

<https://doi.org/10.1038/s40494-025-02146-3>

# Machine learning assisted real-time acoustic monitoring of laser cleaning in Heritage conservation

Check for updates

Georgios D. Barmparis<sup>1</sup> , Anastasios-Nikolaos Raikidis<sup>2</sup>, Kristalia Melessanaki<sup>1</sup>, Giorgos P. Tsironis<sup>1,2</sup>, George J. Tserevelakis<sup>1,3</sup> & Paraskevi Pouli<sup>1</sup>

Laser-assisted cleaning has become an indispensable tool in heritage conservation due to its precision, control, and environmentally friendly nature. However, the complexity of deposition layers and the fragile condition of original surfaces necessitate careful monitoring to avoid irreversible damage. This work explores the integration of machine learning-assisted real-time acoustic monitoring in laser cleaning processes to enhance conservation efforts. By combining acoustic signals generated during laser-material interaction with machine learning, we elevate the precision and reliability of laser cleaning in the delicate context of cultural heritage restoration.

Over the past 30 years, laser-assisted removal of unwanted material has become a common cleaning tool in Heritage conservation, often replacing conventional methods based on chemicals and mechanical action. This is due to its unique advantages, including selective and gradual material removal, high precision and control, and its environmentally friendly nature<sup>1,2</sup>. However, this delicate and irreversible process requires careful selection of irradiation parameters and a thorough understanding of ablation mechanisms, especially given the complex nature of deposition layers and the fragile condition of original heritage surfaces.

Self-limiting processes control the near-infrared (NIR) laser cleaning of black pollution crusts from stonework, ensuring that the cleaning intervention halts immediately after the unwanted crust is removed. Consequently, laser cleaning has become increasingly popular in restoration, as it effectively and safely reveals the original surfaces of cultural heritage assets, such as the Acropolis Sculptures in Athens<sup>2,3</sup>. Successful applications of laser cleaning have also been reported for various materials<sup>1–7</sup>, including the removal of burial crusts from stone sculptures and leather objects, cleaning soiling from paper, and eliminating corrosion layers from historical metal objects<sup>5</sup>. Additionally, it has proven effective in removing tarnishing from gilded silver threads<sup>6</sup>, among other uses.

Laser cleaning is particularly valuable in addressing a wide range of cleaning challenges. These often involve situations in which self-limiting conditions may not apply, i.e., challenges in which the critical parameters needed to remove unwanted material are very close to, or even exceed, the thresholds that could damage the original substrate<sup>7</sup>. One of the most challenging areas of laser cleaning research is the removal of aged or polymerized varnish and overpainting layers from artworks, particularly those that are painted. These layers are composed of polymeric materials that are

transparent to near-infrared (NIR) laser radiation. In contrast, most degraded polymeric overlays highly absorb ultraviolet (UV) laser radiation, making them an effective alternative for layer-by-layer ablation and the controlled removal of unwanted materials. However, the fact that varnish films and paint layers exhibit very similar optical properties, absorbing UV radiation highly, poses limitations to their laser cleaning, as self-limiting conditions do not apply. For this reason, a critical assessment of the cleaning process and real-time monitoring of its progress are crucial.

Over the years, several imaging and spectroscopic techniques have been explored to assess the results and monitor the cleaning process in heritage conservation. Imaging techniques offer a non-destructive and non-invasive approach and are the most widely used. Colorimetry, multispectral reflectography, and optical coherence tomography (OCT) have been employed to assess the color<sup>8,9</sup> and stratigraphy<sup>10</sup> of the objects under study and during restoration. Microscopies, including optical microscopy (OM), scanning electron microscopy (SEM), and atomic force microscopy (AFM), have also been used to investigate the surface morphology of the treated objects<sup>11</sup>. In parallel, the need for minimal chemical alteration to the underlying original surfaces has influenced several studies that employed spectroscopic techniques, such as Raman spectroscopy and Fourier Transform Infrared Spectroscopy (FTIR)<sup>12</sup>, as well as gas chromatography/mass spectrometry (GC-MS), to investigate the chemistry of the cleaned surfaces. Significant research has also been focused on the complementary application of imaging and chemical techniques for a thorough approach to cleaning assessment<sup>13</sup>.

Careful post-treatment assessment and monitoring during the restoration process are becoming increasingly critical in the context of laser cleaning. While laser cleaning offers precision, it also presents challenges,

<sup>1</sup>Institute of Electronic Structure and Laser, Foundation for Research and Technology-Hellas, Heraklion, Crete, Greece. <sup>2</sup>Department of Physics, University of Crete, Heraklion, Greece. <sup>3</sup>Department of Biology, University of Crete, Heraklion, Greece. e-mail: [barmparis@physics.uoc.gr](mailto:barmparis@physics.uoc.gr); [ppouli@iesl.forth.gr](mailto:ppouli@iesl.forth.gr)

especially in finding the right balance between effective cleaning and preserving the substrate. In industrial applications, the effectiveness of the cleaning process is evaluated based on two key factors: temporal aspects, such as cleaning time and removal rates, and spatial aspects, which focus on the complete elimination of unwanted materials<sup>14–16</sup>. However, in heritage conservation, the primary concern is to protect the integrity and chemistry of the original surface.

Significant research has focused not only on the ideal in-situ evaluation of the results<sup>17–20</sup> but also on real-time monitoring of the process<sup>21</sup>. In contrast to industrial applications, where monitoring typically depends on post-process assessments of cleaning effectiveness and removal rates, laser cleaning of heritage objects requires a different focus. The most critical factor is to identify the key parameters to ensure irradiation stops at the right time, thereby preserving the original surface. Having achieved this knowledge, we will then focus our attention on developing the relevant protocol to control (continue or terminate) the procedure.

Acoustic monitoring of the laser cleaning process offers unique advantages, particularly because it provides real-time results<sup>21–25</sup>. Acoustic signals are extremely sensitive to material modifications as they are strongly dependent on the effective optical absorption coefficient of the irradiated region. This property renders acoustic emissions ideal for the accurate detection of the transition between the unwanted layer and the substrate, minimizing thus potential side-effects. On top of this, the technique is completely insensitive to typical external optical noise and does not require clear visual access, therefore, it can be employed outdoors even in dusty and humid environments where the use of traditional optical equipment (e.g., spectrometers, cameras, etc.) is highly challenging. Early studies<sup>26–29</sup> highlighted its potential, and recent research efforts have increasingly focused on this concept<sup>22–25</sup>.

This method involves recording and processing the acoustic signals generated when light interacts with matter. Previous studies<sup>23</sup> have demonstrated a correlation between the amplitude of the acoustic wave and the amount and composition of removed material. This relationship enables monitoring of the cleaning process progress, as the acoustic signal changes gradually when the overlayer is reduced, followed by an abrupt change upon reaching the substrate. One of the key advantages of this method is that the acoustic signals can be easily and non-invasively detected using non-contact air-coupled ultrasonic transducers, which have a frequency response in the MHz range. These transducers can be placed near the laser-treated surfaces, enabling online and real-time measurements without interrupting the cleaning process for analysis and data recording. The effectiveness of this approach has been tested on a series of technical encrusted marble coupons<sup>23</sup> and varnished painting mock-ups<sup>25</sup>, which simulate various cleaning challenges. The relevant results have been cross-checked using other analytical techniques. The proposed acoustic monitoring strategy enables the identification of the critical laser pulse responsible for complete encrustation removal. This capability can assist conservators and restorers (C-Rs) in

monitoring laser treatments and making informed decisions regarding the progress of the crust or overlayer removal process.

Artificial Intelligence and Machine Learning (ML) algorithms are increasingly integrated into scientific research and industrial applications<sup>30,31</sup>, offering powerful tools for data analysis, process optimization, and decision-making. ML models can identify patterns and correlations within experimental data, and they have been employed to various applications in cultural heritage research and conservation<sup>32–38</sup>, such as to identify cracks<sup>34</sup> and other deterioration features<sup>35</sup>, to identify rising damp on monuments promptly<sup>36</sup> or to study certain type of objects<sup>38</sup> to name a few.

Herein, we discuss the implementation of ML in monitoring the laser cleaning process and ensuring a safe intervention on objects of heritage value. Among the available ML algorithms, we choose the Random Forest algorithm (RF), introduced by Breiman in 2001<sup>39</sup>, as an extension to the “the random subspace” method by Tim Kam Ho<sup>40</sup>, which does a random selection of a subset of features to grow each decision tree. RF is an explainable tree-based ensemble learning method that has emerged as one of the most potent and versatile machine learning algorithms for classification and regression tasks. Among its advantages, an RF algorithm can handle high-dimensional datasets, model complex nonlinear relationships, perform well with limited amounts of data, and provide insights through feature importance metrics. When cleaning a surface, unwanted material is removed, leading to changes in the surface properties, such as the absorption coefficient of the surface material, which are reflected in the acoustic signal of the following laser pulse. By using the amplitude at each time step of the acoustic signal as the input feature for the RF, we can identify the time steps that contain critical information about surface changes during cleaning, making it essential for monitoring the cleaning process.

Previous works have applied ML to optimize laser parameters in industrial cleaning applications<sup>41–44</sup> or to monitor and control of the breakthrough stage in laser drilling<sup>45</sup> and evaluate the cleaning results<sup>44–47</sup>. However, in Heritage science and conservation practice, the focus shifts away from material removal efficiency, as seen in industrial applications. The primary goal in heritage conservation is to determine the precise moment to stop the cleaning process, thereby preventing damage to the underlying historical surface that must be preserved and revealed. This approach differentiates conservation work from industrial processes and is consistently followed in this context.

This study presents the first feasibility study to monitor the restoration process of sensitive cultural heritage objects in real-time. It combines acoustic wave signals produced during laser cleaning and explainable ML algorithms to identify the pulse that completely removes the unwanted paint without damaging the substrate.

The methodology, including details on individual irradiation and monitoring parameters, as well as processing procedures and machine learning algorithms used in this study, is outlined in Fig. 1.

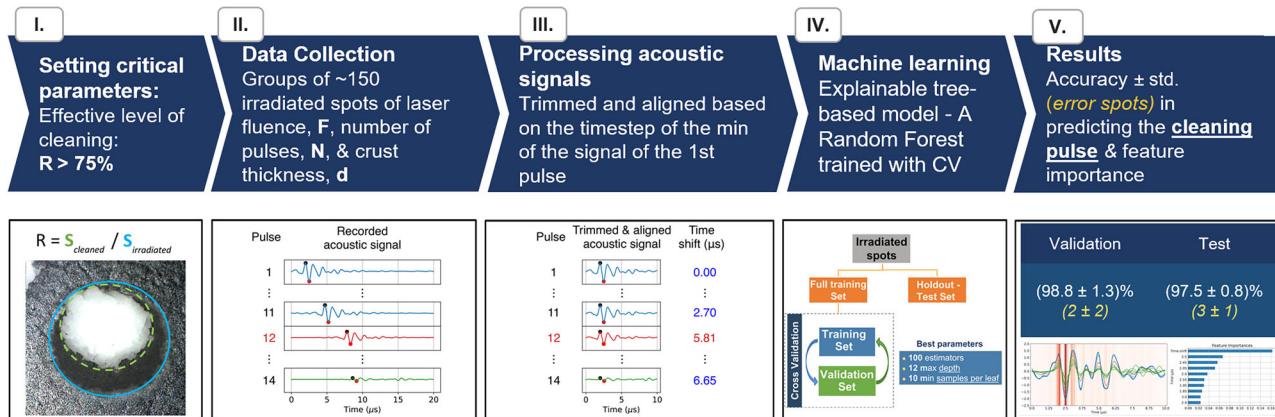


Fig. 1 | The methodology employed in this study. A brief schematic outline.

## Methods

### Technical mock-ups

A series of technical mock-ups were created for data-gathering purposes [Fig. 2 (left)]. The base was made from thin slabs of white marble sourced from the Greek island of Thasos in the Northern Aegean Sea. Black acrylic paint (Motip Matt Black Acrylic Varnish) was sprayed to the marble from approximately 30 cm. The mock-ups were then left to dry naturally for at least one week before conducting irradiation experiments. This black acrylic overlayer, intended to simulate the “unwanted crust,” was applied in multiple layers, resulting in varying thicknesses ranging from 8 to 55  $\mu\text{m}$ . The mock-up design facilitated simple laser ablation processes characterized by self-limiting conditions.

### Cleaning methodology and criteria for quality assessment

Laser cleaning is affected by several parameters that are directly related to the materials involved<sup>2</sup>. The most critical parameters are the appropriate wavelength,  $\lambda$ , and pulse duration,  $\tau_p$ , of the laser system, the laser fluence,  $F$ , and the number of laser pulses applied,  $N$ . The wavelength and the pulse duration ensure the laser ablation mechanism is well-suited to the specific material.  $F$  is the energy,  $E$ , per unit area,  $S$ , of the irradiated surface,  $F = E/S$  ( $\text{J}/\text{cm}^2$ ). In addition to the laser fluence, we need to define the laser fluence threshold,  $F_{\text{thr}}$ , i.e., the minimum fluence required to remove the unwanted material, and the critical laser fluence,  $F_{\text{damage}}$ , i.e., the fluence above which the laser pulse causes damage to the underlying substrate. Finally,  $N$  refers to the total number of sequential laser pulses required to remove the unwanted material, which is closely related to the thickness of the unwanted layer or crust and the selected laser fluence.

In this study, a QS Nd:YAG laser system emitting at  $\lambda = 1064$  nm, with pulses of  $\tau_p = 10$  ns, was employed. The cleaning efficiency of the given technical mock-ups has been well studied and characterized<sup>23</sup>. The laser fluence threshold was found to be  $F_{\text{thr}} = 0.1 \text{ J}/\text{cm}^2$ , and the critical laser fluence,  $F_{\text{damage}} = 2.1 \text{ J}/\text{cm}^2$ .

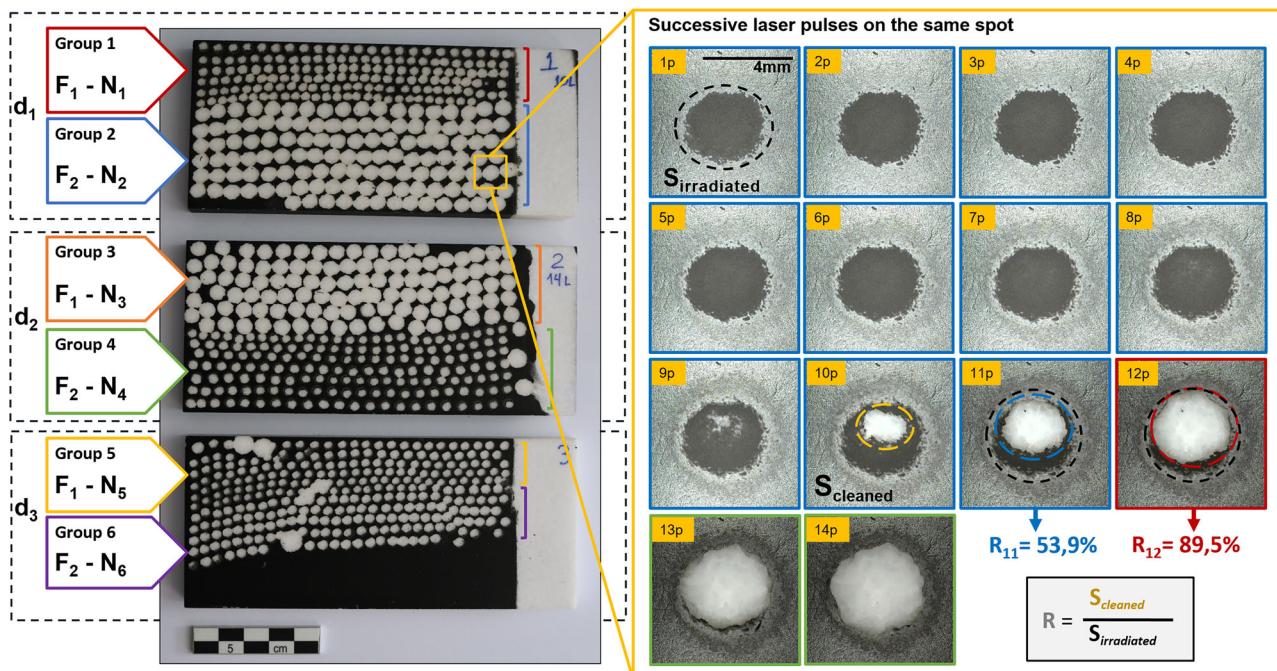
During the data collection process, several values of  $F$  were carefully chosen between the  $F_{\text{thr}}$  and the  $F_{\text{damage}}$  thresholds to investigate the

cleaning process at different paces. Laser pulses with low laser fluence slowly remove the unwanted material; thus, a high number of laser pulses is required to clean the surface, resulting in a slow pace. In contrast, laser pulses with high laser fluence transfer more energy to the irradiated surface, thus requiring fewer laser pulses to clean it, resulting in a high pace. Another parameter that affects the pace of the cleaning process is the thickness,  $d$ , of the unwanted material. The cleaning of materials with different thicknesses requires varying numbers of laser pulses at the same laser fluence.

In this experiment, mock-ups characterized by a thickness,  $d_i$ , of black crust are irradiated with a given laser fluence,  $F_i$ , and a given number of applied laser pulses,  $N_i$ . With this method, we collect data (acoustic signals) from groups of approximately 150 identical irradiated spots, all with the same overlayer thickness,  $d_i$ , laser fluence,  $F_i$ , and number of laser pulses,  $N_i$ . In Fig. 2, we present a schematic outline of the developed mock-ups and the various groups considered in this work, including thickness, laser fluence, and the number of laser pulses.

Successive laser pulses lead to increasingly removing the unwanted material from the surface of the mock-ups, as can be seen in Fig. 2 (right). At this point, it is essential to highlight that an uncontrolled cleaning process can damage the underlying surface in the context of laser cleaning in Heritage conservation.

To avoid damaging the substrate, we introduce the definition of the *cleaning pulse*, which is the pulse that sufficiently cleans the surface without damaging the substrate and is defined as the pulse that removes at least 75% of the irradiated area. The ratio of cleaned to irradiated areas, which reflects cleaning efficiency, has been set at 75% to prevent overexposure of the central area due to the Gaussian profile of the laser system used. Additionally, this percentage facilitates an automated scanning procedure that requires substantial overlap between adjacent irradiated spots. We determine the ratio  $R$  between the *cleaned area*,  $S_{\text{cleaned}}$ , and the *irradiated area*,  $S_{\text{irradiated}}$ , by processing the collected images with the ImageJ software. In our experiments, we irradiate a few additional laser pulses after identifying the cleaning pulse to investigate the material’s response beyond effective cleaning.



**Fig. 2** | Left: Photographs of the developed mock-ups and schematic outline indicating the different groups of parameters that were varied in this study; thickness of the graffiti overlayer ( $d_i$ ), laser fluence values ( $F_i$ ), and number of applied laser pulses ( $N_i$ ). Right: Images of the evolution of the cleaning process upon successive laser

pulses on the same spot. A schematic outline of the calculation of the ratio  $R$ , which reflects cleaning efficiency, and its value calculated for spots #11 and #12 is also shown. The diameter of the irradiated area is 4 mm.

In Fig. 2, right, we present the digital microscope images taken after each of the laser pulses at a given spot with a paint thickness of  $d = 33 \mu\text{m}$ , irradiated with a laser fluence of  $F = 0.7 \text{ J/cm}^2$ . The white area (exposed marble) is the cleaned area, while the black area is the total irradiated area. It's evident that the surface undergoes modification after the first irradiation pulse; however, the graffiti paint layer requires additional pulses for its gradual removal. In this instance (Fig. 2, right), the marble is revealed after the eighth (#8) pulse. By Pulse #11, 53.9% of the irradiated area has been cleaned, and by Pulse #12, 89.5% has been cleaned. Consequently, Pulse #12 is identified as the cleaning pulse.

## Experimental setup

A schematic representation of the experimental setup is shown in Fig. 3. The setup combines laser ablation and acoustic recording modalities. In all experiments, the relative position of the irradiated spot to the laser focusing lens is 32 cm. The acoustic piezoelectric transducer is located approximately 6 cm above and to the right of the spot at a 45-degree angle.

A Q-Switched Nd:YAG laser (LITRON Lasers, TRLi 850 Series, Rugby, Warwickshire, England) was employed for performing the laser irradiation treatments. The laser system operated at the fundamental wavelength of 1064 nm, emitting pulses of 10 ns and with a variable repetition rate ranging from 1 to 4 Hz. The energy fluence values on the mock-ups ranged from  $0.2 \text{ J/cm}^2$  to  $1.2 \text{ J/cm}^2$  and were estimated by measuring the spot size ( $\sim 0.26 \text{ cm}^2$ ) of the focused beam on black photographic paper. All the irradiation experiments were performed in dry conditions.

Optical imaging of the ablated regions was performed by utilizing a portable digital microscope (Dino-Lite Edge AM4113TFV2W) with a magnification in the range of  $\times 50$  to  $\times 200$ . The depth of the laser-induced craters and the thickness of the over-layers were measured by means of a Portable Surface Roughness Tester profilometer (Mitutoyo America Corporation, Surftest SJ-410 Series, Aurora, IL, USA).

The laser-induced acoustic response on the examined mock-ups was detected by an air-coupled transducer (NCT1-D7-P10, The Ultran Group, State College, PA, USA; nominal central frequency: 1 MHz; focal distance: 10 mm; numerical aperture: 0.31) placed at approximately 45 degrees in respect to the horizontal plane and in an out-of-focus position around 4 cm away from the irradiation region to avoid signal saturation effects. The signals were subsequently enhanced by two radio frequency amplifiers (TB-414-8A +, Mini-Circuits, Camberley, England; gain: 31 dB) connected in series prior to their digitization by an oscilloscope (DSO7034A, Agilent Technologies, Santa Clara, CA, USA; bandwidth: 350 MHz; sample rate: up to 2 GSamples/s) which, in turn, was connected to a laptop computer equipped with custom-made software controlling the measurement

procedures. The recorded waveforms were sampled at 1000 points over a  $50 \mu\text{s}$  temporal window (corresponding to a 20 MSamples/s sampling rate) and bandpassed between 0.5 and 2 MHz to reduce mainly high-frequency noise before being saved to the computer as ASCII files. Recording synchronization was achieved through the trigger output of the laser source, which was connected to the oscilloscope's second channel.

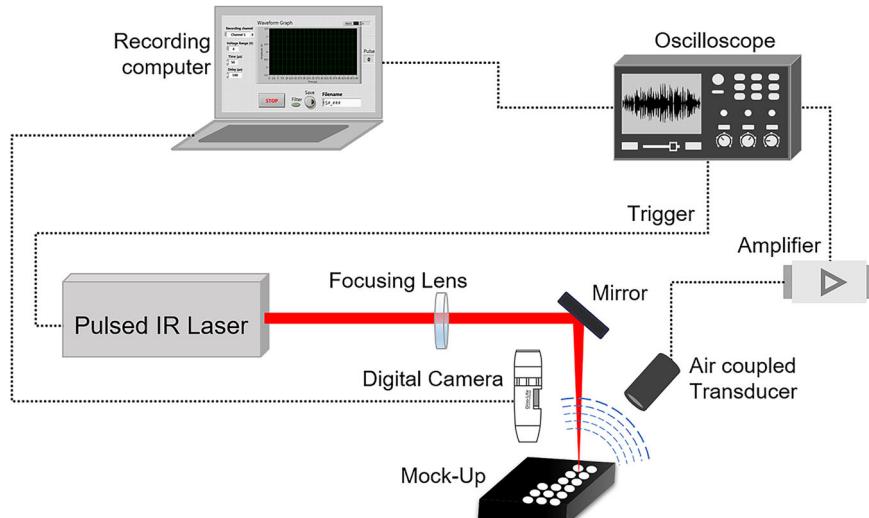
During data collection, a total of 1131 spots were irradiated on the studied mock-ups of black paint overlays of various thicknesses. Each spot was exposed to varying sets of laser fluence,  $F_i$ , and number of laser pulses,  $N_i$ . The recorded waveform corresponding to each laser pulse incidence represents the detected acoustic pressure amplitude as a function of time, where positive and negative values can be interpreted as the compression and rarefaction regions of the propagating ultrasonic wave, respectively. The peak-to-peak amplitude, quantified as the difference between the maximum and minimum pressure values, has been demonstrated<sup>1</sup> to be directly proportional to the effective optical absorption coefficient at the irradiation wavelength of 1064 nm under typical energy fluence conditions. This relationship arises due to the localized thermoelastic expansion and subsequent generation of broadband ultrasonic waves following transient optical energy deposition. Therefore, the peak-to-peak amplitude parameter, along with the acoustic perturbation's time-of-flight, carries essential information on the optical absorption and structural characteristics of the irradiated region, enabling the precise real-time monitoring of the laser cleaning process when processed with the proposed machine learning models. For each laser ablation pulse, the corresponding acoustic signal responses were recorded. Simultaneously, the cleaning result was visually assessed and documented with a portable digital microscope. This approach allowed for a detailed analysis of the surface interactions that occurred with each individual ablation pulse.

In Fig. 4, the acoustic signals recorded upon irradiation on the same spot for successive pulses are shown. The results are in agreement with the fact that upon ongoing irradiation, the peak-to-peak amplitude is much smaller, and there is a further shift to the right. After Pulse #12, a noticeable drop in amplitude can be observed.

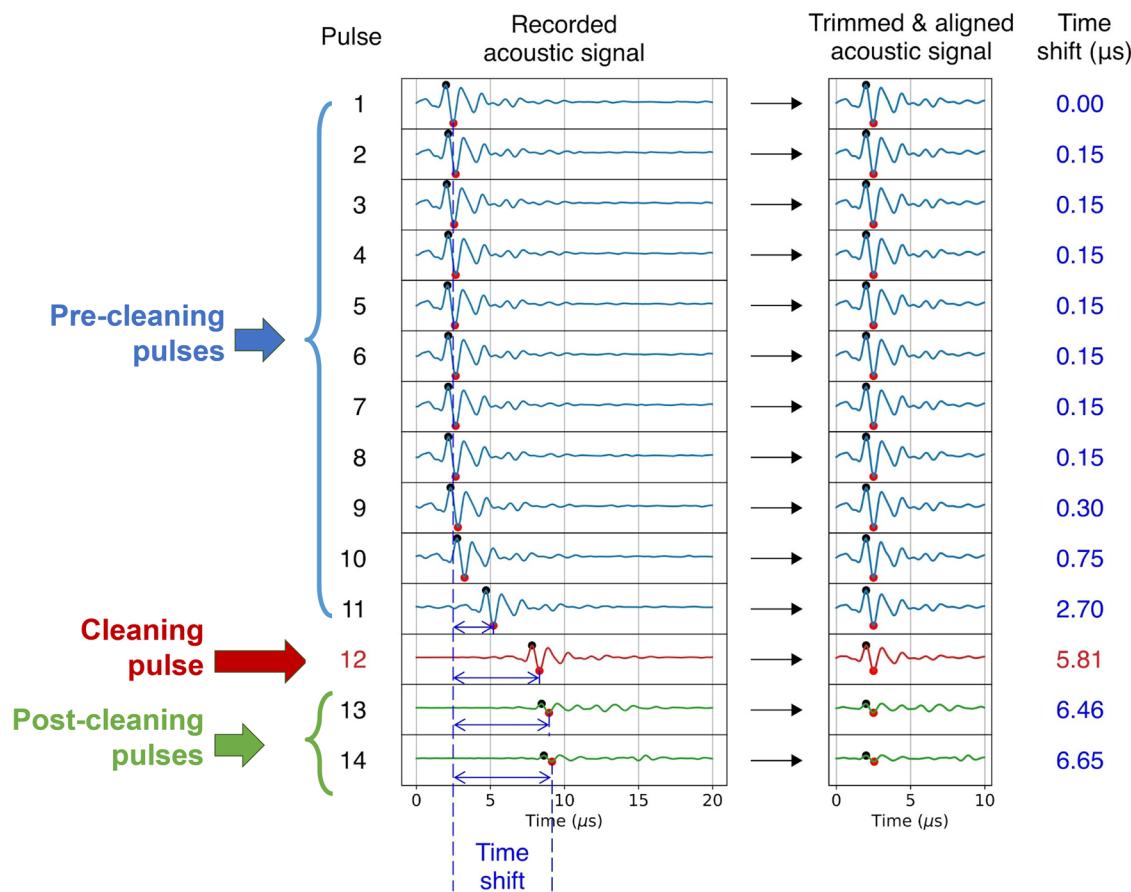
## Data pre-processing

We preprocess the recorded signals to guarantee a common starting point in time and a consistent duration. This standardization process ensures that our data representation remains independent of the setup used (such as the distance between the cleaning surface and the ultrasonic detector), the thickness of the paint being removed (which also affects the distance from the cleaning point to the ultrasonic detector), and the fluence of the cleaning laser pulse. Additionally, events occurring on the surface during the cleaning

**Fig. 3 |** Schematic representation of the experimental setup.



## Successive laser pulses on the same spot



**Fig. 4 | The recorded and the preprocessed acoustic signals of successive laser pulses.** From left to right: Each recorded signal is aligned based on the time step of its global minimum amplitude (the reference time step), and its duration is trimmed between 2  $\mu$ s before and 8  $\mu$ s after the reference time step. The duration between the reference time step of each pulse with respect to the reference time step of the first

pulse is extracted as the time shift of the acoustic signal (blue). From top to bottom: The signal before effective cleaning (cyan, pre-cleaning pulses #1–#11), the signal of the cleaning laser pulse (red, #12), and the signal after the defined critical cleaning level (green, post-cleaning pulses #13–#14).

process are recorded simultaneously across all signals. In other words, among two aligned and trimmed acoustic signals, each time step reflects the same effects on the cleaning surface. Thus, the amplitude of each time step can be used as a descriptor of the cleaning process.

We start by identifying the time step corresponding to the global minimum amplitude of each acoustic signal as the “reference time step” (RTS), and we use it to align and trim the duration of each acoustic signal. The duration of the acoustic signal is defined to be 10  $\mu$ s, consisting of 2  $\mu$ s before the reference time step and 8  $\mu$ s after it. This duration is sufficient to capture all the characteristics of the acoustic wave generated at a given spot during the cleaning process. By aligning the recorded signals based on their RTS, we observe a loss of information on the time delay of each signal reaching the sound detector. This time delay has a physical meaning since it contains valuable information on the amount of paint removed by each laser pulse. During the cleaning process, each laser pulse removes a portion or layer of the unwanted paint. Consequently, each subsequent laser pulse initiates an acoustic signal deeper within the cleaning spot. This means that the acoustic signal from each later laser pulse takes longer to reach the sound detector, resulting in a time delay in the recorded signals. We define the time shift of each acoustic signal as the time duration between the RTS of each signal and the RTS of the signal of the first laser pulse. We extract the time shift of each signal and treat it as an additional input feature in

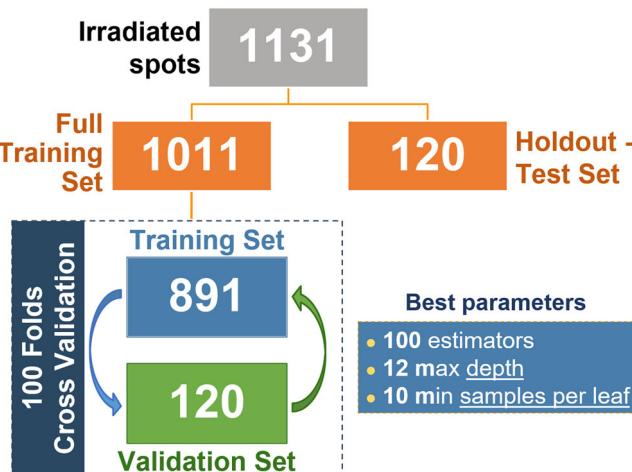
the machine learning algorithm alongside the 200 time steps (10  $\mu$ s at 0.05  $\mu$ s per time step) of each acoustic signal.

In Fig. 4, we present the preprocessing of the acoustic signals of a set of sequential laser pulses at a given cleaning spot. On the left, we present the raw recorded signal of each laser pulse, and on the right, the aligned and trimmed signal with the corresponding extracted time shift.

### Machine learning

We analyze our data using an explainable, tree-based machine learning algorithm, specifically the Random Forest (RF). We chose an RF model due to its ability to handle complex data and perform well even with limited data. Additionally, an RF model can offer insights into how the cleaning process is captured in an acoustic signal. Figure 5 shows a schematic representation of the training process.

The aligned and trimmed acoustic signals from the 1131 cleaned spots were divided into a test set of 120 spots, serving as the hold-out set to evaluate the performance of the trained model, and a training set of 1011 spots to train it (Fig. 5). To eliminate the dependence of the model’s performance on a specific training set, we employed a 100-fold cross-validation technique by further splitting the initial training set into 100 random folds, each consisting of 891 training spots and 120 validation spots. Then, we train an RF model on each fold. This method helps the model remain resilient to potential outliers in the data and provides a



**Fig. 5 | Schematic representation of the 100-fold cross-validation training process of the Random Forest algorithm.** After optimizing the hyperparameters of the model, we conclude with a model that features 100 estimators (decision trees), a maximum depth of 12 for each tree, and a minimum number of samples per leaf of 10.

**Table 1 | Mean accuracy and standard deviation over the RF models for the validation and the test set, and the corresponding number of errors in predicting the cleaning pulse of each spot**

	Validation	Test
Prediction accuracy	$(98.8 \pm 1.3)\%$	$(97.5 \pm 0.8)\%$
Error predictions in 120 spots	$2 \pm 2$	$3 \pm 1$

statistical estimation of its errors. To maintain consistency, we keep all the acoustic signals of an irradiated spot in one set, either the training, validation, or test set. We train each RF model using the amplitude at each of the 200 time-steps of the aligned and trimmed acoustic signal and the time shift of each signal as input features (totaling 201 features) to predict whether the next laser pulse will be the cleaning one [see Fig. 4]. During the training process, the recorded acoustic signal of each consecutive pulse on an irradiated spot is preprocessed by extracting its time shift and aligning it and labeled as either a cleaned pulse or not. Next, the training dataset is created, where the amplitude of the 200 time steps of an acoustic signal and its time shift are used as inputs to predict the label (cleaned or not cleaned) of the next pulse. For each spot, the RF will try to identify the one cleaning signal among several pre- and post-cleaning pulses. The RF was used as implemented in the Scikit-Learn library<sup>48</sup>. Notably, the algorithm has no information regarding the fluence of the laser pulse and the thickness of the unwanted paint.

Subsequently, we fine-tuned the hyperparameters of the RF, including the number of estimators, i.e., the number of decision trees, testing the values [50, 100, 200, and 300], the maximum allowed depth of each tree, testing the values [1, 4, 8, 12, 16, 20, and 22], and the maximum number of samples per leaf of each tree, testing the values [4, 5, 10, and 20]. We define the accuracy of predicting the cleaning pulse of a given spot as the optimization metric. Based on the model's performance on the validation set, i.e., how many of the 120 signals of the cleaning pulses were correctly classified as cleaning signals, we found that the optimal RF model consists of 100 estimators and has a maximum depth of 12 with a maximum number of samples per leaf equal to 10.

After completing the training, the model is ready to monitor the cleaning process and assist conservators in real-time. The trained model then takes the preprocessed acoustic signal of a pulse, which is the amplitude over 200 time steps and its time shift, and predicts in real-time whether the next pulse is a cleaning pulse, informing the conservator. If the model

predicts that the next pulse is not the cleaning one, the conservator can re-irradiate on the same spot. The new acoustic signal is recorded, pre-processed, and classified, creating a sequential real-time monitoring process.

## Results

### Cleaning pulse prediction accuracy

After the training process, the algorithm was evaluated on the 120 test set spots. The results, including the mean accuracy (the average accuracy of all the trained RF models on the 100 validation sets), the number of errors on both the validation and the test sets, together with the corresponding standard deviation across the 100 folds, are presented in Table 1.

The mean accuracy and its standard deviation for all RF models on the test set are  $(97.5 \pm 0.8)\%$ , indicating that they perform well on unseen data, regardless of the pulse fluence and the thickness of the unwanted paint. Additionally, the mean accuracy and its standard deviation of  $(98.8 \pm 1.3)\%$  on the validation set further confirm the robustness of the model's performance. All errors correspond to predictions where the model identified a pulse after the actual cleaning pulse as the cleaning one, indicating that the model tends to pose a minimal risk of over-cleaning.

It is important to note that if the algorithm predicts the pulse just before or just after the designated cleaning pulse, it will still be considered a successful prediction. This approach is taken because the goal of this tool is not to replace heritage scientists but to assist them in making better, more informed decisions.

In our case, we applied a self-limiting criterion, which eliminates the risk of damaging the marble substrate. However, we successfully classified a single pulse from the group that met our predetermined threshold. Given this achievement, we are optimistic that in a laser cleaning scenario where the self-limiting criteria do not apply, we will still be able to accurately predict the onset of cleaning while safeguarding the substrate.

The level of precision required for a clean mock-up, where there is no unwanted paint layer to remove, has not been investigated in terms of data collection from this type of irradiation. It was deemed unnecessary to explore this aspect, as the tool will not be utilized when the surface is already clean. Human supervision will always play a key role in this process.

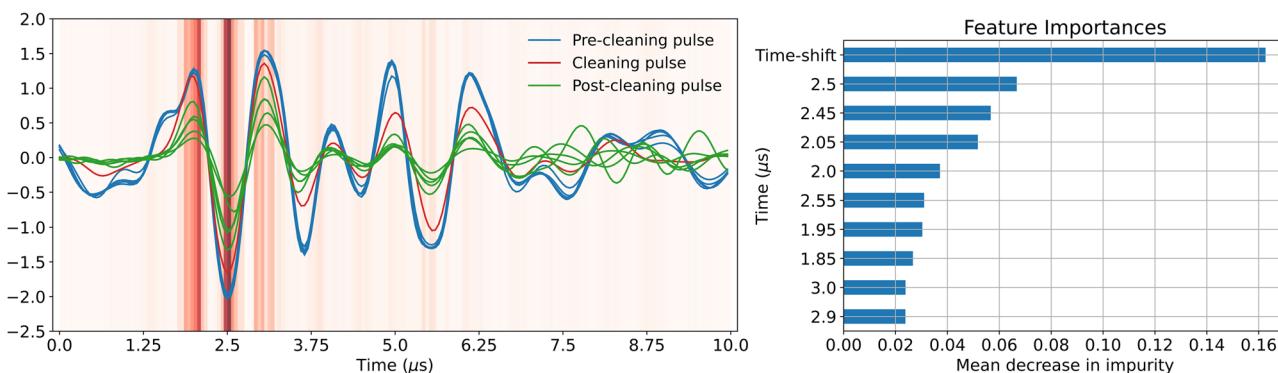
### Feature importance

One of the most significant advantages of the RF algorithm is that it provides insights into which features are most important in the decision-making process.

In Fig. 6, we present the feature importance, based on the mean decrease in impurity, in making decisions as calculated by the Random Forest algorithm. The algorithm demonstrates that the time shift of each pulse with respect to the first pulse at that spot plays the most significant role in identifying the cleaning pulse. In addition, the amplitude of the acoustic signal around the first maximum, the global minimum, and the second maximum is essential for predicting the cleaning pulse. This observation can be understood by considering that the volume fluctuations during the acoustic effect are primarily responsible for the maxima and minima highlighted in Fig. 6. The remainder of the waveform is generally regarded as reflections originating from within (the bulk) or outside the material, arriving later at the detector.

### Modeling the cleaning process

In this section, we present a simple analysis using two Logistic Regression models and the first three most important features, as identified by the Random Forest model. The first model uses only the time shift as input. In contrast, the second model considers the time shift and the amplitude of the signal at its global minimum and its first maximum before the global minimum. The model with only the time shift achieved an accuracy of  $(60.0 \pm 0.1)\%$  (or  $48 \pm 2$  errors in 120 spots, with all of them identifying a pulse after the actual cleaning pulse as the cleaning one), showing that although this feature is the most important, it alone is not enough to distinguish between a cleaning and a non-cleaning pulse. This is because the time shift strongly depends on the thickness of the unwanted material,



**Fig. 6 | The feature importance of the Random Forest algorithm.** Left: The processed acoustic signals of the successive pulses on a spot. In cyan, the signal of the pulses before the cleaning pulse (pre-cleaning pulse), in red, the cleaning pulse, and in green, the signal of the pulses after the cleaning pulse (post-cleaning pulse). The

intensity of the background color demonstrates the importance of that time step. Right: The time shift is the most important feature, based on the mean decrease in impurity, followed by the amplitude of the acoustic signal around its first maximum, its global minimum, and its second maximum.

which is a critical factor in cleaning, but not the only one that determines the process.

On the other hand, the second model, which considers the amplitude of the acoustic signal at two critical moments (as shown in the primary analysis), achieves an accuracy of  $(92.5 \pm 0.4)\%$  (or  $9 \pm 1$  errors in 120 spots, with all of them identifying a pulse after the actual cleaning pulse as the cleaning one).

These results demonstrate that combining the information captured by the acoustic signal (the physical phenomena on the cleaning surface) with an estimation of the unwanted material thickness, from the time shift, can effectively be used as the building blocks for a model to monitor the cleaning process.

## Discussion

This study aims to enable real-time monitoring of laser cleaning interventions by analyzing laser-induced acoustic signals using machine learning algorithms. The results demonstrate that by processing the generated acoustic waves with random forest algorithms, it is possible to identify the pulse responsible for removing a specific amount of material set at 75% of the irradiated surface. This critical pulse serves as a threshold for terminating the process in a timely manner, preventing any damage to the underlying surface of historical or artistic significance. While there have been studies on online monitoring for laser cleaning in industrial applications, to our knowledge, this research is the first to monitor laser cleaning interventions in the heritage field, where specific rules and limitations apply, using machine learning algorithms. Our aim is to determine the optimal moment to stop the irradiation and move the laser beam to an adjacent area.

The RF model demonstrated robustness and accuracy in predicting the cleaning pulse. Furthermore, it provides insights into how the cleaning process is captured by the acoustic signal. The time shift emerged as the most significant feature in monitoring the cleaning process. However, this feature cannot be extracted from images. This finding underscores an additional advantage of using acoustic signals for monitoring the cleaning process, as they provide a highly informative feature for this purpose. Future work will focus on using machine learning (ML) algorithms to identify the feature importance that characterizes the acoustic monitoring of laser cleaning. This will allow us to refine this application for multi-layered heterogeneous contaminants and uneven surfaces, such as high-relief sculptures. We will also address cleaning challenges related to various types of heritage materials, including varnished paintings, corroded metals, and biodeteriorated stonework, among others. In this work, we have demonstrated a proof of concept showing that the combination of acoustic signals and machine learning allows for real-time monitoring of the cleaning process in heritage conservation. In real-world scenarios, conservators can choose to use a pre-trained model as is, applying it under the same cleaning conditions in which

it was trained, or they can fine-tune it for their specific cleaning situations, such as when using new materials, by collecting data from a few irradiated spots. Once the model is fine-tuned it can process acoustic signals in real-time and assist the conservator in deciding whether to continue or to stop the cleaning process.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 30 April 2025; Accepted: 31 October 2025;

Published online: 04 December 2025

## References

1. Cooper, M. *Laser Cleaning in Conservation: An Introduction* (Oxford: Butterworth- Heinemann, 1998)
2. Pouli, P. Laser cleaning on stonework: principles, case studies, and future prospects. in: *Conserving Stone Heritage* (ed. Gherardi, F., Maravelaki, P. N.) 75–100 (Cultural Heritage Science. Springer, 2022).
3. Pouli, P. et al. The two-wavelength laser cleaning methodology: theoretical background and examples from its application on CH objects and monuments with emphasis to the Athens Acropolis sculptures. *Herit. Sci.* **4**, 9 (2016).
4. Siano, S. & Salimbeni, R. Advances in laser cleaning of artwork and objects of historical interest: the optimized pulse duration approach. *Acc. Chem. Res.* **43**, 739–750 (2010).
5. Bertasa, M. & Korenberg, K. Successes and challenges in laser cleaning metal artefacts: a review. *J. Cult. Herit.* **53**, 100–117 (2022).
6. Kono, M., Baldwin, K. G. H., Wain, A. & Rode, A. V. Treating the untreatable in art and heritage materials: ultrafast laser cleaning of “cloth-of-gold”. *Langmuir* **31**, 1596–1604 (2015).
7. Pouli, P., Selimis, A., Georgiou, S. & Fotakis, C. Recent studies of laser science in paintings conservation and research. *Acc. Chem. Res.* **43**, 771–781 (2010).
8. Cucci, C., Delaney, J. K. & Picollo, M. Reflectance hyperspectral imaging for investigation of works of art: old master paintings and illuminated manuscripts. *Acc. Chem. Res.* **49**, 2070–2079 (2016).
9. Fontana, R. et al. Application of non-invasive optical monitoring methodologies to follow and record painting cleaning processes. *Appl. Phys. A* **121**, 957–966 (2015).
10. Iwanicka, M., Sylwestrzak, M. & Targowski, P. Optical coherence Tomography (OCT) for examination of artworks in *Advanced Characterization Techniques, Diagnostic Tools and Evaluation Methods in Heritage Science* (ed. Bastidas, D., Cano, E.) 49–59 (Springer, Cham. 2018)

11. Crina Anca Sandu, I., de Sá, M. H. & Pereira, M. C. Ancient 'gilded' art objects from European cultural heritage: a review on different scales of characterization. *Surf. Interface Anal.* **43**, 1134–1151 (2011).
12. Moretti, P. et al. Non-invasive reflection FT-IR spectroscopy for on-site detection of cleaning system residues on polychrome surfaces. *Microchem. J.* **157**, 105033 (2020).
13. Iwanicka, M. et al. Complementary use of optical coherence tomography (OCT) and reflection FTIR spectroscopy for in situ non-invasive monitoring of varnish removal from easel paintings. *Microchem. J.* **138**, 7–18 (2018).
14. Yang, G. et al. An adaptive laser processing method of multilayer heterogeneous materials by monitoring dynamic spectrum at kHz-MHz level. *Opt. Lasers Eng.* **170**, 107792 (2023).
15. Sun, T. et al. In-situ monitoring of hole evolution process in ultrafast laser drilling using optical coherence tomography. *J. Manuf. Process.* **133**, 1290–1299 (2025).
16. Li, Z., Wang, S., Zheng, W., Wang, Y. & Pan, Y. A review of dynamic monitoring technology and application research of laser cleaning interface. *Measurement* **238**, 115311 (2024).
17. Melessanaki, K., Stringari, C., Fotakis, C. & Anglos, D. Laser cleaning and spectroscopy: a synergistic approach in the conservation of a modern painting. *Laser Chem.* **2006**, 42709 (2006).
18. Striova, J. et al. Optical devices provide unprecedented insights into the laser cleaning of calcium oxalate layers. *Microchem. J.* **124**, 331–337 (2015).
19. Moretti, P. et al. Laser cleaning of paintings: in situ optimization of operative parameters through non-invasive assessment by optical coherence tomography (OCT), reflection FT-IR spectroscopy and laser induced fluorescence spectroscopy (LIF). *Herit. Sci.* **7**, 44 (2019).
20. Kokkinaki, O. et al. Laser-induced fluorescence as a non-invasive tool to monitor laser-assisted thinning of aged varnish layers on paintings: fundamental issues and critical thresholds. *Eur. Phys. J.* **136**, 938 (2021).
21. Tserevelakis, G. J. et al. On-line photoacoustic monitoring of laser cleaning on stone: evaluation of cleaning effectiveness and detection of potential damage to the substrate. *J. Cult. Herit.* **35**, 108–115 (2019).
22. Tserevelakis, G. J., Pouli, P. & Zacharakis, G. Listening to laser light interactions with objects of art: a novel photoacoustic approach for diagnosis and monitoring of laser cleaning interventions. *Herit. Sci.* **8**, 98 (2020).
23. Papanikolaou, A. et al. Development of a hybrid photoacoustic and optical monitoring system for the study of laser ablation processes upon the removal of encrustation from stonework. *Opto-Electron. Adv.* **3**, 190037 (2020).
24. Chen, Y., Deng, G. L., Zhou, Q. & Feng, G. Acoustic signal monitoring in laser paint cleaning. *Laser Phys.* **30**, 066001 (2020).
25. Dimitroulaki, E. et al. Photoacoustic real-time monitoring of UV laser ablation of aged varnish coatings on Heritage objects. *J. Cult. Herit.* **63**, 230–239 (2023).
26. Cooper, M. I., Emmony, D. C. & Larson, J. Characterization of laser cleaning of limestone. *Opt. Laser Technol.* **27**, 69–73 (1995).
27. Lee, J. M. & Watkins, K. G. In-process monitoring techniques for laser cleaning. *Opt. Lasers Eng.* **34**, 429–442 (2000).
28. Bregar, V. B. & Mozina, J. Optoacoustic analysis of the laser-cleaning process. *Appl. Surf. Sci.* **185**, 277–288 (2002).
29. Jankowska, M. & Śliwiński, G. Acoustic monitoring for the laser cleaning of sandstone. *J. Cult. Herit.* **4**, 65–71 (2003).
30. Sun, T. et al. Femtosecond laser drilling of film cooling holes: quantitative analysis and real-time monitoring. *J. Manuf. Process.* **101**, 990–998 (2023).
31. Sun, T. et al. Hybrid Machine learning and temporal-spatial fusion decision for real-time monitoring of drilling stage in ultrafast laser drilling. *Opt. Laser Technol.* **184**, 112354 (2025).
32. Fiorucci, M. et al. Machine learning for cultural heritage: a survey. *Pattern Recognit. Lett.* **133**, 102–108 (2020).
33. Argyrou, A. & Agapiou, A. A review of artificial intelligence and remote sensing for archaeological research. *Remote Sens.* **14**, 6000 (2022).
34. Zhang, C. H. et al. Application of deep learning algorithms for identifying deterioration in the ushnisha (Head Bun) of the Leshan Giant Buddha. *Herit. Sci.* **12**, 399 (2024).
35. Kwon, D. & Yu, J. Automatic damage detection of stone cultural property based on deep learning algorithm. *Int. Arch. Photogramm., Remote Sens. Spat. Inf. Sci.* **XLII-2/W15**, 639–643 (2019).
36. Alexakis, E. et al. A novel application of deep learning approach over IRT images for the automated detection of rising damp on historical masonries. *Case Stud. Constr. Mater.* **20**, e02889 (2024).
37. Twarek, A., Halicz, L., Matwin, S. & Wagner, B. Machine learning in analytical chemistry for cultural heritage: a comprehensive review. *J. Cul. Her.* **70**, 64–70 (2024).
38. Festa, G. et al. Studying ancient Egyptian copper-alloy objects via X-ray diffraction and Machine Learning. *J. Cul. Her.* **72**, 48–58 (2025).
39. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
40. Ho, T. K. Random decision forests. In *Proceedings of Third International Conference on Document Analysis and Recognition*, vol. 1 (ICDAR '95), (IEEE Computer Society, 1995)
41. Wang, G. et al. Multi-objective optimization of laser cleaning quality of Q390 steel rust layer based on response surface methodology and NSGA-II Algorithm. *Materials* **17**, 3109 (2024).
42. Fang, C. et al. Effect of laser cleaning parameters on surface filth removal of porcelain insulator. *Photonics* **10**, 269 (2023).
43. Yu, J. & Chen, Y. Prediction of simulation parameters of fiber laser cleaning Range hood based on BP neural network. *J. Phys. Conf. Ser.* **1820**, 012118 (2021).
44. Chu, D. et al. Real-time acoustic monitoring of laser paint removal based on deep learning. *Opt. Express* **33**, 1421–1436 (2025).
45. Sun, T. et al. Real-time monitoring and control of the breakthrough stage in ultrafast laser drilling based on sequential three-way decision. *IEEE Trans. Ind. Inform.* **19**, 5422–5432 (2022).
46. Sun, B. et al. Cleanliness prediction of rusty iron in laser cleaning using convolutional neural networks. *Appl. Phys. A* **126**, 179 (2020).
47. Lin, D. et al. Real-time LIBS monitoring of laser-based layered controlled paint removal from aircraft skin based on random forest. *Appl. Opt.* **62**, 2569–2576 (2023).
48. Pedregosa, F. & Varoquaux, G. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2018).

## Acknowledgements

This research was conducted without any financial assistance or external funding.

## Author contributions

G.D.B., G.J.T., and P.P. conceived the concept and designed the experiments. G.D.B. performed the ML analysis. G.J.T. developed the experimental setup for collecting the data. A.-N.R. collected the data. K.M. developed the tested mock-ups and contributed to the assessment of the cleaning tests. G.P.T. is involved in the analysis. G.D.B. and P.P. were major contributors to the manuscript writing. P.P. coordinated all phases of the study. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Georgios D. Barmparis or Paraskevi Pouli.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025