-driving scenarios [68,69]. Thus, the creation of benchmark datasets for cultural heritage [52,70] and the use of synthetic data to make the model robust [71,72] have become prominent methods to cope with the dual problem of ‘data hunger’ and ‘scenario diversity’.

This paper reveals a crucial technological bottleneck in cultural heritage digitization research. We define this as the “semantic-BIM transformation bottleneck”. Presently, research on point cloud semantic segmentation is mostly limited to “labeling point clouds” [73,74]. There is a considerable technological gap concerning how to automatically transform these discrete, semantically labeled point clusters into historic building information model

(HBIM) objects with topological relationships, geometric constraints, and parametric attributes. In other sectors (such as industrial manufacturing and modern architecture), the level of automation of Scan-to-BIM processes is considerably higher [75]. The abnormal and irregular geometric shapes of historic buildings are essentially incompatible with the modeling logic of existing BIM software based on regular geometric primitives, thereby leaving the semantic recognition to parametric reconstruction problem unsolved [76]. This is not just a technological challenge; it illustrates a mismatch between the demands of computer vision research and architectural heritage conservation practices. Although some studies have begun to explore generating line drawings from point clouds [77] or propose semi-automated workflows [78], achieving end-to-end automated HBIM generation remains a distant goal. More explicitly, this semantic to HBIM automation gap is a major drawback that prevents end-to-end deployment in practical conservation workflows. Current systems can segment or semantically enrich point clouds, but they do not yet reliably convert those segmented point clusters into editable, parameterized HBIM objects with stable object identities, topology, level-of-detail definitions, material attributes, and conservation-oriented metadata. As a result, substantial manual intervention is still required for object verification, boundary correction, parametric modeling, ontology mapping, and integration with BIM authoring environments. The cited workflows should therefore be interpreted as semi-automatic research prototypes or task-specific pipelines rather than fully deployable end-to-end systems for routine heritage practice.

In damage identification, this study finds that the field is evolving from simple “defect detection” to “multidimensional condition assessment.” Early research focused more on identifying conditions with obvious geometric features such as cracks and spalling [79,80], while current research trends pay more attention to conditions with low contrast and similar textures (such as bio-erosion and salinization) as well as material-level degradation assessment. This is like semantic segmentation tasks in fields such as forestry [81,82] or bridge inspection [83], which all require identifying specific targets in complex contexts. However, the damage patterns of cultural heritage are more diverse and lack fixed patterns, placing higher demands on the model’s fine-grained identification capabilities and environmental adaptability. Furthermore, this study finds that the application of multimodal data fusion in this field is still in its early stages [84,85], with most studies still relying on a single data source. How to effectively integrate the rich texture of images, the precise geometry of point clouds, and the temporal data of sensors to achieve a deeper understanding from “what” to “why” and “how” is a direction that future research urgently needs to break through.

Finally, this study observes that the field is shifting from isolated algorithmic research to integrated system construction, with digital twins being the core vehicle of this trend. This aligns with the digital transformation paths of other industries, such as architecture and manufacturing [86]. However, the construction of digital twins for cultural heritage places greater emphasis on the accumulation and interaction of “knowledge.” Unlike industrial digital twins, which primarily focus on geometric and physical simulation, heritage digital twins require the integration of knowledge from multiple disciplines, including history, art, and materials science, and must provide interpretable and traceable decision support for conservation experts. Therefore, this study observes a growing emphasis on lightweight models and interpretability [31], indicating a matured phenomenon whereby research paradigms are shifting away from only chasing accuracy metrics toward human–machine collaboration and practical utility. The shift forms a bridge between the technology of artificial intelligence and the ethics of heritage conservation, ensuring that technology serves heritage conservation.

Limitations and Future Work

This review has several limitations. First, although a systematic search and screening process was used, relevant studies may have been missed because of database coverage, language restrictions, indexing differences, or rapidly emerging publications after January 2026. Second, the included studies were highly heterogeneous in terms of conservation tasks, data types, model architectures, validation settings, and reported metrics, which prevented formal quantitative meta-analysis. Third, methodological quality appraisal was based on reporting and design features relevant to ML-based heritage conservation, but some judgments inevitably involved reviewer interpretation. Fourth, the review primarily relied on published academic literature and may underrepresent industry projects, unpublished datasets, and conservation practice reports. Finally, because many included studies used single-site or small-scale datasets, the synthesized conclusions should be interpreted as evidence of current research tendencies rather than as definitive rankings of algorithms for all conservation contexts. Regarding practical readiness, the present evidence suggests that the cited systems are mainly at the proof-of-concept, laboratory-validation, or single-case demonstration stage. Their methodological shortcomings include limited cross-site validation, dependence on small or project-specific training datasets, inconsistent semantic taxonomies, weak interoperability between point-cloud processing and HBIM platforms, and insufficient testing under real conservation constraints such as incomplete scans, occlusions, irregular historic geometry, and changing documentation standards. These shortcomings currently block real-world practical implementation because they make model outputs difficult to audit, reproduce, update, and integrate into professional conservation decision-making.

In the future, machine learning research in the field of architectural heritage conservation should focus on breakthroughs in the following directions: First, constructing domain-specific, large-scale, high-quality, multimodal benchmark datasets and establishing standardized evaluation systems to promote fair comparison and reproducible research of algorithms. Second, focusing on overcoming the bottleneck of “semantic-BIM” transformation, exploring parametric reconstruction methods based on geometric and topological prior knowledge, or utilizing generative artificial intelligence technology to directly generate editable HBIM components from point clouds. Third, we will deepen the multimodal data fusion paradigm, researching how to deeply integrate visual, geometric, and temporal environmental data with unstructured information such as historical documents and expert knowledge to construct a heritage knowledge graph, achieving a leap from data perception to cognitive intelligence. Fourth, we will strengthen the application of explainable artificial intelligence (XAI) in heritage protection decision-support systems, ensuring that the algorithm’s operation is transparent and credible to protection experts, and promoting a new human–machine collaborative protection model. In the end, we will drive these technologies from the lab to the protection site, thus establishing a sustainable digital protection solution that covers the entire cycle of “data collection–intelligent analysis–decision support–intervention assessment.” Future studies should therefore report not only segmentation accuracy, but also the degree of automation after segmentation, the amount of manual editing required, the success rate of parametric object generation, interoperability with HBIM-based platforms, and validation in operational heritage management settings.

5. Conclusions

This systematic review synthesized 33 studies on machine learning in architectural cultural heritage conservation and constructed a three-level framework of data interpretation, damage identification, and system integration. The findings indicate that ML is gradually transforming architectural heritage conservation from an experience-driven and

reactive model toward data-driven, preventive, and decision-support-oriented practice. At the micro level, point cloud semantic segmentation and image-based recognition improve the automation and precision of historical component identification and surface damage detection. At the meso level, multi-scale damage identification and material degradation assessment support dynamic perception and quantitative condition evaluation. At the macro level, HBIM and digital twin frameworks provide a basis for predictive conservation and full life-cycle management. Nevertheless, the evidence also shows that limited generalization, scarce labeled data, insufficient external validation, weak semantic-BIM conversion, and shallow multimodal fusion continue to restrict large-scale application.

Future research urgently needs to shift from an “algorithm-driven” to a “system-driven” and “knowledge-driven” approach. To maintain continuous optimization of model performance, there should be increasing considerations of cross-scenario generalization capabilities, domain knowledge embedding, and the construction of human–machine collaboration mechanisms. Efforts to create high-quality, multimodal benchmark datasets and to investigate parametric modeling paths that integrate geometric constraints and semantic rules should be prioritized to overcome a critical barrier in the automatic conversion of point clouds to HBIM. At the same time, strengthening the integration of explainable artificial intelligence and expert-in-the-loop approaches should improve model-decision transparency and credibility. Moreover, sensor data is collected from the environment and processed by platforms inside and outside the museum, such as coupled cloud-edge intelligence, enabling various technologies to be integrated. All in all, this study not only provides a systematic cognitive logic for the application of machine learning to architectural cultural heritage protection, but also points the way for future interdisciplinary research and practice. This is of great significance for promoting the development of this field towards a more mature, sustainable and practical stage.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/buildings16142745/s1>, Table S1: Strings used for the publications search. Table S2: Study-level methodological quality and reliability matrix.

Author Contributions: Conceptualization, K.L. and D.L.; methodology, K.L., B.D. and Z.L.; software, K.L. and B.D.; validation, K.L. and X.Z.; formal analysis, K.L. and B.D.; data curation, K.L., B.D. and X.Z.; writing—original draft preparation, K.L., B.D. and X.Z.; writing—review and editing, K.L. and Z.L.; visualization, K.L.; supervision, K.L. and H.K.; funding acquisition, K.L. and H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by BK21 Four Service Design based Glocal Social Innovation Educational Research Team in Dongseo University.

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author on request.

Conflicts of Interest: The authors declare that they have no competing interests.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine Learning
HBIM	Heritage Building Information Models
CNNs	Convolutional Neural Networks
DGCNN	Dynamic Graph Convolutional Neural Network
GANs	Generative Adversarial Networks
AI	Artificial Intelligence

References

1. Silverman, H. Heritage and authenticity. In *Palgrave*; Macmillan UK eBooks: London, UK, 2015; pp. 69–88. [\[CrossRef\]](#)
2. Granata, F.; Di Nunno, F. Artificial Intelligence models for prediction of the tide level in Venice. *Stoch. Environ. Res. Risk Assess.* **2021**, *35*, 2537–2548. [\[CrossRef\]](#)
3. Lanzoni, N. Cultural heritage protection versus social and economic development: Where does customary international law stand? *Int. J. Cult. Prop.* **2024**, *31*, 299–315. [\[CrossRef\]](#)
4. Croce, V.; Caroti, G.; De Luca, L.; Jacquot, K.; Piemonte, A.; Véron, P. From the semantic point cloud to Heritage-Building information modeling: A semiautomatic approach exploiting machine learning. *Remote Sens.* **2021**, *13*, 461. [\[CrossRef\]](#)
5. Sandak, J.; Sandak, A.; Legan, L.; Retko, K.; Kavčič, M.; Kosel, J.; Poohphajai, F.; Diaz, R.H.; Ponnuchamy, V.; Sajinčič, N.; et al. Nondestructive Evaluation of Heritage Object Coatings with Four Hyperspectral Imaging Systems. *Coatings* **2021**, *11*, 244. [\[CrossRef\]](#)
6. Teruggi, S.; Grilli, E.; Russo, M.; Fassi, F.; Remondino, F. A hierarchical machine learning approach for Multi-Level and Multi-Resolution 3D point cloud classification. *Remote Sens.* **2020**, *12*, 2598. [\[CrossRef\]](#)
7. Colace, F.; Limongiello, M.; Lorusso, A.; Pellegrino, M.; Santaniello, D.; Santoriello, A. Digital twin for cultural heritage: A computational approach to predictive conservation. *Digit. Appl. Archaeol. Cult. Herit.* **2026**, *40*, e00519. [\[CrossRef\]](#)
8. Abusaleh, S.W. Enhancing preservation outcomes for architectural heritage buildings through machine learning-driven future search optimization. *Asian J. Civ. Eng.* **2024**, *25*, 5277–5292. [\[CrossRef\]](#)
9. Grilli, E.; Dinunno, D.; Petrucci, G.; Remondino, F. From 2D to 3D Supervised Segmentation and Classification for Cultural Heritage Applications. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci./Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 399–406. [\[CrossRef\]](#)
10. Grilli, E.; Remondino, F. Classification of 3D digital Heritage. *Remote Sens.* **2019**, *11*, 847. [\[CrossRef\]](#)
11. Matrone, F.; Grilli, E.; Martini, M.; Paolanti, M.; Pierdicca, R.; Remondino, F. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 535. [\[CrossRef\]](#)
12. Grilli, E.; Özdemir, E.; Remondino, F. Application of Machine and Deep Learning Strategies for the Classification of Heritage Point Clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci./Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *42*, 447–454. [\[CrossRef\]](#)
13. Feng, C.-C.; Guo, Z. A hierarchical approach for point cloud classification with 3D contextual features. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 5036–5048. [\[CrossRef\]](#)
14. Song, H.; Chen, Y.; Zheng, L. The Non-Destructive Testing of Architectural heritage surfaces via Machine Learning: A case study of flat tiles in the Jiangnan region. *Coatings* **2025**, *15*, 761. [\[CrossRef\]](#)
15. Muccioli, M.F.; Di Giuseppe, E.; D’Orazio, M. Decay detection and classification on architectural heritage through machine learning methods based on hyperspectral images: An overview on the procedural workflow. In *Proceedings of the 11th International Conference of Ar.Tec. (Scientific Society of Architectural Engineering); Lecture Notes in Civil Engineering*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 507–525. [\[CrossRef\]](#)
16. Um-e-Habiba; Shaikh, F.K.; Chowdhry, B.S. Cultural Heritage Monitoring and Predictive Maintenance using Internet of Things: Assessment and Future Aspects. *J. Mob. Multimed.* **2025**, *20*, 1211–1250. [\[CrossRef\]](#)
17. Zou, H.; Ge, J.; Liu, R.; He, L. Feature recognition of regional architecture forms based on Machine learning: A case study of Architecture heritage in Hubei Province, China. *Sustainability* **2023**, *15*, 3504. [\[CrossRef\]](#)
18. Karadag, İ. Machine learning for conservation of architectural heritage. *Open House Int.* **2022**, *48*, 23–37. [\[CrossRef\]](#)
19. Güzelci, O.Z.; Alaçam, S.; Bekiroğlu, B.; Karadag, I. A machine learning-based prediction model for architectural heritage: The case of domed Sinan mosques. *Digit. Appl. Archaeol. Cult. Herit.* **2024**, *35*, e00370. [\[CrossRef\]](#)
20. Grilli, E.; Remondino, F. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 379. [\[CrossRef\]](#)
21. Zhang, K.; Teruggi, S.; Ding, Y.; Fassi, F. A Multilevel Multiresolution Machine Learning Classification Approach: A generalization test on Chinese heritage architecture. *Heritage* **2022**, *5*, 3970–3992. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Flemming, K.; Noyes, J. Qualitative Evidence Synthesis: Where Are We at? *Int. J. Qual. Methods* **2021**, *20*, 1609406921993276. [\[CrossRef\]](#)
24. Gupta, A. *Qualitative Methods and Data Analysis Using ATLAS.TI*; Springer: Berlin/Heidelberg, Germany, 2023. [\[CrossRef\]](#)
25. Levitt, H.M. How to conduct a qualitative meta-analysis: Tailoring methods to enhance methodological integrity. *Psychother. Res.* **2018**, *28*, 367–378. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Cha, Y.-J.; Ali, R.; Lewis, J.; Büyüköztürk, O. Deep learning-based structural health monitoring. *Autom. Constr.* **2024**, *161*, 105328. [\[CrossRef\]](#)

27. Sharma, H. Applications of Artificial intelligence and Machine Learning in the preservation and analysis of heritage Structures: A Comprehensive review. *Arch. Comput. Methods Eng.* **2025**, *33*, 3001–3033. [\[CrossRef\]](#)
28. Şenol, H.İ.; Gökğöz, T. Integration of Building Information Modeling (BIM) and Geographic Information System (GIS): A new approach for IFC to CityJSON conversion. *Earth Sci. Inform.* **2024**, *17*, 3437–3454. [\[CrossRef\]](#)
29. Hangloo, S.; Arora, B. Multimodal fusion techniques: Review, data representation, information fusion, and application areas. *Neurocomputing* **2025**, *649*, 130827. [\[CrossRef\]](#)
30. Bahrami, M.; Albadvi, A. Deep learning for identifying Iran's cultural heritage buildings in need of conservation using image classification and Grad-CAM. *J. Comput. Cult. Herit.* **2023**, *17*, 1–20. [\[CrossRef\]](#)
31. Rahimi, F.B.; Demers, C.M.H.; Dastjerdi, M.R.K.; Lalonde, J.-F. Agile digitization for historic architecture using 360° capture, deep learning, and virtual reality. *Autom. Constr.* **2025**, *171*, 105986. [\[CrossRef\]](#)
32. Battini, C.; Ferretti, U.; De Angelis, G.; Pierdicca, R.; Paolanti, M.; Quattrini, R. Automatic generation of synthetic heritage point clouds: Analysis and segmentation based on shape grammar for historical vaults. *J. Cult. Herit.* **2023**, *66*, 37–47. [\[CrossRef\]](#)
33. Boesgaard, C.; Hansen, B.V.; Kejser, U.B.; Møllerup, S.H.; Ryhl-Svendsen, M.; Torp-Smith, N. Prediction of the indoor climate in cultural heritage buildings through machine learning: First results from two field tests. *Herit. Sci.* **2022**, *10*, 1–12. [\[CrossRef\]](#)
34. Bouchachi, M.; Jiménez-Delgado, A.; De-Gracia-Soriano, P.; Nemroudi, R. Architectural Heritage and Artificial Intelligence: Diagnosis and solutions proposed by CHATGPT for Algerian Historical Monuments. *Heritage* **2025**, *8*, 139. [\[CrossRef\]](#)
35. Bounouioua, F.; Saffidine, D.R.; Korichi, A. An Enhanced HBIM Framework Integrating Advanced Technologies to strengthen the Cultural Heritage. *J. Inf. Technol. Constr.* **2025**, *30*, 570–602. [\[CrossRef\]](#)
36. Buldo, M.; Agustín-Hernández, L.; Verdoscia, C. Semantic Enrichment of Architectural Heritage Point Clouds Using Artificial Intelligence: The Palacio de Sástago in Zaragoza, Spain. *Heritage* **2024**, *7*, 6938–6965. [\[CrossRef\]](#)
37. Cao, Y.; Teruggi, S.; Fassi, F.; Scaioni, M. A comprehensive understanding of machine learning and deep learning methods for 3D architectural cultural heritage point Cloud semantic segmentation. In *Communications in Computer and Information Science*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 329–341. [\[CrossRef\]](#)
38. Casillo, M.; Colace, F.; Gaeta, R.; Lorusso, A.; Santaniello, D.; Valentino, C. Revolutionizing cultural heritage preservation: An innovative IoT-based framework for protecting historical buildings. *Evol. Intell.* **2024**, *17*, 3815–3831. [\[CrossRef\]](#)
39. Fang, T.; Hui, Z.; Rey, W.P.; Yang, A.; Liu, B.; He, Y. Machine Learning-Based Crack Detection Methods in Ancient Buildings. In *Proceedings of the 2024 Guangdong-Hong Kong-Macao Greater Bay Area International Conference on Digital Economy and Artificial Intelligence (DEAI '24)*, Hongkong, China, 19–21 January 2024; pp. 885–890. [\[CrossRef\]](#)
40. Fiorini, L.; Conti, A.; Pellis, E.; Bonora, V.; Masiero, A.; Tucci, G. Machine Learning-Based Monitoring for Planning Climate-Resilient Conservation of built Heritage. *Drones* **2024**, *8*, 249. [\[CrossRef\]](#)
41. Galantucci, R.A.; Musicco, A.; Verdoscia, C.; Fatiguso, F. Machine Learning for the Semi-Automatic 3D decay segmentation and mapping of heritage assets. *Int. J. Archit. Herit.* **2023**, *19*, 389–407. [\[CrossRef\]](#)
42. Gbran, H.; Rukayah, S.; Suprapti, A.; Pandelaki, E.E. A Hybrid Framework Employing Deep Learning for 3D Decay Segmentation and Adaptive Mapping of Heritage Structures: Insights from an Experiment. *Int. Soc. Study Vernac. Settl.* **2025**, *12*, 101–135. [\[CrossRef\]](#)
43. Gokak, S.; Khichade, P.; Heggalagi, N.; Billowria, A.; Hegde, S. Defect Detection and Classification of Cultural Heritage Buildings Using Deep Learning. In *Proceedings of the 3rd International Conference on Futuristic Technology*, Pune, India, 21–22 February 2025; pp. 619–626. [\[CrossRef\]](#)
44. Karimi, N.; Mishra, M.; Lourenço, P.B. Deep learning-based automated tile defect detection system for Portuguese cultural heritage buildings. *J. Cult. Herit.* **2024**, *68*, 86–98. [\[CrossRef\]](#)
45. Kulkarni, A.; Naduvanamath, P.; Naik, G.; Totad, S.; Kulkarni, U.; Hegde, S. Deep Learning Techniques for Archaeological Image Restoration. In *Proceedings of the 3rd International Conference on Futuristic Technology*; Science and Technology Publications: Setúbal, Portugal, 2025; Volume 3, pp. 26–32. [\[CrossRef\]](#)
46. Llamas, J.; Leronés, P.M.; Medina, R.; Zalama, E.; Gómez-García-Bermejo, J. Classification of architectural heritage images using deep learning techniques. *Appl. Sci.* **2017**, *7*, 992. [\[CrossRef\]](#)
47. Mercado, R.J.M.; Kabeer, M.; Al-Obaidy, H.; Nordin, R. Corrosion Risk Estimation for Heritage Preservation: An Internet of Things and Machine Learning approach using temperature and humidity. *arXiv* **2025**, arXiv:2510.02973.
48. Mesanza-Moraza, A.; García-Gómez, I.; Azkarate, A. Machine Learning for the Built Heritage Archaeological Study. *J. Comput. Cult. Herit.* **2020**, *14*, 1–21. [\[CrossRef\]](#)
49. Mishra, M.; Barman, T.; Ramana, G.V. Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage. *J. Civ. Struct. Health Monit.* **2022**, *14*, 103–120. [\[CrossRef\]](#)
50. Pan, X.; Lin, Q.; Ye, S.; Li, L.; Guo, L.; Harmon, B. Deep learning based approaches from semantic point clouds to semantic BIM models for heritage digital twin. *Herit. Sci.* **2024**, *12*, 1–17. [\[CrossRef\]](#)
51. Patrucco, G.; Setragno, F.; Spanò, A. Synthetic Training Datasets for Architectural Conservation: A Deep learning approach for decay Detection. *Remote Sens.* **2025**, *17*, 1714. [\[CrossRef\]](#)

52. Pierdicca, R.; Paolanti, M.; Matrone, F.; Martini, M.; Morbidoni, C.; Malinverni, E.S.; Frontoni, E.; Lingua, A.M. Point Cloud semantic segmentation using a deep learning framework for cultural heritage. *Remote Sens.* **2020**, *12*, 1005. [\[CrossRef\]](#)
53. Snehapriya, M.; Umamageswari, A. Protecting Historical Treasures: Deep learning for structural health assessment. *Int. J. Basic Appl. Sci.* **2025**, *14*, 840–852. [\[CrossRef\]](#)
54. Wu, J.; Ying, Y.; Tan, Y.; Liu, Z. Innovative Framework for historical Architectural recognition in China: Integrating SWIN Transformer and Global Channel–Spatial Attention Mechanism. *Buildings* **2025**, *15*, 176. [\[CrossRef\]](#)
55. Xu, S.; Chen, H. Deep learning and digital twin integration for structural damage detection in ancient pagodas. *Sci. Rep.* **2025**, *15*, 28408. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Ying, W.; Khoshelham, K.; Kemp, J. Assessment of rock and stone decay in heritage sites using machine learning. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, X-G-2025, 1019–1026. [\[CrossRef\]](#)
57. Yu, T.; Lin, C.; Zhang, S.; Wang, C.; Ding, X.; An, H.; Liu, X.; Qu, T.; Wan, L.; You, S.; et al. Artificial intelligence for Dunhuang Cultural Heritage Protection: The project and the Dataset. *Int. J. Comput. Vis.* **2022**, *130*, 2646–2673. [\[CrossRef\]](#)
58. Yu, Q.; Yuan, X.; Xu, L. Cross-Material damage detection and analysis for architectural heritage images. *Buildings* **2025**, *15*, 3100. [\[CrossRef\]](#)
59. Cotella, V.A. From 3D point clouds to HBIM: Application of Artificial Intelligence in Cultural Heritage. *Autom. Constr.* **2023**, *152*, 104936. [\[CrossRef\]](#)
60. Basu, A.; Paul, S.; Ghosh, S.; Das, S.; Chanda, B.; Bhagvati, C.; Snasel, V. Digital Restoration of Cultural Heritage with Data-Driven Computing: A survey. *IEEE Access* **2023**, *11*, 53939–53977. [\[CrossRef\]](#)
61. Mishra, M. Machine learning techniques for structural health monitoring of heritage buildings: A state-of-the-art review and case studies. *J. Cult. Herit.* **2020**, *47*, 227–245. [\[CrossRef\]](#)
62. Standoli, G.; Salachoris, G.P.; Masciotta, M.G.; Clementi, F. Modal-based FE model updating via genetic algorithms: Exploiting artificial intelligence to build realistic numerical models of historical structures. *Constr. Build. Mater.* **2021**, *303*, 124393. [\[CrossRef\]](#)
63. Laohaviraphap, N.; Waroonkun, T. Integrating artificial intelligence and the internet of things in Cultural Heritage Preservation: A Systematic Review of risk management and environmental monitoring strategies. *Buildings* **2024**, *14*, 3979. [\[CrossRef\]](#)
64. Mazzetto, S. Integrating Emerging Technologies with Digital Twins for Heritage Building Conservation: An Interdisciplinary Approach with Expert Insights and Bibliometric Analysis. *Heritage* **2024**, *7*, 6432–6479. [\[CrossRef\]](#)
65. Xie, Y.; Tian, J.; Zhu, X.X. Linking points with labels in 3D: A review of Point Cloud Semantic Segmentation. *IEEE Geosci. Remote Sens. Mag.* **2020**, *8*, 38–59. [\[CrossRef\]](#)
66. Zhang, J.; Zhao, X.; Chen, Z.; Lu, Z. A review of Deep Learning-Based Semantic Segmentation for Point Cloud. *IEEE Access* **2019**, *7*, 179118–179133. [\[CrossRef\]](#)
67. Betsas, T.; Tsarpalis, H.; Georgopoulos, A. Assessing Generalization Capability of 3D Semantic Segmentation Algorithms using 3D Point Clouds of Cultural Heritage. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci./Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *48*, 111–118. [\[CrossRef\]](#)
68. Yan, H.; Lau, A.; Fan, H. Evaluating deep learning advances for point Cloud semantic segmentation in urban environments. *KN—J. Cartogr. Geogr. Inf.* **2025**, *75*, 3–22. [\[CrossRef\]](#)
69. Balado, J.; Martínez-Sánchez, J.; Arias, P.; Novo, A. Road Environment Semantic Segmentation with Deep Learning from MLS Point Cloud Data. *Sensors* **2019**, *19*, 3466. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Sridhar, M.; Paygude, A.; Pande, H.; Tiwari, P.S. 3DITA—A 3D Benchmark Dataset for Nagara-Style Indian Temple Architecture: India’s first point cloud dataset for semantic segmentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci./Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2025**, *48*, 1435–1441. [\[CrossRef\]](#)
71. Ma, J.W.; Czerniawski, T.; Leite, F. Semantic segmentation of point clouds of building interiors with deep learning: Augmenting training datasets with synthetic BIM-based point clouds. *Autom. Constr.* **2020**, *113*, 103144. [\[CrossRef\]](#)
72. Pierdicca, R.; Mameli, M.; Malinverni, E.S.; Paolanti, M.; Frontoni, E. Automatic generation of point Cloud synthetic dataset for historical building representation. In *Augmented Reality, Virtual Reality, and Computer Graphics; Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 203–219. [\[CrossRef\]](#)
73. Singh, D.P.; Yadav, M. Deep learning-based semantic segmentation of three-dimensional point cloud: A comprehensive review. *Int. J. Remote Sens.* **2024**, *45*, 532–586. [\[CrossRef\]](#)
74. Zhang, R.; Wu, Y.; Jin, W.; Meng, X. Deep-Learning-Based Point Cloud Semantic Segmentation: A survey. *Electronics* **2023**, *12*, 3642. [\[CrossRef\]](#)
75. Li, L.; Chen, J.; Su, X.; Han, H.; Fan, C. Deep learning network for indoor point cloud semantic segmentation with transferability. *Autom. Constr.* **2024**, *168*, 105806. [\[CrossRef\]](#)
76. Wang, F.; Yang, Z.; Liu, Y.; Zhang, M.; Ye, Q.; Luo, J. Reconstructing Heritage Interiors from Point Clouds via Architectural Element Segmentation. *Photogramm. Eng. Remote Sens.* **2026**. [\[CrossRef\]](#)
77. Dong, S.; Wu, D.; Kong, W.; Liu, W.; Xia, N. Research on Intelligent Generation of Line Drawings from Point Clouds for Ancient Architectural Heritage. *Buildings* **2025**, *15*, 3341. [\[CrossRef\]](#)

78. Buldo, M.; Agustín-Hernández, L.; Verdoscia, C.; Tapolare, R. A Scan-to-BIM Workflow Proposal for Cultural Heritage. Automatic Point Cloud Segmentation and Parametric-Adaptive Modelling of Vaulted Systems. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci./Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *48*, 333–340. [[CrossRef](#)]
79. Du, L.; Wang, Y. Bi-YOLO: A novel object detection network and dataset for components of China heritage buildings. *J. Build. Eng.* **2024**, *97*, 110817. [[CrossRef](#)]
80. Li, Q.; Zhang, G.; Yang, P. CL-YOLOV8: Crack Detection Algorithm for Fair-Faced Walls based on Deep Learning. *Appl. Sci.* **2024**, *14*, 9421. [[CrossRef](#)]
81. Ruoppa, L.; Oinonen, O.; Taher, J.; Lehtomäki, M.; Takhtkeshha, N.; Kukko, A.; Kaartinen, H.; Hyypä, J. Unsupervised deep learning for semantic segmentation of multispectral LiDAR forest point clouds. *ISPRS J. Photogramm. Remote Sens.* **2025**, *228*, 694–722. [[CrossRef](#)]
82. Krisanski, S.; Taskhiri, M.S.; Aracil, S.G.; Herries, D.; Turner, P. Sensor agnostic semantic segmentation of structurally diverse and complex forest point clouds using deep learning. *Remote Sens.* **2021**, *13*, 1413. [[CrossRef](#)]
83. Yang, X.; Del Rey Castillo, E.; Zou, Y.; Wotherspoon, L. Semantic segmentation of bridge point clouds with a synthetic data augmentation strategy and graph-structured deep metric learning. *Autom. Constr.* **2023**, *150*, 104838. [[CrossRef](#)]
84. Xie, Z.; Liu, H.; He, Y.; Shi, Y.; Yu, P.; Ai, J.; Zhong, L. Cross modal networks for point cloud semantic segmentation of Chinese ancient buildings. *npj Herit. Sci.* **2025**, *13*, 131. [[CrossRef](#)]
85. Zhang, R.; Li, G.; Li, M.; Wang, L. Fusion of images and point clouds for the semantic segmentation of large-scale 3D scenes based on deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 85–96. [[CrossRef](#)]
86. Yu, Y.; Verbree, E.; Van Oosterom, P.; Pottgiesser, U.; Peng, Y.; Poux, F. From comparison to integration: A workflow evaluation of 3D Gaussian splatting and LiDAR point cloud for modern architectural heritage. *Autom. Constr.* **2025**, *180*, 106509. [[CrossRef](#)]

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.