Journal of Cultural Heritage 69 (2024) 78–85

Contents lists available at ScienceDirect

Journal of Cultural Heritage

journal homepage: www.elsevier.com/locate/culher

Original article

Deep learning for the detection and classification of adhesion defects in antique plaster layers

Michele Lo Giudice, Francesca Mariani, Giosuè Caliano*, Alessandro Salvini

Department of Civil, Computer Science and Aeronautical Technologies Engineering, Universiá degli Studi Roma Tre, Via V. Volterra 62, Roma, 00146, Italy

ARTICLE INFO

Article history:
Received 12 April 2024
Revised 4 July 2024
Accepted 29 July 2024

Keywords:
Artificial intelligence
Convolutional neural network
Deep learning
Detachments
Non destructive testing
PICUS

ABSTRACT

This paper aims is to show an automated intelligent measurement system for the detection of adhesion defects between architectural antique plaster layers. The method emulates the traditional conservators' procedure based on acoustical perturbations, auscultation, detection and classification. The system makes use of a hardware device, known in literature as PICUS, for the generation and acquisition of acoustic signals, while the processing of the acquired signals is handled by a deep learning (DL) architecture designed ad hoc. After a brief description of the PICUS system and the acoustic data acquisition procedure, the whole architecture of the DL system is carefully described. The proposed method has been validated by a significant case study. The system shows an accuracy of up to 82% (± 2%) in multi-class classification and up to 99% (± 1%) in binary classification. In particular, the obtained results suggest a satisfactory precision in the detection of areas where stabilization is necessary.

1. Introduction and research aim

Depending on the conservator's sensitivity, the "detachment system" produces a response detected as an acoustic wave and vibration. The presence of air between layers of a poorly preserved antique plaster, which would otherwise be adjacent and adherent, results in a easily perceptible "hollow" sound. The operator can hear the sound and then interprets it based on their subjective perception. They naturally analyze the sound's characteristics, considering its variations in volume and pitch over time [1].

An inherent challenge in the conservator's evaluation process is its susceptibility to personal biases, which may result in subjective classifications. This can lead to either the rejection of defect-free parts (false negatives) or the acceptance of defective parts that surpass the defined limits outlined by relevant standards (false positives).

These subjective evaluations underscore the importance of comprehensive considerations of both human and technical elements in assessing the reliability and repeatability of the inspection process.

This paper aims to introduce a complete automatic system (hardware and software) that can be calibrated and adjusted according to the surface under detection, as a support instrument for professionals involved in this task.

Recently a hardware system, called PICUS, which emulates the tapping and hearing behavior of an expert conservation professional, has been proposed in [1]. The system consists of an electro-acoustic device that gently taps the surfaces and detects the sound produced using a transducer, returning it in the form of a sampled signal. The PICUS system will be here used as an auscultation acquisition tool.

In terms of handling the acoustic signals obtained by PICUS, we will introduce a software solution built on Deep Learning principles. Deep Learning (DL), due to its ability to learn directly from raw data [2], has been garnering considerable attention in many fields of signal processing from medical to engineering, and so on [3–5]. DL bases its working by using Neural Networks. Convolutional Neural Networks (CNNs) are the algorithms most implemented for detecting patterns by rough data [6,7]. CNNs can extract latent features inside the signal by means of suitable filters (learnable filters). Combinations of hundreds of learnable filters allow to capture different features (latent features) or patterns hidden into acquired raw signal [8]. This capability circumvents the laborious process of feature extraction and reduces biases inherent in handcrafted feature selection, thereby enhancing scalability.

* Corresponding author.
E-mail address: giosue.caliano@uniroma3.it (G. Caliano).

https://doi.org/10.1016/j.culher.2024.07.012

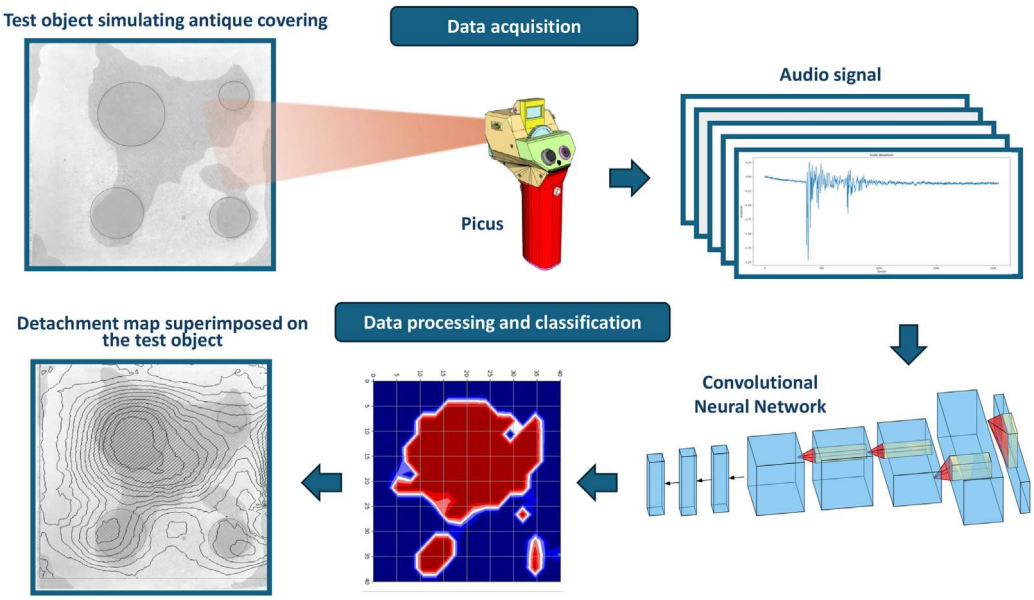


Fig. 1. Flowchart of the proposed method.

For defect recognition, an approach supervised by artificial intelligence [9] is therefore presented. The designed AI is trained to use data from well known cases acquired by PICUS. For each individual case, the nature of a defect, or its absence, was labeled by conservation experts according to a pre-established protocol described in detail below.

Although training these algorithms takes a long time, running a trained algorithm is extremely fast. To create a baseline for identifying audio signals, the approach uses a mock-up object on purpose that simulates an antique plaster, also described below.

We have initially established eight different classes: one with a gap value of zero, which indicates a stable state that does not require restoration work, and seven classes progressively closer to the maximum depth of the defect. We subsequently established two different classes for defect identification only, i.e. a class with a zero gap and a class with the presence of a defect, thus a generic defect identification. The datasets used in this work consist of N records for M classes (with N= 800 and M=8 for dataset 1 and N= 1500 and M=2 for dataset 2). The data provides the basis for training a CNN specifically designed for this purpose. This neural network takes audio samples as input and provides a class prediction of the membership inspected area.

The following sections will offer a detailed description of the proposed method. In particular, the tool used for signal acquisition and the configuration of audio acquisition is explained in Section 2.1. The acquired data, used for the CNN train proposal are explained in Section 2.3; and the details of the neural network architecture, including the parameters and validation mechanisms used, are explained in Section 2.4. In Section 3 the experimental results obtained are presented. The 4 and 5 sections address the discussion and conclusions respectively.

2. Materials and methods

The proposed methodology is visually depicted through a comprehensive flowchart, facilitating a clear understanding of the procedural framework in Fig. 1. The signal is acquired using the PICUS system on a mock-up test object (top left). The audio files (on the top right) were stored and then fed as input to the 1D convolutional neural network (CNN1D), which classifies its class and enables the creation of a color map (bottom centre). This color map is subsequently interpolated and overlaid on the test object (bottom left).

2.1. Data acquisition setup

An electro-acoustic system [10] called “PICUS” has been used as an acquisition system of artificial auscultation. PICUS mechanizes the auscultation technique used by experts in the field of cultural heritage conservation. The PICUS experimental setup includes a probe equipped with an electro-mechanical percussion element, which gently taps the surface to generate a sound converted into a voltage signal by a suitable microphone. A low-cost Arduino-like board acquires and processes the acoustic signals. The analogue signal is sampled at a frequency of 48 kHz and the system acquires 2048 samples: they contain the entire sound response of the point stimulated by the system [11–14]. The position of the probe is identified through an infrared camera on-board which detects an infrared LED source that lights up the scene.

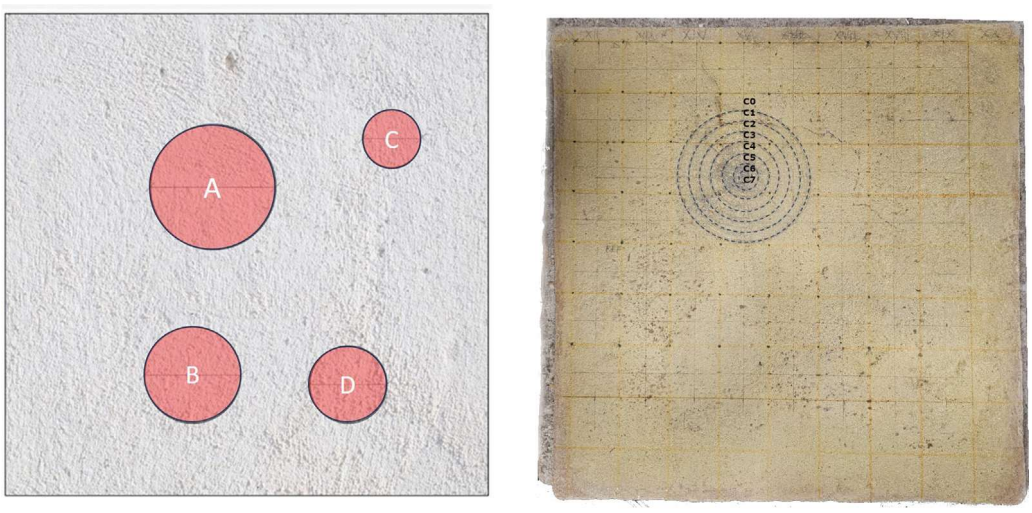
2.2. Experimental configuration for supervised acquisition

Fig. 2 shows the test object, a mock-up reproducing a stratified antique plaster according to antique and direct sources and a bibliography on the art of plastering [14]. The support is a 50 × 50 × 7.5 cm handmade brick tile.

The preparation, or “*arriccio*”, has been applied on the support in a 3 cm thick layer and is composed of lime, pozzolana and river sand. Using circular shapes, with diameters of 13, 10, 6 and 8 cm (A, B, C and D in Fig. 2(a)), four cavities with sharp edges and known profiles were created in the overlapping layer of “*intonaco*” (plaster), 2.5, 2, 2.5 and 2 cm deep, respectively. Next, on the *intonaco*, a superficial layer of mortar, of 1 cm thickness and finer aggregates called “*intonachino*”, has been applied. Once the mock-up was dry, defect A was chosen to be classified for the DL training. To capture the variability in defect levels, a setup of multiple concentric circles was conceived. (Fig. 2(b)).

The defects were categorized into eight classes (next indicated by C0, C1... C7), where C0 represents the defect-free area and from C1 to C7 it represents the gravity of the detachment from the edges towards the centre as illustrated in Fig. 2(b).

Every class represents a different degree of structural problems. To ensure the accuracy of the ground truth labels for supervised



(a) Mock-up of an antique plaster (b) Classes related to the defect area A

Fig. 2. (a) The test-object: a mock-up of an antique plaster with three layers of mortar, with four defective areas (A–D). (b) Classes related to the defect area A.

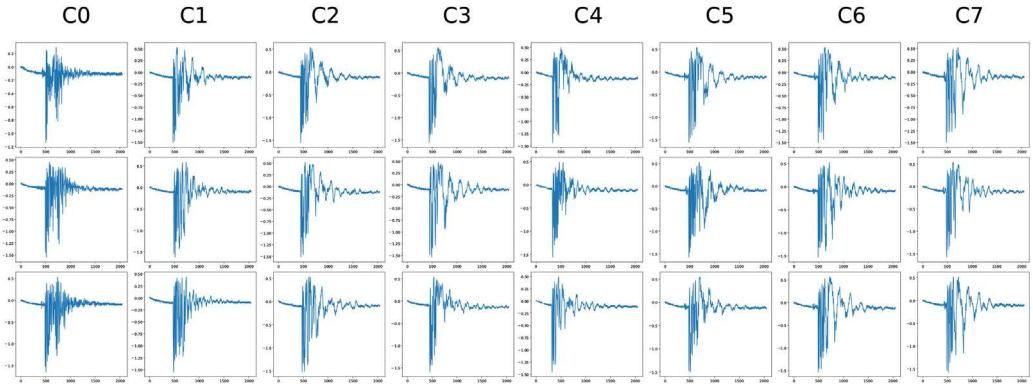


Fig. 3. 3 Random signals for each class.

training, each instance of the induced defect was meticulously labeled by an expert.

2.3. Data set acquisition

This study employed two distinct datasets. The first dataset has been used as a training and test set for the first neural network, tasked for classifying detachment type. For each of the eight designated classes, a collection of 100 audio samples was acquired. Representative examples from this dataset are visually illustrated in Fig. 3. Each class corresponded with a level of structural defect, and their audio signals were recorded using the microphone of the PICUS strategically positioned within the concentric circle setup. The second dataset was used to identify the presence of detachments across the entire surface of the material based on binary-labeled data from a single section, thus including the portion of the solid surface containing no detachment between the 0-label levels and acquisitions on the region of the detachment with label 1. The entire second dataset comprised 700 instances classified as defect-free areas and 800 instances classified as defective areas, for a total of 1,500 acquisitions.

2.4. Proposed convolutional neural network architecture

In the field of machine learning, convolutional neural networks (CNNs) have emerged as the standard over the past decade. These networks, characterised by feed-forward neural networks with convolutional and sub-sampling layers, have demonstrated an exceptional ability to discern complex patterns and objects within large visual datasets, after proper training with labelled data. As a result, they have become indispensable tools in various engineering applications, particularly those dealing with 2D signals such as images and video frames. For processing 1D signals as in this case, researchers introduced 1D CNNs, which quickly proved their worth by achieving state-of-the-art performance in a multitude of fields. In particular, these applications include classification of biomedical data [15], early diagnosis of diseases [16], structural health monitoring [17], anomaly detection in power electronics and fault detection in electric motors.

The 1D forward propagation in each CNN layer is stated as follows:

$$x_k^l = b_k^l + \sum_{i=1}^{N_{l-1}} \text{conv1D} \quad \left(w_{ik}^{l-1}, s_i^{l-1}\right) \tag{1}$$

Given that x_k^l represents the input, b_k^l is the bias of the kth neuron at layer l, s_i^{l-1} denotes the output of the ith neuron at layer $l-1$, and w_{ik}^{l-1} represents the kernel from the ith neuron at layer $l-1$ to the kth neuron at layer l. $Conv1D()$ is used to perform 1D convolution. Therefore, the dimension of the input array, x_k^l, is less than the dimension of the output arrays, s_i^{l-1}. The intermediate output, y_k^l, can be expressed by passing the input x_k^l through the

Table 1
The first convolutional neural network (CNN) architecture designed for classifying defects in ancient coverings, categorized from class C0 to C7. The size of x is defined by the batch size.

Layer (type)	Output Shape	Param #	Connected to
Conv1D	(x, 2048, 32)	128	
MaxPooling1D	(x, 1024, 32)	0	Conv1D
Conv1D	(x, 1024, 32)	3104	
Dropout	(x, 1024, 32)	0	Conv1D
MaxPooling1D	(x, 512, 32)	0	Dropout
Flatten	(x, 16384)	0	MaxPooling1D
Dense	(x, 32)	524320	Flatten
Dense	(x, 8)	264	Dense

activation function, $f(.)$, as,

$$y_k^l = f\left(x_k^l\right) \text{ and } s_k^l = y_k^l \downarrow ss \tag{2}$$

Here, s_k^l represents the output of the kth neuron of layer l, and "$\downarrow ss$" denotes the down-sampling operation with a scalar factor ss [16]. After the CNN layers, which perform the newly defined 1D forward propagation (1D-FP), the back-propagation (BP) algorithm is implemented to perform error back-propagation from the output MLP layer [18].

Given the significant network abilities introduced, a customized architecture of 1D CNN was developed to classify the audio signals to detect and classify the state of ancient coverings. Several architectures were examined, and the empirically chosen model was fine-tuned to achieve optimal performance.

The model architectural design follows a sequential structure using the Keras Sequential API [19] with TensorFlow [20] in the back-end. It begins with a 1D convolutional layer, serving as the cornerstone for feature extraction. This layer employs 32 filters with a kernel size of 3, utilizing the Rectified Linear Unit (ReLU) activation function for introducing non-linearity.

Following the convolutional layer, a max-pooling layer with a pool size of 2 is introduced. This strategic down-sampling operation helps reduce spatial dimensions while retaining crucial features. The subsequent layer is a second 1D convolutional layer, again with 32 filters and a kernel size of 3. To mitigate the overfitting, a dropout layer with a dropout rate of 0.3 is inserted.

Continuing the architectural flow, another max-pooling layer with a pool size of 2 is applied, further enhancing the network's ability to recognize invariant spatial patterns. The flattened layer then transforms the multidimensional output from the previous layers into a form suitable for fully connected layers.

The first dense layer, consisting of 32 units and ReLU activation, fosters global relationships among the learned features. Finally, the output layer, comprising a number of units equal to the total classes in the dataset, employs the softmax activation function for multi-class classification. Table 1 shows the neural network architecture and its parameters.

For the identification of detachments, the model architecture was slightly modified to better adapt to the second dataset. Since binary classification requires fewer neurons in the output layers, the structure was revised as follows: the first dense layer was reduced from 32 to 10 neurons and the output layer was reduced from 8 to 1 neuron. The activation function of the output layer was changed from *softmax* to *sigmoid* function as it is more suitable for binary classification. Reducing the number of neurons in the output layer improves the adaption of the model to the number of classes.

2.5. Metrics

For a complete evaluation of the effectiveness of the classifier in all categories, the confusion matrix was used, to effectively indicate the correspondence between actual and predicted class labels, identifying cases of misclassification with an immediate visual understanding, useful for statistical analysis. The correct classification of a sample is referred to as true positive, while an error in the classification of a sample is referred to as true negative. When an algorithm incorrectly identifies a sample as contending faults, it is referred to as a false positive. Conversely, when the algorithm incorrectly classifies a sample as having defects when in fact it does not, it is called a false negative. Table 2 shows a confusion matrix representation.

Table 2
An example of a confusion matrix.

	Predicted Class 1	Predicted Class 2
Actual Class 1	True Positive (TP)	False Negative (FN)
Actual Class 2	False Positive (FP)	True Negative (TN)

The Accuracy, Precision, Recall and F1-score metrics provide a detailed analysis of the classifier's performance for each individual class, clarifying the extent of its effectiveness within individual categories.

Accuracy measures the overall correctness of a classification model. It is calculated as the ratio of correctly predicted instances to the total number of instances. It represents the proportion of true results (both true positives and true negatives) among all the cases examined.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

The Precision quantifies the accuracy of positive predictions made by the model. It is calculated as the ratio of true positive predictions to the total number of positive predictions made by the model.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

Recall, also known as *sensitivity*, measures the ability of the model to find all the relevant cases within a dataset. It is calculated as the ratio of true positive predictions to the total number of actual positive instances in the dataset.

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{5}$$

The F1-score is the harmonic mean of precision and recall. The F1-score reaches its best value at 1 (perfect precision and recall) and worst at 0. It is calculated as 2 times the product of precision and recall, divided by the sum of precision and recall.

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

The macro/micro-average metrics offer aggregated assessments across all classes, providing a comprehensive perspective on the classifier's effectiveness. The macro average measures are also reported to computes the mean of performance metrics across all classes, assigning equal importance to each class irrespective of its size. Conversely, the weighted average adjusts for class imbalances by assigning weights proportional to the class sizes, thereby providing a more representative measure of overall performance. The receiver Operating Characteristic (ROC) curve is also used along the multi-class Area Under the ROC Curve (AUC).

This statistical approach evaluates the detectable flaw and aligns quantitative and qualitative parameters.

2.6. Data augmentation

We have introduced a data augmentation procedure applied to audio arrays to increase the dataset by 10 times. Specifically this

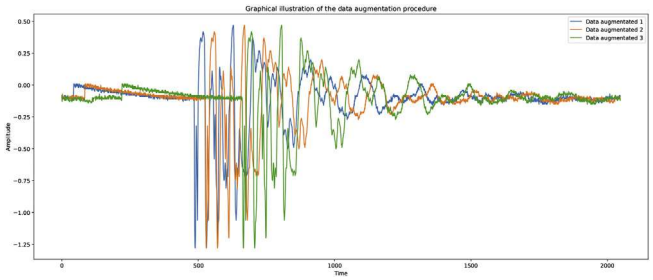


Fig. 4. This figure visually depicts the data augmentation procedure applied to audio arrays using the rolling technique. The illustration shows three of the ten examples involved in rolling the array by acquired data, highlighting how this specific augmentation method contributes to improving the diversity of the audio dataset for training and validation.

is performed using a specific function in NumPy used to roll array elements along a given axis [21]. In detail, each input vector was randomly shifted 10 times with values in the range of 0 to 256. In this way, temporally shifted versions of the original data are created. The evaluation of the augmented samples is empirically scaled to ensure adequate representation. Data augmentation is a crucial technique in DL, significantly increasing the training dataset, therefore improving the network's ability to generalize effectively. In our case, given the limited size of the dataset, the data augmentation reduces overfitting, as a model trained on a small dataset can store specific patterns in the data but fails to generalise to new examples [22]. Data augmentation introduces variations, causing the model to learn more robust features. Enhance generability by training the model on a wider array of data variations, thereby increasing its adaptability to real-world scenarios with nuanced data characteristics. [23,24]. In the Fig. 4 we can see an example of a data augmentation procedure. To improve visual understanding only 3 samples of 10 generated are plotted. This approach not only validates the best model but also strengthens its generalization, leading to more reliable results in real-world applications.

2.7. Training, test split and cross-validation technique

We provide a two-step approach for the optimal evaluation of the model. The data was first divided into training (70%) and testing (30%) sets, allowing us to explore a variety of model architectures. To validate the model in a thorough and rigorous way, we implemented a k-fold cross-validation (k=5) technique on the datasets. It was performed, by partitioning the dataset into five distinct subsets. During each iteration, four subsets were used for training the model (80%), while the remaining subsets were used for testing (20%). This process was repeated five times, ensuring that each subset was used exactly once as the test set.

The results of each cross-validation iteration were averaged to improve the reliability of the performance metrics, ensuring that our study's findings are statistically significant and generalizable.

The adoption of K-Fold Cross-Validation [25] in our work provides a comprehensive assessment of the model's performance by leveraging different subsets of the data for the test, thus mitigating the potential bias and variance associated with a single train-test split. Hence it ensures a thorough evaluation of the proposed method's performance, enhancing the reliability of our findings in the domain of architectural coverings' detachment monitoring with AI.

3. Results

The proposed model shows interesting classification capabilities in the classification task. In particular, the results of the CNN 1D are illustrated in the confusion matrix and ROC curve shown in Fig. 5.

Table 3
AVG Classification Report k-fold cross validation.

Class	Precision	Recall	F1-score	Support
C0	0.98 ± 0.02	0.97 ± 0.04	0.98 ± 0.02	20
C1	0.75 ± 0.09	0.69 ± 0.14	0.70 ± 0.06	20
C2	0.58 ± 0.08	0.64 ± 0.12	0.59 ± 0.06	20
C3	0.78 ± 0.03	0.73 ± 0.12	0.75 ± 0.07	20
C4	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	20
C5	0.81 ± 0.06	0.76 ± 0.10	0.78 ± 0.07	20
C6	0.82 ± 0.05	0.87 ± 0.07	0.84 ± 0.06	20
C7	0.94 ± 0.04	0.93 ± 0.02	0.94 ± 0.02	20
Accuracy			0.82 ± 0.02	160
Macro Avg	0.83 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	160
Weighted Avg	0.83 ± 0.02	0.82 ± 0.02	0.82 ± 0.02	160

The confusion matrix (Fig. 5(a)) gives a representation of statistical classification accuracy with the advantage of an immediate visual representation of the performance of the inter-class model. In this case, we observe a clear distinction between the outermost class of the detachment (C0) and the classes inside the defect (C1-C7). In particular, the C0 class reported a high number of true positives and a low number of false positives and false negatives. Classes C1-C7 still presented good performance, with a number of true positives higher than false positives and false negatives. It is worth noting that most of the classification errors were reported on contiguous classes, which are physically very close to each other. In particular, classes C5 and C6 exhibited classification errors almost exclusively between the two contiguous classes.

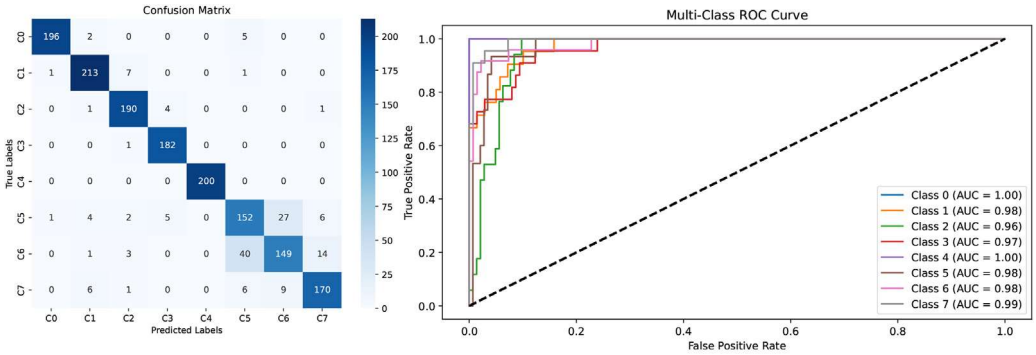
Similar outcomes were achieved by the analysis of AUC reported in Fig. 5(b), where the CNN classifier showed the highest AUC value in the C0 and C4 classifications. The curve showcases the trade-off between sensitivity and specificity, providing a comprehensive visualization of the classification performance and the model's ability to discriminate between different classes.

The results of the classification task are promising. The neural network has shown good potential for classifying audio with latent features. The diversified and augmented dataset has played a pivotal role in refining the model's generalization capabilities, ultimately contributing to an advanced and robust audio processing framework.

In view of the interesting results shown on the single test of the chosen 1D CNN architecture model, we use K-fold cross-validation to validate the proposed model, extending the results in a comprehensive report shown in Table 3 that emphasises its inter-class classification capabilities.

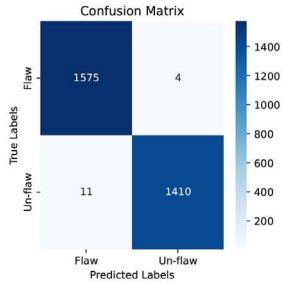
As shown in the report, the proposed neural network achieved remarkable performance in terms of accuracy, precision, and F1 score. Specifically, the neural network achieved an average accuracy of 82% (± 2% of standard deviation). This suggests that the data derived from inspection in regions with defects exhibits common characteristics among areas with defects of the same depth (concentric areas), enabling the network to accurately classify them. Notably, there is an excellent inter-class sensitivity, particularly in recognizing class C0, which is perfectly distinguished from other classes. The high sensitivity in distinguishing this class is crucial for the correct identification of solid zones and areas with potential detachments.

In light of the practical feedback from deploying the system in situ and considering the excellent classification results of class C0 compared to others, a subsequent analysis led to the development of another classification type - binary classification. The aim was therefore to generalise the results of a single defect to others, in particular between C0 and classes C1, C2, C3, C4, C5, C6 and C7. To



(a) Confusion matrix (b) ROC curve

Fig. 5. (a): Confusion matrix shows the accuracy and precision of the performance in class prediction with easy identification of false positives or false negatives. (b): The ROC curve provides a concise assessment of the model's ability to identify different states of preservation of the architectural layer. The representation of these metrics provides a concise assessment of the model's ability to identify the classes related to the different states of preservation of antique plaster layers.



(a) Confusion matrix for binary classification

Class	Precision	Recall	F1-score	Support
Flaw	0.99	1.00	1.00	1579
Un-flaw	1.00	0.99	0.99	1421
Accuracy			0.99	3000
Macro Avg	1.00	0.99	0.99	3000
Weighted Avg	1.00	0.99	0.99	3000

(b) Classification report for binary classification

Fig. 6. (a): The confusion matrix shows accurate performance for binary classification of the area without flaws compared to the area with flaws. (b): The classification report in terms of mean of 5-fold cross-validation for the binary classification. In this case, the standard deviation was not made explicit due to the low variance values.

balance the dataset, additional samples of class C0 were required. Consequently, dataset 2 was used for training a different neural network with a modified structure. The small architectural changes of the first one used were necessary to adapt it to the new purpose of binary classification, i.e. with a single output neuron.

Test results showed average values of 99% (± 1%) on the test area. Detailed results are shown in Fig. 6.

Once the correct architecture was implemented and validated, the whole dataset was used to train the neural network. The neural network trained on the entire dataset was then used to construct a c-map on the entire portion of the model. In particular, the subsequent acquisition of data on the entire model, independent from the test area was used as the test database needed for the c-map. In detail, the test object was divided into a 16×16 grid, resulting in 256 acquisitions that were classified by the binary classification model. The results belonging to class 0 or 1 were mapped onto the acquisition coordinates and thus allowed the generation of the c-map. This approach made it possible to use the network trained on a single portion (defect A) to inspect the entire test object. This facilitated the creation of a c-map highlighting other defects present in the test artefact, which were not used for training the neural network. The neural network provided consistent results from portions of the material that had never been seen before, meaning it was trained in a different region than the training area. Fig. 7 illustrates the results: the left portion shows the neural network's raw output without interpolation, whereas the right part shows the outcome of interpolation, with the darker zone defining the detachment area identified by the traditional analysis performed by the restorer.

It is noteworthy that the defect with which the neural network was trained was perfectly recognized, along with two other defects, albeit the fourth defect in the top right was partially recognized, as can be observed. The reason for this partial recognition of defect C may be related to the internal collapse of the structure, partly due to the laboratory environmental conditions that affected the setting and curing of the mortars of the test-object.

4. Discussion

The conservation and stabilization of covering layers, whether decorated or not, rely heavily on the expertise of professionals in the field. Their refined skills and extensive experience enable them to detect and classify regions that need of stabilization compared to stable areas. However, the current process is complex and faces numerous challenges.

Entrusting the task to a single individual, especially in the case of extensive mappings, consumes considerable time. On the other hand, involving many restoration specialists, while addressing timing issues, introduces a potential bias in measurement due to variations in experience and precision of inspection. Moreover, the transcription on a map of the information acquired through auscultation is another step prone to interpretation, so it is not repeatable. In addition, the complexity and arduousness of this task are further compounded by the physical strains that arise from prolonged exertion, leading to discomfort and fatigue. Consequently, the current procedure suffers from biases and lacks scalability.

An innovative and ergonomic instrument was therefore developed, together with a specifically designed measuring method (the PICUS System) that replicates the action of the restorer, tapping on the surface to analyse the sound emitted. Sound recognition capabilities, developed through extensive on-site experience by conservators-restorers, were replicated using a convolutional

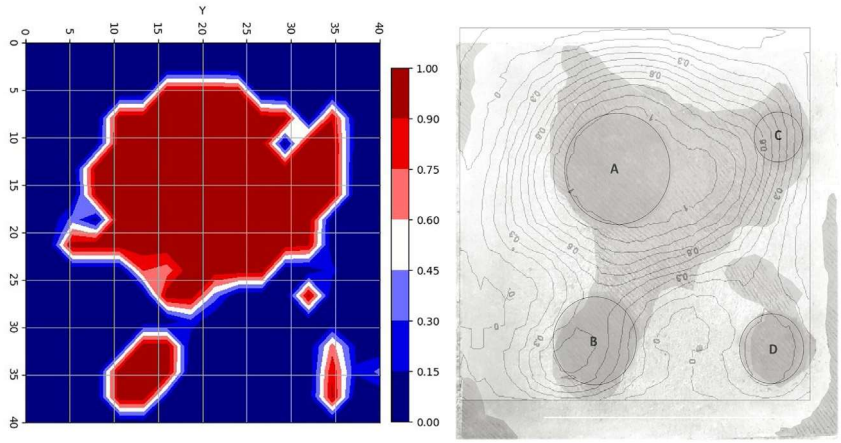


Fig. 7. On the left: the output of the neural network in the test matrix, without interpolation. On the right, a c-map with the (interpolated) defect detection levels based on the outputs of the PICUS system, superimposed on a traditional auscultation map (in grey) and the defect location of the test object.

neural network (CNN) trained on a comprehensive dataset of samples, so far based on a test-object simulating an ancient covering with known defects. These samples were meticulously labelled by experts focusing on detachment severity.

The implementation of neural network has proven highly effective in identifying and classifying detachment types, thereby enhancing the defect mapping process for ancient covering.

To this end, the study examined the potential of a deep learning algorithm to be integrated into the PICUS system to assist restorers in analysing surfaces. The acquisition process carried out by experienced conservators included eight proposed class, from C0 (defect-free area) to C1-C7 (areas with defects of increasing depth). To improve the training process, the experimentally acquired datasets were subjected to a data augmentation procedure designed for this purpose. Subsequently, a k-fold cross-validation was conducted to ensure the reliability and stability of the accuracy measurements. This classification produced interesting results regarding inter-class classification. Notably, during the evaluation process, a particular accuracy was observed in distinguishing the un-flaw class (C0) from others. Additional analysis of results made possible to study the nature of the errors, which were mainly concentrated in the classes contiguous to those under investigation. Although this reduces the overall accuracy, it is still a significant result for the researchers, as the error in the contiguous classes has less impact on the analysis results. Building upon these findings, the decision was made to unify flaws-containing classes, leading to the creation of a useful binary classification system for defect identification. This binary classification approach was then applied in conjunction with the first neural network for the multi-class classification of defect gravity. The second classification performed with a customized neural network achieved remarkably high accuracy results and enabled the identification of defects by means of binary classification. In fact, the C0 class, i.e. the defect-free class, was recognised in both classifications with high accuracy. This aspect is of great relevance in the routine application of the tool, as the distinction between areas that require intervention and areas that do not is of primary importance. High accuracy leads restorers to trust the classification made, with considerable advantages. Following thorough testing and validation procedures, the network was deployed to explore and categorize regions of the test material not included in the training dataset. Encouragingly, the results remained consistent with the desired output, demonstrating the network's ability to accurately identify and categorize defects.

This classification method enabled the generation of a comprehensive defect map (c-map) that covered the entire surface of the test-object, offering valuable insights into the presence and distribution of defects within the material. This comprehensive mapping facilitates informed decision-making during the restoration process, ensuring targeted interventions to preserve and protect ancient layers effectively. Hence, this development represents a significant advance that offers cultural heritage professionals a valuable tool, especially when dealing with large areas.

Moreover, by incorporating tracking capabilities via an infrared camera, it becomes possible to map the area with unparalleled precision. This integration of advanced technology not only enhances the accuracy and detail of the defect map but also provides real-time monitoring and feedback, allowing for prompt and targeted interventions during the restoration process.

The use of the PICUS can also be extended to different surfaces, as the force applied in the percussor unit can be calibrated, so it can be adjusted according to the specific surface to be analysed, thus applying only the minimum force necessary to replicate the restorers' gentle touch. It is important to note that the device is intended for use exclusively by professional restorers. Only professionals possess the necessary expertise to operate the PICUS safely and effectively, ensuring both accurate results and the preservation of the analyzed surfaces. The system now integrates digital technologies, leading to a futuristic AI-PICUS, that can automate repetitive tasks and enhance precision, thereby mitigating measurement biases. The findings of this study align with previous research in the field of cultural heritage conservation. So far, AI can be employed to meticulously analyze images and data to monitor [26] and assess damage to cultural heritage objects [27]. This allows for the proactive identification of potential problems or defects [28], enabling timely interventions to prevent further damage [29]. AI can also play a role in raising public awareness about cultural heritage [30]. This valuable information can then be used to develop and implement targeted conservation plans [8,31]. A revolutionary implementation is in defect detection in materials and infrastructure: leak detection [32], crack identification in metals [33], concrete strength prediction [29], bridge health evaluation [34,35] ultrasonic flaw detection [36], steel corrosion risk modelling [27], and AI-powered structural health monitoring [26].

Despite the above-cited AI applications in cultural heritage, to the best of our knowledge, this implementation represents the pioneering use of a listening device for the identification of plaster detachments across multiple levels, facilitated by AI. This groundbreaking achievement serves as a catalyst for further advancements in the conservation of our invaluable artistic heritage, a legacy from our ancestors. It underscores the imperative to defend and conserve these treasures in the most effective manner possible, leveraging innovative tools such as artificial intelligence.

Overall, the synergy between human expertise, classification techniques, probability assessment, interpolation methods, and infrared tracking presents a comprehensive solution that empowers restorers with the tools and insights necessary to monitor and predict the evolution of the state of conservation together with helping to design complex conservation and restoration projects, with confidence and precision.

5. Conclusions

The AI-based method outlined in this paper offers numerous advantages compared to the traditional auscultation approach in use within the community of professionals dedicated to conservation and restoration, who daily have to face sensitivity to boundary conditions and potential physical strain. Although operated by humans, the AI-based method is less susceptible to biases and boundary conditions. It operates on objective criteria, enhancing the repeatability of condition assessments. This can lead to better resource planning and more efficient use of funds.

The AI method demonstrates remarkable efficiency and accuracy, thanks to the advanced capabilities of artificial intelligence. By leveraging AI algorithms, the method can swiftly and precisely identify defects and areas in need of intervention, streamlining the overall process.

We have introduced an innovative AI-based method tailored for the conservation of cultural heritage multilayered structures and renders, demonstrating its effectiveness through real-world case studies. This method has proven instrumental in generating detailed defect maps, empowering conservators-restorers to work with enhanced efficiency and precision.

The significance of this AI-based approach extends beyond its immediate applications. Future endeavours will concentrate on refining the method's accuracy and implementing tests on cultural heritage cases of study. The journey towards harnessing AI's full potential in restoration endeavours is just beginning, promising exciting prospects for the future of heritage conservation.

The AI-based method offers efficiency, accuracy, objectivity, accessibility, and versatility, representing a significant advancement in the field of restoration. By embracing this innovative approach, the restoration community can enhance their preservation efforts and safeguard our cultural heritage for future generations.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.culher.2024.07.012.

References

[1] G. Caliano, F. Mariani, P. Calicchia, PICUS: a pocket-sized system for simple and fast non-destructive evaluation of the detachments in ancient artifacts, Appl. Sci. 11 (8) (2021) 3382.

[2] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT press, 2016.

[3] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE Trans. Neural Netw. Learn.Syst. 33 (12) (2021) 6999–7019.

[4] N. Aloysius, M. Geetha, A review on deep convolutional neural networks, in: 2017 International Conference on Communication and Signal Processing (ICCSP), IEEE, 2017, pp. 0588–0592.

[5] M. Krichen, Convolutional neural networks: a survey, Computers 12 (8) (2023) 151.

[6] T. Wiatowski, H. Bölcskei, A mathematical theory of deep convolutional neural networks for feature extraction, IEEE Trans. Inf. Theory 64 (3) (2017) 1845–1866.

[7] Y.H. Liu, Feature extraction and image recognition with convolutional neural networks, in: Journal of Physics: Conference Series, 1087, IOP Publishing, 2018, p. 062032.

[8] X. Ye, T. Jin, C. Yun, A review on deep learning-based structural health monitoring of civil infrastructures, Smart Struct. Syst 24 (5) (2019) 567–585.

[9] B. Zhu, Nondestructive testing method of engineering material defects based on artificial intelligence algorithm, in: 2023 2nd International Conference on 3D Immersion, Interaction and Multi-sensory Experiences (ICDIIME), IEEE, 2023, pp. 302–307.

[10] G. Caliano, F. Mariani, P. Calicchia, PICUS: a pocket-sized system for simple and fast non-destructive evaluation of the detachments in ancient artifacts, Appl. Sci. (Switzerland) 11 (8) (2021), doi:10.3390/app11083382.

[11] F. Mariani, A.S. Savoia, G. Caliano, An innovative method for in situ monitoring of the detachments in architectural coverings of ancient structures, J. Cult. Heritage 42 (2020), doi:10.1016/j.culher.2019.07.013.

[12] G. Caliano, F. Mariani, A. Salvini, A portable and autonomous system for the diagnosis of the structural health of cultural heritage (PICUS), 2023. ISBN: 978-92-990090-6-2

[13] G. Caliano, F. Mariani, F. Vitali, P. Pogliani, The "PICUS" system in the detection of defects on panel paintings and wooden boards, 2022 IEEE International Ultrasonics Symposium (IUS), 2022, doi:10.1109/IUS54386.2022.9958892.

[14] F. Mariani, C. Giosué, F. Di Stasio, P. Pogliani, Evaluation of detachment between layers of ancient plaster renderings: comparison between the traditional technique and a new and innovative automated procedure called PICUS, 2024, doi:10.26650/B/AA9PS34.2024.006.014.

[15] F. Li, M. Liu, Y. Zhao, L. Kong, L. Dong, X. Liu, M. Hui, Feature extraction and classification of heart sound using 1D convolutional neural networks, EURASIP J. Adv. Signal Process. 2019 (1) (2019) 1–11.

[16] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D.J. Inman, 1D convolutional neural networks and applications: a survey, Mech. Syst. Signal Process. 151 (2021) 107398.

[17] O. Avci, O. Abdeljaber, S. Kiranyaz, D. Inman, Structural damage detection in real time: implementation of 1D convolutional neural networks for SHM applications, in: Structural Health Monitoring & Damage Detection, Volume 7: Proceedings of the 35th IMAC, A Conference and Exposition on Structural Dynamics 2017, Springer, 2017, pp. 49–54.

[18] J. Li, J.-h. Cheng, J.-y. Shi, F. Huang, Brief introduction of back propagation (BP) neural network algorithm and its improvement, in: Advances in Computer Science and Information Engineering: Volume 2, Springer, 2012, pp. 553–558.

[19] Keras Sequential Model, 2024, (https://keras.io/guides/sequential_model/). Accessed: April 10.

[20] G. TensorFlow, Large-scale machine learning on heterogeneous systems, 2015, (https://tensorflow.org).

[21] N. Developers, NumPy: The fundamental package for scientific computing with Python, 2023. https://numpy.org/doc/stable/reference/generated/numpy.roll.html.

[22] C.F.G.D. Santos, J.P. Papa, Avoiding overfitting: a survey on regularization methods for convolutional neural networks, ACM Comput. Surv. (CSUR) 54 (10s) (2022) 1–25.

[23] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.

[24] L. Taylor, G. Nitschke, Improving deep learning with generic data augmentation, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 1542–1547.

[25] Scikit-learn: KFold documentation, 2024, (https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html). Accessed: April 10.

[26] A. Sabato, S. Dabetwar, N.N. Kulkarni, G. Fortino, Noncontact sensing techniques for AI-aided structural health monitoring: a systematic review, IEEE Sens. J. 23 (5) (2023) 4672–4684.

[27] J.-S. Chou, N.-T. Ngo, W.K. Chong, The use of artificial intelligence combiners for modeling steel pitting risk and corrosion rate, Eng. Appl. Artif. Intell. 65 (2017) 471–483.

[28] Z. Wang, D. Zhu, An accurate detection method for surface defects of complex components based on support vector machine and spreading algorithm, Measurement 147 (2019) 106886.

[29] Y.R. Wang, Y.L. Lu, D.L. Chiang, Adapting artificial intelligence to improve in situ concrete compressive strength estimations in rebound hammer tests, Front. Mater. 7 (2020) 568870.

[30] D. Schreiber, C. Picus, D. Fischinger, M. Boyer, The defalsif-AI project: protecting critical infrastructures against disinformation and fake news, Elektrotech. Informationstech. 138 (7) (2021) 480.

[31] P.G. Asteris, V.G. Mokos, Concrete compressive strength using artificial neural networks, Neural Comput. Appl. 32 (15) (2020) 11807–11826.

[32] B. Ahn, J. Kim, B. Choi, Artificial intelligence-based machine learning considering flow and temperature of the pipeline for leak early detection using acoustic emission, Eng. Fract. Mech. 210 (2019) 381–392.

[33] A. Ali, B. Hu, O.M. Ramahi, Intelligent detection of cracks in metallic surfaces using a waveguide sensor loaded with metamaterial elements, Sensors 15 (5) (2015) 11402–11416.

[34] S. Yehia, O. Abudayyeh, I. Abdel-Qader, A. Zalt, Ground-penetrating radar, chain drag, and ground truth: Correlation of bridge deck assessment data, Transp. Res. Record. 2044 (1) (2008) 39–50, doi:10.3141/2044-05.

[35] S. Yehia, O. Abudayyeh, I. Abdel-Qader, A. Zalt, Ground-penetrating radar, chain drag, and ground truth: correlation of bridge deck assessment data, Transp. Res. Record. 2044 (1) (2008) 39–50.

[36] S. Cantero-Chinchilla, P.D. Wilcox, A.J. Croxford, Deep learning in automated ultrasonic NDE–developments, axioms and opportunities, NDT & E Int. 131 (2022) 102703.