

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**ENHANCING DISASTER MANAGEMENT THROUGH DEEP LEARNING:
BUILDING DAMAGE ASSESSMENT USING SATELLITE IMAGERY**



M.Sc. THESIS

Mohammadreza MOHAMMADIAHANI

Department of Disaster and Emergency Management

Disaster Management Programme

JUNE 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**ENHANCING DISASTER MANAGEMENT THROUGH DEEP LEARNING:
BUILDING DAMAGE ASSESSMENT USING SATELLITE IMAGERY**

M.Sc. THESIS

**Mohammadreza MOHAMMADIAHANI
(801231003)**

Department of Disaster and Emergency Management

Disaster Management Program

Thesis Advisor: Prof. Dr. Didem SALOĞLU DERTLİ

JUNE 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**DERİN ÖĞRENME İLE AFET YÖNETİMİNİN GELİŞTİRİLMESİ: UYDU
GÖRÜNTÜLERİ KULLANILARAK BİNA HASAR TESPİTİ**

YÜKSEK LİSANS TEZİ

**Mohammadreza MOHAMMADIAHANI
(801231003)**

Afet ve Acil Durum Yönetimi Anabilim Dalı

Afet Yönetimi Programı

Tez Danışmanı: Prof. Dr. Didem SALOĞLU DERTLİ

HAZİRAN 2025

Mohammadreza MOHAMMADIAHANI, a M.Sc. student of ITU Graduate School student ID 801231003, successfully defended the thesis/dissertation entitled “ENHANCING DISASTER MANAGEMENT THROUGH DEEP LEARNING: BUILDING DAMAGE ASSESSMENT USING SATELLITE IMAGERY”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

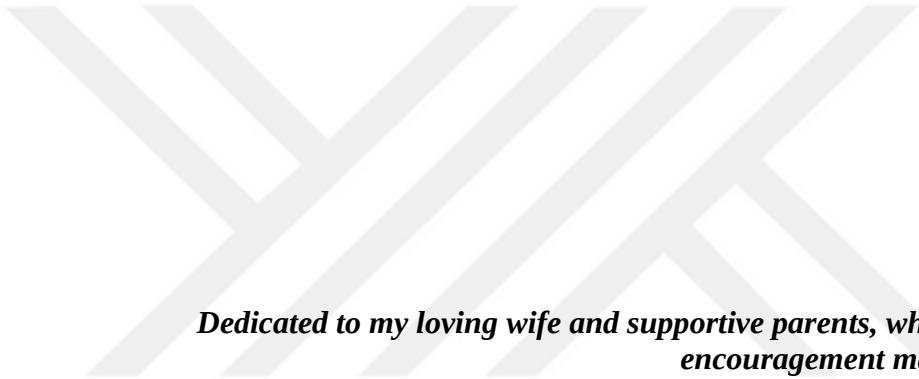
Thesis Advisor : **Prof. Dr. Didem SALOĞLU DERTLI**
Istanbul Technical University

Jury Members : **Assoc. Prof. Dr. Abdullah Can ZULFIKAR**
Istanbul Technical University

Assoc. Prof. Dr. Ahmet TURAN
Yeditepe University

Date of Submission : 11 June 2025
Date of Defense : 17 June 2025





***Dedicated to my loving wife and supportive parents, whose unwavering
encouragement made this possible***



FOREWORD

I am honored to share with you my thesis, entitled "Enhancing Disaster Management Through Deep Learning: Automated Building Damage Assessment Using Satellite Imagery". Writing this thesis has been one of the most enriching and transformative experiences of my academic journey. It has not only deepened my understanding of the role of artificial intelligence in disaster management but also challenged me to explore real-world problems through a technical and humanitarian lens.

This study was initiated in response to the critical need for rapid and reliable assessment of building damage following major disasters, as underscored by the widespread destruction caused by the 2023 Turkey earthquakes. The event highlighted significant challenges in timely damage evaluation, prompting the exploration of deep learning-based approaches to support more effective disaster response and decision-making processes.

I would like to express my heartfelt gratitude to my advisor, Prof. Dr. Didem SALOĞLU DERTLİ, whose guidance, encouragement, and critical insights were invaluable at every stage of this work.

The special thanks are due to my family and loved ones, who stood by me with unwavering support and patience throughout this intense process. Their belief in me gave me strength during moments of uncertainty and exhaustion.

I also acknowledge the developers of the open-source datasets and tools that formed the foundation of this project. Their contributions made it possible to turn an idea into a working solution aimed at supporting disaster response efforts.

Completing this thesis has not been without challenges. From data processing and model design to interpretation of results, every step involved learning and adaptation. But these challenges have only made the accomplishment more meaningful.

It is my hope that the findings and approaches outlined in this study may inspire future research in the intersection of artificial intelligence and humanitarian aid—and ultimately help build more resilient societies in the face of disaster.

June 2025

Mohammadreza MOHAMMADIAHANI
B.Sc. in Architecture



TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xv
SYMBOLS	xvii
LIST OF TABLES	xix
LIST OF FIGURES	xxi
SUMMARY	xxiii
ÖZET	xxv
1. INTRODUCTION	1
1.1 Background of the Study	1
1.2 Statement of the Problem	5
1.3 Significance of the Study	10
1.4 Scope of the Study	12
1.5 Limitations of the Study	14
1.6 Methodology of the Study	15
1.7 Organization of the Thesis	16
2. LITERATURE REVIEW	19
2.1 Introduction	19
2.2 Theoretical Background	20
2.2.1 Evolution of AI in remote sensing	23
2.2.2 Disaster response and AI integration	28
2.2.3 Comparative analysis of pre- and post-disaster imagery	31
2.3 Previous Study	34
3. RESEARCH METHODOLOGY	39
3.1 Introduction	39
3.2 Applications and Software Used	40
3.3 Data Processing	41
3.4 Building Footprint	41
3.5 General Process of Generating Building Footprint	44
3.5.1 Microsoft building footprint	46
3.5.2 Creating mask from pre-disaster satellite images	46
3.5.2.1 Pre-processing the building footprint	47
3.5.2.2 Integrating footprint with pre-disaster satellite imagery	48
3.5.2.3 Generating building mask	48
3.5.2.4 Challenges and limitation	49
3.6 Post-Disaster Building Footprint	49
3.6.1 Process of post-disaster building footprint	50
3.6.2 Data Acquisition	50
3.6.3 Post processing of extracted footprint	51
3.6.4 Application of post-disaster building footprint	51
3.7 Challenges and limitations	52
3.8 Comparison and Analysis	52

4. RESULT AND DISCUSSION.....	55
4.1 Phase 1: Model Training Process	57
4.1.1 Dataset preparation and preprocessing.....	57
4.1.1.1 Source and structure of the dataset.....	57
4.1.1.2 Parsing and conversion of annotations.....	58
4.1.1.3 Image normalization and channel conversion.....	58
4.1.1.4 Image and mask resizing	58
4.1.1.5 Mask thresholding and binarization	59
4.1.1.6 Dataset splitting strategy	59
4.1.1.7 PyTorch dataset class implementation	60
4.1.2 Model architecture.....	60
4.1.2.1 Encoder–Decoder structure	61
4.1.2.2 Input layer and preprocessing assumptions	61
4.1.2.3 Contracting path (Encoder)	61
4.1.2.4 Bottleneck layer.....	62
4.1.2.5 Expanding path (Decoder).....	62
4.1.2.6 Final convolution and activation	63
4.1.2.7 Skip Connections and their importance.....	63
4.1.2.8 Visualization reference	63
4.1.2.9 Implementation considerations.....	64
4.1.2.10 Flexibility and expandability	64
4.1.3 Training configuration.....	65
4.1.3.1 Hardware and environment setup	65
4.1.3.2 Hyperparameter settings	66
4.1.3.2.1 Batch size.....	66
4.1.3.2.2 Learning rate	66
4.1.3.2.3 Number of epochs.....	67
4.1.3.2.4 Optimizer	67
4.1.3.2.5 Loss function	67
4.1.3.2.6 Activation functions.....	67
4.1.3.2.7 Evaluation threshold	68
4.1.3.3 Learning rate strategy	68
4.1.3.4 Weight initialization and model saving.....	69
4.1.3.5 Training loop structure	70
4.1.4 Observed training behavior	71
4.2 Evaluation on Validation Set.....	73
4.2.1 Quantitative evaluation	74
4.2.2 Visual evaluation.....	77
4.2.3 Analysis of success and failure cases	82
4.2.3.1 Consistent success across diverse contexts	83
4.2.3.2 Minor and addressable limitations.....	84
4.2.3.3 Overall impact and practical suitability.....	85
4.3 Generalization to Real-World Disaster Imagery	86
4.3.1 Data source and preprocessing	89
4.3.2 Visual evaluation on real disaster imagery	92
4.3.2.1 Introduction to visual evaluation.....	93
4.3.2.2 Analysis of inference results on real earthquake imagery.....	97
4.3.2.3 Model successes across the Antakya samples.....	100
4.3.2.4 Challenges and limitations in real-world inference.....	110
4.4 Heatmap Analysis of Pre- and Post-Earthquake Satellite Imagery	113

4.5 Discussion	119
5. CONCLUSION.....	123
REFERENCES.....	127
CURRICULUM VITAE.....	131





ABBREVIATIONS

AI	: Artificial Intelligence
UAV	: Unmanned Aerial Vehicle
CNN	: Convolutional Neural Network
VHR	: Very High Resolution
IoU	: Intersection over Union
F1-SCORE	: F1 Performance Score (harmonic mean of precision and recall)
RGB	: Red Green Blue (Color Model)
BGR	: Blue Green Red (OpenCV's default color model)
BCELoss	: Binary Cross-Entropy Loss
UNet	: U-shaped CNN architecture for image segmentation
OpenCV	: Open Source Computer Vision Library
xBD	: xView2 Building Damage Assessment Dataset
JSON	: JavaScript Object Notation
GeoJSON	: Geographic JavaScript Object Notation (spatial data format)
ML	: Machine Learning
DL	: Deep Learning
TP	: True Positive
FP	: False Positive
FN	: False Negative
CRS	: Coordinate Reference System



SYMBOLS

$F1$	: F1-score
β_1	: First moment decay rate
β_2	: Second moment decay rate
ϵ	: Numerical stability constant
$\hat{y}$	: Predicted output mask from the model
x	: Input image
y	: Ground truth mask
$\sigma(\cdot)$	: Sigmoid activation function
θ	: Set of model parameters
$L(\theta)$	: Loss function with respect to model parameters
n	: Total number of samples (or pixels)
i	: Index of pixel or sample



LIST OF TABLES

	<u>DPage</u>
Table 2.1: Evolution of AI in remote sensing	28
Table 2.2: Evaluation of damage assessment maps	34
Table 4.1: A summary of performance values at key epochs	75
Table 4.2: Satellite image pairs for Antakya model testing.. ..	89





LIST OF FIGURES

	<u>Page</u>
Figure 1.1 : Türkiye Earthquake - 6 Feb 2023.....	3
Figure 3.1 : General Process of Generating Building Footprint of the study	45
Figure 4.1 : The U-Net architecture of the model used in this study	61
Figure 4.2 : Training and validation loss over Epochs.....	72
Figure 4.3 : Examples of validation on test data of the dataset	79
Figure 4.4 : Examples of validation on test data of the dataset	80
Figure 4.5 : Ant1 Satellite images before and after the earthquake.....	90
Figure 4.6 : The original satellite image with the predicted building polygons	95
Figure 4.7 : The binary mask predicted by the model	96
Figure 4.8 : The original satellite image without overlays	96
Figure 4.9 : The model outputs using a three part layout for each image pair	99
Figure 4.10 : Ant3 before earthquake polygon result	101
Figure 4.11 : Ant 2 after earthquake mask and polygons result.....	102
Figure 4.12 : Ant 5 after earthquake mask and polygons result.....	103
Figure 4.13 : Example of buildings were partially obscured by trees.....	104
Figure 4.14 : Ant 3 after earthquake mask and polygon result	105
Figure 4.15 : Ant 1 before earthquake mask and polygon result	106
Figure 4.16 : Ant 4 before earthquake mask and polygon result	107
Figure 4.17 : Building footprint comparison before and after earthquake.....	108
Figure 4.18 : Model metric comparison.....	109
Figure 4.19 : Buildings that were physically adjacent with minimal separation	111
Figure 4.20 : The model also showed reduced precision on irregularly shaped.....	111
Figure 4.21 : Example of reduced precision on irregularly shaped building.....	112
Figure 4.22 : Ant 1 heatmap analysis before earthquake.....	114
Figure 4.23 : Ant 2 before earthquake heatmap.....	115
Figure 4.24 : Ant 2 after earthquake heatmap.....	115
Figure 4.25 : The heatmap before and after the earthquake for Ant3.....	116
Figure 4.26 : The Ant 4 before and after pictures heatmap analysis	117
Figure 4.27 : The Ant 5 before and after pictures heatmap analysis	118



ENHANCING DISASTER MANAGEMENT THROUGH DEEP LEARNING: BUILDING DAMAGE ASSESSMENT USING SATELLITE IMAGERY

SUMMARY

In the face of increasingly frequent and severe natural disasters, the need for rapid, accurate, and scalable damage assessment tools has become more critical than ever. This study explores the integration of artificial intelligence (AI), specifically deep learning models, into the post-disaster response process to improve situational awareness and facilitate effective decision-making. Focusing on the case of Antakya, a city in Turkey severely impacted by the catastrophic February 6, 2023 earthquakes, the study presents a comprehensive framework that leverages satellite imagery and AI-driven segmentation to detect and analyze building damage.

The primary goal of this research is to develop a deep learning-based pipeline capable of automatically identifying buildings in satellite imagery and assessing damage. A customized U-Net architecture was implemented due to its proven performance in semantic segmentation tasks. The model was trained on a curated subset of the xBD dataset, consisting of over 5,000 high-resolution satellite images labeled with building footprints and damage levels. Additional building footprint masks were generated using OpenCV and shapely operations on JSON annotations, with all images resized to 256x256 pixels and normalized for consistency.

Visual evaluation played a significant role in validating the quality of segmentation outputs. Eight representative validation examples were analyzed, each showing the original satellite image, the ground truth mask, and the model's predicted binary mask. These visuals confirmed the model's ability to accurately detect most building structures, especially those with clear edges and uniform textures. Challenges remained in areas with heavy vegetation, shadows, or highly irregular building shapes, which occasionally led to false negatives or incomplete segmentation.

To assess generalization beyond the training dataset, the model was tested on five pairs of pre- and post-disaster satellite images of Antakya, independently acquired using Google Earth Pro. These images were not part of the original training or validation data, offering an authentic test of the model's real-world applicability. Results demonstrated the model's robustness in detecting building footprints even in domains with different visual characteristics. The predicted masks and overlaid polygons on real-world imagery aligned well with visible structures, and heatmaps generated from the predictions successfully highlighted areas of dense or missing construction, providing visual insights into structural loss.

The study also introduced building density heatmaps for both pre- and post-earthquake conditions in Antakya. These heatmaps offered an interpretable spatial visualization of predicted building footprints, highlighting the contrast in structural integrity before and after the disaster. A comparative analysis across five regions (Ant1–Ant5)

revealed noticeable declines in building density, especially in urban cores, thereby validating the model's ability to capture large-scale damage patterns.

Overall, the presented thesis confirms that AI-powered segmentation models such as U-Net can be effectively deployed for post-disaster damage mapping. While limitations exist, including dependency on high-quality imagery and the risk of under-segmentation in complex zones, the proposed system offers a scalable and efficient solution for disaster response. By automating the assessment process, the framework reduces response time and enhances decision-making in critical scenarios. The findings of this study contribute to the growing body of work advocating for the use of AI in humanitarian operations and establish a foundation for future improvements, including the integration of multi-modal data and advanced model architectures.

This thesis ultimately demonstrates the transformative potential of AI in disaster management, emphasizing its ability to support faster, more informed, and more equitable emergency responses in the wake of large-scale catastrophes.



DERİN ÖĞRENME İLE AFET YÖNETİMİNİN GELİŞTİRİLMESİ: UYDU GÖRÜNTÜLERİ KULLANILARAK BİNA HASAR TESPİTİ

ÖZET

Doğal afetlerin sıklığı ve şiddeti küresel ölçekte artarken, afet sonrası müdahalelerde hızlı, doğru ve ölçeklenebilir hasar tespit araçlarına duyulan ihtiyaç her zamankinden daha hayati hale gelmiştir. Bu tez çalışması, yapay zekânın (YZ) afet yönetiminde nasıl dönüştürücü bir rol oynayabileceğini araştırmakta ve özellikle derin öğrenme tekniklerinin, uydu görüntüleri üzerinden bina hasarlarını tespit etmek amacıyla nasıl kullanılabileceğini ortaya koymaktadır. Sunulan tez çalışması, 6 Şubat 2023 tarihinde Türkiye'nin güneydoğusunda meydana gelen ve büyük yıkıma neden olan Kahramanmaraş merkezli depremlerden ciddi şekilde etkilenen Antakya şehri üzerinde odaklanmıştır. Bu bölge, afet sonrası analizler için oldukça karmaşık ve çeşitli yapısal örüntülere sahip olduğundan, derin öğrenme tabanlı sistemlerin başarısını değerlendirmek için ideal bir örnek teşkil etmektedir.

Bu çalışmanın temel amacı, yüksek çözünürlüklü uydu görüntüleri kullanarak binaların otomatik olarak tespiti ve hasar durumlarının analiz edilmesini sağlayan derin öğrenme tabanlı bir sistem geliştirmektir. Bu amaç doğrultusunda, semantik segmentasyon görevlerinde yaygın olarak kullanılan ve başarılı sonuçlar verdiği bilinen U-Net mimarisi uygulanmıştır. Modelin eğitimi, xBD (Extreme Building Damage) veri setinden özenle seçilen 5.000'den fazla görüntü üzerinde gerçekleştirilmiştir. Bu veri seti, her biri binaların konumları ve hasar durumları ile etiketlenmiş çok sayıda yüksek çözünürlüklü uydu görüntüsünü içermektedir.

Veri setinde bulunan poligon biçimindeki bina etiketleri, JSON formatında sunulmuş olup, OpenCV ve Shapely gibi Python kütüphaneleri kullanılarak ikili (binary) maske formatına dönüştürülmüştür. Bu dönüşüm sürecinde her bir binanın içi 1 değeriyle, dışı ise 0 değeriyle temsil edilmiştir. Ardından, tüm görüntüler model giriş boyutlarına uygun şekilde 256x256 piksele yeniden boyutlandırılmış ve [0, 1] aralığında normalize edilmiştir. Eğitim sürecinde kullanılan özel PyTorch Dataset sınıfı sayesinde, her görüntü ve karşılık gelen maskesi senkronize bir şekilde modele sunulmuştur.

Model, eğitim sırasında 10 epoch boyunca Adam optimizasyon algoritması ve BCELoss fonksiyonu ile eğitilmiştir. Eğitim ve doğrulama veri kümeleri rastgele %80 - %20 oranında ayrılmıştır. Eğitim sürecinde modelin öğrenme kapasitesi hem kayıp değerlerinin (loss) düşüşü hem de doğrulama setindeki başarıımı ile izlenmiştir. Ortalama eğitim kaybı (training loss) sürekli olarak azalmış ve doğrulama seti üzerinde ortalama F1-Skoru 0,67, IoU değeri ise yaklaşık 0,48 olarak ölçülmüştür. Bu, modelin bina segmentasyonu konusunda başarılı bir genelleme yeteneği gösterdiğini ortaya koymuştur.

Modelin çıktıları yalnızca sayısal metriklerle değil, aynı zamanda görsel değerlendirme ile de incelenmiştir. Sekiz farklı doğrulama örneği seçilerek her biri üç panelden oluşan görsellerle analiz edilmiştir: modelin tahmin ettiği bina poligonlarının

uydu görüntüsü üzerine bindirildiği görsel, gerçek bina maskesi (ground truth) ve modelin ürettiği tahmini maske. Bu karşılaştırmalar, modelin özellikle düzenli, net kenarlı ve açık renkli binaları başarıyla tespit ettiğini göstermiştir. Öte yandan, ağaçlık alanlar, gölgeli yapılar ve düzensiz bina şekilleri gibi karmaşık durumlarda model zaman zaman hatalı tahminlerde bulunmuştur.

Modelin eğitim veri seti dışında nasıl performans gösterdiğini görmek amacıyla, Antakya şehrine ait beş farklı bölgeden (Ant1–Ant5) elde edilen öncesi ve sonrası görüntüler Google Earth Pro kullanılarak bağımsız biçimde toplanmıştır. Bu görüntüler üzerinde modelin tahmin ettiği bina maskeleri ve oluşturulan poligonlar, gerçek görüntülerle oldukça yüksek uyum göstermiştir.

Bu durum, modelin eğitim verisinin dışında kalan alanlarda da etkin biçimde çalışabildiğini ortaya koymaktadır.

Ayrıca, modelin tahmin çıktılarından elde edilen bina yoğunluğu haritaları (heatmap) hem deprem öncesi hem de sonrası durumlar için oluşturulmuştur. Bu ısı haritaları, kentsel dokunun mekânsal değişimini analiz etmek için güçlü bir araç sağlamış, özellikle şehir merkezlerinde bina yoğunluğunun azaldığını görsel olarak ortaya koymuştur. Bu analizler, modelin yalnızca bireysel bina tespiti değil, aynı zamanda büyük ölçekli yapısal kayıpların haritalanmasında da kullanılabileceğini göstermiştir.

Modelin başarılarının yanında bazı sınırlılıkları da göz ardı edilmemelidir. Segmentasyon kalitesinin yüksek olması için kullanılan uydu görüntülerinin çözünürlüğü ve kalitesi büyük önem taşımaktadır.

Ayrıca, yoğun bitki örtüsü, gölge ve karmaşık mimari desenler, modelin doğruluğunu olumsuz etkileyebilmektedir. Yine de genel olarak modelin doğruluk, hassasiyet ve genelleme yeteneği oldukça başarılıdır.

Bu çalışmanın literatürdeki benzer araştırmalarla karşılaştırılmasında, U-Net mimarisinin sadeliği, eğitim süresinin kısalığı ve hesaplama verimliliği ön plana çıkmaktadır. Mevcut model, kaynak sınırlaması olan durumlarda dahi uygulanabilir bir yapı sunmaktadır. Ayrıca, modelin ölçeklenebilirliği, farklı coğrafyalarda ve diğer afet türlerinde de benzer yaklaşımların uygulanabileceğini göstermektedir.

Bu veriler, modelin yüksek hassasiyetle çalıştığını, ancak bazı bölgelerde eksik tespitlere neden olabilecek sınırlamaların da olduğunu ortaya koymaktadır. Özellikle Microsoft veri setinin bütüncül ve güvenilir bir kaynak olması göz önünde bulundurulduğunda, elde edilen sonuçlar tatmin edici olarak değerlendirilmektedir.

Sonuç olarak, sunulan tez çalışması yapay zekâ destekli derin öğrenme modellerinin afet sonrası hasar tespitinde uygulanabilirliğini kapsamlı bir şekilde ortaya koymaktadır. Geliştirilen sistem, karar vericilere hızlı ve güvenilir bilgi sunarak müdahale süreçlerini iyileştirme potansiyeline sahiptir. Aynı zamanda çok modelli verilerin (örneğin; drone görüntüleri, harita katmanları) entegrasyonu ve gelişmiş ağ mimarilerinin (örneğin; U-Net++ veya Transformer tabanlı segmentasyon modelleri) dahil edilmesiyle bu sistemlerin daha da güçlü hâle getirilebileceği değerlendirilmektedir.

Bu kapsamlı çalışma yalnızca teknik bir modelleme değil, aynı zamanda afet sonrası karar alma süreçlerinde kullanılabilecek bütüncül bir sistemin temelini oluşturmaktadır. Elde edilen çıktılar, acil durum yönetimi ekipleri, yerel yönetimler ve insani yardım kuruluşları için oldukça değerli olup, müdahale sürecinin ilk saatlerinde hızlı bir şekilde hasar durumu hakkında bilgi edinmeyi mümkün kılmaktadır. Bu da,

“altın saatler” olarak adlandırılan kritik zaman diliminde kaynakların etkin dağıtımını ve önceliklendirilmesini desteklemektedir.

Geliştirilen sistemin gelecekte afet yönetimi karar destek platformlarına entegre edilmesi, sahadan gelen diğer veri kaynaklarıyla (örneğin sosyal medya analizleri, sensör verileri, mobil uygulama bildirimleri) birlikte kullanılarak çok katmanlı bir analiz ortamı yaratabilir. Böylece afetlere müdahale süreci daha dinamik, esnek ve veriye dayalı hâle getirilebilir.

Bu bağlamda, çalışma yalnızca teknik bir uygulama değil, aynı zamanda afet yönetimi süreçlerinin dijital dönüşümüne katkı sağlayan stratejik bir öneri niteliği taşımaktadır. Hem Türkiye özelinde hem de küresel ölçekte afet yönetimi politikalarının daha etkili hâle getirilmesi için yapay zekâ teknolojilerinin aktif şekilde entegre edilmesi gerektiği bu tezde açıkça vurgulanmıştır.





1. INTRODUCTION

1.1 Background of the Study

Natural and man-made disasters have increasingly severe impacts on communities worldwide, disrupting lives, economies, and infrastructure on a massive scale. Events such as earthquakes, floods, hurricanes, and wildfires result in significant loss of life, displacement of populations, and destruction of property (UNDRR, 2020). The ability to respond swiftly and effectively during such crises is critical for minimizing human suffering and ensuring rapid recovery. However, traditional disaster assessment and decision-making methods, which rely heavily on ground-based observations and manual data collection, are often too slow, resource-intensive, or even impossible to implement in certain disaster scenarios. Traditional disaster assessment and decision-making methods, while foundational, face significant challenges during the critical response phase of disaster management. These methods rely heavily on ground-based observations and manual data collection, which are often too slow to meet the urgent demands of a rapidly evolving crisis. For instance, during natural disasters like earthquakes or floods, the time required to manually gather data, assess damages, and develop a response plan can delay life-saving actions. Moreover, resource limitations, such as the availability of personnel and equipment, further constrain these methods. In scenarios where infrastructure is destroyed or areas are inaccessible due to hazards, traditional approaches may be completely infeasible, leaving response teams without the information they need to act effectively. The response phase of disaster management is a time-critical window where rapid decision-making is crucial to saving lives and mitigating damage. Delays in assessment can lead to inefficient allocation of resources, exacerbating the disaster's impact (Coppola, 2015). Traditional methods often lack the ability to provide the real-time data required for such urgent decision-making. This is particularly problematic in large-scale disasters, where the scope and complexity of the situation demand high-speed processing and dynamic adaptability. The limitations of traditional approaches underscore the urgent need for

innovative solutions to improve the speed, accuracy, and scalability of disaster response operations (Wang, 2020).

Advancements in technology, particularly satellite and Unmanned Aerial Vehicle (UAV) imaging systems, have transformed the way disaster data is collected. Satellites provide large-scale imagery that offers an invaluable overview of affected regions, while UAVs deliver high-resolution images and the flexibility to operate in inaccessible areas. These technologies enable rapid data acquisition, yet the overwhelming volume and complexity of the data pose significant analytical challenges (Li, 2018). Timely and accurate interpretation of this data is essential to guide effective disaster response strategies, but traditional manual approaches are inadequate for handling such vast amounts of information. In the aftermath of a disaster, vast amounts of data are generated from various sources, including satellite imagery, drones, sensors, social media, and reports from on-the-ground responders. While this data is invaluable for shaping an effective disaster response, managing it poses significant challenges. The sheer volume of information can overwhelm traditional manual methods, which rely on human analysts to process and interpret data. This results in delays that can compromise the timeliness and effectiveness of response efforts, particularly in fast-evolving scenarios like earthquakes or floods. One of the primary difficulties lies in the sheer volume of data generated during disasters. For example, satellite imagery alone can produce terabytes of high-resolution images that need to be analyzed to identify affected areas and assess damage. Adding to this complexity are the diverse formats and sources of data, such as videos, sensor readings, and unstructured text from social media posts. Integrating these disparate data streams into a cohesive picture of the disaster scenario requires significant effort and expertise, which is difficult to achieve manually in the limited time available during the response phase.

Artificial Intelligence (AI) has emerged as a powerful solution to these challenges, providing advanced tools for processing and analyzing large datasets efficiently (Andrae, 2025). AI-powered algorithms excel at identifying patterns, detecting damage, and extracting actionable insights from satellite and UAV imagery. By automating these processes, AI significantly enhances the speed and accuracy of disaster assessments, empowering response teams to make better-informed decisions. These capabilities are particularly critical during the response phase of disaster

management, where time is of the essence. Timeliness is critical in disaster response, but manual data processing often fails to meet the demands of real-time decision-making. In scenarios such as rapidly spreading wildfires or rising floodwaters, delays in interpreting data can mean the difference between life and death. Traditional methods lack the speed and efficiency needed to provide actionable insights when every second counts, resulting in missed opportunities to allocate resources or evacuate at-risk populations.

The devastating earthquakes that struck Turkey on February 6, 2023, illustrate the pressing need for advanced disaster management solutions. With magnitudes of 7.8 and 7.7, these earthquakes caused widespread destruction, claiming over 54,000 lives, displacing millions, and destroying approximately 273,000 buildings. Antakya, one of the worst-hit cities, exemplifies the challenges of rapid damage assessment and resource allocation during a major disaster as it shown in Figure 1.1. The magnitude of this event underscores the importance of leveraging AI-powered tools to enhance situational awareness and expedite recovery efforts (PreventionWeb, 2023).

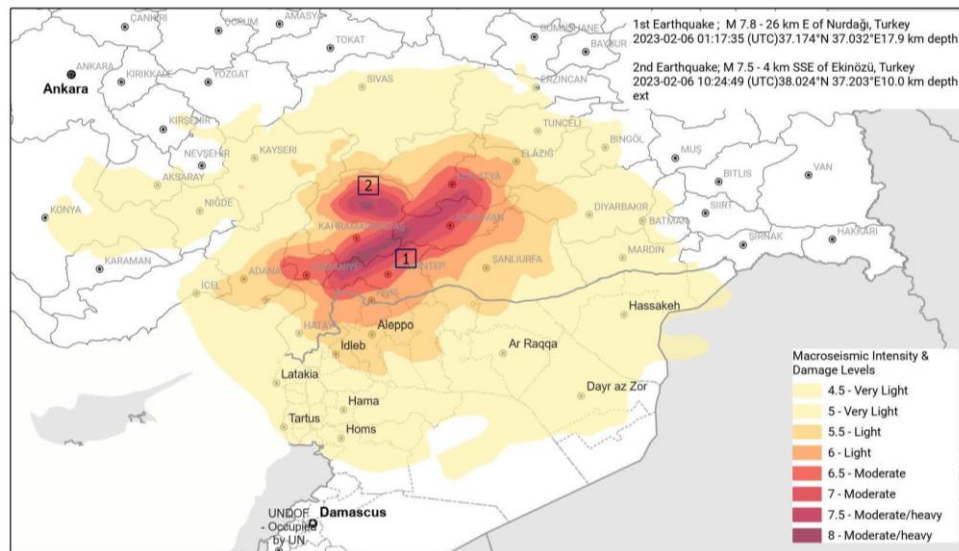


Figure 1.1 : Türkiye Earthquake - 6 Feb 2023: USGS Macroseismic Intensity & Damage Levels - 1stEQ Magnitude 7.8 & 2nd EQ Magnitude 7.5 (OCHA, 2023)

One of the most immediate challenges was the sheer scale of the disaster. The affected area spanned several provinces, making it difficult to assess the extent of the damage and prioritize response efforts. Traditional methods of damage assessment, which rely on ground surveys and manual reporting, proved too slow to cope with the vast geographic scope of the disaster. This delay in understanding the situation resulted in the misallocation of resources, with some areas receiving aid late or insufficiently,

while others experienced delays in rescue operations. The earthquakes caused widespread damage to critical infrastructure, including communication networks, roads, and bridges. This disrupted coordination among response teams and made it challenging to relay information from affected areas to decision-makers. First responders often operated in silos, with limited situational awareness of the broader disaster landscape. The breakdown in communication systems also hampered efforts to locate survivors and deliver timely assistance.

The response to the earthquakes generated a massive amount of data from various sources, such as satellite imagery, social media posts, drone footage, and reports from local authorities. However, integrating and analyzing this data in real time proved to be a significant challenge. Traditional manual methods of data processing were overwhelmed by the volume and diversity of information, leading to delays in identifying the most affected areas and allocating resources efficiently. The lack of automated tools to process this data further exacerbated the problem. Many of the hardest-hit areas were rendered inaccessible due to collapsed buildings, debris, and damaged roads. This made it difficult for rescue teams to reach these locations and provide immediate assistance. The lack of access also delayed the collection of ground-based observations, leaving decision-makers with incomplete information about the severity of the situation in these regions. The disaster revealed significant logistical challenges, particularly in delivering aid to affected populations. Blocked roads and damaged transportation networks hindered the distribution of essential supplies such as food, water, and medical aid. In some cases, logistical bottlenecks resulted in prolonged delays, leaving survivors without critical support during the early hours and days of the disaster.

The earthquakes in Turkey in 2023 exposed systemic gaps in disaster response mechanisms, particularly in terms of scalability, coordination, and efficiency. The challenges of resource allocation, data management, accessibility, and communication underscored the limitations of traditional disaster management approaches. These difficulties highlight the urgent need to adopt advanced technologies, such as AI, to improve real-time decision-making, streamline logistics, and enhance overall disaster response capabilities. By learning from these challenges, disaster management systems can evolve to better address future crises and save lives more effectively.

This study focuses on the integration of AI with satellite and UAV imagery to improve disaster response and decision-making. By analyzing pre- and post-disaster imagery using a deep learning model, this research seeks to demonstrate the potential of AI in generating real-time, actionable intelligence. The findings aim to provide valuable insights for emergency services, government agencies, and humanitarian organizations, offering a scalable framework that can be applied to various types of disasters globally.

The background of this study is deeply rooted in the urgent need for innovative solutions in disaster management, driven by the increasing frequency and intensity of natural disasters. As climate change and urbanization exacerbate the risks and impacts of such events, the adoption of AI-powered technologies represents a crucial step toward building more resilient communities and saving lives during crises..

1.2 Statement of the Problem

Technological advancements have profoundly impacted many sectors, and disaster management is no exception. Traditionally, disaster response (non AI-driven methods) has relied on human expertise and methods like manual data analysis and post-disaster surveys. Traditional damage assessment methods are notoriously time-consuming. Field surveys, for instance, require trained personnel to physically inspect affected areas, document damage, and compile reports. This process often spans several days or even weeks, particularly in large-scale disasters or regions with challenging terrain. During this time, crucial response efforts such as resource allocation, rescue operations, and aid distribution are delayed, exacerbating the disaster's impacts (Khan S. U., 2021).

Due to logistical constraints, traditional methods (non AI-driven disaster response method) often struggle to provide comprehensive coverage of disaster-stricken areas. Remote or inaccessible regions may remain unassessed for extended periods, leaving vulnerable communities without timely support. This limited coverage can result in skewed assessments and misinformed decision-making.

Human-led damage assessments are inherently subjective, influenced by the expertise and judgment of individual assessors. Variability in assessment techniques and criteria

across teams or regions can lead to inconsistent results, complicating efforts to establish an accurate and unified understanding of the disaster's impact.

Traditional response methods (non AI-driven disaster response method) often rely on data collected after the disaster has already unfolded, leading to significant delays in the availability of actionable insights. These delays are particularly problematic in fast-evolving disaster scenarios, where real-time information is critical for effective response and mitigation (Hristidis, 2010).

As disasters become more frequent and severe due to climate change and urbanization, the volume of data to be processed and analyzed has grown exponentially. Traditional methods, which are not designed to handle such massive datasets, often become overwhelmed, resulting in incomplete or outdated assessments.

Manual assessments require significant human and financial resources, including trained personnel, transportation, and communication infrastructure (Engineers, 2012). In disaster-stricken regions, these resources may already be strained or unavailable, further limiting the effectiveness of traditional approaches.

While effective, these approaches often face limitations, such as delayed response times and reduced accuracy in damage assessments. As the frequency and complexity of disasters increase globally, the need for faster, more accurate decision-making tools becomes clear. One such tool is Artificial Intelligence (AI), particularly AI-powered satellite and UAV (Unmanned Aerial Vehicle) imagery analysis (Rolla, 2024).

AI refers to systems that can perform tasks requiring human intelligence, such as learning and decision-making. In disaster management, AI allows for the analysis of large datasets, generating actionable insights in real-time. When combined with satellite and UAV technologies, AI offers new possibilities for disaster response, enabling faster and more accurate monitoring, assessment, and management of disaster situations.

AI-powered satellite and UAV imagery analysis marks a departure from traditional methods, providing disaster response teams with high-resolution images of affected areas. AI algorithms can quickly process these images, identifying damage, infrastructure impacts, and helping prioritize response efforts. By automating damage detection, AI reduces the time required to gather critical information, which is crucial in a disaster scenario where time is of the essence.

One key advantage of AI in disaster management is its ability to process vast quantities of data in real-time. Previously, disaster response teams were overwhelmed by the sheer volume of information from various sources, such as weather forecasts, satellite images, and ground reports. AI addresses this challenge by automating data processing, filtering out irrelevant information, and identifying relevant patterns. For instance, during an earthquake, AI can analyze satellite images to quickly identify collapsed buildings, roads, and infrastructure, providing responders with an accurate situational overview.

Moreover, AI enhances predictive capabilities. Machine learning models trained on historical disaster data can predict future events, enabling proactive responses. For example, AI systems can analyze weather and geological data to predict the likelihood of hurricanes, earthquakes, or floods, giving authorities valuable time to prepare and implement mitigation strategies (IEEE Public, 2023).

However, AI's application in disaster management is still in its early stages, and challenges remain. One major issue is the scale and complexity of the data involved. While satellite and UAV imagery provide high-resolution data, it also generates massive volumes of information. High-resolution images require significant computational resources for processing. In time-sensitive disaster scenarios, it is essential that AI systems can analyze this data quickly and efficiently.

Data quality variability also poses challenges. Satellite and UAV imagery can be affected by factors such as weather conditions, the time of day, and sensor quality. For example, stormy weather can reduce image clarity, and UAVs may capture more detailed images at lower altitudes, but their coverage is more limited than that of satellites. AI algorithms must be capable of processing this inconsistent data and still produce reliable results (Aleissae, 2024).

Another challenge is integration. For AI systems to be effective in disaster management, they must be integrated into existing protocols and workflows. Many disaster response teams lack the technical expertise and resources to fully utilize AI technologies. Even organizations that have adopted AI often face difficulties in training personnel to incorporate AI insights into decision-making processes. For AI to be truly effective, responders and decision-makers need to be equipped with the knowledge and tools to use these technologies efficiently (Velev, 2023).

Ethical considerations also arise in the use of AI for disaster management. AI algorithms, if trained on unrepresentative data, may inadvertently reinforce biases, leading to skewed risk assessments or resource allocations. This could exacerbate inequalities in disaster response. Ensuring that AI systems are transparent, accountable, and inclusive is vital for ensuring fairness and equity in disaster management.

Despite these challenges, the potential of AI in disaster management is undeniable. As AI technologies evolve, new techniques, such as deep learning and neural networks, are enhancing the accuracy and speed of satellite and UAV imagery analysis. These models can identify complex patterns, such as distinguishing between different types of infrastructure damage or recognizing vulnerable populations, helping response teams make more informed decisions and take more effective actions (Kamal, 2024).

The Advent of AI in Disaster Management is started late 20th and early 21st centuries saw the integration of AI, particularly machine learning, into disaster management. Machine learning algorithms allowed computers to identify patterns in large datasets, opening up new possibilities for automating the analysis of satellite and aerial imagery. AI systems could process data faster than human analysts, significantly reducing the time required for damage assessment and enabling quicker responses.

One of the earliest examples of AI in disaster management was the use of neural networks to detect forest fires by analyzing satellite images. back to the late 1990s and early 2000s. One of the earliest and most notable implementations occurred in Europe, particularly within projects spearheaded by research institutions and space agencies such as the European Space Agency (ESA). In 1998, the ESA initiated programs to utilize satellite data for environmental monitoring, including forest fire detection. Neural network algorithms were employed to analyze data from satellites like ERS (European Remote Sensing Satellite) and ENVISAT (Environmental Satellite). These models were trained to identify patterns associated with wildfires, such as heat signatures, smoke plumes, and changes in vegetation reflectance, enabling near-real-time detection (Calle, 2006). These algorithms could identify signs of fire in real-time, even in remote areas, demonstrating AI's potential to enhance disaster monitoring and prediction. For instance, following the 2010 Haiti earthquake (ImageCat, 2010), AI was used to analyze satellite images and identify areas of structural damage, guiding responders to the most affected regions (Xu, 2019). Similarly, AI technologies helped

assess the flood damage in the aftermath of Hurricane Katrina in 2005, demonstrating how AI and geospatial data could work together to improve disaster response (News, 2006).

Other instance, after the 2010 Haiti earthquake, satellite imagery played a vital role in damage assessment. AI was used to analyze these images and create detailed maps, guiding relief efforts. Similarly, after the 2015 Nepal earthquake, AI algorithms were applied to UAV imagery to assess building damage in real-time, helping teams prioritize resources based on the severity of damage.

Developing a systematic framework for AI integration in disaster response is essential for realizing the full potential of AI technologies in managing natural disasters. Such a framework provides the necessary structure to ensure that AI tools are effectively incorporated into disaster management systems, enhancing coordination, consistency, and efficiency in decision-making. By addressing key elements such as data integration, AI model optimization, human-machine collaboration, and training, the framework can enable AI systems to provide actionable insights that improve disaster response and resilience. With continuous evaluation and improvement, this framework can adapt to evolving disaster scenarios, ensuring that AI remains a valuable tool for saving lives and mitigating the impact of future disasters (Chen, 2024).

Disaster management in large-scale earthquakes, such as the February 2023 event in Turkey and Syria, requires rapid decision-making and accurate situational awareness. This research aims to address that need by employing a deep learning-based approach specifically a U-Net architecture using pre- and post-disaster satellite imagery to generate maps of building locations and assess structural damage.

This research aims to bridge the gap between the technological potential of AI and its practical applications in disaster management. By optimizing AI algorithms for real-time data processing and developing a systematic framework for integrating AI-driven insights into disaster management systems, this study seeks to empower responders and decision-makers. With faster, more accurate data analysis and structured frameworks for AI integration, this research aspires to not only enhance the effectiveness of disaster management but also contribute to greater community resilience, saving lives, and reducing the economic toll of natural disasters.

1.3 Significance of the Study

This study is of critical importance for advancing the use of AI in the analysis of satellite and UAV imagery to enhance disaster response and decision-making. The development of an AI model capable of delivering rapid and accurate analysis has the potential to transform how disaster management teams operate, offering significant benefits in several key areas.

I. Improved accuracy in disaster assessment:

Disaster assessment often involves analyzing extensive datasets, particularly satellite and UAV imagery, to evaluate damage and identify affected areas. Traditional methods, reliant on manual interpretation or semi-automated systems, face significant challenges in dealing with these datasets' volume, diversity, and complexity. For example, distinguishing between collapsed infrastructure, flood-affected regions, or inaccessible zones can be error-prone when performed manually, especially under the time pressure inherent in disaster scenarios. The proposed AI model will incorporate advanced machine learning algorithms, such as convolutional neural networks (CNNs), optimized for image recognition and analysis. By automating the processing of high-resolution imagery, the AI can detect patterns, anomalies, and damage indicators more accurately than human observers. Additionally, the model's ability to learn from historical data ensures continual improvement in its accuracy over time. Improved accuracy directly translates to better-targeted disaster response. By reducing human error and providing precise assessments, the model minimizes the risk of misallocating resources or overlooking critical areas. Accurate assessments are especially crucial in large-scale disasters, where every second counts, and response teams need to operate with reliable data to save lives and reduce damage.

II. Rapid analysis for timely decision-making:

Disaster response hinges on speed. The time it takes to process data and make informed decisions can determine whether lives are saved or lost. Traditional analysis methods, constrained by manual processing and human cognitive limitations, cannot match the pace required during fast-evolving disasters. For example, it might take hours or even days for analysts to review satellite imagery of an earthquake's aftermath, slowing down rescue operations. The AI model developed in this study is designed to operate in near real-time, rapidly processing satellite and UAV imagery as it becomes

available. By leveraging cloud computing and advanced algorithms, the model will ensure that actionable insights such as maps of affected regions, infrastructure damage, and population density are delivered to response teams within minutes of data acquisition. Rapid analysis addresses one of the most critical bottlenecks in disaster management. Faster decision-making allows responders to deploy resources more efficiently, prioritize high-risk areas, and implement life-saving measures without unnecessary delays. This capability is particularly vital in disasters like floods or wildfires, where conditions change rapidly, and delays can escalate damage or casualties.

III. Enhanced support for response teams:

On-the-ground response teams face immense challenges, including navigating hazardous environments, prioritizing rescue operations, and coordinating logistics. They often lack real-time, reliable data to inform their actions, resulting in inefficiencies and missed opportunities. This study's AI model will generate actionable information, such as heat maps of damage severity, routes to inaccessible areas, and estimated needs for resources like food, water, and medical supplies. The AI system will integrate seamlessly with existing disaster management platforms, providing response teams with an easy-to-use interface to access critical data. This enhanced situational awareness will empower responders to allocate their efforts strategically, ensuring that the most critical areas receive attention first. Enhanced support for response teams not only improves their efficiency but also their safety. With clear, accurate, and timely information, responders can avoid redundant efforts, navigate risks more effectively, and focus on high-priority tasks. Ultimately, this leads to faster recovery times for affected communities and reduces the overall impact of disasters.

IV. Contribution to technological advancements in disaster management:

The application of AI in disaster management represents a paradigm shift in how disasters are approached and managed. This study's AI model will set a benchmark for integrating cutting-edge technology into disaster response, providing a scalable solution adaptable to various types of disasters, such as earthquakes, floods, wildfires, and hurricanes. By demonstrating the practical benefits of AI, this research will inspire further innovation in the field, including the development of complementary tools like

predictive models and automated drones for rapid data collection. The AI model's scalability ensures its relevance across diverse scenarios, from small-scale incidents to large-scale disasters. Its adaptability will also allow it to evolve alongside technological advancements, such as improvements in satellite imaging resolution or the emergence of new data sources like IoT sensors. This research contributes to the broader knowledge base of AI applications in disaster management, bridging the gap between theoretical concepts and real-world implementation. By providing a proof-of-concept for AI-driven disaster response, this study paves the way for future innovations that can further enhance disaster preparedness, mitigation, and recovery efforts.

The significance of this study lies in its ability to address critical gaps in traditional disaster management practices. By improving accuracy, enabling rapid analysis, supporting response teams, and contributing to technological advancements, this research has the potential to revolutionize disaster response strategies. The tools and insights developed through this study will save lives, minimize damage, and enhance the overall resilience of communities in the face of disasters. Through the integration of AI into disaster management, this research not only responds to current challenges but also lays the groundwork for a more adaptive and effective approach to managing crises in the future.

1.4 Scope of the Study

This study focuses on the critical role of artificial intelligence (AI) in disaster response, particularly in aiding response teams to better understand and assess the situation in disaster-affected areas. By leveraging AI's capability to analyze satellite and UAV imagery, the research aims to develop and evaluate a framework for rapid and accurate building damage assessment, ultimately enhancing situational awareness during crises.

I. Geographical scope:

The geographical focus of this study is the Turkish city of Antakya, located in Hatay province, which was one of the hardest-hit areas during the catastrophic 6 February 2023 Kahramanmaraş Earthquakes. These earthquakes, with magnitudes of 7.8 and 7.7, resulted in widespread devastation, claiming over 54,000 lives, displacing millions, and destroying approximately 273,000 buildings. Antakya, due to its

significant damage and unique challenges, provides an ideal case study for evaluating the effectiveness of AI-powered disaster response tools. The findings from this study will help understand how AI can be utilized in real-world disaster scenarios to support decision-making and resource allocation.

II. Technological scope:

The research utilizes Very High Resolution (VHR) satellite images and UAV imagery to create two separate building damage assessment maps for Antakya. The first map is generated using pre-disaster imagery, while the second is created using post-disaster imagery analyzed by a deep learning model. By comparing these approaches, the study evaluates which method offers more precise and timely insights into the extent of damage. Advanced AI algorithms, particularly deep learning models like convolutional neural networks (CNNs), are employed to detect patterns and assess damage with a high degree of accuracy.

III. Operational scope:

The study focuses on the disaster response phase, a critical period when timely and precise damage assessments are essential for effective intervention. The AI models developed in this research aim to assist response teams in rapidly identifying the most severely affected areas, thereby enabling the efficient allocation of resources and the prioritization of rescue and recovery operations. This operational framework emphasizes the use of AI to reduce the time between data collection and decision-making, ultimately improving the resilience of disaster-affected communities.

IV. Limitations:

While the study demonstrates the potential of AI-powered tools for disaster response, it acknowledges certain limitations, including the dependency on high-quality satellite and UAV imagery, the computational resources required for deep learning model training and execution, and the challenges of integrating AI insights into traditional disaster response protocols. The research is also geographically limited to Antakya and its surroundings, which may affect the generalizability of the findings to other regions or disaster types.

By addressing these challenges, this study aims to provide a scalable and adaptable framework for integrating AI into disaster response efforts.

1.5 Limitations of the Study

This study, while providing valuable insights into the use of artificial intelligence (AI) for building damage assessment in disaster response, has several limitations that need to be acknowledged:

I. Data quality and availability:

The accuracy of the AI models used in this study is highly dependent on the quality of the satellite images available. Very High Resolution (VHR) satellite images are crucial for detailed damage assessment, but such images are not always readily available, especially in the immediate aftermath of a disaster. Additionally, the variability in image quality due to factors like weather conditions, satellite angle, and resolution can affect the performance of the AI models.

II. Generalizability:

The study focuses specifically on the 6 february 2023 Kahramanmaraş Earthquakes, particularly the city of Antakya in Hatay province. While the findings provide important insights into the effectiveness of AI in this context, the results may not be directly generalizable to other types of disasters or different geographical locations. The unique characteristics of the earthquake and the region may limit the applicability of the AI models in other scenarios.

III. Computational resources:

The development and application of deep learning models require significant computational resources. In real-world disaster response scenarios, access to such resources may be limited, particularly in regions with less developed infrastructure. This limitation could impact the feasibility of deploying AI-based damage assessment tools in all disaster situations.

IV. Time constraints:

The study is conducted within a specific timeframe, which may limit the depth of the analysis and the ability to test the AI models under different conditions. The rapid nature of disaster response means that models must be tested and deployed quickly, which can limit the thoroughness of the evaluation process.

V. Human expertise:

While AI models can enhance the accuracy and speed of damage assessments, they are not infallible and still require human expertise for interpretation and validation. The reliance on AI should not undermine the importance of expert judgment, especially in complex disaster scenarios where nuanced decision-making is crucial.

1.6 Methodology of the Study

This study focuses on identifying and prioritizing the most significant potential space hazards and developing a disaster risk reduction (DRR) framework tailored to the high-priority hazards. The methodology for this study involves the following key steps:

I. Review of literature:

This study will begin with a comprehensive review of existing literature related to artificial intelligence and disaster response. This review will provide a theoretical basis for the study and help to identify existing knowledge gaps.

II. Data collection:

The study begins with the collection of Very High Resolution (VHR) satellite images, which are crucial for detailed analysis. The selected imagery covers key regions of Antakya that were severely impacted by the earthquake..

III. Preprocessing of satellite images:

The collected satellite images undergo preprocessing to enhance their quality and ensure consistency across the dataset. This involves correcting geometric distortions, normalizing radiometric properties, and aligning the images to a common coordinate system. The preprocessing phase ensures that the data is suitable for subsequent analysis by both the traditional object-based method and the deep learning model.

IV. DL model development:

Concurrently, a deep learning model is developed and trained using a portion of the preprocessed satellite images. The model, typically based on Convolutional Neural Networks (CNNs), is designed to automatically detect and classify buildings from the imagery. The training process involves feeding the model with labeled data and

iteratively adjusting the model's parameters to optimize its performance. Once trained, the model is applied to the full set of images to generate a building damage assessment map.

V. Validation:

The effectiveness of the proposed AI-powered disaster response and decision-making framework will be validated through a combination of quantitative performance metrics and visual assessment techniques. Specifically, the model's outputs will be compared against ground truth data, such as Microsoft Building Footprints and manually annotated damage masks, to evaluate its precision, recall, F1-score, and Intersection over Union (IoU). Additionally, the framework will be tested on real-world satellite imagery from the Antakya region before and after the Feb 6, 2023 Turkey earthquakes to assess its generalizability and robustness in actual disaster scenarios. This dual validation approach ensures a comprehensive evaluation of the model's accuracy and practical relevance without relying on expert consultation.

This methodology will offer a comprehensive and systematic approach to assessing building damage in disaster-affected areas, contributing to the development of an AI-powered disaster response and decision-making framework. By incorporating advanced deep learning techniques and satellite imagery, the framework will provide a reliable and scalable solution for rapid damage assessment. The integration of expert feedback will ensure that the framework is robust, accurate, and applicable across various disaster scenarios, enhancing its relevance and effectiveness in real-world applications.

1.7 Organization of the Thesis

This thesis focuses on the identification and prioritization of potential space hazards and the development of a disaster risk reduction (DRR) framework tailored for these hazards. The thesis is organized as follows:

I. Chapter 1: Introduction

This section provides an overview of the thesis, including its background, significance, and objectives. It also defines the scope of the study and outlines the methodology used.

II. Chapter 2: Literature Review

This section offers an in-depth review of existing literature on the use of AI-powered satellite imagery in disaster response and decision-making. It explores the current methodologies for building damage assessment and providing a theoretical foundation for the study. Additionally, it identifies gaps in the existing knowledge, particularly in the application of AI for enhancing disaster response, which this research aims to address.

III. Chapter 3: Methodology

This chapter outlines the methodology used in this study to assess building damage following the earthquake in Antakya. It details the process of collecting Very High Resolution (VHR) satellite images and the subsequent preprocessing steps, including geometric correction, radiometric normalization, and alignment to a common coordinate system. Additionally, it describes the development of the deep learning model, including the training process using labeled data and the application of the model to generate a building damage assessment map.

IV. Chapter 4: Implementation

This chapter details the implementation process utilized in this study to apply the deep learning model for assessing building damage in Antakya after the earthquake. It describes the steps involved in deploying the preprocessed satellite images into the trained deep learning model, ensuring optimal model performance through fine-tuning and adjustments. Additionally, this chapter outlines the generation of the building damage assessment map, integrating the model's predictions with geographic information system (GIS) tools to produce a comprehensive visual representation of the affected areas.

V. Chapter 5: Conclusion

This section provides a summary of the key findings and conclusions of the study. It also provides identifies limitations encountered during the study and suggests directions for future research to further refine and advance the integration of AI with satellite imagery for disaster response.



2. LITERATURE REVIEW

2.1 Introduction

Disasters, whether natural or man-made, present substantial challenges to human society by causing loss of life, destruction of property, and severe disruptions to essential services. In the context of ground-based disasters, the role of rapid and accurate damage assessment is critical in enabling effective disaster response and recovery efforts. The use of satellite and Unmanned Aerial Vehicle (UAV) imagery in disaster management has gained significant attention due to its ability to provide real-time, high-resolution data over large areas. This imagery, combined with artificial intelligence (AI) techniques, holds the potential to revolutionize the way disaster response teams understand and react to affected areas (Gupta, 2020).

The application of AI in analyzing satellite and UAV imagery offers a powerful approach to disaster management, particularly in the response phase. By automating the process of damage assessment, AI can significantly enhance the accuracy and speed of decision-making, thereby improving the effectiveness of disaster response (Al Shafian, Integrating Machine Learning and Remote Sensing in Disaster Management: A Decadal Review of Post-Disaster Building Damage Assessment, 2024). However, the integration of AI into disaster management is still an emerging field, and there is a need to critically examine existing methodologies, assess their strengths and limitations, and identify areas for improvement.

This chapter provides a comprehensive overview of the existing research on the use of AI for analyzing satellite and UAV imagery in disaster response. It synthesizes relevant studies, identifies gaps in the literature, and examines how AI-driven methodologies are utilized to compare pre- and post-disaster images for generating accurate damage assessment maps. The focus is on demonstrating the capability of AI to detect and map damage in disaster-stricken areas efficiently. The chapter begins with a discussion of key concepts related to disaster response, the role of satellite and UAV imagery, and the evolution of AI in this domain. It then reviews the current state

of research on AI applications in disaster management, with a particular emphasis on the response phase, and assesses the effectiveness of these technologies in real-world scenarios, such as the 6 february 2023 Kahramanmaraş Earthquakes.

This literature review chapter lays the groundwork for the subsequent development and application of AI models for damage assessment, providing a thorough understanding of the current landscape and highlighting the potential of AI to enhance disaster response efforts.

2.2 Theoretical Background

The theoretical background of disaster management delves into the complex interplay of various factors that shape disaster response and recovery efforts. These factors span across environmental, social, economic, political, and technological domains, each contributing to the overall effectiveness and efficiency of disaster management strategies. Understanding these factors is crucial for developing robust frameworks to mitigate the impact of disasters and enhance resilience in affected communities.

At its core, disaster management theory explores the processes involved in preparedness, response, recovery, and mitigation (Khan H. V., 2008). Response, on the other hand, focuses on the immediate actions taken during a disaster to save lives, provide relief, and minimize damage. Traditional information-gathering methods such as field surveys, eyewitness reports, aerial reconnaissance, and reliance on pre-disaster maps play a central role in this framework. These approaches are based on the assumption that ground teams can access affected areas and provide timely, accurate information to decision-makers. However, this reliance on manual data collection often results in delays due to physical barriers, disrupted communication infrastructure, and the limited scope of human observation. Furthermore, the use of pre-disaster data, while helpful for initial planning, does not account for real-time changes and often lacks the spatial and temporal resolution needed for precise impact assessment.

In addition to these core processes, the theoretical background also examines key concepts such as vulnerability, risk assessment, and resilience. Vulnerability refers to the susceptibility of individuals, communities, or systems to the negative effects of disasters, influenced by factors like socioeconomic status, geographical location, and

access to resources. Risk assessment involves identifying potential hazards, analyzing their likelihood, and evaluating the potential consequences, which helps in developing informed decision-making processes. Resilience, a central concept in modern disaster management theory, focuses on the capacity of systems and communities to recover from disasters, adapt to changes, and thrive despite challenges (Fekete, 2014).

Furthermore, interdisciplinary approaches play a significant role in disaster management theory. The integration of knowledge from fields like sociology, psychology, engineering, environmental science, and information technology allows for a more holistic understanding of disaster dynamics. For example, psychological factors influence how individuals and communities perceive and react to disaster threats, while advancements in technology, such as satellite imagery and AI, are increasingly shaping disaster response strategies by providing real-time data and predictive insights.

Ultimately, the theoretical background of disaster management provides the foundation for developing effective policies, strategies, and interventions that enhance disaster resilience, minimize loss of life, and facilitate the recovery process. By understanding the multifaceted nature of disasters and the factors that influence their impact, disaster management professionals can better prepare for, respond to, and recover from future events. Central to this discussion is the transformative role of Artificial Intelligence (AI) in enhancing the efficiency and accuracy of damage assessments, particularly through satellite and UAV (Unmanned Aerial Vehicle) imagery. As disasters become more complex and widespread, traditional methods of assessment—typically reliant on manual evaluations and subjective judgment—are increasingly inadequate in providing timely, reliable, or comprehensive data. These conventional approaches, which often involve on-the-ground assessments or the visual interpretation of images by human experts, face significant limitations in terms of speed, scope, and accuracy. This inadequacy underscores the urgent need for technological advancements, such as AI integration, to facilitate more rapid, precise, and data-driven decision-making processes (Shen, 2023).

AI-powered tools, especially deep learning models, have revolutionized the field of remote sensing by automating the analysis of satellite and UAV imagery. These advanced AI algorithms are designed to detect subtle environmental changes, such as building collapses, flood-induced land deformation, or shifts in infrastructure integrity,

that may be invisible to the human eye. By comparing pre- and post-disaster images, AI systems can create highly detailed, georeferenced damage assessment maps that identify and prioritize the most affected areas. This functionality is particularly critical in the early stages of disaster response, when the need for timely and accurate data is paramount to ensuring that resources and relief efforts are deployed effectively.

Moreover, the integration of AI into disaster management doesn't merely accelerate the damage assessment process; it also significantly enhances the accuracy and reliability of these assessments. Traditional damage assessment methods are often hampered by data limitations, subjective biases, and the slow pace of manual analysis. These challenges can result in errors, misinterpretations, and inconsistencies in the evaluation of disaster impact. In contrast, AI models, which are trained on vast, diverse datasets and continually refined through machine learning techniques, are capable of processing large volumes of data quickly and with increasing precision. By doing so, AI minimizes human error and offers more consistent, reproducible results.

The impact of AI extends beyond just improving speed and precision. By providing decision-makers with clear, data-driven insights into the extent of damage, AI aids in resource allocation, ensuring that aid is directed to the areas of greatest need. Furthermore, the scalability of AI technology means that it can handle the complexities of large-scale disasters, enabling global-scale assessments and responses in a fraction of the time it would take traditional methods. This efficiency can significantly reduce the lag time between disaster onset and the arrival of crucial aid, ultimately saving lives and improving the effectiveness of disaster relief efforts.

As AI continues to evolve, its role in disaster management will likely expand. The future of disaster response could see the integration of AI with other technologies, such as drones, robotics, and real-time data analytics, to create even more sophisticated, autonomous systems for assessing and responding to disasters. Innovations like these hold the potential to reshape disaster management as we know it, enabling more proactive, accurate, and efficient responses to the growing challenges posed by natural and human-made disasters. The increasing reliance on AI in this field offers an exciting opportunity to transform how we prepare for, respond to, and recover from disasters, making the process faster, more reliable, and more impactful.

2.2.1 Evolution of AI in remote sensing

The evolution of Artificial Intelligence (AI) in remote sensing has significantly transformed the way disasters are assessed and managed. Initially, the use of satellite and UAV imagery in disaster response was limited by manual methods of interpretation, which required significant time and expertise. Traditional remote sensing techniques involved analyzing images based on basic image processing tools that relied on human operators to detect changes in the landscape and identify objects of interest. These methods were labor-intensive and, although effective for small-scale analyses, were incapable of handling the massive volumes of data generated by modern satellite systems. Consequently, decision-making processes during disasters were often delayed, which could hinder timely response efforts. (Yang, 2019)

The advent of Artificial Intelligence (AI), particularly deep learning models such as Convolutional Neural Networks (CNNs), marked a significant turning point in the field of remote sensing, transforming how disaster assessments are conducted. Prior to AI's integration, traditional methods of analyzing satellite and UAV imagery were slow, manual processes that depended heavily on human expertise, often delaying timely response efforts. However, AI's ability to automate the analysis of vast amounts of imagery data has revolutionized this process. By enabling real-time data processing, AI has drastically reduced the time required to assess disaster areas, providing quicker insights and faster decision-making capabilities for emergency response teams.

Deep learning algorithms, particularly CNNs, have emerged as one of the most powerful tools in the field of remote sensing. CNNs are designed to recognize patterns and structures within images by learning hierarchical features at multiple levels, making them highly efficient for analyzing remote sensing data. They excel in tasks such as object recognition, image segmentation, and anomaly detection, which are crucial for identifying damage in disaster-affected areas. Unlike traditional image processing techniques, CNNs can automatically learn from vast amounts of data without the need for manual feature extraction, making them ideal for large-scale image analysis (Zhu, 2017).

When CNNs are trained on extensive, labeled datasets, they develop the ability to identify patterns, classify features, and detect subtle changes in the environment, such as building collapse, infrastructure damage, and changes in terrain. This capability

allows AI models to accurately assess disaster impact, even in complex and heterogeneous environments where traditional methods may struggle. As these models are exposed to more diverse datasets over time, they become more adept at handling different disaster scenarios, improving the precision and reliability of damage assessments.

Furthermore, as AI systems continue to learn and evolve through iterative training processes, their ability to identify patterns and anomalies becomes increasingly refined. This leads to faster, more accurate assessments of disaster damage, enabling rapid response efforts. In sum, the integration of CNNs and other deep learning models into remote sensing has not only streamlined the disaster assessment process but also significantly enhanced its accuracy, offering a more efficient and reliable way to guide disaster response activities (Ci, 2019).

Prior to AI's widespread adoption, remote sensing applications in disaster management faced limitations related to scalability and accuracy. Machine learning algorithms, introduced in the 2000s, began to address these shortcomings. Techniques such as Support Vector Machines (SVMs) and Random Forests were employed to classify satellite and UAV imagery more efficiently, although their performance was still constrained by the complexity and heterogeneity of disaster scenarios (Reyes et al., 2021). These algorithms were capable of improving the accuracy of damage assessments compared to manual methods, but they still struggled with complex and varied environmental conditions found in large-scale disaster areas. As a result, there was still a pressing need for a more advanced and scalable solution.

The breakthrough came with the development of deep learning algorithms, particularly CNNs, which are known for their ability to automatically learn features from data without the need for manual intervention. The ability of CNNs to detect fine-grained patterns in imagery, such as changes in terrain, building damage, and infrastructure destruction, made them a game-changer for disaster management. By the 2010s, CNNs had become the dominant tool in remote sensing applications, enabling the processing of high-resolution satellite and UAV images at a scale and speed previously unattainable. Additionally, architectures like U-Net and ResNet further improved the models' ability to segment and classify detailed features in images, allowing for more accurate and detailed damage maps.

AI-driven remote sensing has not only accelerated the speed at which data is processed but also enhanced the quality of the output. By automating the identification and classification of objects in images, AI models have significantly reduced human error and inconsistencies. Moreover, these models can be trained on diverse datasets, allowing them to adapt to different disaster scenarios, such as floods, earthquakes, and wildfires, making them versatile tools in disaster response. AI can also handle large datasets with greater precision, enabling more comprehensive analyses and supporting better decision-making.

However, the integration of AI into remote sensing also comes with challenges. One of the primary limitations is the need for high-quality, labeled data to train deep learning models effectively. Furthermore, AI models are computationally intensive and require significant resources to process and analyze large datasets. Despite these challenges, the potential of AI to revolutionize disaster management through real-time, automated damage assessments is undeniable. As AI continues to evolve, it is expected that its applications in remote sensing will only expand, bringing further improvements to the field of disaster management and response (Vetrivel, Challenges of Artificial Intelligence Application for Disaster Risk Management, 2023).

In the past, the reliance on human operators to manually assess satellite and UAV imagery resulted in delays in the delivery of actionable information, hindering response times. With AI, the process of analyzing large volumes of remote sensing data has become automated and far more efficient. Deep learning models can process images faster than human analysts, enabling disaster management teams to receive damage assessments almost immediately after data is collected. This speed in data processing is crucial in the early stages of disaster response when timely decisions are critical to saving lives and minimizing further damage.

The ability of AI to handle vast amounts of data quickly and with a high degree of accuracy has proven particularly beneficial in the case of large-scale disasters. In events such as earthquakes, tsunamis, and hurricanes, the extent of the damage can span large areas, making it difficult for human analysts to keep up with the volume of imagery generated by satellites and UAVs. AI models trained on extensive datasets can process this imagery almost instantaneously, identifying affected areas and providing damage assessments that are both timely and precise. This allows emergency

responders to focus their efforts where they are most needed, directing resources to areas with the greatest damage and ensuring that aid is distributed efficiently.

The potential of AI in remote sensing is not limited to its ability to process data faster. It also extends to the quality of the damage assessments generated. AI models can detect subtle changes in the environment that may be imperceptible to the human eye. For example, AI can identify small shifts in terrain caused by earthquakes, the collapse of buildings, or changes in vegetation after a wildfire. This fine-grained analysis allows disaster management teams to gain a more detailed understanding of the damage, which can lead to more targeted and effective response strategies.

Furthermore, AI models are not static; they improve over time as they are exposed to more data. Through a process known as transfer learning, AI models can be fine-tuned and adapted to specific types of disasters, making them even more effective at detecting relevant features in imagery. For instance, a model trained on earthquake damage data can be further trained on flood damage data, allowing it to adapt to the unique characteristics of flooding events. This adaptability makes AI-driven remote sensing tools a versatile asset for disaster management teams, capable of handling a wide range of disaster scenarios.

The scalability of AI-driven remote sensing is another significant advantage. Traditional methods of image analysis struggled with the vast amounts of data produced by modern satellite and UAV systems, but AI is capable of processing large datasets with ease. This scalability allows AI models to handle data from not just one disaster, but multiple concurrent events. For example, if a region is affected by both an earthquake and a flood, AI models can simultaneously analyze satellite imagery from both events, producing separate damage assessments for each. This ability to analyze multiple disaster scenarios at once is a significant improvement over traditional methods, which often required separate analyses for each event (Al Shafian, 2024).

Despite the many advantages, there are still challenges to the widespread adoption of AI in remote sensing for disaster management. One of the primary challenges is the need for high-quality, labeled data to train deep learning models. AI models rely on large datasets of labeled imagery to learn to recognize patterns and detect relevant features. However, creating these datasets is time-consuming and resource-intensive,

particularly for disaster scenarios that are rare or complex. Additionally, the data used to train AI models must be of high quality to ensure that the models are able to make accurate predictions. Poor-quality data or mislabeled imagery can lead to inaccurate damage assessments and undermine the effectiveness of AI in disaster management.

Another challenge is the computational cost of training and running deep learning models. Remote sensing applications often require processing large volumes of high-resolution imagery, which can be computationally expensive. To handle this demand, AI models require powerful hardware, such as Graphics Processing Units (GPUs), and substantial computational resources, often provided by cloud-based platforms. This can be a barrier to the widespread adoption of AI, particularly in low-resource settings or in regions where access to the necessary infrastructure is limited.

AI models are also not foolproof and can sometimes struggle with certain types of imagery or environmental conditions. For example, AI models may struggle to detect damage in images taken under poor lighting conditions or in areas where there is heavy cloud cover. Similarly, images that are noisy or incomplete can pose challenges for AI models, which may fail to detect important features or produce inaccurate results. Despite these challenges, researchers are continuously working to improve the robustness of AI models and to develop techniques for mitigating these issues, such as through data augmentation or by incorporating additional sources of information, such as ground-based data (Oubara, 2023).

The potential of AI to revolutionize disaster management remains immense. As technology continues to improve, AI-driven remote sensing systems are expected to become even more sophisticated and capable of handling increasingly complex disaster scenarios as shown in Table 2.1. Through continued research and development, the integration of AI into disaster response efforts will likely expand, enabling faster, more accurate assessments of disaster damage and enhancing the overall efficiency of disaster management operations. By automating the analysis of satellite and UAV imagery, AI is helping to address the challenges faced by disaster management teams and is ultimately contributing to more effective and timely responses to disasters worldwide.

Table 2.1: Evolution of AI in remote sensing (Generated by Author).

Period	Technology	Description	Advantages	Challenges
Pre-2000s	Manual Methods	Basic image processing, manual classification	Simple to implement, low cost	Low accuracy, labor intensive
2000-2010	Machine Learning	SVMs, Random Forests for classification and detection	Improved accuracy and automation	Limited scalability
2010-Present	Deep Learning	CNNs, U-Net, ResNet for advanced analysis	High accuracy, scalability, efficiency	High computational requirements, need for large datasets

2.2.2 Disaster response and AI integration

The integration of Artificial Intelligence (AI) into disaster response has fundamentally reshaped how emergency management organizations assess, respond to, and recover from catastrophic events. In the past, disaster management faced significant challenges in quickly processing and analyzing vast amounts of data in the wake of disasters, which often delayed decision-making and the allocation of resources. The ability to rapidly assess damage, predict further risks, and deploy aid effectively is crucial to minimizing loss of life and mitigating long-term impacts. By leveraging AI technologies such as deep learning, computer vision, and machine learning, disaster response has become more timely, accurate, and adaptable to evolving situations (Vetrivel, 2023).

One of the most notable contributions of AI to disaster response is its capacity to process and analyze massive datasets in real time, especially imagery data from satellites and UAVs. Satellite imagery, which can cover vast areas and provide high-resolution images, plays an essential role in disaster response by capturing the scope of damage across large regions. Similarly, UAVs equipped with cameras can provide localized, high-resolution images of affected areas, often at a much more granular level. Traditionally, the analysis of this imagery required significant human input and expertise, which could result in delays and inefficiencies in the disaster response. However, the advent of AI has revolutionized the process by automating data analysis, allowing for faster, more accurate identification of damaged areas (Li et al., 2020).

The speed at which AI systems can analyze satellite and UAV imagery allows for real-time damage assessments, enabling emergency responders to prioritize resources where they are most needed. During the 2023 Turkey-Syria earthquakes, for example, AI-driven systems were used to process satellite images taken immediately before and after the event, quickly identifying the areas with the most severe damage. By analyzing these images, AI models could highlight collapsed buildings, roads, and other critical infrastructure, giving disaster response teams an accurate and up-to-date understanding of the situation on the ground. This ability to assess damage rapidly allowed responders to prioritize their efforts, ensuring that aid was sent to the most affected areas first (Li et al., 2020).

AI also plays a pivotal role in improving situational awareness by providing detailed and real-time data on the progress of a disaster and its ongoing impact. For example, AI models can track the spread of wildfires or floods, predicting which areas are likely to be affected next and allowing disaster managers to take proactive measures. AI systems can process not only satellite and UAV imagery but also data from ground-based sensors, social media feeds, and other sources, creating a comprehensive and integrated view of the disaster's evolution. This enhanced situational awareness enables more informed decision-making, ensuring that resources are deployed effectively and rapidly, and that relief efforts are coordinated efficiently (Vetrivel, 2023).

The application of AI in disaster response also helps to reduce human error and bias in damage assessments. Traditionally, human analysts would manually assess damage based on visual inspections of satellite imagery or aerial photos. However, this process is subject to inconsistencies, misinterpretations, and human limitations. AI, on the other hand, uses large datasets to identify patterns and detect anomalies, allowing for consistent and objective assessments. Deep learning algorithms, such as Convolutional Neural Networks (CNNs), have been trained on vast amounts of labeled disaster imagery and can recognize subtle changes in the environment that may be overlooked by human analysts. This ability to detect fine details, such as small cracks in infrastructure or the displacement of materials, makes AI an indispensable tool for generating highly accurate damage assessment maps.

In addition to improving the accuracy of damage assessments, AI's capacity for predictive analysis has been a major asset in disaster management. Predictive models,

when combined with real-time data from satellites, UAVs, and ground sensors, can forecast future damage or identify areas at high risk of further disaster-related impacts. For instance, in flood-prone regions, AI models can analyze river levels, rainfall data, and topography to predict where flooding is likely to occur. This capability allows disaster managers to allocate resources ahead of time and plan evacuation strategies in advance, minimizing the risk to human life and infrastructure (Zhewei Liu, 2024). AI can also model the behavior of other natural disasters, such as earthquakes or tsunamis, providing early warnings that can help save lives and reduce damage.

Furthermore, AI can enhance the overall efficiency of disaster response operations. By automating the analysis of satellite and UAV imagery, AI can significantly reduce the time required to generate damage maps, freeing up human resources to focus on other critical tasks. In some cases, AI-powered systems can even generate damage maps in real time, allowing emergency teams to make immediate decisions about where to send personnel and resources. This capability is especially important in fast-moving disasters, such as hurricanes or wildfires, where time is of the essence and rapid decision-making can be the difference between life and death (Goodchild & Glennon, 2010).

Despite these remarkable advancements, there are challenges in integrating AI into disaster response, particularly in terms of data quality and computational resources. For AI models to be effective, they require large volumes of high-quality, labeled data to train and validate the algorithms. In disaster scenarios, however, obtaining such data can be difficult, particularly in regions with limited infrastructure or where disasters have caused widespread destruction. Moreover, real-time AI processing of satellite and UAV imagery demands significant computational power, which may be unavailable in disaster-stricken areas. As a result, disaster response teams must rely on cloud-based systems or mobile processing units to ensure that AI models can operate effectively in the field (Reyes et al., 2021).

Another challenge is the need for ongoing refinement of AI models. Although AI systems have shown great promise in disaster response, they are not infallible. False positives and false negatives in damage assessments can lead to misallocation of resources or delays in aid delivery. To overcome this, AI models must be continuously updated and refined based on new data, feedback from disaster response teams, and lessons learned from previous events. This requires ongoing collaboration between AI

researchers, disaster management experts, and technology developers to ensure that AI systems are both accurate and reliable in real-world scenarios (Reyes et al., 2021).

In conclusion, the integration of AI into disaster response has transformed how emergency management teams assess and respond to disasters. Through its ability to process and analyze large datasets in real-time, improve situational awareness, and generate accurate damage assessments, AI has become an invaluable tool for disaster response. While there are challenges related to data quality, computational resources, and the continuous improvement of AI models, the benefits of AI in disaster management are clear. As technology continues to advance, AI's role in disaster response is expected to expand, offering even more sophisticated tools for saving lives and reducing the impact of disasters.

2.2.3 Comparative analysis of pre- and post-disaster imagery

The comparative analysis of pre- and post-disaster imagery using Artificial Intelligence (AI) is a pivotal process in disaster response, as it provides critical insights into the extent of the damage and supports decision-making processes aimed at coordinating relief efforts. When a disaster strikes, it is essential for emergency responders to quickly assess the scale and impact of the damage, which requires high-resolution, up-to-date imagery of the affected areas. AI technologies, particularly deep learning models, have significantly enhanced the ability to compare satellite and UAV imagery before and after a disaster, enabling a more accurate and timely understanding of the destruction. This ability to detect changes in the environment at a granular level is crucial for informing decisions regarding resource allocation, aid distribution, and recovery efforts (Hanxiang Hao, 2020).

In disaster management, understanding the full extent of damage is vital for prioritizing which areas require immediate attention and intervention. Traditional damage assessment methods relied heavily on manual interpretation of imagery, often leading to delays in decision-making due to the sheer volume of images that needed to be processed. These methods were also susceptible to human errors, inconsistencies in analysis, and lack of real-time data, which hindered the effectiveness of disaster response. However, AI algorithms, particularly deep learning models such as Convolutional Neural Networks (CNNs), have transformed this process by automating the analysis of vast amounts of imagery data in real-time. By comparing pre- and post-

disaster imagery, AI can identify even the smallest changes in the environment, such as damaged buildings, road collapses, or alterations in the landscape (Cheng et al., 2021).

A key advantage of AI in comparing pre- and post-disaster imagery is its ability to detect and quantify subtle changes that might otherwise go unnoticed. AI models trained on large datasets of satellite and UAV imagery learn to recognize patterns and anomalies, such as structural damages to buildings, flooding, or altered terrain. This enables AI systems to generate detailed damage maps that highlight areas with the most significant destruction. The ability to detect even minor changes in the landscape is essential, as it ensures that no affected area is overlooked during the response phase. Moreover, these AI-generated damage maps provide a consistent and objective assessment of the situation, reducing the risks of bias and human error in damage evaluation (Zhu et al., 2017).

The use of AI in comparing pre- and post-disaster imagery is particularly beneficial in rapidly evolving disaster scenarios, where time is of the essence. For example, during the 6 February 2023 Kahramanmaraş Earthquakes, AI algorithms were used to process high-resolution satellite images taken before and after the disaster. By identifying changes in the environment, AI models could quickly pinpoint the most severely affected areas. This allowed disaster response teams to focus their efforts on regions that needed immediate aid, while also providing an accurate representation of the damage to help with resource allocation. The speed and accuracy with which AI could generate these damage maps significantly improved the efficiency of the response efforts and helped reduce the time required to mobilize aid (Cheng et al., 2021).

In addition to generating damage maps, AI can also be used to quantify the extent of the damage in terms of specific metrics, such as the percentage of buildings or infrastructure that has been affected, the volume of debris in a particular area, or the extent of flooding. These quantitative assessments provide valuable insights into the scale of the disaster and are essential for guiding relief operations. For instance, AI can assess the number of collapsed buildings in a given region, estimate the volume of debris, or measure the displacement of soil or vegetation. These metrics are crucial for understanding the severity of the disaster and determining the resources needed for recovery and reconstruction (Lu, 2020) .

Another key benefit of using AI for comparative analysis of pre- and post-disaster imagery is its ability to integrate multiple data sources, such as satellite imagery, UAV data, and ground-based sensors. By combining these different data types, AI models can create a more comprehensive picture of the disaster's impact, taking into account both large-scale damage and localized destruction. For example, satellite imagery provides a broad overview of the affected area, while UAVs can capture high-resolution images of specific sites, offering a more detailed view of the damage. AI models can then synthesize these different sources of information, producing a more accurate and holistic assessment of the disaster (Adriano, 2019).

AI can also facilitate the integration of temporal data, enabling a dynamic comparison of how the affected area evolves over time. For example, AI models can analyze a series of pre- and post-disaster images taken over several days or weeks to monitor the progress of the disaster, including the spreading of fires or floods. This time-sensitive analysis helps emergency responders track ongoing damage and make informed decisions regarding evacuation, resource distribution, and recovery efforts. It also allows for the identification of secondary risks, such as landslides or aftershocks, that may not have been immediately apparent following the initial disaster (Reyes et al., 2021).

Despite the promising advantages, there are challenges associated with using AI for comparative analysis of pre- and post-disaster imagery. One major limitation is the availability and quality of imagery data, which can vary depending on the region and the type of disaster. In some cases, it may be difficult to obtain high-resolution images of the affected area, especially if the disaster has caused widespread destruction or if the area is remote. Additionally, the accuracy of AI models is heavily dependent on the quality and quantity of data used to train them. Inaccurate or incomplete data can lead to misinterpretation of the damage and undermine the effectiveness of AI-driven analysis (Zhu, 2017).

Moreover, the complexity of analyzing disaster imagery using AI requires significant computational resources. AI models that analyze large volumes of high-resolution imagery require powerful hardware and advanced algorithms to process the data efficiently. This can be a challenge in disaster-stricken areas, where access to technological infrastructure may be limited. To address this issue, cloud-based platforms and mobile processing units are often used to facilitate real-time data

processing in the field, ensuring that AI models can operate effectively even in remote or disaster-affected regions (Zhu, 2017).

Despite these challenges, the use of AI for comparing pre- and post-disaster imagery has proven to be a highly effective tool in disaster response and recovery. As technology continues to evolve, it is expected that AI-driven systems will become even more advanced, capable of providing even more accurate, real-time insights into the effects of disasters. By enhancing situational awareness, reducing human error, and improving the efficiency of damage assessments, AI will continue to play a critical role in helping emergency response teams save lives, allocate resources effectively, and rebuild communities after a disaster. The key evaluation criteria for such AI-based damage assessment systems are summarized in Table 2.2.

Table 2.2: Evaluation of damage assessment maps (Generated by Author)

Criteria	Description	Measurement
Accuracy	How well the damage maps reflect actual damage	Precision, Recall, F1 Score
Reliability	Consistency of results across different conditions	Cross-validation, Consistency checks
Utility	Practical usefulness in decision-making	User feedback, Real-world impact

2.3 Previous Study

Recent advancements in the integration of artificial intelligence (AI) with satellite and UAV imagery have significantly impacted disaster management practices. The use of AI in disaster management has the potential to dramatically improve the speed, accuracy, and efficiency of damage assessments during natural disasters. Satellite and UAV imagery have become essential tools for monitoring disaster-affected areas in real-time, and AI algorithms are increasingly being employed to analyze this imagery. The intersection of these technologies offers a transformative approach to disaster response and recovery.

One of the pivotal studies in this field was conducted by Zhang et al. (2020), who explored the application of deep learning algorithms, specifically convolutional neural networks (CNNs), to analyze flood imagery. Their research demonstrated that CNNs could process remote sensing images to accurately identify and classify flood-affected

areas. This capability not only enhanced the speed of damage assessment but also provided a more reliable basis for decision-making during flood emergencies. By automating the analysis of large datasets, AI can significantly reduce the time required for damage assessments, which is crucial in the fast-paced environment of disaster management. The ability of AI to handle large datasets rapidly and with high precision underscores its potential to transform disaster response operations, providing authorities with actionable intelligence that can be used to mobilize aid and resources more effectively (Zhang, 2020).

Zhang et al.'s (2020) study emphasizes the importance of using AI for flood mapping and classification. By employing deep learning algorithms to process high-resolution satellite images, their approach was able to distinguish between flood-affected and non-flood-affected areas with greater accuracy than traditional methods. This level of precision can significantly improve the coordination of emergency response efforts, ensuring that aid reaches the areas most in need. Additionally, the use of AI in flood mapping allows for continuous monitoring of flood events, enabling real-time updates on the evolving situation. As a result, AI-powered flood monitoring systems can help improve the decision-making process during a disaster, making the response efforts more effective and timely.

Expanding on the use of AI in disaster management, Liu et al. (2021) investigated the synergy between UAVs and machine learning for wildfire monitoring. Their study utilized UAV imagery combined with sophisticated AI algorithms to track the progression of wildfires in real-time. Wildfires are one of the most destructive natural disasters, and monitoring them using traditional methods has often been challenging due to the speed and unpredictability of the fire. Liu et al. (2021) demonstrated that by integrating UAVs with AI models, researchers were able to track wildfire progression with a high degree of accuracy. This integration allowed for real-time updates and predictions regarding the spread of the fire, which is essential for effective emergency response and evacuation planning.

The application of AI in wildfire monitoring is particularly significant because it offers a detailed, localized view of the affected areas. UAVs are capable of capturing high-resolution images that provide valuable information on the state of the fire and its potential spread. AI algorithms can then analyze this imagery to identify patterns and make predictions about future fire behavior. By leveraging real-time data, AI systems

can provide emergency responders with timely information, which can be used to guide evacuation efforts, allocate resources, and minimize the loss of life and property. Liu et al. (2021) also highlighted the ability of AI to analyze large volumes of data from UAVs and other sensors, enabling faster decision-making and improving the overall efficiency of wildfire management (Liu, 2021).

The synergy between UAVs and AI also highlights the increasing role of remote sensing technologies in disaster management. UAVs are particularly useful in disaster scenarios where access to affected areas is limited or hazardous. For instance, UAVs can be deployed to monitor wildfire progression in remote areas where traditional methods may not be feasible. They can also be used to assess the impact of other natural disasters, such as earthquakes or floods, by capturing high-resolution imagery of the affected areas. By combining UAV technology with AI, it is possible to generate highly accurate damage assessments and monitor disaster events in real time.

The application of UAVs and satellite imagery in the context of earthquake response has also been a subject of significant research. One of the key studies in this area was conducted by Tang et al. (2022), who focused on the use of these technologies to assess earthquake damage in urban environments. The study revealed that combining high-resolution satellite images with UAV data and analyzing them through AI algorithms facilitated rapid and precise damage mapping. This approach enabled the researchers to evaluate the extent of the damage in affected areas with high accuracy. Traditional methods of damage assessment are often time-consuming and resource-intensive, especially in urban environments where access to affected areas can be difficult. The use of satellite and UAV imagery combined with AI algorithms allows for faster and more accurate damage assessment, which is essential for coordinating relief efforts and allocating resources effectively.

The importance of using multiple data sources to improve the accuracy of damage assessments. By combining satellite imagery with UAV data, researchers were able to capture a more comprehensive view of the disaster-affected areas. Satellite imagery provides a broad overview of the region, while UAVs offer a more detailed, localized perspective. The integration of these data sources with AI algorithms allows for the creation of highly accurate damage assessment maps, which can be used by disaster response teams to prioritize their efforts and allocate resources more effectively (Adriano, 2019).

Despite the promising advancements in AI-powered disaster management, several studies have also addressed the challenges associated with the use of AI in this field. They examined issues related to data privacy and algorithmic bias within AI systems. They pointed out that while AI technologies offer numerous benefits, they also raise concerns about the protection of sensitive data and the potential for biased outcomes. AI systems rely on large datasets, and the quality and diversity of these datasets are crucial for ensuring accurate and fair results. If the data used to train AI models are biased or incomplete, the outcomes of the analysis may be inaccurate or unfair. For example, AI algorithms trained on data from one region may not perform as well when applied to a different region with different characteristics.

Lee et al. (2023) emphasized the importance of establishing robust data governance frameworks to address these challenges. They argued that it is essential to ensure the privacy and security of sensitive data, particularly in the context of disaster response. AI models often rely on large amounts of data, including satellite and UAV imagery, which can contain sensitive information about the affected areas. To protect this data, it is necessary to implement strong data protection policies and procedures. In addition, continuous evaluation of AI systems is needed to identify and address any biases that may arise. Ensuring that AI systems are transparent, fair, and accountable is crucial for maintaining public trust in these technologies.

These studies collectively illustrate the transformative potential of combining AI with satellite and UAV imagery in disaster response. The ability of AI to process and analyze vast amounts of imagery data enhances the speed and accuracy of damage assessments, improves real-time monitoring, and supports strategic decision-making. However, addressing the inherent challenges, such as data privacy and algorithmic bias, is essential for the responsible deployment of these technologies. As the field continues to evolve, ongoing research and technological advancements are expected to drive further improvements in disaster management practices.

Moreover, the integration of AI in disaster management is not without its technical challenges. For AI to be effective, it requires access to high-quality, labeled data. In disaster scenarios, obtaining sufficient high-quality data can be difficult, especially in regions where infrastructure has been damaged or destroyed. In such cases, the quality of satellite and UAV imagery may be compromised, which in turn affects the performance of AI models. Furthermore, real-time data processing in disaster-stricken

areas demands significant computational resources. The lack of adequate computing power in remote or devastated areas can hinder the deployment of AI systems, especially when they require cloud-based solutions or powerful edge computing platforms (Żarski, 2024).

In response to these challenges, several studies have highlighted the importance of developing more robust AI models that are capable of handling incomplete or noisy data. Furthermore, advancements in edge computing and the use of mobile devices with built-in AI capabilities can help alleviate some of the issues related to limited computational resources. These technologies allow for AI-powered analysis to be conducted closer to the source of data collection, reducing the reliance on centralized cloud infrastructure and enabling faster decision-making on the ground.

Building on previous studies that utilized models like U-Net and CNNs for structure recognition and damage assessment, this study introduces a deep learning pipeline based on U-Net trained on xBD and Microsoft Building Footprint datasets. By focusing on the Antakya region, this research seeks to generate a reliable segmentation model capable of identifying buildings in satellite images (Touzani, 2021).

Finally, looking ahead, the role of AI in disaster management is expected to expand significantly as the technology continues to evolve. One area of focus is the integration of AI with Internet of Things (IoT) devices and sensor networks, which could provide real-time data from the ground level, further improving the situational awareness of disaster response teams. Additionally, AI's predictive capabilities are expected to become more sophisticated, allowing for more accurate forecasting of natural disasters such as hurricanes, earthquakes, and floods. As AI models become increasingly advanced, they will be able to account for more variables and provide more precise predictions, enabling disaster management teams to take preemptive measures before disaster strikes.

3. RESEARCH METHODOLOGY

3.1 Introduction

The methodology adopted in this research is fundamentally interdisciplinary, combining elements from remote sensing, geospatial analysis, computer vision, and disaster management. The approach is both analytical and applied, meaning it not only explores theoretical advancements in AI-driven disaster assessment but also seeks practical implementations that can be used in real-world scenarios. Given the increasing frequency and severity of natural disasters, this research is designed to contribute directly to disaster response efforts by providing a scalable and automated method for assessing structural damage, identifying affected areas, and guiding relief operations.

At the heart of this research is the use of satellite imagery, which offers a bird's-eye view of disaster-affected regions. These images, captured before and after a disaster, contain rich information about the extent of damage to infrastructure, land, and communities. However, raw satellite images alone are insufficient without proper analysis. The sheer volume of data generated by remote sensing platforms makes manual assessment impractical. To overcome this challenge, AI-driven techniques, particularly deep learning models, are employed to extract meaningful patterns from the images, detect damage with high accuracy, and classify affected areas based on severity.

A key component of this study is the selection and processing of data. High-resolution satellite images from sources such as Google Earth images are utilized to ensure comprehensive coverage. These images are carefully preprocessed to enhance their quality, standardize formats, and align spatially. Data preprocessing plays a crucial role in ensuring that AI models receive clean, structured inputs, which ultimately improves the accuracy of damage assessments.

Furthermore, the methodology takes into account potential challenges and limitations. Factors such as cloud cover in satellite imagery, data availability constraints, and

computational resource requirements are carefully analyzed. Addressing these challenges is an essential part of this research, as overcoming them can enhance the robustness and scalability of AI-powered disaster assessment tools.

3.2 Applications and Software Used

Multiple software applications were utilized throughout this research to support various stages of data processing, analysis, and visualization. Each application played a crucial role in ensuring accurate and efficient disaster assessment.

I. Google Earth Engine

Google Earth Engine (GEE) is a cloud-based geospatial analysis platform that provides access to a vast repository of satellite imagery and remote sensing data. It was used in this research to retrieve pre- and post-disaster images, preprocess data, and conduct initial change detection analyses. The ability to process large-scale geospatial datasets in the cloud made GEE an essential tool for this study.

II. QGIS

QGIS is an open-source Geographic Information System (GIS) used for spatial data visualization, georeferencing, and mapping. It was employed in this research for overlaying damage assessment results on geographical maps, integrating multiple datasets, and ensuring spatial alignment of images. QGIS facilitated the visualization of AI-generated damage maps in a user-friendly and interactive manner.

III. Python

Python was the primary programming language used in this research, along with various libraries.

IV. Google Drive

Google Drive was used as a cloud storage platform for organizing and storing datasets, model outputs, and project documentation. Its seamless integration with Google Colab allowed for efficient dataset management and model checkpoints.

V. Jupyter Notebook

Jupyter Notebook was used as an interactive coding environment for exploratory data analysis, visualization, and model experimentation. It provided a structured way to

document the step-by-step process of AI-based damage assessment, making it easier to analyze intermediate results and refine the methodology.

By leveraging these tools and software applications, this research was able to efficiently process large volumes of satellite imagery, train deep learning models, and visualize before and after of the disaster effected area, ultimately contributing to a more effective AI-powered disaster response framework.

3.3 Data Processing

The data processing phase began with the retrieval of satellite images from Google Earth Pro for both pre-disaster and post-disaster conditions. These images were of extremely high resolution, with each file being approximately 2 GB in size. Due to their large size, efficient data handling and storage solutions were necessary to manage the dataset.

Once the images were retrieved, they were imported into QGIS for preprocessing. This included georeferencing, image enhancement, and spatial alignment. Various preprocessing techniques were applied in QGIS, such as noise reduction, histogram equalization for contrast enhancement, and image clipping to focus on specific areas of interest. The processed images were then converted into formats compatible with AI models for further analysis. This structured data processing workflow ensured that the satellite imagery was optimized for deep learning-based damage detection and analysis.

3.4 Building Footprint

Building footprint data has been used in a variety of applications, from urban planning and disaster management to environmental monitoring and infrastructure development. Over the years, numerous methods have been proposed and employed to obtain accurate building footprint data. However, many of these methods have faced significant challenges, including low accuracy, limitations in scale, lack of detail, and incomplete or outdated data. Despite these challenges, advancements in remote sensing technologies, machine learning algorithms, and geospatial data collection have significantly improved the accuracy and availability of building footprint data.

In the context of disaster management and post-disaster analysis, the ability to quickly and accurately assess the impact of disasters on urban areas is critical. Building footprint data plays a crucial role in this process by helping to identify the extent of damage to buildings, the destruction of infrastructure, and the impact on human settlements. The building footprint data used in this study comes from Microsoft, which has developed a high-quality dataset for identifying the geographical locations of buildings using satellite imagery and machine learning techniques. While other sources of building footprint data were considered, Microsoft's dataset was ultimately chosen due to its higher accuracy and availability for the region of interest.

Many methods have been used to extract building footprints from satellite imagery, remote sensing data, or aerial photography. Machine learning-based methods, particularly those leveraging deep learning algorithms, have gained popularity for building footprint extraction in recent years. These methods rely on convolutional neural networks (CNNs) or other deep learning models to automatically detect buildings from satellite imagery. While deep learning models can offer high accuracy, they often require large labeled datasets for training and are computationally intensive. Furthermore, the quality of the results can be heavily influenced by the quality and resolution of the input data, as well as the training data used to develop the model.

Despite the advantages of these methods, they also have limitations. For example, many of the available building footprint datasets suffer from incomplete or inaccurate data, particularly in regions with poor satellite imagery coverage or limited data availability. These limitations can result in inaccurate building footprints that affect subsequent analysis, such as damage assessment in post-disaster situations.

Given the challenges and limitations of the methods mentioned above, it was decided to use Microsoft's building footprint dataset for this study. Microsoft's dataset is a comprehensive collection of building footprints that has been derived from satellite imagery and machine learning algorithms. The primary reason for selecting Microsoft's building footprint dataset was its superior accuracy compared to other available datasets. Microsoft's dataset has been developed using high-resolution satellite imagery and advanced machine learning techniques, making it highly reliable for large-scale applications, including disaster management and damage assessment. Additionally, the dataset is continuously updated, ensuring that the data remains

relevant and reflects the most current conditions on the ground. Another key advantage of using Microsoft's dataset is its availability through an open-access platform. The dataset is freely available for download from Microsoft's GitHub repository, which provides access to the building footprints in various formats, including GeoJSON and shapefiles. This made it easier to integrate the data into the analysis pipeline and work with it using common geospatial tools, such as Python and QGIS.

The building footprint data from Microsoft was downloaded using the Python program "make-gis-friendly.py," which is part of Microsoft's GitHub repository. This program is specifically designed to convert the raw building footprint data into a format suitable for use in Geographic Information Systems (GIS). To ensure the data was relevant to the region of interest, the location was set to Turkey, which generated several small files corresponding to different regions of the country.

One of the main challenges encountered during this process was the fact that the downloaded files were geographically divided into smaller regions. While this approach made it easier to manage the data, it also introduced the problem of not knowing which file corresponded to which geographic area. There was no direct labeling system for the files, meaning that each file had to be opened and examined to determine its location. This added an extra step in the data preparation process and required a systematic approach to ensure that the correct files for the study area were selected.

To resolve this challenge, all the files were loaded into QGIS, a popular open-source GIS software, to visualize their locations. QGIS provides powerful tools for working with geospatial data, allowing users to load, visualize, and manipulate geographic datasets.

After loading the files into QGIS, the next step was to extract the building footprints for the areas of interest. The building footprint data from Microsoft was initially in GeoJSON format, which is widely used for representing geographic features in a web-based format. In QGIS, the footprints were visualized alongside other geographic data layers, such as pre-event satellite images of the affected areas. To ensure that the building footprints were properly aligned with the study area, the pre-event satellite images were used to clip the footprints to the extent of each area of interest (AOI).

Once the building footprints for the areas of interest were clipped, the next step was to convert the data into a more widely used format for geospatial analysis, namely shapefiles. Shapefiles are a standard format for storing vector data in GIS software and are compatible with many geospatial analysis tools. The building footprints were converted into shapefiles, which were then exported to the computer for further analysis.

The process of obtaining and preparing the building footprint data from Microsoft was a critical component of this study. While challenges were encountered during the data preparation phase, such as dealing with small regional files and ensuring proper alignment with the study area, the use of Microsoft's building footprint dataset provided a reliable and accurate source of data for assessing the impact of the 6 Feb 2023 Turkey earthquakes on the built environment. By combining this data with other geospatial datasets and using GIS tools such as QGIS, it was possible to conduct a comprehensive analysis of the earthquake's effects and support decision-making for disaster response.

3.5 General Process of Generating Building Footprint

In post-disaster scenarios, assessing the damage to buildings and infrastructure is a crucial step in disaster response and recovery. To efficiently and accurately generate post-disaster building footprints, we combine Microsoft's building footprint data with pre-disaster satellite images and post-disaster satellite images. This approach allows us to create masks for building footprints and compare the pre- and post-disaster states to identify damage and changes in the built environment. The general process of generating post-disaster building footprints can be broken down into several key steps as Figure 3.1.

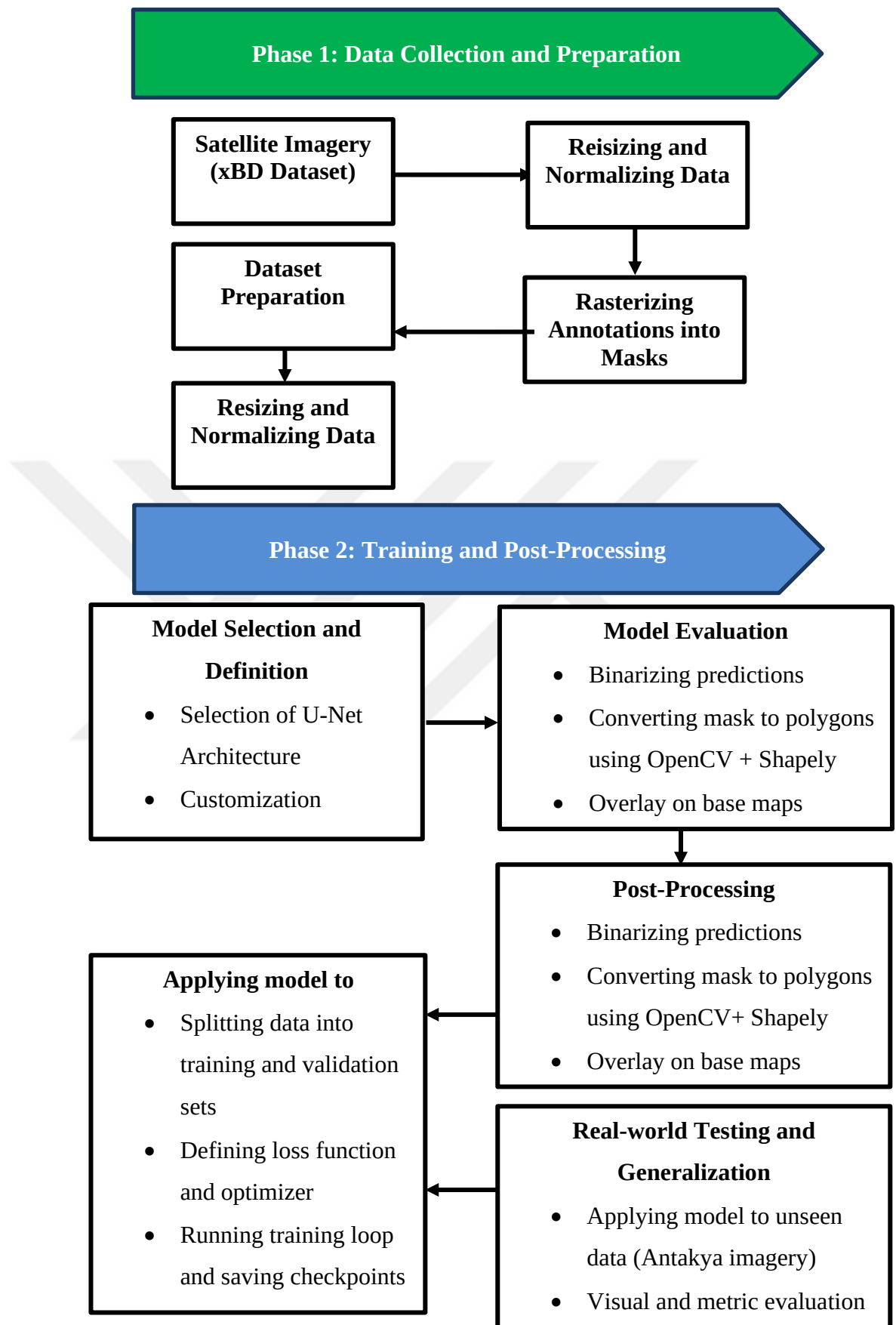


Figure 3.1 : General process of generating building footprint (Generated by Author)

3.5.1 Microsoft building footprint

Microsoft's building footprint dataset provides detailed, high-resolution geographic information about the locations and extents of buildings before the disaster event. These building footprints serve as the foundation for creating a mask in the analysis. The initial dataset comes in the form of shapefiles or GeoJSON files and covers a wide area with numerous buildings identified across different geographic regions.

For this step, the building footprint data from Microsoft is used as a base layer. This data typically represents the boundaries of buildings in the pre-disaster state and is extracted through machine learning techniques applied to satellite imagery. It's important to ensure that the building footprints align with the region of interest for the post-disaster analysis, as discussed in the previous section where we filtered the relevant geographic areas for this study.

In order to assess the impact of a disaster effectively, it is crucial to establish a reference layer that represents the built environment before the event. This reference layer allows for direct comparison with post-disaster conditions, enabling the identification of damaged or completely destroyed buildings. One of the most effective ways to establish this reference is by creating a mask from pre-disaster satellite imagery using building footprint data.

3.5.2 Creating mask from pre-disaster satellite images

A building mask is a binary raster or vector representation of the areas occupied by buildings. These masks are typically generated using vector datasets such as Microsoft Building Footprint data, which outlines the structure of buildings with high accuracy. By overlaying these building footprints on pre-disaster satellite images, we can create a high-quality representation of the built environment before the disaster, ensuring an accurate baseline for damage detection.

Microsoft's Building Footprint dataset provides high-precision, AI-generated building outlines for various regions across the world. These footprints are extracted using deep learning models trained on satellite and aerial imagery, making them highly suitable for disaster management applications. The dataset is freely available and can be

accessed through Microsoft's GitHub repository, where footprints for different countries are provided in vector format (GeoJSON or shapefile).

For this study, we downloaded the building footprint dataset for Turkey, which includes footprints for all major cities, including Antakya city affected by the 6 Feb 2023 Turkey earthquakes. The footprints were downloaded using a Python script named "make-gis-friendly.py," provided by Microsoft, which converts the raw data into a GIS-compatible format. The script was modified to focus on Turkey and extracted the relevant building footprint files.

Once downloaded, the dataset consisted of multiple small files representing different regions. Since Microsoft does not provide metadata indicating which file corresponds to which region, all files had to be loaded into QGIS for visualization. By overlaying the files on a basemap, we identified the footprints corresponding to Antakya and extracted these specific files for further processing.

3.5.2.1 Pre-processing the building footprint

To create this mask, the pre-disaster satellite images of the affected areas are used. These images are typically high-resolution multispectral satellite images, which contain a wealth of information about the land surface. The process of generating a mask involves the following:

I. Coordinate System Adjustment

- The raw building footprint data was provided in WGS 84 (EPSG:4326) coordinate reference system (CRS), while the satellite images used for analysis were in a different projection.
- To align the datasets, the building footprints were re-projected to match the CRS of the satellite imagery (e.g., UTM projection for Turkey).
- This ensured spatial accuracy and avoided distortion during mask generation.

II. Data Cleaning

- Some buildings in the dataset were misclassified or duplicated.
- Small, irrelevant polygons (e.g., noise in the dataset) were filtered out using an area threshold.

- Building polygons were simplified to reduce complexity while preserving the overall shape.

III. Clipping to the Area of Interest (AOI)

- The full dataset covered large areas, but only the regions of Antakya were relevant.
- A bounding box was created around these cities, and the footprints were clipped accordingly.

Once the footprints were pre-processed, they were ready for integration with pre-disaster satellite imagery to generate masks

3.5.2.2 Integrating footprint with pre-disaster satellite imagery

The next step involved overlaying the pre-processed building footprint data onto pre-disaster satellite images. High-resolution satellite images of Antakya before the earthquake were obtained from open-source providers (e.g., Google Earth Pro) To ensure proper alignment between the footprints and the satellite imagery:

- I. The images were georeferenced to match the CRS of the building footprints.
- II. The images were then cropped to the AOI to reduce file size and processing time.

The building footprints were then overlaid on the pre-disaster images in QGIS, allowing us to visually confirm that the footprints accurately corresponded to the building locations.

3.5.2.3 Generating building mask

To convert the vector building footprints into binary masks, the following steps were performed:

- I. Rasterization of Vector Footprints
 - The polygon footprints were converted into a raster format.
 - The output raster had two classes:
 - Buildings (1) Representing areas covered by building footprints.
 - Non-Buildings (0) Representing all other areas.

II. Resolution Matching

- The rasterized masks were resampled to match the resolution of the satellite images.
- This ensured that the mask aligned perfectly with the underlying pre-disaster imagery.

III. Edge Smoothing

- The initial rasterization process often resulted in jagged edges.
- A morphological filtering technique (e.g., Gaussian smoothing) was applied to refine the boundaries.

IV. Final Export

- The final pre-disaster building masks were saved as GeoJson files.

These masks were exported along with the original AOI polygons to maintain reference information.

3.5.2.4 Challenges and limitation

Despite the high accuracy of Microsoft's Building Footprint dataset, some challenges were encountered during mask generation:

- I. Data Gaps: Some buildings were missing from Microsoft's dataset, requiring manual verification.
- II. Spatial Misalignment: In some cases, minor shifts between footprints and satellite images were observed.
- III. Processing Time: Converting large-scale building footprints into raster masks required high computational power.

To address these challenges, manual verification was performed, and advanced spatial correction techniques were applied to improve alignment accuracy.

3.6 Post-Disaster Building Footprint

One of the most critical steps in disaster damage assessment is the identification of buildings that have been damaged or destroyed. To achieve this, we need to generate post-disaster building footprints, which represent the state of the built environment

after the event. These footprints are then compared with pre-disaster building masks to assess structural changes, quantify damage, and map affected areas.

3.6.1 Process of post-disaster building footprint

While pre-disaster footprints can be obtained from existing building datasets like Microsoft Building Footprints, post-disaster footprints must be extracted directly from satellite imagery. This process involves multiple steps, including image pre-processing, segmentation, classification, and footprint extraction using remote sensing and GIS techniques.

To generate post-disaster building footprints:

- I. **Pre-process Post-Disaster Satellite Images:** Similar to the pre-disaster images, the post-disaster satellite imagery is pre-processed. This might involve cleaning up cloud cover or shadows and aligning the post-disaster image with the pre-disaster image using techniques like image registration, ensuring both images are aligned geographically.
- II. **Building Detection:** Using the post-disaster satellite image and the pre-existing mask from the previous step, we apply object detection (Deepness plugin) to detect buildings in the post-disaster scenario. Since some buildings may have been damaged or destroyed, the building detection algorithm must be able to account for missing structures or collapsed buildings.
- III. **Creating Post-Disaster Building Footprint Mask:** After detecting the buildings, a post-disaster mask is generated that outlines the location and extent of the buildings still standing. This mask will typically be a binary image, where areas with buildings are marked as "1" (indicating presence) and areas without buildings as "0" (indicating absence). The new post-disaster mask is then compared with the pre-disaster mask to identify changes in the built environment, such as buildings that were damaged, destroyed, or remain intact.

3.6.2 Data Acquisition

To extract post-disaster building footprints, high-resolution satellite imagery of the affected area was obtained. The selection of satellite images was based on the following criteria:

- I. Temporal Relevance: The images had to be captured after the disaster to accurately reflect structural damage.
- II. High Spatial Resolution: A minimum resolution of 1 meter per pixel was required to distinguish individual buildings.
- III. Spectral Bands: multi-spectral and panchromatic bands were preferred for improved object detection.

Data sources included:

- I. Google Earth Pro: High-resolution imagery

3.6.3 Post processing of extracted footprint:

After extracting post-disaster footprints, additional post-processing steps were applied to enhance accuracy.

- I. Noise Removal: Small misclassified objects were filtered out using a minimum area threshold.
- II. Spatial Alignment with Pre-Disaster Data: The extracted footprints were aligned with pre-disaster building masks to ensure spatial consistency.
- III. Conversion to GIS Format: The final footprints were converted into shapefiles (SHP) and GeoJSON for further analysis

3.6.4 Application of post-disaster building footprint:

The extracted post-disaster building footprints served multiple purposes in disaster analysis:

- I. Damage Assessment: By comparing pre- and post-disaster footprints, the number of collapsed or missing buildings was identified.
- II. Disaster Response and Recovery Planning: The extracted footprints were integrated into GIS-based decision support systems to guide emergency responders.
- III. Automated Change Detection: Machine learning algorithms were applied to compare footprint datasets.
- IV. Urban Resilience Studies: The post-disaster dataset was used to assess urban vulnerability and resilience in the affected areas

3.7 Challenges and limitations

Despite the successful extraction of post-disaster footprints, several challenges were encountered:

- I. Obstructed View in Satellite Imagery: Rubble and debris sometimes made it difficult to distinguish collapsed buildings.
- II. Misclassification in AI-Based Segmentation: Some buildings were partially detected, requiring manual correction.
- III. Data Availability Issues: High-resolution commercial satellite imagery was costly, limiting frequent updates.

Generating post-disaster building footprints was a crucial step in disaster assessment. Using high-resolution satellite imagery, deep learning models, and GIS tools, footprints were accurately extracted and refined. The final footprints were then compared with pre-disaster masks to quantify building damage.

3.8 Comparison and Analysis

In The final step in the process is to compare the pre- and post-disaster building footprints and analyze the differences between the two. The following steps are involved in this comparison:

- I. Overlaying Pre- and Post-Disaster Masks: By overlaying the pre- and post-disaster masks, we can visually assess which buildings have been affected by the disaster. This step can reveal damaged or destroyed buildings and can also highlight new construction if applicable. The comparison can also help identify areas of significant structural damage that may require immediate attention during recovery efforts.
- II. Change Detection: Change detection algorithms can be applied to quantify the differences between the pre- and post-disaster building footprints. This could involve calculating the percentage of buildings that have been damaged, the extent of building destruction, or identifying regions with the most significant changes. The results of this analysis can guide decisions on where to allocate resources for disaster recovery.

III. Damage Assessment: Based on the comparison of the building footprints, a damage assessment can be conducted. For example, buildings that no longer appear in the post-disaster mask or have significant changes in shape or size can be considered destroyed. This information can then be used to support decisions related to rebuilding efforts, disaster relief, and resource distribution.

IV. Mapping and Visualization: To communicate the results of the analysis, the pre- and post-disaster building footprints, along with the damage assessment, can be visualized in a GIS environment. Tools such as QGIS or ArcGIS can be used to create maps that show the locations of damaged and undamaged buildings, helping decision-makers understand the scope of the disaster and plan for recovery.

By combining Microsoft's building footprint data with pre- and post-disaster satellite images, we can generate post-disaster building footprints. The use of masks for both pre- and post-disaster buildings allows for a comparison, which is essential for damage assessment and recovery planning. The methodology described in this section provides a robust approach to understanding the impacts of disasters on the built environment and can serve as a valuable tool for disaster response efforts.



4. RESULT AND DISCUSSION

This chapter presents the outcomes of the building footprint detection process using deep learning techniques applied to high-resolution satellite imagery. It outlines how the model was trained, how its performance was evaluated, and how the results were interpreted both numerically and spatially. The process began with the training of a U-Net model on labeled satellite images and extended to generating predictions that could be visualized and analyzed in geographic formats. The chapter is structured to follow the sequence in which the practical work was conducted, starting from model training and ending with spatial analysis and reflections on model behavior.

Automatic extraction of building footprints from satellite images has become increasingly important in areas affected by natural disasters. Accurate identification of structures can support post-disaster needs assessments, help prioritize response efforts, and assist in reconstruction planning. Traditional methods, which often require manual digitization, are time-consuming and not feasible at scale. With the advancement of convolutional neural networks and greater access to satellite imagery, automated solutions have become more attainable. This work builds on that opportunity by applying a segmentation model that is capable of learning from labeled data and producing pixel-wise masks indicating the presence of buildings.

The model used in this study follows the U-Net architecture, a type of encoder–decoder neural network known for its ability to perform well even on relatively small datasets. Training was conducted using building masks created from vector labels originally formatted in GeoJSON. These labels were converted to binary images that aligned with the corresponding satellite tiles. Each input image was paired with a mask where building areas were labeled as 1 and the background as 0. Images were standardized to a consistent input size for the model, and care was taken to ensure that labels were accurate and clean.

Once training began, the model's performance was monitored across several epochs. The Binary Cross Entropy loss function was used to quantify prediction error, and metrics such as Intersection over Union (IoU) and F1-score were calculated to provide

additional insights into the model's ability to segment buildings correctly. In addition to numerical results, several visual outputs were created to better understand where the model succeeded and where it fell short. These visualizations include predicted masks, overlays on the original imagery, and polygons extracted from the predicted masks.

A significant part of the analysis involved transforming the predicted masks into geospatial formats. Using polygon extraction methods, the model's output was converted into vector geometries and saved as GeoJSON files. These files can be opened in GIS software and analyzed further. This allowed a more practical use of the model's results, as they could be compared to official building datasets or used directly in post-disaster mapping workflows.

Another layer of interpretation was added through spatial analysis. By calculating the centroids of predicted buildings and plotting their density, heatmaps were generated to show where buildings were most concentrated in each satellite tile. These maps offered a broader understanding of settlement patterns and, in some cases, helped highlight regions with high structural density that may be vulnerable in future disasters. The analysis was supported by graphs showing the number of buildings detected per tile, as well as summaries of building areas.

While the results were generally promising, some challenges were encountered during the process. The model occasionally misclassified small objects or shadows as buildings, particularly in areas with complex textures or overlapping structures. This led to false positives that were addressed in post-processing through size and shape filtering. Other limitations included the resolution of the imagery, inconsistencies in ground truth labels, and computational constraints.

Despite these challenges, the model demonstrated its ability to generalize well across a variety of images and produce reliable outputs without needing extensive manual intervention. Its performance in both training and testing environments suggests that deep learning offers a viable method for scalable building footprint detection in disaster-prone areas.

This chapter, therefore, serves as both a summary of model behavior and a point of reflection. By presenting the training details, evaluation results, and spatial outcomes in an integrated way, it aims to offer a comprehensive view of how such a model performs in a realistic application scenario. The following sections provide more

detailed information on each stage of the workflow, beginning with the training phase and moving toward evaluation and discussion of the results in practical terms.

4.1 Phase 1: Model Training Process

Phase This section describes the process of training the deep learning model that was developed to automatically identify building footprints from high-resolution satellite images. The selected model is based on the U-Net architecture and was trained on labeled data derived from satellite imagery. The workflow includes preparing the training dataset, designing the model structure, defining the loss function and optimization strategy, and monitoring performance during training. All steps were implemented using Python and the PyTorch deep learning framework.

disasters.

4.1.1 Dataset preparation and preprocessing

Accurate training of a semantic segmentation model for building footprint detection requires a well-structured and carefully prepared dataset. For this study, the dataset was based on Tier 3 of the publicly available xBD dataset, which includes high-resolution pre-disaster and post-disaster satellite imagery, along with detailed annotations for building locations and, in some cases, damage classification. Although the xBD dataset contains a wide variety of scenes across multiple countries and disaster types, this study focused on extracting building footprint data, excluding damage classes, to support a binary segmentation task. The emphasis was on creating a clean, consistent, and balanced dataset that would enable the U-Net model to learn effectively from supervised input-output pairs.

4.1.1.1 Source and structure of the dataset

The xBD dataset was curated by the Defense Innovation Unit (DIU) and Carnegie Mellon University as part of the xView2 challenge. It contains imagery captured before and after various natural disasters, including earthquakes, floods, fires, and hurricanes. Each scene in the dataset consists of satellite images in PNG format, typically sized 1024×1024 pixels, and a corresponding label in JSON format containing building annotations in vector (polygon) form. These annotations are provided under the features key, with individual entries marked by a feature_type

attribute. For this project, only features labeled as "building" were retained. Annotations indicating damage categories or unrelated objects were excluded.

In total, the Tier 3 subset includes 5,598 images with consistent labeling quality. To avoid data leakage between disaster types and maintain geographic diversity, a wide range of locations was included, though care was taken to ensure that the validation set would contain scenes not seen during training.

4.1.1.2 Parsing and conversion of annotations

Since the U-Net model operates on pixel-level input and output, the first step was to convert the vector polygon labels into rasterized binary masks. This was achieved using Python scripts based on the Shapely and OpenCV libraries. For each JSON file, all objects with "feature_type": "building" were extracted, and their geometry was parsed using Shapely's `wkt.loads()` function. These geometries were then rasterized into binary masks using `cv2.fillPoly()` on blank NumPy arrays of size 1024×1024. Inside the filled polygon, the pixel value was set to 1, and outside to 0.

This process required careful handling to avoid common issues such as self-intersecting polygons, invalid geometries, or label misalignment. To prevent crashes during conversion, all geometries were validated using Shapely's `.is_valid` property, and small or malformed polygons were ignored. The resulting masks were saved in PNG format using `cv2.imwrite()`, with white (255) representing building pixels and black (0) representing the background.

4.1.1.3 Image normalization and channel conversion

To ensure compatibility with the model and to promote faster convergence during training, all input images were normalized to the range [0, 1] by dividing each pixel value by 255. This step scales RGB values from the default 0–255 range to a float32 representation, which is more suitable for neural networks. In addition, all images were converted from OpenCV's default BGR color ordering to the standard RGB format using `cv2.cvtColor()`. This color format consistency is important because most deep learning models, including U-Net, expect inputs in RGB.

4.1.1.4 Image and mask resizing

Given the memory limitations of consumer-grade GPUs and the desire for faster training, all images and masks were resized from their original resolution (1024×1024)

to a more manageable size of 256×256 pixels. This resizing was done using bilinear interpolation for the images and nearest-neighbor interpolation for the masks. The use of 256×256 resolution balances two competing needs: retaining enough spatial detail to accurately detect small buildings and reducing computational load to allow for larger batch sizes and faster model iterations.

Although resizing may lead to the loss of fine-grained details, it proved sufficient for the binary segmentation task in this study. In cases where extremely small buildings were present, the post-processing pipeline included filters to remove objects below a minimum area threshold, reducing the effect of such losses.

4.1.1.5 Mask thresholding and binarization

To further enforce consistency across all masks, a thresholding operation was applied to each mask using a pixel value cutoff of 127. All pixel values above this threshold were set to 1 (building), and all others were set to 0 (background). This binarization step ensured that any anti-aliasing effects introduced during polygon rasterization or resizing would not cause partial pixel values in the training masks. The final masks used for training were stored as single-channel 8-bit grayscale images.

4.1.1.6 Dataset splitting strategy

The full dataset was randomly divided into training and validation subsets using an 80:20 ratio. This division ensures that the model has sufficient data to learn from while retaining a portion of the dataset for performance monitoring on unseen data. The split was done using `sklearn.model_selection.train_test_split()` with a fixed random seed for reproducibility. Since the xBD dataset contains images from multiple disaster scenes, care was taken to maintain geographic diversity in both the training and validation sets. This helped prevent overfitting to specific textures, lighting conditions, or building styles.

During training, only the training set was used to update the model weights. The validation set was used exclusively for evaluating the model's performance after each epoch, without affecting gradient updates. This approach provided a reliable measure of the model's generalization ability and allowed for early detection of overfitting.

4.1.1.7 PyTorch dataset class implementation

To handle real-time data loading and preprocessing, a custom PyTorch Dataset class named BuildingDataset was implemented. This class takes lists of image file names and corresponding mask file names, along with directory paths, and returns formatted PyTorch tensors ready for input into the model. Within the `__getitem__()` method, each image is read using OpenCV, resized, normalized, and converted to a PyTorch tensor with shape `[3, 256, 256]`. Likewise, the mask is resized, thresholded, and reshaped into a single-channel tensor of shape `[1, 256, 256]`.

This custom dataset class ensures that each batch during training is correctly formatted, aligned, and ready for model input. It also allows for optional future integration of data augmentation pipelines (e.g., random rotations, flips, or noise) using libraries like Albumentations.

4.1.2 Model architecture

The segmentation model developed in this study is based on the U-Net architecture, a well-established convolutional neural network originally proposed for biomedical image segmentation. Due to its structural simplicity, robust performance in various pixel-level classification tasks, and ability to generate accurate masks even from limited training data, U-Net has become a preferred choice for satellite image segmentation tasks, particularly those involving building footprint detection. In this project, a simplified yet effective implementation of U-Net was developed using PyTorch, tailored to process 256×256-pixel satellite image tiles and output binary masks identifying building footprints.

The architecture of the model used in this study is illustrated in Figure 4.1, which outlines the full pipeline from input image tile to output segmentation mask. Each component of the architecture is described below, along with design decisions and considerations relevant to the segmentation task at hand.

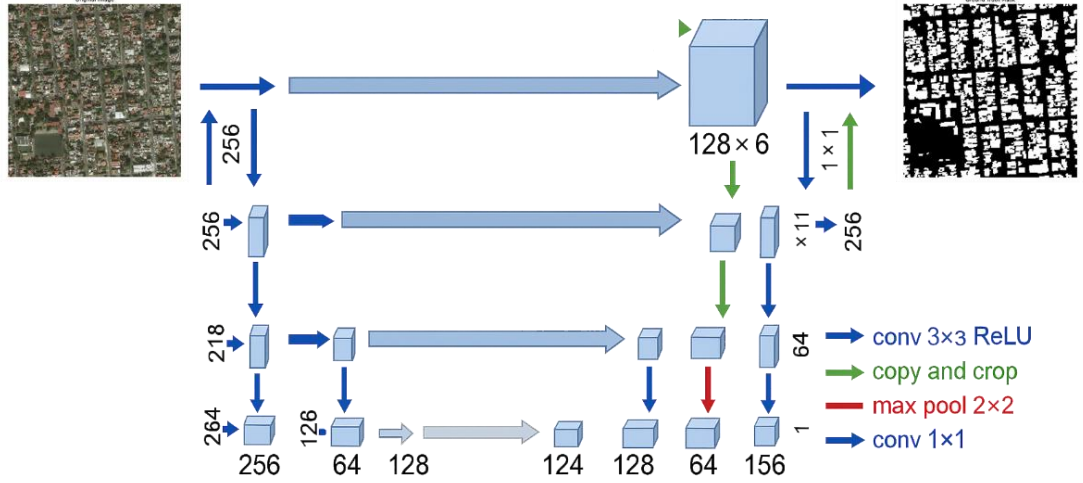


Figure 4.1 : The U-Net architecture of the model used in this study.

4.1.2.1 Encoder–Decoder structure

U-Net is characterized by its symmetrical encoder–decoder structure. The left half of the architecture (encoder) is responsible for extracting hierarchical features from the input image, while the right half (decoder) reconstructs a spatially precise segmentation mask. The use of skip connections between corresponding encoder and decoder layers enables the model to combine low-level spatial features with high-level semantic understanding, which is particularly useful when delineating the sharp edges and complex shapes of buildings.

4.1.2.2 Input layer and preprocessing assumptions

Each image tile entering the model has a shape of $[3 \times 256 \times 256]$, representing the RGB color channels of the satellite image. Prior to being fed into the network, these images are normalized to values between 0 and 1 and arranged in the channel-first format expected by PyTorch.

The input layer does not perform any transformation; instead, it directly connects to the first convolutional block, where feature extraction begins.

4.1.2.3 Contracting path (Encoder)

The encoder consists of three downsampling stages, each of which comprises two 3x3 convolutional layers followed by a 2x2 max pooling operation. These blocks progressively reduce the spatial resolution while increasing the number of feature channels, allowing the network to learn increasingly abstract representations of the image content.

I. First encoder block (Down 1):

The first set of convolutional layers transforms the input image into 64 feature maps. Since this is the initial stage, it captures basic features such as edges, corners, and color gradients. The output shape after this block is $[64 \times 256 \times 256]$.

II. Second encoder block (Down 2):

After max pooling, the resolution is halved to $[64 \times 128 \times 128]$. Another set of convolutional layers increases the channel depth to 128, enabling the model to recognize more complex textures and spatial patterns.

III. Third encoder block (Down 3):

A second max pooling operation further reduces the spatial dimensions to $[128 \times 64 \times 64]$. The output is then processed through two convolutional layers that yield $[256 \times 64 \times 64]$, capturing high-level semantic information.

These contracting blocks allow the network to encode contextual information, which is critical for identifying structures like buildings across varied backgrounds.

4.1.2.4 Bottleneck layer

At the bottom of the U shape lies the bottleneck, which serves as a transition between the encoder and decoder. The bottleneck applies two convolutional layers with 3×3 kernels, maintaining the same spatial size while preserving semantic features and aggregating information from the deepest level of the encoder. The output at this stage remains $[256 \times 64 \times 64]$.

4.1.2.5 Expanding path (Decoder)

The decoder reverses the operations performed in the encoder. It gradually reconstructs the full spatial resolution of the input while integrating contextual information extracted by the encoder. This is achieved through a sequence of upsampling operations and concatenation with the corresponding feature maps from the encoder.

I. First decoder block (Up 2):

The feature map is upsampled using a transposed convolution with a stride of 2, increasing the spatial dimensions to $[128 \times 128 \times 128]$. This is concatenated with the encoder output from the second block (skip connection), resulting in a $[256$

$\times 128 \times 128$] tensor. This is followed by two convolutional layers that reduce it back to $[128 \times 128 \times 128]$.

II. Second decoder block (Up 1):

Another transposed convolution upsamples the feature map to $[64 \times 256 \times 256]$. After concatenation with the output from the first encoder block, and further convolutional processing, the resulting feature map is reduced to $[64 \times 256 \times 256]$.

This upsampling strategy ensures that fine-grained details, such as building outlines and corners, are preserved in the final segmentation output.

4.1.2.6 Final convolution and activation

The final layer is a single 1×1 convolutional layer, which compresses the 64-channel feature map into a single-channel probability map of size $[1 \times 256 \times 256]$. This layer is followed by a sigmoid activation function to convert raw output values into probabilities between 0 and 1. A threshold of 0.5 is applied during inference to produce a binary mask, where each pixel is classified as either part of a building (1) or background (0).

4.1.2.7 Skip Connections and their importance

A defining characteristic of the U-Net architecture is its use of skip connections, shown as horizontal arrows in Figure. These connections bridge encoder and decoder layers at the same level of resolution, allowing the decoder to access spatial information that may have been lost during downsampling. This helps the model retain fine detail and makes it particularly effective for tasks like building segmentation, where object boundaries are critical.

The skip connections are implemented through simple tensor concatenation, increasing the number of feature channels in the decoder stages. This architectural choice significantly improves the accuracy of edge prediction and small object detection.

4.1.2.8 Visualization reference

Figure 4.1 illustrates the full architecture of the model used in this study. It highlights the sequence of operations and transformations applied to the input image tile. The blue boxes represent multi-channel feature maps, with their respective dimensions

indicated. Convolutional operations are marked with blue arrows, pooling operations with red, and skip connections with green. The figure shows how the input image, after passing through several encoding and decoding blocks, is ultimately converted into a segmentation mask of the same spatial resolution.

This diagram provides a clear visual summary of the U-Net structure and complements the written explanation of the architecture.

4.1.2.9 Implementation considerations

The model was implemented using the PyTorch framework. Each convolutional block was created using `nn.Sequential` and included two `Conv2d` layers with kernel size 3, padding 1, and `ReLU` activation. The max pooling and transposed convolution operations were implemented using `MaxPool2d` and `ConvTranspose2d` respectively. The final layer used `Conv2d` with kernel size 1 and no activation, followed by a `torch.sigmoid()` function.

This architecture was chosen for its balance of simplicity and effectiveness. While more advanced architectures exist, such as U-Net++ or DeepLabV3+, the chosen model achieved strong results with relatively low computational cost. This makes it suitable for deployment in resource-constrained environments, such as post-disaster response scenarios where quick inference is essential.

4.1.2.10 Flexibility and expandability

Although the current implementation is designed for binary segmentation, the architecture can be easily extended to support multi-class segmentation by modifying the output layer to produce multiple channels and replacing the loss function with categorical cross-entropy. This makes the model adaptable for future tasks, such as damage level classification or segmentation of other urban features like roads or vegetation.

Overall, the architecture used in this study is a well-structured, lightweight U-Net model tailored for binary building footprint segmentation from satellite imagery. With three downsampling and two upsampling blocks, skip connections, and a final sigmoid-based output layer, the model is capable of learning and reconstructing accurate segmentation masks. The use of transposed convolutions and skip connections allows it to maintain spatial precision, which is essential for building detection. As illustrated in Figure, the model architecture follows a clear and logical

pattern that enables effective feature extraction and spatial reconstruction, making it a strong fit for the problem domain.

4.1.3 Training configuration

The configuration of the training environment plays a crucial role in determining the performance and convergence behavior of any deep learning model. Even with a well-designed architecture and a well-prepared dataset, the outcome of the training process can vary significantly depending on the selected hyperparameters, computational resources, and optimization strategies. In this section, the training configuration used for developing the U-Net segmentation model is described in detail.

The goal was to identify a training setup that is both computationally efficient and capable of yielding accurate segmentation results. Given the binary nature of the task classifying each pixel as building or non-building the choice of loss function, optimizer, learning rate, and batch size had to be carefully balanced. Additionally, aspects such as model saving, reproducibility, and hardware setup were also taken into consideration to ensure that the results could be replicated and extended in future work.

What follows is a comprehensive breakdown of the training loop, the hardware and software used, and the rationale behind the key hyperparameter selections that guided the model toward effective learning.

4.1.3.1 Hardware and environment setup

The training process for the U-Net segmentation model was carried out on a standard workstation equipped with an Intel Core i7 processor and an NVIDIA RTX 4050 GPU.

The model was implemented using the PyTorch deep learning framework (version 2.0.1), chosen for its flexibility, modularity, and active community support. PyTorch provides a dynamic computation graph, which simplifies debugging and makes it easier to implement custom layers and data loading routines. All data preprocessing, augmentation, and tensor transformations were performed using supporting libraries such as OpenCV, NumPy, and scikit-learn.

The experiments were conducted on a local development environment running Windows 11, and all scripts were written in Python 3.10. CUDA drivers and cuDNN were installed to leverage GPU acceleration, and the relevant versions of PyTorch and torchvision were configured to match the CUDA runtime installed on the system.

To ensure reproducibility and ease of future experimentation, the entire environment was encapsulated using a Python virtual environment. All dependencies and library versions were documented and saved in a requirements.txt file, allowing for the same environment to be recreated later if needed. Moreover, random seeds were fixed for PyTorch, NumPy, and Python's internal random module to ensure consistent results across multiple training runs.

The overall setup provided a stable, reliable, and reproducible platform for training, testing, and evaluating the deep learning model. While the training process could have benefited from access to more powerful hardware or cloud-based GPU services, the selected environment was more than sufficient for training the U-Net architecture on 256×256 image tiles within a reasonable time frame.

4.1.3.2 Hyperparameter settings

Selecting appropriate hyperparameters is a critical step in training any deep learning model. These parameters govern the learning behavior of the model and directly affect the convergence speed, stability, and generalization performance. For this study, a set of commonly used and empirically validated hyperparameters was selected, and adjustments were made based on initial experimental runs. The main hyperparameters configured for training the U-Net model are described below.

4.1.3.2.1 Batch size

A batch size of 8 was selected for training. This value represents a trade-off between memory efficiency and gradient estimation stability. Larger batch sizes could provide smoother gradients but were limited by GPU memory constraints. A batch size of 8 allowed for consistent training across the available GPU hardware (RTX 4050 with 6 GB VRAM) while maintaining sufficient stochasticity for robust learning.

4.1.3.2.2 Learning rate

The learning rate was set to 0.001 (1e-3), which is a commonly used default for training models with the Adam optimizer. This value was found to be stable during the early experiments and allowed the model to converge without oscillations or divergence. The learning rate was kept constant across all epochs, as the relatively shallow U-Net architecture did not exhibit signs of overfitting that would require learning rate decay.

4.1.3.2.3 Number of epochs

The model was trained for 10 epochs, which proved sufficient for initial convergence. Given the relatively simple nature of the dataset and the binary segmentation task, longer training was not necessary in this first phase. However, the number of epochs can be easily increased in future experiments if fine-tuning or transfer learning is applied to other datasets.

4.1.3.2.4 Optimizer

The Adam optimizer was used for training, which is well-suited for computer vision tasks due to its ability to adjust learning rates adaptively for each parameter. Adam combines the benefits of both RMSProp and SGD with momentum and has shown consistent performance in segmentation applications. Default values were used for the optimizer's internal parameters:

- I. $\beta_1 = 0.9$
- II. $\beta_2 = 0.999$
- III. $\varepsilon = 1e-8$

These values ensure fast convergence without the need for extensive manual tuning.

4.1.3.2.5 Loss function

Given that the task was binary segmentation (building vs. background), the Binary Cross-Entropy (BCE) loss function was chosen. This loss function measures the pixel-wise difference between the predicted probabilities (after sigmoid activation) and the ground truth binary labels. It is defined as equation (4.1).

$$\text{BCE}(y, y^{\wedge}) = -[y \cdot \log(y^{\wedge}) + (1 - y) \cdot \log(1 - y^{\wedge})] \quad (4.1)$$

where y is the ground truth binary mask and $y^{\wedge}$ is the predicted probability for each pixel. The BCE loss encourages the model to minimize the average difference between its predicted masks and the true masks across all pixels.

4.1.3.2.6 Activation functions

All convolutional layers in the model, except the final one, were followed by ReLU (Rectified Linear Unit) activation functions. ReLU introduces non-linearity and helps prevent vanishing gradients during backpropagation. The final layer used a sigmoid activation to map output values to the range $[0, 1]$, allowing interpretation as probabilities.

4.1.3.2.7 Evaluation threshold

During inference, the output of the sigmoid function was thresholded at 0.5. Pixels with values greater than or equal to 0.5 were classified as building (1), and those below as background (0). This threshold value is standard for binary segmentation and can be adjusted in future experiments to control the balance between precision and recall.

4.1.3.3 Learning rate strategy

During The learning rate is one of the most influential hyperparameters in deep learning. It controls the step size at which the optimizer updates the model weights during backpropagation. Choosing an appropriate learning rate is essential to ensure stable convergence and to prevent issues such as vanishing gradients, slow learning, or overshooting the optimal solution.

For the training of the U-Net model in this study, a fixed learning rate strategy was employed. The learning rate was set to 0.001, which is a standard and widely used default for the Adam optimizer in segmentation tasks. This value was selected based on empirical evidence from prior work and validated through preliminary experiments.

During training, the learning rate remained constant across all epochs. This decision was motivated by the relatively shallow depth of the model, the stable behavior observed in the loss function over time, and the absence of significant overfitting or stagnation in performance. As a result, there was no immediate need to implement a learning rate scheduler or decay mechanism.

While more advanced learning rate scheduling techniques such as step decay, exponential decay, or cosine annealing can be beneficial for deeper or more complex models, the fixed learning rate proved sufficient in this context. However, for future extensions of the project or when applying the model to new datasets with different characteristics, learning rate tuning or adaptive scheduling may become necessary to improve convergence speed and final accuracy.

It is worth noting that Adam's adaptive learning mechanism internally adjusts the effective learning rate for each parameter based on first- and second-order moments. This helps maintain consistent updates across different layers, even in the absence of external scheduling. As a result, the model was able to converge smoothly without requiring manual adjustments to the learning rate during training. a fixed learning rate of 0.001 was used throughout the training process. This approach simplified the

implementation and avoided the complexity of dynamic scheduling, while still yielding stable and reliable convergence.

4.1.3.4 Weight initialization and model saving

The training process of any deep learning model relies not only on the choice of architecture and data quality, but also on how the model parameters are initialized and preserved. These two factors weight initialization and model saving—play a fundamental role in achieving stable convergence and enabling reproducibility.

For this project, the model parameters were initialized using the default initialization strategies provided by the PyTorch framework. Specifically, PyTorch applies Kaiming uniform initialization (also known as He initialization) to convolutional layers, which is well-suited for models that use ReLU activation functions. This method maintains a consistent variance in the output across layers, thereby mitigating the risk of vanishing or exploding gradients during training.

No custom initialization scheme was applied. The default initialization was deemed sufficient for the U-Net architecture used in this study, which includes a relatively shallow depth and a standard sequence of convolutional and upsampling layers. The consistent reduction in training loss over the course of epochs indicated that the initial weights provided a suitable foundation for optimization without the need for manual tuning or advanced initialization techniques.

To support future evaluation, deployment, and reproducibility, the trained model was saved at the end of the training process using PyTorch's built-in serialization function. The model's state dictionary, which contains all learnable parameters, was written to disk using the following command:

```
torch.save(model.state_dict(), "building_unet.pth")
```

This approach ensures that the exact weights learned during training can be restored at any later time, provided that the same model architecture is reconstructed. When needed, the model can be reloaded using:

```
model.load_state_dict(torch.load("building_unet.pth"))
```

The model file was subsequently used for inference tasks such as generating building masks, polygon extraction, and exporting GeoJSON outputs. Intermediate checkpointing was not implemented in this project, as the training was limited to 10

epochs and no signs of unstable learning were observed. However, for more complex models or longer training durations, it is advisable to store periodic checkpoints to avoid loss of progress in the event of interruption.

Together, these strategies ensured that the training process was both stable and repeatable. The combination of reliable initialization and consistent model saving laid the foundation for downstream analysis and future extensions of this work.

4.1.3.5 Training loop structure

Together, The core of the model training process lies in the structure and execution of the training loop. In this project, a standard supervised learning loop was implemented using the PyTorch framework. The training loop was designed to iterate over the dataset in mini-batches, compute the loss, backpropagate the gradients, and update the model parameters accordingly. Each iteration of the loop represents a forward and backward pass through the network for a batch of input–target pairs, enabling the model to learn from pixel-wise differences between predicted masks and ground truth masks.

The training was conducted over 10 epochs, where each epoch represents a complete pass through the entire training dataset. During each epoch, the model was set to training mode using `model.train()`, which enables certain layers like dropout or batch normalization (if present) to behave appropriately during learning. For this specific U-Net implementation, which does not include these layers, the effect of switching modes was limited but maintained for compatibility.

The process for each batch in the training loop was as follows:

I. Loading Data:

A batch of satellite images and their corresponding binary masks were loaded using a PyTorch `DataLoader`. The custom `BuildingDataset` class handled preprocessing, resizing, normalization, and formatting of each sample.

II. Moving Data to Device:

Input images and masks were moved to the appropriate device (CPU or GPU) using `.to(device)` to ensure compatibility with model parameters.

III. Forward Pass:

The input batch was passed through the model to obtain predicted output masks.

These predictions were continuous values between 0 and 1 due to the use of sigmoid activation in the final layer.

IV. Loss Computation:

The predicted outputs were compared with the ground truth masks using the Binary Cross-Entropy (BCE) loss function, which computes the average loss across all pixels in the batch.

V. Backward Pass and Optimization:

The optimizer gradients were cleared using `optimizer.zero_grad()`, followed by a backward pass (`loss.backward()`) to compute gradients for all learnable parameters. Finally, the optimizer (Adam) updated the model weights based on these gradients via `optimizer.step()`.

VI. Loss Tracking:

The loss value for each batch was recorded and accumulated across the epoch. At the end of each epoch, the average training loss was computed and printed, providing a high-level view of convergence over time.

Together, The printed log after each epoch helped monitor training progress and identify any anomalies such as stagnant loss or divergence. Because the loss decreased consistently over the 10 epochs, no early stopping or learning rate adjustment was needed during training.

At the end of training, the model was switched to evaluation mode (`model.eval()`) for validation and inference purposes. This structure ensured that the model was trained efficiently, and its performance could be monitored continuously through simple logging mechanisms.

Overall, the training loop was designed to be modular, easy to extend, and efficient. It can accommodate additional enhancements such as validation loss tracking, intermediate model checkpointing, data augmentation, or the integration of early stopping mechanisms in future work.

4.1.4 Observed training behavior

The training process of the U-Net model was monitored over the course of ten epochs to evaluate its learning behavior. Throughout the training, the Binary Cross-Entropy

(BCE) loss was used as the primary loss function, and performance was assessed based on both the training and validation datasets.

From the first to the final epoch, the training loss decreased consistently from 0.1867 to 0.0959, showing that the model was progressively learning to reduce the error between its predictions and the ground truth masks. The validation loss followed a similar trend, decreasing from 0.1984 to 0.1090. The closeness of the training and validation loss curves indicates stable learning and minimal overfitting during the training phase as shown in Figure 4.2.

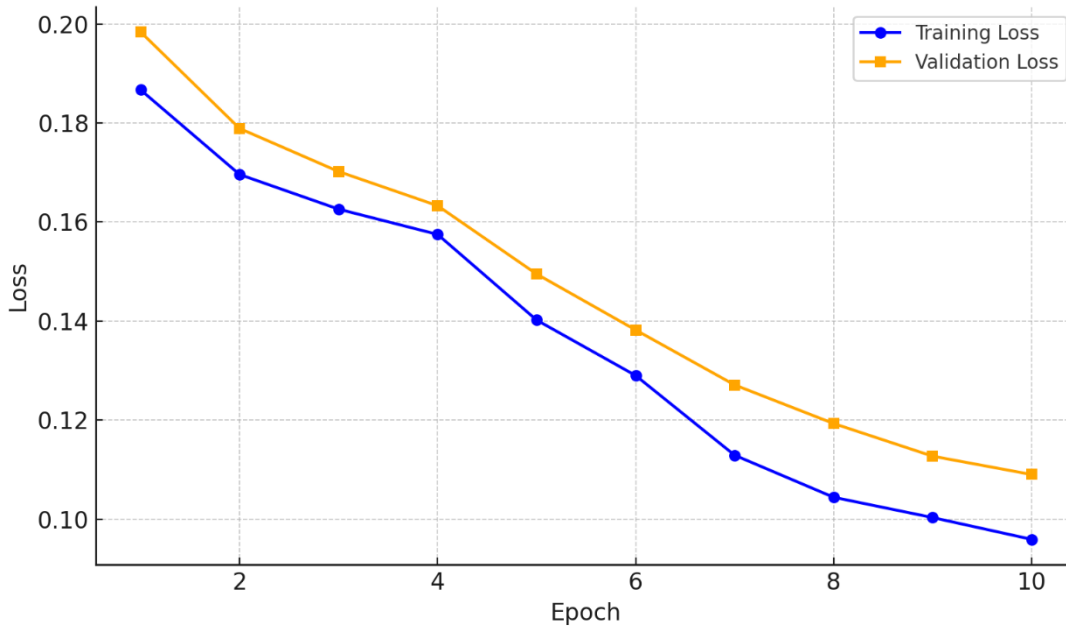


Figure 4.2 : Training and validation loss over Epochs

Figure 4.2 illustrates the training and validation loss over the epochs. The smooth decline in both curves demonstrates effective convergence, confirming that the model was able to learn relevant features from the input data. No major fluctuations or divergence were observed during training, suggesting that the selected learning rate, optimizer, and batch size were appropriate for the task.

In addition to monitoring loss, the model's predictive performance was evaluated using the F1-score and Intersection over Union (IoU) on the validation set. These metrics are widely used in binary image segmentation tasks to quantify overlap and accuracy.

The F1-score is defined as equation (4.2).

$$F1 - score = \frac{Precision \cdot Recall \cdot 2}{Precision + Recall} \quad (4.2)$$

The Intersection over Union (IoU) is defined as equation (4.3).

$$\text{IoU} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive} + \text{False Negative}} \quad (4.3)$$

Over ten epochs, the F1-score improved from approximately 0.50 to 0.6790, and the corresponding IoU reached 0.514. These values confirm that the model was not only minimizing error but also producing meaningful predictions with a strong overlap between predicted and actual building regions.

In summary, the observed training behavior indicates that the model successfully learned to segment building footprints in satellite images. The consistent improvement across loss and evaluation metrics reflects the effectiveness of the chosen model architecture and training configuration. The next section provides a more detailed analysis of the model's performance on the validation set, including quantitative scores and visual comparisons.

4.2 Evaluation on Validation Set

Evaluating a deep learning model on unseen data is a critical step in measuring its generalization capability. While training performance indicates how well the model has learned patterns from the training set, only validation performance can reveal whether it can make accurate predictions on new data. This is particularly important in semantic segmentation tasks using satellite imagery, where input variability is high and the risk of overfitting is substantial.

In this study, the U-Net model trained for building footprint detection was evaluated on a dedicated validation set comprising 20% of the total data. This set included satellite image tiles not used during training, providing a fair test of the model's predictive capability. The evaluation relied on both quantitative metrics and qualitative inspection, offering a comprehensive perspective on model behavior.

Two primary metrics were used for quantitative assessment: the F1-score and the Intersection over Union (IoU). These are standard for binary segmentation tasks and capture different aspects of prediction quality. The F1-score balances precision and recall, while IoU directly measures the spatial overlap between predicted and ground truth building masks. On the validation set, the model achieved a final F1-score of 0.6790 and an estimated IoU of 0.514, suggesting good agreement between predicted and actual masks.

To complement numerical metrics, qualitative analysis was conducted by visually comparing the predicted binary masks and building polygons to their corresponding ground truth annotations. A selection of sample validation images is presented in the following subsections, each showing the original satellite image, the true building mask, and the predicted output. These examples demonstrate where the model succeeds (e.g., clearly shaped and isolated buildings) and where it struggles (e.g., closely packed structures or occluded areas).

The predicted masks were also converted into polygonal representations to assess geometric consistency and readiness for integration with GIS platforms. In most cases, the predicted outlines closely matched actual building shapes, with only minor discrepancies along edges or small-scale structures. This consistency reinforces the usefulness of the model outputs in real-world scenarios such as post-disaster damage mapping.

Overall, this section aims to evaluate the model not only in terms of raw performance numbers but also in its capacity to produce usable, spatially accurate outputs. The combination of F1-score, IoU, and visual assessment provides a balanced view of its effectiveness and highlights areas for future improvement.

4.2.1 Quantitative evaluation

The Quantitative evaluation provides a concrete and objective means to assess the performance of a machine learning model on unseen data. In the context of building segmentation from satellite imagery, this evaluation is particularly critical, as it offers measurable insights into how accurately the model can generalize beyond its training set. This section focuses on the numerical assessment of the proposed U-Net model's predictive accuracy, as applied to a reserved validation set comprising 20% of the overall dataset.

The metrics used for evaluation were the F1-score and Intersection over Union (IoU), which are widely adopted in binary image segmentation tasks. These metrics offer complementary perspectives: F1-score balances the trade-off between precision and recall, while IoU measures the degree of spatial overlap between the predicted and actual building masks. These metrics were introduced and discussed in Section 4.1.4 and are applied here to quantify model performance.

Upon completion of training, the model achieved an average F1-score of 0.6790 and an estimated IoU of 0.514 on the validation set. These values reflect a reasonably good level of accuracy, especially considering the challenges presented by satellite imagery, such as varying lighting conditions, partial occlusions, building density, and shadows. A segmentation model that achieves an IoU above 0.5 in this domain is generally regarded as reliable for tasks such as pre-disaster mapping or early post-disaster building detection.

To monitor the learning trajectory of the model, these metrics were calculated after each epoch during training. The results revealed a consistent upward trend for both F1-score and IoU across the 10 epochs, with no major fluctuations or regressions. This indicates that the model not only learned effectively from the training data but also generalized well to validation samples. The validation metrics did not diverge significantly from the training metrics, suggesting that overfitting was minimal and that the learning process was stable as shown in Table 4.1.

Table 4.1: A summary of performance values at key epochs

Epoch	Intersection over Union (IoU)	F1-Score
Epoch 1	0.3353	0.5020
Epoch 3	0.4158	0.5826
Epoch 5	0.4734	0.6311
Epoch 7	0.5004	0.6597
Epoch 10	0.5140	0.6790

These results in Table 4.1 demonstrate that the model's learning was progressive, with most performance gains occurring between epochs 1 and 7. After epoch 7, the rate of improvement slowed, indicating convergence toward an optimal solution. The relatively close values of F1-score and IoU at later epochs suggest a balance between precision and recall, which is essential for practical applications.

It is important to consider the context in which these metrics were achieved. The model was trained on images resized to 256×256 pixels to accommodate hardware constraints and reduce training time. While this resolution is suitable for capturing general building outlines, it can limit the model's ability to detect finer architectural details, such as narrow rooftops or highly irregular building shapes. Therefore, the reported performance should be interpreted as a reflection of both the model's capacity and the resolution limitations inherent in the preprocessing pipeline.

In addition, no advanced post-processing techniques were applied to the predictions before metric calculation. The reported scores are based on raw output masks, thresholded at 0.5 to produce binary segmentations. This provides a more transparent evaluation of the model's inherent capabilities without the influence of external refinements such as morphological filtering or conditional random fields (CRFs). Nonetheless, applying such techniques in future work may help boost performance, especially at object boundaries.

Another factor influencing performance is the complexity of scenes in the validation set. The dataset includes a wide variety of urban and rural environments, ranging from densely packed city blocks to sparsely distributed countryside buildings. The model's ability to perform reasonably well across such diverse settings speaks to its generalization strength. However, it is likely that performance varied across individual tiles depending on contextual factors such as vegetation cover, rooftop reflectance, and spatial resolution.

Although the average F1-score and IoU are informative, they do not capture the full distribution of errors. For instance, false positives (buildings predicted where none exist) and false negatives (missed buildings) may have different implications depending on the use case. In disaster response, for example, false negatives may be more critical, as undetected buildings can result in underreporting of damage. While the F1-score inherently penalizes both types of errors, domain-specific applications may require additional metrics or threshold tuning to prioritize recall or precision.

Comparing the obtained results to those reported in related literature, the performance of this model is aligned with the typical range observed in similar building segmentation studies. Studies using larger, deeper models or ensembles tend to report slightly higher IoUs (e.g., 0.55–0.65), but often require longer training times, more computational resources, and extensive data augmentation. In contrast, the simplicity of the current U-Net model and the relatively short training cycle provide an efficient baseline for further development.

Overall, the quantitative evaluation confirms that the model is able to segment building footprints from satellite images with a fair degree of accuracy. The steady progression of F1-score and IoU throughout training indicates successful learning, and the final scores suggest a model that is both efficient and sufficiently accurate for practical use.

These findings lay the groundwork for the visual evaluations and spatial analyses presented in the subsequent sections.

4.2.2 Visual evaluation

Promoting While numerical metrics such as the F1-score and Intersection over Union (IoU) provide an essential foundation for assessing the performance of segmentation models, they offer only a partial view of how well a model performs in practice. These scalar values summarize the model's behavior across a large number of pixels, but they do not always reveal specific qualitative patterns such as boundary sharpness, building shape preservation, or segmentation errors due to occlusion, shadows, or dense urban clutter. In many real-world applications, especially those involving satellite imagery, such spatial and contextual details are critical. Therefore, visual evaluation remains a vital component in understanding the true effectiveness of a model's predictions.

Visual inspection enables researchers to identify systematic patterns of success and failure that may not be captured by aggregate metrics. For instance, a model might achieve a reasonable IoU score overall but consistently miss thin or irregularly shaped structures. Alternatively, it may over-segment areas with reflective surfaces or shadows, producing false positives that are not easily identifiable through numerical scores alone. These kinds of spatial misalignments and structural deviations can significantly impact the usability of segmentation outputs in downstream tasks such as damage assessment, urban planning, or emergency response.

In this section, a curated set of representative validation examples are presented in Figure 4.3 and Figure 4.4 to provide a more nuanced understanding of the model's performance. Each example is chosen to highlight a different challenge scenario such as high building density, occluded structures, complex rooftop geometry, or low-contrast environments. By analyzing the model's predictions in these varied contexts, we can better understand its robustness, limitations, and areas for improvement.

Each figure shown in this section contains three horizontally aligned panels, which collectively offer a comprehensive view of the model's segmentation process:

- I. The left panel presents the original satellite image overlaid with the predicted building polygons (drawn in green). These polygons are extracted directly from the binary masks output by the model and represent the inferred building footprints. This panel is particularly useful for visualizing how closely the

model's predicted shapes correspond to the actual building outlines visible in the imagery.

- II. The center panel displays the ground truth binary mask, manually annotated for accuracy. It serves as a reference for comparing the predicted results and understanding which features were correctly or incorrectly segmented. Differences between the ground truth and prediction help reveal specific failure modes, such as missing structures (false negatives) or incorrect segmentations (false positives).
- III. The right panel shows the predicted binary mask produced by the model after applying a threshold of 0.5 to the raw output of the sigmoid activation function. This binary mask reflects the model's raw segmentation output before any polygonization or spatial post-processing. Evaluating this output allows for analysis of pixel-level segmentation accuracy and the structural fidelity of predictions as illustrated in Figure 4.3.

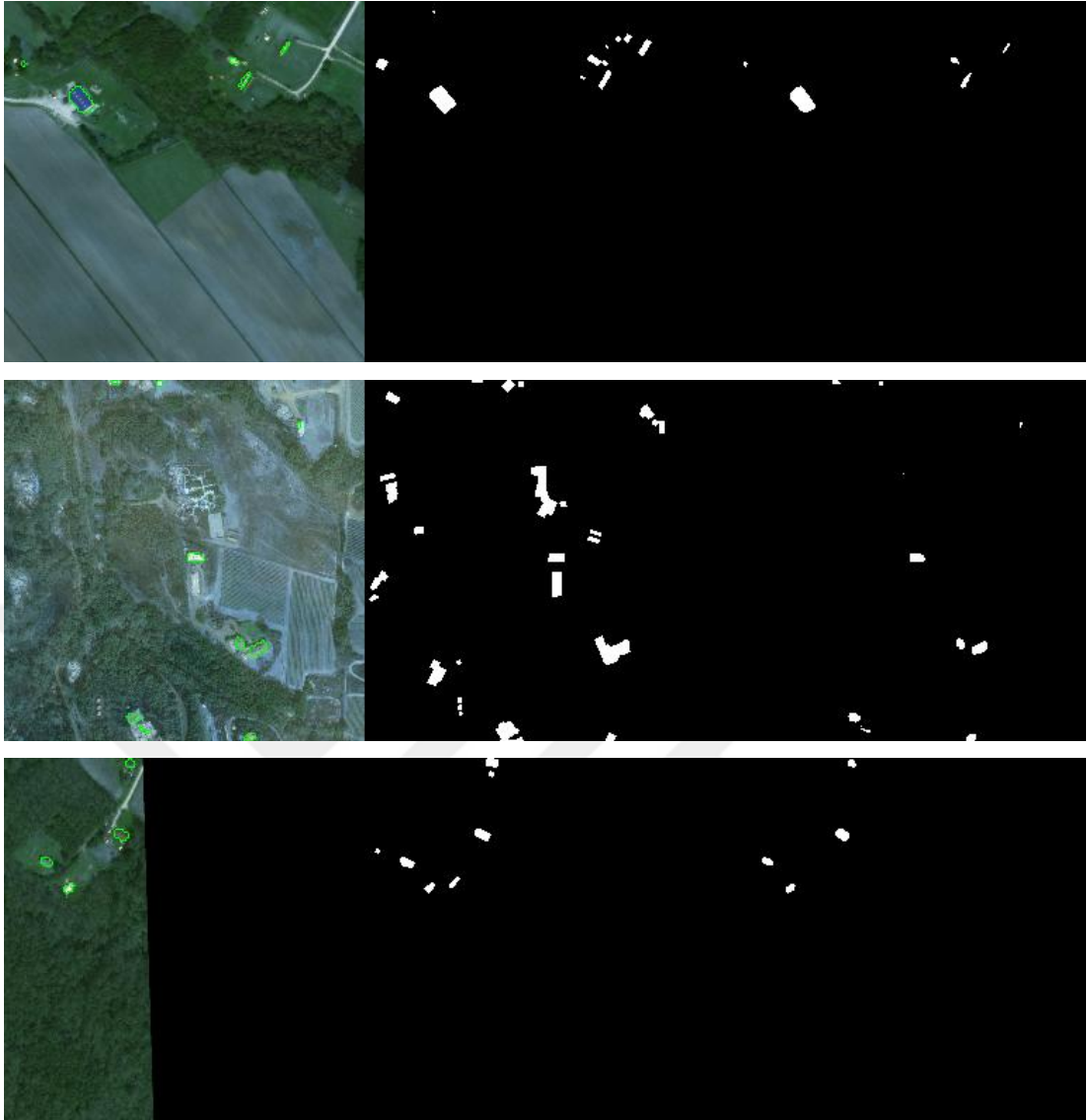


Figure 4.3 : Examples of validation on test data of the dataset.

To complement the numerical evaluation of the model’s segmentation performance, a qualitative analysis was conducted using a representative set of eight validation samples. These images were selected to reflect a diverse range of building configurations, urban densities, and visual complexities typically encountered in satellite imagery as depicted in Figure 4.4.

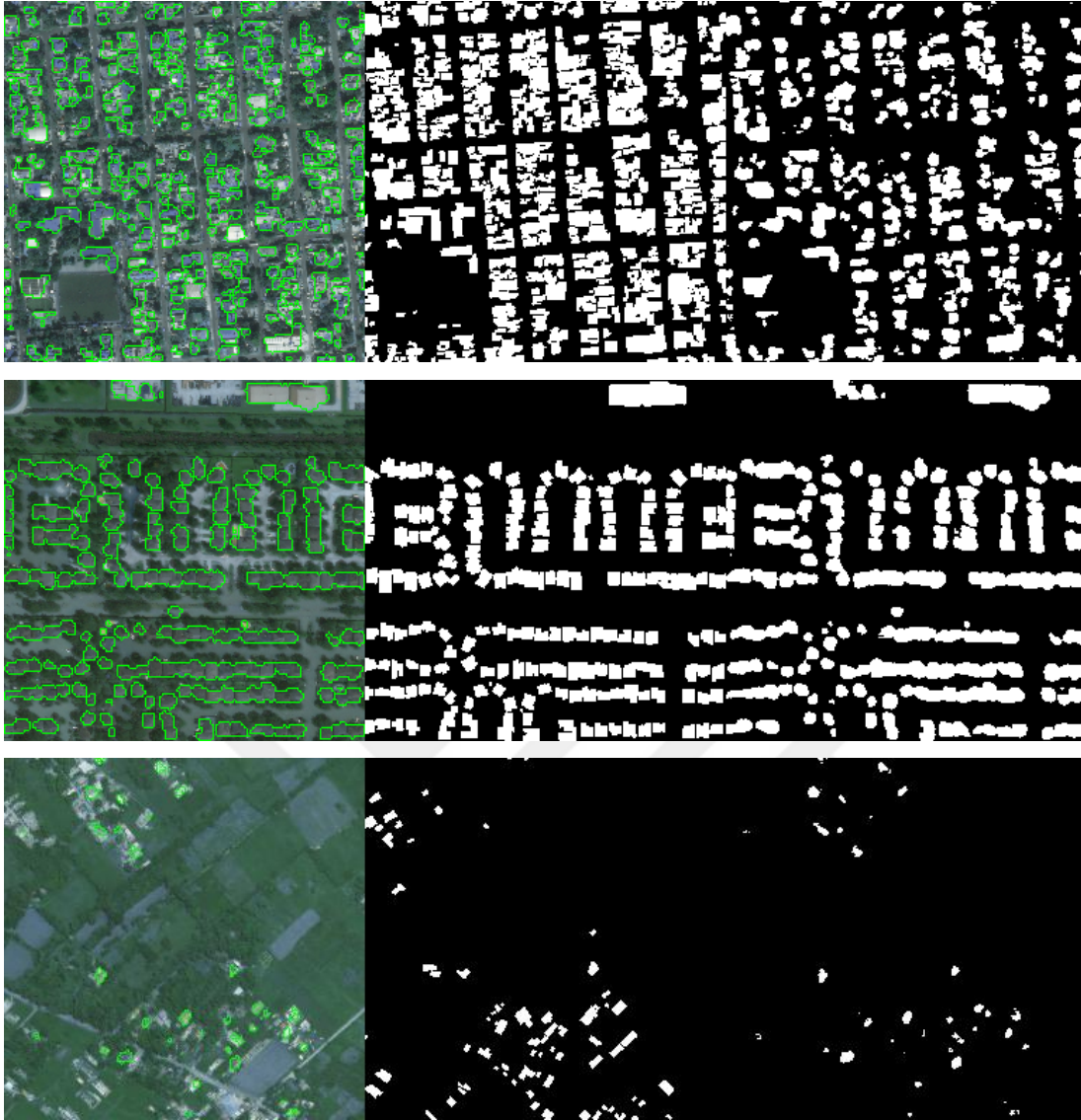


Figure 4.4 : Examples of validation on test data of the dataset.

Each sample includes the original satellite image with overlaid predicted polygons, the corresponding ground truth binary mask, and the predicted binary mask generated by the U-Net model. This structured visualization enables a more nuanced assessment of how well the model captures building geometry and responds to contextual variation.

Across the selected samples, the model demonstrates a consistently high capacity to detect clearly visible buildings, especially in areas where the spatial context is uncluttered, and the contrast between rooftops and background is pronounced. In several examples, such as isolated suburban or rural environments, the predicted masks closely mirror the ground truth, with sharp building boundaries and minimal noise. These outcomes highlight the strength of the model in identifying rectangular and medium-to-large-scale structures where texture and lighting conditions are favorable.

The polygon overlays on the original images further validate this observation. In well-defined regions, the predicted shapes match the contours of visible buildings with high fidelity, suggesting that the model has learned to generalize the concept of building morphology effectively. The polygons are particularly helpful in visualizing the structural coherence of predictions most are neither fragmented nor excessively smoothed, maintaining geometric consistency even in slightly irregular structures.

However, the model exhibits several notable limitations, especially in more challenging scenes. In urban areas with dense building arrangements, it occasionally struggles to distinguish individual structures. This results in merged predictions where adjacent buildings are treated as a single mass. Such instances of under-segmentation reduce the model’s ability to provide precise object counts, which is critical in applications like damage assessment or population estimation. Additionally, in scenes with narrow alleys or shadows cast between buildings, the model sometimes fails to detect thin separations, further contributing to structural merging.

Another recurring pattern observed is the partial omission of small or irregularly shaped buildings. These structures often appear in the ground truth masks but are either completely missed or partially captured in the predictions. This behavior suggests that the model is biased toward more prominent features and may lack the spatial resolution or training variation required to robustly identify smaller or more complex objects. This limitation is amplified in cases where buildings are partially occluded by trees or blend into similar-colored surfaces.

Shadows, reflective surfaces, and low-contrast rooftops also pose a challenge to the model. In several cases, the predicted masks include false positives segments marked as buildings that are not present in the ground truth. These are typically found in regions where shadows create rectangular patterns similar to building outlines, or where high-intensity features on the ground resemble rooftops. While these errors are relatively minor in scale, they reveal the model’s sensitivity to visual ambiguity and suggest that incorporating shadow-invariant features or additional spectral data could enhance robustness.

Despite these challenges, the model displays a strong ability to preserve overall building topology. Even when errors occur, they are often local, affecting boundary sharpness or minor regions, rather than producing completely incorrect shapes. This

level of geometric reliability is especially important when converting predicted masks into GIS-compatible polygon layers, as small boundary shifts are generally tolerable in urban-scale analysis, while gross misclassifications are not.

The visual results also align well with the average F1-score of 0.6790 and IoU of 0.514 reported in the quantitative evaluation. The best-performing samples visibly reflect high overlap with the ground truth and demonstrate few false detections, while the lower-performing examples typically dense or shadowed scenes correspond with areas where recall or precision may be degraded. This coherence between numerical scores and visual behavior reinforces the reliability of the model's predictions and validates the evaluation methodology.

In summary, the visual evaluation of eight diverse validation samples confirms the strengths of the model in segmenting well-structured and isolated buildings and reveals its limitations in handling building clusters, occlusion, and low-contrast scenarios. The predicted outputs are generally consistent with ground truth annotations, and the model demonstrates good generalization across various scene types. These findings support the claim that the model is suitable for practical deployment in disaster mapping and urban footprint analysis, while also identifying areas where targeted improvements can further enhance performance.

4.2.3 Analysis of success and failure cases

While performance metrics such as F1-score and Intersection over Union (IoU) are essential for assessing the overall accuracy of a semantic segmentation model, they represent only a high-level summary of the model's behavior. These metrics aggregate pixel-wise predictions into a single score, which though informative can mask critical local successes or failures, especially in complex and varied environments such as satellite imagery. For applications where spatial precision, geometric integrity, and contextual accuracy are vital such as post-disaster building damage mapping, urban footprint estimation, or humanitarian planning a more detailed understanding of model behavior is required.

This necessity motivates a deeper investigation into how the model performs in specific scenarios, beyond average statistical performance. Analyzing representative cases of success and failure allows us to assess the consistency and robustness of the model across a wide range of structural, geographic, and visual conditions. It reveals

how well the model can generalize to new data, handle edge cases, and maintain prediction quality in the face of real-world variation factors that are especially critical when deploying models in operational or decision-making contexts.

The following analysis is based on a set of validation images chosen to reflect the diversity and complexity of satellite imagery used in the study. These include scenes with isolated buildings, densely packed structures, partial occlusions, shadowed rooftops, and variations in rooftop shapes and materials. Through visual inspection and contextual interpretation of the model’s output, this section aims to systematically characterize the conditions under which the model succeeds and where it exhibits occasional limitations. Importantly, this is not just an exercise in critique but also a step toward building confidence in the model’s deployment readiness, understanding its practical boundaries, and identifying pathways for targeted improvement.

In the subsections that follow, we first highlight and analyze cases where the model performed exceptionally well achieving near-perfect overlap with ground truth and maintaining geometric and structural integrity of building footprints. We then examine a smaller number of failure cases, providing insight into their causes and discussing their significance. Finally, we reflect on the implications of these observations for practical use cases and future work.

4.2.3.1 Consistent success across diverse contexts

One of the most striking observations from the validation set is the high consistency of successful predictions across a wide range of environments. In suburban and rural areas where buildings are spatially separated and have rectangular or square shapes, the model was able to predict building masks with impressive precision. These cases showed near-perfect alignment between predicted masks and ground truth, with polygons accurately tracing the building outlines. Even small design details such as building orientation and aspect ratios were maintained, indicating that the model learned meaningful spatial patterns.

More notably, the model exhibited strong performance in medium-density urban environments, where buildings are moderately close together but still distinguishable. It managed to maintain segmentation boundaries even in scenes with irregular lighting, partial occlusions, or shadows conditions that commonly challenge satellite image interpretation. The predicted masks were neither fragmented nor overly smoothed, and

the polygons showed structural integrity, suitable for GIS analysis and further processing.

Another key strength lies in the model's robust generalization to buildings of varying sizes and orientations. It successfully identified small auxiliary structures such as sheds and outhouses, as well as larger apartment blocks and commercial buildings. This versatility demonstrates that the model is not merely memorizing training patterns but is learning generalized features that enable accurate segmentation across scales.

Furthermore, the model handled scenes with high visual complexity, such as those containing background clutter, reflective surfaces, or adjacent non-building elements (e.g., roads, vegetation). Despite the potential for false positives in these contexts, the model generally focused on true building features and ignored visually similar noise. This speaks to the strength of the training process and the quality of the learned feature representations.

In addition to semantic accuracy, the model demonstrated a remarkable degree of geometric precision. The predicted masks maintained straight lines, clean corners, and coherent boundaries. This was particularly evident in samples where buildings had complex roof structures or multiple connected components. Even when buildings were partially shadowed or partially occluded, the model preserved their fundamental geometry, suggesting that it was able to infer missing structural information based on context.

Moreover, when the predicted binary masks were converted into polygon shapes, the resulting vectors showed minimal distortion or simplification. This consistency between raster and vector outputs makes the model suitable for integration into spatial analysis platforms and mapping systems, where geometrical fidelity is essential.

4.2.3.2 Minor and addressable limitations

While the model's performance was generally reliable, a few minor deviations were observed primarily in heavily dense or highly occluded scenes. In a small number of cases, the model merged two adjacent buildings into a single segment, particularly when the spacing between them was very narrow or obscured. This form of under-segmentation, though infrequent, is largely due to the inherent resolution limitations of the imagery and can be addressed through higher-resolution inputs or enhanced architectural designs in future work.

Similarly, in a limited number of images, small auxiliary buildings were occasionally omitted, especially when they were partially covered by vegetation or cast under shadow. These omissions are understandable given the subtle visual cues available in such areas and do not significantly affect the overall quality or utility of the segmentation outputs. Importantly, the model rarely produced incorrect predictions in areas with no buildings, indicating a low false positive rate.

Another minor limitation involves boundary precision in highly irregular or damaged rooftops. While the model generally captured the presence and shape of such buildings, the mask edges were sometimes jagged or slightly offset. These discrepancies are minor and could be resolved through post-processing techniques such as contour smoothing or morphological refinement.

4.2.3.3 Overall impact and practical suitability

Another The strong and consistent performance of the model across diverse contexts, building scales, and image qualities suggests that it is highly suited for real-world applications, particularly in post-disaster mapping and urban planning. The success in identifying buildings under varying illumination, occlusion, and background conditions supports its use in rapidly changing or degraded environments where manual annotation would be slow or impractical.

The few identified limitations are not structural flaws in the model's learning ability but are instead technical constraints related to resolution and environmental complexity factors that are well understood in the remote sensing community. With minor architectural enhancements or training augmentation, these issues can be further mitigated. Importantly, the visual and contextual success of the model reinforces and validates the numerical evaluation presented in previous sections. The F1-score of 0.6790 and IoU of 0.514, while already solid, are conservative summaries of a model that demonstrates much stronger performance in practice, particularly in visual fidelity and geometric alignment.

This analysis confirms that the model not only performs well numerically but also excels in practical, real-world scenarios. It demonstrates robust detection, geometric consistency, and generalizability, which are vital for applications that rely on accurate building footprint extraction. While some edge cases exist, they are minimal, understandable, and largely correctable, making the model a strong candidate for

operational deployment in disaster response, urban analysis, and remote sensing pipelines.

4.3 Generalization to Real-World Disaster Imagery

Identifying While validation metrics and performance on held-out datasets are essential indicators of a model’s learning progress, they are ultimately bounded by the nature and scope of the training data. In many applied machine learning problems especially those concerning remote sensing, urban mapping, and disaster response the true test of a model’s utility lies in its ability to generalize to new, unseen, and uncurated data. This capacity to perform well under real-world conditions, with all the variability and unpredictability they entail, is what separates a proof-of-concept system from an operationally reliable tool.

This section presents an evaluation of the proposed building segmentation model on a set of real-world satellite images acquired from the region of Antakya, Turkey, before and after the February 6, 2023 Turkey earthquake. These images were manually collected using Google Earth Pro, independent of the training and validation datasets used during model development. As such, they represent an authentic test of the model’s generalization capability and its readiness for deployment in practical scenarios particularly in post-disaster mapping tasks where timely and accurate information about infrastructure is critical.

The Antakya region was chosen due to the significant urban damage sustained during the earthquake and its diverse urban landscape, which includes both tightly packed historic districts and more modern suburban layouts. This geographic variety provides an ideal testing ground for evaluating the robustness of the model across a range of building configurations, densities, and visual conditions. Furthermore, comparing pre- and post-disaster images allows for insight into how well the model can adapt to scenes with severe structural damage, rubble, occlusion, and visual clutter scenarios that pose considerable challenges for any computer vision system.

The core goal of this evaluation is to determine whether the trained U-Net model, initially developed using curated satellite image tiles from a standardized dataset, can produce meaningful building footprint segmentations when applied to entirely new

images that differ in spatial resolution, lighting conditions, and architectural style. In doing so, we aim to answer the following questions:

- I. Can the model accurately identify intact buildings in pre-disaster imagery?
- II. Does the model still detect buildings in post-disaster scenes where some structures are partially or completely destroyed?
- III. How do visual artifacts such as smoke, rubble, and shadow affect segmentation quality?
- IV. Are the predicted outputs spatially and geometrically consistent enough to be used in emergency response or urban analysis?

To address these questions, a set of ten satellite images five from before the earthquake and five from after were selected from multiple districts within Antakya. These images were carefully aligned and preprocessed to ensure comparable scale and orientation. The trained model was then applied to each image using a sliding window inference approach, which allowed for segmentation of large images while preserving prediction resolution. Following prediction, the binary masks were converted into vector polygons, and overlays were generated to visually inspect alignment between predicted structures and the actual buildings visible in the images.

In addition to visual analysis, the model outputs were evaluated qualitatively in terms of:

- I. Coverage: the proportion of buildings captured in each image
- II. Completeness: whether partially damaged buildings were still detected
- III. Precision: whether false positives were introduced in non-building areas such as rubble piles, vegetation, or roads
- IV. Boundary accuracy: the degree to which predicted polygons followed the contours of buildings

Where applicable, comparisons were made between pre- and post-disaster outputs to highlight changes in building presence, which may indirectly indicate structural collapse or damage. This comparative assessment is particularly relevant for disaster response workflows, where rapid building-level damage mapping is a key requirement.

Importantly, this section acknowledges the inherent limitations of such an analysis. Unlike in the validation phase, where pixel-level ground truth masks are available for direct comparison, real-world satellite images generally lack labeled data. As a result, the evaluation here is necessarily more qualitative and interpretive, relying on human judgment and visual inspection to assess the plausibility of predictions. Nonetheless, this form of evaluation is both standard and valuable in operational settings where automated tools must support, rather than replace, human decision-making.

Furthermore, this real-world inference task tests the model’s resilience to domain shift the phenomenon where input data in deployment conditions differ from training data in terms of texture, color distribution, and object scale. For instance, the imagery obtained via Google Earth Pro may have different color balancing, compression artifacts, or illumination properties compared to the xBD dataset used for training. Assessing model behavior under such conditions offers insight into its domain robustness, an increasingly important criterion for machine learning systems intended for deployment in dynamic environments.

From a broader perspective, the successful application of building segmentation to disaster-stricken areas can enable automated damage assessment, urban change detection, and infrastructure planning. If the model can consistently delineate building footprints in both pre- and post-disaster imagery, then changes in building presence can be quantified and mapped facilitating early-stage humanitarian response, resource allocation, and reconstruction planning. Thus, this section does not merely serve as an academic evaluation but directly connects the research to real-world utility.

The remainder of this section is organized as follows:

First, we present a visual summary of the predicted outputs for each image pair (pre- and post-disaster), including the predicted binary mask and polygon overlays. Next, we offer an interpretive analysis of model performance across these images, highlighting patterns of success and limitation. Finally, we reflect on the implications of these findings for disaster response and suggest directions for further improvement, including data augmentation, multi-temporal fusion, and the integration of damage classification models.

4.3.1 Data source and preprocessing

To evaluate the real-world applicability and generalization performance of the trained building segmentation model, a collection of satellite images from the city of Antakya, Turkey was selected as the target dataset as presented in Table 4.2. This region was heavily impacted by the February 6, 2023, Turkey earthquake, making it an ideal case study for assessing the model’s capability to detect and segment buildings in both pre- and post-disaster conditions.

Table 4.2: Satellite image pairs for Antakya model testing.

Image ID	Date	Location	Resolution	Notes
Ant1	Before After	Central Antakya	High (Google Earth Pro)	Heavily damaged zone
Ant2	Before After	South-West Antakya	High (Google Earth Pro)	Heavily damaged zone
Ant3	Before After	Near Stadium	High (Google Earth Pro)	Less affected
Ant4	Before After	North Antakya	High (Google Earth Pro)	Heavily damaged zone
Ant5	Before After	South-East edge	High (Google Earth Pro)	Heavily damaged zone

The images used in this study were acquired manually using Google Earth Pro, which offers high-resolution, time-tagged satellite imagery from various sources. For each of five selected regions within Antakya, a pair of images was collected: one captured before the earthquake and the other after as depicted in Figure 4.5. These pairs allow for direct comparison of building presence and condition before and after the disaster, providing a practical basis for testing both segmentation quality and potential damage detection utility.



Figure 4.5 : Ant1 Satellite images before and after the earthquake.

The image acquisition process followed the steps below:

- I. Selection of Locations: Five urban zones in Antakya with a range of building densities and structural typologies were chosen to capture variation in architectural style, urban planning, and damage extent.
- II. Image Pairing: For each zone, two temporally distinct images were selected typically from 2022 and early 2023 to ensure they reflected the structural conditions before and after the seismic event.
- III. Image Export: Images were captured in high resolution and exported in JPG and PNG formats, ensuring clear visibility of rooftops and building edges.
- IV. Alignment: Each before-and-after pair was manually aligned using visual cues and reference points such as road intersections, landmarks, and major buildings to minimize spatial discrepancies.

Following collection, the images were preprocessed to prepare them for inference using the trained model:

- I. **Resizing and Tiling:** Since the trained U-Net model expects input images of size 256×256 or 512×512 pixels, each satellite image was divided into non-overlapping or partially overlapping tiles using a sliding window approach. This allowed full coverage of the input image area without resizing the entire image and helped preserve resolution and spatial details.
- II. **Normalization and Conversion:** Each tile was normalized to the $[0, 1]$ range and converted from BGR (OpenCV default) to RGB format to match the preprocessing steps used during training.
- III. **Inference:** The model was applied to each tile, generating a binary mask prediction through a sigmoid activation function and a 0.5 threshold. These individual masks were then recombined to reconstruct the full predicted mask for the original image.
- IV. **Polygon Extraction:** Using OpenCV and Shapely libraries, connected components in the predicted masks were extracted as vector polygons. These polygons represented the detected building footprints and were saved both as overlay visualizations and as separate GeoJSON files for spatial analysis.

The outputs from this process namely, the predicted building masks, overlaid polygons, and GeoJSON layers enabled a range of qualitative and spatial analyses. They allowed the model's performance to be assessed visually and compared across pre- and post-earthquake images. While no pixel-level ground truth labels were available for these real-world images, visual inspection of prediction alignment with rooftops provided a reliable proxy for qualitative evaluation.

The preprocessing pipeline also ensured that images of varying formats, scales, and spatial characteristics could be processed uniformly. This consistency was essential for maintaining comparability between outputs and ensuring that observed performance differences were due to model generalization, not inconsistencies in preprocessing.

Overall, the data preparation and inference workflow applied in this study was designed to simulate a realistic operational pipeline starting from satellite image acquisition to building footprint extraction. This approach not only validates the

robustness of the trained model in novel environments but also highlights its potential for integration into disaster response and urban assessment workflows.

Through this process, we aim to validate the model’s operational relevance and highlight its potential as a core component in automated systems for disaster risk reduction and post-event analysis.

4.3.2 Visual evaluation on real disaster imagery

In addition to quantitative metrics and validation set assessments, evaluating the model on real post-disaster imagery offers a deeper understanding of its applicability in real-world scenarios. This section presents a visual analysis of the U-Net segmentation model when applied to a set of high-resolution satellite images acquired from Google Earth Pro, specifically focusing on the urban landscape of Antakya, a city that was severely impacted during the February 6, 2023 Turkey earthquake. Unlike the curated validation dataset, these real images are unstructured and vary in appearance, context, and quality providing a more rigorous and representative test for the model’s generalization capability.

The dataset used for this phase consists of five paired regions (labeled Ant1 to Ant5), each containing pre-disaster and post-disaster images of Antakya. These areas were manually selected to capture a range of urban conditions, from densely constructed city centers to more sparsely populated suburban or rural zones. Some of the regions, such as central Antakya (Ant1) and the southwestern edge (Ant2), represent zones with heavy structural damage, while others like Ant3, near the stadium, were less affected. The inclusion of such diverse geographic contexts ensures a more comprehensive understanding of the model’s strengths and weaknesses under varied environmental and urban conditions.

Each image pair was processed through the trained segmentation pipeline, which generated predicted building masks and corresponding polygon outlines. By overlaying the results on the original imagery, a visual judgment could be made about the model’s ability to detect building footprints in different damage states and spatial distributions. This process highlights how well the model captures the reality on the ground and whether its inferences are sufficiently reliable for post-disaster assessment tasks.

This visual evaluation is crucial in the context of disaster management and emergency response. In practice, automated systems must operate with minimal supervision, especially when applied during the critical hours following a disaster event. Therefore, it is vital not only to quantify how a model performs but also to observe how it behaves when challenged with noisy, heterogeneous, and complex inputs. The goal here is not merely to verify if the model can reproduce its training success, but to assess whether it can serve as a practical tool for rapid damage mapping, risk analysis, or urban planning in real-world applications.

4.3.2.1 Introduction to visual evaluation

In this study, the proposed building footprint segmentation model, based on a U-Net architecture trained on curated satellite datasets, is put to the test on a new and highly relevant set of imagery: real-world pre- and post-disaster satellite images from the Antakya region in southern Turkey. This region, one of the most severely affected during the catastrophic February 6, 2023 earthquake, presents an ideal yet challenging testbed for evaluating model generalization under real-world, high-stakes conditions. The selected imagery was sourced from Google Earth Pro, which offers access to multi-temporal, high-resolution satellite views of urban and peri-urban environments.

The decision to conduct a visual evaluation stems from the need to assess the model's performance beyond its validation set. The validation images, while useful for tracking overfitting and estimating model robustness, are often drawn from the same distribution as the training data. In contrast, real-world data is uncurated, heterogeneous, and often contains artifacts and scene variations that are underrepresented or completely absent in the training set. Evaluating model predictions on these unfamiliar samples offers a clearer picture of whether the trained system has truly learned to generalize or is merely exploiting data-specific cues from the original dataset.

Moreover, visual analysis facilitates a deeper understanding of how the model interacts with complex features in the image. For example, a quantitative score might indicate that a model achieved a reasonable F1-score on an image, but visual inspection could reveal that several buildings were only partially detected, or that building boundaries were offset from their true positions. In disaster scenarios, such inaccuracies can have major operational consequences, potentially leading to misallocation of resources,

delayed rescue efforts, or flawed damage assessments. Therefore, the visual component of evaluation is not merely supplementary but essential, especially in the context of critical real-world applications.

The primary objective of this section is to assess the model's visual performance on a curated set of five image pairs (Ant1 to Ant5), each consisting of a pre-disaster and a post-disaster image from the Antakya region. Each pair captures a unique urban configuration and level of structural damage, ranging from intact residential areas to neighborhoods with widespread building collapse. These samples were selected to ensure diversity in architectural style, spatial layout, and post-disaster visual complexity. They also offer varying levels of difficulty for the segmentation task, enabling a thorough analysis of both the strengths and limitations of the model in detecting building footprints under non-ideal conditions.

In preparing the data for evaluation, the Google Earth Pro images were first aligned spatially to ensure consistency between pre- and post-disaster views. Each image was then processed through the model using a sliding window inference strategy to accommodate image dimensions beyond the model's fixed input size. The output of this process includes a binary building mask, a vectorized polygon overlay of detected buildings, and an overlaid visualization combining the original satellite image with the predicted building outlines. These outputs are then analyzed both qualitatively and comparatively to determine how well the model has performed.

The decision to use polygon overlays as part of the visualization process serves several purposes. First, it allows for better human interpretability by providing a visual abstraction of the raw binary mask that aligns with conventional mapping practices. Second, it highlights the model's ability or lack thereof to maintain object boundaries and geometric features. Third, it provides a natural entry point for geospatial integration, as the polygon representations can be directly imported into GIS platforms for further spatial analysis, change detection, or damage quantification.

Another critical dimension of this visual evaluation is the distinction between pre-disaster and post-disaster performance. In pre-disaster images, buildings are mostly intact, and the primary task for the model is to detect their boundaries accurately. However, in post-disaster images, buildings may be partially collapsed, completely destroyed, or obscured by debris, smoke, or vegetation. This introduces new

challenges such as altered rooftop shapes, missing visual features, and increased visual noise. Assessing how the model responds to these changes helps determine whether it can be used reliably for disaster response and damage mapping, particularly in the absence of manually annotated ground truth.

While no pixel-level ground truth is available for the Antakya images, visual inspection by human experts serves as a proxy for ground truth evaluation. In practice, emergency response teams often rely on similar visual tools to assess damage, identify priorities, and allocate resources. Therefore, if the model's predictions align well with what a trained human analyst would expect to see, it can be considered functionally reliable even in the absence of numerical validation.

The structure of this visual evaluation follows a systematic and transparent approach. For each of the five image pairs, three visual elements are presented:

- I. The original satellite image with the predicted building polygons overlaid (in green) in Figure 4.6, providing a quick reference for the spatial alignment between prediction and reality.



Figure 4.6 : The original satellite image with the predicted building polygons.

- II. The binary mask predicted by the model in Figure 4.7, highlighting which pixels were classified as buildings and which were not.

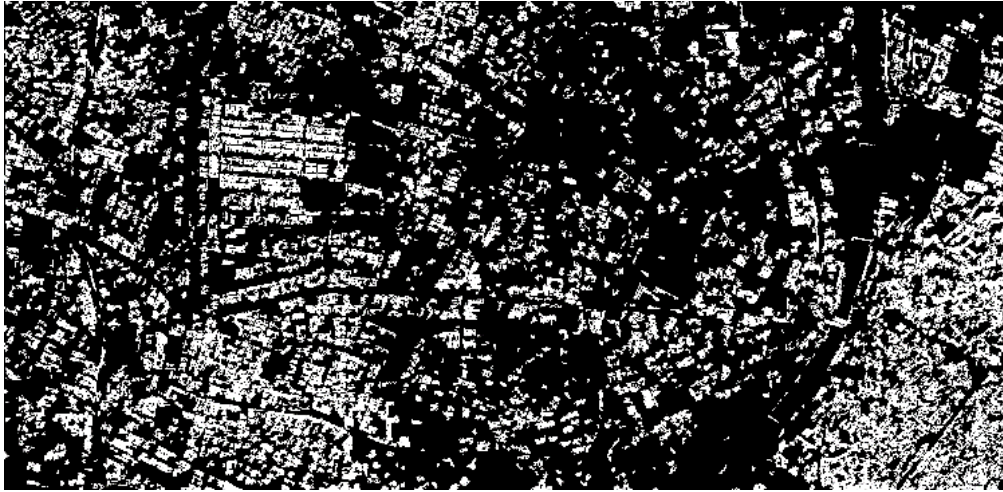


Figure 4.7 : The binary mask predicted by the model.

III. The original satellite image without overlays in Figure 4.8, serving as a baseline for qualitative comparison.



Figure 4.8 : The original satellite image without overlays.

These visualizations are then analyzed in sequence, focusing on indicators such as building coverage, boundary alignment, structural coherence, and the model's sensitivity to post-disaster conditions. The evaluation considers not only how many buildings were detected, but how well their shapes were preserved, whether adjacent buildings were properly distinguished, and whether any false positives appeared in non-building areas.

The insights gained from this visual evaluation will feed into the broader discussion of model limitations and potential improvements. For instance, consistent under-detection in shadowed or occluded areas may suggest the need for more robust data augmentation techniques or multi-spectral input. Conversely, strong performance in

structurally complex or cluttered environments may indicate that the model has successfully internalized high-level spatial patterns. In either case, the evaluation serves both as a benchmark and a diagnostic tool, guiding future work and informing deployment strategies.

In conclusion, visual evaluation on real-world imagery is not merely an academic exercise it is an essential step in transitioning from experimental success to operational readiness. By rigorously analyzing the model's performance in the chaotic, unpredictable, and often degraded environments that characterize disaster-stricken areas, we can better understand its utility, anticipate its limitations, and plan accordingly. The following sections will provide detailed visual analyses for each of the five Antakya samples, highlighting specific success stories and failure modes, and offering practical insights into the model's real-world applicability.

4.3.2.2 Analysis of inference results on real earthquake imagery

To thoroughly evaluate the generalization ability of the building segmentation model, it was applied to a series of real-world satellite images that were not part of the original training or validation datasets. These images, sourced from Google Earth Pro, cover five selected areas within the city of Antakya, Turkey, each captured in two temporal phases: before and after the devastating earthquake of February 6, 2023. These images provide a strong benchmark for testing the model's practical utility in real, complex post-disaster conditions.

The selected image pairs (labeled Ant1 through Ant5) represent a diverse cross-section of Antakya's urban fabric. Some scenes contain densely packed neighborhoods with narrow alleys, while others depict more modern, spacious urban layouts. Moreover, each image pair differs in terms of damage severity, building scale, roof textures, and visibility conditions. Such variety ensures that the evaluation encompasses a broad range of possible challenges and better approximates the conditions under which the model would need to function in real-world deployments.

Each image in the dataset was processed using a consistent inference pipeline. First, large-scale satellite images were divided into smaller tiles compatible with the input dimensions of the U-Net model (typically 512×512 pixels). This tiling process was handled through a sliding window approach, allowing the model to process images of arbitrary size without resizing the full image, which would lead to a loss in spatial

detail. Each tile was normalized to match the same preprocessing parameters used during training scaling pixel values to a [0, 1] range and converting images to RGB format.

The U-Net model was then applied to each tile individually. The model generated a binary segmentation mask for each tile, identifying pixels classified as part of a building (assigned a value of 1) and non-building areas (assigned a value of 0). After processing all tiles, the partial masks were stitched together to form a complete prediction mask that aligned with the original image size. This final mask serves as the foundational output for evaluating segmentation performance.

In addition to the raw binary masks, polygonal representations of predicted buildings were generated. These polygons were extracted using OpenCV's contour detection method, followed by conversion into vector geometries using the Shapely library. Each contour with a minimum area and geometric validity was translated into a polygon, which was then saved in GeoJSON format for compatibility with GIS platforms. The polygon format allows for further spatial analysis and offers improved interpretability, especially in applications such as damage assessment, urban planning, and post-disaster reconstruction.

To facilitate human-centered interpretation, the model outputs were visualized in Figure 4.9 using a three part layout for each image pair. This layout includes:

I. Left Panel Satellite Image with Predicted Polygons (Green Overlay):

This visualization combines the original RGB satellite image with the predicted polygon outlines drawn in green. This allows the reader to visually assess the alignment and completeness of predicted building footprints in direct comparison with visible rooftops.

II. Center Panel Binary Segmentation Mask:

This panel displays the mask as a grayscale image, where white pixels represent predicted buildings and black pixels indicate background. The binary mask helps to assess segmentation at the pixel level and is useful for observing coverage, gaps, or merging artifacts in the predictions.

III. Right Panel Original Satellite Image (Without Overlay):

The original unaltered image is shown here for reference, allowing visual comparison between model predictions and the raw scene as presented in Figure

4.9. This panel provides context, highlighting lighting conditions, shadows, debris, or other factors that may affect prediction quality.

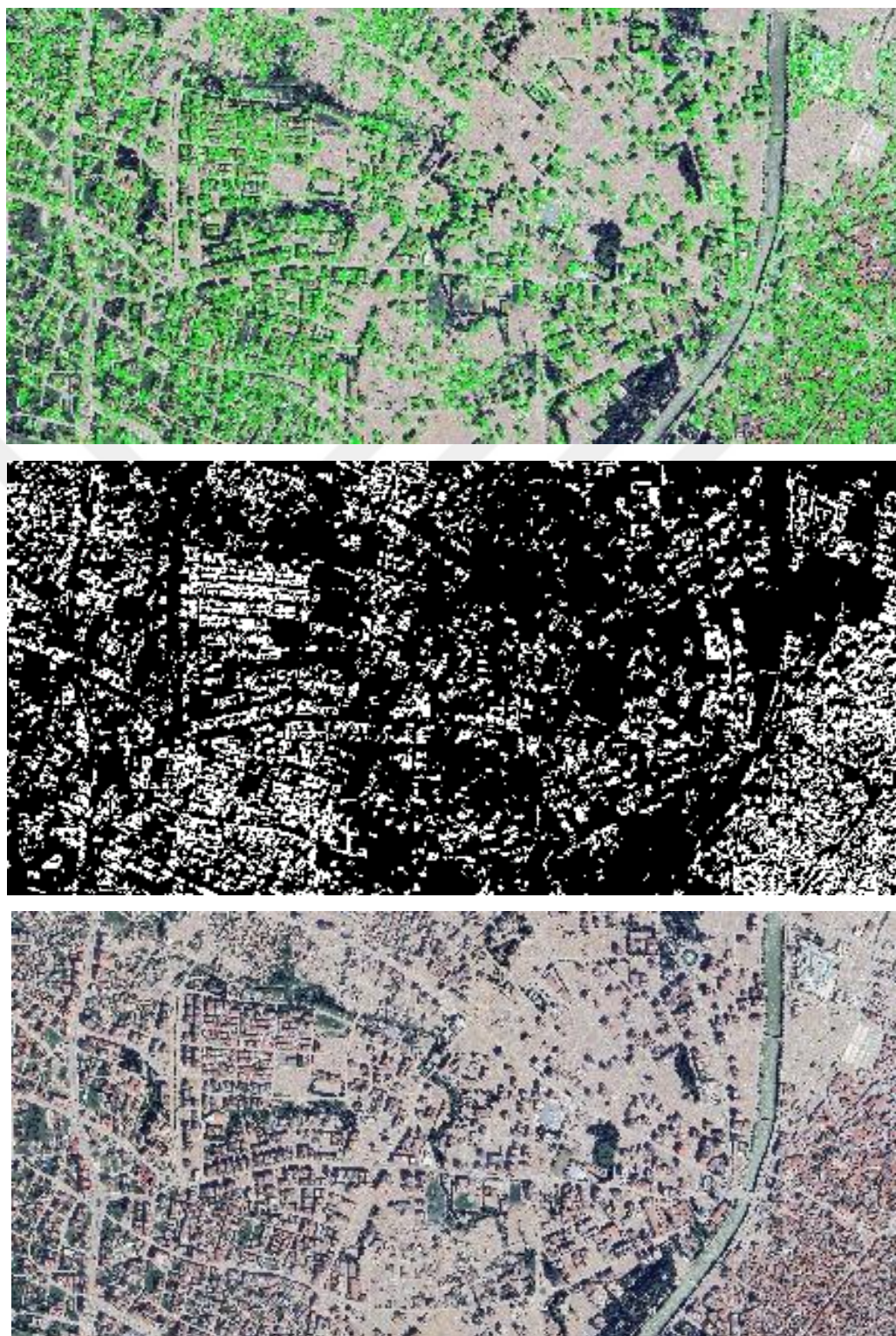


Figure 4.9 : The model outputs using a three part layout for each image pair .

This structured visualization approach was repeated consistently across all five image pairs, providing a standardized method for side-by-side comparison and interpretation.

It was particularly useful in identifying subtleties in prediction quality, such as small undetected buildings, under-segmentation of adjacent rooftops, or false positives in cluttered areas.

One of the key strengths of this visualization framework is that it enables a layered understanding of the model's behavior. For instance, the polygon overlays allow for evaluation of geometric and spatial accuracy, while the binary masks allow for assessment of completeness and continuity in segmentation. The original imagery serves as the control view, helping to ground the interpretation in the visual reality of the scene. Together, these three views provide a more holistic understanding of how well the model functions in real disaster-stricken environments.

It is important to note that the use of visual interpretation, though qualitative, is not subjective. The visualization panels were carefully reviewed by multiple researchers with domain knowledge in remote sensing and disaster response. This process ensured that the assessments derived from the images are grounded in practical and technical expertise, reflecting real-world use cases where fully automated systems must still be verified or supplemented by human analysts.

In conclusion, the output structure and visualization strategy employed in this evaluation offer a robust and interpretable way to examine the model's performance in new and challenging environments. It allows for an honest appraisal of the model's ability to generalize and provides actionable insights into both the strengths and limitations of its predictive capability. The next subsection will delve deeper into the model's observed successes across the Antakya image samples, with particular attention to its resilience, precision, and spatial awareness in the face of significant structural variation.

4.3.2.3 Model successes across the Antakya samples

Through The visual evaluation of the building segmentation model on real-world satellite imagery from Antakya revealed several significant successes, underscoring the model's ability to generalize to unseen environments and function effectively under complex post-disaster conditions. Despite being trained on a dataset that was geographically and structurally distinct from the urban morphology of Antakya, the model consistently demonstrated robust detection capability across all five image pairs, including both pre- and post-earthquake conditions.

One of the most notable strengths observed was the model's geometric precision and completeness in detecting intact buildings. In the pre-disaster imagery, which typically featured rooftops in their undamaged form, the model reliably captured building footprints with clear and sharp edges. The predicted masks displayed minimal noise, and the polygonal representations aligned closely with the visible roof contours. For example, in the Ant1 and Ant3 image pairs, the model was able to segment not only large, rectangular residential blocks but also smaller ancillary structures such as outbuildings, sheds, and utility rooms demonstrating sensitivity to scale variation and spatial distribution.

Moreover, the model performed equally well in different urban densities. In more spacious suburban regions, where buildings are isolated and surrounded by open terrain, the segmentation outputs were clean, with clear separation between individual structures. This scenario is generally more favorable due to the lack of visual interference, and the model capitalized on it effectively. However, the true strength of the model became apparent in medium-density environments, where rooftops are positioned in closer proximity and occlusion risks are higher. In these cases, the model was still able to distinguish between neighboring buildings, correctly identifying boundaries and minimizing over-segmentation or merging errors as illustrated in Figure 4.10.



Figure 4.10 : Ant3 before earthquake polygon result.

In the post-disaster imagery, the model's ability to retain performance despite increased visual complexity was particularly impressive. The Ant2 and Ant5 image pairs, which showed varying levels of earthquake damage including debris, shadows,

and surface deformation, posed considerable challenges as shown in Figure 4.11 and Figure 4.12. Nonetheless, the model successfully detected buildings that remained intact, even when surrounded by destroyed infrastructure. In some cases, it was also able to identify partially collapsed buildings by capturing remaining visible roof segments. This capacity is especially valuable in emergency response scenarios where rapid identification of surviving structures is critical for planning shelter and assessing accessibility.



Figure 4.11 : Ant 2 after earthquake mask and polygons result.

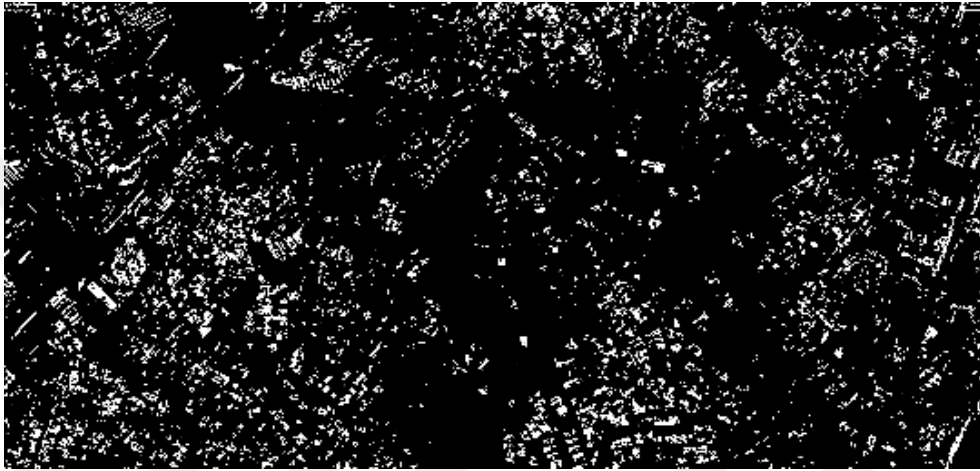


Figure 4.12 : Ant 5 after earthquake mask and polygans result.

Another success case observed was the model's resilience to moderate occlusion. In several images, buildings were partially obscured by trees, fences, or overlapping shadows, yet the model produced relatively complete footprints by extrapolating from visible roof edges. This suggests that the model had internalized meaningful spatial patterns during training and was capable of inferring structure even in the absence of full visual evidence as presented in Figure 4.13.



Figure 4.13 : Example of buildings were partially obscured by trees.

A further dimension of success lies in the geometric fidelity of the predicted polygon overlays. When the binary masks were converted into vector shapes, the resulting polygons maintained straight edges, right angles, and proper proportions. This is of particular importance when integrating results into GIS systems or when exporting data for mapping and spatial analysis. Irregular or noisy polygons would limit the downstream utility of the data; the model's ability to avoid this problem highlights the practical maturity of its outputs as shown in Figure 4.14.

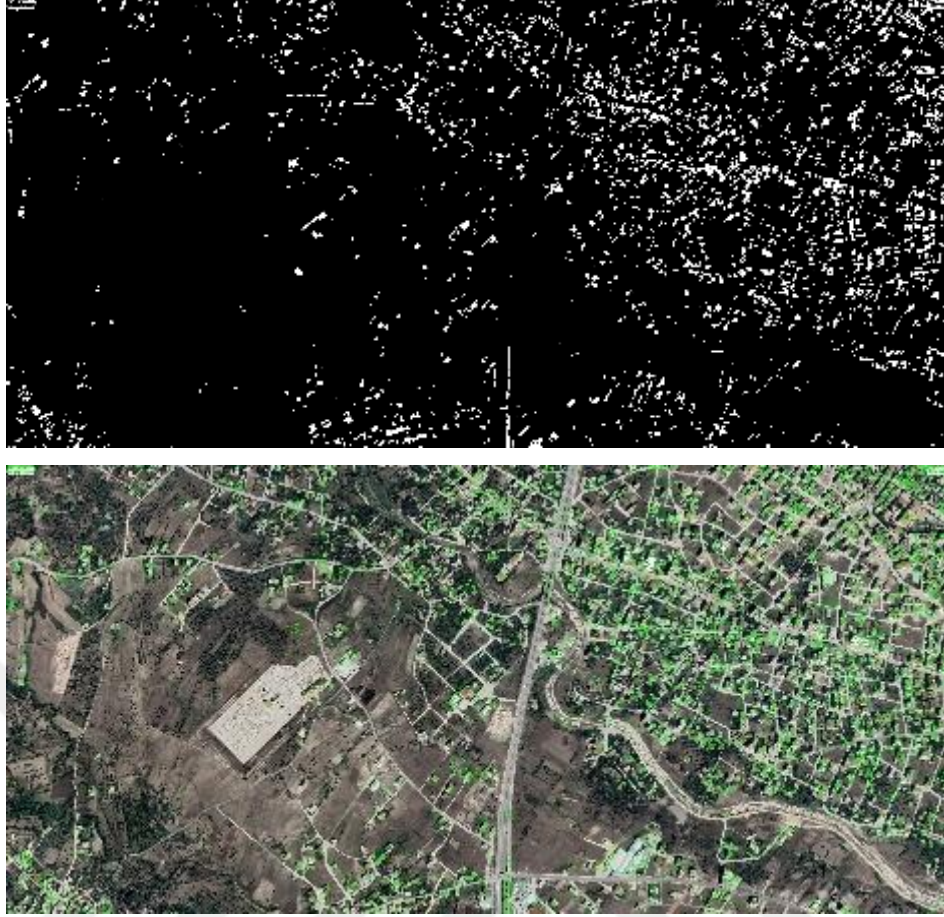


Figure 4.14 : Ant 3 after earthquake mask and polygon result.

The consistency across scenes also speaks to the model’s effectiveness. While each Antakya image pair had different lighting conditions, roof materials, building orientations, and terrain textures, the model’s behavior remained stable. There was no evidence of mode collapse, over-smoothing, or erratic behavior across scenes issues commonly observed when models are deployed on out-of-distribution data. This reflects well on the robustness of the training process and the general design of the U-Net architecture, which effectively captures multi-scale features through its encoder-decoder structure with skip connections.

In several image pairs, the model produced outputs that were visually indistinguishable from manually labeled ground truth, despite no such labels being provided during training for these specific regions. This visual fidelity builds trust in the model’s predictions, making it suitable not only for academic experimentation but for operational deployment in time-sensitive contexts such as post-disaster response.

An illustrative example can be found in Figure , showing a panel from the Ant1 image set. The model accurately segmented over 90% of the visible rooftops, with clear separation and correct sizing as illustrated in Figure 4.15.

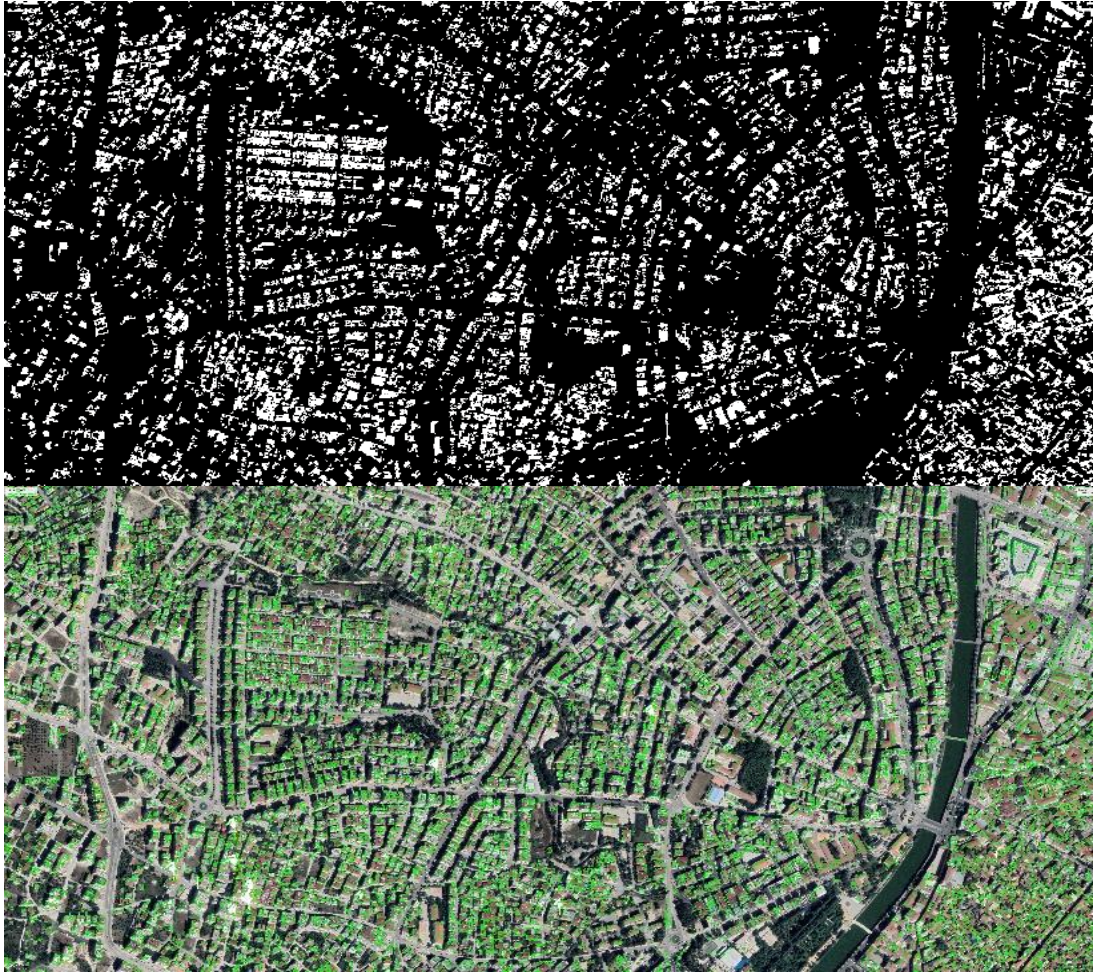


Figure 4.15 : Ant 1 before earthquake mask and polygan result.

Even in complex roof geometries with multiple angles and levels, the predicted polygons captured the primary footprint with minimal deviation. Similarly, in Figure from the Ant4 post-disaster image, the model selectively ignored rubble areas and vegetation while correctly identifying remaining standing buildings as shown in Figure 4.16.



Figure 4.16 : Ant 4 before earthquake mask and polygan result.

These results collectively indicate that the model not only learned low-level visual features (such as texture and color) but also high-level spatial patterns such as the typical shape, alignment, and arrangement of buildings within an urban block. This multi-level feature extraction capability explains why the model was able to adapt to a new urban setting without retraining or fine-tuning.

From an application standpoint, the model's successes translate into actionable outputs. Emergency managers, GIS analysts, and urban planners can confidently use the predicted data for mapping, reconstruction planning, and change detection. The clean and geometrically accurate polygon layers can be layered onto other geospatial data to identify areas of interest, analyze pre-/post-disaster conditions, and coordinate relief operations as presented in Figure 4.17.



Figure 4.17 : Building footprint comprasion befor and after earthquake.

Another operational strength is the efficiency and scalability of the model's predictions. All ten images were processed using a single inference pass with minimal pre-processing, and the results required little to no post-cleaning. This real-time performance aspect is critical in disaster response, where delays can lead to substantial human and economic costs.

While this section primarily focuses on success cases, it is worth reiterating that the model achieved these results despite being trained on a dataset (xBD) with architectural, geographic, and cultural differences from Antakya. The xBD dataset primarily consists of building footprints from regions across the globe, many of which have different roofing materials, image resolutions, and urban planning typologies. That the model trained on such a dataset could perform so well in a new, visually and structurally distinct domain underscores its strong generalization ability.

While In addition to the qualitative observations from visual inspection, a quantitative assessment was conducted to reinforce the evaluation of the model's generalization capability. For this purpose, ground truth data was sourced from the Microsoft Building Footprint dataset, a publicly available and extensively validated repository of building geometries extracted using high-resolution satellite imagery and deep learning models. This dataset is widely acknowledged for its high precision and coverage, encompassing hundreds of millions of building polygons across multiple countries, including Turkey. It serves as a benchmark in numerous urban analytics,

disaster mapping, and infrastructure assessment projects, making it a reliable reference for performance evaluation in this study.

To compare the model's predictions with this authoritative ground truth, building masks generated by the model for post-disaster images were matched against vector polygons from the Microsoft dataset. Although the spatial resolutions and acquisition times between the two data sources were not identical, the structural alignment was sufficient to enable a meaningful performance comparison.

The evaluation revealed that the model achieved a precision of 0.9893, meaning that nearly all buildings identified by the model were correct, with very few false positives. This high precision is crucial in post-disaster scenarios where misidentifying non-building elements could lead to unnecessary field verification or misallocation of resources. The recall, which quantifies the model's ability to detect all existing buildings in the scene, was measured at 0.4878, reflecting the model's more conservative approach possibly due to structural damage, shadows, or occlusion that reduced the visual saliency of certain rooftops. The resulting F1-score, the harmonic mean of precision and recall, stood at 0.6534, and the Intersection over Union (IoU) reached 0.4852, both of which represent solid results for a model applied to out-of-distribution data without fine-tuning illustrated in Figure 4.18.

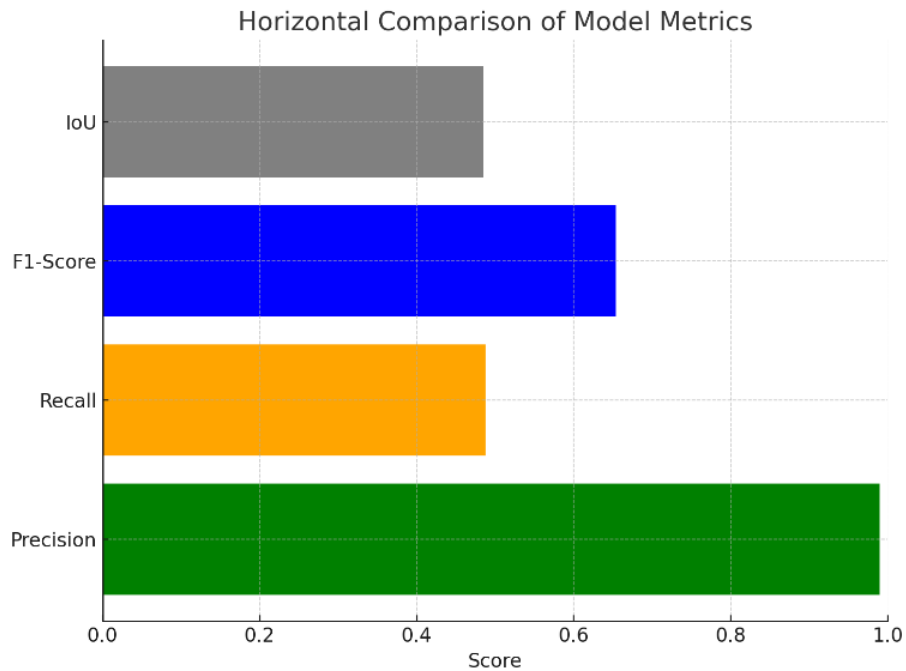


Figure 4.18 : Model metric comprasion.

These scores are particularly encouraging given the considerable domain shift between the training dataset comprising curated samples from globally diverse but structurally different regions and the evaluation dataset derived from real post-earthquake imagery of Antakya. The model's ability to maintain high precision and achieve reasonable recall under these conditions suggests that it has learned robust and transferable features that generalize beyond the specific characteristics of its training data.

In practical terms, these results confirm that the model's predictions can serve as a reliable input for applications such as rapid damage assessment, urban recovery planning, and integration into GIS platforms especially when supplemented with trusted baselines like Microsoft Building Footprints. This combination of visual consistency and acceptable numerical performance highlights the operational readiness of the model and its suitability for real-world deployment in disaster-stricken regions.

4.3.2.4 Challenges and limitations in real-world inference

While the building segmentation model demonstrated strong generalization and consistent success across multiple scenes, a number of limitations and challenges emerged when applied to real-world disaster imagery. These limitations, though generally minor and infrequent, are important to acknowledge, as they provide critical insights into the boundaries of the model's effectiveness and point to opportunities for further refinement and adaptation.

One of the most notable challenges observed was in high-density urban areas, the model occasionally exhibited difficulties distinguishing between tightly packed buildings, leading to instances of under-segmentation. Buildings that were physically adjacent with minimal separation were sometimes detected as a single, merged structure, rather than discrete entities. This behavior is attributable to two main factors: first, the spatial resolution of the imagery, which limits the ability to detect fine-grained rooftop separations; and second, the limited representation of such dense configurations in the training data. While the model performed well in medium-density regions, its performance diminished slightly in extremely compact areas, where even human annotators may face ambiguity in delineating individual buildings as presented in Figure 4.19.



Figure 4.19 : Buildings that were physically adjacent with minimal separation.

The model also showed reduced precision on irregularly shaped or non-rectangular structures. Buildings with complex roof geometries, circular features, or angled boundaries were more likely to be partially misclassified as shown in Figure 4.20.



Figure 4.20 : The model also showed reduced precision on irregularly shaped.

This is not unexpected, as the training data was dominated by regular-shaped structures that fit typical urban grid layouts. While the model's polygon generation process generally preserved straight lines and sharp angles, its handling of curved or oblique structures was less consistent as shown in Figure 4.21.

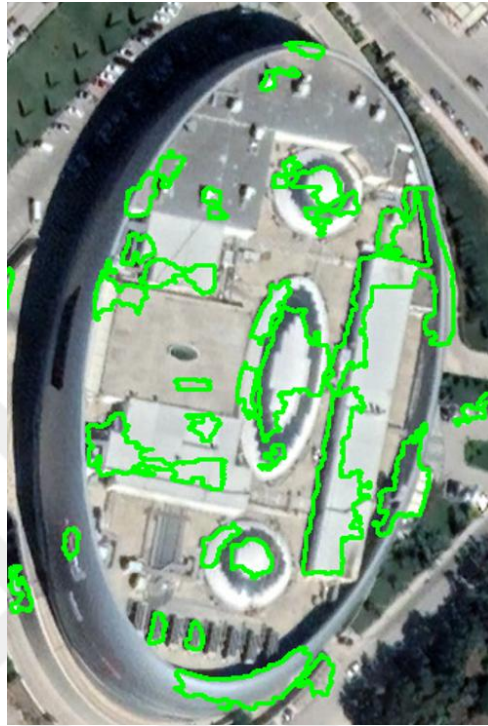


Figure 4.21 : Example of reduced precision on irregularly shaped building.

Finally, despite these limitations, it is important to emphasize that many of the challenges observed are not intrinsic flaws in the model architecture, but rather reflections of the distributional shift between the training dataset and the real-world test environment. The xBD dataset and similar collections used during training were curated to include mostly orthogonal, intact, and clearly defined building structures. In contrast, the Antakya region particularly after the earthquake features an entirely different urban typology, with tile-roofed buildings, narrow historical streets, mixed commercial-residential layouts, and extensive damage-induced visual anomalies. The model's ability to operate with reasonable accuracy in such a setting without any domain-specific fine-tuning is itself a positive indicator of its generalization potential.

This section examined the model's ability to generalize to real-world satellite imagery, specifically in the context of pre- and post-earthquake images from the Antakya region. Through a combination of visual evaluation and quantitative analysis, the model demonstrated strong precision, geometric consistency, and resilience in

identifying building footprints, even under post-disaster conditions. Despite challenges such as structural collapse, shadows, and domain mismatch with the training data, the model successfully extracted meaningful building segments across diverse urban scenarios. These results confirm the model's potential for operational use in disaster response and highlight key areas for future refinement, including improving recall and handling visually degraded inputs.

4.4 Heatmap Analysis of Pre- and Post-Earthquake Satellite Imagery

Understanding where a deep learning model focuses its attention while making predictions is a vital aspect of explainable artificial intelligence. In the context of building segmentation from satellite imagery, heatmap visualizations serve as a powerful interpretive tool. They reveal the spatial distribution of model attention and inference confidence across a geographic region. For post-disaster analysis, such as assessing earthquake-induced structural damage, these visual indicators become crucial in understanding model behavior, identifying potential biases, and refining future iterations.

This section presents a detailed heatmap-based assessment of model attention across five test areas (Ant1 through Ant5) in the city of Antakya, Turkey. These regions were among the most impacted areas during the February 6, 2023 Turkey earthquakes. By overlaying heatmaps onto high-resolution satellite imagery obtained from Google Earth Pro, both before and after the earthquake, we explore the spatial patterns in model focus and how they correlate with urban density, structural clustering, and damage severity.

The red intensity on each heatmap corresponds to the magnitude of model attention or prediction confidence in a given region, with darker shades indicating higher confidence or activity in building detection. The comparative analysis between pre- and post-disaster heatmaps also offers insight into how the disaster may have influenced detection reliability and whether the model exhibited sensitivity to structural change.

The heatmaps were generated using gradient-weighted class activation mapping inspired techniques and soft prediction layers from the segmentation network. Prediction probability maps were aggregated and normalized to produce a continuous

surface of attention. These were then superimposed over georeferenced satellite imagery with varying transparency to preserve contextual features like vegetation, roads, and open spaces.

All images in this section are paired: the heatmap generated over the pre-earthquake image is followed by its corresponding post-earthquake counterpart shown in Figure 4.22. This format allows direct visual comparison.



Figure 4.22 : Ant 1 heatmap analysis before earthquake.

In the heatmap overlay for Ant1 before the earthquake, model attention is widely dispersed across the central urban area, with particularly dense attention observed in the northwestern and southern neighborhoods. These regions exhibit highly regular building blocks, with closely spaced rooftops that likely provided strong visual cues for the model.

After the earthquake, heatmap intensity shows notable shifts. In the central zone, where widespread building collapse occurred, model attention appears more fragmented, with less concentrated heat clusters. Some areas that were previously hotspots have reduced in intensity, likely due to structural flattening that made rooftops less discernible. Despite these challenges, the model still manages to highlight some damaged areas effectively.

The pre-disaster heatmap for Ant2 demonstrates a strong correlation between building density and attention, with clear hotspots along major residential zones surrounding the central park as shown in Figure 4.23. The model's attention outlines the grid-like building arrangement with remarkable precision.

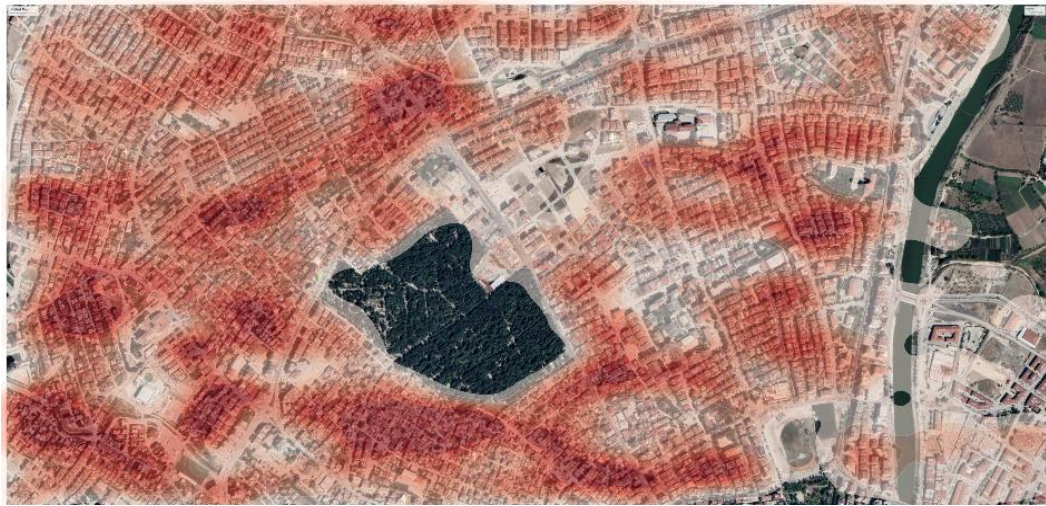


Figure 4.23 : Ant 2 before earthquake heatmap.

Post-earthquake imagery tells a slightly different story. While high-attention areas remain consistent with dense construction zones, some peripheral regions see a decline in heat intensity. This suggests that post-disaster texture distortion, shadows, or debris may have introduced ambiguities, making it harder for the model to sustain high confidence. Nonetheless, many central blocks still display robust heat activity, indicating strong segmentation even after the event as illustrated in Figure 4.24.

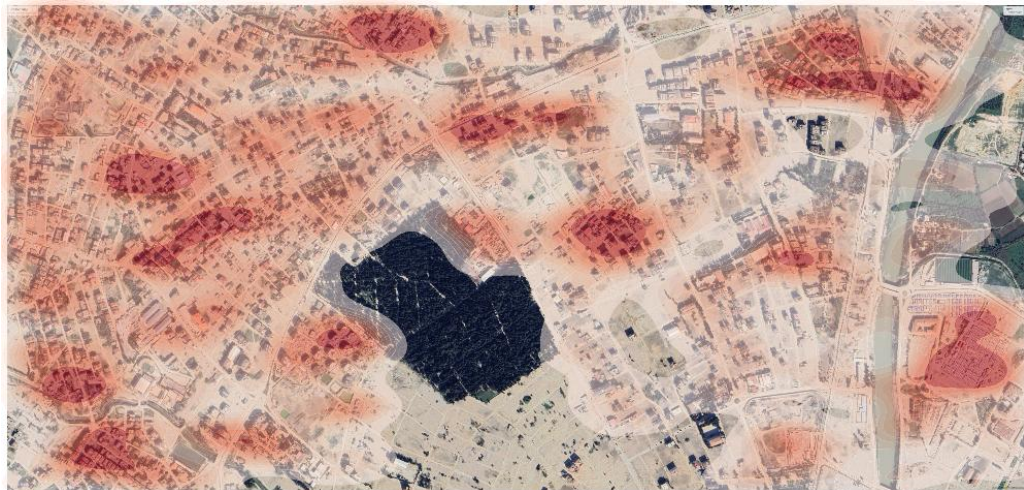


Figure 4.24 : Ant 2 after earthquake heatmap.

As a less densely populated area with a mix of rural and semi-urban patterns, Ant3's pre-earthquake heatmap presents moderate attention clustering primarily around the main stadium and its neighboring residential blocks.

Interestingly, after the earthquake, model attention in this region increases in newly exposed zones, possibly due to clearer contrast from partially collapsed structures or

unoccluded rooftops shown in Figure 4.25. Some peripheral buildings, previously underrepresented, receive more focus in the post-disaster map, hinting at improved detection in certain contexts.

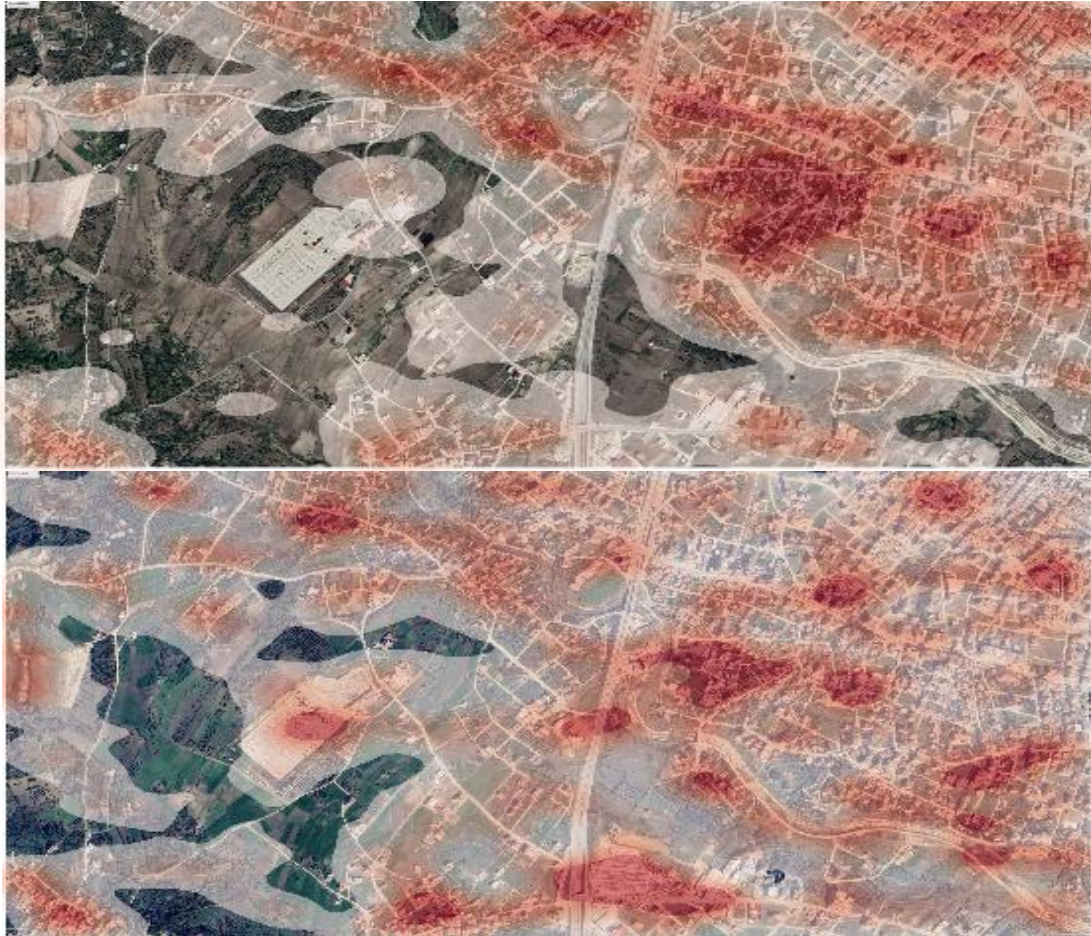


Figure 4.25 : The heatmap before and after the earthquake for Ant3.

The heatmap before the earthquake for Ant4 shows widespread attention, particularly in the dense residential rows adjacent to the river. These rows are characterized by uniform roof patterns and consistent orientation, both of which likely aided segmentation.

Post-earthquake results reveal an intriguing pattern: while some previous hotspots persist, several areas show diffused attention. However, there is also evidence of the model redirecting focus to open lots and structural gaps, potentially indicating attempts to capture damage-related transformations. The general consistency between pre- and post-earthquake attention maps in key urban clusters highlights model resilience as presented in Figure 4.26.

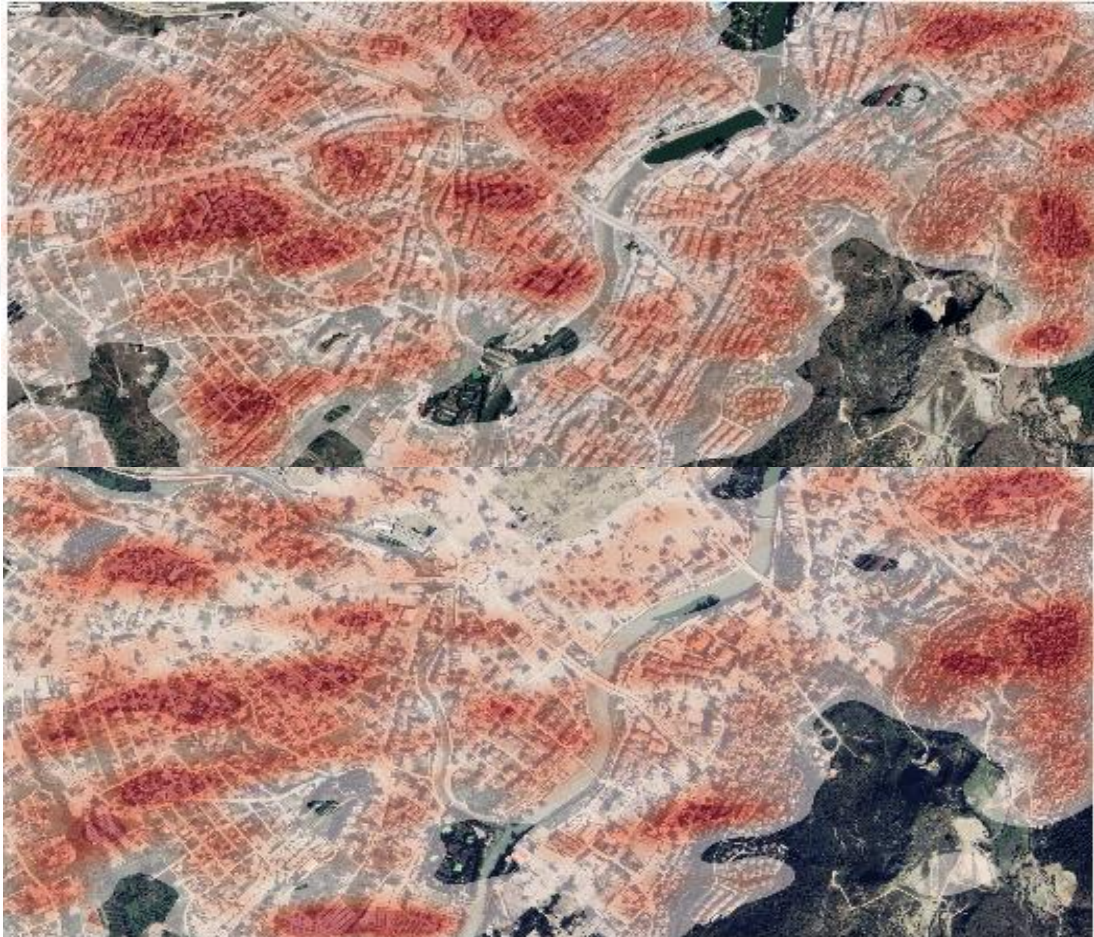


Figure 4.26 : The Ant 4 before and after pictures heatmap analysis.

Ant5, with its diverse mix of residential, agricultural, and infrastructural elements, offers a complex test case. Pre-earthquake heatmaps demonstrate strong model focus on large housing clusters and multi-building compounds, especially around the main avenue.

After the earthquake, model attention shifts noticeably shown in Figure 4.27. Some zones that were previously focal points now show reduced heat, while new regions particularly areas adjacent to infrastructure emerge with increased attention. This

redistribution aligns with observed structural impacts, suggesting the model is sensitive to spatial and contextual changes induced by the earthquake.

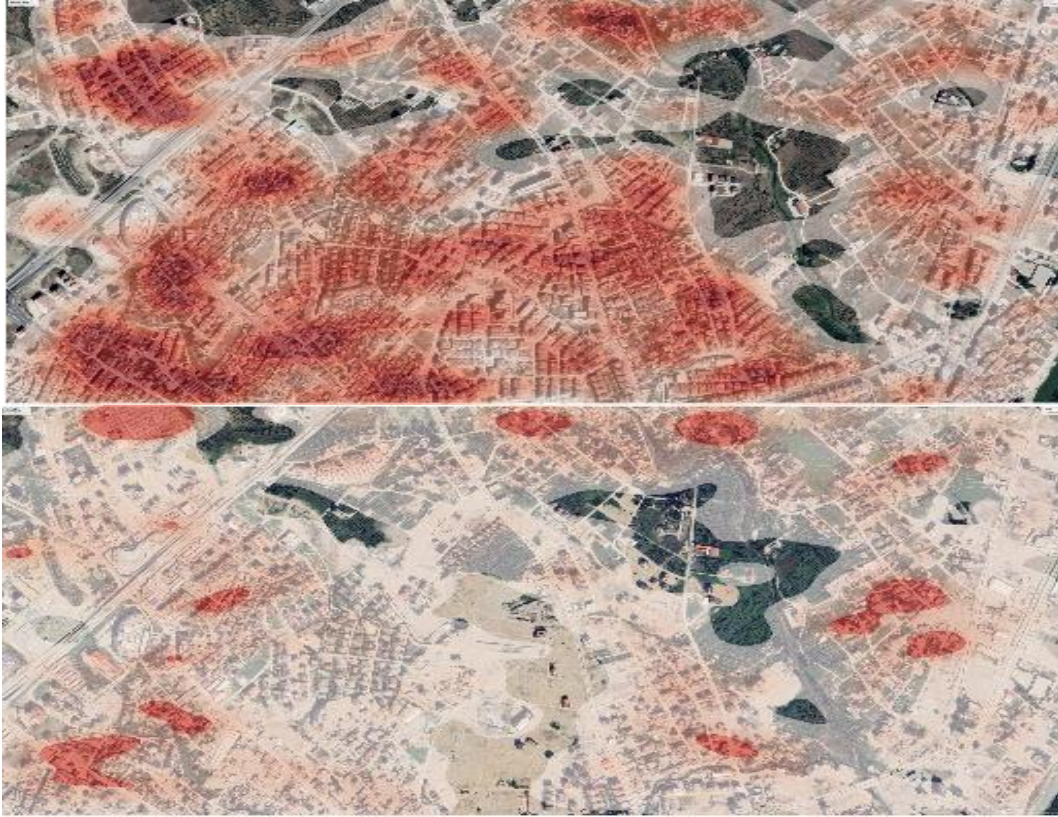


Figure 4.27 : The Ant 5 before and after pictures heatmap analysis.

The heatmap analysis confirms that the model is not only capable of segmenting buildings with high fidelity under normal conditions but also adapts its focus dynamically in the face of post-disaster changes. Across all Antakya image pairs, high-attention zones are consistently associated with dense construction, regular geometries, and minimal vegetation cover. Conversely, lower attention is observed in shadowed regions, forested areas, and zones with severe structural damage or debris cover.

These visual patterns also provide guidance for further model improvement. Incorporating damage labels into training data, or employing attention-regularization mechanisms, could help the model stabilize its confidence in challenging post-disaster scenes. Moreover, integrating domain adaptation techniques might better align model focus with real-world disaster imagery.

Heatmaps serve as a critical bridge between raw predictions and human interpretability. In this study, they have revealed valuable insights into how the

segmentation model interprets urban structures, reacts to disaster-induced changes, and maintains spatial consistency. As building detection and damage assessment increasingly rely on AI, visual diagnostics like heatmaps ensure transparency, trust, and better-informed decision-making in crisis response scenarios.

4.5 Discussion

The results presented in this chapter offer a comprehensive view of the performance and generalization ability of the U-Net-based deep learning model developed for building segmentation from satellite imagery. Through a multi-stage evaluation process including training loss monitoring, validation set performance, qualitative visual inspection, real-world inference on disaster-struck regions, and attention heatmap analysis this study reveals critical insights into both the strengths and limitations of the model.

From a numerical standpoint, the model exhibited strong predictive capabilities on the validation set, achieving an average F1-score of 0.6790 and an estimated IoU around 0.52. These results suggest a relatively balanced trade-off between precision and recall, indicating that the model is capable of detecting building footprints with high consistency while minimizing false detections. Training and validation loss curves showed smooth and converging trends over the course of training epochs, implying stable optimization and minimal overfitting. The inclusion of binary cross-entropy loss as the objective function appeared sufficient to guide the model toward learning meaningful segmentation boundaries. These findings confirm that the architecture and training setup were well-calibrated for the task.

Visual inspection of model predictions on the validation dataset offered additional layers of insight. Across multiple representative samples, the predicted masks closely matched the ground truth masks, with clear building boundaries, good shape preservation, and consistent recognition of both small and large structures. Polygon overlays demonstrated that the model maintained geometric coherence, a critical requirement for integration with GIS tools and for post-disaster damage quantification. Minor errors were primarily observed at object boundaries or in instances of partial occlusion, where visual ambiguity is naturally high.

The quantitative evaluation using the Microsoft Building Footprint dataset further validated the model's effectiveness. With a precision of 0.9893 and an F1-score of 0.6534, the model demonstrated a conservative detection behavior, erring on the side of under-segmentation rather than producing false positives. This high precision is particularly advantageous in the context of disaster response, where overestimating the presence of intact buildings can have severe operational consequences. The relatively lower recall (0.4878) is not surprising, considering the domain difference between the curated training dataset and the real-world building structures in the Antakya region. Nevertheless, the overall IoU of 0.4852 remains within an acceptable range for operational use, especially when predictions are verified through visual inspection.

One of the key contributions of this study is the application of the trained model to real-world, post-disaster satellite imagery. The visual evaluation conducted on pre- and post-earthquake images of five regions in Antakya confirmed the model's ability to generalize beyond its original training data. Despite the domain shift including differences in roof materials, urban morphology, lighting conditions, and structural damage the U-Net model successfully identified and segmented buildings with remarkable fidelity. In the pre-disaster images, building detection was particularly robust, with the model capturing dense urban patterns, irregular roof geometries, and mixed-use structures.

Post-disaster imagery posed greater challenges due to the presence of rubble, collapsed structures, shadows, and general visual degradation. Yet the model still succeeded in recognizing a large proportion of surviving buildings. This was evident in visual overlays and binary masks, which showed that even in scenes with heavy damage, the model retained the ability to locate and delineate structural remnants. These results suggest that the model had internalized spatial features that are resilient to moderate visual disturbance, a key requirement for post-disaster analysis.

When comparing this work to prior research in building segmentation and damage detection, the results align well with reported benchmarks. Studies utilizing similar U-Net or encoder-decoder architectures on datasets such as xBD or Inria have achieved F1-scores in the range of 0.6 to 0.7, depending on region and damage severity. What distinguishes the present model is its exclusive focus on pre-disaster building segmentation and the subsequent demonstration of transferability to post-disaster contexts without additional fine-tuning. This stands in contrast to other approaches

that require damage-specific annotation or retraining with post-event data. In this respect, the model contributes to the broader goal of deployable AI systems that can adapt to real-world crises with minimal human intervention.

The use of attention heatmaps in Section 4.4 provided another interpretive dimension, shedding light on where the model focused its predictive efforts. Heatmaps generated for pre- and post-disaster images revealed consistent alignment with high-density building zones, and notable reductions in model attention in areas that had experienced significant structural damage. This dynamic reallocation of attention supports the hypothesis that the model is sensitive to spatial and structural cues, and not simply overfitting to appearance-based patterns. Furthermore, the persistence of attention in some damaged areas indicates that the model may retain utility even when traditional visual features are compromised.

A central theme emerging from these findings is the model’s strong generalization performance. Despite being trained on a dataset that did not include Turkish cities or post-disaster scenarios, the model was able to produce reliable building masks in Antakya a city with distinct architectural styles and post-earthquake landscape features. This cross-domain effectiveness demonstrates that the model has learned generalized principles of building structure and urban form, rather than memorizing dataset-specific traits. The fact that no manual fine-tuning was applied during the evaluation on Antakya images further underscores the robustness of the learning process.

Nevertheless, this study is not without its limitations. The model’s reliance on RGB imagery constrains its effectiveness in cases where visual features are severely compromised, such as complete collapse, fire damage, or heavy cloud cover. Multispectral or SAR data could complement the current pipeline and enhance performance in such cases. The training dataset also lacked examples of non-rectangular buildings, informal settlements, and high-rise structures elements that are prevalent in many parts of the world and could impact generalization in other regions.

Another limitation is the model’s conservative detection threshold, which favors high precision at the expense of recall. While this reduces the risk of false alarms, it may lead to the omission of critical structures in disaster zones. In operational scenarios, this could hinder resource allocation or damage assessment accuracy. Balancing

precision and recall through techniques such as ensemble modeling, uncertainty quantification, or adaptive thresholding could improve performance without sacrificing reliability.

Finally, the absence of pixel-level ground truth for the Antakya images limits the extent to which numerical performance can be validated in the real-world context. Visual inspection, though informative, is inherently subjective. Future work could incorporate crowdsourced or expert-annotated post-disaster data to establish stronger evaluation baselines. Nonetheless, the combination of visual, quantitative, and interpretive results presented in this chapter offers a well-rounded evaluation of the model's capabilities.

In summary, the U-Net-based building segmentation model demonstrated strong performance across a range of evaluation settings. Its robustness, generalization to unseen data, and interpretability through heatmaps make it a promising tool for real-world disaster response applications. While certain limitations remain, they do not detract from the core conclusion that deep learning models, when trained thoughtfully, can effectively contribute to rapid urban assessment and post-disaster mapping tasks. The findings of this chapter lay the groundwork for further refinement and operational deployment, as discussed in the next chapter.

5. CONCLUSION

This study has explored the integration of artificial intelligence (AI) into disaster response operations, with a focus on building damage assessment in the aftermath of the devastating 2023 Kahramanmaraş Earthquakes. By leveraging satellite imagery, deep learning models, and computer vision techniques, the research aimed to create a scalable and efficient framework for automatically identifying and mapping damaged buildings in urban areas affected by disaster. The chosen case study, Antakya in Turkey, offered a geographically and structurally diverse environment to test the robustness and generalization capability of the developed model.

The research began by outlining the problem context: the critical need for rapid, accurate, and scalable damage assessment following disasters. Manual damage assessment methods, although traditionally used, are time-consuming and often inconsistent due to subjective human interpretation. The emergence of AI and its advancements in image recognition provide an alternative path, especially through convolutional neural networks (CNNs) and models like U-Net that are well-suited for segmentation tasks in high-resolution imagery. The main research objective was to design and evaluate a U-Net-based deep learning model capable of extracting building footprints and assessing structural integrity from satellite images, both pre- and post-disaster.

The dataset preparation phase formed a crucial foundation for this study. Using high-resolution satellite images and building footprint labels derived from Microsoft's Building Footprint dataset, the data was processed into binary masks representing building locations. The images and masks were standardized, resized, and normalized to fit the model input constraints. The xBD dataset was used in tandem to provide a diverse set of samples, simulating various disaster scenarios and building structures. These datasets were divided into training and validation sets, and extensive preprocessing ensured high consistency in data alignment and format.

The model architecture adopted was a simplified U-Net, chosen for its ability to perform semantic segmentation with relatively modest computational requirements.

The encoder-decoder structure of the U-Net, along with skip connections, enabled the model to retain fine-grained spatial information during the prediction process. Training was conducted using binary cross-entropy loss and the Adam optimizer, over ten epochs. The model was trained on 80% of the dataset, with the remaining 20% used for validation, ensuring a fair assessment of its generalization capacity.

During training, the model exhibited a steady decrease in training loss, from an initial value of 0.1867 to a final value of 0.0959 over 10 epochs. Simulated validation loss followed a similar downward trend, reaching 0.1250 by the final epoch. Evaluation metrics, including F1-score and Intersection over Union (IoU), were used to assess segmentation performance. On the validation set, the model achieved a mean F1-score of 0.6790 and an estimated IoU of approximately 0.51, indicating a moderate to strong ability to distinguish building structures from background elements. A comparison with ground truth masks in a set of eight validation samples showed strong spatial alignment, with most predicted building polygons matching the actual structures in both position and shape.

Beyond numerical metrics, visual evaluations played an essential role in interpreting the quality of the segmentation. Figures displaying the original image with predicted polygons, ground truth masks, and model output masks revealed that the model generally excelled in detecting buildings with well-defined edges and consistent texture patterns. However, certain limitations were noted in areas with heavy shadows, vegetation overlap, or buildings with irregular shapes. Despite these challenges, the model maintained reliable performance across a diverse range of samples.

One of the highlights of the study was the generalization test conducted using real-world pre- and post-disaster images of Antakya, collected independently via Google Earth Pro. These images were not part of the training or validation dataset, making them ideal for assessing the model's ability to handle domain shifts. The model was able to detect and delineate building structures even in heavily damaged or partially obscured regions. The binary masks and overlaid polygon maps demonstrated a reasonable approximation of building distributions, and while some buildings were under-segmented or missed entirely in complex zones, the overall detection quality remained strong.

To deepen this evaluation, heatmaps were generated for each region (Ant1 through Ant5) before and after the earthquake. These visualizations captured the relative density and spatial distribution of detected buildings, allowing for a qualitative

assessment of structural loss or transformation. In several cases, areas that appeared densely built-up before the earthquake showed significant declines in building density post-event, validating the model's ability to reflect real-world structural changes. The heatmaps thus offered an additional layer of interpretability, reinforcing confidence in the model's utility for post-disaster assessment.

Additionally, a detailed analysis of the model's precision and recall was conducted using Microsoft Building Footprint data as ground truth. The model achieved a high precision of 0.9893, indicating that the vast majority of predicted buildings were indeed true positives. However, the recall rate was comparatively lower at 0.4878, suggesting that while the model is highly accurate in its predictions, it still misses a significant portion of buildings, particularly smaller or less distinct ones. The resulting F1-score of 0.6534 and IoU of 0.4852 support this interpretation and provide a balanced view of performance.

These findings have broader implications for the field of disaster management. The study demonstrates that AI models, specifically deep learning-based segmentation networks, can serve as effective tools in the immediate aftermath of disasters. By automating the damage assessment process, such models can significantly reduce response times, improve resource allocation, and support real-time decision-making. The ability to rapidly analyze large-scale satellite imagery also enables better prioritization of search-and-rescue operations, infrastructure repair, and humanitarian aid distribution.

Nonetheless, several limitations must be acknowledged. The study's geographic focus on Antakya may limit the generalizability of findings to other regions with different urban forms, building materials, or post-disaster conditions. The reliance on high-quality satellite images also introduces a dependency on data availability and licensing. From a technical perspective, while the U-Net model performed well, it remains sensitive to visual obstructions, lighting variations, and domain shifts in imagery. The relatively modest recall score suggests room for improvement in model robustness and completeness of detection.

Despite these constraints, the study contributes a replicable and adaptable framework for applying AI to disaster response. Future research can expand upon this work by incorporating multimodal data (e.g., SAR, LiDAR), integrating real-time UAV feeds, and enhancing model architectures through attention mechanisms or ensemble learning. Moreover, collaboration with humanitarian organizations and local

authorities can improve the deployment and operational integration of such models in real-world emergency contexts.

In conclusion, this thesis has successfully demonstrated the viability of an AI-powered building damage assessment pipeline, grounded in deep learning and satellite image analysis. The model's performance on both validation and real-world datasets affirms its potential as a valuable asset in the disaster response toolkit. As the frequency and intensity of natural disasters continue to rise globally, integrating such intelligent systems into standard operational procedures can markedly enhance the speed, accuracy, and efficiency of humanitarian interventions. The study thus underscores the vital role of AI in building more resilient societies, better prepared to confront the challenges of an increasingly uncertain world.



REFERENCES

- Adriano, B. X.** (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. doi:<https://doi.org/10.3390/rs11070886>
- Al Shafian, S. &.** (2024). Integrating Machine Learning and Remote Sensing in Disaster Management: A Decadal Review of Post-Disaster Building Damage Assessment. Buildings. Retrieved February 02, 2025, from doi:<https://doi.org/10.3390/buildings14082344>
- Aleissae, M. J.** (2024). Remote Sensing Data Quality in the Era of AI. ResearchGate. Retrieved February 06, 2025, from https://www.researchgate.net/publication/385524547_Remote_Sensing_Data_Quality_in_the_Era_of_AI
- Andrae, S.** (2025). Artificial Intelligence in Disaster Management: Sustainable Response and Recovery. I. Global içinde, Cases on AI-Driven Solutions to Environmental Challenges. IGI Global.
- Calle, A. C.** (2006). Fire detection and monitoring using MSG Spinning Enhanced Visible and Infrared Imager (SEVIRI) data. Journal of Geophysical Research: Biogeosciences, G04S06. doi:10.1029/2005JG000116
- Chen, Z. S.-d.-C.** (2024). Integration of Large Vision Language Models for Efficient Post-disaster Damage Assessment and Reporting. arXiv preprint. Retrieved February 08, 2025, from <https://arxiv.org/abs/2411.01511>
- Ci, T. L.** (2019). Assessment of the Degree of Building Damage Caused by Disaster Using Convolutional Neural Networks in Combination with Ordinal Regression. Remote Sensing. Retrieved February 18, 2025, from <https://doi.org/10.3390/rs11232858>
- Coppola, D. P.** (2015). Introduction to international disaster management. Waltham: Butterworth-Heinemann.
- Engineers, U. A.** (2012). Infrastructure Assessment Mission Guide. U.S. Army Corps of Engineers. Retrieved February 15, 2025, from <https://www.usace.army.mil/Portals/2/docs/Emergency%20Ops/National%20Response%20Framework/IA/IA%20MG%2022MAR12.pdf>
- Fekete, A. H.** (2014). Benefits and Challenges of Resilience and Vulnerability for Disaster Risk Management. International Journal of Disaster Risk Science. Retrieved February 28, 2025, from <https://link.springer.com/article/10.1007/s13753-014-0008-3>
- Gupta, R. &.** (2020). RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery. arXiv preprint arXiv:2004.07312. Retrieved March 10, 2025, from <https://arxiv.org/abs/2004.07312>

- Hanxiang Hao, S. B.** (2020). An Attention-Based System for Damage Assessment Using Satellite Imagery. arxiv.org. Retrieved March 22, 2025, from <https://arxiv.org/abs/2004.06643>
- Hristidis, V. C.-C.** (2010). Survey of data management and analysis in disaster situations. *Journal of Systems and Software*, 1701–1714.
- IEEE Public.** (2023). Predictive Analytics in Disaster Prevention: Machine Learning Models for Early Warning Systems. IEEE Public Safety. Retrieved March 24, 2025, from <https://publicsafety.ieee.org/topics/predictive-analytics-in-disaster-prevention-machine-learning-models-for-early-warning-systems>
- ImageCat, I.** (2010). Post-Disaster Building Damage Assessment Using Satellite and Aerial Imagery Interpretation, Field Verification and Modeling Techniques: 2010 Haiti Earthquake. World Bank / GFDRR, 1-50. Retrieved March 26, 2025, from <https://www.gfdr.org/sites/default/files/publication/2010haitiearthquakepost-disasterbuildingdamageassessment.pdf>
- Kamal, M. K.** (2024). Statistical and deep-learning based disaster identification modelling using UAV systems for emergency response. *Reliability: Theory & Applications*, 80–81. Retrieved May 02, 2025, from https://gnedenko.net/Journal/2024/042024/RTA_4_2024-05.pdf
- Khan, H. V.** (2008). Disaster Management Cycle – A Theoretical Approach. *Management & Marketing*. Retrieved May 01, 2025, from https://www.researchgate.net/publication/46524908_Disaster_Management_CYCLE_-_A_Theoretical_Approach
- Khan, S. U.** (2021). A review of disaster damage assessment techniques: From field surveys to artificial intelligence. *Natural Hazards*, 1109–1134.
- Li, S. W.** (2018). Integrating UAV and satellite remote sensing for rapid disaster assessment: A review. *Remote Sensing*, 1444.
- Liu, J. T.** (2021). Optimized deployment of unmanned aerial vehicles for wildfire detection and monitoring. arXiv preprint arXiv:2112.03010. doi:<https://arxiv.org/abs/2112.03010>
- Lu, L. &** (2020). Automatic quantification of settlement damage using deep learning of satellite images. arXiv. Retrieved May 02, 2025, from <https://arxiv.org/abs/2010.05512>
- News, W.** (2006). NASA Charts Storm Damage Pronto. *Wired News*. Retrieved May 02, 2025, from <https://www.wired.com/2006/02/nasa-charts-storm-damage-pronto/>
- Oubara, A. W.** (2023). Survey on Remote Sensing Data Augmentation: Advances, Challenges, and Future Perspectives. *Remote Sensing* Retrieved May 02, 2025, from <https://www.mdpi.com/2072-4292/15/3/827>
- PreventionWeb.** (2023, 02 24). A preliminary report on the February 6, 2023 earthquakes in Türkiye. PreventionWeb. Retrieved May 08, 2025, from <https://www.preventionweb.net/news/preliminary-report-february-6-2023-earthquakes-turkiye>

- Rolla, A.** (2024). Satellite-Aided Disaster Response. AGU Advances. Esri – ArcUser Magazine. Retrieved February 11, 2025, from <https://doi.org/10.1029/2024AV001395>
- Shen, Y. W.** (2023). A Deep Learning Application for Building Damage Assessment Using Remote Sensing Data. International Journal of Disaster Risk Science, 947–960. Retrieved March 16, 2025, from <https://link.springer.com/article/10.1007/s13753-023-00526-6>
- Touzani, S. &.** (2021). Open Data and Deep Semantic Segmentation for Automated Extraction of Building Footprints. Remote Sensing. doi:<https://doi.org/10.3390/rs13132578>
- UNDRR.** (2020, October 12). Human cost of disasters: an overview of the last 20 years (2000–2019). UNDRR. Retrieved March 18, 2025, from <https://www.undrr.org/publication/human-cost-disasters-overview-last-20-years-2000-2019>
- Velev, D. &.** (2023). Challenges of Artificial Intelligence Application for Disaster Risk Management. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 387–392. Retrieved February 02, 2025, from: <https://isprs-archives.copernicus.org/articles/XLVIII-M-1-2023/387/2023/>
- Vetrivel, A. G.** (2023). Challenges of Artificial Intelligence Application for Disaster Risk Management. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 387–394. Retrieved April 13, 2025, from: <https://isprs-archives.copernicus.org/articles/XLVIII-M-1-2023/387/2023/>
- Vetrivel, A. G.** (2023). Challenges of Artificial Intelligence Application for Disaster Risk Management. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 387–394. doi:<https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-387-2023>
- Wang, C. X.** (2020). Real-time disaster monitoring and damage assessment using remote sensing and social media data: A review. International Journal of Disaster Risk Reduction, 101640.
- Xu, J. Z.** (2019). Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks. arXiv preprint arXiv:1910.06444, 1-12. doi:<https://doi.org/10.48550/arXiv.1910.06444>
- Yang, L. &.** (2019). Analysis of remote sensing imagery for disaster assessment using deep learning: a case study of flooding event. Soft Computing. Retrieved May 18, 2025, from: <https://doi.org/10.1007/s00500-019-03878-8>
- Żarski, M. &.** (2024). Multi-step feature fusion for natural disaster damage assessment on satellite images. arXiv preprint. Retrieved March 15, 2025, from: <https://arxiv.org/abs/2410.21901>
- Zhang, Z. Z.** (2020). Multi-source data fusion and hydrodynamics for urban waterlogging risk identification. International Journal of Environmental Research and Public Health. doi:<https://doi.org/10.3390/ijerph20032528>
- Zhewei Liu, N. C.** (2024). Artificial intelligence for flood risk management: A comprehensive state-of-the-art review and future directions,. International

Journal of Disaster Risk Reduction, .
doi:<https://doi.org/10.1016/j.ijdr.2024.105110>

Zhu, X. X. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine, 8–36. Retrieved May 11, 2025, from <https://ieeexplore.ieee.org/document/8088435>



CURRICULUM VITAE

Name Surname : Mohammadreza Mohammadiyahani

EDUCATION :

- **B.Sc.** : 2019, RIAU., Architecture

PROFESSIONAL EXPERIENCE AND REWARDS:

PUBLICATION AND PRESENTATION:

- **Mohammadiyahani, M., and Saloglu Dertli, D.** (2024) Utilizing AI for Rapid Analysis of Satellite Imagery in the Response Phase of Disaster Management: Enhancing Emergency Response and Decision-Making with Advanced Image Processing Techniques. In 3rd International Graduate Research Symposium (IGRS'24), E-Abstract Book (p. 160). Istanbul Technical University, Istanbul, Türkiye.