Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.2169

Research Article

Crowdsourcing-aided AI for multi-label cultural heritage damage classification

Himanshu SUYAL[1,2,*], **Avtar SINGH**[1]

[1]*Department of Computer Science and Engineering, Dr. B. R. Ambedkar NIT Jalandhar, Punjab, 144008, India*
[2]*School of Computer Science, Engineering and Technology, Bennett University, Uttar Pradesh, 201310, India*

ARTICLE INFO

Article history
Received: 07 February 2025
Revised: 01 June 2025
Accepted: 01 July 2025

Keywords:
Artificial Intelligence; Classification; Crowdsourcing; Deep Learning

ABSTRACT

Cultural heritage monuments have considerable historical, aesthetic and social value, yet they are fragile and increasingly threatened by natural disasters and human-induced degradation. The resulting structural damage is often complex, with several damage types overlapping on the same surface, and its accurate and timely classification is essential for effective restoration and preservation planning. The task is constrained by two limitations: (i) the scarcity of annotated damage data and (ii) the difficulty of obtaining consistent, high-quality labels from crowd workers, whose cultural knowledge varies widely. To address both limitations, we propose a new human-in-the-loop framework for multi-damage classification that combines machine and human intelligence. Images are classified by a hybrid deep neural network consisting of a DenseNet-169 backbone followed by Bi-LSTM and RNN layers, while an active learning strategy selects only the most informative samples for manual annotation, thereby reducing annotation effort while maximizing model performance. Five classes are distinguished: spalling, erosion, loss of material, other damage and no damage. When evaluated on a dataset of cultural heritage images collected in Uttarakhand, India, the model achieved an average classification accuracy of 97% and remained robust with limited labelled data. Compared with random sampling, active learning increased accuracy from 91% to 97%.

Cite this article as: Suyal H, Singh A. Crowdsourcing-aided AI for multi-label cultural heritage damage classification. Sigma J Eng Nat Sci 2026;44(5):3443–3460.

INTRODUCTION

Cultural Heritage Monuments (CHMs) are unique in that they combine aesthetic, historical, holistic and cultural values within a single structure, and their preservation is a challenge worldwide. Surface deterioration is difficult to monitor because it progresses slowly and because comparative data documenting a monument at different stages of damage are seldom available. Many CHMs are built from stone, wood or ancient mortars, which weather and lose strength over time. A small damaged area can gradually deepen and spread until a much larger section of the material is compromised. Such damage must therefore be identified early and precisely, which requires the structure to be inspected for different damage types. Traditionally, this is done through manual

***Corresponding author.**
*E-mail address: suyal.himanshu@gmail.com

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey
© Author. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

visual inspection, in which an inspector examines the CHM on site; this method is generally preferred for its accuracy. However, manual inspection is labour-intensive, costly and inconsistent, and it depends heavily on the availability and expertise of professionals. In some cases, manual inspection is also hindered because parts of the structure are difficult to reach. Manual inspection may itself cause further damage; consequently, technological advances have led to alternative solutions in which images are captured with drones or mobile devices and analysed later. These solutions enable automatic damage assessment at reduced cost. However, the accuracy of such assessment is highly sensitive to image quality, lighting conditions and obstructions, and obstructions may introduce data uncertainty even when drones or other technologies are used for image capture, although preprocessing techniques can improve data quality. Sensor-based damage detection has also been proposed [1, 2]. However, these methods are more expensive.

Advances in computer vision and deep learning can improve the accuracy of damage classification [3]. In AI-assisted inspection, digital images are analysed and interpreted so that anomalies can be detected [4, 5]. Machine learning algorithms can be used to extract features from image data and then classify different types of damage. Many contemporary methods are used to evaluate the structural integrity of CHMs alongside conventional nondestructive approaches [6].

Recent advances in deep learning and computer vision have enabled more scalable and automated damage classification in various domains, such as roads [7], bridges [8], construction [9] and modern buildings [10, 11]. Yiğit et al. [12] stressed the critical role of extensive digital documentation and offered an automated photogrammetry-based solution for recognising fractures in antiquated structures. Gezer et al. [13] investigated damage in the wooden structure of a historical mansion. Peng et al. [14] introduced an RBF-based deep learning compression method for the hyperspectral analysis of stone cultural heritage, which improved identification accuracy, efficiency and speed while maintaining high F1 scores. Liu [15] developed an AI-integrated digital platform for Chinese intangible cultural heritage and reported high predictive accuracy, an efficient user experience, and good performance in domain-specific entity recognition and knowledge coverage.

However, transferring these approaches to CHMs is not straightforward because of several specific difficulties. First, deep learning models typically require large annotated datasets, whereas CHM-specific damage data are scarce and costly to label. Second, annotation by domain experts does not scale, and because it relies on subjective judgment, the resulting ground truth labels are often uncertain. Crowdsourcing offers an alternative, lower-cost means of obtaining annotated data, but label quality varies considerably among crowd workers, particularly in tasks such as CHM damage annotation that require cultural understanding. Label quality can be improved by assigning each task to multiple crowd workers and aggregating their responses into high-quality labels.

To address these gaps, we propose a novel automatic multi-label classification framework for CHM damage assessment. The framework is designed to accurately classify multiple types of damage within a single image, namely spalling, erosion, loss of material, other damage and no damage. In addition, it introduces a human-in-the-loop active learning strategy that minimizes annotation effort by selectively querying only the most informative samples. Unlike existing approaches, our method does not rely solely on large annotated datasets or expensive sensors; instead, it combines AI-based automation with efficient human guidance to provide a label-efficient and domain-adaptable solution that advances the state of the art in cultural heritage damage classification. To the best of our knowledge, publicly available labelled datasets for multi-damage classification of cultural heritage monuments are lacking, particularly for Indian heritage sites such as those in Uttarakhand. To fill this gap, we introduce a new dataset of diverse images with multi-label annotations collected through crowdsourcing, which addresses a serious data shortage in this field and supports the development of scalable, AI-driven preservation methods.

Key Contributions

In this work, the collective intelligence of the crowd is combined with AI to address the CHM damage classification (CHMDC) problem. Specifically, we aim to classify damage efficiently by training a deep learning model on large amounts of imagery data annotated by crowd workers. This paper addresses three main challenges of deep multi-label active learning with crowdsourcing, as described below.

1. The first challenge is the availability of ground truth. Traditional supervised algorithms assume that labelled data are available [16]; however, in practical CHMDC applications, ground truth is either unavailable or very limited.
2. The second challenge is cost-effective data annotation. Crowdsourcing is an inexpensive way to collect labelled data; however, it still incurs costs due to the complexity of multi-label imagery data.
3. The third challenge is correctly inferring the ground truth. Unlike expert annotations, the labels provided by crowd workers are inconsistent. This inconsistency is particularly prominent in CHMDC because of the intricate nature of CHMs, and it becomes more pronounced in multi-label imagery data, in which an image may belong to more than one class simultaneously.

To overcome these challenges, this paper makes the following contributions:

1. The unavailability of labelled data is overcome by collecting annotations from crowd workers for new case studies of various CHMs in Uttarakhand.
2. An active learning strategy is formulated to reduce the annotation cost by selecting the most informative

samples, together with a multi-label aggregation method to infer high-quality labels.
3. A hybrid deep neural network based on DenseNet-169, Bi-LSTM and RNN layers is used for multi-label classification.
4. Extensive experiments demonstrate the validity of the methodology with limited labelled data.

Related Work

Structural Health Monitoring (SHM) involves regular inspections to assess the structural condition of CHMs. Its main objective is to detect abnormalities in a monument's structure, identify the type of damage, and enable timely intervention to preserve it. Damage to a CHM refers to any destruction or deterioration of the structure, whether deliberate or unintended. It can be caused by natural events, intentional damage, combat, environmental contamination, negligence, or flawed restoration practices. Should it remain unremedied, such harm can impair a monument's structural integrity and lessen its historical and artistic importance, obstructing its role in preserving cultural heritage for future generations. The erosion of CHMs affects our mutual cultural heritage and identity, constituting a critical issue for conservationists, historians, and societies globally. Multiple manual strategies are there for recognising such damage.

Machine Learning For Damage Classification of CHM

AI-based visual inspection typically employs a machine learning (ML) framework, most commonly a convolutional neural network (CNN). CNNs are the most widely used neural-network-based ML approach and have been applied by many researchers to identify different types of damage in CHMs. Meklati et al. [17] designed a CNN-based crowd-sensing web application to classify four types of damage: efflorescence, spalling, mould, and crack. Yang et al. [18] used YOLOv4 for object classification of five types of damage. Mishra et al. [19] employed YOLOv5 and Faster R-CNN to analyse CHMs in New Delhi and categorise damages including spalling, fissures, exposed masonry, and discolouration. Wei et al. [20] proposed BFD-YOLO, an improved YOLOv7 model for damage detection. Samhouri et al. [21] used a CNN with feature extraction methods for multi-label classification of various types of damage in Jordan. Mansuri et al. [22] used Faster R-CNN to detect objects.

Table 1 presents a synthesis of prior investigations, specifying the machine learning models applied, the datasets used, and the forms of damage observed. For instance, Wang et al. [23] applied a deep learning (DL)-based CNN to identify deterioration in brick masonry on a wall of a Chinese palace. The same authors later adapted the CNN to monitor damage from video on web and mobile platforms. In a separate study, Wang et al. [24] combined deep learning with mobile crowd sensing to examine damage to the Great Wall of China. Garrido et al. [25] used infrared thermal imaging instead of conventional images to locate defects in marquetry. Pathak et al. [26] used 3D point clouds to detect cracks and spalling in CH structures. Karimi et al. [27] applied an Inception-ResNet model to detect deterioration in bridge CH structures. Cardellicchio et al. [28] applied YOLOv5 to identify defects in RC bridges. More recently, Liu et al. [29]

Table 1. Comparison of ML techniques for damage classification

Reference	Case Study on CHM	ML-technique	Output	Dataset Size	Accuracy
[17]	Algiers Kasbah	CNN (MobileNetV2)	• Efflorescence • Spalling • Mould • Crack	2674 Images	86.80%
[18]	CHM in Macau	YOLOv4	• Cracking • Missing components • Microbial Erosion • Pollution • Yellowing	1355 Images	85.70%
[31]	Clay tiles on location in Macau, China	YOLOv4	• surface wear • Cracks • Stains	363 Images	89.30% for surface wear, 95.40% for crack and 80.90% for stains
[27]	Ancient bridges in the city of Isfahan	Inception-ResNet-v2	• Flaking • Erosion • Cracking • Efflorescence • No defect	8331 images	96.60%
[20]	Photographs of building exteriors	YOLOv7	• Tile loss • Delamination • Spalling	1907 images	77.60-81.60%
[32]	Loire Valley Castle, France	YOLOv5	• Spalling • Deterioration	1012 images	81.00%

Table 1. Comparison of ML techniques for damage classification (countinued)

Reference	Case Study on CHM	ML-technique	Output	Dataset Size	Accuracy
[19]	CHM in India's capital—New Delhi	Faster R-CNN and YOLOv5	• Discolouration • Cracks • Exposed bricks • Spalling	10291 images	93.70%
[22]	Dutch and English graves in Surat, India	Faster R-CNN	• Cracks • exposed bricks • spalling	880 images	91.50%
[33]	Ayutthaya, Thailand's Ancient City	SVMs and deep CNNs	• non-cracks • cracks	6002 Images patches	92.00%
[25]	Two varieties of Marquetry with CHM components	Mask R-CNN	• missing tesserae • honeycombing • subsurface defects • surface defects	A small set of infrared images of marquetry with surface defects	97.10%
[26]	Temple of Ayutthaya, Thailand: Wat Phra Si Sanphet	Faster-RCNN	• spalling • crack	7500 images	44.60% for crack and 71.80% for spalling
[34]	United Kingdom masonry wall pictures	CNNs pre-trained on ImageNet	• crack • no-crack	351 cracked images and 118 crack-free images	95.30%
[35]	Chinese wall and canal around the Palace Museum	Faster R-CNN	• spalling • efflorescence	500 images using two orthophotos	95.00%
[24]	The Great Wall of China	R-CNN-based inspection system	• Crack • Intact • Spalling • Efflorescence	610 images	78.20%
[36]	Wall of the Forbidden City in China	CNN based on sliding window	• Intact • Crack • Efflorescence • Spalling • Crack	2400-20000 images	94.30%
Our method	Heritage monuments in Uttarakhand	DenseNet-169 with Bi-LSTM	• Spalling • Erosion • Loss of material • Other damages • No damage	3530 images	97.00%

designed an efficient multi-branch Transformer that fuses multi-scale features to improve object detection, providing a scalable means of recognizing complex damage patterns of different sizes in cultural heritage images. Yang et al. [30] proposed a Modality Fusion Vision Transformer in which cross-attention and spectral self-attention modules combine hyperspectral and LiDAR data, which overcomes the limitations of earlier fusion methods and improves classification performance. Other studies summarised in Table 1 include Zheng et al. [31], who detected surface wear, cracks and stains on clay tiles in Macau using YOLOv4; Idjaton et al. [32], who detected limestone spalling in 3D survey images; Ravichand et al. [33], who detected cracks on brick walls; Dais et al. [34], who classified and segmented cracks on masonry surfaces; and Wang et al. [35] and Zou et al. [36], who used CNNs to classify damage and locate missing components in historic Chinese masonry structures.

Crowdsourcing

Crowdsourcing is an emerging paradigm in which individual workers collaborate to solve complex problems [37] cost-effectively. Its applications range from disease surveillance to real-time cyber threat detection. Despite its potential, crowdsourcing faces several significant issues, such as data dependability, incentive structure, data insufficiency, human-computer interaction (HCI), and privacy protection [38]. Moreover, multi-label damage classification of imagery data remains challenging. For CHMDC, our approach addresses these issues by combining AI with crowd intelligence, thereby improving both data quality and classification accuracy.

Because crowd workers are not domain experts, data quality, and therefore data dependability, depends on worker quality. Our methodology addresses this in two ways. First, a hidden test was used to assess worker quality, and only workers who passed it were permitted to provide

annotations. Second, repeated labelling was used with a voting limit of k=10 labels per instance. Once an instance had received annotations from 10 different crowd workers who had passed the hidden test, it was removed from the task assignment pool. Label aggregation is then required to estimate higher-quality labels and mitigate the effect of noisy labels; for this purpose, we applied a multi-label aggregation method to obtain high-quality data.

Although our approach does not include an incentive structure, we acknowledge that a well-designed incentive structure would motivate crowd workers to participate in the annotation process. Data insufficiency, the most common obstacle in multi-label classification, is mitigated by active learning, which requests labels selectively to make the most effective use of data at reduced cost. HCI plays a central role in making crowdsourcing platforms more accessible to crowd workers; in crowdsourcing, it mainly concerns task design, interface usability and worker engagement. Our approach uses a task design that provides a basic description, although a better task design could further improve data quality. Finally, privacy is essential in crowdsourcing because CHM data can be sensitive, and stronger privacy-preserving techniques can prevent unauthorized access to data. By taking these issues into account, our approach provides a dependable crowdsourcing application for complex, data-intensive tasks such as CHMDC and supports accurate, scalable and ethically sound heritage preservation.

Crowd-AI Active Hybrid Systems

Our work belongs to an emerging line of research on Crowd-AI hybrid systems that address complex real-world problems [39]. Krivosheev et al. [40] proposed an active hybrid crowdsourcing method that combines active learning and crowdsourcing for classification. Yang et al. [41] introduced a collaborative system that combines crowdsourcing with a probabilistic architecture to remove noise from information in movie reviews and reports. Guo et al. [42] developed a crowd-AI query answering system for smart-home applications that evaluates video streams acquired by intelligent IoT devices. Crowdsourcing applications are often designed for binary or multi-class classification tasks, whereas multi-label classification is more challenging and has been addressed only in some recent studies. For example, Suyal et al. [43] proposed a crowdsourcing-based method for cost-effective multi-label classification, and Tu et al. [44] proposed a joint matrix factorization approach for multi-label crowd consensus. However, these approaches have not been applied to CHMDC. The present study therefore focuses on multi-label classification of imagery data based on active learning and crowdsourcing.

PROPOSED METHODOLOGY

This paper proposes a crowdsourcing-based solution for automatic damage classification. The methodology comprises four major components, as shown in Figure 1. The first component, data collection, is the most critical because such data are difficult to obtain. Unlabelled images were gathered by capturing photographs with mobile devices and drones and by collecting images from various social media sites. The second component, active hybrid crowdsourcing, uses active learning to select the most informative image

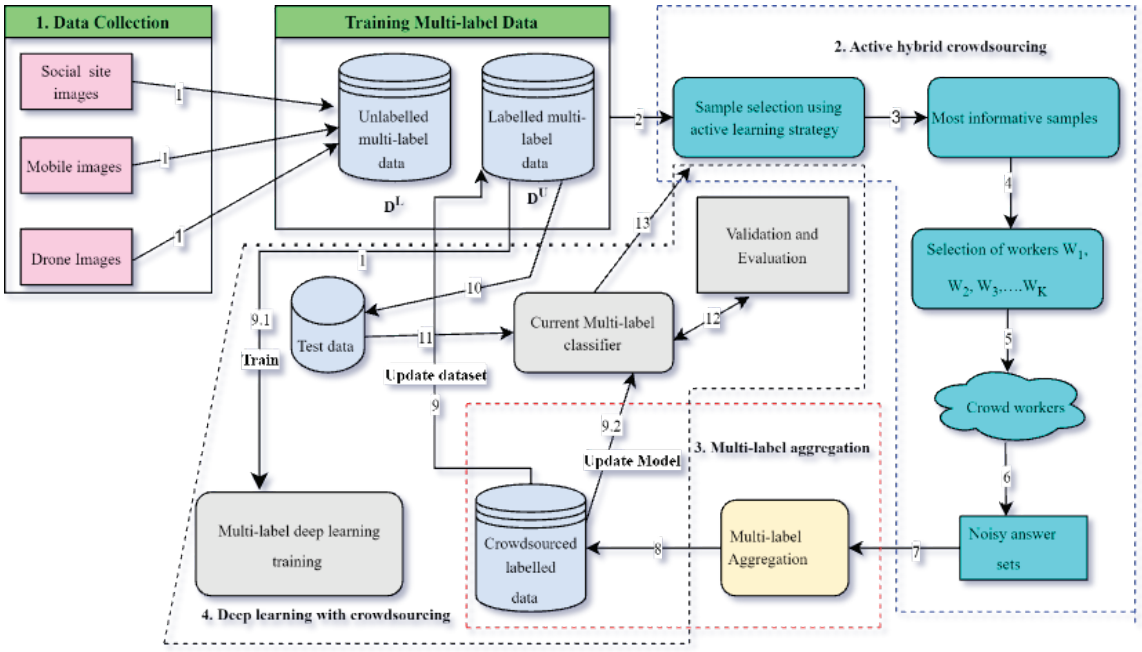


Figure 1. The proposed methodology has four major components: data collection, active hybrid crowdsourcing, multi-label aggregation, and deep learning with crowdsourcing.

samples, which are then sent to multiple crowd workers to obtain low-cost but noisy label sets. The third component, multi-label aggregation, infers the ground truth from these labels to produce high-quality data. The final component is deep learning with crowdsourcing, for which this study proposes a hybrid deep neural network combining a DenseNet-169 backbone with Bi-LSTM and RNN layers.

Motivation

Uttarakhand is a culturally rich, largely hilly state with many heritage sites of historical significance that are frequently affected by extreme weather events, including monsoon rain and snowfall. The resulting damage takes various forms that must be recognized, which motivated us to develop an automated damage classification system. In this paper, damage is grouped into five major categories: spalling, erosion, loss of material, other damage, and no damage. The standard definitions of the damage types, based on the ICOMOS glossary [45], are given below.

1. Spalling is the detachment or flaking of outer stone layers caused by environmental stresses such as freeze-thaw cycles and salt crystallization, both of which are common in the variable climate of the region. Exposure of the underlying material allows moisture ingress and accelerates deterioration. Accurate identification of spalling therefore helps prioritize preventive care for monuments exposed to these conditions.
2. Erosion is the gradual wearing away of surface material by weathering, running water or wind and is most pronounced at exposed mountain locations. Besides reducing the aesthetic value of heritage sites, it threatens their structural integrity. Monitoring its progression supports restoration planning and decisions on where drainage or wind barriers should be installed.
3. Loss of Material refers to any reduction of the stone surface, whether of mechanical, chemical or biological origin. In Uttarakhand, it is frequently caused by visitor contact with the monuments, heavy rainfall and vegetation growth on the structures. Classifying such losses enables targeted interventions, such as restricting visitor access or removing invasive plants, to preserve the authenticity of the structure.
4. Other Damage covers forms of degradation that do not fall into categories such as spalling or erosion, for example biological colonization (lichen, moss) and discolouration. In humid and shaded areas of Uttarakhand, such biological growth can accelerate stone decay.
5. No Damage refers to surfaces that show none of the above forms of deterioration. Including this class allows the model to distinguish intact surfaces from damaged ones and helps reduce false positive predictions.

Problem Statement

This paper considers an active hybrid multi-label classification model in which classification is performed jointly by the crowd and the machine. The problem is defined for a given tuple (D^U, A, B, L, D) where:

- D^U is a pool of unlabelled multi-label image datasets where $D^U = \{<x_i, Y_{i>}\}_{i=1}^{N}$, where (x_i, Y_i), $x_i \in \chi$ (instance space) where $x_i = \{v_1^i, v_2^i, v_3^i, v_4^i, \ldots\ldots\ldots v_d^i,\}$ is a feature vector and v_d^i is d^{th} attribute of instance x_i. $Y_i \subseteq Y$, where $Y = (L_1, L_2 \ldots\ldots L_j)$ set of unknown labels.
- A is a Multi-label Active Learning (MAL) strategy for selecting the most informative samples.
- B is the budget, i.e., the number of crowd-worker queries available for annotation.
- L is a multi-label aggregation method used to infer the ground truth.
- D is an untrained deep-learning model.

This paper aims to train the deep learning model D using the pool of unlabelled data. We apply MAL techniques to select the most informative items $I \subset U$ and assign them to k crowd workers through task assignment. The annotated data are passed to the multi-label aggregation method to infer the ground truth. Finally, we train the deep learning model on the labelled training data (D^T) collected by the label aggregation method and evaluate the performance of the approach on the test dataset (D^{Test}). The overall workflow of our hybrid approach is described in Algorithm 1.

Algorithm 1: Procedure of the proposed approach

Input: U, A, B, L, D
1. $D^U \leftarrow \{\}$, U # unlabelled items
2. $D^T \leftarrow \{\}$, T # labelled training data
3. label_all ← {} # set of crowd labels
4. b ← 0 {} # Budget spent (number of queries processed)
5. while b <B
6. select_batch samples from U via A
7. label_batch # collect labels from the crowd workers
8. label_all← label_all ∪ label_batch
9. end while
10. $D^L \leftarrow$multi-label aggregation (Label_all) via L
11. Update D^T using D^L
12. train D on D^T
13. test D on D^{Test}
14. multi-label classification
15. return

Active Hybrid Crowdsourcing for Multi-Label Damage Classification

Active learning is the process of interactively selecting the most informative samples, obtaining their labels, and training the model on the updated labelled dataset. It is an effective way to reduce labelling cost while producing adequate training data, and it becomes increasingly necessary as the number of unlabelled images grows. In this research, we used a pool-based active learning method because of its effectiveness for image datasets. Any MAL approach has two major components: the sampling strategy, which selects the most informative samples, and the annotation process, in which crowd workers provide the labels. The labelled data are then used to train the multi-label classifier, and

this cycle is repeated until a stopping condition is reached. Algorithm 2 shows the overall process of pool-based active hybrid crowdsourcing.

Algorithm 2. Pool-based Active hybrid crowdsourcing

Input: Unlabelled multi-label data pool $D^U = \{<x_i, Y_{i>}\}_{i=1}^{N}$
Training data $D^T \leftarrow \{\}$ *# initially empty*
Deep learning model $D \leftarrow \{\}$ *# initially untrained*
Output: Final Deep Learning Model D
1. *while stopping condition not reached*
2. *For each unlabelled instance* x_i *in U, do*
3. *Compute the diversity of* x_i *using eq.2*
4. *End*
5. *Sort the unlabelled instances* x_i *based on the diversity*
6. $S \leftarrow$ *select the most diverse image batch samples*
7. *Ask crowd workers for annotations using the strategy in the Crowd Annotations section*
8. $D^L \leftarrow$ *Aggregate the labels to infer the ground truth using the strategy in the Multi-Label Aggregation section*
9. $D^T \leftarrow D^T \cup D^L$ *#update the training data*
10. $D^U \leftarrow D^U/S$ *#update unlabelled data*
11. $D \leftarrow D^L$ *Update the deep learning model based on new training datasets.*
12. *End*

Sampling

MAL sampling involves two aspects: sample granularity, i.e., the sample size, and informativeness, i.e., how informative a sample is. Because traditional MAL is slow, this paper adopts the most straightforward way of accelerating it, namely batch-mode sample selection. Batch-mode selection chooses the most diverse unlabelled samples for each batch but suffers from redundancy; we therefore select unlabelled examples with minimal redundancy. To compute the redundancy function (*Red)* for each pair, each image is first converted into a feature vector. These features characterize the CHM imagery and support the similarity analysis. For instance, colour histograms or edge-detection features clearly show how image similarity influences the selection process of the MAL algorithm. After normalizing each vector to the same scale, we estimate the cosine similarity between the examples of each pair (x_j, x_s) using *eq. 1:*

$$similarity((x_j, x_s)) = \left|(x_j, x_s)\right| \Big/ \sqrt{(x_j, x_j) \cdot (x_s, x_s)} \quad (1)$$

After the pairwise redundancy has been computed, the diversity (diver) of each example is defined in *eq. 2* as follows:

$$diver(x_i) = 1 - Max_{x_s \in U} similarity(x_i, x_s) \quad (2)$$

Crowd Annotations

MAL mainly relies on experts for annotation, but experts are scarce or costly. Moreover, a multi-label deep learning model requires large amounts of labelled data, which makes expert annotation even more expensive. Crowdsourcing is a cost-efficient way to collect multiple labels per data point, although these labels are noisy. In this study, 50 crowd workers were recruited for annotation after an initial qualification test; only those who achieved a minimum accuracy of 80% on a set of gold-standard examples were selected for the main task. Each qualified worker was

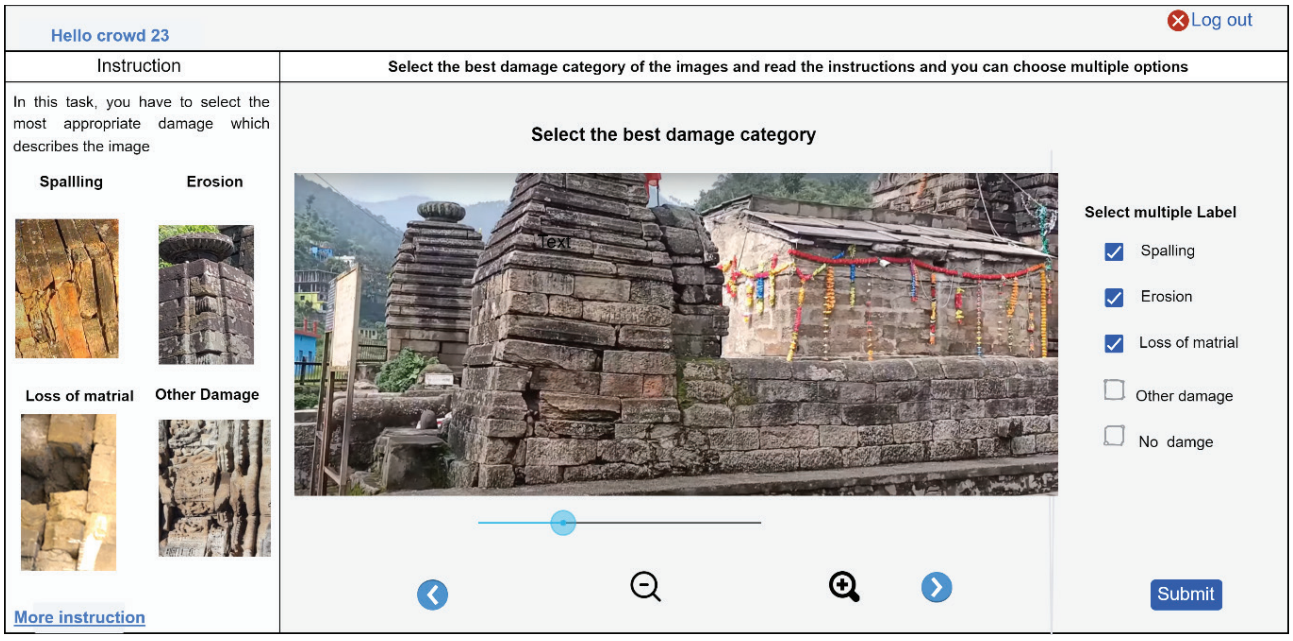


Figure 2. Task assignment given to crowd workers.

assigned approximately 640 images on average, which the proposed MAL framework had selected as the most informative samples. Figure 2 shows the annotation interface, which provided clear instructions, a definition of each label and example images to minimize ambiguity. Workers could assign more than one damage label to an image, because monuments often exhibit several overlapping types of damage. To ensure quality, each image was annotated by multiple workers, and a label aggregation method was applied to resolve inconsistencies. In addition, gold-standard images were periodically embedded to monitor worker performance throughout the task.

Multi-Label Aggregation

Multi-label crowdsourced data are represented as $D^U = \{<x_i, Y_{i>}\}_{i=1}^{N}$ where each instance x_i is associated with a noisy label vector $Y_i = \{Y_{ip}\}_{p=1}^{P}$. where Y_{ip} denotes the label of x_i given by the worker's U_p (p= 1,2, 3….. P) and takes values from Y = (L_1, L_2…….L_z), and each instance may belong to multiple labels. Z represents the number of labels; if the value of x_{ijp} is 1, crowd worker p assigns class j to the instance x_i; a value of 0 indicates that the worker does not. Aggregation aims to infer the ground truth $\check{Y}$ for the instance x_i. In this paper, we first perform label augmentation; then, to handle the noise, we apply distance-based label weighting and finally use ML-kNN to predict the label $\check{Y}$. Label aggregation is discussed in detail below, and Algorithm 3 summarizes the process.

Label Augmentation

This section describes the use of ML-kNN for label augmentation, in which the k nearest neighbours are used to obtain multiple noisy label sets. The ML-kNN used in our method differs slightly from the traditional ML-kNN used for crowdsourced data. In our approach, we first find the nearest neighbours of each image, including the image itself. Because the data are multi-label, the choice of distance measure is important; we used the Hamming distance given in *eq.3*.

$$D(x_i, x_j) = \frac{1}{Z}\sum_{m=1}^{Z} I\,(x_{im} \neq x_{jm}) \tag{3}$$

Z is the number of labels, and in our method, Z=5. x_{im} denotes the value for the instance x_i for label m. I ($x_{im} \neq x_{jm}$) is an indicator function and returns 1 if x_{im} and x_{jm} differ, and 0 otherwise. A smaller Hamming distance indicates that two instances are closer. The instances are then sorted in increasing order of distance to find the k nearest neighbours $\{x_{ik}\}_{k=1}^{K}$ for each instance x_i. The label sets of these neighbours are then merged to create an augmented noisy label set (Y_{i1}, Y_{i2}, …. Y_{ik}). We used majority voting to resolve any conflicts during augmentation. Since crowdsourced data are noisy, handling the noise is essential; we therefore apply label weighting, as described in the next section.

Label Weighting

To handle the noise, the labels are weighted by distance. The distance is calculated from the augmented label set and used to estimate the similarity between each instance and its neighbours; a smaller distance indicates that they are more likely to belong to the same class. We therefore assign lower weights to more distant neighbours to reduce their influence and define the weight W_{ik} of the kth neighbour for the instance x_i as *eq. 4*.

$$W_{ik} = 1 - \frac{D(x_i, x_{ik})}{D(x_i, x_{iK})} \tag{4}$$

Finally, the ground truth is predicted from the posterior probability using the above weights, as given in *eq. 5*.

$$\check{Y} = \arg max_{L_Z} \sum_{k=1}^{K}\sum_{p=1}^{P} \delta\left(l_{ikp} \middle| L_Z\right).W_{ik} \tag{5}$$

where $\delta(l_{ikp} \mid L_Z)$ counts the number of votes for the label L_Z.

Algorithm 3: Multi-label aggregation

Input: labelled crowd data $D^U = \{<x_i, Y_{i>}\}_{i=1}^{N}$
Output: $\{\check{Y}_i\}_{i=1}^{N}$
1. ***for*** *i=1 to n*
2. ***for*** *j=1 to n*
3. *calculate the distance using the eq. 3*
4. ***end*** *for*
5. *find k nearest neighbours* $\{x_{ik}\}_{k=1}^{K}$ *by sorting them based on the distance*
6. *merge the labels to get the noisy label vector.*
7. ***for*** *k=1 to K*
8. *use the distance to apply the weight using the eq. 4*
9. *update the weight using the updated probability*
10. ***end*** *for*
11. *infer the ground truth* $\check{Y}$ *using the eq. 5*
12. ***end*** *for*
13. *return* $\{\check{Y}_i\}_{i=1}^{N}$

Convolutional Neural Network (CNN)

A CNN is a class of deep learning models widely used for image classification. It is composed of three types of layers: convolutional layers, which extract features; pooling layers, which reduce dimensionality; and fully connected layers, which perform classification processes. A CNN first extracts features from an input image through successive convolution, pooling and activation layers. Several pre-trained CNN models are available, such as R-CNN, DenseNet, and ResNet. This paper uses DenseNet rather than ResNet because its architecture is well suited to multi-label classification with few parameters. DenseNet has several variants, including DenseNet-121, DenseNet-161, DenseNet-169 and DenseNet-201; this study focuses on DenseNet-169 [5+(6+12+32+32)×2]. A traditional CNN calculates the output of the nth layer by applying a non-linear transformation Hn(.) to the output of the previous layer Yn-1 as shown in *eq. 6*:

$$Y_n=H_n(Y_n\text{-}1) \quad (6)$$

In contrast to conventional CNNs, DenseNet connects each layer to all subsequent layers, so that the nth-layer receives the feature maps of all preceding layers, and eq. 6 becomes *eq. 7:*

$$Y_n= H_n\,(Y_0,\ Y_1,\ Y_2,\ \ldots\ldots\ Y_n\text{-}1) \quad (7)$$

where $(Y_0, Y_1, Y_2, \ldots\ldots Y_n\text{-}1)$ is the tensor formed by concatenating the output maps of the preceding layers. A non-linear transformation is represented by the function Hn(.) and composed of three operations: batch normalisation (BN), rectified linear units (ReLU) activation, and convolution (CONV). As illustrated in Figure 3, this investigation suggests a hybrid deep neural network that integrates Bi-LSTM and RNN layers with a DenseNet-169 feature extractor. The model is divided into two parts: a backbone feature extractor and a dynamic modelling component. The DenseNet-169 backbone uses dense connections to reduce redundant feature selection and the vanishing-gradient problem and to provide latent feature space representations. The extracted features are fed to a hybrid deep neural network consisting of Bi-LSTM layers with 512 and 128 neurons, an RNN layer with 48 units, and fully connected layers with 128 and 32 neurons. The model is trained with the Adam optimizer and binary cross-entropy loss, with a learning rate of 0.01, a dropout of 0.01 in the hybrid head, and 25 epochs. The long short-term memory network extracts local and global features and performs deep neural conjugation. The final output of the architecture, of size $2 \times 2 \times 512$, is flattened to 2048 neurons and self-convoluted to $1 \times 1 \times 2048$.

Table 2 presents the DenseNet-169 architecture with pre-trained weights. It comprises convolution, dense, max-pooling and transition layers, with ReLU activation throughout and a sigmoid activation layer at the output. The convolutional layers extract features from the image data, whereas the max-pooling layers reduce the spatial dimensions of the inputs.

Experimental Setup

Developing deep CNNs requires substantial computational power because of the large number of repetitive computations, such as backpropagation and convolution. Because image processing involves extensive matrix computation, the experiments were performed on an HP machine with an Intel Core i7-7700HQ 2.80 GHz CPU with 32 GB RAM and an NVIDIA GeForce GTX 1060 GPU. Anaconda Python 3.11 was used as the programming platform. The GPU-enabled version of Keras with a TensorFlow backend was chosen as the machine learning framework because of its effectiveness and wide adoption across different learning styles, design characteristics and hyperparameter settings. Other libraries included pandas for data handling, NumPy for multidimensional arrays and scikit-learn for data analysis. All other classifiers were trained, tested and applied using the scikit-learn machine learning library.

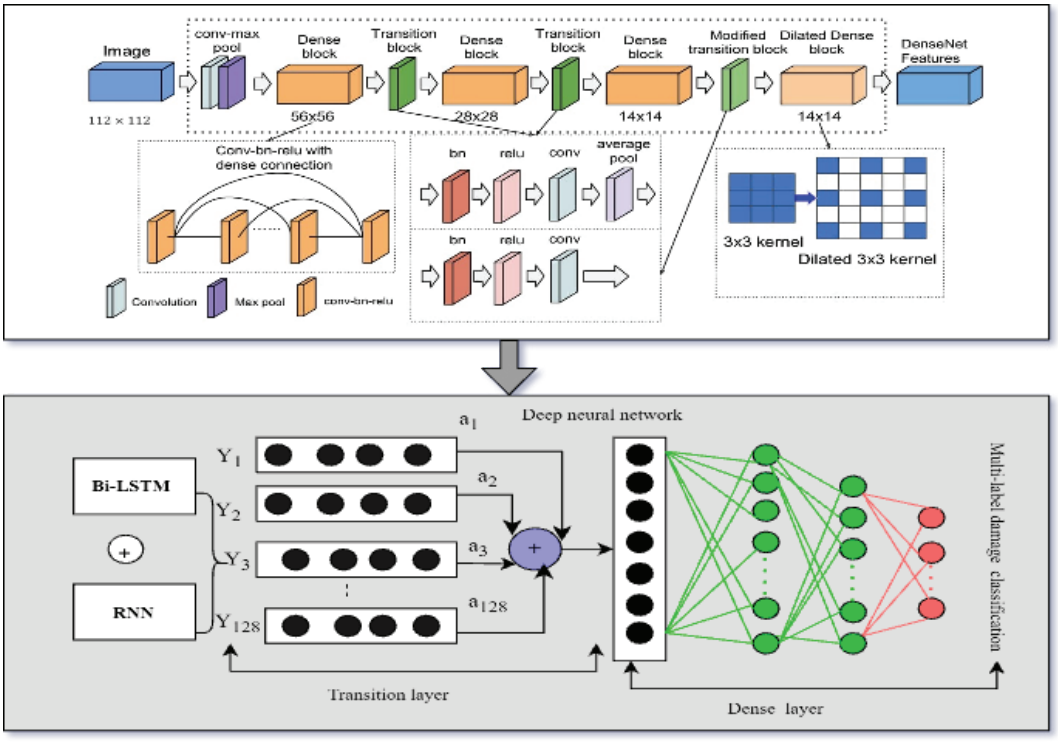


Figure 3. Proposed architecture of a hybrid deep neural network based on a Bi-LSTM network.

Table 2. Layered DenseNet-169 architecture (standard ImageNet pre-trained backbone; in the proposed model, the 1000-D classification layer is replaced by the Bi-LSTM/RNN head with a five-unit sigmoid output)

Layers	Size of Output	Parameters	Kernel Size
Convolution	[112 × 112]	Stride=2, ReLu	7 × 7 (Convolution)
Pooling	[56 × 56]	Stride=2	3× 3 (Max Pool)
Dense Block-1	[56 × 56]	Dropout= 0.2	[1× 1 × 6] (Convolution)
			[3× 3 × 6] (Convolution)
Transition-1	[56 × 56]	Stride= 2	[1× 1] (Convolution)
	[28 × 28]		[2 × 2] (Avg. Pool)
Dense Block-2	[28 × 28]	Dropout= 0.2	[1× 1 × 12] (Convolution)
			[3× 3 × 12] (Convolution)
Transition-2	[28× 28]	Stride= 2	[1× 1] (Convolution)
	[14× 14]		[2 × 2] (Avg. Pool)
Dense Block-3	[14 × 14]	Dropout= 0.2	[1× 1 × 32] (Convolution)
			[3× 3 × 32] (Convolution)
Transition-3	[14× 14]	Stride= 2	[1× 1] (Convolution)
	[7× 7]		[2 × 2] (Avg. Pool)
Dense Block-4	[7 × 7]	Dropout= 0.2	[1× 1 × 32] (Convolution)
			[3× 3 × 32] (Convolution)
Classification layer	[1× 1]		[7 × 7](Global Avg. Pool)
	[1000]		[1000 D] (Fully Connected Sigmoid)

Dataset

The images were collected from heterogeneous heritage sites across Uttarakhand representing diverse architectural styles, including ancient temples, forts, step-wells and colonial-era buildings. The sites span a range of altitudes and terrains, from densely forested valleys to high-altitude stone ruins, which gives the dataset both cultural and structural diversity. Some photographs were taken with mobile devices or drone cameras, and others were obtained from publicly available social media posts, so both ground-level and aerial views are included. The images reflect realistic variation in environmental conditions, having been captured at different times of day (morning, midday, dusk), in clear, cloudy and rainy weather, and in different seasons. From an initial pool of 6,254 images, 3,209 were selected by the active learning strategy and then labelled for training. To reduce false positive predictions, a further 321 images without visible damage were deliberately included, bringing the labelled dataset to 3,530 images. Together with the 50 no-damage images among those selected by active learning, the dataset therefore contains 371 no-damage images (Table 3). This variety makes the dataset more representative and helps the model remain robust across conditions and site types. The characteristics of the multi-label data are listed in Table 3. After multi-label aggregation, the labels collected from the crowd were saved in a CSV file. Because no ground truth labels were available for the dataset, a

Table 3. Characteristics of the labelled data collected from the crowd

Label set	Images	Label set	Images
Spalling	671	Erosion+ Loss of material	21
Erosion	615	Erosion+ other Damage	81
Loss of material	656	Loss of material + other damages	115
other damages	754	Spalling+ Erosion+ Loss of material	26
No damage	371	Spalling+ Loss of material+ other damage	15
Spalling+ Erosion	55	Spalling+ Erosion+ other Damage	31
Spalling+ Loss of material	27	Erosion+ Loss of material + other damage	23
Spalling+ Other Damage	65	Spalling+ Erosion+ Loss of material + other damages	4

hybrid validation strategy combining cross-validation with expert-labelled data was adopted. Stratified 5-fold cross-validation was used so that each fold preserved the distribution of damage classes, allowing a rigorous assessment of how well the model generalizes to different subsets of the data. Furthermore, a subset of 10% of the complete dataset was randomly selected and painstakingly annotated with the help of domain experts in heritage conservation. This subset served as a dependable ground truth benchmark for performance validation. The proposed model demonstrated a high level of predictive capability on this expert-validated subset, achieving an accuracy of 98% in the absence of full-scale ground truth. This two-level validation (cross-validation on noisy or pseudo-labelled data and manual validation on a clean subset) supports the robustness and reliability of the results. During preprocessing, each image was resized to 112 × 112 pixels and paired with its labels. Data augmentation, comprising rotation, flipping, noise addition and lighting changes, was then applied to enlarge the training set and to simulate real-world variation in environmental conditions. Multi-hot encoding was used to encode the ground truth labels. The dataset was split into 63% for training, 7% for validation and 30% for testing.

RESULTS AND DISCUSSION

Evaluation Measures

The model was evaluated using accuracy, precision, recall, and F1-score. These measures are computed from the numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN).

Precision is the proportion of true positive instances among all instances predicted as positive, as given in *eq.8*:

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall: Recall is the proportion of actual positive instances that the model correctly predicts, as defined in *eq. 9:*

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Accuracy: Accuracy is the ratio of correct predictions to the total number of predictions and is defined in *eq.10*:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

F1 Score: In *eq.11,* the F1 score is defined as the harmonic mean of precision and recall (sensitivity).

$$F1\ score\ = \frac{2 * (recall)(precision)}{Precision + Recall} \tag{11}$$

Analysis of Results

Table 4 reports the accuracy, recall, precision and F1 score of the proposed classification model for each of ten runs on individual test sets, together with their averages, with 60% of the image pool labelled through active learning. Figure 4 shows the training and testing accuracy per epoch for five settings in which 20%, 30%, 40%, 50% and 60% of the data were labelled through active learning. Training accuracy increased with the amount of labelled data; however, performance did not improve further when the labelled share increased from 50% to 60%. With 60% labelled data, the final training and testing accuracies were 97% and 96%, respectively. Because the dataset introduced in this work has not previously been used in the literature, we compare the performance of our method with that of similar deep learning-based damage classification techniques applied at other heritage sites. Although a direct benchmark comparison is not possible, our model achieves 97% accuracy with only 60% of the data labelled, demonstrating its robustness and generalizability. Prior studies

Table 4. Classification result for the proposed model

Run	Accuracy	Recall	Precision	F1-score	Loss	Time
Run 1	**98.51**	92.13	96.65	94.34	**0.012**	15.56
Run 2	98.46	96.28	96.23	96.25	0.018	**15.01**
Run 3	98.06	97.14	96.24	96.69	0.021	15.78
Run 4	98.13	96.56	92.92	94.71	0.032	15.89
Run 5	95.52	95.44	**97.54**	96.48	0.097	15.46
Run 6	95.12	93.56	92.66	93.11	0.101	15.46
Run 7	95.16	**97.45**	90.59	**93.89**	0.105	15.67
Run 8	98.02	97.14	97.45	97.29	0.014	15.78
Run 9	98.13	92.45	95.45	93.93	0.019	15.78
Run 10	98.12	92.56	94.56	93.55	0.014	15.98
Average	97.32	95.07	95.03	95.02	0.043	15.64

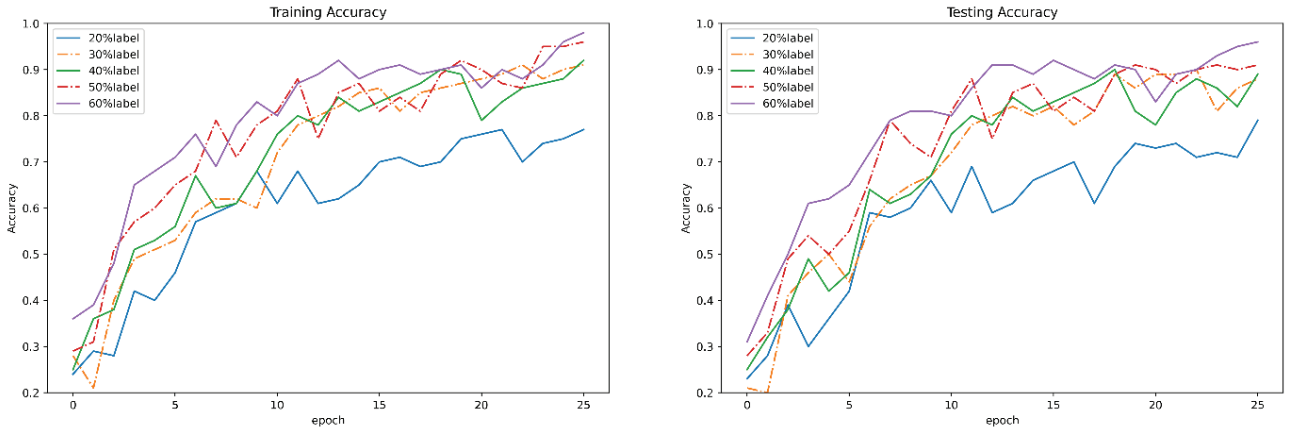


Figure 4. Training and testing accuracy of the deep learning model.

[19, 22, 33, 35] typically report accuracy in the range of 91–95% using larger, fully annotated datasets, which highlights the effectiveness of our crowdsourcing-enhanced, label-efficient approach.

Figure 5 shows the training and testing loss of the proposed approach, computed with binary cross-entropy, which is suitable for multi-label data. After 25 epochs, the final loss falls to about 0.02 for both training and testing (Fig. 5); across the ten runs in Table 4, the average loss is 0.043. The loss decreased consistently on both the training and testing datasets over the later epochs, indicating effective training. The similar learning behaviour on both datasets, despite fluctuations in the testing curve, indicates that the model does not overfit the training data.

Figure 6 shows the performance metrics, including precision, recall and F1 score. To examine how the amount of annotated data affects the deep learning model, experiments were conducted on labelled subsets containing 20%, 30%, 40%, 50% and 60% of the full dataset. All key performance metrics improved steadily as the proportion of labelled data increased. With only 20% of the instances labelled, the accuracy and F1 score were 0.75 and 0.72, respectively, representing a baseline level of generalization under limited supervision. Increasing the labelled proportion to 30% produced a substantial improvement, with the F1 score rising to 0.87. Performance continued to improve at 40% and 50%, with both accuracy and F1 score reaching 0.96 at 50%. Beyond this point, the improvement was marginal. At 60% labelled data, accuracy increased slightly to 0.97 and precision to 0.98, whereas the F1 score remained at 0.96. Recall decreased slightly from 0.96 to 0.95, suggesting that model sensitivity may be saturating. Thus, additional labelled data improves model performance, but with diminishing returns beyond 50%.

Figure 7 presents the per-class evaluation of the proposed deep learning-based framework for multi-label damage classification in CHMs. Overall classification performance is good, with accuracy ranging from 0.86 to 0.98. The highest accuracies are obtained for no damage (0.98) and other damage (0.96), indicating that intact surfaces

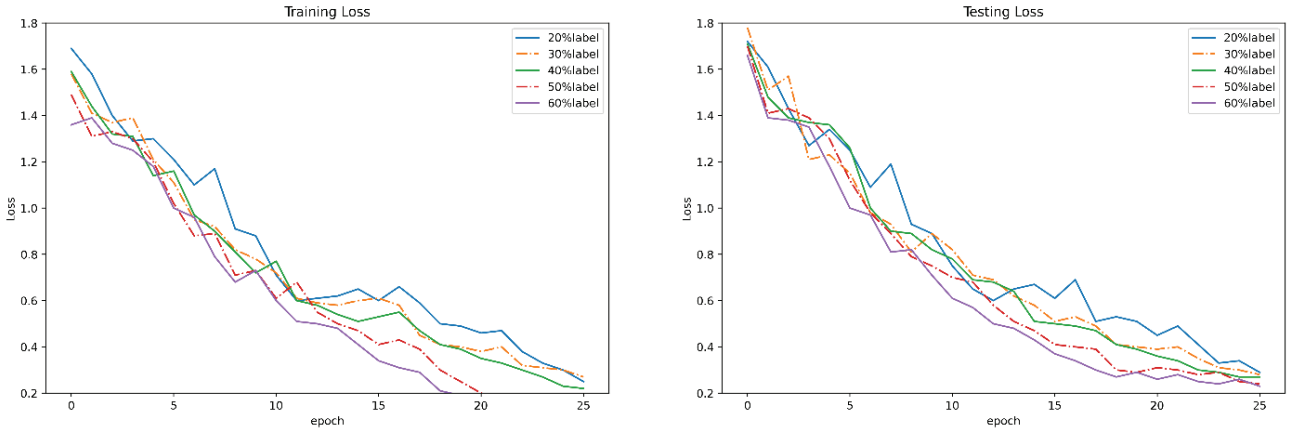
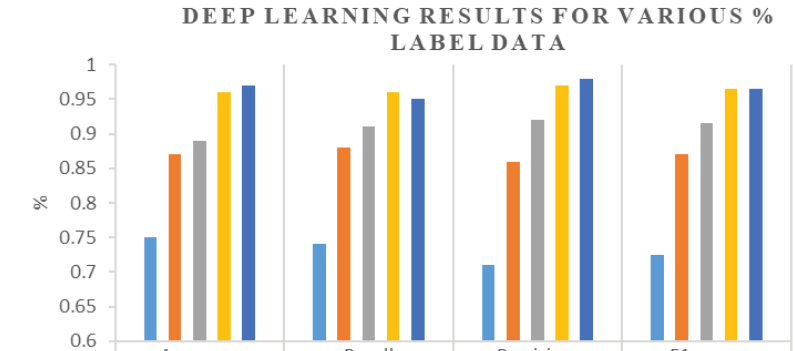
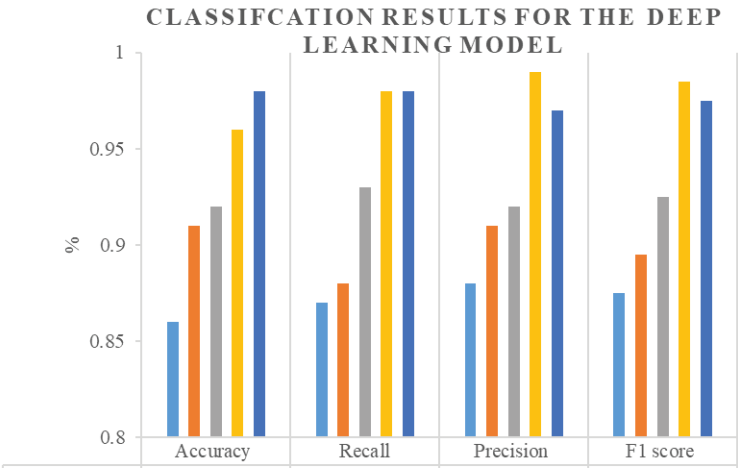


Figure 5. Training and testing loss of the deep learning model.



	Accuracy	Recall	Precision	F1 score
20% label	0.75	0.74	0.71	0.72
30% label	0.87	0.88	0.86	0.87
40% label	0.89	0.91	0.92	0.91
50% label	0.96	0.96	0.97	0.96
60% label	0.97	0.95	0.98	0.96

Figure 6. Performance metrics of the deep learning model.



	Accuracy	Recall	Precision	F1 score
Spalling	0.86	0.87	0.88	0.87
Erosion	0.91	0.88	0.91	0.89
Loss material	0.92	0.93	0.92	0.92
Other damage	0.96	0.98	0.99	0.98
No damage	0.98	0.98	0.97	0.97

Figure 7. Per-class performance of the deep learning model.

are reliably distinguished from generalized deterioration. Spalling has the lowest accuracy (0.86), possibly because subtle texture differences or occlusions in the images make finer structural defects harder to recognize. Recall is 0.98 for both no damage and other damage, indicating few false negatives in these classes. Recall is also high for loss of material (0.93) and erosion (0.88), further indicating consistent behaviour across damage types. Precision is likewise high, at 0.99 for other damage and 0.97 for no damage, indicating few false positives. Precision is somewhat lower for spalling (0.88) and erosion (0.91), probably because their visual features overlap with those of other classes, which suggests that the feature representation could be further refined. The F1-score, which balances precision and recall, confirms the overall stability and reliability of the model. Other damage (0.98) and no damage (0.97) again achieve the highest values, followed by loss of material (0.92), whereas spalling has the lowest F1-score (0.87), consistent with the other metrics. Overall, the proposed model effectively identifies damaged and undamaged architectural features in CHMs, even with limited labelled data and noisy annotations.

The normalized confusion matrix in Figure 8 shows that all damage categories are classified well. Other damage and no damage have the greatest true positive rates (0.98 each), suggesting the precise identification of generally undamaged and deteriorated areas. Spalling and erosion exhibit similar surface degradation patterns, resulting in occasional confusion between them. Their rates are slightly lower, at 0.89 and 0.91, respectively. Loss of material reaches 0.92 but shows slight confusion with erosion and other damage, suggesting that localized detachment is sometimes difficult to distinguish from broader wear. Overall, the small off-diagonal values confirm the robustness of the model for multi-label damage classification on real-world heritage data.

Figure 9 compares the proposed deep neural network with several established architectures, namely VGG16, ResNet50 and several DenseNet variants. The proposed model outperforms all baselines, achieving an accuracy of 0.97, recall of 0.95, precision of 0.98 and F1-score of 0.96. By comparison,

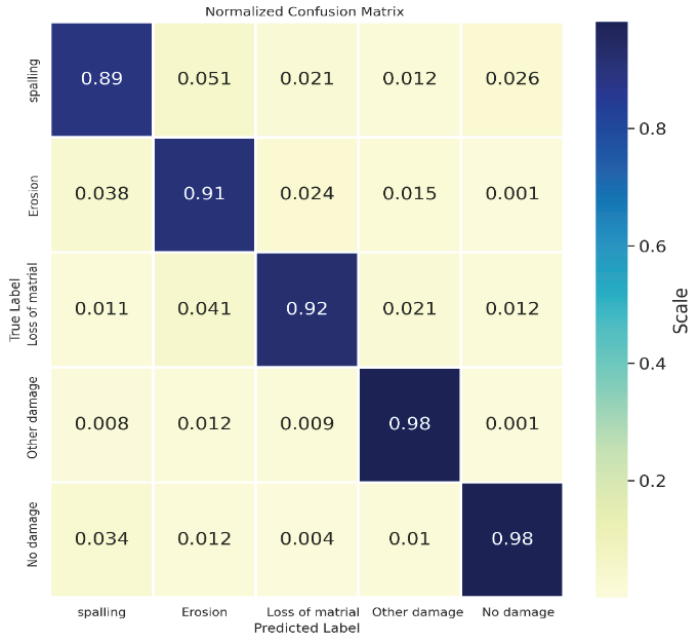
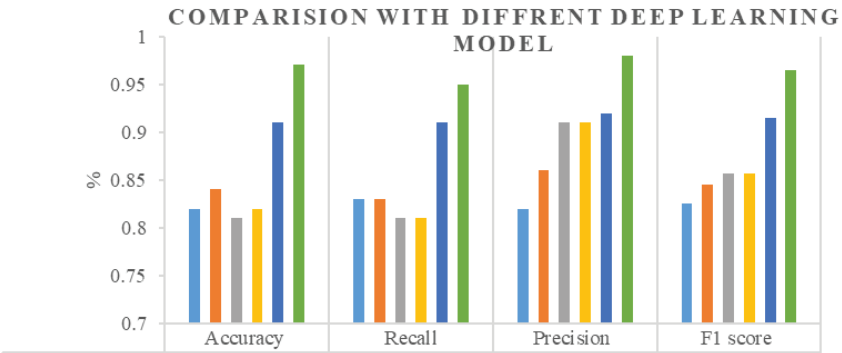


Figure 8. Normalized confusion matrix of the proposed deep learning model.



	Accuracy	Recall	Precision	F1 score
VGG 16	0.82	0.83	0.82	0.82
ResNet 50	0.84	0.83	0.86	0.84
DenseNet-121	0.81	0.81	0.91	0.86
DenseNet-160	0.82	0.81	0.91	0.86
DenseNet-169	0.91	0.91	0.92	0.91
Proposed deep Neural Network	0.97	0.95	0.98	0.96

Figure 9. Average performance comparison of the proposed model with various deep learning models.

the strongest baseline, DenseNet-169, reaches an accuracy of 0.91 and an F1-score of 0.90. This gain indicates that the proposed model is better at capturing complex multi-label damage features in heritage images. Its high precision and recall further indicate low false positive and false negative rates across the different damage categories. Table 5 illustrates the multi-label deep learning model's ability to identify multiple types of damage in a single cultural heritage image and assign a probability to each. Table 5 provides sample outputs. These outputs demonstrate that the model is capable of identifying co-occurring damage patterns. One image, for instance, has high probabilities of spalling (0.72), other damage (0.66), and attrition (0.61), suggesting that multiple forms of deterioration are identified concurrently. In another image, loss of material dominates with a probability of 0.827, confirming the model's sensitivity to pronounced damage. The probability assigned to "no damage" remains relatively low in all cases, further indicating that actual damage is precisely distinguished from intact surfaces. Such probabilistic outputs capture the complex, overlapping damage typical of heritage structures and can therefore support detailed, well-grounded conservation decisions.

An ablation study was conducted to quantify the contribution of each component of the proposed damage

Table 5. Damage classes detected by the multi-label deep learning model

Multi-label deep learning model classifying image	Damage classification	Probability
	Spalling	0.72
	Erosion	0.61
	Loss of material	0.25
	Other Damage	0.66
	No damage	0.01
Multi-label deep learning model classifying image	**Damage classification**	**Probability**
	Spalling	0.513
	Erosion	0.346
	Loss of material	0.321
	Other Damage	0.367
	No damage	0.251
Multi-label deep learning model classifying an image.	**Damage classification**	**Probability**
	Spalling	0.671
	Erosion	0.565
	Loss of material	0.553
	Other Damage	0.511
	No damage	0.125
Multi-label deep learning model classifying an image.	**Damage classification**	**Probability**
	Spalling	0.164
	Erosion	0.127
	Loss of material	0.827
	Other Damage	0.176
	No damage	0.013

Table 6. Ablation study results for the proposed model

Model Configuration	Accuracy	F1-Score	Precision	Recall
Full model (Active learning + Label aggregation + Hybrid CNN)	**0.97**	**0.96**	**0.98**	**0.95**
Without Active Learning (Random sampling)	0.91	0.88	0.89	0.87
Without Label Aggregation (Raw crowd labels)	0.89	0.86	0.87	0.85
Without Hybrid Architecture (Standard CNN only)	0.92	0.90	0.91	0.89

classification framework; the results are summarized in Table 6. The complete model, comprising active learning, label aggregation and the hybrid deep neural network architecture, achieved the best overall results: 97.0% accuracy, 0.96 F1-score, 0.98 precision and 0.95 recall. Substituting active learning with random sampling resulted in a noticeable decrease in performance (accuracy: 91.0%, F1-score: 0.88, precision: 0.89, recall: 0.87), highlighting the significance of judiciously choosing the most useful samples for effective learning. Excluding label aggregation and training just on unrefined crowd labels significantly diminished performance (accuracy: 89.0%, F1-score: 0.86, precision: 0.87, recall: 0.85), emphasising the necessity of mitigating annotation noise via aggregation. Ultimately, substituting the hybrid deep network with a conventional CNN diminished accuracy to 92.0%, F1-score to 0.90, precision to 0.91, and recall to 0.89, so validating that the hybrid design more effectively encapsulates the multi-scale and contextual attributes essential for precise damage classification. Each element enhances the resilience and efficacy of the suggested framework for evaluating cultural heritage deterioration. Note that Table 4 reports the average over ten independent runs, each evaluated on its own test set, whereas Figures 6 and 9 and Table 6 report the results of the final trained model; the precision and F1-score values therefore differ slightly between the two evaluations.

CONCLUSION

This study presented a new human-in-the-loop framework for multi-label damage classification of cultural heritage monuments that addresses key problems such as scarce ground truth annotations, noisy crowdsourced labels and high annotation costs. Through pool-based active learning and label aggregation, the method efficiently selects the most informative samples and improves label quality. Trained on a diverse set of crowdsourced images, the model achieved a classification accuracy of 97% with only 60% of the data labelled, demonstrating that it is both effective and label-efficient. The approach enables scalable, cost-effective and accurate damage assessment and thereby supports timely preventive maintenance of cultural heritage sites.

Although the results are promising, several aspects can be improved. Larger multi-label datasets with higher resolution and more damage categories would considerably improve generalization and coverage. Improved crowdsourcing protocols and adaptive model tuning could make the annotations more consistent and reliable. In future work, we intend to integrate multimodal data, such as LiDAR for structural profiling and thermal imaging for revealing hidden or heat-induced degradation, to further increase detection accuracy and resilience. Sensitivity to subtle or overlapping damage might also be improved with advanced image processing techniques such as semantic segmentation, denoising and multi-scale feature extraction. Finally, the present study addresses classification only; future research will focus on the precise localization and spatial mapping of damage within architectural elements to enable targeted conservation and restoration planning.

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Aras F, Krstevska L, Altay G, Tashkov L. Experimental and numerical modal analyses of a historical masonry palace. Constr Build Mater 2011;25:81–91. [CrossRef]

[2] Gentile C, Gallino N. Condition assessment and dynamic system identification of a historic suspension footbridge. Struct Control Health Monit 2008;15:369–388. [CrossRef]

[3] Mishra M. Machine learning techniques for structural health monitoring of heritage buildings: a state-of-the-art review and case studies. J Cult Herit 2021;47:227–245. [CrossRef]

[4] Agyemang IO, Zhang X, Acheampong D, Adjei-Mensah I, Kusi GA, Mawuli BC, Agbley BLY. Autonomous health assessment of civil infrastructure using deep learning and smart devices. Automation Construct 2022;141:104396. . [CrossRef]

[5] Ye XW, Jin T, Yun CB. A review on deep learning-based structural health monitoring of civil infrastructures. Smart Struct Syst 2019;24:567–585.

[6] Soleymani A, Jahangir H, Nehdi ML. Damage detection and monitoring in heritage masonry structures: systematic review. Constr Build Mater 2023;397:132402. [CrossRef]

[7] Bibi R, Saeed Y, Zeb A, Ghazal TM, Rahman T, Said RA, et al. Edge AI-based automated detection and classification of road anomalies in VANET using deep learning. Comput Intel Neurosci 2021;1;6262194. [CrossRef]

[8] Zinno R, Haghshenas SS, Guido G, Rashvand K, Vitale A, Sarhadi A. The state of the art of artificial intelligence approaches and new technologies in structural health monitoring of bridges. Appl Sci 2022;13:97. [CrossRef]

[9] Pan Y, Zhang L. Roles of artificial intelligence in construction engineering and management: a critical review and future trends. Autom Constr 2021;122:103517. [CrossRef]

[10] Bikku T. Multi-layered deep learning perceptron approach for health risk prediction. J Big Data 2020;7:50. [CrossRef]

[11] Batchu RK, Bikku T, Thota S, Seetha H, Ayoade AA. A novel optimization-driven deep learning framework for the detection of DDoS attacks. Sci Rep 2024;14:28024. [CrossRef]

[12] Yiğit AY, Uysal M. Detection of cracks in cultural heritage buildings using UAV photogrammetry-based digital twin. J Comput Cult Herit 2025;18:1–17. [CrossRef]

[13] Gezer ED, Aydoğan Selçuk B. Determination of damage and defects in historical wooden structures using nondestructive test devices. Sigma J Eng Nat Sci 2020;11:127–134.

[14] Peng L, Bo W, Yang H, Li X. Deep learning-based image compression for enhanced hyperspectral processing in the protection of stone cultural relics. Expert Syst Appl 2025;271:126691. [CrossRef]

[15] Liu Z. The construction of a digital dissemination platform for the intangible cultural heritage using convolutional neural network models. Heliyon 2025;11. [CrossRef]

[16] Suyal H, Singh A. Improving multi-label classification in prototype selection scenario. Computational Intelligence and Healthcare Informatics. 2021. p. 103–119. [CrossRef]

[17] Meklati S, Boussora K, Abdi MEH, Berrani SA. Surface damage identification for heritage site protection: a mobile crowd-sensing solution based on deep learning. J Comput Cult Herit 2023;16:1–24. [CrossRef]

[18] Yang X, Zheng L, Chen Y, Feng J, Zheng J. Recognition of damage types of Chinese gray-brick ancient buildings based on machine learning-taking the Macau World Heritage Buffer Zone as an example. Atmosphere (Basel) 2023;14:346. [CrossRef]

[19] Mishra M, Barman T, Ramana GV. Artificial intelligence-based visual inspection system for structural health monitoring of cultural heritage. J Civ Struct Health Monit 2024;14:103–120. [CrossRef]

[20] Wei G, Wan F, Zhou W, Xu C, Ye Z, Liu W, et al. BFD-YOLO: A YOLOv7-based detection method for building façade defects. Electronics 2023;12:3612. [CrossRef]

[21] Samhouri M, Al-Arabiat L, Al-Atrash F. Prediction and measurement of damage to architectural heritages facades using convolutional neural networks. Neural Comput Appl 2022;34:18125–18141. [CrossRef]

[22] Mansuri LE, Patel DA. Artificial intelligence-based automatic visual inspection system for built heritage. Smart Sustain Built Environ 2022;11:622–646. [CrossRef]

[23] Wang N, Zhao X, Zhao P, Zhang Y, Zou Z, Ou J. Automatic damage detection of historic masonry buildings based on mobile deep learning. Autom Constr 2019;103:53–66. [CrossRef]

[24] Wang N, Zhao X, Wang L, Zou Z. Novel system for rapid investigation and damage detection in cultural heritage conservation based on deep learning. J Infrastruct Syst 2019;25. [CrossRef]

[25] Garrido I, Erazo-Aux J, Lagüela S, Sfarra S, Ibarra-Castanedo C, Pivarčiová E, et al. Introduction of deep learning in thermographic monitoring of cultural heritage and improvement by automatic thermogram pre-processing algorithms. Sensors 2021;21:750. [CrossRef]

[26] Pathak R, Saini A, Wadhwa A, Sharma H, Sangwan D. An object detection approach for detecting damages in heritage sites using 3-D point clouds and 2-D visual data. J Cult Herit 2021;48:74–82. [CrossRef]

[27] Karimi N, Valibeig N, Rabiee HR. Deterioration detection in historical buildings with different materials based on novel deep learning methods with focusing on Isfahan historical bridges. Int J Archit Herit 2023:1–13. [CrossRef]

[28] Cardellicchio A, Ruggieri S, Nettis A, Mosca N, Uva G, Renò V. On the use of YOLOv5 for detecting common defects on existing RC bridges. In: Stella E, Ceglarek D, Kemao Q, Soldovieri F, editors. Multimodal Sensing and Artificial Intelligence: Technologies and Applications III. SPIE; 2023. p. 15. [CrossRef]

[29] Liu F, Zheng Q, Tian X, Shu F, Jiang W, Wang M, et al. Rethinking the multi-scale feature hierarchy in object detection transformer (DETR). Appl Soft Comput 2025;175:113081 .. [CrossRef]

[30] Yang B, Wang X, Xing Y, Cheng C, Jiang W, Feng Q. Modality fusion vision transformer for hyperspectral and LiDAR data collaborative classification. IEEE J Sel Top Appl Earth Obs Remote Sens 2024;17:17052–17065. [CrossRef]

[31] Zheng L, Chen Y, Yan L, Zhang Y. Automatic detection and recognition method of Chinese clay tiles based on YOLOv4: a case study in Macau. Int J Archit Herit 2023:1–20. [CrossRef]

[32] Idjaton K, Janvier R, Balawi M, Desquesnes X, Brunetaud X, Treuillet S. Detection of limestone spalling in 3D survey images using deep learning. Autom Constr 2023;152:104919. [CrossRef]

[33] Ravichand M, Kumar R, Hazela B, Suthar T. Crack on brick wall detection by computer vision using machine learning. In: 2022 6th International Conference on Electronics, Communication and Aerospace Technology. IEEE; 2022. p. 1017–1020. [CrossRef]

[34] Dais D, Bal İE, Smyrou E, Sarhosis V. Automatic crack classification and segmentation on masonry surfaces using convolutional neural networks and transfer learning. Autom Constr 2021;125:103606. [CrossRef]

[35] Wang N, Zhao Q, Li S, Zhao X, Zhao P. Damage classification for masonry historic structures using convolutional neural networks based on still images. Comput Aided Civ Infrastruct Eng 2018;33:1073–1089. [CrossRef]

[36] Zou Z, Zhao X, Zhao P, Qi F, Wang N. CNN-based statistics and location estimation of missing components in routine inspection of historic buildings. J Cult Herit 2019;38:221–230. [CrossRef]

[37] Suyal H, Singh A. Comparative study of truth inferences algorithms in crowdsourcing. Optimization Methods for Engineering Problems. 2023. p. 99. [CrossRef]

[38] Zhang D, Ma Y, Hu XS, Wang D. Toward privacy-aware task allocation in social sensing-based edge computing systems. IEEE Internet Things J 2020;7:11384–11400. [CrossRef]

[39] Yu G, Tu J, Wang J, Domeniconi C, Zhang X. Active multi-label crowd consensus. IEEE Trans Neural Netw Learn Syst 2021;32:1448–1459. [CrossRef]

[40] Krivosheev E, Casati F, Bozzon A. Active hybrid classification. arXiv [Preprint]. 2021:2101.08854.

[41] Yang J, Smirnova A, Yang D, Demartini G, Lu Y, Cudre-Mauroux P. Scalpel-CD: leveraging crowdsourcing and deep probabilistic modeling for debugging noisy training data. In: The World Wide Web Conference. New York: ACM; 2019. p. 2158–2168. [CrossRef]

[42] Guo A, Jain A, Ghose S, Laput G, Harrison C, Bigham JP. Crowd-AI camera sensing in the real world. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018;2:1–20. [CrossRef]

[43] Suyal H, Singh A. Multi-label classification using crowdsourcing under budget constraints. Knowl Inf Syst 2023. [CrossRef]

[44] Tu J, Yu G, Domeniconi C, Wang J, Xiao G, Guo M. Multi-label crowd consensus via joint matrix factorization. Knowl Inf Syst 2020;62:1341–1369. [CrossRef]

[45] Sites ME. ICOMOS-ISCS Illustrated glossary on stone deterioration patterns. 2011.