UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Developing a Spatial Framework for Earthquake-Induced Building Damage Assessment Using Remote Sensing Data and Advanced Machine Learning methods

by Sahar Soleimanimatin

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Distinguished Professor Biswajeet Pradhan and Dr. Mohsen Naderpour

Faculty of Engineering and Information Technology

June 2023

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Sahar Soleimanimatin declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Sahar Soleimanimatin

Date: 6/06/2023

DEDICATION

I would like to dedicate this thesis to Reyhan Jafarizade who was trapped under debris for six hours after the 2003 Bam/Iran earthquake. Young Reyhan was rescued but she lost all her family in that incident. She claims her sister was alive as she could hear her just before the rescuers arrived. Now she is a psychologist helping people like herself who went through trauma in their life. Her story was my motivation for working on this research topic. May this research contribute to the automation of the building damage detection process to speed up the rescue time.

ACKNOWLEDGMENT

There are many people without whom this thesis might not have been completed, and to whom I am greatly indebted. I express my deep and sincere sense of indebtedness to my supervisor Distinguished Professor Biswajeet Pradhan, at the School of civil and environmental engineering, Faculty of Engineering and IT for his invaluable guidance, painstaking effort, constant encouragement, and inspiration during every step of my project work throughout the Ph.D. period. Despite his extremely busy schedule, I have always found him accessible for suggestions and discussions.

I would like to thank Dr. Mohsen Naderpour for providing me the academic facilities and constant motivation. I am also grateful to all candidature assessment panel members for their discussions and suggestions.

I wish to convey my heartiest thanks to my lab mate Abhirup Dikshit and all my friends from university for their encouragement, and moral and technical support during my thesis work.

I am very much thankful to the UTS Women in Engineering and IT (WiEIT) community and the team for always being there for me through my Ph.D. journey.

I would like to thank Dr. Amir Asghari for always believing in me and inspiring me to be my best. I truly appreciate his continued support, words of wisdom, and encouragement.

Further, I would like to express my gratitude to the office staff of the department for their help in administrative matters.

Last but not least, I am grateful to my family and my friends (especially Eman Dasi) for their consistent emotional support and inspiration.

LIST OF PAPERS/PUBLICATIONS

1. Matin, S. S., & Pradhan, B. (2021). Challenges and limitations of earthquake-induced building damage mapping techniques using remote sensing images-A systematic review. Geocarto International, 1-27.

2. Matin, S. S., & Pradhan, B. (2021). Earthquake-induced building-damage mapping using Explainable AI (XAI). Sensors, 21(13), 4489.

PUBLICATIONS INCLUDED IN THIS THESIS

Publication citation – incorporated within chapters in a conventional form.

Sahar S Matin, Biswajeet Pradhan (2021) 'Challenges and Limitations of Earthquake-Induced Building Damage Mapping Techniques Using Remote Sensing Images-A Systematic Review' Geocarto International. **https://doi.org/10.1080/10106049.2021.1933213**

Contributor	Statement of contribution	Thesis Chapters
Sahar Soleimanimatin	Reviewing and analysing the research literature (100%) Designing the research and writing the manuscript (75%)	Chapter 1, 2
Biswajeet Pradhan	Directing the research and reviewing the manuscript (15%)	

Sahar S Matin, Biswajeet Pradhan (2021) 'Earthquake-Induced Building-Damage Mapping Using Explainable AI (XAI)' Sensors, 21, 4489 **https://doi.org/10.3390/s21134489**

Contributor	Statement of contribution	Thesis Chapters
Sahar Soleimanimatin	Reviewing literature and performing analysis (100%) Designing the research and writing the manuscript (75%)	Chapter 2, 3, 4, 5
Biswajeet Pradhan	Directing the research and reviewing the manuscript (15%)	

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
DL	Deep Learning
LiDAR	Light Detection and Ranging
ML	Machine Learning
MLP	Multilayer Layer Perceptron
ML-user	Machine Learning User
OSM	Open Street Map
ReLU	Rectified linear unit
RF	Random Forest
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SVM	Support Vector Machine
USGS	The United States Geological Survey
VIM	Variable Important Measurement
XAI	Explainable Artificial Intelligence

ABSTRACT

Building-damage mapping using remote sensing images is crucial to support post-earthquake rescue and relief activities. Further, in the long term, knowledge regarding details of damage after an earthquake aids decision-makers in better designing buildings and infrastructure which can withstand destruction in future seismic events. Building damage mapping in real-world earthquake scenarios is limited mainly to the manual interpretation of optical satellite images, which is costly and labour-intensive but transparent and reliable. In this regard, a large body of literature proposes novel machine learning-based frameworks for automating earthquake-induced building damage mapping. These frameworks in this domain are promising yet unreliable for several reasons, including but not limited to the dependency on the availability of both pre and post-earthquake images, lack of quality and the limited number of labelled images, along with transferability and explainability of the machine learning models.

Therefore, this research addresses these issues by developing a transferable and explainable machine learning-based framework for earthquake-induced building damage assessment using a single post-earthquake satellite image. The city of Palu, Indonesia and Salina Cruz, Mexico, was selected for this research because of the high number of collapsed buildings after the 2018 Indonesia and 2017 Mexico earthquakes, respectively, and the availability of the post-earthquake optical satellite images for these events.

In the first step, the performance of the U-Net, multilayer perceptron (MLP), Random Forest (RF) and Support Vector Machine (SVM) was evaluated and discussed to select the best model for identifying the collapsed and non-collapsed buildings in Palu City. The result showed that i) a single post-earthquake image can be used for building damage mapping, and ii) MLP outperformed the other models with an overall accuracy of 84%. Next, to build an explainable model, the Shapley additive explanation (SHAP) was utilised to analyse the impact of each feature descriptor on the output of the proposed MLP model. The results showed that i) the spectral features contribute the most to classifying the collapsed and non-collapsed buildings in the Palu City study area, and ii) SHAP enables us to analyse the data in detail and guides us to build a more diverse dataset. Next, Salina Cruz, Mexico, was considered a geographically different study area to assess the transferability of the proposed model. The conclusions of this research can be used by the researchers and the decision makers towards the use of explainable AI for post-earthquake building-damage mapping.

Keywords: Building-damage Mapping; Machine Learning; Explainable Machine Learning; Remote Sensing

CHAPTER 1

INTRODUCTION

This chapter provides a general overview of the research subject, research background, and research goals and objectives. This chapter also reveals the main context of the study, structure, problem statement, the research main objective, research plan, research questions, research motivation, research limitations and challenges, and thesis organization. It addresses the earthquake impacts and provides the importance of post-earthquake building damage assessment. This chapter briefly analyses the studies using different types of remote sensing data, gives an overview of the process of automatic building damage detection, summarises the current trend of the methods in the domain of building damage detection using optical remote sensing images. The chapter addresses some challenges and limitations that hinder the current damage mapping and also the automatic building damage mapping through an operational lens.

1.1 General introduction

Earthquake is one of the most destructive environmental disasters that affect mankind (Wijkman & Timberlake, 2021). On average, two earthquakes per minute occur across the world, summing up to a million earthquakes per year (Dong & Shan, 2013). On the other hand, rapid urban population growth, combined with socioeconomic factors, increases the risk of earthquake-induced damages in urban areas. According to studies, damage to buildings accounts for most of the casualties after an earthquake event (Shohet et al., 2014). Although earthquakes are inevitable, proper and timely response and rescue can mitigate the damages caused by such a devastating natural disaster. The behavior and response of a building to an earthquake can determine the number of people who might get injured or killed as most earthquake causalities is associated with severe damage to the buildings. For that reason, rapid and efficient earthquake impact assessment is crucial for search and rescue operations in a short period (Nex et al., 2019). In the long term, knowledge regarding details of damage after an earthquake aids decision-makers in better designing buildings and infrastructure which can withstand destruction in future seismic events (Xuanmei Fan et al., 2019). Remote sensing technologies, both satellite-borne and airborne, can provide valuable data in this regard, thanks to their synoptic view, repetitive coverage, and providing low-cost multi-spectral and multiresolution data which we will discuss further in this paper (Kerle et al., 2019).

A building footprint is the perimeter of a building that outlines the exterior walls and provides information about the building's location, shape, and area of the building. The building footprint is essential information in earthquake vulnerabilities and damage assessment. The building footprint is also used in creating the post-earthquake damage map to show the location of the damaged and collapsed buildings.

There are different ways to obtain the building footprint dataset including the available updated datasets such as the detailed cadastral maps (Pfeifer et al., 2007), creating a new building footprint dataset by performing a ground survey (Zhou & Troy, 2008), digitizing the available remote sensing imagery (LaJeunesse Connette et al., 2016), or using an open-source platform (Hecht et al., 2013).

Although the cadastral maps and ground surveying provide an accurate and reliable local building footprint, these data sources are expensive, time-consuming, and resource-intensive and there is no guarantee that they are available in all the earthquake-affected countries. Therefore, such datasets generally cannot be used worldwide in the rapid building damage mapping process in case of a natural hazard where time is considered a critical factor.

Digitizing an available optical satellite or aerial image is another common method to obtain the building footprint (Gavankar & Ghosh, 2018). However, in this method, the footprint of the building roof is outlined which might not be necessarily the same as the actual building footprint. Although having access to high-resolution satellite images is costly, free access to high-resolution satellite images is possible through Bing Maps under the Open Street Map (OSM) project (Haklay & Weber, 2008).

Remote sensing-based damage mapping is limited to the visual interpretation of the optical satellite images in real-world scenarios which are labor-intensive and time-consuming (Ma et al., 2016; Pham et al., 2014). Therefore, there is an interest in automatic post-earthquake damage mapping. Artificial intelligence (AI) has been widely used in automatic dam-age mapping studies using remote sensing images (Abdollahi, Pradhan, et al., 2020b; Dikshit et al., 2020; Syifa et al., 2019). However, AI-based solutions have not been yet entirely adapted to earthquake-induced damage mapping (Bai et al., 2018). One reason that hinders AI to be used reliably in real-world applications is the lack of explainability of AI models to justify the AI decisions (Ahmad et al., 2020).

Recently, explainability has become extremely crucial for the practical deployment of AI models, and there is a growing interest in this domain (Arrieta et al., 2020). The Explainability of the AI model is even more important in the domains, such as post-

earthquake rapid damage mapping that has a high level of resistance to error. It means that any error in detecting collapsed buildings may cause a delay in rescuing the trapped victims and consequently can affect human life (Syifa et al., 2019). Exploring how the AI-model inputs impact the output leads us to better understand the behavior of the model and aids us in building a more robust explainable AI model (Adadi & Berrada, 2018). Explainable AI (XAI) refers to the AI models of which the output can be understood by humans which is the opposite of the term "black box" for machine learning models. Therefore, XAI models provide insight into the causes of the model output (Gunning, 2017). Further, the explainability can assist in identifying the unknown vulnerabilities and flaws of the AI system and therefore, enabling us to correct the errors (Adadi & Berrada, 2018).

1.2 Research background

The first preliminary attempt to use remote sensing imagery for damage mapping dates back to 1906 when San Francisco was hit by a 7.9 Mw earthquake and a series of kites were used as aerial platforms to carry a 20 kg camera to map part of the affected area (Dell'Acqua & Gamba, 2012), while the first application of satellite images in the seismic field probably dates back to 1972 when six satellite images were used to investigate the cause of the 1964 Alaska earthquake (Moher et al., 2009). These make earthquake-induced damage mapping one of the oldest applications of remote sensing, yet one of the few challenging methods in terms of operational image-based mapping, and therefore to date remains a topic of active research (Kerle et al., 2019).

There are several studies using different automatic and semi-automatic methods to generate a post-earthquake building damage map. We have compiled Figure 1.1, which illustrates damage mapping workflow through an operational lens while considering the time and accuracy in each stage affecting the overall efficiency of each method. For this workflow to function, a series of requirements need to be met in each stage. These stages and requirements are listed below:

Identifying the right pre-event data within the archives and updating it if necessary;

- Post-event data acquisition with the required resolution;
- Defining damage scale by following standard operational guidelines;
- Pre-processing the image including preparing the training dataset;
- Damage detection processing using various methods;

• Quality Assurance (QA) and validation; and finally

• Generating a damage map.

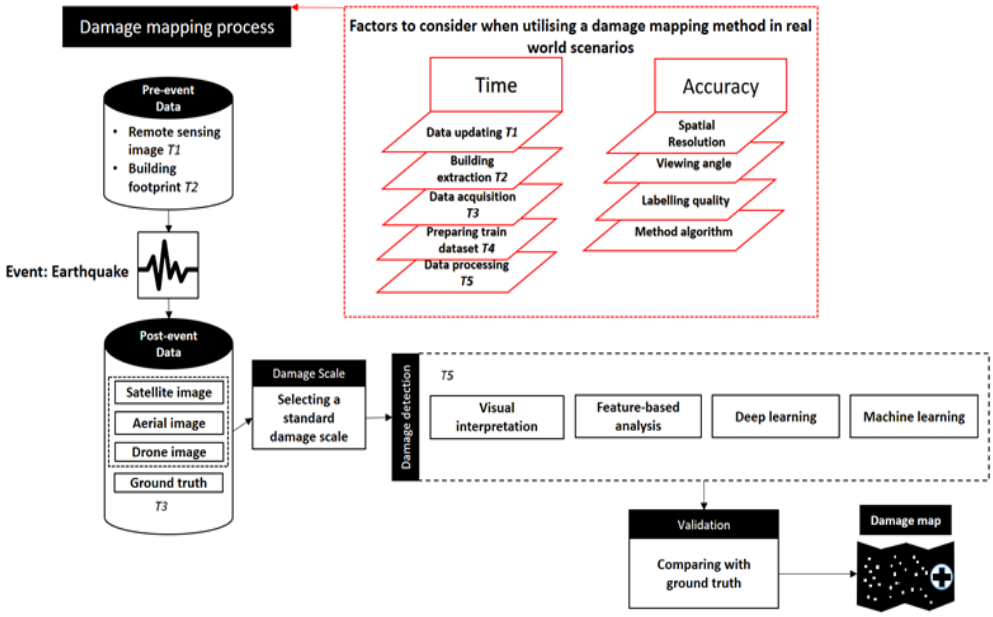

Figure 1.1. Damage mapping workflow through an operational lens. Ti is the time for each stage.

The study conducted by (Abdollahi, Pradhan, et al., 2020b) presented a time scale for operational automatic damage mapping and estimated the time required for each stage based on the framework they designed (Figure 1.2.). Further details about this study will be provided in chapter 2.

Figure 1.2. The presented time scale for operational damage mapping (Abdollahi, Pradhan, et al., 2020b)

Therefore, this research addresses this issue by developing a novel machine-learning model using optical satellite images.

Moreover, the majority of the studies develop building damage detection models based on using both pre and post-earthquake remote sensing images. Despite the successful performance of these methods, they cannot be generally applied to another study area in

real-world scenarios as there is a need for the availability of both pre and post-earthquake. Further, the use of only post-event images avoids pre-processing work like co-registration after a major earthquake where ground distortion happens. Hence, to fill this gap, a single post-earthquake image was used in this research.

1.3. Earthquake around the world

Earthquake is some of the most destructive environmental hazard to affect mankind. During 2001 to 2011 alone structural collapses resulting from massive quakes were responsible for about 60% of all disaster-related casualties (Bartels & VanRooyen, 2012). Although earthquakes are inevitable, proper respond to them can mitigate the damages caused by such a devastating natural disaster. For example, rapid and effective earthquake impact assessment greatly assist the decision makers to better respond to the disasters in a short amount of time. In addition, in long term the knowledge regarding to the details of the damage to the buildings aids the decision makers in better managing the future constructions and mitigate the damage to the structures and infrastructures in future seismic events, considering the fact that most of the causalities caused by earthquake is associated with the damages to the buildings (Xiwei Fan et al., 2019).

After a severe earthquake the immediate response is often hindered due to the lack of knowledge regarding to the extent, distribution of the damage and the location of the affected buildings. Fortunately utilizing the modern remote sensing technology along with the advanced machine learning methods creates a valuable opportunity for preparing building damage map (Cooner et al., 2016; Dong & Shan, 2013).

1.4. Earthquakes in Indonesia and Mexico

On 28th September 2018 at 17:02 UTC, a massive earthquake with a moment magnitude of 7.5 occurred on Sulawesi Island in Indonesia located on 5.36 LU-7.48 S and 117.02 -125.74 E. Subsequently, Palu City which is located approximately 80 km to the south of the epicenter of the earthquake-affected extensively. The ground vibration was also followed by devastating multiple Tsunamis and extensive soil liquefaction causing more than two thousand casualties and severe damage to the buildings and infrastructure along the Palu coast of central Sulawesi. It was reported that more than 68451 houses were damaged around Palu Bay which was the most affected urban area. The National Board for Disaster Management (BNPB) of Indonesia reported 2101 casualties and 95 million USD damage (Syifa et al., 2019). In the Palu neighbourhood of Balaroa around 1700 houses were swallowed up due to the liquefaction. To the best of our knowledge, a few research works

investigated the damages after Indonesia's 2018 earthquake. Since this study area is complex and a variety of events such as tsunamis, collapses, and liquefaction occurred as a result of this devastating earthquake, it would be a valuable opportunity to challenge the ability of the proposed method in detecting the damages to the buildings. Further, to assess the ability of the proposed model to adapt to a previously unseen dataset, Part of Salina Cruz city in the south of Mexico was considered as a second study area. Located among five tectonic plates, Mexico is also one of the most seismically active countries in the world (Godínez-Domínguez et al., 2020). On the 7th of September 2017 at 23:49 CDT a strong earthquake with a moment magnitude of 8.2 occurred in the Gulf of Tehuantepec, in the south of Mexico. Subsequently, the southern and central states of Mexico including Oaxaca were affected by this massive earthquake. As a result of this massive earthquake, more than 110,000 houses were damaged from which 41000 houses were considered collapsed. 65,044 of the damaged houses were located in the state of Oaxaca. Salina Cruz is one of the affected cities in the state of Oaxaca which was considered a second study area in this research (Godínez-Domínguez et al., 2020).

1.5. Problem statement

The increasing availability of various technical studies, programming tools, decrease in the cost of hardware, and step-by-step tutorials and instructions have made machine learning accessible to the public, however, there is a possibility that the required domain knowledge and understanding of the dataset, meaningful data features and how models can be improved or generalized to another dataset can be ignored (Horwath et al., 2020). One way to tackle this problem is to use an explainable machine learning model to shed light on how the model operates, understand the underlying reasons behind the model decision, and gain insight into which features play a more important role in the model outcome.

1.6. Research gaps

Using remote sensing images in the domain of natural hazard management, in particular, building damage detection has increased significantly and hence there is a potential for building damage mapping to be an automatic task (Duarte et al., 2018). However, despite the significant number of studies proposing different automatic or semi-automatic methods and significant technological developments, the emergency management services (EMS) such as the International Charter “Space and Major Disasters” or the Copernicus Emergency Management service, still widely rely on the visual interpretation of the optical satellite images to assess the damages to the buildings (Dell'Acqua & Gamba, 2012; Kerle et al., 2019).

Structural collapses are responsible for most of the casualties after the earthquake. Producing a damage map is essential for the response phase to mitigate the loss of life and distribute relief after the event. Although time is one of the critical components of the response phase, visual interpretation of the very high-resolution dataset is still widely used in operational practice to produce the damage map after an earthquake which is very time-consuming. To automatically produce a reliable damage map of the affected area the availability of the free data, type, and resolution of the data, the classification method, and the processing time are the most important factors to be considered. Researchers have conducted several invaluable researches to investigate every aspect of damaged building detection to rapidly and accurately produce a building damage map after the earthquake. However, a handful of studies considered all the factors and the majority of the proposed methods in the literature are not applicable in real-world rapid damage mapping. In this research, a framework for building damage mapping was proposed to leverage the advantage of freely available datasets. Multilayer perceptron (MLP) was proposed as an advanced machine learning classification model which requires a small size train dataset and is suitable for rapid damage mapping.

There is a significant number of studies in this domain that focus on developing novel ML-based models or optimizing the ML-based models to achieve higher accuracy. However, other model properties such as explainability, transparency, interpretability, and the possible domain knowledge are of importance in the domain of remote sensing applications. Especially, in using advanced machine learning models such as deep neural networks which produce high accuracy results but, lack these properties (Roscher et al., 2020).

One of the challenges of using remote sensing data is the limited number of annotated data. One reason behind this issue is that manually annotating the image dataset is time-consuming and expensive. Moreover, in cases such as post-earthquake assessment, the image samples are not always accessible. Developing an explainable and interpretable machine learning model along with using the interpretability tools can identify valuable new samples to increase the generalization ability and transferability of the model.

Therefore, in this section, we highlight some of the gaps that were identified in the extensive literature.

- Only a handful of studies in this domain investigated the capabilities of using a single post-earthquake satellite image.

- The studies used a specific image dataset and did not use open source image datasets.

- Most studies developed the classification models using a specific dataset from a particular geographic area.

- To the best of our knowledge, there is no study to discuss the advantages of explainable AI (XAI) in the domain of earthquake-induced building damage mapping.

- Assessing the transferability of the model is a huge gap in the study that has been ignored by most of the studies in this domain.

1.7. Scope of the study

Generally, the post-earthquake damage assessment that has evolved for a rescue mission, relief distribution, and damage estimation through image analysis, modeling, and mapping are the major scopes of this research.

The main objective of this study is to develop a machine learning-based framework to create an earthquake-induced building damage map using single post-earthquake optical remote sensing images. Further, this research attempts to assess the efficiency of several machine learning models to classify collapsed and non-collapsed buildings after an earthquake using post-earthquake satellite images. This study also covers the topic of transparency and explainability of the machine-learning models and discusses the importance of developing an explainable machine-learning model for assessing post-earthquake damages.

In terms of the scope of the data, the purpose of the study is to use only publicly available datasets. To this end, the post-earthquake satellite images from the open data program by Maxar Company (former Digital Glob) were used. The vector building footprint for the study area was extracted from Open Street Map (OSM). The study area is limited to an area of 3606 x 3148 pixels, about 3 km^2 that covers part of Palu, Indonesia, and an area of 4577 x 4864 pixels, about 1.8 km^2 covering part of Salina Cruz in Mexico.

Therefore, the scope of the research contains:

1. Earthquake-induced building damage assessment

2. Developing machine learning models

3. Explainable-AI

4. Damage mapping

1. High-resolution optical satellite images

2. Open Street Map (OSM) vector data

3. Feature-based image analysis

4. Transparent and explainable AI

1.8. Research aims and objectives

The aim of this study is to develop a machine learning-based framework to create an earthquake-induced building damage map using a single post-earthquake optical remote sensing image. The current research follows the three objectives to fulfil the gaps in the literature. The objectives include:

1. To assess advanced machine learning methods on damage building detection using publicly available single post-earthquake optical satellite images.

2. To develop an explainable machine learning model for building damage detection.

3. To assess the Transferability of the proposed building damage detection model on a different geographical area.

1.8.1. Objective 1

There are many studies in the literature that have used very high-resolution aerial images. Although these images provide more accurate information about the affected area, such images are not always available in a short amount of time after the earthquake. Further, a significant number of studies used multi-temporal images to compare the pre and post-earthquake images for detecting the changes between the images. However, a pair of high resolution pre and post-earthquake images that are cloud free and have and acquired from the same platform and sensor are not always available. Therefore, in this study the focus is on developing a machine-learning-based model using the capabilities of the publicly available data and a single post-earthquake optical satellite image.

Many studies in this domain have investigated the performance of numerous machine learning and deep learning models using different data that was collected from different platforms and sensors. Therefore, in this objective, the performance of different machine

learning and deep learning models for post-earthquake building damage mapping will be investigated.

1.8.2. Objective 2

To develop an explainable machine-learning model. The gap is that in this field (building damage detection) to the best of our knowledge no study considered the importance of explainability and interpretability of the model.

This objective emphasizes explaining a building damage detection model output which leads us towards an understanding of 1) how/which features contribute to the model output? 2) how the dataset characteristics affect the model output, and 3) how features behave differently in geographically diverse study areas? The insight we gain from using XAI for building damage detection models can be used as a rationality check for the training dataset, which is fundamental to machine learning models. To the best of our knowledge, there is no study examining the earthquake-induced building damage assessment models regarding the explainability to enhance the model and identify the existing limitations in the dataset, which is essential for building a transferable model.

1.8.3. Objective 3

To assess the transferability of the built model using a geographically different image dataset and add the dataset to the previous dataset to explore the efficiency of the model and show how explainability can assist us in increasing the performance of the model.

This study explored the 2017 Salina Cruz post-earthquake image as a geographically different new image dataset to assess the transferability of the proposed model.

1.9. Research questions

As mentioned the main objective of this study is to develop a machine learning-based framework to create an earthquake-induced building damage map using a single post-earthquake optical remote sensing image. To identify the steps, need to be taken to achieve this objective different question has been developed for each objective of this study.

1.9.1. Questions about objective 1

The following research questions were developed for objective 1.

i. What is the advantage of detecting earthquake-damaged buildings using publicly available remote sensing images?

ii. What are the required publicly available remotely sensed images and complementary datasets to fulfil the aim of the study?

iii. What is the definition of collapsed and non-collapsed buildings?

iv. Are the building labels and borders consistent throughout the dataset?

v. How to select the training dataset to gain enough information from the data?

vi. How to select the test dataset that represents a real-world earthquake scene?

vii. What machine learning models can be employed to classify the post-earthquake optical satellite image?

viii. What are the features that should be extracted from the image?

ix. What deep learning models can be employed to segment the post-earthquake optical satellite image?

x. What are the evaluation metrics to assess the performance of the classification and segmentation models?

1.9.2. Questions about objective 2

The following questions were developed for objective 2.

i. What are the advantages of using an explainable model?

ii. How to develop an explainable model?

iii. What are the most important features that contribute to the model outcome?

iv. What are the redundant features that should be removed from involving in the model?

v. How local interpretability can help to improve the quality of the dataset?

1.9.3. Questions about objective 3

The following questions were developed for objective 3.

i. Is the proposed model in objective 2, transferrable?

ii. What is the accuracy of the model on the unseen dataset?

iii. What is the accuracy of the model trained on both datasets and tested on the new dataset?

iv. To what extent adding a new set of data can increase the performance of the model?

1.10. The motivation behind this research

Remote sensing technology has revolutionized the way we respond to natural disasters such as an earthquake. Damage building mapping using remote sensing satellite images provides better emergency response to minimize the number of casualties and easier distribution of the relief. To prepare such maps visual interpretation of the very high resolution, optical satellite images are often used in practice to assess the damages to the buildings due to the

high accuracy of this method. However visually detecting the affected areas is very time-consuming and is not an efficient method for immediate response to the disaster particularly if the affected area is significantly large. Therefore, automatic detection methods such as machine learning have been suggested by previous studies to classify earthquake-induced building damages. Recently open data programs such as Digital Globe have provided several high-resolution free satellite images with limited spectral bands for large-scale natural disasters. Using automatic building damage classification is the subject of active research.

Despite the earthquake-induced damage mapping being one of the oldest applications of remote sensing, the increasing amount of publicly available optical satellite images, and the advancement in developing novel machine learning models, visual interpretation is used in operational image-based mapping, and therefore to date, this field remains a topic of active research.

Although extracting the information regarding the buildings in the affected area is essential, classifying the post-earthquake images is challenging because of their complexity. Further, there is a need for the models to be explainable to increase the reliability and open room for improving the quality of the training dataset.

Therefore, considering the impact and importance of operational automatic building damage mapping, developing a transferable and explainable machine learning model trained on a comprehensive building damage database deserves further attention and effort.

1.11. Novelty and main contribution of the research

The main objective of this research is to develop a machine-learning-based post-earthquake building damage framework that considers: i) using publicly available single-post earthquake optical satellite image, ii) utilizing the capabilities of Open Street Map (OSM) data, iii) the consistency of the building labels, iv) the standard damage scale, v) developing an explainable machine learning model and vi) adding a new set of a dataset from a different geographical area to assess the transferability of the proposed model.

To the best of our knowledge, no study considers the explainability of the machine learning model in the domain of natural hazards. Therefore, developing an explainable model which provides insight into the dataset is the main contribution of this research. Further, there are a few studies that utilized single post-earthquake satellite images for post-earthquake building damage mapping.

1.12. Thesis organization

Chapter 1 includes a general introduction to post-earthquake damage mapping and the importance of researching this topic. The research background will be discussed in this chapter to provide context to the topic, problem statement, and research gaps. The scope of the study will be explained. The research aims and objectives followed by research questions will be discussed. The motivation of the research, novelty, main contribution of the research, and the structure of the research will be explained in detail.

Chapter 2 demonstrates the literature on post-earthquake building damage assessment. The type of data in terms of platform, temporal and spatial resolution will be discussed. The available standard guidelines for defining damage levels will be explained. Different methods for change detection, classifying, and segmenting post-earthquake remote sensing images will be explained. The methods will be compared with respect to the accuracy, amount of time and data that is required for training the model, and complexity of the model. The overall advantages and limitations of the methods will be discussed in this chapter. Further, the explainability, transparency, and interpretability concepts will be explained and the necessity of considering explainability for machine learning models will be discussed.

Chapter 3 discusses the methodology of the research and the proposed models in detail. This chapter demonstrates and discusses the study area, data acquisition, the overall methodology, and the implementation of the models for building damage assessment. Further, this chapter describes the image features used in this study and feature analysis using Shapley Additive Explanation (SHAP) plots.

Chapter 4 includes the results of the study that have been obtained from developing several machine learning models for earthquake-induced building damage assessment. The limitation and challenges of these models will be discussed. Further, the results of SHAP analysis that lead to improving the proposed model will be discussed in this chapter. The results of assessing the transferability of the model will be shown. Finally, the generated building damage maps will be presented.

Chapter 5 includes the conclusion of each objective, the general conclusion of the research, the limitations of the study, and the future direction of the research.

CHAPTER 2

LITERATURE REVIEW

2.1. Introduction

This chapter provides a comprehensive review of post-earthquake building damage mapping methods by highlighting the data types and platforms, labels, various machine learning models, and explainable and interpretable machine learning models. This chapter introduces the most common ML models used in this field and the capabilities and limitations of these studies are discussed.

Further, to have a quantitative overview of the related academic literature, a systematic review was conducted. The research methodology is based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement which describes selecting items for systematic reviews, meta-analysis, and both (Moher et al., 2009). The review's scope is limited to the studies related to operational earthquake-induced building damage detection using optical remote sensing images. The Scopus and Web of Science (WoS) were used for extracting the relevant manuscripts. Several combinations of expressions were applied to search the manuscripts such as "Deep Learning, Remote Sensing, Earthquake; Machine Learning, Remote Sensing, Earthquake" and "Damage mapping, automatic, earthquake, building" over the period, from 2010 to August 2020. 154 records were initially found. After removing the repetitive articles, a set of inclusive and exclusive criteria was set as eligibility factors to determine the studies which contribute to the review objective. For checking these factors, the article title, keywords, and abstracts were evaluated by the authors. The exclusion factors were as follows:

i. Studies using ground-based remote sensing images.
ii. The full text of the article was not available.

For inclusion the following factors were established:
a. Articles published in peer-reviewed journals or conferences.
b. Written in English.
c. Published between 2010 to August 2020.
d. Developed a framework for automatic building damage mapping.

Among the 154 initial articles, 32 articles were eliminated as redundant or not related to the remote-sensing-based classification. In the final step, the articles aligned with the

objective of the present review were selected as the result of the process and will be presented in the following sections in forms of graphs in more detail.

Evaluating the existing automatic and semi-automatic methods can lead us to identify the issues in exploiting them in real scenarios, directing us to develop more reliable methods that can be applied in operational damage mapping practice using remote sensing images. Therefore, the last section of this chapter summarises the conclusions of these studies and current issues and outlines ideas for future research in this domain.

2.2. Space borne and aerial data Platforms

Considering the need for assessing the earthquake damage over a vast region, a very high-resolution (VHR) satellite image (spatial resolution <50 cm) which is usually made available within a few hours after the event serves as an ideal source of information in the response phase of the earthquake management (Wu et al., 2019). The satellite images have been extensively used in operational damage mapping in events such as the 2010 Haiti earthquake (Corbane et al., 2011), 2016 central Italy (Cotrufo et al., 2018), and 2018 Indonesia (Adriano et al., 2019).

Due to the limited spatial resolution (0.3-1m) and the viewpoint (near nadir looking) of optical satellite images, some types of damage such as pancake collapses and damage on the walls cannot be detected. The low resolution of the satellite images raises the question of whether this source of information can lead to accurate and certain damage mapping (Kerle & Hoffman, 2013) even when they are manually interpreted (Kerle, 2010; Saito et al., 2010). Aerial imagery, on the other hand, with manned aircraft, is often costly and faces difficulty in remote areas in the emergency timeframe (Watts et al., 2012). Therefore, in situ damage mapping is the potential solution that is possible by using unmanned aerial vehicle (UAV) images (Kerle et al., 2019).

However, due to the time-pressured nature of first responders to detect damaged areas, damage-detection methods including UAV image-based detection which requires manual intervention of an operator are considered time-consuming when potential lives are in danger. Industry experts and practitioners in this domain are suggesting a series of guidelines, which form a framework of requirements to assist in future systems that detect damages automatically to alert first responders, saving time and saving lives. The outlined system in this framework needs to deliver real-time or near-real-time updated data, which is easily interpreted by even non-experts within the field. The system should also be inexpensive and easy to deploy while taking into consideration operational limitations

which could be legal or otherwise introduced due to emergency conditions. An example of such a system was tested by first responders in INACHUS (Technological and Methodological Solutions for Integrated Wide Area Situation Awareness and Survivor Localization to Support Search and Rescue Teams) which is a European union-funded research project aimed at providing timely and detailed information on potential damaged areas and location of survival. In this project, which was proposed by (Nex et al., 2019) as the UAV starts capturing images across the area, the images are streamed and progressively are classified into two classes of damaged and intact.

Figure 2.1. shows the number of earthquake-induced building damage detection studies using images from space-borne, airborne, or both platforms between the years 2010 to 2020.

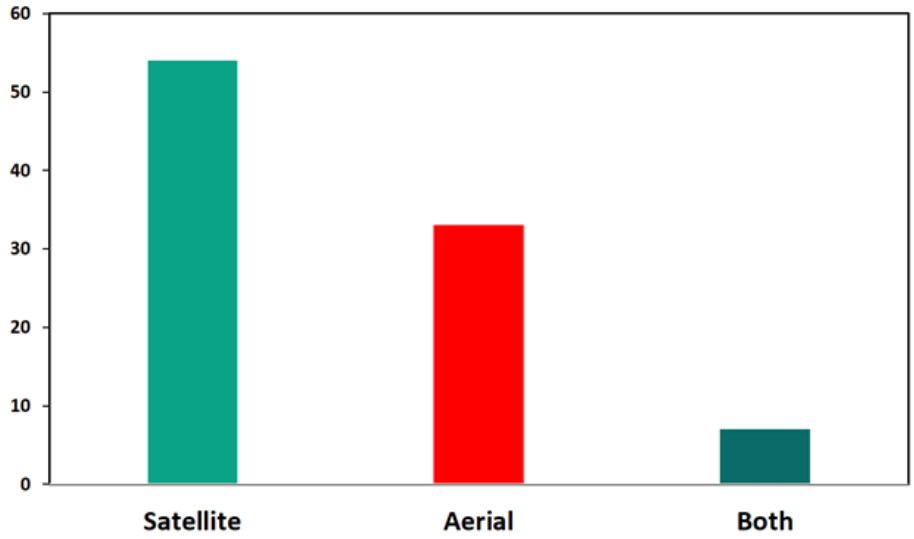

Figure 2.1. The number of earthquake-induced building damage detection studies using images from space-borne, airborne, or both platforms between the years 2010 to 2020.

2.3. Data sensors

To produce a damage map, different types of remote sensing images have been utilised depending on the sensor that captures the information. The donut chart below (Figure 2.2.) indicates the number of studies using a single image type or a combination of different image types.

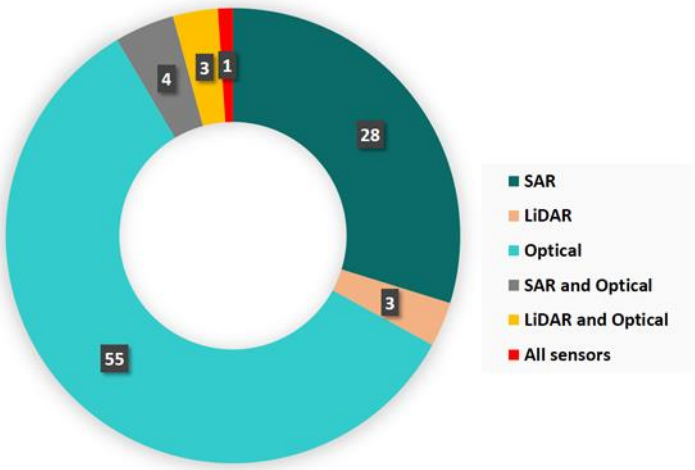

Figure 2.2. The number of studies using different image types

2.3.1. Optical Images

Optical images are the most favourable image types for operational damage mapping as these types of data are easier to interpret compared to other remote sensing data such as synthetic aperture radar (SAR) images (Duarte et al., 2018; Wen et al., 2014). Building characteristics such as texture, geometric shape, and spectral reflectance can be obtained easily from optical images. They also have the potential to allow 3D reconstruction if they are captured with enough overlap. The generated 3D building model after an earthquake can deliver valuable information about the geometric deformation of the building resulting in more accurate detection of damages to the building. However, automatic 3D building reconstruction through dense image matching is very time-consuming and stays theoretical rather than to be used in operational rescue and relief practice (Gong et al., 2016). To overcome the drawbacks of dealing with capturing multiple images, light field technology was introduced to enable 3D building reconstruction using a single snapshot. This advanced technology provides multiple viewpoints in one snapshot resulting in a 3D image of a scene. In recent work, (Farhood et al., 2020) proposed a novel method to accurately generate a 3D point cloud using this technology. Although the proposed method has shown great potential to be used in detecting damage to buildings, it is still far from being adopted widely in the domain of building damage detection during natural hazard events.

2.3.2. SAR Images

SAR (Synthetic Aperture Radar) images on the other hand are a reliable source of information for discriminating the damages vs. buildings (Cotrufo et al., 2018). The regular and cube shape of the buildings make them distinctive to detect in SAR images (Bai et al., 2018). Furthermore, building texture features are relatively stable in this type of image.

Therefore, changes in the shape and texture of the building are distinguishable in SAR images. Moreover, the backscattering intensity and intensity correlation coefficients can be used in building damage assessment (Gupta et al., 2019). For example, in a preliminary study to exploit SAR images to detect the earthquake-affected area, (Vetrivel et al., 2015) utilised multi-temporal ALOS-2 PALSAR-2 obtained after the 2015 Nepal earthquake considering the high values of the difference and low values of correlation coefficient in SAR backscattering coefficient. The information from L-band SAR images showed promising potential in assessing the damages to individual buildings however, the low spatial resolution made the study limited to assessing the damage in the block level (25mX25m). To evaluate the result, pre and post-high-resolution optical images and drone-based video footage were used as ancillary data. In the research conducted by (Schweier & Markus, 2006), a TerraSAR-X image captured after the Wenchuan Earthquake, China was utilized to estimate the damages to the buildings at the city block level. The objective of the study was to analyse the damaged building characteristics in SAR images considering its direction, geometric features, and the context of the building. Since loss in the building elevation can lead to changes in the building's linear characteristics, such information captured from SAR images can be used as a proxy for different levels of damage. Furthermore, strong scattering from the building is considered a damage cue. In this study, the block samples including different damage levels were selected using ground survey and aerial images over the study area, and four different texture features were calculated for each building block. Using the obtained texture information, fuzzy classification was implemented in the next step to extract the intact buildings because intact building blocks show more consistency in the texture characteristic in contrast to collapsed buildings with a variety in texture values. The rest of the building blocks were then divided into two classes moderate damage and severely damaged blocks. However, getting a pair of SAR images captured before and after the earthquake is not easy in real-world scenarios. The standing buildings usually have low backscattering in SAR images and appear dark compared to the debris and collapsed buildings. Therefore, in cases where the building footprint is available, single-post SAR images can facilitate the damage assessment (Bai et al., 2018).

2.3.3. LiDAR Images

Airborne LiDAR (Light Detection and Ranging) sensors are another less-common type of data being used in detecting the damage to buildings. Due to the ability of LiDAR sensors in providing information about the height change and geometry of the building this type of data is used in detecting multilayer and pancake collapses which is usually difficult to detect using nadir-looking optical images (Schweier & Markus, 2006). There are a limited number of studies investigating automatic building damage detection using LiDAR data

only (Janalipour & Mohammadzadeh, 2015; Wen et al., 2014). Such data type is usually used in the absence of optical data after an earthquake event or as a piece of additional information to enhance the performance of the building damage classification by generating a more accurate digital surface model (DSM) (Bittner & Korner, 2018; Pollino et al., 2020).

2.4. Multi-temporal images and single-image

There have been numerous studies proposing different methods for extracting information from remote sensing images to assess earthquake-induced damages to buildings. These methods can be divided into multi-temporal or single-post-image assessments. Figure 2.3. indicates the number of studies that have used single post-event images and pre and post-event remote sensing images for building damage mapping.

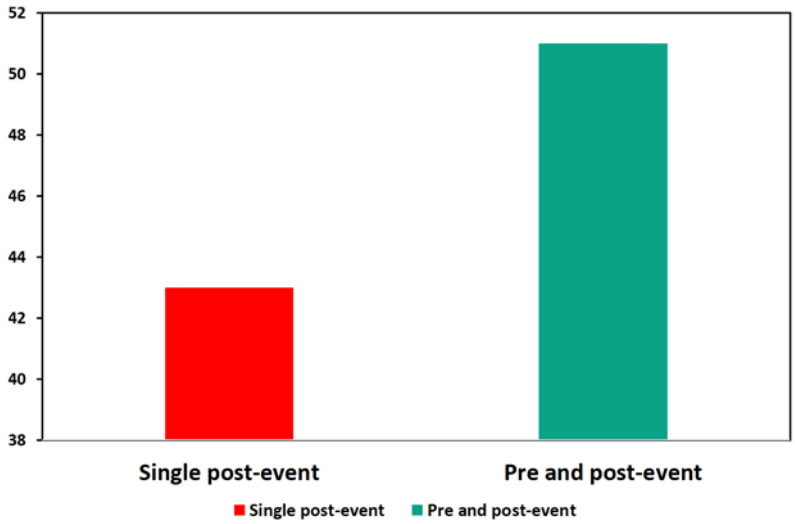

Figure 2.3. The number of earthquake-induced building damage detection studies using a single post-event image or a pair of images captured before and after the earthquake, between the years 2010 to 2020.

The multi-temporal assessment methods traditionally evaluated the damage by considering meaningful changes in the relevant image features such as texture, intensity changes, etc. based on a pre-defined threshold set by experts in the field or according to the ground truth (Adriano et al., 2019). Change detection methods can be performed using object-oriented approaches, pixel-wise, and principal component analysis approaches (Ma, Liu, Ren, & Yu, 2020). In another work, (Gong et al., 2013) employed different change detection techniques to extract the building damages following the 2010 Yushu earthquake using very high-resolution multi-temporal images. Their results indicated that object-oriented damage detection outperforms the other two methods. (Hedjam et al., 2019) proposed a

semi-supervised method to detect the changes after the disasters by automatically generating a training dataset using a pair of non-labeled images captured before and after the earthquake. In this method, a sliding window scan the images to detect the genuine pairs (the patches not affected by the event), and impostor pairs (the patches affected by the event). To automatically generate the labeled patches from the images, a deep learning architecture consisting of two identical convolutional neural networks were used. This particular architecture is called the Siamese network (Bromley et al., 1993). The proposed method was tested for five pairs of real event images to generate a difference- image by localizing the genuine and impostor pairs. Next, the threshold was determined to produce a binary map where the bright areas depict the change after the event. This method detects the actual changes more accurately, and it is less sensitive to unwanted changes compared to similar studies which consider pixel-by-pixel differences. In this technique, the focus is more on labeling the training data. The method is more suitable for geometric change rather than spectral changes. However, the natural changes caused by the differences in weather and image acquisition platform can negatively affect the performance of the method.

Despite the successful performance of these methods, they cannot be generally applied to another study area in real-world scenarios as there is a need for the availability of both pre and post-earthquake, the threshold used in these methods is often site-specific and the accuracy of the method is sensitive to co-registration (Gokon et al., 2015; Karimzadeh & Matsuoka, 2018; Pozos-Estrada et al., 2019). Therefore, the methodologies using only the single post-event images have been developed to classify the damages. These methods are generally based on the statistical features extracted from the objects in the image, such as texture (Gong et al., 2016) and morphology (Vu, 2012). However, some studies detected collapsed buildings from patches of a single post-event image using the deep learning method (Bittner & Korner, 2018).

After exploring the title and the abstract of the published papers, the most used data type was optical-based remote sensing images, used individually or in combination with SAR or LiDAR. Figure 2.4. presents the result of the initial analysis of the above-mentioned papers based on the year they have been published and the method they employed. The methods include deep learning (DL), machine learning (ML), and other methods. The other methods include non-machine learning methods such as feature-based image analysis, image processing techniques, statistical analysis, etc.

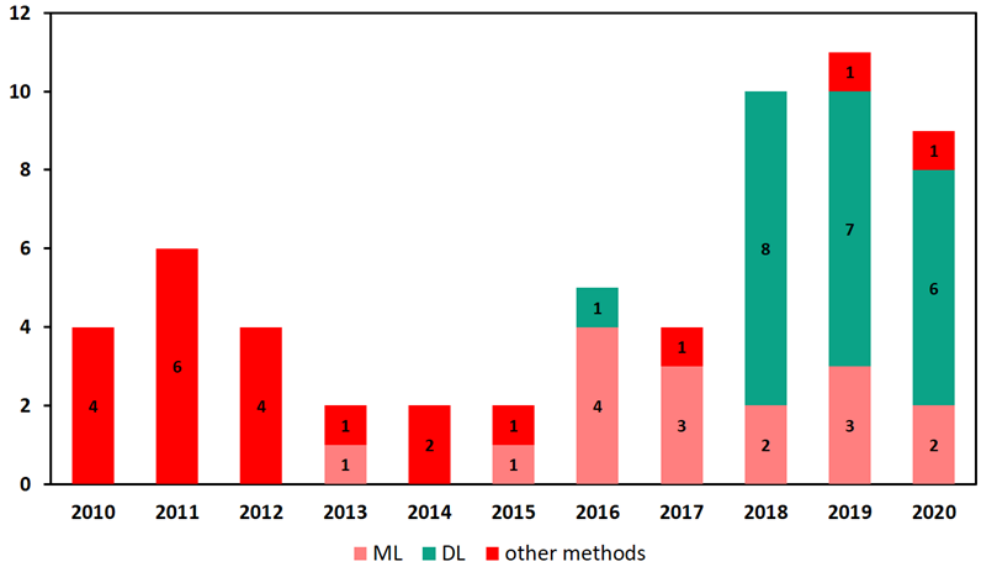

Figure 2.4. The number of earthquake-induced building damage detection studies using optical, optical and SAR, optical and LiDAR, optical and SAR, and LiDAR remote sensing images and different classification methods per year.

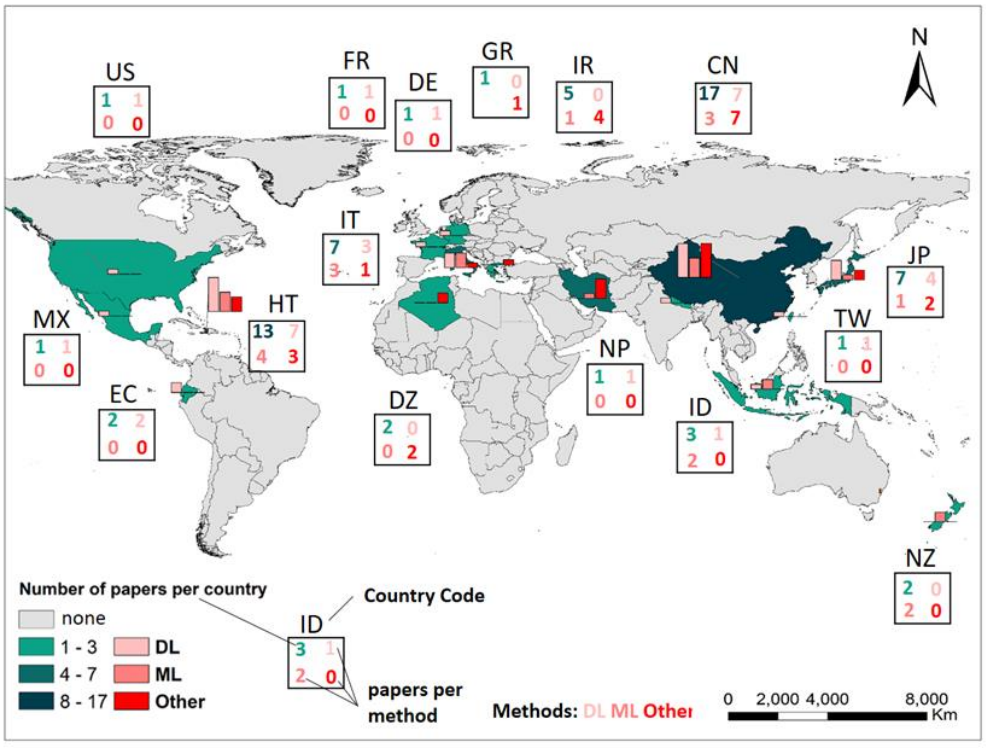

Figure 2.5. Geographic distribution of earthquake-induced building damage detection studies using optical remote sensing images. The country's green shade colour represents the number of published papers, in three classes. The exact number of published papers considering a study area in each country and the number of studies using deep learning (DL), machine learning (ML), and other (O) methods are given. Countries showed based on ISO 3166-1 alpha-2 code. It should be noted that the images from some study areas were simultaneously utilised in the studies.

The above-mentioned studies used remote sensing images after different earthquake events around the world and proposed various classification methods (Figure 2.5.).

2.5. Earthquake-induced Building Damage level estimation

Although accurate damage to individual buildings can only be identified by teams of experts through the ground inspection where the damages to both internal structural elements of the buildings (e.g.; columns, beams) and external elements (e.g., façade, roof) are considered, such detailed damage assessment is time-consuming, resource-intensive and can be even dangerous as it requires ground inspection in-person (Gupta et al., 2019). Furthermore, in the case of an earthquake in vast urban areas, a building-to-building inspection of the whole affected area is not possible in a short time, which is against the requirement of stakeholders for mitigation and rescue operations (Dong & Shan, 2013). Remote sensing images due to their synoptic view can be a precious source of information by delivering images from the affected area used during the response and recovery phase for rescue missions, resource allocation and aid routing, or even later for the reconstruction phase (Gupta et al., 2019; Vetrivel et al., 2018) .

Although tools such as The United States Geological Survey (USGS) Shakemaps provide essential information about the potential damage intensity of the earthquake in the affected region, it is not representative of what has happened there. To have a clear understanding of the damage in the region, it is of great importance to consider different levels of damage to individual buildings in the damage map (Monfort et al., 2019). To assess different levels of damage, regardless of the method (i.e., manual or automatic damage detection using remote sensing images or ground inspection), first, these levels should be defined and described clearly to develop a common understanding of each level of damage for emergency management services, decision-makers, rescuers and researchers.

Several damage-level- classifications have been studied and proposed for different types of structures (Cotrufo et al., 2018; Grünthal, 1998; Gupta et al., 2019). One of the earliest and most popular damage-classification guidelines is European Macroseismic Scale 1998 (EMS98) which quantifies building damages by describing five commonly used damage grades including negligible to slight damage (Grade1), moderate damage (Grade2), substantial to heavy damage (Grade3), very heavy damage (Grade 4) and destruction (Grade5) (Grünthal, 1998).

EMS98 guideline is mainly designed for ground inspection, where the damages to the building are assessed in person. However, characteristics of the building in remote sensing images are different from those observed in the ground inspection as in nadir-looking satellite images only the roof of the buildings can be observed (Cotrufo et al., 2018). To overcome the limitation of the vertical imagery where the façade and walls cannot be

assessed, several damage detection techniques have been suggested. For instance detecting the debris, rubbles, and spalling surrounding the building as a proxy to detect the damages to the buildings (Ghaffarian & Kerle, 2019). However, detecting these cues requires the post-event image to be captured in a short time after the earthquake when the rubbles were not removed yet. Moreover, when using remote-sensing images, the spatial resolution of images plays a key role in classifying earthquake-induced building damages. For example, while it is only possible to detect basic damage at a block level with a resolution of 10 meters and coarser, precise building damage grading can be conducted at a building level as the resolution reaches 1 meter (Dong & Shan, 2013). One study claimed that the damage level differences can be visually detected when the resolution of the image is below 0.8m (Gupta et al., 2019). Further resolution improvement in such images anywhere below 0.3 meters can facilitate the detection of intermediate earthquake-induced damages at a building level (Cotrufo et al., 2018).

Generally, by increasing the access to higher resolution remote sensing images, the classification of the lower damage level per building (grade 1 to grade 3) would be more likely to be possible. For example, (Fernandez Galarreta et al., 2015) attempted to create a methodology based on object-based image analysis (OBIA) to assess the intermediate level of damage using VHR UAV multi-view oblique imagery. Taking advantage of the high resolution of the image along with the image angle, the authors assessed damage features in the façade and roof of individual buildings and detected some cracks and holes, assigning the damage level to each building was challenging though. Identifying very-heavy damage (grade 4) is difficult to be detected as it is usually described as heavy structural damage such as failure of the walls, which cannot be seen using nadir-looking remote sensing images (Corbane et al., 2011). To distinguish very heavy damages from collapsed buildings, images from before and after the earthquake are required. Utilizing oblique imagery to detect the damage on the walls might make detecting very heavy damages to the buildings possible (Gerke & Kerle, 2011).

Although high-resolution images allow researchers to develop new methods to more accurately classify the damages to the buildings, in some research work the common guidelines have not been followed for damage classification and the authors described the damage level based on the other criteria and their perception of damage level. For example, exploiting the ancillary data such as building height before the earthquake, (Tu et al., 2016) classified the damage to the building rooftops into four classes trivial damaged, severely damaged, and totally collapsed, and no damage after the Wenchuan earthquake in China. In another study, the researchers used image processing techniques to classify the damages

in the urban area (Korkmaz & Abualkibash, 2018). In this study, four classes of no damages, slight damages, Medium damages, and severe damages were described according to the percentage of collapsed buildings in the area. Such a creative choice of definition for each damage level is further hindering the comparison between different studies. An example of the creative damage level description for a block of buildings is shown in Table 2.1.

Table 2.1. The proposed classification of the damaged block of buildings in post-earthquake images (Ma, Liu, Ren, Wang, et al., 2020).

Category	**Character description**	**Collapse rate**
Serious damage	All destroyed or most Collapsed	>70%
Moderate damage	About half collapsed	30-70%
Slight damage	Generally intact or a small part collapsed	<30%

With the emergence of high-resolution UAV and satellite images classifying the building damages into different damage, levels are easier, thus necessitating the need for an internationally accepted guideline tailored for remote sensing images. In 2017, (Cotrufo et al., 2018) introduced a novel damage level guideline which is recently followed by emergency management services such as Copernicus.

(Gupta et al., 2019) presented xBD dataset which contains 850,736 labeled buildings using pre and post-event satellite images across various natural hazards from 15 countries. Along with introducing the xBD dataset, a novel building damage scale - The Joint Damage Scale was presented. The Joint Damage Scale was inspired by several building damage guidelines available for each hazard type, including EMS-98 for earthquake-induced building damage assessment. Considering the coverage and the large size of the xBD dataset. It is expected that this database will be used in many future studies to develop novel methodologies or assess the performance of different machine learning methods and also to practice transfer learning in the domain of building damage detection (Kerle et al., 2020). Therefore, although the Joint Damage Scale is not an authoritative damage assessment guideline (Gupta et al., 2019), future studies may consider using this scale as a unified damage guideline for building damage assessment using satellite imagery. Table 2.2 shows the comparison of the available common damage level guidelines.

Table 2.2. Comparison of the EMS-98, Copernicus EMS classes, and the joint damage

European Microseismic Scale 1998 (EMS98)	**Copernicus EMS classes**	**The Joint damage scale**
Grade 0 No damage (absence of damage)	**No visible damage**	**No damage** Undisturbed. No sign of water, structural damage, shingle damage, or burn marks.
	Possibly damaged Uncertain interpretation due to image quality Presence of possible damage proxies. Building surrounded by damaged/destroyed buildings	
Grade 1 Negligible to slight damage (no structural damage, slight non-structural damage)	**Damaged** Minor: The roof remains largely intact, but presents partial damage	**Minor damage** Building Partially burnt, the water surrounding the structure, volcanic flow nearby, roof elements missing, or visible cracks.
Grade 2 Moderate damage (slight structural damage, moderate non-structural damage)	Major: Partial collapse of the roof; serious failure of walls	
Grade 3 Substantial to heavy damage (moderate structural damage, heavy non-structural damage)		**Major damage** Partial wall or roof collapse, encroaching volcanic flow, or the structure is surrounded by water or mud.
Grade 4 Very heavy damage (heavy structural damage, very heavy non-structural damage)	**Destroyed** Total collapse, the collapse of part of the building (>50%); Building structure not distinguishable (the walls have been destroyed or collapsed)	**Destroyed** The structure is scorched, completely collapsed, partially or completely covered with water or mud, or no longer present
Grade 5 Destruction (very heavy structural damage)		

2.6. Methods employed for earthquake-induced building damage assessment

Various methods have been proposed to assess the earthquake-induced building damages using optical remote sensing images. These methods can be categorized into three groups of feature-based, conventional machine learning, and deep learning classification methods.

2.6.1. Image features

There have been a significant number of studies proposing different automatic and semi-automatic frameworks using a variety of image types (optical, Synthetic Aperture Radar – SAR, and Light Detection and Ranging - LiDAR) for earthquake-damage assessment (Dell'Acqua & Gamba, 2012; Kerle et al., 2019). To understand which features, contribute to detecting the damaged building, researchers examined the use of different types of features and the combination of them in assessing the damages after the earthquake. These features vary across different studies in the literature, depending on input image type, using multi-source or multi-resolution images, availability of the ancillary data (e.g., Digital Elevation Model - DEM), and the capability of the method in handling a multi-dimensional dataset. Apart from the type of the features, the number of features is also different across the studies and ranges from four (Syifa et al., 2019) to about forty (Adriano et al., 2019).

One of the popular approaches to assessing the damages to the buildings is calculating the changes in features extracted from a pair of images captured before and after the earthquake, which is called change detection (Dong & Shan, 2013). These features include but are not limited to spectral, textural, edge, spatial relationship, and geometry information (Adriano et al., 2019). After calculating the value of the changes in feature values, a threshold is set by the experts or according to the ground truth to identify the meaningful change and hence identify the damaged area. Taking this approach, researchers (Table 2.3) used different features to detect damaged or collapsed buildings. For example, (Gamba & Casciati, 1998) proposed an automatic change detection method that could detect about 70% of collapsed buildings based on the shape analysis. (Yusuf et al., 2001) compared the brightness value of the pre-and post-earthquake Landsat-7 images to detect the dam-aged areas. (Sugiyama & Abe, 2002) considered the differences in color of the pair of co-registered images to detect the damaged area. (Zhang et al., 2002) used the mean and variance of the intensity of grayscale pre-and post-event images to detect the damage grade of the building blocks. (Rathje et al., 2005) used a combination of texture and a correlation coefficient to detect the changes after the earthquake, whereas (Pesaresi et al., 2007) used radiometric and morphological features to detect the collapsed buildings. (Turker &

Cetinkaya, 2005) used the temporal DEMs generated from the pre and post-event aerial images to detect the collapsed buildings. (Rezaeian & Gruen, 2007) and (Rezaeian, 2010) used the combination of the digital surface model (DSM), pixel density, and segment shape to identify the level of damage to the building.

Studies show that using the information obtained from SAR or LiDAR images combined with the features extracted from optical images can improve the damage assessment (Rehor et al., 2008; Stramondo et al., 2006). Height information extracted from LiDAR image and spectral and textural features from optical post-event image has been used by (Ural et al., 2011) and (Haiyang et al., 2010).

As the cloud-free pre-event optical image might not always be available, using change detection techniques is limited. Therefore, with the emergence of very high-resolution remote sensing images such as Worldview 3 and TerraSAR-X extracting the relevant features using only post-event images is an alternative solution. However, a few studies used only post-earthquake satellite images from an urban complex scene to assess the damages to the buildings. Some of these studies have been summarized in Table 2.3.

Table 2.3. Studies that have used different feature descriptors, methods, and image types.

Main Findings	Features	References
Detected the collapsed buildings	Shape Radiometric and morphological DSM radiometric and morphological Colour, shape, texture, height	(Gamba & Casciati, 1998) (Pesaresi et al., 2007) (Turker & Cetinkaya, 2005) (Haiyang et al., 2010)
Classified buildings as damaged and non-damaged	Spectral, textural and spatial relations variance and the direction of edge intensity, texture Spectral, textural, Structural Textural and Colour Spectral and textural Spectral, textural and height geometric	(S. Li et al., 2011) (Mitomi et al., 2002) (Cooner et al., 2016) (Rasika et al., 2006) (Rathje et al., 2005) (Ural et al., 2011) (Wang & Jin, 2011)
Classified buildings as damaged, destroyed, possibly damaged, no damage	Spectral, Intensity, coherence, DEM	(Adriano et al., 2019)
Classified the image into debris, non-debris, no change	Spectral, textural, brightness, shape	(Khodaverdizahraee et al., 2020)
classified buildings as collapsed, partially collapsed, and undamaged	Normalised DSM, pixel intensity and segment shape	(Rezaeian & Gruen, 2007) (Rezaeian, 2010)
Classified buildings as non-damaged or slightly damaged, low level of damage	SAR texture	(Dell'Acqua & Gamba, 2012)
Identified damaged areas	Spectral	(Syifa et al., 2019)

Considering the variety of the image features used in assessing the damages to the building, it is crucial to understand what features contribute to the detection of the damages. Identifying the important features involved in the analysis has long been the researchers'

interest. For instance, to understand which feature leads to the best damage assessment result, (Tomowski et al., 2010) compared the overall accuracy of four change detection scenarios using different texture features (contrast, correlation, in-verse distance moment, and energy) and concluded that utilizing principal component analysis (PCA) with energy achieves the best result. However, using PCA to reduce the dimensionality, eliminates the interpretability of the features as PCA features are the combination of the actual features and hence is not interpretable.

(Mansouri & Hamednia, 2015) employed a genetic algorithm (GA) optimiser to select the optimum features among 8 texture features to classify the building damages into three classes of no to slight, moderate, and heavy to destruction. Three features of "mean", "dissimilarity" and "second moment" were selected by GA optimiser and subsequently used the support vector machine (SVM) classification method. In an attempt to select the most impactful features that contribute to distinguishing the changed and unchanged areas, (Khodaverdizahraee et al., 2020) utilized the GA optimiser. All the features were extracted from a pair of optical satellite images captured before and after the Bam earthquake. The initial features included 29 texture and spectral features from which only 9 features have been selected after performing GA. However, performing a genetic algorithm is computationally expensive, as the model should be trained for each feature (Javadi et al., 2005).

Employing multi-source multi-resolution images can slightly increase building damage detection accuracy. However, exploring the effect of different datasets and the related features involved in the analysis becomes crucial when the dataset dimension increases due to utilising data fusion. (Adriano et al., 2019) designed a comparison study in which different scenarios were considered to compare the classification results in the presence or absence of different datasets. A total of eight different scenarios were defined using different combinations of pre and post-event datasets. In their study, the dataset included 11 images (five pre-event images and six post-event images) acquired from different sensors including ALOS-2 PALSAR-2, Sentinel-1, and Sentinel-2, and Planet Scope. The DEM and the building footprint data were considered ancillary data to contribute to generating the damage map. Three machine learning classifiers, including Random Forest (RF), Rotation Forest (RoF), and Canonical Correlation Forest (CCF) were used for pixel-based classification and to generate a damage map. The result from this study showed improvement in accuracy as more datasets from both SAR and optical images were added to the learning process. As a result of this change, the maximum accuracy was achieved for all three classifiers in the scenario in which all the datasets were utilized. Regarding the

importance of the image features in classification, the result from computing variable importance measurement (VIM) indicated that coherence information derived from multi-temporal SAR images contributed slightly more than the optical-derived features. Further, the VIM results showed that DEM was found to be the most important feature used in the classification model. Although DEM can be an earthquake risk factor for specific areas with significant variation in the elevation in Tsunami-prone areas (Jena et al., 2020), it cannot be generally transferred to all geographic areas.

Although using both pre-and post-event images increases the classification accuracy, the use of only post-event images avoids pre-processing works like co-registration and histogram matching (Khodaverdizahraee et al., 2020; Moya et al., 2020). Moreover, remote sensing images contain a huge amount of information that can be extracted and used for different purposes. Including a large number of features leads to a large dimension dataset. Handling such a dataset is time-consuming and computationally expensive. Furthermore, the redundant features may add a layer of complexity to the model and thus de-creasing the accuracy (Duangsoithong & Windeatt, 2009). Therefore, identifying the optimum features as well as re-moving the redundant ones is a necessary pre-processing step to decrease the dimensionality of the data without a decrease in model performance (Bolon-Canedo & Remeseiro, 2019).

Therefore, the main objective of this study is to address the limitation described above and construct an explainable model by i) using a single post-event optical image for detecting the collapsed buildings, ii) selecting the best image features based on the model output, and iii) investigating the performance of multilayer perceptron (MLP) in distinguishing the collapsed and non-collapsed buildings.

2.6.2. Feature-based classification methods

Preliminary studies in the domain of automatic building damage detection were based on changes in spectral and texture information between the images captured before and after the earthquake (Tomowski et al., 2010). The change detection method has been explained earlier in from the data perspective, however in this section different approaches of the change detection and the selection of various image features used in both the change detection method and single post-event image analysis will be discussed.

Change detection methods can be performed by either pixel- or object-based approaches. In the pixel-wise approach, each pixel of the image is considered an independent object. The pixel-wise approach is adaptable to detect the damaged areas in the study area with

simple terrain. However, it usually leads to noisy results in a complex train due to the high spatial distribution of various spectral values. Moreover, setting an appropriate threshold to identify the damaged area is challenging (Fei et al., 2015). Whereas, in object-based image analysis (OBIA) a meaningful group of pixels in the image is considered as an object in the image (Blaschke, 2005). Lots of studies reported satisfactory results, taking an object-based image classification approach in the domain of building damage assessment (S. Li et al., 2011). Taking the object-based image analysis approach, texture and geometric information of the objects of the image can be leveraged. Among different texture features (i.e. Haralick, Gabor, Fractal, and first-order statistical features (Ranjbar et al., 2014)), Haralick texture features, calculated from gray level co-occurrence matrix (GLCM) or gray-level difference vector (GLDV) have been extensively used in the literature.

Using GLCM texture features, (Miura et al., 2011) developed an automatic building damage detection to distinguish the collapsed and non-collapsed buildings after Haiti earthquake using the pre and post-event images. The result indicated that the dissimilarity of the collapsed buildings is larger than non-collapsed buildings. (X. Li et al., 2011) used the texture and spectral information to classify a post-earthquake areal image with the spatial resolution of 0.5m. The result showed that the object-based classification proposed in this study can distinguish the damaged and non-damaged buildings with about 90% accuracy.

Other studies attempted to use the shape and texture of the objects in the image to detect damage or collapsed buildings (X. Li et al., 2011; Miura et al., 2011). For example, following the 2008 Sichuan earthquake, (Vu, 2012) used spectral and texture information from QuickBird satellite images to detect the debris and rubble using the K-means clustering technique. The study defined ‘damage ratio’ as the number of damaged buildings by the total number of buildings, which was used to classify damage levels as heavy, moderate, and light, and indicated an accuracy of 84%. After the 2010 Haiti earthquake, (Dubois & Lepage, 2013) used a semi-automatic approach to apply object-based classification by comparing the texture and the shape information of the buildings from a pair of QuickBird satellite images. The study achieved an overall accuracy of 84% in classifying the level of the damage into three classes no damage, heavy damage, and destruction.

2.6.3. Conventional machine learning based methods

Various studies leverage the machine learning classifiers to generate a damage map using the image features mentioned in the previous section. Following this approach generally,

first, numbers of image features are manually extracted by the user, then the classifier model classifies the image based on the relationship between those features and the target class (Chauhan & Singh, 2018). In another commonly used approach, arbitrary features can also be extracted using deep learning networks as a feature extractor. Next, those extracted features can be fed into a machine learning classifier (Ji et al., 2019). Figure 2.6. shows the workflow involved in earthquake-induced building damage mapping using the machine learning method.

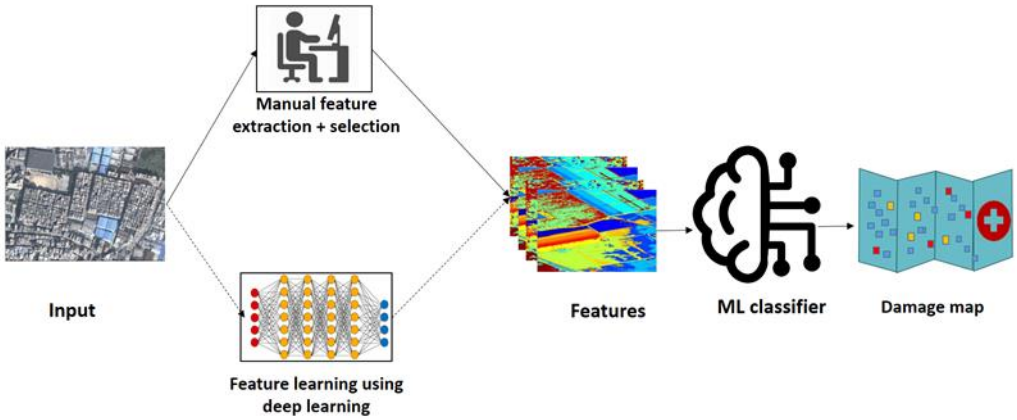

Figure 2.6. General schematic representation of the workflow involved in earthquake-induced building damage mapping using the machine learning method. The solid line indicates the workflow that the majority of the papers followed. However, a few works used the features learned using deep learning (the dashed line). Either one of the paths or both can be followed to generate a damage map.

Taking this approach, a combination of different image features and several machine learning models were reported in the literature (Abdollahi, Pradhan, et al., 2020b; Chauhan & Singh, 2018; Miura et al., 2011). Followed by the 2018 earthquake in Central Sulawesi Province in Indonesia, (Syifa et al., 2019) used pre and post-earthquake satellite images from Landsat-8 and Sentinel-2 satellites to generate damage maps delineating the assessment of damage in the affected area. For comparison purposes, two different ML models, SVM and ANN were used for satellite image classification. The outcome was four damage maps based on the image and the method being used. Comparing the generated damaged maps with the field data map, indicated that both SVM and ANN coupled with the decorrelation method performed well (85.83% for Landsat-8) in classifying satellite images. However, large areas of sentinel-2 images were covered by clouds which is a hinder in optical satellite-based classification.

(Cooner et al., 2016) utilised two texture features of Entropy and dissimilarity from GLCM along with structural features (Laplacian of Gaussian (LoG) and rectangular fit) as additional texture information to enhance the classification. The authors applied 3 different

ML models, namely, feedforward neural networks (FNN), radial basis function neural networks (RBFNN) and RF for damage classification into damage and non-damaged buildings post 2010 Haiti earthquake event. The classification result for each method was compared with the united nation (UN) damage validation points indicating that despite the subtle difference in overall accuracy (74.14%,77.26%,76.14% for ANN, RBFNN, and RF respectively) both ANN outperform RF in detecting the damages. Moreover, the variable important measure (VIM) indicated that rectangular fit contributed more to image classification using ANN and RF compared to the rest of the variables used in this study. The results indicated RBFNN to be performing the best.

(Adriano et al., 2019) employed multi-temporal SAR and Optical satellite images and ensemble learning techniques to map the damages to the buildings caused by the 2018 Indonesia earthquake. In this study, the dataset included 11 images (five pre-event images and six post-event images) acquired from different sensors including ALOS-2 PALSAR-2, Sentinel-1, and Sentinel-2 and Planet Scope. The digital elevation model (DEM) and the building footprint data were considered ancillary data to contribute to detecting the collapsed, damaged, possibly damaged, and intact buildings. In this study, three machine learning classifiers, including random forest (RF), rotation forest (RoF), and canonical correlation forest (CCF) were used for pixel-based classification and to generate a damage map.

To investigate the contribution of each dataset in the classification, 8 different scenarios were defined based on using different combinations of pre and post-event datasets. The result from this study showed improvement in accuracy as more datasets from both SAR and optical images are added to the learning process, resulting in the maximum accuracy being achieved for all three classifiers in the scenario in which all the datasets were utilized. Regarding the importance of the image features contributing to the classification, the result from computing variable importance measurement (VIM) indicated that coherence information derived from multi-temporal SAR images is slightly contributing more than optical-derived features. Further, (Mansouri & Hamednia, 2015) employed a Genetic Algorithm (GA) to find the optimum features for classifying the damages to the buildings. The authors proposed a framework based on a soft computing algorithm to detect the damage level to the buildings after the 2003 Bam earthquake in Iran. A pair of panchromatic images captured before and after the earthquake were used to identify the three classes of no to slight, moderate, and heavy destruction. In this study first, the difference map was generated by subtracting the images. Next, the GLCM features were extracted from the difference map. In the next stage, GA was used to select the optimum

features. The selected features in this stage were used as the input for the SVM method to classify the image into three classes of low change, moderate change, and severe change. Leveraging the available building footprint data, the damage level label was assigned to a random subset of the buildings. This labeled dataset was used as an input for the Fuzzy Inference System (FIS) to classify the buildings in the area. The overall accuracy of the proposed method was 76%, thereafter, the accuracy increased to 86% for binary classification of the buildings.

To explore the performance of CNN features and hand-crafted features contributing to building damage detection, (Dubois & Lepage, 2013) designed a comparative study using pre and post-event satellite images from the 2010 Haiti earthquake and a Random Forest (RF) as a classification method. In this study, GLCM texture features were extracted for each building in both pre and post-event data. In the next step, random forest (RF) was adopted to classify the image into collapsed and non-collapsed buildings using the features extracted from CNN and the hand-crafted texture features. The final result showed that the combination of the features from CNN and the random forest method outperforms the other methods in distinguishing the collapsed and non-collapsed buildings by achieving the highest accuracy (OA = 87.6%). However, this method was tested using the image from a particular geographical region, the same should be applied to other regions and also using oblique images to obtain more information about the texture features of the damaged and collapsed buildings.

The conventional techniques of determining the optimum image features have been restricted to statistical techniques like the Gini score (S. Li et al., 2011) or ablation study (Cotrufo et al., 2018). However, the recent development of SHapley Additive exPlanations (SHAP) by (Lundberg & Lee, 2017), has shown promising results for feature selection in the fields of medical imagery. The use of SHAP could prove to be immensely helpful for optimum feature selection in the field of earthquake building damage detection and future researchers are encouraged to test the algorithm. To the best of our knowledge, this technique is yet to be applied in this domain. Also, the SHAP algorithm has the capability to address the detailed interaction among the features and their impact on the result (Lundberg & Lee, 2017).

2.6.4. Deep learning methods

In recent years, deep learning methods as state-of-the-art techniques in the field of machine learning are recognized as one of the most popular methods in extracting information from remote sensing images (Abdollahi, Pradhan, et al., 2020a; Abdollahi, Pradhan, Shukla, et

al., 2020; Dikshit et al., 2020). These methods are based on neural networks. The superiority of the deep learning method over conventional machine learning methods is that it learns low to high-level features from the raw data directly by applying different nonlinear filters on the image and therefore decreases the user intervention in choosing the arbitrary image features (Abdollahi, Pradhan, Shukla, et al., 2020). Three common approaches of detection, segmentation, and classification were reported in the literature to detect the earthquake-induced damages in the affected area using optical remote sensing data. Figure 2.7. shows the workflow for earthquake-induced building damage mapping using deep learning methods.

To train a segmentation model, each pixel of the image is labeled according to the damage level guidelines discussed in section 3.2. Whereas in the training phase for the classification approach, each patch of the image is labeled based on the number of damaged buildings in the image patch. More details about the detection, segmentation, and classification approach can be found in (Sublime & Kalinicheva, 2019) and (Ma, Liu, Ren, Wang, et al., 2020) respectively. In the following para, the summary of the studies using the deep learning approach will be explored.

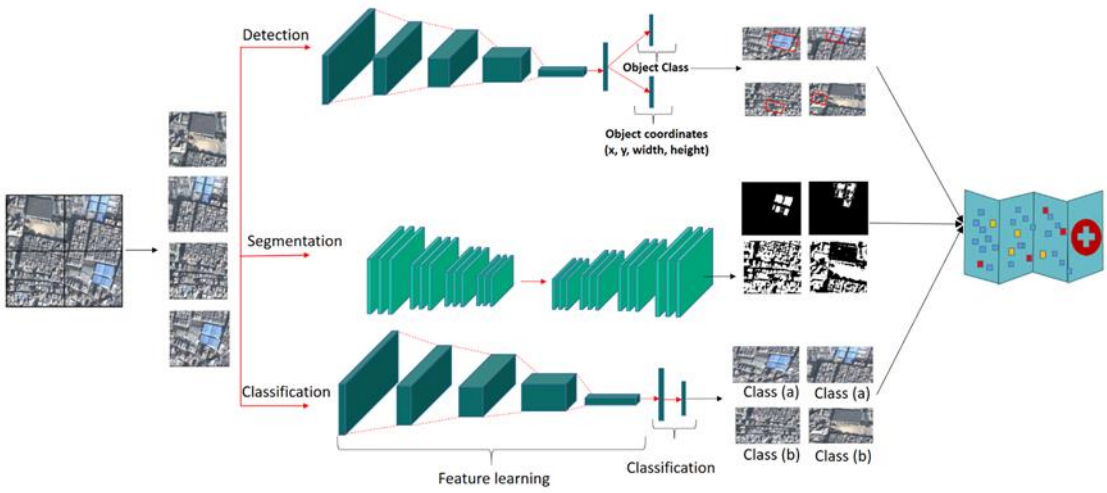

Figure 2.7. General schematic representation of the workflow for earthquake-induced building damage mapping using deep learning methods. Depending on the type of the available label (mask or the patch class) for training the model, the result can be the segmented image or classified patches of the image.

In an attempt in producing a near-real-time damage map taking the classification approach, (Nex et al., 2019) proposed a deep learning-based new solution using low-cost commercial UAVs to fulfill the first responder requirements during the response phase of the earthquake management cycle. In this study, the shallow ResNet architecture was used to

detect the damages to the building. The presented network in this study has achieved an overall accuracy of 92.1%. It should be mentioned that morphological filters, and vegetation indexes are used to identify the potential damage areas in each image to reduce the deep learning computational time. Despite the high accuracy of the proposed method, there are limitations in using the proposed framework. Firstly, legal restrictions on the deployment of the UAVs weren't considered in the study. Secondly, the short flight duration (maximum 25 minutes) and the narrow coverage of the UAV make this method limited compared to using satellite images in operational building damage mapping.

(Ma, Liu, Ren, & Yu, 2020) improved a CNN-based model (called YOLOv3) to detect the collapsed building after 7.1Mw Yushu and Wenchuan earthquakes using post-earthquake aerial images. YOLO algorithms have been developed in three versions and have shown rapid and high accuracy performance in detection tasks. In this study, YOLOv3 architecture and the loss function were improved and its performance was compared to the original YOLOv3 using average precision (AP) and precision-recall curve (PRC) as evaluation metrics. The result from this study showed a 5% increase in precision (93%), 10% (88%) in recall rate, and 5.05% in average precision (90.89%). Besides, the improved version performed faster compared to the original version based on FPS (Frames per second) metric to evaluate the detection speed. Although the improvement in the results and using only post-event images are considered the superiority of the proposed method, the limitation in the number of training samples may lead to errors in real-world scenarios. (Song et al., 2020) proposed a classification framework to identify the degree of damage using post-event aerial images captured after the 2008 Wenchuan earthquake. The framework integrated Deeplab v2 neural network, the simple linear iterative cluster (SLIC) method for super-pixel segmentation and mathematical morphology. The evaluation of the method indicated more than 90% of overall accuracy for over eight different test areas from the region. However, this method didn't consider the limitation in the number of the labeled dataset in real-world scenarios.

(Xu et al., 2020) used a method based on a pre-trained Single Shot MultiBox Detector (SSD) deep learning method to detect the land cover while detecting the ruins and houses from the post-earthquake google image captured in the Yunnan province of China. Considering the lack of the post-earthquake labeled training dataset, the authors used augmentation and balancing techniques to tackle the imbalanced data and simultaneously enhance the accuracy of the detection. Moreover, they pre-trained the model using the PASCAL VOC dataset to enhance the classification accuracy while using the small size training dataset to fine-tune the network parameters. The resulted F1 score even improved

to 87.88% and 89.02% after utilizing a data balancing technique. Finally, the optimized network was developed based on previous results. Although the results are encouraging, the authors didn't provide a damage map of the entire region. Moreover, considering the sparse locations of the houses in the study area, the performance of the proposed method should be examined in a more complex urban area. To distinguish the intact and collapsed buildings in a study area with densely distributed buildings, (Aali et al., 2019) proposed a multi-task deep learning method to detect the buildings in good shape (main task) and to detect the other objects such as roads, trees, etc. (supplementary task). The architecture of the proposed network includes a shared layer based on the PSPNET model followed by the feature map to train two segmentations separately for each task (Zhao et al., 2017). In this study, 34 image samples were used to train the model and 10 other images were selected in the test dataset. The result of the method indicated the F-1 score of 82.73% and 93.65% for damaged and buildings in good shape, respectively.

(Shao et al., 2020) proposed a general optimized deep learning method that can be employed in automatic damage mapping after five different types of disasters including earthquakes, floods, hurricanes, volcanic eruptions, and wildfire. The proposed method is based on U-Net called building damage detection network (BDD-NET). In this study, the sample images are the subset of xBD dataset which includes pairs of pre and post-event images that covers different disasters around the world. The single post-event images can be used as an input of the proposed network however the accuracy of the classification increases using the pair of images from before and after the events. Utilizing the pair of images significantly affects the accuracy of the classification, particularly in the events such as floods where the building footprint is not clear in the post-event image. To optimize the network, several types of loss functions were evaluated, including weighted cross-entropy and the combination of that with dice loss or focal loss. It should be noted that using different loss functions doesn't affect the accuracy significantly when only a single post-event image is considered as an input as the image offers less semantic information. However, optimizing the network can positively affect the F1 score in the case of using a pair of images. The resulted F1 score utilizing the optimized network for classifying the images of the earthquake event was 86.5% when the pair of the images was used while this score is 36.6% when using a single post-earthquake image. This study reflects the necessity to use both pre and post-event images for achieving a more accurate damage classification. However, it's important that datasets like xBD are updated regularly, which would enable for damage mapping in the near or long run future.

Convolutional neural network (CNN) has shown satisfactory performance in classifying the post-earthquake remote sensing images. However, due to the fixed receptive field in the structure of CNN and the fact that the objects in post-earthquake images appear on different scales, the result may not be delineated accurately leading to coarse-resolution classification results where the edges of the objects are not clear. To overcome this issue, (Huang et al., 2019) designed a CNN-based architecture that takes advantage of multiscale training samples produced by the Multiple Sclerosis (MS) segmentation algorithm at three different scales. Images (spatial resolution of 0.67m) from three subareas after the Wenchuan earthquake in China were considered in this study and the target classes included intact buildings, broken buildings, forest, grasslands, debris flow, concrete, bare soil, road, landslides, dammed lake. The result of the classification was compared to SVM and conventional CNN. The classification result shows that the overall accuracy of the proposed method for all three study areas is significantly higher than that for SVM and conventional CNN, ranging between 92.49% to 96.88% which is about 10% and about 20% higher compared to SVM and CNN, respectively. (Frank et al., 2017) followed the same approach to detect the collapse and non-collapsed buildings using the aerial images captured after the Yunnan area in China and achieved an overall accuracy of 87%. Although the proposed method shows high overall accuracy, the effectiveness of this method is largely affected by the accuracy of the preliminary segmentation which limits the performance of this method in practice.

To improve the performance of the CNN method in identifying the damage level to the group of buildings, (Ma, Liu, Ren, Wang, et al., 2020) used the combination of GIS boundary data of the block buildings and a CNN-based method called Inception V3. The Inception V3 architecture consists of a global average pooling instead of the last fully connected layer in other types of CNN-based methods to make it perform faster and avoid overfitting. To evaluate the performance of this method, the aerial image (0.5 m) captured after the Yushu earthquake in China was collected to be classified into three classes of serious damage, moderate damage, and slight damage. The result indicated that the proposed method outperforms the conventional machine learning methods, achieving an overall accuracy of 90.07%. The number of samples was limited and there was confusion between the collapsed buildings and the bare land. The labels were based on visual interpretation and there is a possibility of wrong labeling of some samples where the labeling is not supported by the actual ground truth.

While most studies described in this section focused on proposing the new architecture to improve the segmentation, classification, and detection accuracy, few studies focused on

the localisation of the damages by generating a building damage map. A good example of the study that leveraged deep learning models with operational damage mapping is the study conducted by (Bai et al., 2018). In this study, the authors designed a U-Net-based framework to detect the "washed away", "survived" and "collapsed" buildings after the 2011 Tohoku Tsunami earthquake. They exploited the capability of U-Net architecture in achieving a considerably good result while using a small training dataset. They implemented the designed U-Net convolutional network on Microsoft's new GeoAI platform to fulfill the need of generating a rapid map within 2-15 minutes in the condition that the training dataset is available. Such mapping platforms could override the involvement of manual mapping using software like ArcGIS resulting in faster and more efficient rapid damage mapping.

2.7. Explainability, Transparency, Interpretability

Explainability provides reasons for the behaviour of a machine learning model to ML users. The reasons why a high level of explainability and transparency is required are diverse and vary from application to application (Adadi & Berrada, 2018). Depending on the context, in general, explainability can help with trust in the AI-based system decisions, accountability, creating opportunities for improving the model, and transferability of the model (Roscher et al., 2020). Although many conventional machine learning algorithms are easy to be explained or utilise features such as Variable Importance Measurements (VIM) to explain their decisions, explainability and interpretability can be an issue when it comes to using more sophisticated models. Developing an explainable model aids in identifying the most important features and observations in the dataset which helps in boosting the model performance (Janik et al., 2019).

Transparency is the first step to opening a Machine learning model's black box (Pedreschi et al., 2019). Transparency provides information about the model development (e.g., code and training and testing dataset) and the model algorithm (Bhatt et al., 2021). One of the examples of a non-transparent model is neural networks. However, these networks can be transparent once the structure is mathematically formulated. This can be achieved by releasing the model architecture and the hyper parameters that have been used in the model.

Interpretability provides information about how the input affects the model output. Feature importance is one of the examples of interpretability tools for ML models. It shows how each feature contributes to the model output and how strongly the model output changes in response to the changes in that feature.

We summarised the reviewed studies using optical images in Table 2.4. The table reports the information about the input image, the classification method, the spatial resolution of the image, and the level of the classification. It should be mentioned that some studies did not mention the spatial resolution of the input image.

Table 2.4. The summary of the studies using optical remote sensing images. For the studies that used more than one image with different scales, the range of the spatial resolution has been shown in the table.

Data Platform	Multi-temporal or Post-event	Classification Method	Classification Result	Spatial Resolution	Author(s)
Satellite	Pre- and post-event data	Deep Learning	Collapsed/not-collapsed	0.5m	(Ji et al., 2019)
			Damaged area	15m	(Sublime & Kalinicheva, 2019)
			Damaged/not-damaged	0.5m	(Shao et al., 2020)
				0.15m	(Aali et al., 2019)
			Collapsed/washed away/survived	0.6m	(Bai et al., 2018)
		Conventional	Damaged area		(Hedjam et al., 2019)
		Machine		30m	(Syifa et al., 2019)
		Learning	No to slight/moderate/heavy to destruction	0.6m	(Mansouri & Hamednia, 2015)
			Collapsed/not-collapsed	0.6m<<2.4m	(Anniballe et al., 2018)
			Damaged/not-damaged	0.5m<<2.4m	(Cooner et al., 2016)
		Texture	Collapsed/partly-collapsed		(Tiede et al., 2011)
		Analysis	Damaged/heavy damaged/destruction		(Korkmaz & Abualkibash, 2018)
	Post-event	Deep Learning	Damaged area	0.5m	(Bhangale et al., 2019)
			Intact/Ruins	0.3m	(Xu et al., 2020)

		Conventional Machine Learning	-	-	-
Aerial	Post-event	Deep Learning	Collapsed/not-collapsed	0.5m	(Ma, Liu, Ren, & Yu, 2020)
				0.3m	(Chen et al., 2018)
			Damaged/not-damaged		(Vetrivel et al., 2016)
			Intact/broken	0.67m	(Huang et al., 2019)
			Slight damage / moderate damage/serious damage	0.5m	(Ma, Liu, Ren, Wang, et al., 2020)
			Intact/Semi-collapsed/collapsed	0.5m	(Song et al., 2020)
		Conventional Machine Learning	Damaged/not-damaged	0.1m	(Frank et al., 2017)
Satellite & Aerial	Post-event	Deep Learning	Damaged/not-damaged	0.02m<<0.6m	(Duarte et al., 2018)
				0.4m<<0.6m	
		Other	Damaged area	0.2m<<30m	(Nie et al., 2016)

The studies used several types of remote sensing images in terms of platform. Satellite images by far are the most used images in earthquake-induced building damage mapping. Using multi-resolution of images by combining the information derived from satellite and aerial images can improve the accuracy of the classification and facilitate the detection of more detailed damage. The deep learning architecture such as ResNet was designed to deal with multi-resolution images (Nex et al., 2019). Regarding the use of conventional machine learning, no study was found using only post-earthquake satellite images, thus further study is required to evaluate the performance of conventional machine learning methods using a single post-event satellite image in this domain.

The literature review also showed that the studies were carried out based on various damage level definitions. Some came up with their own definition, some followed the available standard guidelines while some studies merged the detailed damage level categories to more general classes such as collapsed and non-collapsed or damaged and non-damaged buildings. There is an urgent need for designing an internationally accepted damage level guideline to be followed in creating a comprehensive building damage database.

Regarding the methods employed, the review has shown how the developments in deep learning methods have facilitated multi-source multi-resolution image analysis, which is the case in real scenarios, and how the deep learning extracted features coupled with machine learning classifier can lead to higher accuracy. Diverse classification accuracy using different evaluation metrics was reported for building damage detection. The most common evaluation metric was overall accuracy which range varies from about 70% to about 98% depending on the spatial resolution of the image, the availability of the pre-earthquake image, the number of target classes, and the performance of the model. However, the performance of the methods in producing a damage map cannot be compared only based on the overall accuracy or other common evaluation metrics, as the study areas are quite different. For example, (Xu et al., 2020) proposed the Single Shot MultiBox Detector (SSD) method to detect the intact and ruined buildings from a post-earthquake image and achieved the F-1 score of 89.02%; while in an attempt at classifying the damaged and not-damaged building, the F-1 score of 86.5% was achieved. In another study by (Shao et al., 2020) used a building damage detection network (BDD-Net) and relied only on the F-1 score, one can conclude that the SSD method outperformed the BBD-Net. However, it should also be noted that the buildings considered by (Xu et al., 2020) are sparsely located making it easier to detect the collapsed buildings, compared to the buildings in the dense urban region, considered by (Shao et al., 2020). Moreover, in the study by (Xu et al., 2020),

one specific area was considered while (Shao et al., 2020) considered detecting the damages to the buildings using images from different hazard events around the world.

Table 2.5 shows the advantages and the limitations of the three categories of the methods explained in this section. Generally, feature-based analysis is the simplest compared to the other two methods in terms of the process. It can be directly applied to the image and it does not need the training stage. However, feature-based methods provide low accuracy, and the method designed for specific data cannot be usually generalized to be used on another dataset – i.e. non-transferable. Conventional machine learning techniques, on the other hand, perform well on a small training dataset. Moreover, Machine learning models can benefit from several fusion techniques to deal with imperfect input datasets. The conventional machine learning methods are dependent on the knowledge of the user in choosing the features which are time-consuming and labour intensive. Moreover, the changes in the parameters and tuning process do not affect the performance significantly. Deep learning methods on the other hand, have the privilege of extracting low to high-level features directly from the input image. The hyperparameters used in the structure of the method are flexible, hence making this method adaptable to be used for a new dataset. Compared to the conventional machine learning method, CNN-based deep learning methods require larger labeled datasets to avoid over-fitting which created a challenge in using these methods (Wang et al., 2017). In real scenarios, the number of damaged and collapsed buildings is far less than what is required for training a CNN-based architecture (Farabet et al., 2012).

The need for large training datasets is becoming less challenging by developing the deep learning methods which show successful performance using a relatively small training dataset by leveraging the augmentation method and also the emergence of Generative Adversarial Networks (GAN). Besides, introducing new architecture such as U-Net indicated that deep learning methods can perform well even when there is a small dataset available. Moreover, balancing techniques enhance the classification accuracy in case the training dataset is imbalanced. However, the performance of the proposed methods requires to be evaluated using the image from complex urban areas which is similar to real earthquake scenarios. Table 2.5 summarises the advantages and limitations of the discussed earthquake-induced damage mapping methods.

Table 2.5. Advantages and limitations of reviewed earthquake-induced damage mapping methods using remote sensing data

Method	Advantages	Limitations
Feature-based analysis	• Simple computation • Doesn't need training	• Provide Lower accuracy • Cannot be generalised
Conventional Machine Learning	• Fusion ability • Perform well even on a relatively small training dataset	• Do not perform well in complex urban scenes • Have limited tuning capability
Deep Learning	• Extract high-level features • Provide high accuracy • Can be fine-tuned	• Require a large training dataset • Require large amounts of processing power • Training is time-consuming

The recent availability of large datasets like xBD, which has a large number of labeled images, would significantly reduce the time spent on labeling the images, which are crucial for effective damage assessment. This can shift the focus from preparing the dataset to developing more efficient architectures. However, this database covers a limited number of earthquake events and a limited spatial resolution (0.5m). More effort is needed for transfer learning and detecting the level of the damages to the buildings through building a database including the multi-resolution labeled data of the previous earthquake-induced damages from earthquake susceptible regions. This way, for future earthquake events, the method can be fine-tuned using a small training dataset from the post-event remote sensing images.

2.8. Summary of the Chapter

We presented this progress through analysing the studies conducted in this domain which utilise diverse remote sensing images, following different damage level definition guidelines, and proposed a wide range of damage mapping from feature-based methods to conventional machine learning to more advanced deep learning methods. Although a fair quantitative comparison of the studies is not possible due to a variety of reasons like i)

different input image datasets (SAR, LiDAR, optical, areal, etc.); ii) site-specific design of the proposed methods, and iii) following different guidelines for damage assessment. The outcomes of the reviewed studies are:

• Satellite optical image is the most efficient remote sensing image for earthquake-induced damage mapping considering its quick access and wide coverage.

• Higher resolution of the aerial remote sensing images allows detection of different levels of damage to the buildings.

• Ancillary data such as the building footprint vector data, the pre-earthquake image, and multi-scale images can enhance the performance of the building damage detection.

• The quality of the labeled sample images in terms of accurately labeling the buildings through following a united standard damage level guideline affects the accuracy of the damage map.

• The capabilities of deep learning methods for automatic building damage detection are more efficient than other traditional approaches. However, the performance of the deep learning method is dependent on the spatial resolution of the data, the complexity of the study area, the number of labeled samples, the architecture of the networks, and the number of target damage levels. Further, the use of pre-processing methods such as augmentation or data balancing can improve the accuracy of the deep learning methods. Most studies that used deep learning methods, focused on modifying the model to enhance the accuracy. It is also important to notice that only a few studies considered the location of the damaged buildings by generating a damage map.

• The critical issue in using a general transferable deep learning model is the lack of large datasets of labeled buildings, covering different building types globally. Therefore, there is an urgent necessity for creating a comprehensive building damage database of remote sensing images that are labeled accurately based on a specific damage level standard guideline. Preparing the labeled samples for training the architecture is the most time-consuming step in the process of operational automatic building damage map. The availability of xBD dataset could be considered as a first step towards achieving a general transferable model, as well as considerably reducing the training time. As and when more comprehensive datasets would be available, the ultimate goal of a generalized transferable artificial intelligence-based model can be achieved.

In summary, considering the impact and importance of operational automatic building damage mapping, developing a transferable and explainable AI-based method trained on a comprehensive building damage database following the unique damage level standard deserves further attention and effort.

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1. Introduction

This chapter illustrates the overall design of the methodology. In this chapter, first the characteristics of the study area will be discussed. Then different types and sources of the satellite image will be explained. Finally, the vector datasets used in this study will be discussed in detail. The examples of the limitations of the datasets will be briefly illustrated in this chapter. The architecture of the segmentation and classification models including U-Net, Random Forest, Support Vector Machine, and MultiLayer Perceptron that have been applied to the dataset will be explained. The explainability of the model and the use of Shapley Additive Explanations (SHAP) for feature analysis will be discussed. The model performance evaluation metrics for the classification and segmentation used in this study will be discussed. The image features will be described in detail in this chapter. Further, the process of implementing the models for the individual objectives considering the software required for the implementation will be discussed.

3.2. Study area

3.2.1. Palu city- Indonesia

Sulawesi Island is located at the convergence of the Indian-Australian Plate, the Pacific Plate, and the Eurasian Plate, and there are several active faults in this region, making this area one of the most vulnerable regions to earthquake. Sulawesi Island is located at the convergence of the Indian-Australian Plate, the Pacific Plate, and the Eurasian Plate. Therefore, there are several active faults in this region, making this area one of the most vulnerable regions to earthquake. The largest earthquake in this region, happened in 1996 (7.9 Mw), affecting about 100 km^2 of Sulawesi Island (Jena et al., 2020).

On 28 September 2018, a strong earthquake (7.4 Mw) hit Central Sulawesi in Indonesia. As a result of this massive earthquake, 2081 casualties and 68,451 damaged houses were reported (ASEAN, 2018). A large area of Palu Bay in this region was affected by the resultant tsunami and, as a result of the intense shaking, liquefaction occurred in the southwest of the bay (Syifa et al., 2019).

By using remote-sensing images, several emergency management agencies began to publish a rapid damage map to support the rescue and relief activities in this region. For

instance, the Copernicus Emergency Management Service was activated on the day of the event and soon published a satellite-based map of Palu City showing a preliminary assessment of damage to buildings and roads (Copernicus, 2021). International Charter also was activated a day after the event and provided several damage-assessment maps by comparing pre-event and post-event images of the region (Space, 2021). In this research, in order to conduct an earthquake-induced building-damage assessment, part of the affected region in the southwest of Palu Bay was considered as the study area. Figure 3.1., shows the location of the study area.

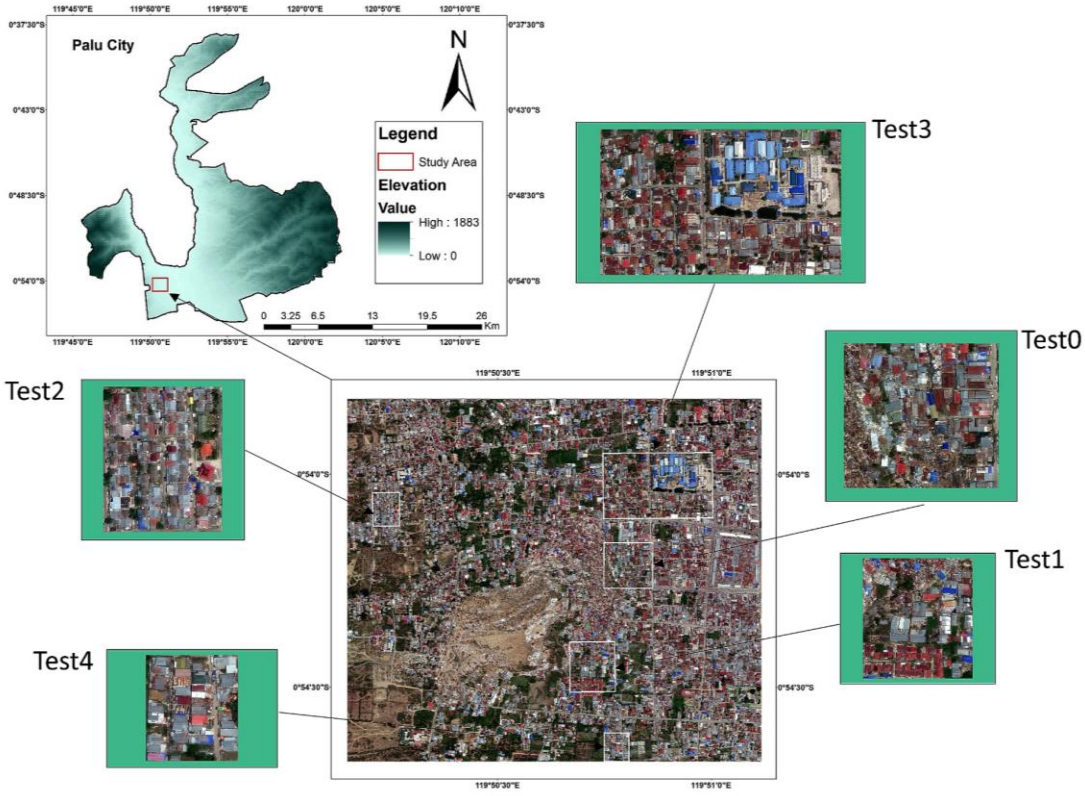

Figure 3.1. Study area location including the training and test areas.

3.2.2. Salina Cruz – Mexico

On 8th of September 2017, a strong earthquake (8.2 Mw) hit Southern Mexico about 700 Km away from Mexico City. The earthquake was followed by more than 4326 aftershocks and the 5.8 Mw after shock occurred 72 Km away from southeast of Salina Cruz (Pozos-Estrada et al., 2019). The devastating earthquake caused about 100 fatalities, mostly in Oaxaca state in the south of Mexico and 41000 buildings were damaged (Ye et al., 2017).

Salina Cruz is located in the south of the Oaxaca state in the south of Mexico and is one of the cities that badly impacted by the 2017 Mexico earthquake. Many old bricks and adobes buildings such as government buildings, church, schools and self-build houses suffered from partial damage or complete collapsed in this area. In this research, in order to evaluate the proposed earthquake-induced building-damage model, part of the affected region in the south of Salina Cruz was considered as the second study area. Three different scenarios were defined to check the transferability of the model and based on one of the scenarios a subset of Salina Cruz study area was considered as a test dataset. Figure 3.2. shows the location of this study area and the correspondence test area.

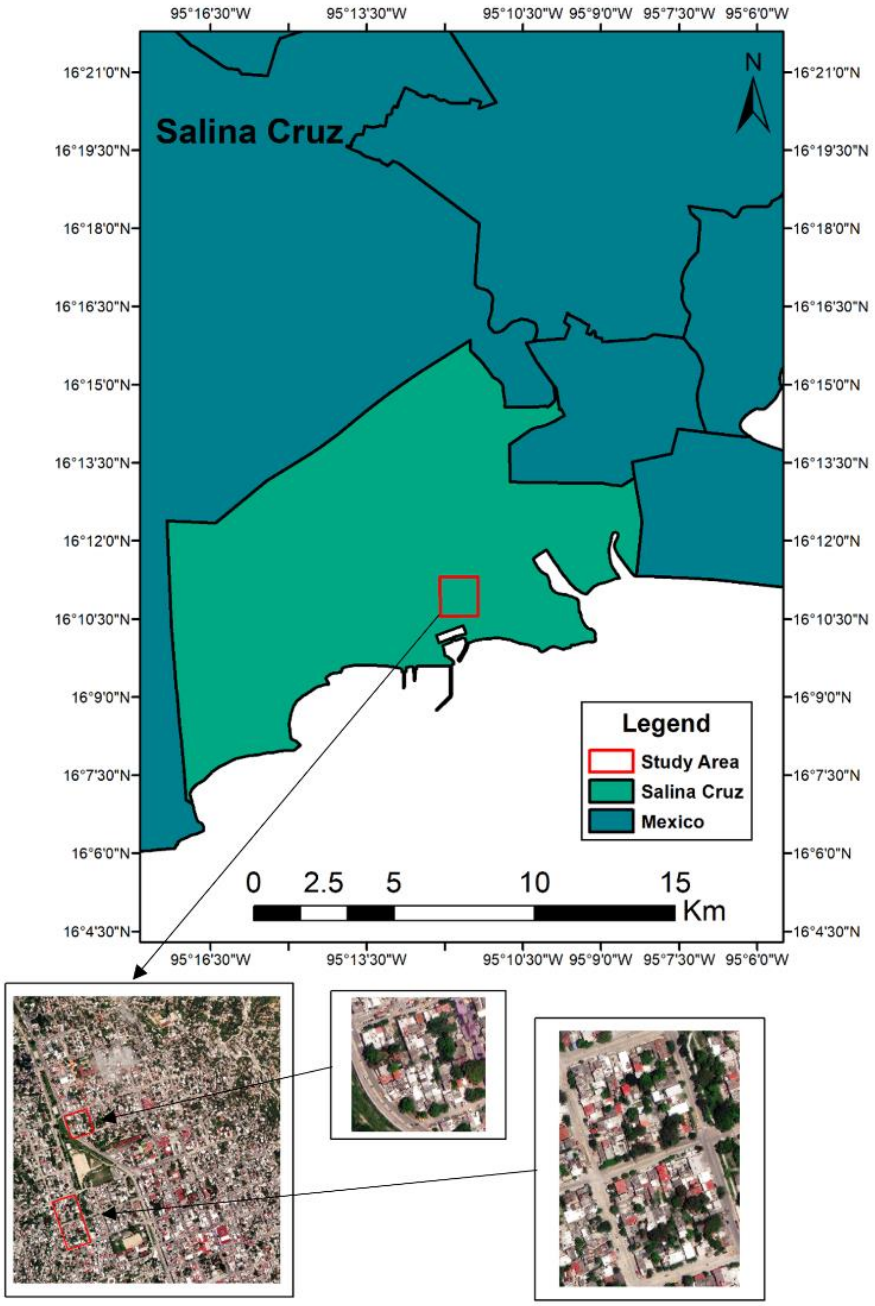

Figure 3.2. The study area location including the train and test area

3.3. Data acquisition

The Maxar (former DigitalGlobe) open data programme aims to provide high-resolution satellite imagery products to support rescue and relief activities in affected regions after major natural or human-made hazards (Maxar, 2021). The pre-and post-event satellite images for affected areas in Indonesia and Mexico are available in the Maxar open data programme (Figure 3.3.). However, due to the complex distortion of the ground surface after the Indonesia 2018 massive earthquake, the co-registration of the images can be erroneous in some parts of the study area (Moya et al., 2020). Therefore, in this study, only

a post-event image was considered for analysis. The WorldView-3 satellite image captured on 2 October 2018 after the Indonesia earthquake was acquired from the Maxar open data programme for this study.

Figure 3.3. Samples of the pre-event buildings and the corresponding post-event buildings for the 2018 Indonesia earthquake.

From the affected area in Indonesia, an area of 3606 x 3148 pixels, equal to about 3 km^2 in the southwest of Palu Bay was selected as the study area. The built-up area in the selected region was about 2821 m^2. Due to the severity of the earthquake, ground distortion, and the liquefaction in the region, the number of collapsed buildings was significant in this study area (Moya et al., 2020). However, the majority of the collapsed buildings were located in the middle of the study area. Therefore, in order to explore the performance of the model in the test areas with different proportions of collapsed and non-collapsed buildings, five test areas were considered, as shown in Figure 3.2.

Further, from the affected area in Mexico, an area of 4577 x 4864, about 1.8 km^2 covering part of Salina Cruz in Mexico was selected to assess the transferability of the model.

3.4. Building footprint

With the increase in geo-information services and the effort of digital volunteers in collaborative projects such as missing maps (Maps), MapSwipe App (MapSwipe), and Open Street Map (Street) building-footprint data will be more in developing countries (Scholz et al., 2018). In this study, we utilised the Open Street Map (OSM) building-footprint data. It is important to mention that the quality of the OSM vector data may vary for different areas and in the absence of a building layer, there are land-cover classification techniques that can be used for extracting built-up areas from Earth observation data (Esch et al., 2011; Esch et al., 2009).

Open Street Map (OSM) provides geospatial data across the world; thus, the geocoded editable map is available for the public and updates weekly. Although having access to OSM data is easy, there are several challenges regarding using this data in training a machine learning model and making a building damage map. During conducting this research the following challenges were identified in using OSM building footprint data for Palu city in this project.

1. The Inhomogeneous distribution of the OSM building footprint data in some remote areas or in the congested buildings in city. In this case, the polygons for the missing buildings were added to the existing OSM data manually.

2. Features are updated at different times and the building footprint data may not be available for some of newly constructed buildings. This can be an issue in the fast growing or dynamic part of the cities like some part of Palu city. Therefore, the new buildings should be detected visually, and their building footprints should be identified and added to the OSM dataset.

3. Topology error. There are some overlapped polygons in some areas. Some polygons were shifted to cover the right building. However, some polygons were deleted and re-drown as the major edits were needed (Figure 3.4.).

4. Co-registering the OSM building footprint and the remote sensing image. Due to the angle of the image, the building vector data may not cover the buildings in the raster image correctly. In this project, some of the polygons were edited to fall onto the corresponding buildings correctly.

5. Inconsistency in the building boundaries. In some areas, particularly in the congested neighbourhoods, there are different boundaries for a single building or there is a single polygon for a group of buildings. These boundaries were edited for such buildings so that each building has a specific polygon based on its roof boundary. As it is shown in figure 3.5. the subject area that has been covered by polygon 1 is part of polygon 2 as well. This results in confusion in labelling the buildings. As shown, polygon 2 covers many smaller scale buildings that is required to be labeled separately.

Figure 3.4. An example of the misaligned OSM building footprint for Salina Cruz study area.

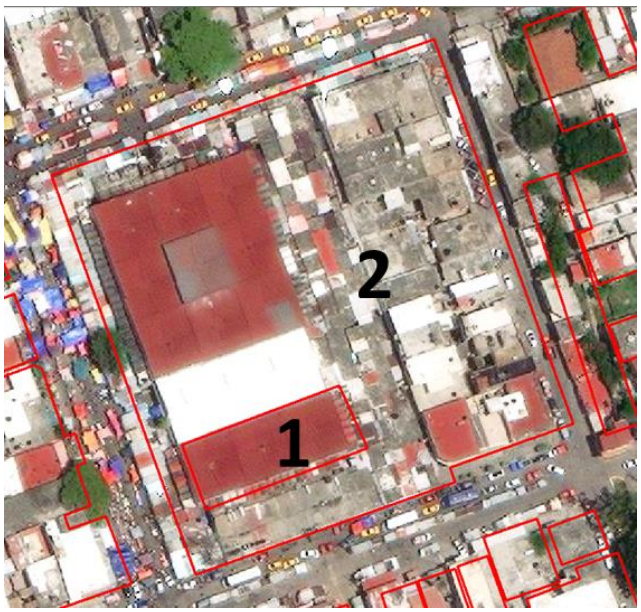

Figure 3.5. An example of inconsistency in the building boundaries for Indonesia study area. Polygon 1 shows a single building while polygon number 2 covers number 1 polygon and other buildings as well.

In this study, we created an area of interest to download the OSM built-up area layer in both regions (Street). Due to the distortion of the ground after the earthquake in Indonesia, co-registration of the polygons and the optical image captured by Maxar was challenging and hence, some of the polygons were manually shifted to align the buildings and the corresponding polygon in the image (Moya et al., 2020). In the same manner, the misalignments and inconsistency of the OSM built-up area dataset for Salina Cruz study were assessed and fixed extensively.

3.5. Copernicus Management Services

The Copernicus Emergency Management Service (EMS) is part of the European Union's Earth Observation programme that provides different mapping products, including but not limited to, natural hazard rapid maps and digital geoformation on a global scale (Copernicus). Copernicus EMS published a grading rapid-damage map along with a building damage label dataset, in a point format, for the earthquake in Indonesia on 2 October 2018 (Copernicus, 2021). The dataset shows different levels of damage to buildings (i.e., destroyed, damaged, possibly damaged and no visible damage). The number of buildings in each class is very different for the region of interest, and hence, the extracted dataset is highly skewed. Further, the low-resolution of the images do not allow us to detect the degree of damage to buildings. Therefore, we considered the damaged, possibly-damaged, and no-visible-damage building classes as non-collapsed and limited the study to classifying them into collapsed and non-collapsed buildings. Some of the damaged buildings in the region of interest were not assigned by a label, as some label points do not correspond to the buildings after co-registration of the dataset and the image. Therefore, we manually labeled the remaining buildings based on the Copernicus EMS damage-assessment classes (Table 3.1). The overview of the quantity of collapsed and non-collapsed buildings in the training and test datasets is shown in Table 3.2. As mentioned above, the selected test areas are diverse in terms of containing different portions of collapsed and non-collapsed buildings. Further, due to the contiguous built-up area, cropping a test area that covers the complete buildings in the border of the test area is very challenging; hence, this difficulty led us to selecting areas with a limited number of buildings, such as in test 2 and test 4.

Table 3.1. Damage classification used in this study for preparing the training dataset and the number of buildings in each class.

This study	Copernicus EMS guideline
Non- Collapsed (No.6373)	No visible damage
	Possibly damaged Uncertain interpretation due to image quality Presence of possible damage proxies. Building surrounded by damaged/destroyed buildings
	Damaged Minor: The roof remains largely intact, but presents partial damage Major: Partial collapse of the roof; serious failure of walls
Collapsed (No.2135)	Destroyed Total collapse, the collapse of part of the building (>50%); Building structure not distinguishable (the walls have been destroyed or collapsed)

Table 3.2. Overview of the quantity of collapsed and non-collapsed buildings in the training and test datasets.

Study area	Number of collapsed buildings	Number of non-collapsed buildings
Train	1854	5638
Test 0	119	78
Test 1	127	107
Test 2	0	109
Test 3	28	401
Test 4	1	47

As explained, the Copernicus publishes rapid building damage map in point format after major natural hazard events. In some parts of the affected area the Copernicus EMS points were not aligned with the collapsed buildings in the Indonesia image dataset. Therefore, the points were manually aligned with the buildings to later assign the labels to the buildings. An example of such misalignment has been shown in Figure 3.6.

Figure 3.6. An example of the misalignment of Copernicus EMS points and the collapsed buildings in Indonesia study area.

Table 3.3. summarises different datasets employed to prepare the dataset used in this study.

Table 3.3. Overview of the source and characteristics of the datasets used in this study.

Dataset	**Source**	**Type**	**Acquisition/ Published Date**	**Resolution**
Indonesia Earthquake and Tsunami	Maxar	Raster (RGB)	2/10/2018	0.5 m (resample)
Built-up area	OSM	Shapefile (Polygon)	16/07/2019	-
EMSR317 Earthquake in Indonesia	Copernicus	Shapefile (point)	30/9/2018	-
Southern Mexico Earthquake	Maxar	Raster (RGB)	9/9/2017	0.3 m (resample)
Built-up Area	OSM	Shapefile (Polygon)	6/7/2022	

EMSR240 Earthquake in Southern Mexico	Copernicus	Shapefile (point)	13/9/2017	

3.6. Machine learning models

3.6.1. U-Net model for segmentation

Generally, deep learning is a statistical technique for identifying patterns in a labeled dataset using neural networks with multiple layers. Neural networks typically consist of input units (exp. image samples), multiple hidden layers, and a set of output units. The connection between the units is adjusted by an algorithm called back-propagation so that a given input results in the corresponding output (i.e. input-output mapping). In most deep learning architectures, a technique called convolution is used to capture the features from the training image samples. Deep learning models are often used in classification tasks to decide which class each test input belongs to (Marcus, 2018).

Generally training a deep neural networks model requires thousands of annotated image samples. In many image processing applications such as biomedical science using medical images or building damage detection using remote sensing images, a large number of annotated image samples are not available. To fill this gap, U-Net architecture was presented in 2015 to use the small training dataset efficiently (Ronneberger et al., 2015). U-Net is a fully convolutional network (FCN) which means it doesn't have fully connected layers. In total U-Net has 23 convolutional layers. The symmetric architecture of the U-Net consists of an encoder (the left side) and decoder part (the right side). The encoder has a typical convolution network architecture. This side consists of 3x3 unpadded convolutions followed by a ReLU and a 2x2 max pooling operation with stride 2 for downsampling. The number of feature channels is doubled at each downsampling step. On the other hand, each step in the decoder consists of upsampling of the feature map followed by a $2\ X\ 2$ up-convolution, a concatenation, and two $3\ X\ 3$ unpadded convolutions followed by a ReLU in the final layer of the U-Net At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes (Ronneberger et al., 2015).

Table 3.4. shows the hyperparameters used in the proposed U-Net model.

Table 3.4. Hyperparameters used in the proposed U-Net model

Initial learning rate	0.1
Number of epochs	100
Filter size	3 x 3
Pooling size	2 x 2
Activation function	ReLU

3.6.2. Multilayer Perceptron

Artificial neural networks are an algorithm that attempts to emulate the structure of the biological neurons in the brain. The multilayer perceptron is an example of an artificial neural network that has been used extensively in image classification (Noriega, 2005). The architecture of the multilayer perceptron (MLP) consists of three main elements, input, hidden layers, and output; the inputs are the image features. Each node in MLP is a neuron with non-linear activation function (Tang et al., 2015). MLPs utilize a supervised learning technique called backpropagation for training. MLPs have a fully connected structure wherein each perceptron is connected to all other perceptrons (Figure 3.7). In terms of input, different image features, including six spectral, six textural, and three shape features were extracted from the image. The feature extraction method and the features used in training the MLP model has been explain in section 3.7 in detail. Further, the extracted features are consistent in all the images used in this study.

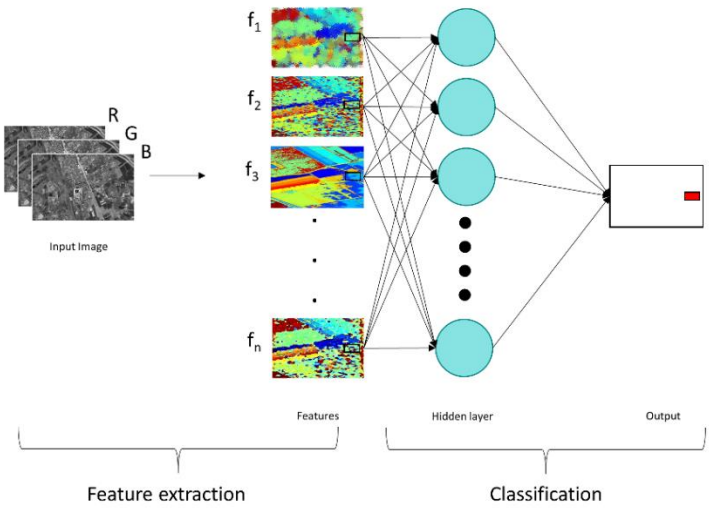

Figure 3.7. Structure of MLP for building-damage classification (the images/features in this figure are symbolic and are not the real images/features used in this study).

In each epoch, the network-predicted output is compared to the actual output and the calculated error is propagated back to the networks and therefore, the weights are updated accordingly to minimise the error. Table 3.5. shows the hyperparameters used in the proposed model.

Table 3.5. Hyperparameters used in training the proposed MLP model in this study.

Learning Rate	0.01
Hidden layer	2
Activation Function	ReLU
Solver	Adam
Input Nodes	50
Iterations	500

3.6.3. Random Forest (RF)

The random forest algorithm (Breiman, 2001) has been widely used to solve many classification problems. Random Forests (RF) model, developed in 1990 (Breiman, 2001), is well-known for its capability in classification and regression and commonly used for

measuring the importance of the features (Janitza et al., 2013). RF model has found to be accurate and stable in classifying high-dimensional data (Ma et al., 2017). Further, this model had been used in many building damage detection studies (Adriano et al., 2019; Bialas et al., 2019). Therefore, in this study the RF performance along with its built-in feature importance scores was evaluated in classifying the post-earthquake image into two classes of collapsed and non-collapsed buildings. Further, the Variable Importance Measurement (VIM) plot was drawn to show the ranking of the features. The RF classification results will be shown and discussed in the next chapter, results and discussion, in section 4.2.2.

3.6.4. Support Vector Machine (SVM)

The support vector machine (SVM) is another well-performed supervised machine learning algorithm that have been widely used to solve many classification problems (Noble, 2006). This algorithm, developed by Vladimir Vapnik in 1992 (Cortes & Vapnik, 1995). SVM construct a set of hyperplanes in an infinite-dimensional space that separate the data into different classes. This algorithm has shown good results in classifying and segmenting the images (Barghout, 2015). Therefore, in this study SVM has been used to classify the post-earthquake image into two classes of collapsed and non-collapsed buildings.

3.7. Feature extraction methods

The fundamental step of image classification is selecting the meaningful features that describe the information which can be extracted from an individual or a group of pixels in an image (Ranjbar et al., 2014). In this study, the spectral and textural features were calculated for each pixel and then averaged for the group of pixels within the building polygon boundary (i.e. image objects) within each building boundary. To avoid over- and un-der-segmentation and considering the multi-scale nature of geographical objects in the image, non-homogeneous materials used in the building rooves, and the large uniform region in the liquefaction area, we integrated the building vector data in the segmentation process. The details of this method can be found in the study by (Kato, 2019).

Texture features are referred to as the spatial distribution of the intensity values in the image (Guide, 2009). These features can represent the smoothness or roughness of a group of pixels in the image. Texture features can be calculated for each pixel and have been widely used in the domain of building damage detection (Haiyang et al., 2010; Khodaverdizahraee et al., 2020). Texture features can be extracted using several texture analysis methods such as Gabor filter, fractal dimension, and Haralick. Among these, the Grey-level co-

occurrence matrix (GLCM) and gray-level difference vector (GLDV) have been reported to perform well in building damage detection (Haiyang et al., 2010; Janalipour & Mohammadzadeh, 2015; Khodaverdizahraee et al., 2020). GLCM is a matrix of gray-level distributions for a pair of pixels at a specific distance and orientation while GLDV counts up the number of in-stances of a given difference between pixels. These features contain information about the position of the pixels with the same gray-level values. Therefore, two sets of GLCM and GLDV texture statistical features were calculated for each pixel in this study. The features were calculated over all the shifts of 1x1 of 0, 45, 90, and 135 degrees, using a 3x3 pixel window, and 32 as the number of gray levels for each pixel in the image. Then the average of these statistics was returned for each pixel. These features and the corresponding mathematical expressions are shown in Table 3.6, where $P_{(i,j)}$ is the normalized value of the pixel i, j and V_K is the image object level and N is the number of rows and columns in the image object.

Table 3.6. Extracted GLCM and GLDV features used in this study

Number	Feature Name	Expression
1	GLCM Dissimilarity	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$
2	GLCM contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$
3	GLCM Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$
4	GLCM Entropy	$\sum_{i,j=0}^{N-1} P_{i,j}\,(-lnP_{i,j})$
5	GLDV entropy	$\sum_{K=0}^{N-1} V_K\,(-lnV_K)$
6	GLDV Mean	$\sum_{K=0}^{N-1} K(V_K)$

Spectral features represent the mean intensity value variation in different image bands. These features are used to describe different image objects based on their spectral

information (Ranjbar et al., 2014). The spectral features used in this study and their relative mathematical expressions are shown in Table 3.7, where $\bar{C}_i(v)$ is the mean intensity of image band i for object v, P_v is the set of pixels in that object and $\#P_v$ is the number of pixels in the set. w_k^B is the brightness weight of image band k, $c_k(x,y)$ is the intensity value at the pixel (x,y) and $B_v(d)$ is the bounding box around the object v with the distance d.

Table 3.7. Extracted spectral features to be used as the input features of the initial classification

Feature Name	**Expression**
Maximum Differences	$\frac{max_{i,j\epsilon K_B}\left\|\bar{C}_i(v)-\bar{C}_j(v)\right\|}{\overline{C(v)}}$
Brightness	$\frac{1}{W^B}\sum_{k-1}^{K} w_k^B\,\bar{c}_k(v)$
Mean Layer	$\frac{1}{\#P_v}\sum_{(x,y,z,t)\epsilon P_v} c_k(x,y)$
Standard deviation	$\sqrt{\frac{1}{\#P_v}\left(\sum_{(x,y,z,t)\epsilon P_v} c_k^2(x,y)-\frac{1}{\#P}\left(\sum_{(x,y,z,t)\epsilon P_v} c_k(x,y)\right)^2\right)}$
Skewness	$\frac{\sum_{(x,y)\epsilon P_v}(c_k(x,y)-\bar{c}_k(v))^3}{(\sum_{(x,y)\epsilon P_v}(c_k(x,y)-\bar{c}_k(v))^2)^{\frac{3}{2}}}$
Contrast to Neighbour Pixels	$1000\,(1-\frac{\bar{c}_k(B_v(d)-P_v)}{1+\bar{c}_k(P_v)})$

Several image-object shape features have been used in many post-earthquake building damage detection studies (Haiyang et al., 2010; Khodaverdizahraee et al., 2020; Wang & Jin, 2011). Three different shape features were extracted in this study. Density describes the distribution of the pixels in the object image (Guide, 2009). Asymmetry is the ratio of length and width of the image objects, and rectangular fit indicates how well the object can fit into a rectangle. The selected shape features are shown in Table 3.8, where $\rho_v(x,y)$the elliptic distance at a pixel is (x,y), $Var\,X$ and $Var\,Y$ are the variance of X and Y , respectively.

Table 3.8. Shape features used in the study

Feature name	Expression
Rectangular fit	$\frac{\{\#(x,y)\epsilon P_v: \rho_v(x,y) \leq 1\}}{\#P_v}$
Asymmetry	$\frac{2\sqrt{\frac{1}{4}(Var\ X + Var\ Y)^2 + (Var\ XY)^2} - Var\ X - Var\ Y}{Var\ X + Var\ Y}$
Density	$\frac{\sqrt{\#P_v}}{1 + \sqrt{Var\ X + Var\ Y}}$

3.8. Explainable machine learning models

Using the explainable machine learning model, we attempt to answer the questions such as 1- how the model decision is associated with the value of each feature, 2- are these associations meaningful according to the domain knowledge and 3- what features has the most effect on the model decision?

The interpretability of the ML models includes global and local interpretation. Global interpretability provides insights about the general relationship between the features and the model output. Therefore, the AI user can understand how the complex ML model works. Local interpretability, on the other hand, considers the individual ML decision (Du et al., 2019).

Figure 3.8. shows the workflow diagram of XAI for earthquake damage mapping. The initial steps are similar to the building a conventional machine learning model. However, the explanation and interpretation of the model decisions will be added to the workflow to provide insight about the data. XAI aids to detect the errors, biases and artefacts in the data that may have been unnoticed and used by the model. The types of the errors in the training dataset include (i) biases dataset (for example capture bias, label bias, negative bias) and (ii) spurious correlations in the training dataset. More details about such errors were

explained in the recently published papers by (Anders et al., 2022).

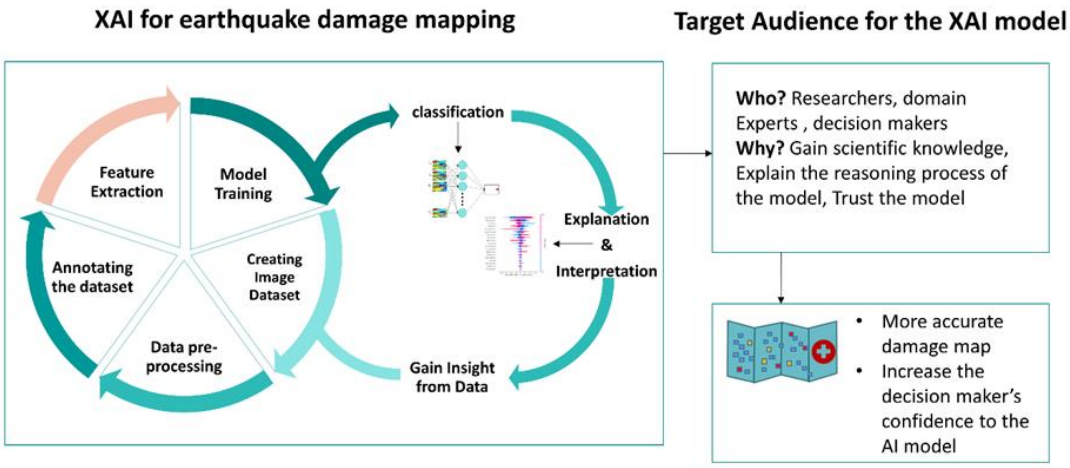

Figure 3.8. The workflow diagram of XAI for earthquake damage mapping

3.8.1. Shapley Additive Explanations (SHAP)

The Shapley additive explanations value (SHAP) was first introduced for calculating the contribution of the individual players in a coalition game (Shapley, 1953). Recently, (Lundberg & Lee, 2017) raised this concept as the basis for an algorithm measuring the importance of the features in machine learning models to explain the output of any ML model. In this case, reproducing the model output considered as the coalition game, in which the players are all the features used in the model, and the number of the observations (number of samples) is the number of times the game is played.

Shapley value provides the global interpretability of the model by estimating the general contribution of each feature on the model output, as well as local interpretability by estimating for each observation. The Shapley value is calculated based on the average of the marginal contribution across all the possible permutations of the features. The mathematical expression of the classical SHAP value is given in Eq. 1.

$$\emptyset_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n-|S|-1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad \text{- (1)}$$

where, $\emptyset_i$ is the contribution of feature i, N is the set of all the features, n is the number of features in N,N\{i} is the set of all features except i, S is any subset of N not containing the feature i and $v(N)$ is the base value meaning the predicted output for each feature in N without knowing the feature values.

The model output for each observation is estimated by summing up the SHAP value of each feature for that observation. Therefore, the explanation model is formulated as below:

$$g(z') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i z_i' \quad \text{- (2)}$$

where $z' \in \{0,1\}^M$ and M is the number of features and $\emptyset_i$ can be obtained from the equation 1. As shown in equation 2, summing up the effect of each feature (SHAP value) approximates the output of the model. SHAP provides multiple AI model explainers. The details about different model explainers are beyond the scope of this study.

The readers are referred to (Molnar, 2020). Using the relevant model explainer, the global and local importance of the features can be achieved using the summary plot and the force plot, respectively.

3.8.2. SHAP Summary Plot

The summary plot depicts both the feature importance ranking and the feature effects. The term feature effect describes how each feature contributes to the classification output (the X-axis on SHAP summary plot) while the feature importance describes how each feature contribute to the classification performance (the Y-axis on SHAP summary plot). Each point in the summary plot represents the Shapley value for a feature and an observation in the dataset. The position of the feature on the Y-axis determines the importance of that feature, and the X-axis represents the Shapley value of each observation. The color on the right of the plot shows the value of each observation from low to high, hence the direction of the relationship between the feature and prediction is shown. An example of the application of the summary plot will be shown in the next chapter to discuss the result of the MLP model (Figure 4.3).

3.8.3. The individual force-plot

Apart from the global interpretability, SHAP provides local interpretability by enabling us to identify the impact of each feature on classifying the individual objects in the image. This capability allows us to analyse the classified and misclassified objects in more detail, which leads us to a deeper understanding of the dataset and the model.

3.8.4. Feature analysis

Remote sensing images contain a huge amount of information that can be extracted and used for building damage detection (Bolon-Canedo & Remeseiro, 2019). However, the inclusion of more features in any machine-learning-based classification model may not yield a better result. More number of features could lead to the presence of redundant or irrelevant features, adding a layer of complexity to the model, thereby increasing the

computing time and may even lead to a decrease in accuracy (Duangsoithong & Windeatt, 2009). By selecting the most relevant features to distinguish between collapsed and intact buildings, there is more likely to achieve a good result even without using an advanced machine learning model and hyperparameter tuning. Therefore, feature selection is a critical pre-processing step in classification studies (Ranjbar et al., 2018).

This stage aims to explore the impact of the extracted features on the model output and therefore remove the redundant features and select the best features to include in the modeling. To this end, SHAP summary plot has been used to rank the impact of the features on the capability of the model in distinguishing the collapsed and non-collapsed buildings.

3.9. Software for model implementation

In this research, the main dataset included optical satellite image and the OSM building footprint in raster and vector format respectively, which was obtained from Maxar Open data program and Open Street Map (OSM). QGIS was used to extract the building footprint dataset from OSM website. ArcMap software was used to pre-processing, and post-processing, and raster to vector, point raster calculation, co-registering the raster and the vector dataset, and mapping Python language was used to develop U-Net, MLP and RF classification models for distinguishing the collapsed and non-collapsed buildings. In order to extract the image features eCognition software was used which is primarily used for object image analysis for satellite images (Nussbaum & Menz, 2008). Then the extracted features were exported in CSV (Comma-separated Values) format which pre-processed in Microsoft Excel spreadsheet. Next, the Python language was used to analyse the data and develop the models.

The U-Net model has been trained on Google Collaboratory (Google Colab) where free GPUs and TPUs are provided by Google. The data could be uploaded, stored, created, edited and shared on this platform. In order to analyse the data and implement machine learning models, a Jupyter notebook is provided to use python language. Further, 128 GB of Disk space and 12 GB of RAM is offered by Google.

Table 3.9. The software used in this study and their features and applications in this research

Software	**Features and Functionality**	**Application in This Research**
Microsoft Excel	Storing, filtering, and editing the data with the	Used in storing and transferring the data from

	capability of handling large data.	the eCognition software and to the ArcMap. Filtering the data features.
QGIS	Free Open-Source GIS software to store, edit and analyse the geospatial dataset. QGIS supports vector and raster data formats.	Used to extract the building footprint dataset from the OSM website and extract the vector building footprint in .shp format.
ArcGIS 10.4.1	Commercial GIS software to store, manipulate and analyse the geospatial dataset.	Used to co-register the satellite image and the building footprint, select the training and test area, edit the building footprint polygons, add labels to the buildings, join the prediction results to the building attributes, visualise the data and generate a damage map.
Python 3.7	High-level programing language for small and large-scale projects.	Used to generate more data through augmentation. Used to read, edit and analyse the data and develop the proposed ML and DL models.

3.10. Performance evaluation metrics for the models

3.10.1. Performance measures for the segmentation model

One way to evaluate the segmentation model performance is using Pixel Accuracy in which the number of correctly classified pixels is considered as the accuracy of the model. However, the high Pixel Accuracy is not necessarily the indicator of a well-performed segmentation model. The high Pixel Accuracy may occur in highly imbalanced datasets

where most pixels in an image patch belong to only one of the classes in the dataset. Because in real-world post-earthquake image datasets the number of collapsed and damaged buildings is far more than non-collapsed and non-damaged buildings, using Pixel Accuracy for evaluating the segmentation model for post-earthquake damage assessment is not a suitable metric (Long et al., 2015).

The Dice Coefficient (F1 score) (Anuar & Sultan, 2010) on the other hand can be a better option for evaluating the performance of an image segmentation model. The dice coefficient is the ratio of the area of intersection of the predicted and the ground truth pixels over the union of the predicted and ground truth pixels. Dice Coefficient metric ranges between 0 and 1 where 0 means no similarity between the ground truth image and the predicted image whereas 1 means 100% similarity. The equation 3 Formulate the Dice Coefficient in which A is the prediction and B is the ground truth image. Accordingly, the numerator of the ratio is twice the number of pixels in the overlap area and the denominator is the total number of pixels in both the predicted and the ground truth image.

$$D.C = \frac{2\times|A\cap B|}{|A|+|B|} \qquad \text{-(3)}$$

3.10.2. Performance measures for the classification model

In binary machine learning classification, each input and output have belonged to one of two non-overlapping classes (Asuncion & Newman, 2007). Since in this study the buildings were labeled as either collapsed or non-collapsed, the inputs (i.e. the buildings) were labeled as 0 for non-collapsed and 1 for collapsed buildings.

There are different metrics to measure the performance of a machine learning model. The most common metric is overall accuracy which is the ratio of the correctly classified samples to the total number of samples in the test dataset.

The objective of the ML classifier in this study is to detect collapsed buildings. Therefore, a true positive case is where the ML model correctly predicts the building to be collapsed. Similarly, a true negative case is where the ML model correctly predicts the building to be non-collapsed. A false positive case on the other hand is where the ML model incorrectly predicts a building to be collapsed. Similarly, a false negative case is where the ML model incorrectly predicts a building to be non-collapsed. The performance of the ML classification can be assessed by calculating the true positives, true negatives, false positives, and false negatives. These four numbers constitute a confusion matrix (Table

3.10). The most often used evaluation metrics are defined according to the values of the corresponding arrays in the confusion matrix.

Table 3.10. Confusion matrix for the ML binary classification

Data class	Collapsed observed	Intact Observed
Collapsed predicted	True Positive	False Negative
Intact predicted	False Positive	True Negative

Precision is defined as the number of true positive cases divided by the total number of true positive and false-positive cases. Precision can be formulated as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives+False Positives}} \quad \text{-(4)}$$

In the imbalanced dataset, precision represents the accuracy of the minority class. Recall on the other hand is defined as the number of true positive cases divided by the total number of true positive and false-negative cases. Recall can be formulated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives+False Negative}} \quad \text{-(5)}$$

Recall provides insights into the missing positive predictions. Since in this study identifying the missing collapsed building are of importance, the goal is to decrease the false negatives. According to the recall formula, the recall increases by decreasing the false negatives. Therefore, in this study, the goal is to increase the recall.

F1 score is another common evaluation metric which is the combination of precision and recall and can be formulated as:

$$\text{F1 score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{-(6)}$$

In order to evaluate the performance of the model, the achieved results should be compared to the ground truth dataset. The building-states dataset collected through ground inspection after the earthquake is usually considered the ground truth. However, in this study, access to the ground-inspection dataset was not possible, and therefore, the Copernicus grading map was considered the ground truth.

In this study, the confusion matrix was created and its parameters (i.e., overall accuracy, precision, recall, and F1 score) were calculated to quantitatively summarise the performance of the models (Sokolova & Lapalme, 2009).

3.11. The overall methodology of the research

The overall methodology is presented in figure 3.9. which briefly describes the three objectives: Assessing different machine learning models for post-earthquake building damage classification, developing an explainable machine learning model, and assessing the transferability of the proposed building damage detection model on a different geographical area.

In the first stage, the building footprint dataset was obtained from different sources such as OSM building footprint vector data, and the data was manually edited to achieve more accurate building footprint data. Then a single post-earthquake optical satellite image was obtained from Maxar (former Digital globe) open data program. The building footprint vector data and the satellite image were co-registered. Training and test datasets were determined and data augmentation has done to increase the image samples in the training dataset. Some images were manually labeled and the consistency of the labels was assessed. After preparing the dataset several machine learning models (U-Net, RF, SVM, and MLP) were developed and the performance of them was assessed to select the most suitable model for classifying the post-earthquake. The evaluation metrics and the visual assessment shows that MLP is the most suitable model for post-earthquake building damage classification.

In the second step, shapley additive explanations (SHAP) was used to provide insight into the training dataset and also to shed some light on how MLP classifies the dataset. The SHAP summary plot was used to rank the contributions of the features and also to show the direction of the feature's importance. According to the SHAP summary plot, the redundant features were identified and removed from the dataset. Then the most important features were fed into the MLP classification model to decrease the complexity of the model and increase its performance.

In the third step, the dataset from another geographic area was added to the previous dataset to assess the transferability of the proposed MLP model. The same data preparation steps were taken to prepare the dataset. The selected features from objective two were extracted from the new image, added to the previous dataset, and fed into the proposed MLP model to train the model. The performance of the newly trained model was evaluated on the same

test dataset and compared to the previously trained model using common evaluation metrics. Finally, the building damage map was generated.

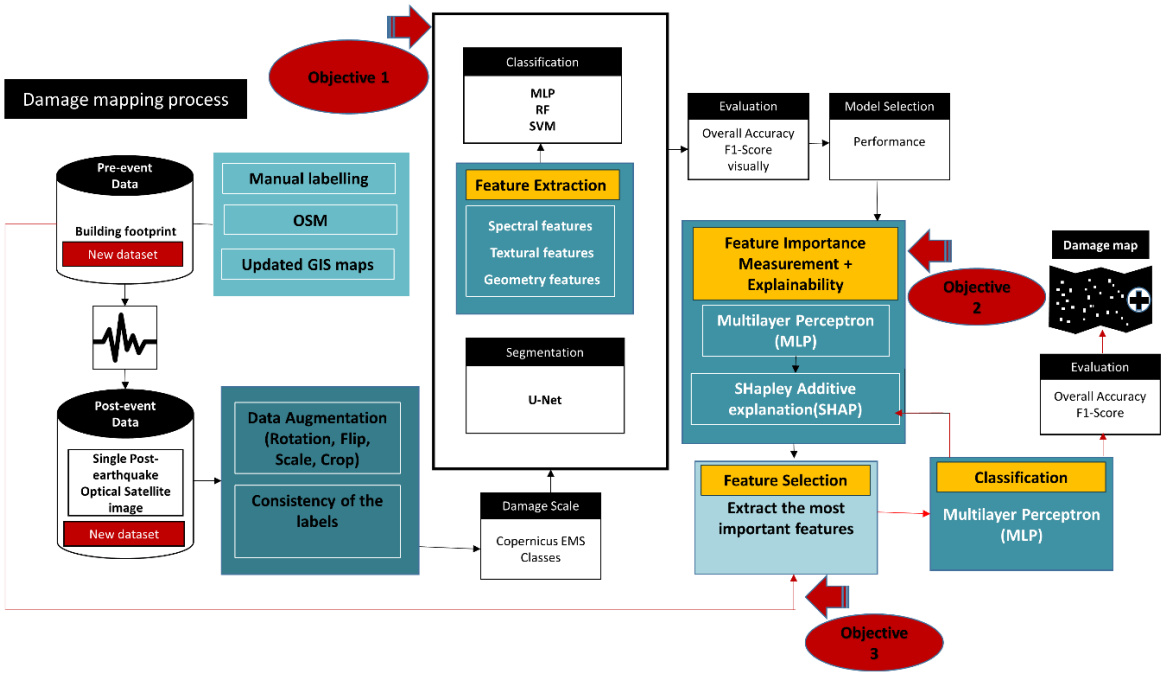

Figure 3.9. Overall methodology for machine-learning based post-earthquake building damage mapping.

3.11.1. Objective 1

In order to prepare the training dataset, each pixel is required to be assigned with a collapsed or non-collapsed label which is called a mask. Due to the intensity of the earthquake and the consequent liquefaction in a large part of the study area, the post-earthquake building footprints for the majority of the collapsed buildings do not match the pre-earthquake building footprints. Therefore, the OSM building footprints are not a good representative of the locations of the buildings after the earthquake and hence cannot be used as masks for the training stage. To solve this issue, the collapsed buildings along with the debris were manually labeled as collapsed whereas the OSM building footprints were used for the collapsed and non-collapsed buildings outside the liquefaction area.

The overall methodology of this objective has been shown in figure 3.10. In the first step, the built-up area is extracted. In the next step collapsed and non-collapsed buildings were labeled. OSM building footprint and the corresponding building labels were used to create the training mask for most of the areas. However, as explained earlier, due to the misalignment of the OSM building footprint with the buildings in the liquefaction area, the collapsed buildings and the debris were masked manually in that area. In the next step,

image augmentation was used to increase the number of image patches and their corresponding masks. Supervised deep learning models are data-hungry (Marcus, 2018) and as the number of image patches is limited, data augmentation has been used to create more image patches. The satellite image samples were flipped, rotated, and sharpened to create more training image samples. Next, the image patches and masks were randomly divided into 70% train and 30% test image dataset. The training dataset was fed into the U-Net model and the test dataset was used to show the performance of the model to segment the image into collapsed and non-collapsed buildings.

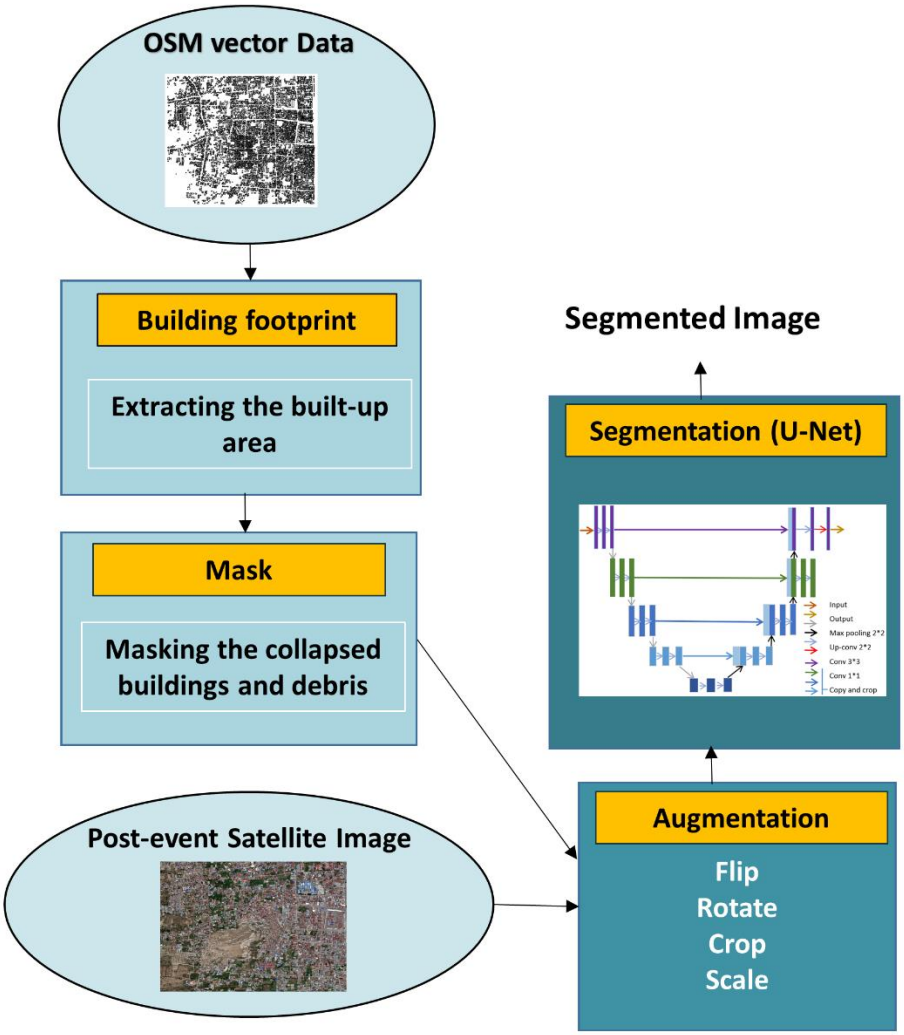

Figure 3.10. Overall methodology flowchart for part of objective 1, using the U-Net architecture for segmenting the RGB satellite image to collapsed and non-collapsed regions.

The parameters that have been used in the U-Net model are shown in the table 3.11.

Table 3.11. The parameters used in training the proposed U-Net model in this study

Initial learning rate	0.1

Number of epochs	100
Filter size	3 x 3
Pooling size	2 x 2
Activation function	ReLU

The second part of objective 2 includes assessing the performance of different classification models for building damage classification. Multilayer Perceptron (MLP), Random Forest (RF), and Support Vector Machine (SVM) were considered for this part. In order to develop MLP, RF and SVM the first step is to extract the relevant features from the images. As explained in the section 3.7, a series of spectral, textural, and geometry features have been extracted from the image. The features for the built-up area were obtained and fed into the proposed machine learning models. In the next step, the performance of the three models was compared using the common evaluation metrics. Finally, a building damage map was generated based on the classification result of the selected model.

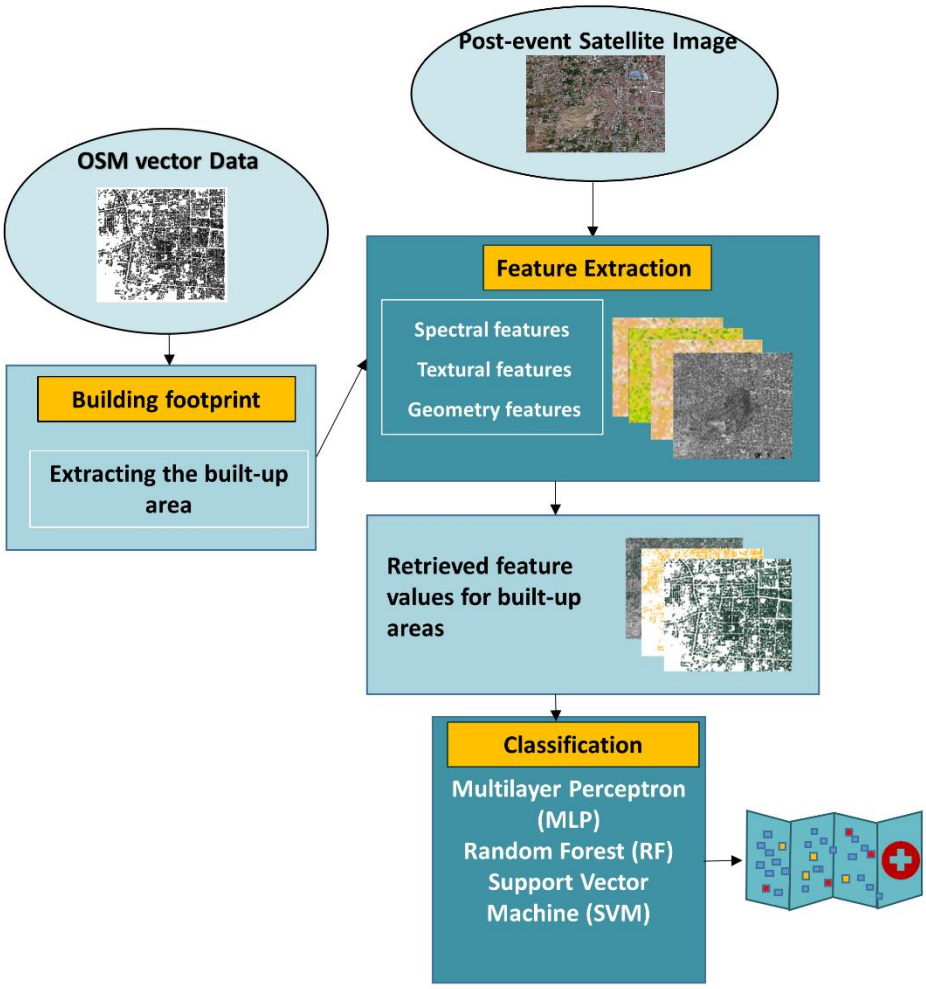

Figure 3.11. Methodological flowchart for objective 1, using different classification models for classifying the collapsed and non-collapsed buildings.

The parameters that have been used for developing the MLP are shown in the Table 3.12.

Table 3.12. The parameters that have been used for implementing the MLP

Learning Rate	0.01
Hidden layer	2
Activation Function	ReLU
Solver	Adam
Input Nodes	50
Iterations	500

3.11.2. Objective 2

In this subsection, the objective is to build an explainable model. To this end, Shapley additive explanation (SHAP) was used to rank the importance of the features and give insight into the dataset. The SHAP summary plot has the capability of showing the impact of each feature on the model outcome. It also shows the direction of the correlation between each feature and the model outcome. The contribution of some of the features at the bottom of the SHAP summary plot is negligible and those features can be removed from the dataset as redundant features. Removing the redundant features usually decreases the dimension of the dataset and improves the model performance. Therefore, the model will be trained using the features that contribute the most to the model.

In the final stage, the common evaluation metrics were used to evaluate the performance of the model in different test areas from the Palu city dataset. Next, a building damage map was generated based on the result of the proposed classification model.

Moreover, SHAP force plot was used to explore the model decisions on some of the sample images. SHAP force plot assists the local interpretability of the model where the contribution of the features on the individual decision for each observation will be shown. Understanding how features contribute to the model decision for each image sample aids us in finding the root cause of the error in case of an incorrect classification result and also gives us insight into the features of the data we need to add to the training to improve the model decisions.

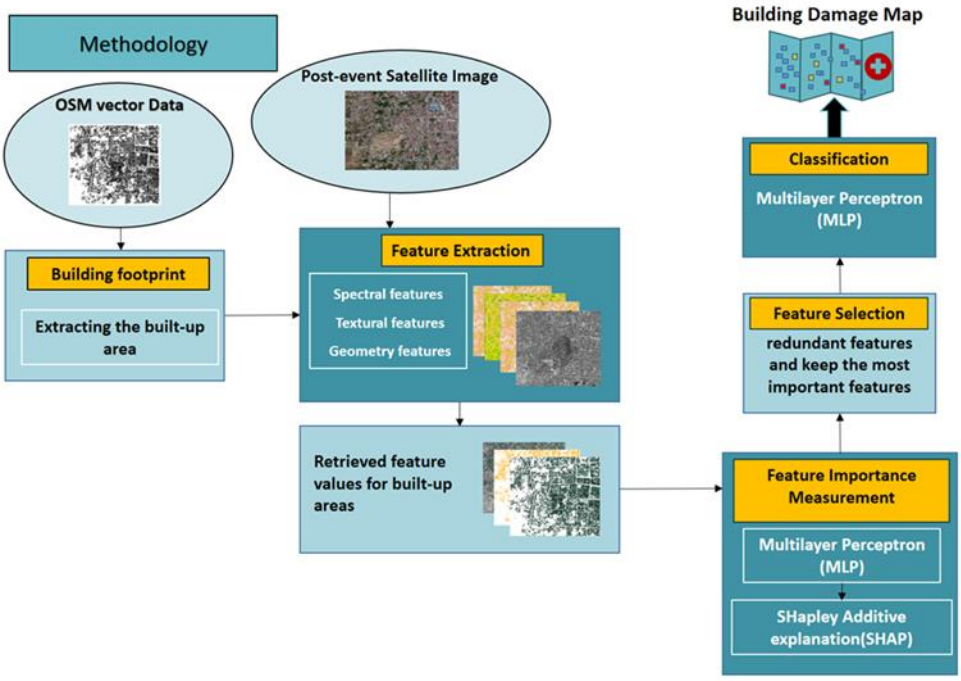

Figure 3.12. Methodological flowchart for objective 2, using Shapley Additive Explanation (SHAP) to build an explainable classification model.

3.11.3. Objective 3

In this subsection, the new image samples from Salina Cruz located in southern Mexico were added to the previous dataset. The study area was divided into two separate training and testing areas. In the next step, the same spectral and textural image features from the training and testing areas were extracted from the satellite image. In order to check the transferability of the proposed MLP, different scenarios were defined. In scenario one, the performance of the trained MLP model from the previous objective was evaluated on the unseen Salina Cruz dataset. In scenario two, the model is trained on both the Palu City training dataset and the Salina Cruz training dataset. Then the performance of the model is evaluated in the Palu City and the Salina Cruz testing areas.

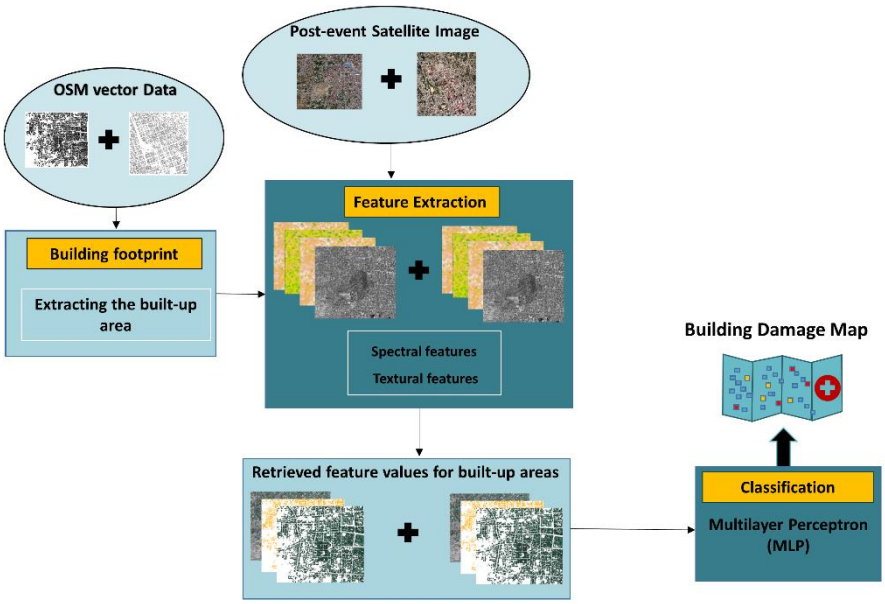

Figure 3.13. Methodological flowchart for objective 3, adding a new dataset from a geographically different area to examine the transferability of the proposed model.

3.12. Summary

The summaries of the characteristics of the study areas and the developed models for earthquake-induced building damage assessment are as follow:

- Because of the high number of collapsed buildings and the availability of the optical post-earthquake satellite image, Palu City/Indonesia and Salina Cruz were selected as two geographically different study areas to develop the model and examine its transferability of the model.
- The post-earthquake satellite images were collected from MAXAR open data program which is publicly available for humanitarian purposes.
- The building footprint dataset for the study area was collected from OSM website. The collected building footprint data was vastly edited to increase its quality.
- The Copernicus building damage dataset from the study area was collected to identify the status of the buildings after the earthquake. Pre-processing was performed on the collected dataset to increase the quality of the dataset. Damaged, possibly-damaged, and no-visible-damage building classes in the Copernicus dataset were considered as non-collapsed and therefore we limited the study to classifying the buildings into collapsed and non-collapsed buildings.

- U-Net model for segmentation and RF, SVM and MLP for classifying the post-earthquake satellite image were introduced and the algorithm, hyperparameters and the features of these models were discussed in this chapter.
- Explainable and interpretable machine learning models are necessary to understand the decisions of the model and also to gain insight about the training dataset.
- Shapley additive explanations value (SHAP) was used to increase the explainability of the model.
- The Dice Coefficient (F1 score) and precision, recall and F1 score were used for evaluating the performance of the segmentation and the classification models respectively.
- QGIS, ArcGIS and Python programming was used to extract, manipulate, develop the model, draw the plots and to create the maps.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Introduction

This chapter illustrates and discusses the results of this study that have been obtained from developing various machine learning models for earthquake-induced building damage assessment. In this chapter, the result of implementing a U-Net segmentation model is presented and the limitations and challenges of using segmentation models will be discussed for segmenting the post-earthquake satellite image from Palu City. In the following subsections, the results of the developed MLP, SVM, and RF using several spectral, textural, and shape features will be presented and compared. The improvement of the MLP model using SHAP will be discussed in this chapter. Moreover, the transferability of the proposed model will be assessed by adding a new dataset from a geographically different area.

The produced earthquake-induced building damage map is presented in this chapter to show the collapsed and non-collapsed buildings in the study area.

4.2. Results of objective 1

4.2.1. U-Net results for image segmentation

The result of the building damage segmentation using the U-Net model was presented in Figure 3.11., including the satellite image samples from the test dataset, ground truth (i.e. masks), and the resulted segmented image samples. The black pixels show the collapsed buildings while the white pixels show the non-collapsed areas including the built-up area and the background. As shown, the segmented test images are not a very good match with the ground truth. However, they could identify the collapsed regions to some extent. Moreover, the boundaries of the buildings weren't correctly separated and some of the individual buildings are missing in the segmented images.

In order to give a quantitative description of the segmentation result, the F1 score was calculated for the test dataset. The test dataset achieved an F1 score of 68% which indicates an average performance of the collapsed buildings' segmentation.

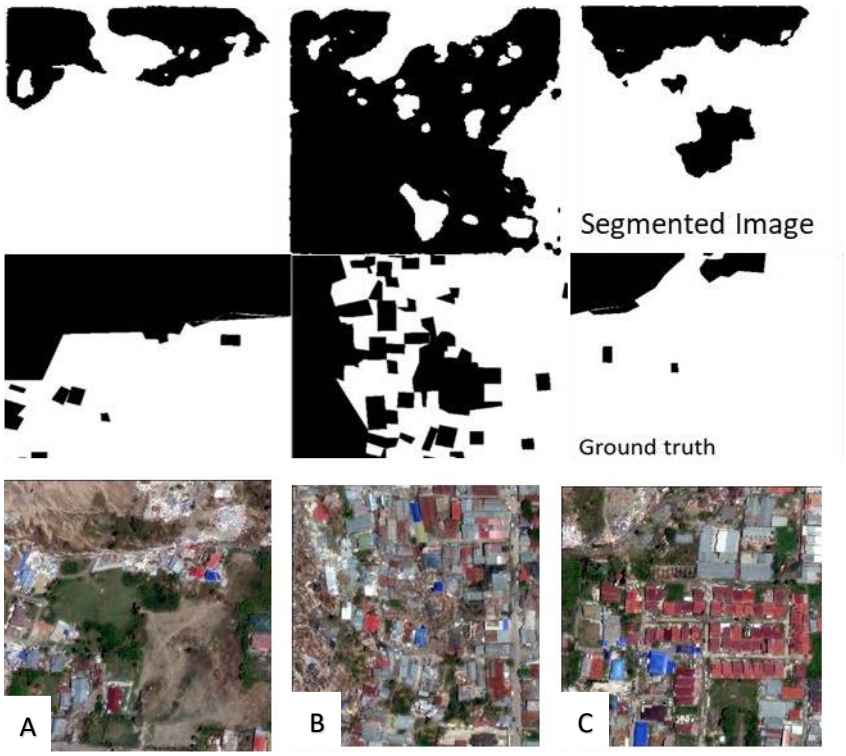

Figure 4.1. A, B, and C are samples of satellite images and their corresponding ground truth and segmented images from the test dataset.

Following the recommendation of several studies in the literature, U-Net architecture performs well when the number of samples in the training dataset is limited (Abdollahi, Pradhan, et al., 2020b; Bai et al., 2018; Ronneberger et al., 2015). Therefore, in this study, U-Net was used to identify the collapsed buildings. Although data augmentation was used to create more image samples, the complexity of the study area along with the limited number of training samples created hurdles for achieving a satisfying segmentation accuracy. For example, sample A in figure 4.1., contains the non-built area (at the bottom right of the image) and the liquefaction area (at the top left of the image) which appear similar in terms of color and texture. Hence distinguishing these two parts is challenging. Despite this complexity, the proposed U-Net could distinguish the non-built area from the liquefaction area.

Among the three samples shown in figure 4.1., sample B contains the most complex scene due to the ground distortion and the density of the buildings. As shown, the segmented image for this image sample is not a good match for the ground truth. Sample C, contains intact building blocks, green areas, and collapsed areas. For this sample, the proposed U-Net could segment most of the collapsed buildings correctly. However, it wrongly segmented some of the non-collapsed buildings as collapsed. The mixture of the collapsed road and the green area at the top of the image and the pattern of the dense buildings in the middle of the image can be the source of error for segmenting this image sample.

Overall, although the segmentation accuracy wasn't significant, this case study showed the feasibility and the potential of using the U-Net algorithm to segment earthquake-induced collapsed regions. In order to achieve better accuracy, more image samples are required. Apart from the accuracy, the process of creating the masks involved manual editing and therefore was time-consuming.

Generally, deep learning models are data-hungry and in most real-world problems the related data is never infinite. On the other hand, although the deep learning model often generalise beyond the specific training dataset, the model decision is heavily dependent on the training dataset. Therefore, there is no guarantee that the model decision will be accurate for the test dataset (Marcus, 2018).

4.2.2. MLP, RF, and SVM results for image classification

In the pre-processing step, the post-event satellite images acquired from the Maxar open data program and OSM polygon dataset were co-registered. The built-up area included 8508 buildings, of which 2135 buildings were labeled as collapsed and the rest were considered non-collapsed. After identifying the built-up area, 15 features, including six spectral, six textural, and three shape features (dataset A) were extracted from all of the building polygons. Spectral and textural features were calculated for each of the three bands in the image (i.e., RGB). Then the binary, SVM, RF, and MLP were trained using these features. Table 4.1 shows the overall accuracy, precision, recall, and F1 score of the developed MLP, RF, and SVM models.

Table 4.1. Overall accuracy, precision, recall, and F1 score parameters over the test area for the MLP, RF, and SVM models.

Model	**Overall Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
MLP	83.68	80.11	52.72	63.59
RF	73.84	99	32.72	6.38
SVM	72.5	98	30.56	5.83

As shown in table 4.2, MLP achieved the highest accuracy compared to two other models. To have a more detailed look at the MLP performance, a confusion matrix has been made for each test area for the proposed MLP model (Figure 4.2).

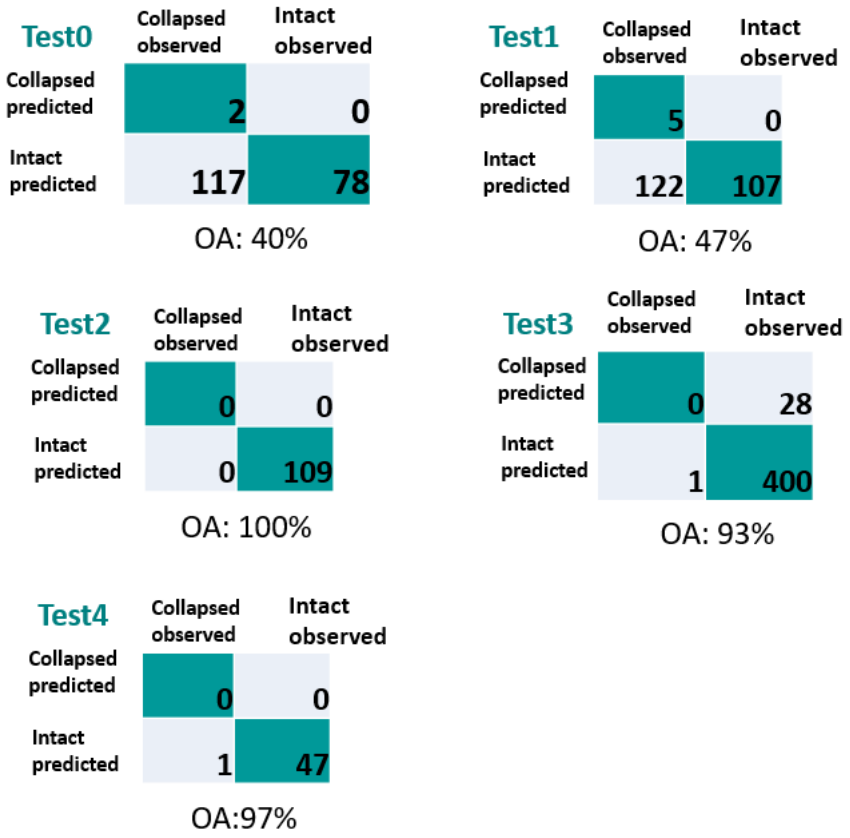

Figure 4.2. The confusion matrix of the result of the proposed MLP model for different test areas.

Regarding the performance of the machine learning-based building damage models, although both MLP and RF have similar overall accuracies, RF failed to detect almost all of the collapsed buildings. As discussed in section 3.10.2., detecting the collapsed buildings is of more importance therefore recall is a suitable evaluation metric to compare the performance of the models.

As shown in table 4.1, RF and SVM have achieved high precision of 99% and 98% respectively. However, the recall for those models is 32.72% and 30.56% which is quite low. High precision and low recall result from a small true positive and large false negative meaning that the model incorrectly classified most of the collapsed buildings. MLP outperformed other models with 80.11% precision and 52.72% recall.

The classification results show that the proposed MLP model performed well in detecting non-collapsed buildings. However, it wasn't very successful in detecting most of the collapsed buildings in test 0 and test 1. One reason behind the high number of true-negative results for MLP could be the fact that the dataset is skewed toward non-collapsed buildings. The high accuracy is more evident in test 2 and test 4, as they have more non-collapsed

buildings. The other reason for the misclassification of collapsed buildings could be related to using irrelevant image features. To have a better understanding of which image features contribute to the classification of each building an explainable MLP model should be developed. Using SHAP as an explainable algorithm opens avenues for improving the MLP model's performance.

4.3. Results of objective 2

4.3.1. SHAP result

Regarding feature importance, as shown in the initial SHAP summary plot (Figure 4.3), the spectral features have more impact on distinguishing the collapsed and non-collapsed buildings. This result is likely related to the significant difference in the roof colour of collapsed and non-collapsed buildings. In contrast, the shape features have the least impact on the model output. The low contribution of the shape features in the classification can be derived from the high similarity in the intact building shapes, which are mostly rectangles. Therefore, due to the low variance in the shape features, these features do not provide enough information to contribute to the classification, which became evident in the SHAP summary plot (Figure 4.3). Moreover, the majority of the collapsed buildings occurred in the liquefaction area, where the pre-event building footprint was used to indicate the buildings as objects for analysis. Therefore, although there might be a correlation between some of the building shapes in the study area and their condition after the earthquake, generally the shape features are considered irrelevant in this study and can be removed from the model input dataset. Further, GLCM-homogeneity, GLCM-dissimilarity, and GLDV-entropy are likely to be redundant due to the low contribution of these features.

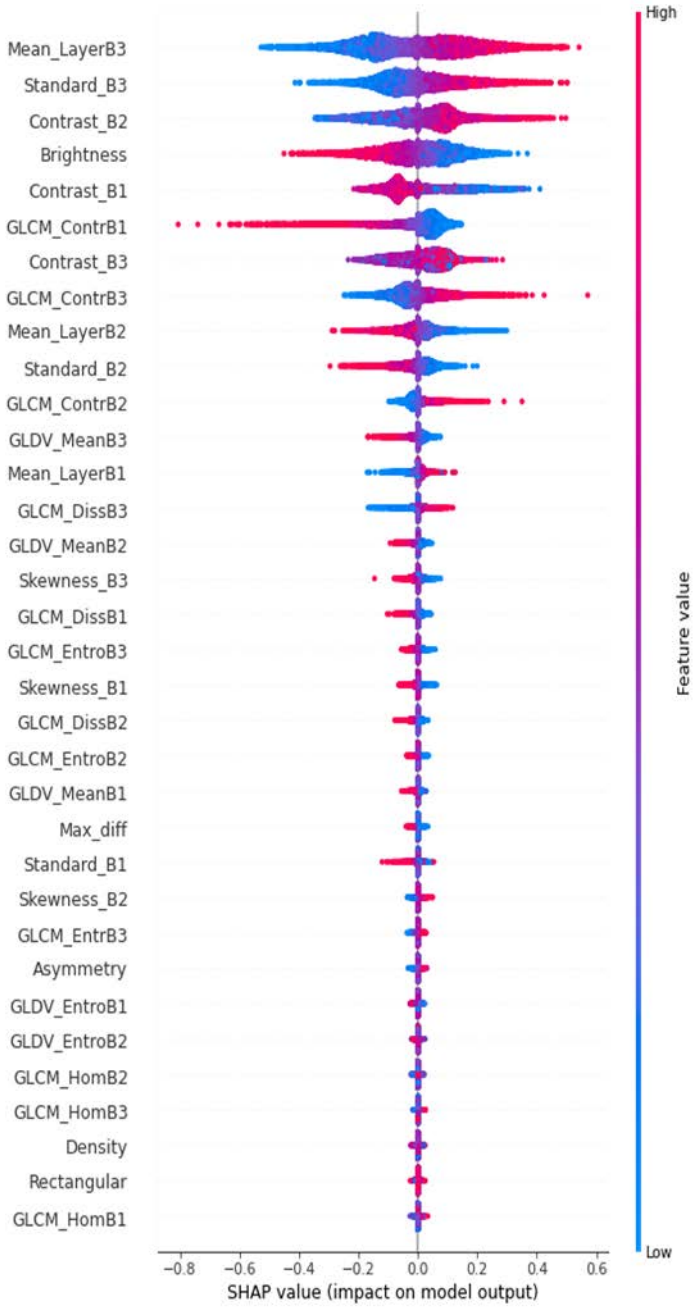

Figure 4.3. SHAP summary plot for MLP model using dataset A (including 15 spectral, textural, and shape features)

After removing the irrelevant and redundant features, the eight remaining spectral and textural features (dataset B) were fed into the model; and the SHAP values for all the features were calculated. Figure 4.4 shows the SHAP summary plot for the MLP model that has been trained on dataset B. As expected, according to the new SHAP summary plot in Figure 4.4, the spectral features were considered the most important features for this classification. Considering the low rank of the mean-layer for band 2 compared to the mean-layer for the other bands, one may argue that the mean-layer is not generally an

important feature. However, the variance of the green band (i.e., band 2) is significantly lower than the other bands, leading to its lower contribution to the model outcome.

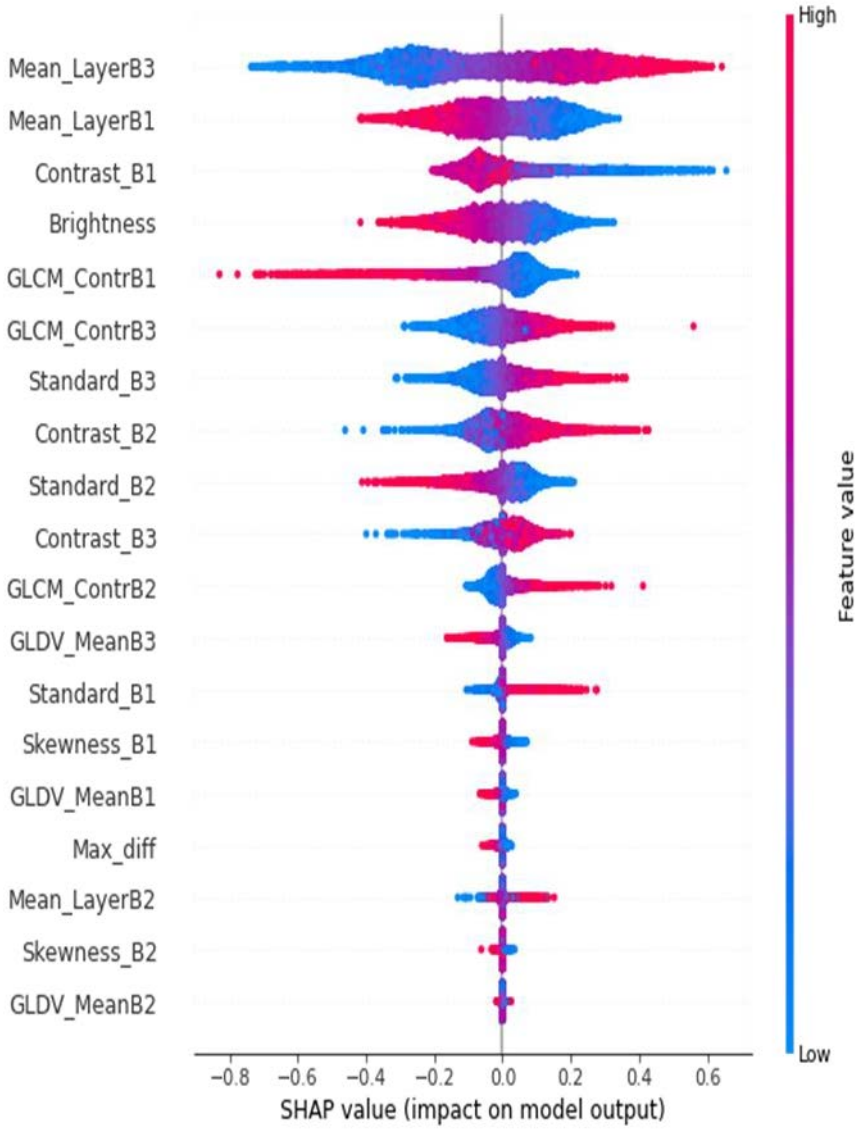

Figure 4.4. SHAP summary plot for MLP model using dataset B (including eight spectral and textural features)

The blue, purple, and red colour in the SHAP summary plot denotes the low, average, and high value of each feature for all the samples in the training dataset, respectively. The values in the X-axis represent the SHAP values of each observation. Therefore, the relationship between the features and the target (i.e., the building being identified as collapsed) can be explored using this plot. As shown in figure 4.4, it can be concluded that, generally, the increase in the high-rank features for band 3 (e.g., mean layer, contrast, GLCM contrast) leads to the increase in SHAP value. However, the increase in the features related to band 1 decreases the likelihood of the building to be collapsed. The reason for this might be because of the dominant red colour of the roof for most non-collapsed

buildings. It is worth mentioning that the results shown above are the general feature analysis for the majority of the samples in the dataset, and the features may impact the classification output differently for individual buildings. In order to locally interpret the effect of the features on the prediction, a force plot was used to investigate the impact of the features on classifying each building (Figure 4.5). As shown, the MLP model could classify the buildings (a) and (c) correctly with a 6% and 16% probability of the building to be collapsed, respectively.

The MLP misclassified building (b) as collapsed with a 63% probability. The source of the error could be due to the angle of the image, which did not allow the building footprint to be placed correctly on the building image. Building (d) was labeled as collapsed based on the guideline in Table 2.3. However, it was misclassified as non-collapsed with a 26% probability. As shown (Figure 4.5d), half of the building roof is collapsed and the other half is completely intact, which makes classifying this building challenging. However, the model output (0.26) is very close to the base value (0.32), indicating that the impact of the features increasing the output (shown by the red arrow) and the features decreasing the output (shown by blue) are almost equal. Therefore, although the result shows the building is non-collapsed, the model is not as confident as for buildings (a) and (c).

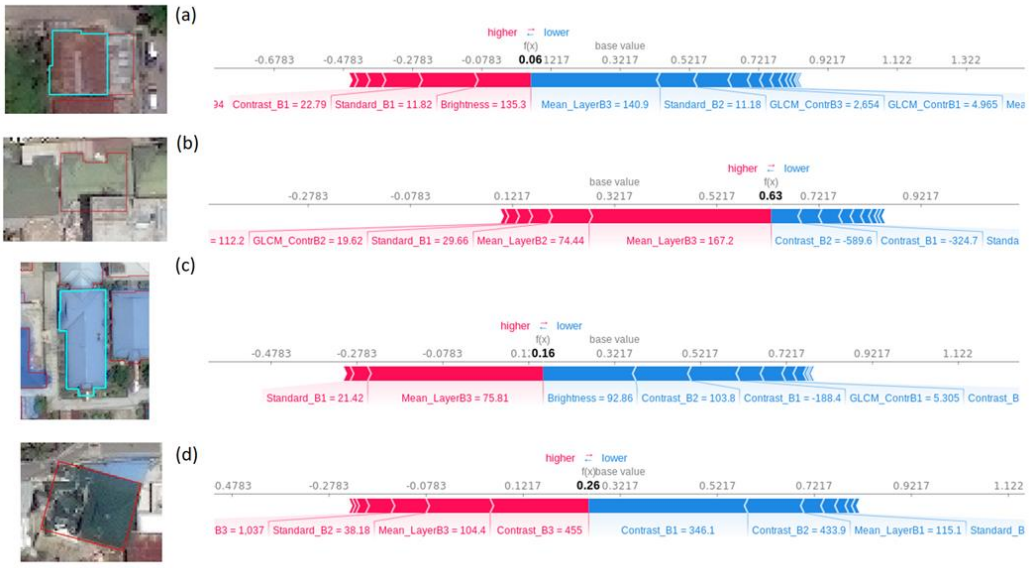

Figure 4.5. Comparison of the force plots of different buildings in the test area (a–d).

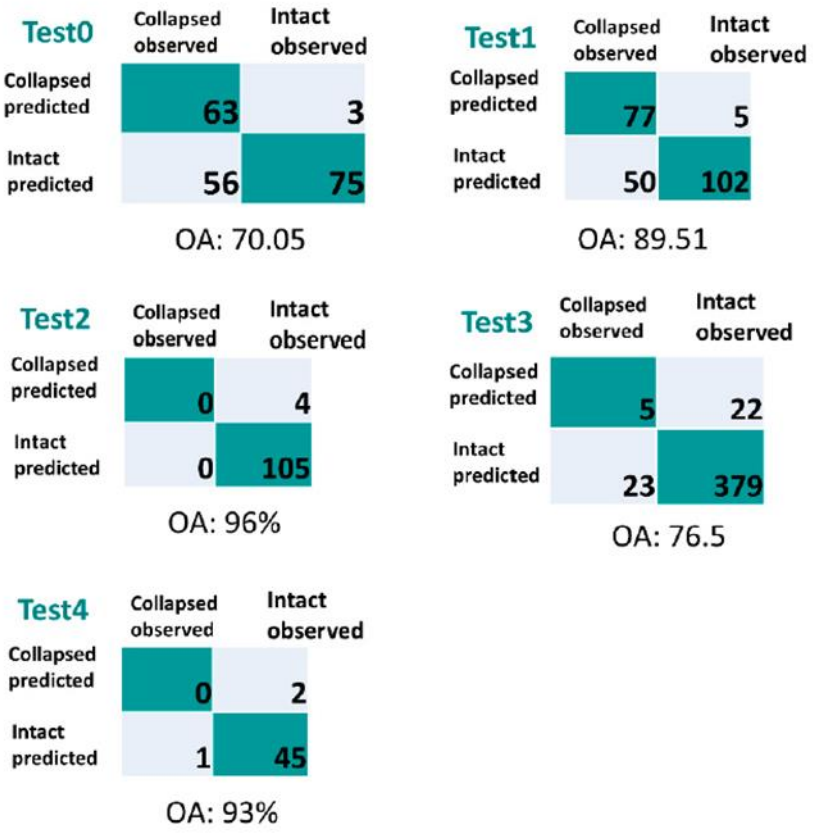

Figure 4.6. The confusion matrix of the result of the proposed MLP model (using nine spectral and textural features) for different test areas.

4.3.2. Comparing the performance of MLP and RF

Since in the analysis in objective 1, MLP and RF performed better than SVM, the effectiveness of the proposed explainable MLP model was compared with Random Forest (RF) model after SHAP analysis. The Random Forests (RF) model, developed in 1990 (Breiman, 2001), is well-known for its capability in classification and regression and commonly used for measuring the importance of the features (Janitza et al., 2013). RF model was found to be accurate and stable in classifying high-dimensional data (Ma et al., 2017). Further, this model has been used in many building damage detection studies (Adriano et al., 2019; Bialas et al., 2019). Therefore, in this study, the RF's performance, along with its built-in feature importance scores, was compared with the proposed MLP model.

Dataset A, including 15 spectral, textural, and shape features, was fed into the RF model with 5000 decision trees. It is worth mentioning that different numbers of decision trees ranging from 200 to 5000 were examined, and the best result was achieved by 5000 trees.

As shown in Figure 4.7., the RF model ranked the shape features as the least important, which is similar to the results from the MLP SHAP summary plot. However, the most important features contributing to the RF model are quite different from that of MLP. Further, according to the RF plot, there was no significant difference between the rankings of shape features and some of the low-ranked spectral and textural features. This shows that features such as brightness were not able to contribute to the model output, as opposed to in MLP, wherein brightness was ranked as one of the important features.

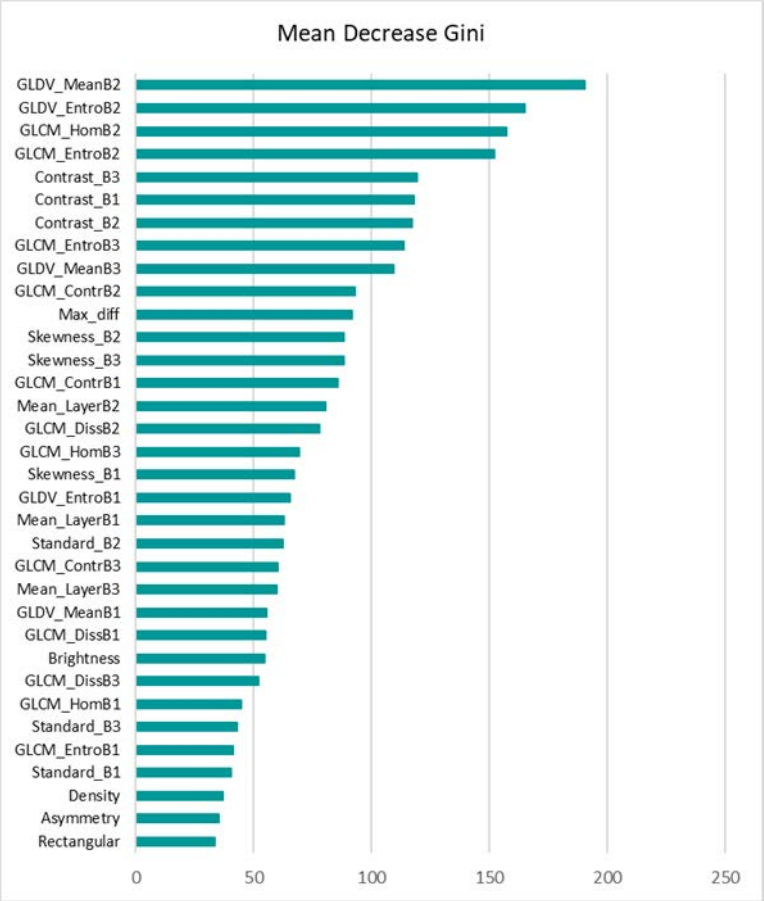

Figure 4.7. VIM plot for the RF model using 15 spectral, textural, and shape features.

Regarding the performance, although both MLP and RF have similar overall accuracies, RF failed to detect almost all of the collapsed buildings. Further, using dataset B, RF did not improve the model accuracy. However, the model's performance improved slightly after removing the three low-ranked shape features. Both the RF ranking plot and the SHAP summary plot show the importance of the features. However, the SHAP summary plot shows the direction of the correlation (feature effect) as well.

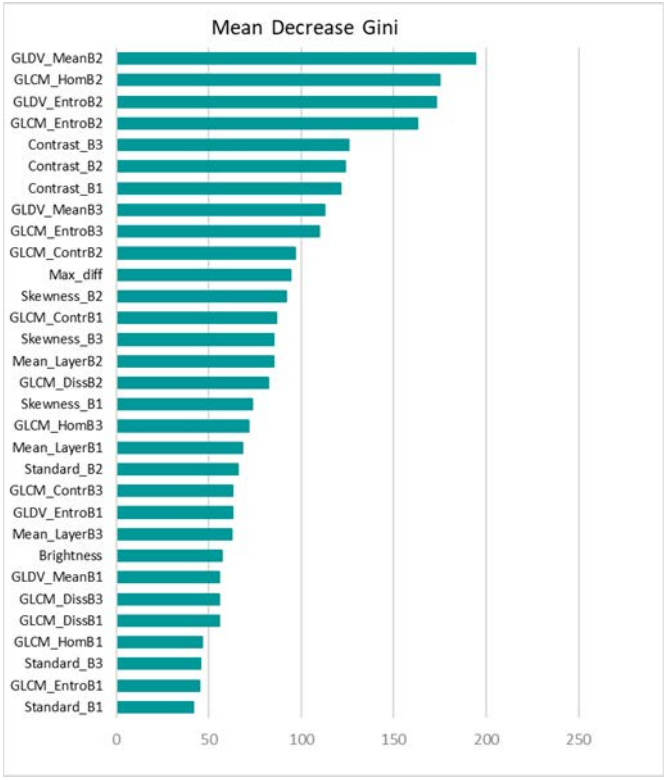

Figure 4.8. VIM plot for the RF model using 8 spectral and textural features

In order to evaluate the performance of the model, the achieved results should be compared to the ground truth dataset. The building-states dataset collected through ground inspection after the earthquake is usually considered the ground truth. However, in this study, access to the ground-inspection dataset was not possible, and therefore, the Copernicus grading map was considered the ground truth. In this study, the confusion matrix was created and its parameters (i.e., overall accuracy, precision, recall, and F1 score) were calculated to quantitatively summarise the performance of the models (Sokolova & Lapalme, 2009). The evaluation metrics were calculated for the MLP and RF models using 15 features (dataset A), the MLP that was built after removing some of the features (dataset B), and the RF model after removing the three low-ranked features. Table 4.2 summarises the overall accuracy, precision, recall, and F1 score parameters over the test area for the abovementioned models.

As shown in Table 4.2, the MLP outperforms the RF in correctly classifying the collapsed buildings.

Table 4.2. Overall accuracy, precision, recall, and F1-score parameters over the test area for the proposed MLP.

Accuracy	Overall Accuracy (%)	Precision (%)	Recall (%)
MLP (using dataset A)	73.55	87.5	2.54
MLP (using dataset B)	83.68	80.11	52.72
RF (using dataset A)	73.5	99	2
RF (after removing the 3 low-ranked features)	73.84	99	32.72

4.4. Results of objective 3

In this subsection, the performance of the proposed MLP model that will be trained on the Palu training dataset will be evaluated on the unseen Salina Cruz dataset. Therefore, the previous Palu training dataset is considered the training dataset, and the whole Salina Cruz dataset is considered a test dataset.

As shown in Table 4.3, a confusion matrix has been made for the Salina Cruz area to measure the model performance. The Overall accuracy, precision, recall, and F1-score parameters were calculated as common evaluation metrics to show the performance of the model (Table 4.3). As shown, the overall accuracy of the proposed MLP model for the unseen Salina Cruz dataset is about 62% which indicates that the model could classify more than half of the buildings correctly. Further, the confusion matrix shows that the model could classify more than half of the collapsed buildings resulting in a high recall. However, a large number of non-collapsed buildings were classified as collapsed. The resulted high recall and low precision show that the model is skewed toward the collapsed class. The reason behind this can be a large number of collapsed buildings in the Palu City training dataset compared to the small number of non-collapsed buildings in the Salina Cruz dataset.

In general, although the overall accuracy seems low, it should be considered that the model hasn't been fine-tuned using the Salina Cruz dataset and the whole test dataset for the model is new.

Table 4.3. Overall accuracy, precision, recall, and F1-score parameters over the Salina Cruz/ Mexico dataset as the test area for the MLP trained on Palu City/Indonesia dataset and the corresponding confusion matrix.

	Collapsed Observed	**Intact Observed**
Collapsed Predicted	154	1620
Intact Predicted	139	2715

Accuracy	OA:	61.99%
	Precision:	8.6%
	Recall:	52.5%
	F1:	14%

To investigate the performance of the model on a dataset that has been enriched with samples from both Palu City and Salina Cruz datasets, a new training dataset was built. In this experiment, the Salina Cruz dataset was divided into training and test datasets, as mentioned in 3.2.2 section. Then the Salina Cruz training was added to the Palu City training dataset to make a new training dataset. Next, the proposed MLP model was trained on the new training dataset and evaluated on the Salina Cruz test dataset.

A confusion matrix has been made for the Salina Cruz test area to measure the model performance. The Overall accuracy, precision, recall, and F1-score parameters were calculated as common evaluation metrics to show the performance of the model (Table 4.4). As shown, more than 90% of the buildings were classified correctly. About 90% of the non-collapsed buildings were classified correctly as well. As mentioned, the number of collapsed buildings in the Salina Cruz dataset is far less than that in the Palu City dataset and it reflects in the test area as well. The number of collapsed buildings in the Salina Cruz test area is only 15 but as seen, 80% of the collapsed buildings were classified correctly which is very significant.

Table 4.4. Overall accuracy, precision, recall, and F1-score parameters over part of the Salina Cruz/ Mexico dataset as the test area for the MLP trained on Palu City/Indonesia and Salina Cruz/Mexico training dataset and the corresponding confusion matrix.

	Collapsed Observed	**Intact Observed**
Collapsed Predicted	12	8
Intact Predicted	3	130

Accuracy	OA:	92%
	Precision:	60%
	Recall:	80%
	F1:	68%

To investigate the performance of the model in Palu City's test area, a confusion matrix has been made for the Palu City test area. The Overall accuracy, precision, recall, and F1-score parameters were calculated as common evaluation metrics to show the performance of the model (Table 4.5). As shown, 80% of all buildings and more than 65% of collapsed buildings were classified correctly. Further, there is no major difference between the percentage of precision and recall.

Regarding classifying the collapsed buildings, the MLP that was trained on both Palu City and the Salina Cruz datasets performed better compared to the MLP that was only trained on the Palu City dataset. The confusion matrix shows that the new MLP could classify 181 collapsed buildings whereas this number is 145 for the MLP that was trained only on the Palu City training dataset.

In terms of classifying non-collapsed buildings across the Palu City test areas, the MLP that was trained on only the Palu City training dataset performed slightly better. However, the latter MLP model is more reliable as it has been trained on a larger dataset, and it has

learned the features from two different geographic study areas. In general, the results show that the proposed MLP model is transferrable and the performance of the model improves when a new training dataset is added to the previous training dataset.

Table 4.5. Overall accuracy, precision, recall, and F1-score parameters over Palu City/Indonesia test dataset for the MLP trained on Palu City/Indonesia and Salina Cruz/Mexico training dataset and the corresponding confusion matrix.

	Collapsed Observed	**Intact Observed**
Collapsed Predicted	181	102
Intact Predicted	94	640

Accuracy	OA:	80.73%
	Precision:	63.95%
	Recall:	65.81%
	F1:	64.87%

4.5. Results of Building damage mapping

Building-damage mapping is crucial to support post-earthquake rescue and relief activities. Further, in the long term, knowledge regarding details of damage after an earthquake aids decision-makers in better designing buildings and infrastructure which can withstand destruction in future seismic events. In this research two building damage maps were generated to show the classification results of analysis for objectives 2 and 3. In objective 2, the model was trained on part of the Palu City building dataset and was tested on five different built-up areas in the study area. The generated damage map for the analysis in objective 2 is shown in figure 4.9., where red and green indicate the classified collapsed and non-collapsed buildings, respectively. The correctly classified and misclassified buildings in the test areas are shown in blue and orange, respectively.

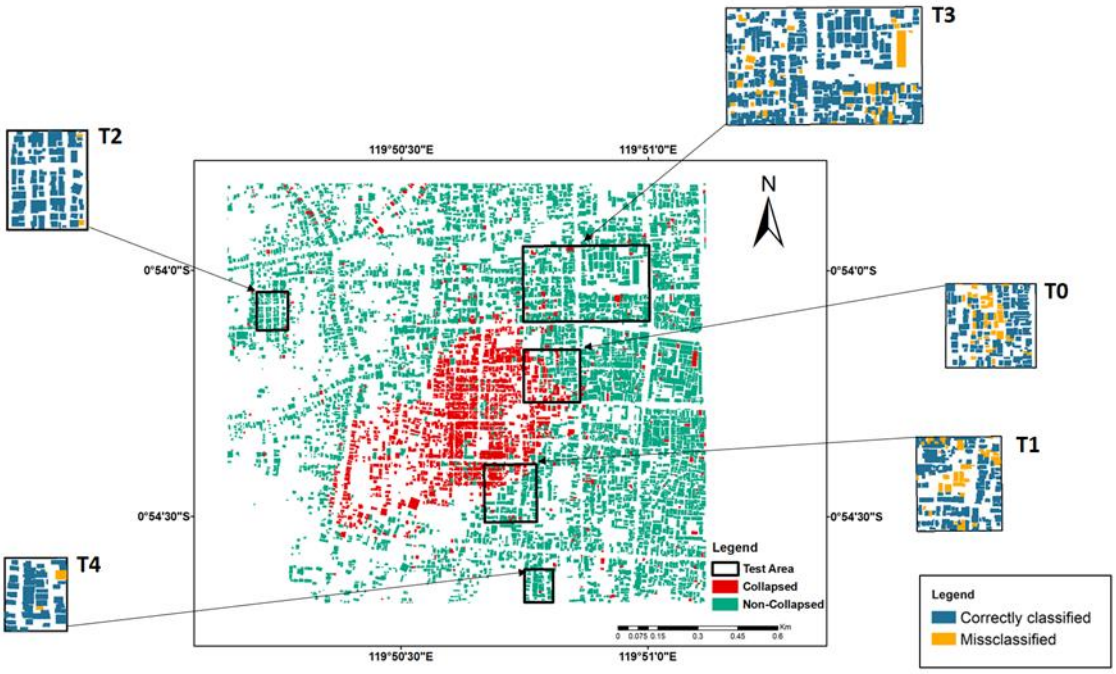

Figure 4.9. Results of building damage mapping in objective 2. The blue-orange map is the color-coded map of the correctly classified and misclassified for all five test study areas.

In objective 3, the proposed MLP model was trained on a larger training dataset containing both Palu City and Salina Cruz training datasets. The performance of the model then evaluated using the Salina Cruz test area. Figure 4.10., shows the resulted generated map for the analysis in this objective. Red and green indicate the classified collapsed and non-collapsed buildings, respectively. The correctly classified and misclassified buildings in two test areas are shown in blue and orange, respectively.

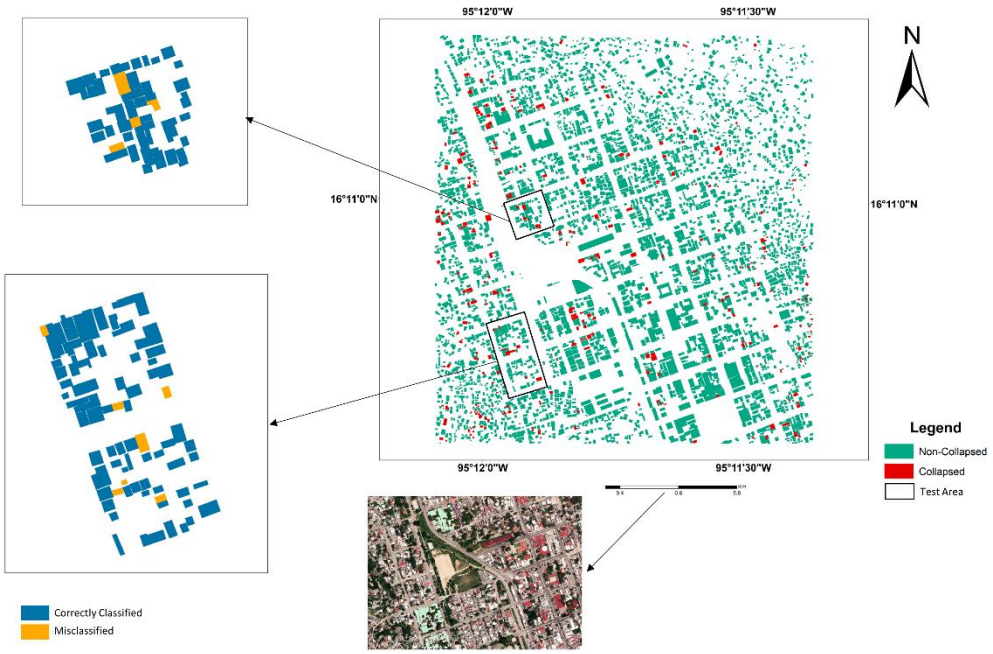

Figure 4.10. Results of building damage mapping in objective 3. The blue-orange map is the color-coded map of the correctly classified and misclassified test study areas.

4.6. Summary

In chapter 4, the development of U-Net, MLP, RF, and SVM models, the explainable MLP model, and assessing the transferability of the model leads to the following results:

- Image patches and the corresponding masks were created for training dataset. Further, augmentation was used to increase the size of the training dataset.
- For the first objective, a U-Net model was developed to segment the post-earthquake satellite image from the Palu City study area. The developed U-Net model segmented the image with 68% accuracy which showed that the model is only suitable for identifying the collapsed region.
- Using 15 different spectral, textural, and shape image features from the Palu City dataset, MLP, RF, and SVM models were developed to classify the collapsed buildings. MLP achieved the highest accuracy compared to two other models.
- SHAP was used to rank the contribution of the image features on the model output using the SHAP summary plot. The plot showed that the spectral and textural features are the most important features in classifying the collapsed buildings. The plot also showed that the shape features do not contribute to the classification and therefore can be removed to decrease the complexity of the training dataset.

- The MLP model performance increased after removing the redundant features from the training dataset. Further, SHAP summary plot showed that the spectral features played an important role in classifying the collapsed buildings. To gain more insight about the model decisions, SHAP force plot was used for some of the buildings. The SHAP analysis showed that there is correlation between the building roofs' spectral feature and the status of the building.
- The same spectral and textural features were extracted from the post-earthquake Salina Cruz, Mexico dataset to create a new dataset. The performance of the MLP model was evaluated on the new dataset to assess the model's transferability. The trained MLP could classify about 62% of the unseen buildings in the Salina Cruz dataset. Further, the model was trained on both the Salina Cruz and Palu City dataset and tested on test datasets from both study areas. The new MLP model could achieved about 92% and 82% overall accuracy.
- Building damage map was created for both the Palu City and the Salina Cruz dataset.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS

5.1. General conclusions

Deploying machine learning-based frameworks for earthquake-induced building damage mapping in real-world scenarios is largely limited to the manual interpretation of optical satellite images. Although manual interpretation is costly and labour-intensive, it is preferred over automatic and semi-automatic building damage mapping using ML-based frameworks. These frameworks in this domain are promising, yet are not reliable for several reasons, including but are not limited to the use of site-specific image datasets, dependency on the availability of both pre and post-earthquake images, lack of transferability and transparency of the machine learning model, the use of irrelevant descriptor features in building the model and the lack of quality in the labeled images. This research employed single post-earthquake optical satellite images and the open-source dataset for building footprint developed an explainable model and examined the transferability of the model to fill the gaps in the previous studies. Data pre-processing was performed to assess the quality of the data, increase the size of the dataset and extract the relevant image features. Several machine learning models (U-Net, MLP, RF, and SVM) were trained and their efficiency was evaluated. The results showed that the developed MLP model outperformed the other models. SHAP algorithm was used to build an explainable MLP model and therefore to better understand the classification model decision for each observation. This led to improving the model performance by removing the irrelevant image features and decreasing the complexity of the model. Finally, the transferability of the model was assessed on an image dataset from a different geographic region. The results showed that the model is transferrable and the performance of the developed MLP model improves when it is trained on a larger dataset.

In addition, the proposed model outperforms the other comparative models in terms of accuracy and provides an explainable and transferrable model. The proposed framework can be used as a guide for future research to show how to gain insight from the explainability of the model to improve the dataset and hence the model performance by either editing the quality of the labels or adding new image samples to the data. In short, the suggested approach is both meaningful and efficient for operational post-earthquake building damage mapping.

5.2. Specific conclusions

In this subsection the conclusion of each objective will be discussed.

5.2.1. Conclusions of objective 1

In this study, a single post-earthquake optical satellite image was used to develop several machine learning segmentation and classification model. In the first step, patches of images were created from the Palu City image dataset. Then the collapsed buildings and the debris in the image patches were masked manually to create a corresponding ground truth image dataset. Then a U-Net segmentation model was trained using 70% of the image patches. The performance of the model was evaluated on the remaining image patches. The segmentation result shows that the U-Net model could classify 68% of the collapsed pixels correctly. Visual comparison of the segmented images and the grand truth (masks) shows that the developed U-Net could identify the collapsed regions to some extent and it is not an accurate model to classify the individual buildings. Although augmentation was done to increase the number of samples in the training dataset, the U-Net model requires more samples to achieve a good segmentation result. Moreover, annotating the image patches to build the masks dataset is time-consuming and requires manual editing which is not suitable for rapid damage mapping.

To investigate the performance of the MLP, RF, and SVM classification model, several textural, spectral, and shape features were extracted from the image dataset. The image was divided into a training area and five different test areas. The models were trained and their classification accuracies were calculated. The result of the evaluation metrics showed that the MLP outperformed the other models.

In this study classification models using the image features, were preferred over the U-Net segmentation model because they required less pre-processing and the classification models led to a more promising result. The results of this study showed that classifying the earthquake-induced collapsed buildings is possible with using a single post-earthquake satellite image. The conclusion of this study can be a good source of information for the model developers and the researchers in the field of natural hazard assessment.

5.2.2 Conclusions of objective 2

For objective 2, a multilayer perceptron (MLP) was employed to produce a damage map using a single post-event satellite image captured after Indonesia, Palu earthquake in 2018. Each building was used as an object in the image using the available open street map (OSM)

vector data for this region. Then MLP model was trained using the spectral, textural, and shape features to distinguish the collapsed and non-collapsed buildings. Furthermore, SHAP (SHapley Additive exPlanation) algorithm was employed to analyse the impact of the features on the model output.

Based on the SHAP summary plot, the irrelevant and redundant features were removed from the model and the MLP was retrained using the new set of features resulting in better performance of the model in detecting the collapsed buildings. The SHAP force plot was utilised for some of the building samples in the dataset to analyse the effect of the features on classifying the individual buildings. The performance of the model was evaluated over five diverse test areas (i.e. in terms of the number of collapsed and non-collapsed buildings within the area) using the usual evaluation metrics. The model achieved an overall accuracy of 83.68%, precision of 80.11%, and an F1 score of 63.59%. The model performed better in the test areas where the number of non-collapsed buildings is dominant.

The SHAP feature analysis results showed that in general, the spectral features have the most impact on detecting the collapsed buildings in the study area. As discussed in section 5, this can be the source of error in deploying the model for a geographically different study area where the spectral information of the roof image is different. Therefore, there is a need to create a training dataset that represents the variety of buildings and different collapsed types from geographical areas. As the AI industry is shifting from model-centric to data-centric, there is a necessity to focus more on the dataset being used to train the model. To this end, SHAP enables us to analyse the data in detail and guides us to build a more generic database and AI model in the future. Therefore, future work will examine the impact of the features extracted from the images covering different urban areas. The current study was a small step towards the use of explainable AI for post-earthquake building damage mapping. Regarding the need for building a large, diverse training dataset in this domain, explainable AI can be used for detecting the limitations in the training dataset that has led to a biased result. The result of this study can be used as a guideline for the researchers and the model developers in the field of natural hazard assessment.

5.2.3 Conclusions of objective 3

This study showed that the MLP model that was trained on the Palu City training dataset could classify about 62% of the unseen Salina Cruz dataset correctly. The classification accuracy increased dramatically when the model was trained on an enriched training dataset containing the data from both Palu City and Salina Cruz datasets. The later model could classify about 92% of the buildings in the Salina Cruz test dataset correctly. To further

assess the performance of the model, the model was evaluated on the Palu City test dataset. The confusion matrix and the evaluation results showed that the new MLP model outperform the previous MLP model that was trained only on the Palu City dataset. Further, the model is more reliable as it was trained on two geographically different training datasets. The study shows that the developed MLP model is transferable and can be utilized to model a new unseen post-earthquake dataset. Further adding new data samples to the training dataset can equip the trained MLP model with more information about the variety of different image samples. Therefore, the model will perform more efficiently in classifying a new image dataset.

5.3. Research drawbacks and limitations

Several limitations were encountered in objectives 1 and 3. Most of the limitations of this study were rooted in the quality and the size of the dataset. For example, the pre-processing of the data including cleaning and labelling can be challenging and time-consuming. Human annotation typically takes a long time, and it is costly. In terms of reducing the annotating's time, the task is usually done by a team of trained annotators to expedite this process. There are also tools to support image annotating. One of the common open-source tools used in large-scale projects is Computer Vision Annotation Tool (CVAT). Such tools support the annotation process by enabling drawing complex shapes and providing a structured labelling system. Therefore, the annotators can assign the correct labels to image artifacts. However, each tool requires a trained operator to annotate the image dataset making the human and the annotating technology equally important.

As mentioned in chapter 3, the co-registration of the OSM building footprint and the satellite image was a challenge and it involved a significant amount of editing. Most of the OSM building footprint polygons were required to be updated or shifted to cover the individual buildings.

Labeling the individual buildings according to the damage level standard was a necessary but time-consuming step in the pre-processing. Each building condition was required to assess visually to be assigned to the correct collapsed or non-collapsed buildings.

In addition, although a fair amount of image samples was used in this study, the number of image samples was considered limited particularly for training the U-Net model. Further, the number of collapsed buildings were far less than the number of the non-collapsed building which makes the dataset imbalanced. Although the oversampling technique was used to increase the minority building class, the accuracy did not improve significantly.

5.4. Recommendations for future work

In this research, the proposed models were implemented and all three objectives were achieved. Moreover, further work can be done on earthquake-induced building damage assessment by applying a detailed dataset using the latest models of artificial intelligence. The recommendations for future work are:

1. To build a database of geographically diverse post-earthquake image samples with consistence labeling, particularly from earthquake-prone areas. As and when more comprehensive datasets would be available, the ultimate goal of a generalized transferable artificial intelligence-based model can be achieved.

2. To explore the capabilities of different convolutional neural networks filters along with the mentioned features in this research (for example edge detection etc.) in distinguishing the collapsed and non-collapsed buildings.

3. To build a Rashomon dataset (i.e. a set of machine learning models that perform equally well for a specific classification/prediction task) to select the best classification model. Further, the capabilities of the Fuzzy classification model can be investigated for classifying earthquake-induced collapsed buildings.

4. To contribute to updating and editing the open street map (OSM) project, especially in developing countries where the updated building footprint might not be available.

5.5. Summary

• Classifying post-earthquake collapsed buildings is possible by using a single post-earthquake optical satellite image. However, the quality of the training dataset plays an important role in the accuracy of the classification and therefore an extensive pre-processing is required for creating a training dataset.

• Although U-Net could segment the image to some extent, larger training dataset is required to achieve a satisfying image segmentation accuracy. Further, the creating the annotated image patches is time-consuming.

• The proposed MLP model outperformed the other machine learning models developed in this study to classify the collapsed buildings.

• Explainable MLP model gives insight about the reasons behind the model's decisions and also about the dataset. Explainability frameworks such as SHAP, sheds light on the correlation of the image features and the status of the buildings.

• The developed MLP model is transferrable and the accuracy of the classification can increase by fine-tuning the model using the new training dataset from a geographically different area.

• Limited number of image samples, imbalanced dataset, limiting the study to only two geographic areas and extensive data pre-processing are considered as the limitation of this research.

• The study recommends building a comprehensive annotated image dataset that covers diverse geographical areas. Developing other explainable and interpretable models such as Fuzzy classification models and using convolutional neural networks filters will include in the future research.

References

Aali, H., Sharifi, A., & Malian, A. (2019). Earthquake Damage Detetction using Satellite Images (Case Study: Sarpol-Zahab Earthquake). *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 1-5.

Abdollahi, A., Pradhan, B., & Alamri, A. (2020a). VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data. *Ieee Access*, *8*, 179424-179436.

Abdollahi, A., Pradhan, B., & Alamri, A. M. (2020b). An ensemble architecture of deep convolutional Segnet and Unet networks for building semantic segmentation from high-resolution aerial images. *Geocarto International*, 1-16.

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2020). Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. *Remote sensing*, *12*(9), 1444.

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *Ieee Access*, *6*, 52138-52160.

Adriano, B., Xia, J., Baier, G., Yokoya, N., & Koshimura, S. (2019). Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 sulawesi earthquake and tsunami in Palu, Indonesia. *Remote sensing*, *11*(7), 886.

Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., & Al-Fuqaha, A. (2020). Developing Future Human-Centered Smart Cities: Critical Analysis of Smart City Security, Interpretability, and Ethical Challenges. *arXiv preprint arXiv:2012.09110*.

Anders, C. J., Weber, L., Neumann, D., Samek, W., Müller, K.-R., & Lapuschkin, S. (2022). Finding and removing clever hans: Using explanation methods to debug and improve deep models. *Information Fusion*, *77*, 261-295.

Anniballe, R., Noto, F., Scalia, T., Bignami, C., Stramondo, S., Chini, M., & Pierdicca, N. (2018). Earthquake damage mapping: An overall assessment of ground surveys and VHR image change detection after L'Aquila 2009 earthquake. *Remote sensing of environment*, *210*, 166-178.

Anuar, N., & Sultan, A. B. M. (2010). Validate conference paper using dice coefficient. *Computer and Information Science*, *3*(3), 139.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., & Benjamins, R. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82-115.

ASEAN. (2018). *Situation Update No.15-Final 7.4 Earthquake and Tsunami.* ASEAN Coordinating Centre for Humanitarian Assistance on Disaster Management. Retrieved 20

November 2020 from https://ahacentre.org/situation-update/situation-update-no-15-sulawesi-earthquake-26-october-2018/

Bai, Y., Mas, E., & Koshimura, S. (2018). Towards operational satellite-based damage-mapping using u-net convolutional network: A case study of 2011 tohoku earthquake-tsunami. *Remote sensing*, *10*(10), 1626.

Barghout, L. (2015). Spatial-taxon information granules as used in iterative fuzzy-decision-making for image segmentation. In *Granular computing and decision-making* (pp. 285-318). Springer.

Bartels, S. A., & VanRooyen, M. J. (2012). Medical complications associated with earthquakes. *The Lancet*, *379*(9817), 748-757.

Bhangale, U., Durbha, S., Potnis, A., & Shinde, R. (2019). Rapid Earthquake Damage Detection Using Deep Learning from VHR Remote Sensing Images. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium,

Bhatt, U., Antorán, J., Zhang, Y., Liao, Q. V., Sattigeri, P., Fogliato, R., Melançon, G., Krishnan, R., Stanley, J., & Tickoo, O. (2021). Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,

Bialas, J., Oommen, T., & Havens, T. C. (2019). Optimal segmentation of high spatial resolution images for the classification of buildings using random forests. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101895.

Bittner, K., & Korner, M. (2018). Automatic large-scale 3d building shape refinement using conditional generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,

Blaschke, T. (2005). Towards a framework for change detection based on image objects. *Göttinger Geographische Abhandlungen*, *113*, 1-9.

Bolon-Canedo, V., & Remeseiro, B. (2019). Feature selection in image analysis: a survey. *Artificial Intelligence Review*, 1-27.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, *6*.

Chauhan, N. K., & Singh, K. (2018). A review on conventional machine learning vs deep learning. 2018 International conference on computing, power and communication technologies (GUCON),

Chen, M., Wang, X., Dou, A., & Wu, X. (2018). THE EXTRACTION OF POST-EARTHQUAKE BUILDING DAMAGE INFORMATIOM BASED ON

CONVOLUTIONAL NEURAL NETWORK. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(3).

Cooner, A. J., Shao, Y., & Campbell, J. B. (2016). Detection of urban damage using remote sensing and machine learning algorithms: Revisiting the 2010 Haiti earthquake. *Remote sensing*, *8*(10), 868.

Copernicus. *Copernicus Emergency Management Service*. Retrieved 23 March 2021 from https://emergency.copernicus.eu/

Copernicus. (2021). *Copernicus Emergency Management Service* https://emergency.copernicus.eu/mapping/list-of-components/EMSR317

Corbane, C., Saito, K., Dell'Oro, L., Bjorgo, E., Gill, S. P., Emmanuel Piard, B., Huyck, C. K., Kemper, T., Lemoine, G., & Spence, R. J. (2011). A comprehensive analysis of building damage in the 12 January 2010 MW7 Haiti earthquake using high-resolution satelliteand aerial imagery. *Photogrammetric Engineering & Remote Sensing*, *77*(10), 997-1009.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Cotrufo, S., Sandu, C., Giulio Tonolo, F., & Boccardo, P. (2018). Building damage assessment scale tailored to remote sensing vertical imagery. *European Journal of Remote Sensing*, *51*(1), 991-1005.

Dell'Acqua, F., & Gamba, P. (2012). Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives. *Proceedings of the IEEE*, *100*(10), 2876-2890.

Dikshit, A., Pradhan, B., & Alamri, A. M. (2020). Pathways and challenges of the application of artificial intelligence to geohazards modelling. *Gondwana Research*.

Dong, L., & Shan, J. (2013). A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *Isprs Journal of Photogrammetry and Remote Sensing*, *84*, 85-99.

Du, M., Liu, N., & Hu, X. (2019). Techniques for interpretable machine learning. *Communications of the ACM*, *63*(1), 68-77.

Duangsoithong, R., & Windeatt, T. (2009). Relevant and redundant feature analysis with ensemble classification. 2009 Seventh International Conference on Advances in Pattern Recognition,

Duarte, D., Nex, F., Kerle, N., & Vosselman, G. (2018). Satellite Image Classification of Building Damages using Airborne and Satellite Image Samples in a Deep Learning Approach. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *4*(2).

Dubois, D., & Lepage, R. (2013). Automated building damage classification for the case of the 2010 Haiti earthquake. 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS,

Esch, T., Schenk, A., Ullmann, T., Thiel, M., Roth, A., & Dech, S. (2011). Characterization of land cover types in TerraSAR-X images by combined analysis of speckle statistics and

intensity information. *Ieee Transactions on Geoscience and Remote Sensing*, *49*(6), 1911-1925.

Esch, T., Thiel, M., Schenk, A., Roth, A., Muller, A., & Dech, S. (2009). Delineation of urban footprints from TerraSAR-X data by analyzing speckle characteristics and intensity information. *Ieee Transactions on Geoscience and Remote Sensing*, *48*(2), 905-916.

Fan, X., Nie, G., Deng, Y., An, J., Zhou, J., Xia, C., & Pang, X. (2019). Estimating earthquake-damage areas using Landsat-8 OLI surface reflectance data. *International Journal of Disaster Risk Reduction*, *33*, 275-283.

Fan, X., Scaringi, G., Korup, O., West, A. J., van Westen, C. J., Tanyas, H., Hovius, N., Hales, T. C., Jibson, R. W., & Allstadt, K. E. (2019). Earthquake-induced chains of geologic hazards: Patterns, mechanisms, and impacts. *Reviews of Geophysics*, *57*(2), 421-503.

Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2012). Learning hierarchical features for scene labeling. *IEEE transactions on pattern analysis and machine intelligence*, *35*(8), 1915-1929.

Farhood, H., Perry, S., Cheng, E., & Kim, J. (2020). Enhanced 3D point cloud from a light field image. *Remote sensing*, *12*(7), 1125.

Fei, G., Lu, Z., Jun, W., & Jingyuan, M. (2015). Change detection in remote sensing images of damage areas with complex terrain using texture information and SVM. 2015 AASRI International Conference on Circuits and Systems (CAS 2015),

Fernandez Galarreta, J., Kerle, N., & Gerke, M. (2015). UAV-based urban structural damage assessment using object-based image analysis and semantic reasoning. *Natural Hazards and Earth System Sciences*, *15*(6), 1087-1101.

Frank, J., Rebbapragada, U., Bialas, J., Oommen, T., & Havens, T. C. (2017). Effect of label noise on the machine-learned classification of earthquake damage. *Remote Sensing*, *9*(8), 803.

Gamba, P., & Casciati, F. (1998). GIS and image understanding for near-real-time earthquake damage assessment. *Photogrammetric engineering and remote sensing*, *64*, 987-994.

Gavankar, N. L., & Ghosh, S. K. (2018). Automatic building footprint extraction from high-resolution satellite image using mathematical morphology. *European Journal of Remote Sensing*, *51*(1), 182-193.

Gerke, M., & Kerle, N. (2011). Automatic structural seismic damage assessment with airborne oblique Pictometry© imagery. *Photogrammetric Engineering & Remote Sensing*, *77*(9), 885-898.

Ghaffarian, S., & Kerle, N. (2019). Towards post-disaster debris identification for precise damage and recovery assessments from UAV and satellite images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 297-302.

Godínez-Domínguez, E., Tena-Colunga, A., Archundia-Aranda, H., Gómez-Bernal, A., Ruíz-Torres, R., & Escamilla-Cruz, J. (2020). Structural damage in housing and apartment

buildings during the September 7, 2017 Tehuantepec earthquake. Proceedings, 17th World Conference on Earthquake Engineering,

Gokon, H., Post, J., Stein, E., Martinis, S., Twele, A., Mück, M., Geiß, C., Koshimura, S., & Matsuoka, M. (2015). A method for detecting buildings destroyed by the 2011 Tohoku earthquake and tsunami using multitemporal TerraSAR-X data. *Ieee Geoscience and Remote Sensing Letters*, *12*(6), 1277-1281.

Gong, L., Li, Q., & Zhang, J. (2013). Earthquake building damage detection with object-oriented change detection. 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS,

Gong, L., Wang, C., Wu, F., Zhang, J., Zhang, H., & Li, Q. (2016). Earthquake-induced building damage detection with post-event sub-meter VHR TerraSAR-X staring spotlight imagery. *Remote sensing*, *8*(11), 887.

Grünthal, G. (1998). European macroseismic scale 1998 (EMS-98).

Guide, U. (2009). Definiens AG. *Germany: Definiens developer XD*, *2*(4).

Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, *2*(2).

Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., & Gaston, M. (2019). xbd: A dataset for assessing building damage from satellite imagery. *arXiv preprint arXiv:1911.09296*.

Haiyang, Y., Gang, C., & Xiaosan, G. (2010). Earthquake-collapsed building extraction from LiDAR and aerophotograph based on OBIA. The 2nd International Conference on Information Science and Engineering,

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, *7*(4), 12-18.

Hecht, R., Kunze, C., & Hahmann, S. (2013). Measuring completeness of building footprints in OpenStreetMap over space and time. *Isprs International Journal of Geo-Information*, *2*(4), 1066-1091.

Hedjam, R., Abdesselam, A., & Melgani, F. (2019). Change detection from unlabeled remote sensing images using Siamese ANN. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium,

Horwath, J. P., Zakharov, D. N., Mégret, R., & Stach, E. A. (2020). Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images. *npj Computational Materials*, *6*(1), 1-9.

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. *Journal of Applied Remote Sensing*, *13*(2), 022007.

Janalipour, M., & Mohammadzadeh, A. (2015). Building damage detection using object-based image analysis and ANFIS from high-resolution image (Case study: BAM earthquake, Iran). *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(5), 1937-1945.

Janik, A., Sankaran, K., & Ortiz, A. (2019). Interpreting black-box semantic segmentation models in remote sensing applications.

Janitza, S., Strobl, C., & Boulesteix, A.-L. (2013). An AUC-based permutation variable importance measure for random forests. *BMC bioinformatics*, *14*(1), 1-11.

Javadi, A. A., Farmani, R., & Tan, T. P. (2005). A hybrid intelligent genetic algorithm. *Advanced Engineering Informatics*, *19*(4), 255-262.

Jena, R., Pradhan, B., Beydoun, G., Alamri, A. M., & Sofyan, H. (2020). Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. *Science of the Total Environment*, *749*, 141582.

Ji, M., Liu, L., Du, R., & Buchroithner, M. F. (2019). A Comparative Study of Texture and Convolutional Neural Network Features for Detecting Collapsed Buildings After Earthquakes Using Pre-and Post-Event Satellite Imagery. *Remote Sensing*, *11*(10), 1202.

Karimzadeh, S., & Matsuoka, M. (2018). Building damage characterization for the 2016 Amatrice earthquake using ascending–descending COSMO-SkyMed data and topographic position index. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(8), 2668-2682.

Kato, L. V. (2019). *Integrating openstreetmap data in object based landcover and landuse classification for disaster recovery* University of Twente].

Kerle, N. (2010). Satellite-based damage mapping following the 2006 Indonesia earthquake—How accurate was it? *International Journal of Applied Earth Observation and Geoinformation*, *12*(6), 466-476.

Kerle, N., & Hoffman, R. R. (2013). Collaborative damage mapping for emergency response: the role of Cognitive Systems Engineering. *Natural Hazards and Earth System Sciences*, *13*(1), 97-113.

Kerle, N., Nex, F., Gerke, M., Duarte, D., & Vetrivel, A. (2019). UAV-based structural damage mapping: A review. *Isprs International Journal of Geo-Information*, *9*(1), 14.

Kerle, N., Nex, F., Gerke, M., Duarte, D., & Vetrivel, A. (2020). UAV-based structural damage mapping: A review. *Isprs International Journal of Geo-Information*, *9*(1), 14.

Khodaverdizahraee, N., Rastiveis, H., & Jouybari, A. (2020). Segment-by-segment comparison technique for earthquake-induced building damage map generation using satellite imagery. *International Journal of Disaster Risk Reduction*, *46*, 101505.

Korkmaz, K. A., & Abualkibash, M. (2018). Earthquake damage detection using before and after earthquake satellite images. 2018 IEEE International Conference on Electro/Information Technology (EIT),

LaJeunesse Connette, K. J., Connette, G., Bernd, A., Phyo, P., Aung, K. H., Tun, Y. L., Thein, Z. M., Horning, N., Leimgruber, P., & Songer, M. (2016). Assessment of mining extent and expansion in Myanmar based on freely-available satellite imagery. *Remote sensing*, *8*(11), 912.

Li, S., Wu, H., Wan, D., & Zhu, J. (2011). An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine. *Knowledge-Based Systems*, *24*(1), 40-48. https://doi.org/https://doi.org/10.1016/j.knosys.2010.07.003

Li, X., Yang, W., Ao, T., Li, H., & Chen, W. (2011). An improved approach of information extraction for earthquake-damaged buildings using high-resolution imagery. *Journal of Earthquake and Tsunami*, *5*(04), 389-399.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition,

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Ma, H., Liu, Y., Ren, Y., Wang, D., Yu, L., & Yu, J. (2020). Improved CNN classification method for groups of buildings damaged by earthquake, based on high resolution remote sensing images. *Remote sensing*, *12*(2), 260.

Ma, H., Liu, Y., Ren, Y., & Yu, J. (2020). Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3. *Remote Sensing*, *12*(1), 44.

Ma, L., Fu, T., Blaschke, T., Li, M., Tiede, D., Zhou, Z., Ma, X., & Chen, D. (2017). Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers. *Isprs International Journal of Geo-Information*, *6*(2), 51.

Ma, Y., Chen, F., Liu, J., He, Y., Duan, J., & Li, X. (2016). An automatic procedure for early disaster change mapping based on optical remote sensing. *Remote sensing*, *8*(4), 272.

Mansouri, B., & Hamednia, Y. (2015). A soft computing method for damage mapping using VHR optical satellite imagery. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(10), 4935-4941.

Maps, M. *Missing Maps*. Retrieved 20 November 2020 from http://www.missingmaps.org/

MapSwipe. *MapSwipe*. Retrieved 20 Novemeber 2020 from http://mapswipe.org/

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Maxar. (2021). *Digital Globe: Satellite imagery for natural disasters*. Retrieved 31 March 2021 from https://www.digitalglobe.com/ecosystem/open-data

Mitomi, H., Matsuoka, M., & Yamazaki, F. (2002). Application of automated damage detection of buildings due to earthquakes by panchromatic television images. The 7th US national conference on earthquake engineering,

Miura, H., Modorikawa, S., & Chen, S. H. (2011). Texture characteristics of high-resolution satellite images in damaged areas of the 2010 Haiti earthquake. Proceedings of the 9th International Workshop on Remote Sensing for Disaster Response, Stanford, CA, USA,

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*, b2535. https://doi.org/10.1136/bmj.b2535

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Monfort, D., Negulescu, C., & Belvaux, M. (2019). Remote sensing vs. field survey data in a post-earthquake context: Potentialities and limits of damaged building assessment datasets. *Remote Sensing Applications: Society and Environment*, *14*, 46-59.

Moya, L., Muhari, A., Adriano, B., Koshimura, S., Mas, E., Marval-Perez, L. R., & Yokoya, N. (2020). Detecting urban changes using phase correlation and ℓ1-based sparse model for early disaster response: A case study of the 2018 Sulawesi Indonesia earthquake-tsunami. *Remote Sensing of Environment*, *242*, 111743.

Nex, F., Duarte, D., Steenbeek, A., & Kerle, N. (2019). Towards real-time building damage mapping with low-cost UAV solutions. *Remote sensing*, *11*(3), 287.

Nie, J., Du, S., Fan, Y., Yang, S., He, H., Cui, Y., & Zhang, W. (2016). Estimating the numbers and the areas of collapsed buildings by combining VHR images, statistics and survey data: A case study of the Lushan Earthquake in China. *Journal of the Indian Society of Remote Sensing*, *44*(1), 101-110.

Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, *24*(12), 1565-1567.

Noriega, L. (2005). Multilayer perceptron tutorial. *School of Computing. Staffordshire University*.

Nussbaum, S., & Menz, G. (2008). eCognition image analysis software. In *Object-based image analysis and treaty verification* (pp. 29-39). Springer.

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful explanations of black box AI decision systems. Proceedings of the AAAI conference on artificial intelligence,

Pesaresi, M., Gerhardinger, A., & Haag, F. (2007). Rapid damage assessment of built-up structures using VHR satellite data in tsunami-affected areas. *International Journal of Remote Sensing*, *28*(13-14), 3013-3036.

Pfeifer, N., Rutzinger, M., Rottensteiner, F., Muecke, W., & Hollaus, M. (2007). Extraction of building footprints from airborne laser scanning: Comparison and validation techniques. 2007 Urban Remote Sensing Joint Event,

Pham, T.-T.-H., Apparicio, P., Gomez, C., Weber, C., & Mathon, D. (2014). Towards a rapid automatic detection of building damage using remote sensing for disaster management: The 2010 Haiti earthquake. *Disaster Prevention and Management*, *23*(1), 53-66.

Pollino, M., Cappucci, S., Giordano, L., Iantosca, D., De Cecco, L., Bersan, D., Rosato, V., & Borfecchia, F. (2020). Assessing earthquake-induced urban rubble by means of multiplatform remotely sensed data. *Isprs International Journal of Geo-Information*, *9*(4), 262.

Pozos-Estrada, A., Chávez, M. M., Jaimes, M. Á., Arnau, O., & Guerrero, H. (2019). Damages observed in locations of Oaxaca due to the Tehuantepec Mw8. 2 earthquake, Mexico. *Natural Hazards*, *97*(2), 623-641.

Ranjbar, H. R., Ardalan, A. A., Dehghani, H., & Saradjian, M. R. (2018). Using high-resolution satellite imagery to provide a relief priority map after earthquake. *Natural Hazards*, *90*(3), 1087-1113.

Ranjbar, H. R., Azmoude Ardalan, A. R., Dehghani, H., Serajeyan, M. R., & Alidousti, A. (2014). Evaluation of physical data extraction of damaged buildings due to earthquake and proposing an algorithm using GIS and remote sensing layers. *Scientific-Research Quarterly of Geographical Data (SEPEHR)*, *23*(91), 21-42.

Rasika, A., Kerle, N., & Heuel, S. (2006). Multi-scale texture and color segmentation of oblique airborne video data for damage classification. Proceedings of ISPRS 2006: ISPRS Midterm Symposium 2006 Remote Sensing: From Pixels to Processes,

Rathje, E. M., Crawford, M., Woo, K., & Neuenschwander, A. (2005). Damage patterns from satellite images of the 2003 Bam, Iran, earthquake. *Earthquake Spectra*, *21*(1_suppl), 295-307.

Rehor, M., Bähr, H. P., Tarsha-Kurdi, F., Landes, T., & Grussenmeyer, P. (2008). Contribution of two plane detection algorithms to recognition of intact and damaged buildings in lidar data. *The Photogrammetric Record*, *23*(124), 441-456.

Rezaeian, M. (2010). *Assessment of earthquake damages by image-based techniques* (Vol. 107). ETH Zurich.

Rezaeian, M., & Gruen, A. (2007). Automatic classification of collapsed buildings using object and image space features. In *Geomatics solutions for disaster management* (pp. 135-148). Springer.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention,

Roscher, R., Bohn, B., Duarte, M., & Garcke, J. (2020). Explain It to Me–Facing Remote Sensing Challenges in the Bio-and Geosciences With Explainable Machine Learning. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *3*, 817-824.

Saito, K., Spence, R., Booth, E., Madabhushi, M., Eguchi, R., & Gill, S. (2010). Damage assessment of Port au Prince using Pictometry. 8th International Conference on Remote Sensing for Disaster Response,

Scholz, S., Knight, P., Eckle, M., Marx, S., & Zipf, A. (2018). Volunteered geographic information for disaster risk reduction—the missing maps approach and its potential within the red cross and red crescent movement. *Remote sensing*, *10*(8), 1239.

Schweier, C., & Markus, M. (2006). Classification of collapsed buildings for fast damage and loss assessment. *Bulletin of earthquake engineering*, *4*(2), 177-192.

Shao, J., Tang, L., Liu, M., Shao, G., Sun, L., & Qiu, Q. (2020). BDD-Net: A general protocol for mapping buildings damaged by a wide range of disasters based on satellite imagery. *Remote sensing*, *12*(10), 1670.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, *2*(28), 307-317.

Shohet, I. M., Wei, H.-H., Levy, R., Shapira, S., Levi, O., Levi, T., Salamon, A., Bar-Dayan, Y., Zohar, M., & Vilnay, O. (2014). Earthquake Casualty Loss Assessment in a Major City of Israel- The case of Tiberias. In *2014 European Seismological Commission 34rd Assembly*.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, *45*(4), 427-437.

Song, D., Tan, X., Wang, B., Zhang, L., Shan, X., & Cui, J. (2020). Integration of super-pixel segmentation and deep-learning methods for evaluating earthquake-damaged buildings using single-phase remote sensing imagery. *International Journal of Remote Sensing*, *41*(3), 1040-1066.

Space, C. (2021). *Charter Space and Majors disasters*. https://disasterscharter.org/web/guest/activations/-/article/earthquake-in-indonesia-activation-587

Stramondo, S., Bignami, C., Chini, M., Pierdicca, N., & Tertulliani, A. (2006). Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *International Journal of Remote Sensing*, *27*(20), 4433-4447.

Street, O. *OpenStreetMap Contributors*. Retrieved 10 February 2021 from https://www.openstreetmap.org

Sublime, J., & Kalinicheva, E. (2019). Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. *Remote Sensing*, *11*(9), 1123.

Sugiyama, M., & Abe, H. S. K. (2002). Detection of earthquake damaged areas from aerial photographs by using color and edge information. Proceedings of the 5th Asian Conference on Computer Vision, Melbourne, Australia,

Syifa, M., Kadavi, P. R., & Lee, C.-W. (2019). An artificial intelligence application for post-earthquake damage mapping in Palu, central Sulawesi, Indonesia. *Sensors*, *19*(3), 542.

Tang, J., Deng, C., & Huang, G.-B. (2015). Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, *27*(4), 809-821.

Tiede, D., Lang, S., Füreder, P., Hölbling, D., Hoffmann, C., & Zeil, P. (2011). Automated damage indication for rapid geospatial reporting. *Photogrammetric Engineering & Remote Sensing*, *77*(9), 933-942.

Tomowski, D., Klonus, S., Ehlers, M., Michel, U., & Reinartz, P. (2010). Change visualization through a texture-based analysis approach for disaster applications. ISPRS Proceedings,

Tu, J., Sui, H., Feng, W., & Song, Z. (2016). Automatic Building Damage Detection Method Using High-Resolution Remote Sensing Images and 3D GIS Model. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *3*(8).

Turker, M., & Cetinkaya, B. (2005). Automatic detection of earthquake-damaged buildings using DEMs created from pre-and post-earthquake stereo aerial photographs. *International Journal of Remote Sensing*, *26*(4), 823-832.

Ural, S., Hussain, E., Kim, K., Fu, C.-S., & Shan, J. (2011). Building extraction and rubble mapping for city port-au-prince post-2010 earthquake with GeoEye-1 imagery and lidar data. *Photogrammetric Engineering & Remote Sensing*, *77*(10), 1011-1023.

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 Journal of Photogrammetry and Remote Sensing*, *140*, 45-59.

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 Journal of Photogrammetry and Remote Sensing*, *105*, 61-78.

Vetrivel, A., Kerle, N., Gerke, M., Nex, F., & Vosselman, G. (2016). Towards automated satellite image segmentation and classification for assessing disaster damage using data-specific features with incremental learning.

Vu, T. (2012). Rapid disaster damage estimation. *ISPRS—Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci*, *39*, b8.

Wang, G., Fan, B., Xiang, S., & Pan, C. (2017). Aggregating rich hierarchical features for scene classification in remote sensing imagery. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(9), 4104-4115.

Wang, T.-L., & Jin, Y.-Q. (2011). Postearthquake building damage assessment using multi-mutual information from pre-event optical image and postevent SAR image. *Ieee Geoscience and Remote Sensing Letters*, *9*(3), 452-456.

Watts, A. C., Ambrosia, V. G., & Hinkley, E. A. (2012). Unmanned aircraft systems in remote sensing and scientific research: Classification and considerations of use. *Remote sensing*, *4*(6), 1671-1692.

Wen, X., Miao, C., Lu, L., Zhang, H., Zhang, F., & Bi, X. (2014). Accelerated extraction technology research on damaged building information by earthquake based on LiDAR image. International Conference on Geo-Informatics in Resource Management and Sustainable Ecosystem,

Wijkman, A., & Timberlake, L. (2021). *Natural disasters: acts of God or acts of man?* Routledge.

Wu, F., Wang, C., Zhang, B., Zhang, H., & Gong, L. (2019). Discrimination of collapsed buildings from remote sensing imagery using deep neural networks. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium,

Xu, Z., Chen, Y., Yang, F., Chu, T., & Zhou, H. (2020). A Postearthquake Multiple Scene Recognition Model Based on Classical SSD Method and Transfer Learning. *ISPRS International Journal of Geo-Information*, *9*(4), 238.

Ye, L., Lay, T., Bai, Y., Cheung, K. F., & Kanamori, H. (2017). The 2017 Mw 8.2 Chiapas, Mexico, earthquake: energetic slab detachment. *Geophysical Research Letters*, *44*(23), 11,824-811,832.

Yusuf, Y., Matsuoka, M., & Yamazaki, F. (2001). Damage assessment after 2001 Gujarat earthquake using Landsat-7 satellite images. *Journal of the Indian Society of Remote Sensing*, *29*(1), 17-22.

Zhang, J.-f., Xie, L.-l., & Tao, X.-x. (2002). Change detection of remote sensing image for earthquake-damaged buildings and its application in seismic disaster assessment. *Journal of natural disasters*, *11*(2), 59-64.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition,

Zhou, W., & Troy, A. (2008). An object-oriented approach for analysing and characterizing urban landscape at the parcel level. *International Journal of Remote Sensing*, *29*(11), 3119-3135.