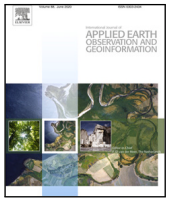
International Journal of Applied Earth Observation and Geoinformation 122 (2023) 103406

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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



Classification of structural building damage grades from multi-temporal photogrammetric point clouds using a machine learning model trained on virtual laser scanning data

Vivien Zahs [a,*], Katharina Anders [a], Julia Kohns [b], Alexander Stark [b], Bernhard Höfle [a,c,d]

[a] 3D Geospatial Data Processing (3DGeo) Research Group, Institute of Geography, Heidelberg University, Im Neuenheimer Feld 368, Heidelberg, 69120, Germany
[b] Institute of Concrete Structures and Building Materials, Karlsruhe Institute of Technology (KIT), Gotthard-Franz-Straße 3, Karlsruhe, 76131, Germany
[c] Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Im Neuenheimer Feld 205, Heidelberg, 69120, Germany
[d] Heidelberg Center for the Environment (HCE), Heidelberg University, Im Neuenheimer Feld 229, Heidelberg, 69120, Germany

ARTICLE INFO

Dataset link: https://doi.org/10.11588/data/D3WZID/

Keywords:
Change detection
UAV
3D
Damage classification
Earthquake
Natural hazards

ABSTRACT

Automatic damage assessment by analysing UAV-derived 3D point clouds provides fast information on the damage situation after an earthquake. However, the assessment of different damage grades is challenging given the variety in damage characteristics and limited transferability of methods to other geographic regions or data sources. We present a novel change-based approach to automatically assess multi-class building damage from real-world point clouds using a machine learning model trained on virtual laser scanning (VLS) data. Therein, we (1) identify object-specific point cloud-based change features, (2) extract changed building parts using k-means clustering, (3) train a random forest machine learning model with VLS data based on object-specific change features, and (4) use the classifier to assess building damage in real-world photogrammetric point clouds. We evaluate the classifier with respect to its capacity to classify three damage grades (heavy, extreme, destruction) in pre-event and post-event point clouds of an earthquake in L'Aquila (Italy). Using object-specific change features derived from bi-temporal point clouds, our approach is transferable with respect to multi-source input point clouds used for model training (VLS) and application (real-world photogrammetry). We further achieve geographic transferability by using simulated training data which characterises damage grades across different geographic regions. The model yields high multi-target classification accuracies (overall accuracy: 92.0%–95.1%). Classification performance improves only slightly when using real-world region-specific training data (< 3% higher overall accuracies). We consider our approach especially relevant for applications where timely information on the damage situation is required and sufficient real-world training data is not available.

1. Introduction

The timely assessment of building damage after an earthquake is of utmost importance for the effective planning of rescue and remediation actions. Automatic damage assessment based on the analysis of 3D point clouds can provide fast and objective information on the damage situation within few hours (Vetrivel et al., 2018). As building damage can be of different type and degree, a detailed assessment of multiple damage grades enables efficient use and distribution of resources, and supports the evaluation of structural stability of buildings and repair measures.

The assessment of different damage grades, beyond binary damage detection, is a challenging task. There is large variety in possible damage characteristics and the transferability of methods developed for a study site to other geographic regions is limited, as is the transferability to other data sources, especially for machine learning classifiers (Kerle et al., 2020). An approach for classification of building damage grades that is transferable both geographically and with respect to the source and characteristics of point clouds used for training and evaluation would strongly support damage assessment for earthquake response.

We therefore present a method for classifying damage grades with supervised machine learning using a random forest (RF) classifier which is trained on simulated point clouds. We assess how the use of generic simulated training data, instead of region-specific building and damage structures, influences the accuracy of identifying damage grades. In case of an earthquake event, training on simulated data can save valuable time as pre-trained classifiers can directly be applied

* Corresponding author.
E-mail address: zahs@uni-heidelberg.de (V. Zahs).

https://doi.org/10.1016/j.jag.2023.103406
Received 10 March 2023; Received in revised form 29 May 2023; Accepted 24 June 2023
Available online 15 July 2023

to assess damage in event-specific real-world datasets without time-consuming manual labelling and further training using event-specific data. The presented method, thus, contributes to a timely assessment of multi-class structural building damage in disaster response.

1.1. UAV-based point clouds for damage assessment

New possibilities for structural building damage assessment have emerged in recent years with the increasing availability of UAV-borne remote sensing strategies. UAV-borne acquisition strategies allow a dense 3D representation of the scene with point spacings down to a few centimetres. Derived UAV-borne laser scanning (ULS) or photogrammetric point clouds provide 3D data of urban quarters and entire cities within reasonable time frames (few hours). This allows change detection on the scale of individual building parts. Those are important for identifying damage patterns that are typical for higher damage grades (heavy damage, extreme damage, destruction), which are target classes of our method (cf. Section 3).

1.2. Approaches for multi-class damage classification in 3D point clouds

Structural building damage assessment so far has mainly been performed using point clouds acquired after an earthquake event, i.e., using mono-temporal approaches (Vetrivel et al., 2018; Khoshelham et al., 2013). These approaches lack pre-event information on the building structure. This can be compensated with assumptions on the pre-event building shape, for example, to detect missing elements in the post-event point cloud. However, it leads to misclassification where these assumptions do not hold true, and thereby limits the applicability and usability of mono-temporal approaches (Vetrivel et al., 2015).

With the increasing availability of 3D city models and country-wide point cloud acquisitions (e.g., through airborne laser scanning), the development of 3D methods for damage assessment through change detection has increased (de Gélis et al., 2021; Xu et al., 2021). Current multi-temporal approaches directly compare a pre-event dataset and a post-event dataset and thereby extract different types of change features, e.g., change in geometric and radiometric properties (Tran et al., 2018; Awrangjeb et al., 2015). Change can also be classified, e.g., based on a set of geometric or histogram-based features (Roynard et al., 2016).

For binary classification, deep learning approaches today represent the state-of-the-art (de Gélis et al., 2023; Xiu et al., 2023; Qing et al., 2022). They distinguish between damaged and non-damaged buildings, or between two damage grades with very different damage characteristics. All classification approaches are still challenged, though, by the variety of damage characteristics (millimetre width of cracks and spalling up to partial failure modes and complete collapse) and by the transfer of trained algorithms to unseen data and other geographic regions with different building and hence damage characteristics (Huang et al., 2019; Vetrivel et al., 2018).

1.3. Training data generation with virtual laser scanning (VLS)

A prerequisite for machine learning-based damage classification is the availability of sufficient amounts of labelled training data covering the full range of expected damage. Lack of it can lead to poor classification performances when transferred to an unseen dataset or a different geographic region (Munawar et al., 2021; Vetrivel et al., 2018). Training data demands of state-of-the-art machine learning approaches are difficult to meet when multiple damage grades shall be classified (Alzubaidi et al., 2021). Resulting inter-class confusion might lead to missing damaged buildings (i.e., classifying damaged buildings as undamaged). Moreover, economic constraints still limit the availability of real-world region-specific training data prior to the earthquake. Time-consuming manual labelling of point clouds acquired of the affected area and training of the model would then have to be performed after the event. If no or insufficient labelled real-world data is available, training and evaluation of machine learning classifiers can benefit from simulated data (de Gélis et al., 2021).

Virtual laser scanning (VLS) provides simulated point clouds with known properties from labelled 3D input scenes (Hildebrand et al., 2022). Even if yielding lower classification accuracies, training purely on simulated training data might provide adequate performance in time-sensitive situations, such as an earthquake event. Pre-trained classifiers can then be directly applied to assess damage in event-specific real-world datasets without further training. Adequate modelling of damage patterns in the input scenes is thereby crucial for the accurate representation of damage in the simulated point clouds. This modelling process can, for example, be supported by domain-knowledge in earthquake engineering. One big advantage of our approach is that multiple operational tools for the simulation of laser scanning point clouds exist (Winiwarter et al., 2022; Gastellu-Etchegorry et al., 2016; North et al., 2010). Methods for damage assessment using simulated laser scanning point clouds as training data therefore have to deal with different point cloud sources being used for model training and application, as point clouds of an earthquake-affected area are more commonly derived from photogrammetric surveys, due to lower costs and wider availability of the instruments.

1.4. Objective

In this research, we automatically classify multi-class building damage from multi-temporal real-world photogrammetric point clouds using a RF model trained on VLS data (Fig. 1). We develop our method considering the following aspects:

1. Damages are assessed per building by deriving change of geometric features between pre-event and post-event point clouds. We are thereby independent from modelling of pre-event building shapes, but can derive change through the comparison of multi-temporal point clouds.
2. Domain knowledge from earthquake engineering is integrated in the process of training data generation from virtual scenes. We thereby ensure that our training data covers the full spectrum of damage patterns expected in the real-world dataset.
3. By using VLS for the generation of simulated point clouds, labelled building-specific training data with realistic point cloud characteristics is automatically obtained.
4. Through the use of object-specific change features, our machine learning model is trained on VLS point clouds to classify damage in real-world photogrammetric point clouds. We thereby aim to achieve transferability with respect to the source of input point clouds used for model training and application.

Our method contributes to multi-class structural building damage assessment especially for applications where timely damage information is required, and sufficient pre- and post-event real-world training data is not available. To further increase timeliness of response, we investigate how the use of non-region-specific training data influences the classification accuracy, thereby assessing transferability between geographic regions.

2. Study site and data

In this paper, we use UAV-borne photogrammetric point clouds of the city of L'Aquila (Italy) to classify earthquake-induced building damage. L'Aquila was hit by an earthquake on Monday, 6 April 2009, with a moment magnitude Mw = 6.3 (Geo-Engineering Extreme Events Reconnaissance, 2009). Up to 15,000 buildings were damaged, and more than 24,000 people were left homeless (Earthquake Engineering Research Institute, 2009). Collapsed and damaged structures in L'Aquila included both masonry buildings and reinforced concrete structures. Damage generally involved minor cracking to relatively

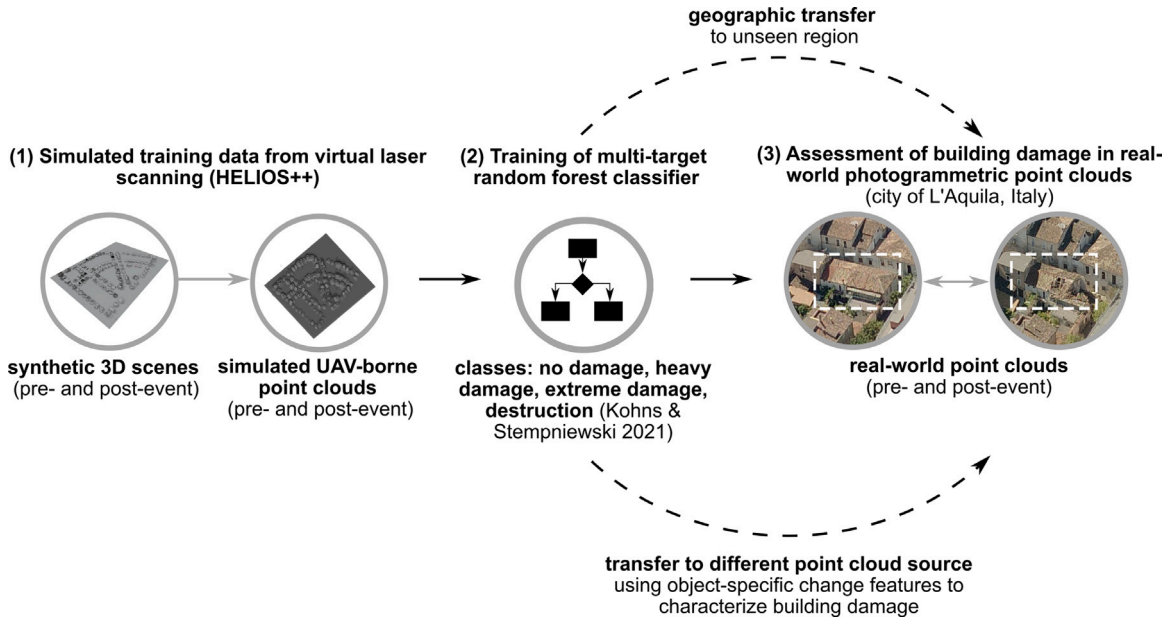


Fig. 1. Overview of the approach to classify building damage in photogrammetric point clouds using a machine learning model trained on simulated laser scanning point clouds.

severe cracking or collapse of the masonry infill walls. Old unreinforced masonry buildings made of mortar and multi-wythe rubble-stone or clay bricks were significantly damaged, ranging from wall cracking to extreme damage and collapse.

Photogrammetric point clouds of L'Aquila (Fig. 2) were generated based on nadir RGB images captured before (2008-08-30) and after (2009-04-29) the earthquake. Reconstructed point clouds contain around 75 million points with an average point spacing of 0.1 m. Resulting from the nadir perspective of the images, horizontal building elements (roofs) were sampled with a higher density (point spacing: 0.09 m; standard deviation SD: 0.03 m) compared to vertical elements (point spacing: 0.12 m; SD: 0.04 m). We improved the alignment of the entire pre- and post-event point clouds using an iterative closest point algorithm (Besl and McKay, 1992) applied to stable areas (streets and unchanged walls) and achieved a final registration error of 7.4 cm (derived as the SD of local point cloud distances in stable areas).

3. Methods

We develop a method to classify multiple grades of structural building damage in UAV-borne photogrammetric point clouds using a machine learning model trained on VLS point clouds (Fig. 3). Following domain knowledge in earthquake engineering, we consider four damage grades in our approach (Fig. 4): (1) No damage, (2) heavy damage, (3) extreme damage, (4) destruction. Classes of slight and moderate damage are not considered, as the geometric representation of their typical damage patterns (e.g., crack widths of few millimetres) in the point clouds requires a higher spatial data resolution than typically available by UAV acquisitions. Our approach consists of five main steps:

1. Simulation of pre- and post-event point clouds through UAV-borne laser scanning of virtual scenes
2. Coarse identification of changed and unchanged building points using k-means clustering
3. Extraction of robust object-specific change features
4. Training of an RF machine learning model with simulated point clouds and object-specific change features
5. RF classification of multi-class building damage in real-world pre- and post-event photogrammetric point clouds

We evaluate the performance of classifiers trained on (1) generic VLS point clouds, (2) region-specific VLS point clouds, (3) generic VLS point clouds and real-world region-specific photogrammetric point clouds, and (4) real-world region-specific photogrammetric point clouds regarding their accuracy of building damage classification in the real-world photogrammetric dataset. Evaluation is based on a reference dataset derived from the photogrammetric point clouds of the L'Aquila earthquake (cf. Section 2). A set of evaluation metrics (cf. Section 3.5) is used to assess the performance for each target damage grade separately and the overall capacity of the classifiers to separate buildings of any damage grade from undamaged buildings.

3.1. Generation of real-world training and evaluation data

The photogrammetric dataset of L'Aquila is used to generate real-world training data and to evaluate the performance of all classifiers. We introduce an 80:20 random split of the dataset into training and test data. Areas of individual buildings in the dataset are manually segmented and labelled by an expert in earthquake engineering, and by using the damage catalogue developed in Kohns et al. (2022) to identify typical damage patterns for the target classes. The training dataset consists of 112 labelled building models per damage class. The evaluation dataset consists of 125 labelled buildings (35 no damage, 19 heavy damage, 32 extreme damage, 35 destruction). The uneven distribution of damage grades in the evaluation dataset results from the uneven number of buildings per damage grade that could be confidently assessed in the expert-based labelling.

3.2. Generation of simulated training data

The generation of simulated training data consists of (1) the preparation of virtual scenes of 3D building models with various damage patterns labelled with the damage grade, and (2) the simulation of point clouds through VLS of these scenes.

3.2.1. Preparation of virtual scenes

To investigate the importance of region-specific training data, two types of virtual scenes in pre-event and post-event state are generated:

1. Region-specific scene: This scene mimics the characteristics of the real-world scene of this study (L'Aquila) with respect to building types and construction materials typical for this region (cf. Section 2). It also exhibits main characteristics of the urban

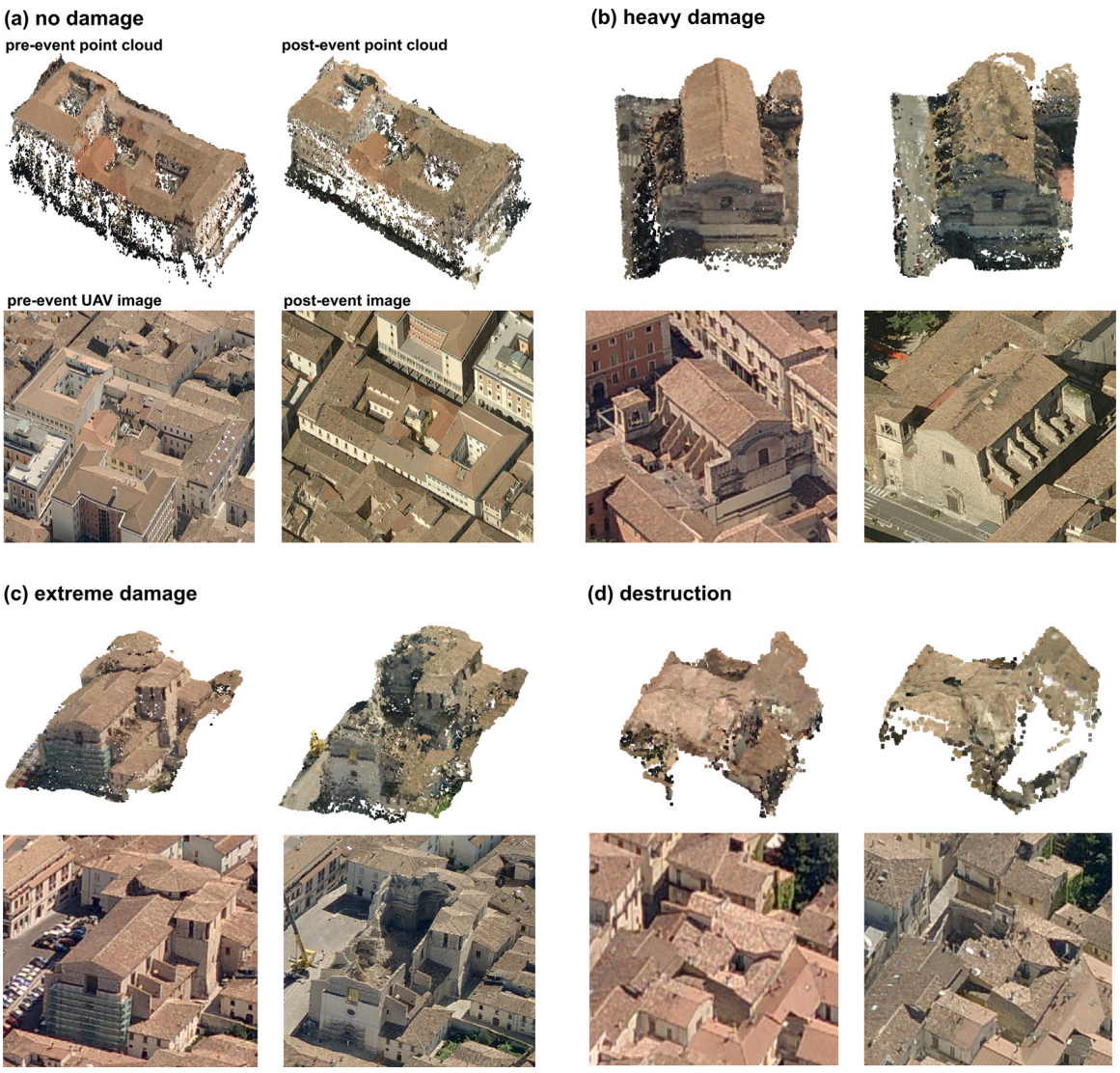


Fig. 2. (a)–(d): Pre-event (2008-08-30) and post-event (2009-04-29) 3D building point clouds (upper rows) and corresponding RGB images (lower rows) obtained by UAV-borne acquisitions of L'Aquila (cf. Section 2). Images and point clouds represent the four target classes (a) no damage, (b) heavy damage, (c) extreme damage, and (d) destruction.

structure when assembling the individual 3D building models in the scene. Damage patterns implemented in the post-event point cloud are typical damage patterns for this geographic location.

2. Generic scene: This scene contains a broader range of building types (single family houses up to large apartment buildings), construction material (masonry and reinforced concrete), and built structure (both loose and narrow development), all of which typically occur in small to medium-sized European cities. Consequently, the post-event state of this scene contains a greater variety of damage patterns.

Both types of scenes consist of 112 undamaged individual buildings, respectively. The post-event scenes with 112 buildings for each damage grade are generated based on their corresponding pre-event scenes. Therein, we introduce damage representative of the respective damage grade to each building in duplicates of each pre-event scene. This provides the data basis for a direct comparison of pre-event and post-event scenes to extract change features and classify building damage in a later step. 3D building models used to assemble the virtual scenes are taken from open source online repositories (TurboSquid, Inc., 2023). Further buildings are generated manually in the open source 3D creation suite software Blender (Blender Online Community, 2018, version 2.93.0). The number of the originally 28 different building models is augmented by applying modifications to building size or parts of the buildings. This results in a total number of 448 buildings composed of 112 buildings per damage grade. Damage is modelled into the buildings based on the damage catalogue presented in Kohns et al. (2022) and Kohns and Stempniewski (2021). The damage catalogue is a descriptive framework for the classification of five damage grades, following the European Macroseismic Scale (EMS-98): Slight, moderate, heavy, extreme, and destruction. It defines distinct geometric properties typical for our target damage grades, which are: Separation of horizontal structural components, local spallings, partial collapse of structural and non-structural components, pressure failure in corner areas, up to 15 mm crack widths (heavy damage), collapse or misalignment of one or more stories, pancake collapse of single stories, tilt or tip over of the building (extreme damage), and collapse of whole building, pancake collapse of all stories, debris pile, tip over of whole building (destruction).

The damage catalogue has been specifically designed for UAV-based damage assessment and therefore focuses on damage patterns that are recognisable from the outside. For each target damage grade, we manually introduce the typical damage patterns into 3D building models of the virtual post-event scenes. Following the damage catalogue, we focus on the two predominant construction materials in Europe, i.e., masonry and reinforced concrete.

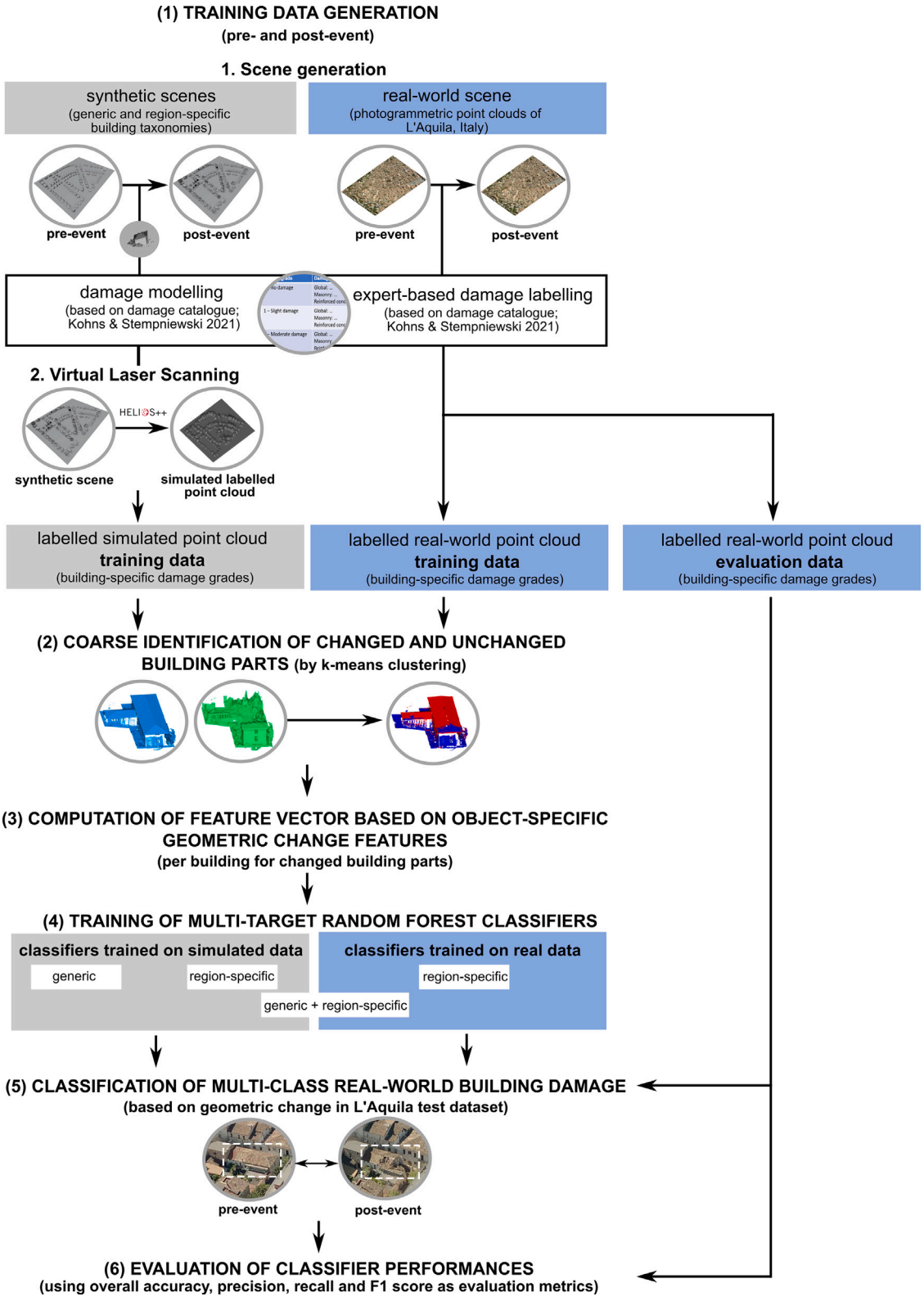


Fig. 3. Full workflow of the developed approach to classify building damage in pre- and post-event photogrammetric point clouds using point cloud-based change features and a random forest classifier trained on simulated point clouds from virtual laser scanning.

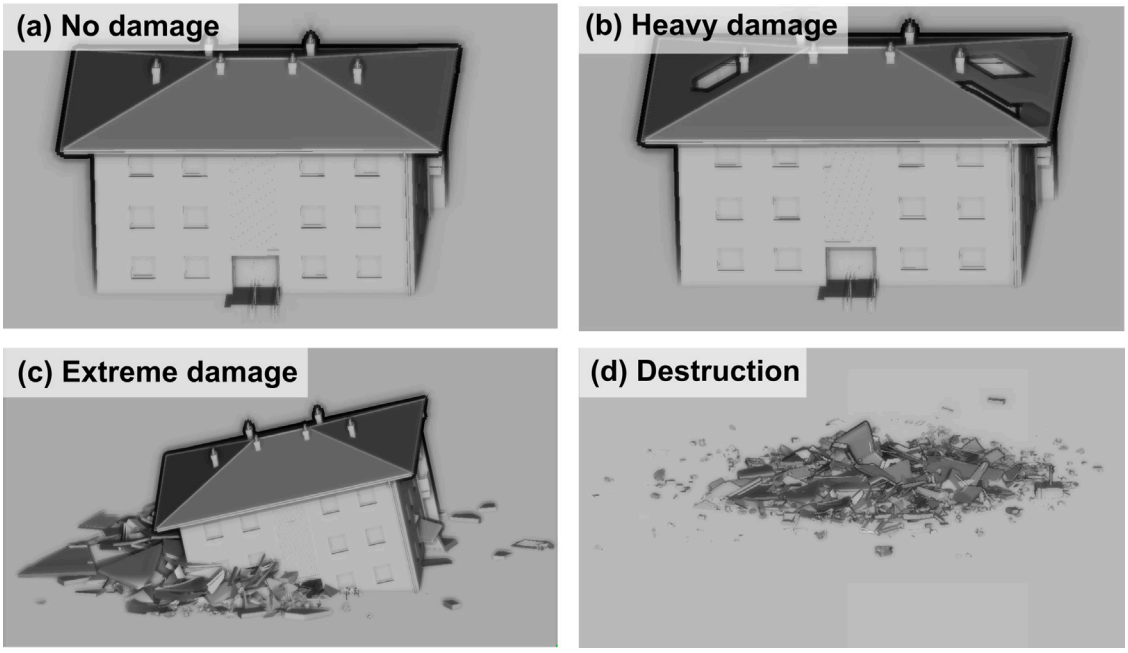


Fig. 4. Example 3D building models representing the four target damage grades (a) no damage, (b) heavy damage, (c) extreme damage, and (d) destruction considered in our classification.

Table 1
Acquisition parameters used for the laser scanning simulation in HELIOS++ using a RIEGL VUX-1UAV scanner model.

Scan rate [lines/s]	Pulse repetition rate [kHz]	Strip overlap [%]	Field of view [deg.]	Flight altitude [m AGL]	Flight speed [m/s]
89	300	60	120	100	8

3.2.2. Simulated point cloud generation using virtual laser scanning (VLS)

We perform VLS of the generated scenes using the open source software HELIOS++ (Winiwarter et al., 2022). HELIOS++ is a general-purpose ray tracing-based simulation framework with support for multiple platforms, sensors, and scene types that can be flexibly combined in a modular manner.

Acquisition parameters (Table 1) using a RIEGL VUX-1UAV scanner model for our simulations are selected to achieve similar horizontal point densities as the real-world dataset (cf. Section 3.1) in all simulated point clouds. The influence of different acquisition parameters between pre- and post-event acquisitions on the geometric representation of a building is not in the focus of our study but, for example, assessed by de Gélis et al. (2021).

As input for each simulation, we specify one virtual scene. The damage label and unique ID annotated to each building are stored with the output VLS point cloud. As output of the simulation, we obtain pre- and post-event ULS point clouds with per-building damage grade as class label in the post-event point clouds.

3.3. Object-specific change feature selection

We assess structural building damage through change analysis of geometric point cloud features between pre- and post-event point clouds. For this, we derive change features, which we define as the change of a feature between two epochs (i.e., points in time). A change feature is obtained by computing the difference between the feature value of a point in the pre-event epoch to the feature value of its nearest point in the post-event epoch. We investigate the change of the following hand-crafted features which are commonly used in classical machine learning approaches of building damage assessment (de Gélis et al., 2021; Tran et al., 2018): Anisotropy, curvature, eigenentropy, eigenvalues, linearity, omnivariance, planarity, point density, sphericity, sum of eigenvalues, surface density, surface variation, verticality, volume density (Weinmann et al., 2015), normal vector, roughness (Dorninger and Nothegger, 2007), echo ratio (Höfle et al., 2009), no. neighbours, z range (highest minus lowest z value), and z rank (relative vertical position of the feature point within its neighbourhood) (Otepka et al., 2013).

To evaluate the use of different input point cloud sources for training (VLS) and classification (photogrammetric), we investigate this transfer of change features in an experimental study of real-world ULS and photogrammetric point clouds. These datasets were acquired from a damaged building at three epochs during a demolition process (Fig. 5). Using these point clouds, we identify object-specific change features that are robust to variable properties of different point cloud sources. For this, geometric point cloud features are computed per point within local neighbourhood radii ranging from 1.0 m to 4.0 m with a step size of 0.5 m, according to the point density of the dataset. Based on this, we select one neighbourhood radius for the computation of each feature where the mean value of the derived change feature is most similar between point cloud sources. Finally, we consider a change feature to be robust if its relative difference between the values of ULS and photogrammetric point clouds is low.

Before the classification of damage grades, we coarsely filter out non-damaged building points using a k-means clustering on each building point cloud. This is done because even for heavy or extreme damage, larger parts of the building point cloud can still be unchanged. Hence, geometric change following typical damage patterns occurs only in local parts of a building. Descriptive statistical values per building are then not suitable to represent the actual degree of damage (Fig. 6). Changed and unchanged building points are separated in feature space using the two pointwise features change in curvature (to identify holes and large cracks) and change in height (to identify collapsed roofs and stories), separating the building point cloud into two clusters, i.e. one for unchanged and one for changed points of a building. To reflect the share of damaged area of a building after filtering, we include the

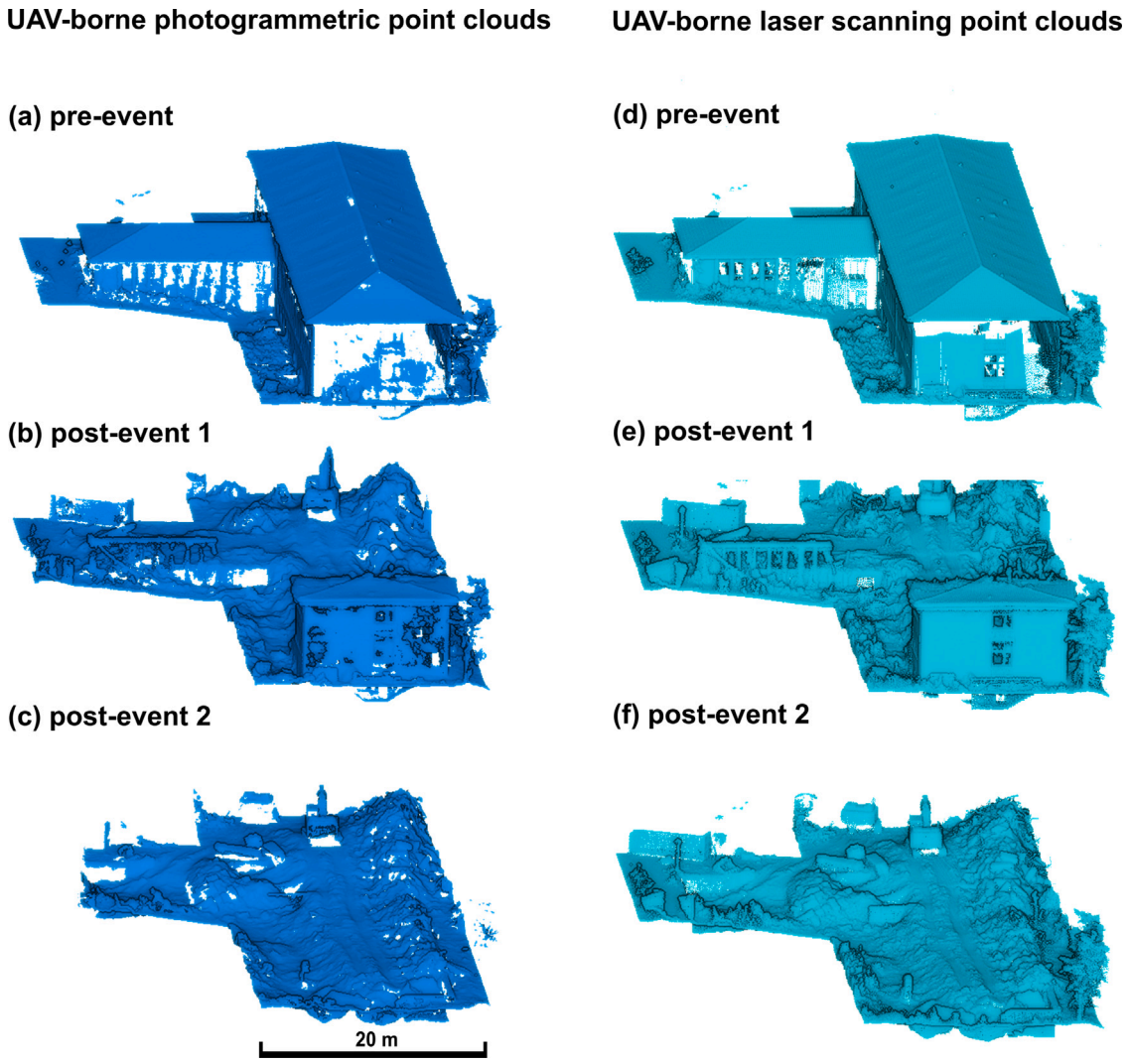


Fig. 5. (a)–(c) Real-world UAV-borne photogrammetric point clouds and (d)–(f) real-world UAV-borne laser scanning point clouds acquired of a building before demolition (a, d) and at two different demolition stages (b, c, e, f).

percentage of clustered changed points to all points of a building as additional feature in the classification.

3.4. Classification of building damage grades

To classify structural building damage, we perform supervised classification using the object-specific geometric change features for the changed points per building. As our method targets damage classification on building level, change features are computed based on all changed building points, although they might represent different locations of damage (cf. Fig. 6). We use RF decision trees for classification, which are mostly uncorrelated due to high variations of the trees. Moreover, as no automatic workflows exist for modelling of the relevant damage patterns into 3D building models, our approach benefits from the low demand of training data by RF models (Breiman, 2001).

To investigate the influence of using region-specific and real-world training data for damage classification, we train multiple RF classifiers with different input training data: (1) Simulated generic data (VLS generic), (2) simulated region-specific data (VLS region specific), (3) simulated generic data and real-world region-specific photogrammetric data (VLS generic + real-world photogrammetric), and (4) real-world region-specific photogrammetric data (real-world photogrammetric).

Each classifier is trained and tested using the 448 labelled damaged and undamaged buildings of the respective dataset with an equal number of 112 buildings per damage grade. In the selection of the buildings used for training, we ensure that the full range of damage patterns is included in the training dataset. For each classifier, building objects of the entire training dataset are randomly split into 70% training data and 30% testing data to evaluate the accuracy of the trained classifier. The RF classifier is trained with a set of 100 trees and a maximum depth of 5. All four classifiers are finally applied to the real-world photogrammetric dataset.

3.5. Evaluation of classifier performances

The performance of the classifiers with respect to their accuracy of classifying structural building damage in a real-world photogrammetric point cloud is assessed using the labelled evaluation dataset. We evaluate the performance for each target damage grade separately as a binary case, as the correct discrimination of multiple damage grades is of great relevance to our study. Further, we evaluate the overall capacity of the classifiers to correctly separate buildings of any degree of damage from undamaged buildings. We use overall accuracy, precision, recall, and F1 score as classification metrics.

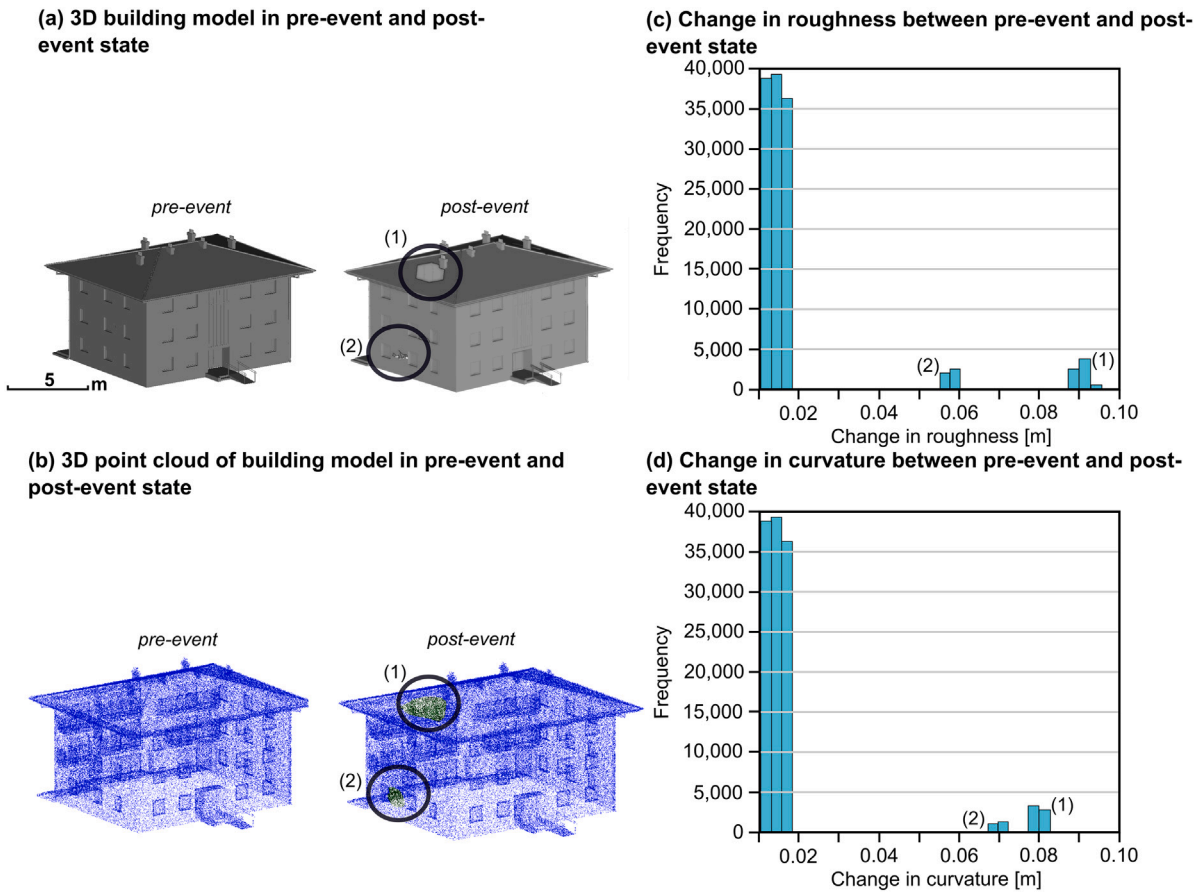


Fig. 6. (a)–(b) 3D building model and derived point cloud of a building in a pre-event and post-event state. (c)–(d) Histogram of change in roughness and curvature between pre-event and post-event state. Small building parts exhibiting change (1 and 2) result in a multi-modal distribution of change values, which cannot be adequately described by descriptive statistical values derived for the whole building.

4. Results

4.1. Generation of simulated training data

The two types of virtual scenes (generic and region-specific) are shown in Fig. 7 in pre-event and post-event states along with exemplary damage patterns of various 3D building models with different damage grades. Corresponding point clouds are shown in Fig. 8. The figure also compares real-world and simulated point clouds of buildings with similar damage patterns to compare their geometric representation of change. It is clearly visible that the spatial sampling of the buildings differs due to the different acquisition strategies. The geometric representation of damage patterns is, however, not considerably affected by these differences. For higher damage grades, the change in geometry occurs on a larger spatial scale than the differences in sampling due to the overall dense sampling of a building. We can therefore expect that geometric change between two epochs is in the same order of magnitude in VLS and photogrammetric point clouds.

4.2. Object-specific change feature selection

Change features with low difference between real-world ULS and photogrammetric point clouds are shown in Fig. 10. These features are considered to be robust object-specific features that are suitable for building damage classification both in ULS and photogrammetric point clouds. We consider only those geometric features as input for damage classification where the difference of the feature between ULS and photogrammetric point clouds is maximum 10%, as this yields a good compromise between the similarity of geometric change between ULS and photogrammetric point clouds and the number of change features available for damage assessment (Fig. 9). Accordingly, the final selected change features used for building damage classification are: Change in planarity, surface variation, point density, number of neighbours, surface density, volume density, roughness, z rank, z range, and in normal vector.

We can easily separate changed and unchanged building points in feature space using the two pointwise features (1) change in curvature (to identify holes and large cracks) and (2) change in height (to identify collapsed roofs and stories) to separate the building point cloud into two clusters (changed/unchanged). We only use points of the changed cluster for the classification of damage grades (Fig. 11).

4.3. Evaluation of classifier performances

The results of all metrics used to quantitatively evaluate the trained models are given in Table 2 and Fig. 12. When applied to the real-world evaluation dataset, all classifiers yield high classification accuracies for the target damage grades (overall accuracies: 92.0%–96.8%, F1 scores: 73.2%–94.6%).

Inter-class confusions for all classifiers mainly occur between neighbouring damage grades. Only few buildings with no damage are misclassified as heavy damage or vice versa by all classifiers, which are significantly different degrees of damage. We hence assume this relates to the damage catalogue being a suitable descriptive framework for the generation of realistic training data but not exclusively relating to geometric change in the point cloud. Consequently, the model might learn certain geometric representations of damage grades differently from how expert analysts would categorise them. For one of these

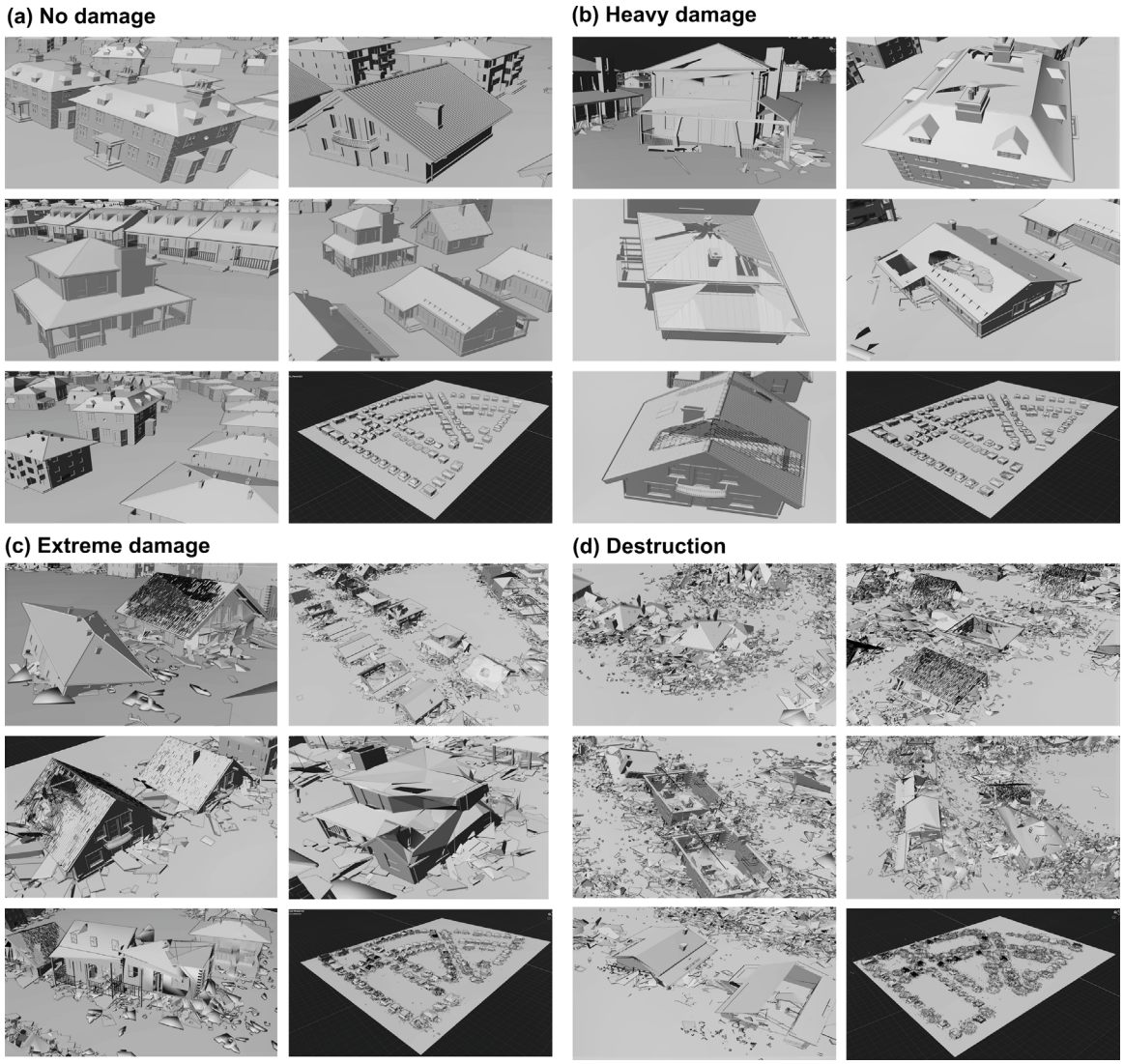


Fig. 7. Examples of 3D building models in (a) pre-event state and (b)–(d) post-event states of the target damage grades (a) no damage, (b) heavy damage, (c) extreme damage, and (d) destruction.

misclassified buildings, a visual inspection of the point clouds reveals occlusion effects in the area of damage, which results from the narrower built structure in the region-specific simulated scene. For the other buildings, close inspection does not show occlusion effects. However, the class probabilities of no damage and heavy damage in these cases are similarly high, which suggests that damage patterns of these building objects are not significantly different from no damage.

Using generic simulated training data yields good classification results for all target classes with overall classification accuracies between 92.0% and 95.1%. Using region-specific simulated training data instead of generic simulated training data does not strongly reduce inter-class confusion or increase the completeness of detected damaged buildings (+3%), neither does the use of real-world region-specific training data (+6% higher completeness of damaged buildings). This indicates that our model trained purely on generic simulated data has a high transferability to unseen regions and that the benefit of adding site-specific real-world training data is low for the classification task at hand. We attribute this to the fact that for the damage grades considered in our study, damage patterns do not vary considerably for different building types and built structure. Change features learned from the generic simulated training dataset hence generalise appropriately to be used for damage assessment in datasets with different site characteristics.

For binary classification of damaged and undamaged buildings, the VLS classifier achieves an accuracy of 84.6%. Using real-world region-specific training data yields only slightly higher accuracies (89.7%). This is an important result, as applications can rely to identify damaged buildings with high completeness.

These results support our hypothesis that the transfer of a supervised machine learning model trained purely on simulated training data to an unseen real-world dataset with specific site characteristics is a valid approach for timely damage assessment in earthquake response.

5. Discussion

5.1. Transferability of the method with respect to the source of input point clouds

Our approach achieves transferability of input point clouds from different sources for training and application of the RF classifier. We are able to transfer the classifier trained on simulated ULS point clouds to classify damage in real-world photogrammetric point clouds. We achieve this through the identification of object-specific geometric change features, which show to be representative for the investigated

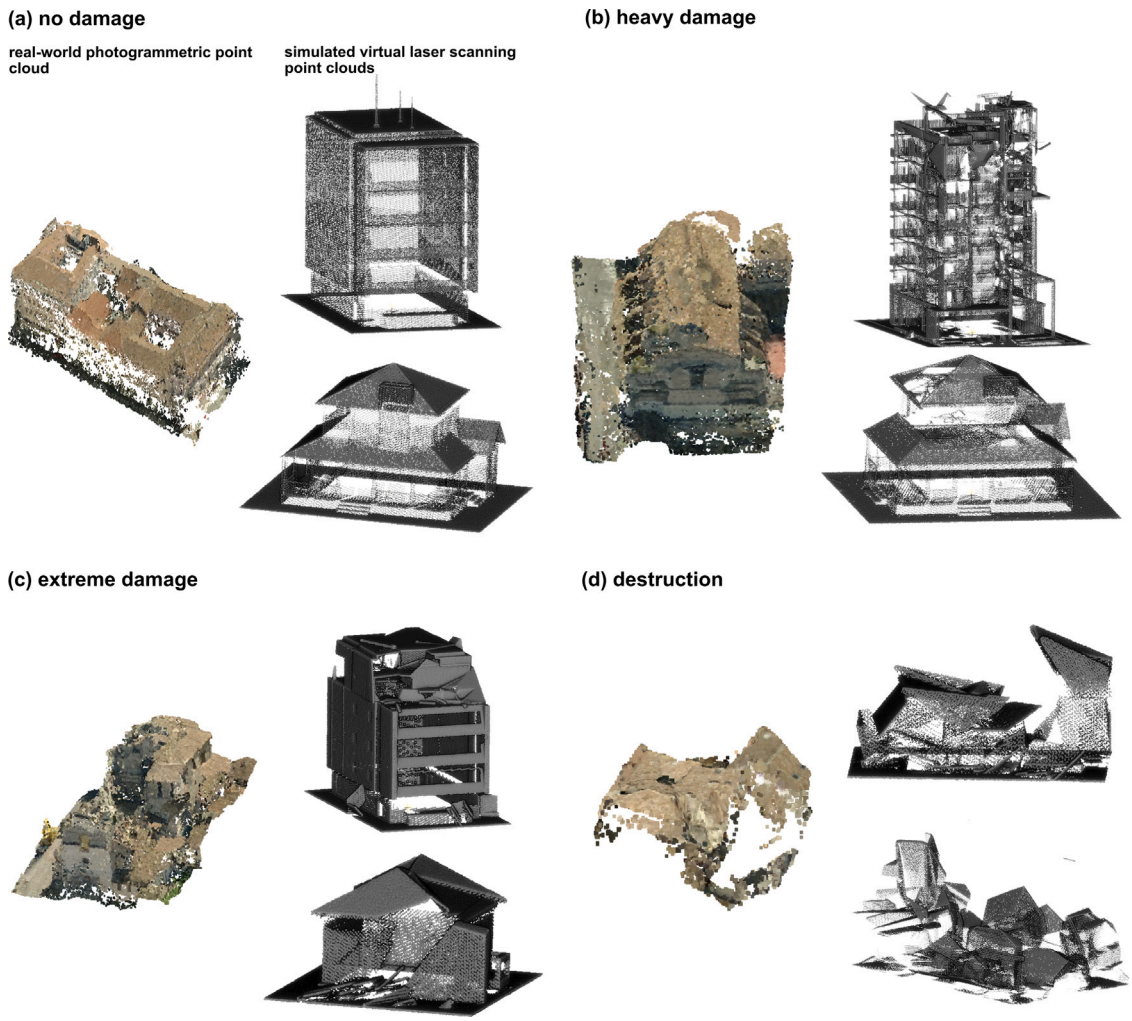


Fig. 8. Real-world photogrammetric and simulated virtual laser scanning point clouds for the target damage grades (a) no damage, (b) heavy damage, (c) extreme damage, and (d) destruction.

damage grades. Handling multi-source point clouds is especially relevant when training models with simulated data, due to the lack of available tools to simulate photogrammetric point clouds. In our study, we identify object-specific change features that are robust to variable properties of input point clouds from different sources (UAV-borne laser scanning and photogrammetry) and hence different point cloud characteristics used for training and application of the model. For other real-world data properties, the VLS framework used in our approach allows simulating pre-event and post-event point clouds with different acquisition parameters and measurement noise. By this, the model could be trained to deal with different representations of damage between input point clouds and even between the two epochs.

5.2. Transferability of the method to different site characteristics

An important strength of our approach is the transferability of a trained classifier to data from an unseen region. We find that the RF classifier trained on generic simulated building point clouds achieves high classification accuracies in the real-world dataset, which are comparable to those achieved in studies training on region-specific inputs (e.g., de Gélis et al., 2021). Using region-specific simulated building point clouds does not strongly improve the classification results in our study (increase of overall accuracy < 2%), neither does training purely based on real-world region-specific data (increase of overall accuracy < 2%). Although we did not see a notable improvement in classification accuracy when using real-world region-specific training data, in some cases this might be beneficial to add. For individual buildings with very site-specific damage patterns, the model trained on generic simulated training data might fail to assign the correct damage grade. Re-training the model with additional region-specific real-world training data can then improve the classification accuracy without big efforts.

Model transferability is still a major challenge for the assessment of binary or multi-class building damage. For example, the transferability of a CNN model using 3D point cloud features is limited already when scene characteristics vary slightly (Vetrivel et al., 2018). We achieve model transferability through the integration of domain knowledge in the process of simulated training data generation. Using the concept of a damage catalogue, we are able to model damage patterns which are characteristic across different geographic regions into the 3D building models. Therefore, our classifier is trained on geometric change features which generalise adequately to discriminate target damage grades in point clouds of other geographic regions. Comparable classification accuracies and generalisation properties were reported by de Gélis et al. (2021) for deep neural networks for building change detection on a 2D patch level using simulated airborne laser scanning training data. Moreover, they find these approaches to be less sensitive to measurement noise compared to RF approaches working directly in the point clouds. These results motivate future studies to investigate the performance of deep learning architectures in full 3D and for multi-class classification tasks, especially if the availability of training data increases. The application of our approach is currently

Table 2
Accuracy measures for the trained random forest classifiers (VLS generic, VLS region-specific, VLS generic + real-world photogrammetric, and real-world photogrammetric) for multi-class damage classification of 125 buildings in the real-world photogrammetric dataset of L'Aquila, Italy. The performance is evaluated for each damage grade separately and for all damage grades as a binary case of change detection. Therein, overall accuracies are calculated as the share of correctly classified buildings of the respective damage grade, and for all damage grades to separate damaged from undamaged buildings.

	All damage grades	No damage	Heavy damage	Extreme damage	Destruction
VLS generic					
Overall accuracy	**95.12**	**95.12**	**92.80**	**92.00**	**93.60**
Precision	1.0	84.62	72.73	78.13	91.43
Recall	84.62	1.0	84.21	89.29	86.49
F1 score	**91.67**	**91.67**	**78.05**	**83.33**	**88.89**
VLS region-specific					
Overall accuracy	94.40	92.00	92.00	92.00	92.00
Precision	82.05	84.21	69.57	78.13	82.05
Recall	1.0	69.57	84.21	89.29	91.43
F1 score	90.14	76.19	76.91	83.33	86.49
VLS generic + real-world photogrammetric					
Overall accuracy	94.40	94.40	91.20	92.00	92.00
Precision	82.05	1.0	68.18	81.25	83.78
Recall	1.0	82.05	78.95	86.67	88.57
F1 score	90.14	90.14	73.17	83.87	86.11
Real-world photogrammetric					
Overall accuracy	**96.80**	**96.80**	**93.60**	**92.00**	**93.60**
Precision	1.0	89.74	78.95	81.25	84.62
Recall	89.74	1.0	78.95	86.67	94.29
F1 score	**94.59**	**94.59**	**78.95**	**83.87**	**89.19**

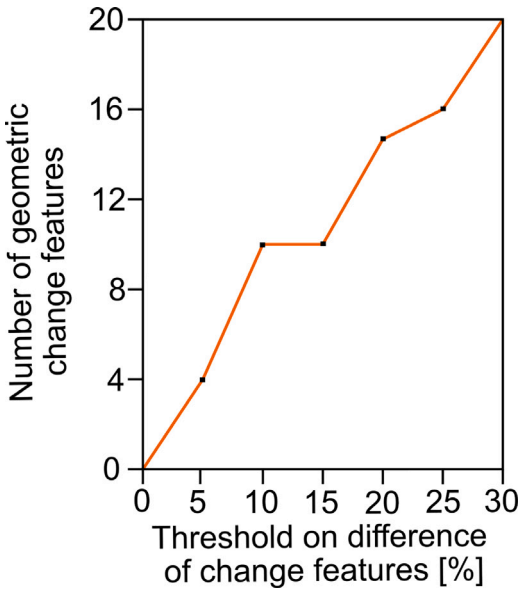


Fig. 9. Relation between the threshold on maximum allowed difference of change features between ULS and photogrammetric point clouds and the resulting number of geometric change features. The relative difference (in %) is derived between the mean values of a per-point change feature. Lower thresholds consequently lead to a smaller number of features identified as robust. Here, we select a threshold of maximum 10% difference to obtain robust change features as input for the random forest classification.

targeted to European countries because the damage patterns included in the damage catalogue have been developed for built structures and building materials typical of European countries. In the future, the damage catalogue can be extended to cover damage patterns for areas outside Europe, which then allows to apply the method also to these areas. Moreover, as for all change-based approaches (e.g., de Gélis et al., 2021; Xu et al., 2021), our approach requires pre-event and post-event 3D point clouds to be available of the affected area to classify building damage. The increasing availability of pre-event point clouds, due to decreasing costs of UAV-borne point cloud acquisitions, however, provides great potential for the application of the approach.

Generally, our approach could be modified to be applicable to other natural disasters. As characteristic damage patterns are different for other disaster types, this transfer requires domain knowledge and an adaption of damage patterns that are typical for the respective disaster type. Similar to our use case of earthquakes, the damage patterns have to generalise adequately to discriminate target damage grades in point clouds. Subsequently, damage patterns can be modelled into 3D building models for training in the same way to classify damage also for other natural disasters.

5.3. Automatic generation of large 3D training datasets

Simulated training data in this study comprises the manual modelling of damage patterns into 3D building models, and the annotation of buildings with damage labels. As such training data can be generated to train machine learning models before an earthquake occurs, we consider the time effort for manual modelling in our approach acceptable. The trained models can then be directly applied to classify damage in real-world datasets acquired after an earthquake occurs, which saves valuable time for rescue and remediation actions (Kohns et al., 2022). Future approaches of UAV damage assessment in earthquake response might integrate databases of existing damaged and undamaged 3D building models. Such databases could integrate both building models generated from real-world city models (e.g., Uggla et al., 2023; Zihang et al., 2018) and synthetically generated building models from other sources, including VLS as in our study. The database could be connected with automatic damage modelling procedures, which might be developed in the future, and with the VLS module to assemble a multitude of different scenes where building objects can be flexibly interchanged. Such an approach could be especially valuable for classification methods with high demands for training data, such as deep learning (de Gélis et al., 2023). While we classify damage on

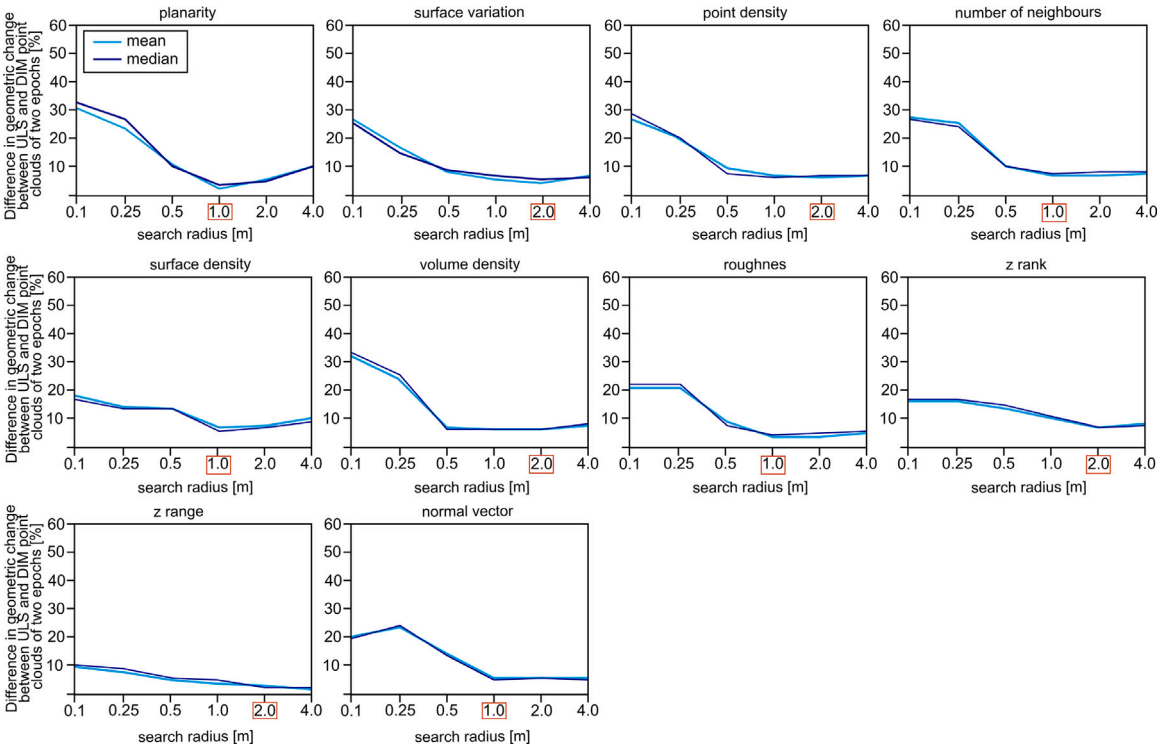


Fig. 10. Robust object-specific change features which show less than 10% difference between ULS and photogrammetric point clouds of two epochs in our experimental investigation. Red boxes indicate the local neighbourhood size for which differences are lowest. These features yield a good compromise between the similarity of change features between both sources of input point clouds and the number of features available for damage assessment.

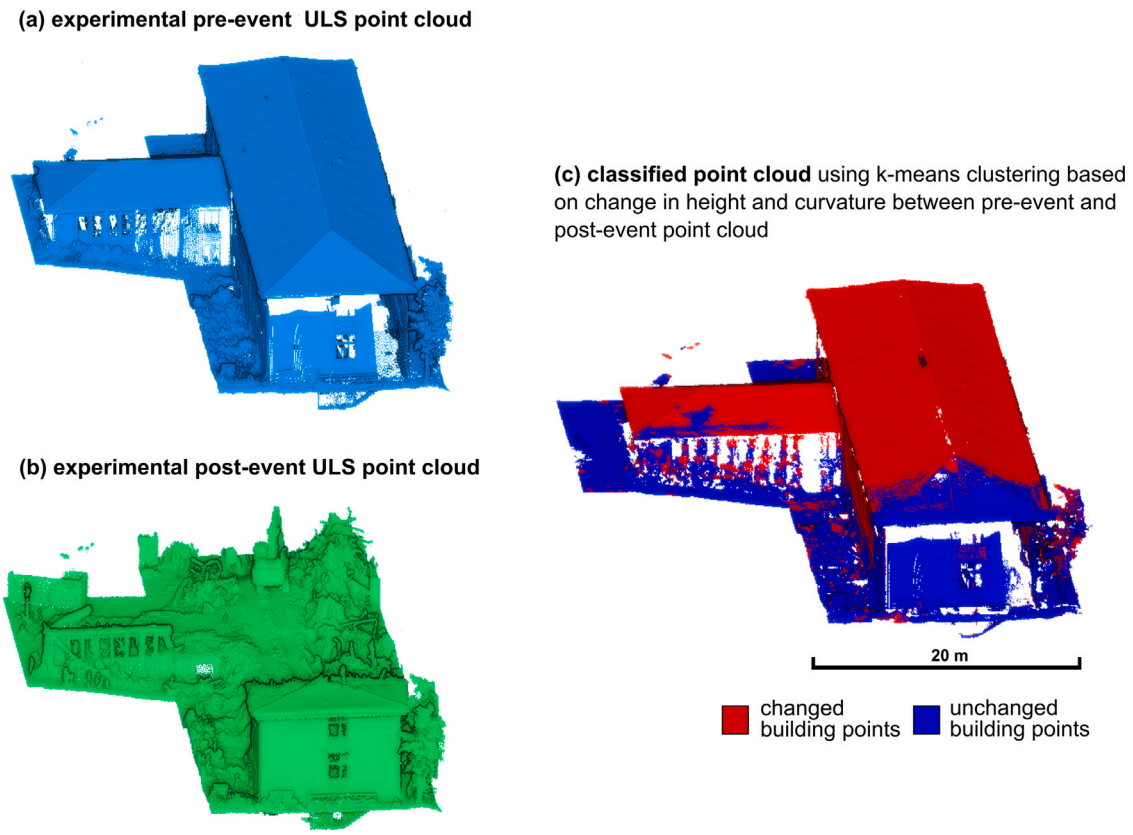


Fig. 11. (a) Pre-event and (b) post-event UAV-borne laser scanning (ULS) point clouds of the experimental building and (c) classified changed (red) and unchanged (blue) building points by k-means clustering. Changed building points are input for the extraction of change features as input for the random forest classification. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(a) VLS generic classifier

Classes (predicted / actual)	No damage	Heavy damage	Extreme damage	Destruction	Total
No damage	33	0	0	0	33
Heavy damage	6	16	2	0	24
Extreme damage	0	3	25	3	31
Destruction	0	0	5	32	37
Total	39	19	32	35	125

(b) VLS region-specific classifier

Classes (predicted / actual)	No damage	Heavy damage	Extreme damage	Destruction	Total
No damage	32	0	0	0	32
Heavy damage	7	16	0	0	23
Extreme damage	0	3	25	3	31
Destruction	0	0	7	32	39
Total	39	19	32	35	125

(c) VLS generic + real-world photogrammetric classifier

Classes (predicted / actual)	No damage	Heavy damage	Extreme damage	Destruction	Total
No damage	32	0	0	0	32
Heavy damage	7	15	0	0	22
Extreme damage	0	4	26	4	34
Destruction	0	0	6	31	37
Total	39	19	32	35	125

(d) Real-world photogrammetric classifier

Classes (predicted / actual)	No damage	Heavy damage	Extreme damage	Destruction	Total
No damage	35	0	0	0	35
Heavy damage	4	15	0	0	19
Extreme damage	0	4	26	2	32
Destruction	0	0	6	33	39
Total	39	19	32	35	125

Fig. 12. Confusion matrices for the trained random forest classifiers (a) VLS generic, (b) VLS region-specific, (c) VLS generic + real-world photogrammetric, and (d) real-world photogrammetric for multi-class damage classification of 125 buildings in the real-world photogrammetric dataset of L'Aquila, Italy.

building level in our approach, it generally offers flexibility with respect to the spatial unit of extracting changed building parts and subsequent damage classification. Instead of entire building objects, the approach might be tested to assess damage only for coherent units representing partial buildings of, for example, large building complexes.

6. Conclusion

We present a novel approach to automatically classify multi-class structural building damage using pre- and post-event point clouds of an earthquake event. We evaluate a supervised machine learning model

trained on simulated point clouds from virtual laser scanning (VLS) with respect to its capacity to classify damage grades no damage, heavy damage, extreme damage, and destruction in a real-world photogrammetric dataset. Damage is thereby assessed through change of geometric point cloud features between multi-temporal point clouds.

Our results reveal transferability with respect to multi-source point clouds used for training and application of the model. This is possible using a set of robust object-specific change features. Consequently, VLS provides a valuable source of training data for the classification task at hand.

We further achieve transferability of the model between region-specific site characteristics by integrating domain knowledge from earthquake engineering in the generation of realistic simulated training data. This allows training the model on geometric change which characterises the target damage grades across different geographic regions. The classification of multiple damage grades in the real-world dataset yields high accuracies (overall accuracy: 92%–95%). Accuracies only slightly improve when using real-world region-specific training data (+ 2%). The same applies for the binary case of detecting damaged buildings for which the classifier trained on generic simulated training data detects 85% of damaged buildings. Using real-world region-specific training data increases the detection rate by +3.1%.

Therefore, our developed approach provides a powerful assessment of multi-class structural building damage. We consider it especially relevant for applications where timely information on the damage situation is required, often linked to the situation that sufficient real-world training data is not available.

CRediT authorship contribution statement

Vivien Zahs: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Software, Visualization, Writing – review & editing. **Katharina Anders:** Conceptualization, Methodology, Formal analysis, Investigation, Validation, Software, Writing – review & editing. **Julia Kohns:** Writing – review & editing. **Alexander Stark:** Review & editing. **Bernhard Höfle:** Conceptualization, Methodology, Review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The 3D building models, simulated point clouds and Python scripts used in this article will be openly available on heiDATA (https://doi.org/10.11588/data/D3WZID/) after publication of this manuscript.

Acknowledgements

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF), Federal Ministry of Education and Research under Grant 03G0890 in the frame of the project LOKI and by the Deutsche Forschungsgemeinschaft (DFG), German Research Foundation, in the frame of the project VirtuaLearn3D under Grant 496418931. We thank CGR S.p.A. (Compagnia Generale Ripreseaeree) for providing the image data of L'Aquila for deriving the photogrammetric point clouds used in this study.

References

Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 1–74. http://dx.doi.org/10.1186/s40537-021-00444-8.

Awrangjeb, M., Fraser, C.S., Lu, G., 2015. Building change detection from LiDAR point cloud data based on connected component analysis. ISPRS Ann. Photogrammm. Remote Sens. Spat. Inf. Sci. II-3/W5, 1–8. http://dx.doi.org/10.5194/isprsannals-II-3-W5-393-2015.

Besl, P., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14 (2), 239–256. http://dx.doi.org/10.1109/34.121791.

Blender Online Community, 2018. Blender — A 3D modelling and rendering package (Version 2.93.0). http://www.blender.org.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.

de Gélis, I., Lefèvre, S., Corpetti, T., 2021. Change detection in urban point clouds: An experimental comparison with simulated 3D datasets. Remote Sens. 13, 1–29. http://dx.doi.org/10.3390/rs13132629.

de Gélis, I., Lefèvre, S., Corpetti, T., 2023. Siamese KPConv: 3D multiple change detection from raw point clouds using deep learning. ISPRS J. Photogramm. Remote Sens. 197, 274–291. http://dx.doi.org/10.1016/j.isprsjprs.2023.02.001.

Dorninger, P., Nothegger, C., 2007. 3D segmentation of unstructured point clouds for building modelling. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 36, 191–196, https://www.pf.bgu.tum.de/isprs/pia07/puba/PIA07_Dorninger_Nothegger.pdf. (Accessed 22 May 2023).

Earthquake Engineering Research Institute, 2009. Learning from Earthquakes: The Mw 6.3 Abruzzo, Italy, Earthquake of April 6, 2009. EERI Special Earthquake Report, June, EERI, Oakland, CA, pp. 1–12, https://www.reluis.it/doc/pdf/Aquila/EERI_L_Aquila_report.pdf. (Accessed 7 March 2023).

Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Grau, E., Rubio, J., Cook, B.D., Morton, D.C., Sun, G., 2016. Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing. Remote Sens. Environ. 184, 418–435. http://dx.doi.org/10.1016/j.rse.2016.07.010.

Geo-Engineering Extreme Events Reconnaissance, 2009. Preliminary Report on the Seismological and Geotechnical Aspects of the April 6 2009 L'Aquila Earthquake in Central Italy, Version 2.0. GEER Association Report No. GEER-016, pp. 1–166, https://www.academia.edu/67093517/Preliminary_Report_on_the_Seismological_and_Geotechnical_Aspects_of_the_April_6_2009_LAquila_Earthquake_in_Central_Italy_Version_2_0_GEER_Association_Report. (Accessed 7 March 2023).

Hildebrand, J., Schulz, S., Richter, R., Döllner, J., 2022. Simulating LiDAR to create training data for machine learning on 3D point clouds. ISPRS Ann. Photogrammm. Remote Sens. Spat. Inf. Sci. X-4/W2-2022, 105–112. http://dx.doi.org/10.5194/isprs-annals-X-4-W2-2022-105-2022.

Höfle, B., Mücke, W., Dutter, M., Rutzinger, M., Dorninger, P., 2009. Detection of building regions using airborne LiDAR: a new combination of raster and point cloud based GIS methods. In: Proc. of the Third Geoinformatics Forum Salzburg: Geoinformatics on Stage. July 7–10, 2009, pp. 66–75.

Huang, H., Sun, G., Zhang, X., Hao, Y., Zhang, A., Ren, J., Ma, H., 2019. Combined multiscale segmentation convolutional neural network for rapid damage mapping from postearthquake very high-resolution images. J. Appl. Remote Sens. 13 (2), 1–15. http://dx.doi.org/10.1117/1.JRS.13.022007.

Kerle, N., Nex, F., Gerke, M., Duarte, D., Vetrivel, A., 2020. UAV-based structural damage mapping: A review. ISPRS Int. J. Geo-Inf. 9 (1), 1–23. http://dx.doi.org/10.3390/ijgi9010014.

Khoshelham, K., Oude Elberink, S., Xu, S., 2013. Segment-based classification of damaged building roofs in aerial laser scanning data. IEEE Geosci. Remote Sens. Lett. 10 (5), 1258–1262. http://dx.doi.org/10.1109/LGRS.2013.2257676.

Kohns, J., Stempniewski, L., 2021. Classification of earthquake-induced building damage using innovative methods. IABSE Congr. Struct. Eng. Future Soc. Needs 1366–1374. http://dx.doi.org/10.2749/ghent.2021.1366.

Kohns, J., Stempniewski, L., Stark, A., 2022. Development of damage catalogues for visual assessment of buildings in the event of an earthquake. Bauingenieur 97 (12), 403–412. http://dx.doi.org/10.37544/0005-6650-2022-12-39.

Munawar, H.S., Ullah, F., Qayyum, S., Heravi, A., 2021. Application of deep learning on UAV-based aerial images for flood detection. Smart Cities 4 (3), 1220–1242. http://dx.doi.org/10.3390/smartcities4030065.

North, P.R.J., Rosette, J.A.B., Suárez, J.C., Los, S.O., 2010. A Monte Carlo radiative transfer model of satellite waveform LiDAR. Int. J. Remote Sens. 31, 1343–1358. http://dx.doi.org/10.1080/01431160903380664.

Otepka, J., Ghuffar, S., Waldhauser, C., Hochreiter, R., Pfeifer, N., 2013. Georeferenced point clouds: A survey of features and point cloud management. ISPRS Int. J. Geo-Inf. 2 (4), 1038–1065. http://dx.doi.org/10.3390/ijgi2041038, https://www.mdpi.com/2220-9964/2/4/1038.

Qing, Y., Ming, D., Wen, Q., Weng, Q., Xu, L., Chen, Y., Zhang, Y., Zeng, B., 2022. Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level. Int. J. Appl. Earth Obs. Geoind. 112, 102899. http://dx.doi.org/10.1016/j.jag.2022.102899.

Roynard, X., Deschaud, J.-E., Goulette, F., 2016. Fast and robust segmentation and classification for change detection in urban point clouds. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XLI-B3, 693–699. http://dx.doi.org/10.5194/isprs-archives-XLI-B3-693-2016.

Tran, T.H.G., Ressl, C., Pfeifer, N., 2018. Integrated change detection and classification in urban areas based on airborne laser scanning point clouds. Sensors 18 (2), 1–21. http://dx.doi.org/10.3390/s18020448.

TurboSquid, Inc., 2023. Free3D. https://free3d.com/. (Accessed 1 March 2023).

Uggla, M., Olsson, P., Abdi, B., Axelsson, B., Calvert, M., Christensen, U., Gardevärn, D., Hirsch, G., Jeansson, E., Kadric, Z., Lord, J., Loreman, A., Persson, A., Setterby, O., Sjöberger, M., Stewart, P., Rudenå, A., Ahlström, A., Bauner, M., Hartman, K., Pantazatou, K., Liu, W., Fan, H., Kong, G., Li, H., Harrie, L., 2023. Future Swedish 3D city models — Specifications, test data, and evaluation. ISPRS Int. J. Geo-Inf. 12 (2), 47. http://dx.doi.org/10.3390/ijgi12020047.

Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G., 2018. Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS J. Photogramm. Remote Sens. 140, 45–59. http://dx.doi.org/10.1016/j.isprsjprs.2017.03.001.

Vetrivel, A., Gerke, M., Kerle, N., Vosselman, G., 2015. Identification of damage in buildings based on gaps in 3D point clouds from very high resolution oblique airborne images. ISPRS J. Photogramm. Remote Sens. 105, 61–78. http://dx.doi.org/10.1016/j.isprsjprs.2015.03.016.

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS J. Photogramm. Remote Sens. 105, 286–304. http://dx.doi.org/10.1016/j.isprsjprs.2015.01.016.

Winiwarter, L., Esmorís Pena, A.M., Weiser, H., Anders, K., Martínez Sánchez, J., Searle, M., Höfle, B., 2022. Virtual laser scanning with HELIOS++: A novel take on ray tracing-based simulation of topographic full-waveform 3D laser scanning. Remote Sens. Environ. 269, 1–18. http://dx.doi.org/10.1016/j.rse.2021.112772.

Xiu, H., Liu, X., Wang, W., Kim, K.S., Shinohara, T., Chang, Q., Matsuoka, M., 2023. DS-Net: A dedicated approach for collapsed building detection from post-event airborne point clouds. Int. J. Appl. Earth Obs. Geoind. 116, 103150. http://dx.doi.org/10.1016/j.jag.2022.103150.

Xu, N., Huang, D., Song, S., Ling, X., Strasbaugh, C., Yilmaz, A., Sezen, H., Qin, R., 2021. A volumetric change detection framework using UAV oblique photogrammetry — A case study of ultra-high-resolution monitoring of progressive building collapse. Int. J. Digit. Earth 14 (11), 1705–1720. http://dx.doi.org/10.1080/17538947.2021.1966527.

Zihang, Y., Nagel, C., Kunde, F., Hudra, G., Willkomm, P., Donabauer, A., Adolphi, T., Kolbe, T.H., 2018. 3DCityDB — 3D city database version 3.0. Open Geospat. Data Softw. Stand. 3 (5), http://dx.doi.org/10.1186/s40965-018-0046-7.