

# **Automated Classification of House Damage in Earthquake-Prone Areas Using AI-Driven Image and Data Analysis**

**Abdulrahman Bazbouz**  
**Supervisor, Nurullah Bektas**

## **Abstract**

Recent earthquakes, such as the 2023 Kahramanmaraş earthquake in Turkey and the Syrian earthquake, have highlighted the importance of assessing damage to buildings after major seismic events. Building collapses during earthquakes can result in fatalities and significant financial losses, particularly in areas where many structures were built before modern seismic regulations were implemented. This project aims to create AI models where the model can be used to identify potential faults and determine failure modes. It can be used as a support tool for engineers or practitioners. This project will be done using AI methods and data analysis techniques. The dataset includes both categorical data and images. The first step will be data analysis, getting useful information from the data, handling the wrong values and dropping the useless features in it, and then trying to build conclusions based on the figures and the histograms we get from the data to understand the data's nature. The next stage will be using AI models to predict the probability of house damage. We will work with deep learning models. The third stage is to integrate these models to create an effective tool that can take image input to predict the extent of damage to a house. Finally, we will evaluate the models we got using metrics such as accuracy, precision, recall, and F1 score to ensure that the predictions are reliable. We will perform Cross-validation and testing on unseen data to measure the model's performance in real-world scenarios.

**Keywords:** Earthquakes; Image Processing; Building Vulnerability; Disaster Recovery.

## **1. Introduction**

Natural disasters, earthquakes in particular, is one of the major perils to human life, infrastructure, and economies. Seismic events can cause building collapses, which may lead to loss of life, injuries, and financial losses, especially when these buildings were constructed prior to modern seismic safety standards. The unprecedented growth of urban populations is placing an urgent demand for effective methodologies for post-disaster damage assessment. Rapid Visual Screening methodologies have been crucial in the rapid assessment of the structural integrity of the building stock immediately after an earthquake [4]. These methods give the engineers and practitioners the capability to identify and categorize the potential structural vulnerabilities of the structures, thus enabling timely decisions regarding resource allocation in disaster recovery. RVS methods can be divided into two main categories: pre-earthquake and post-earthquake approaches.

Pre-earthquake RVS considers addressing the problems of seismic building vulnerability assessments in advance of any seismic event, so that risk mitigation strategies could be developed and put into practice. In contrast, post-earthquake RVS methods will assess damage and prioritize safety inspections post-seismic event, such that resources can be directed to the most affected structures. In the last couple of years, the inclusion of AI algorithms in RVS methodologies has brought a sea change in building vulnerability estimation. The improved methodologies of RVS using AI make use of data analytics and machine learning to perform even better in damage assessments. Such algorithms make use of computer vision and image processing for the analysis of post-event imagery regarding the identification and classification of the damage states in buildings [2]. Therefore, this technological advancement not only accelerates disaster assessment activities but also increases the credibility of the information obtained, which is very important in disastrous responses.

That was an important investigation since it corresponded to the urgent need of defining buildings damage states with some degree of reliability: sometimes, after earthquakes, on-site investigations it's hard to discern different states; it is a common scenario when the manifestation of damage is complex. Therefore, there is a need for the development of AI-assisted tools that will, in turn, allow efficient decision-making by engineers and other specialists on the issue of disaster recovery. For example, it can be used to quickly and accurately determine damage to buildings without requiring extensive specialization so that it is easy to assess damage and help in better safety and recovery after seismic events.

## **2. Site Seismicity**

Seismicity provides insight into the understanding of a region with respect to assessing the impact that earthquakes may have and the planning of appropriate disaster preparedness and response strategies. Seismicity describes the frequency, type, and size of earthquakes that have occurred in an area during a particular period of time. This would provide the necessary input toward assessing seismic hazards and risks associated with structures that lie within an earthquake-prone zone. The most current case that comes to mind is that of the earthquake in Haiti in 2021, which happened on August 14, 2021. This earthquake was of magnitude 7.2, hence so devastating throughout the southern part of the island. It came with enormous structural damage, claimed several lives, and left several thousand of the inhabitants displaced. The need for an operative methodology of assessment and recovery became particularly urgent in dealing with a vulnerable seismic region after this event [5].

Seismic hazard maps are the representation that generally deals with the regions known for earthquakes to help out in seismic risk analysis effectively. Those maps divide the areas into the level of expected ground shaking based on the historical seismicity, geology, and seismic sources. The seismic hazard map of Haiti shows that the southern peninsula, hosting the capital city affected by the earthquake, is located within a high seismic hazard zone, requiring stringent building codes and fast assessment methodologies with a view to mitigating the risks due to future seismic events [6]. Such maps are incredibly important tools for engineers, urban planners, and policy makers, providing them with the vital information they need on the possible intensity of shaking that the buildings will experience during an earthquake. Through

the use of seismic hazard data integrated into building vulnerability assessments, stakeholders can prioritize areas for detailed inspection and targeted repairs and retrofitting, enhancing the ability of communities to resist future earthquakes.

### **3. Building Characteristics**

#### **a. Dominant Building Typologies**

Building stock in study areas generally comprises low-rise buildings, usually masonry and RC structures of 1 to 3 stories. The majority of the structures are single-family residential, although there are also commercial, religious, medical, government, and educational buildings. In most of these construction practices, one of the four major typologies predominates: unreinforced masonry (URM), reinforced masonry, confined masonry, and masonry-infilled RC frames. However, due to the limited resources for structural engineering, many buildings exhibit hybrid characteristics. For example, while confined masonry designs are common, they often lack sufficient reinforcement in columns and beams, impacting their resilience in seismic events.

Challenges such as poor material quality, limited budgets, and restricted access to quality labor feed into these hybrid typologies. This is evident in the variable reinforcement quality in masonry walls, often assembled without optimal tie columns, beam placement, or proper grouting. Concrete, normally hand mixed and hand placed and hand pressed CMUs introduce additional variability in building quality. The high cost of, and limited availability of, imported steel further limits the use of proper reinforcements while alternative construction materials such as structural steel or metal systems are rare.

Similarly, quality lumber is scarce, particularly affecting rural and low-income constructions.

#### **b. Foundation and Roofing Systems**

Because of the poor-quality soil and high water tables of the area, most low-rise buildings are founded on shallow foundations, often stone masonry slabs with grade beams to provide stability in coastal areas and areas subject to flooding. Many buildings are also elevated slightly (~0.5 m) to mitigate flood risks in rainy seasons.

Two roofing systems prevail: (1) corrugated galvanized iron (CGI) sheets supported by wood or tack-welded steel hollow structural sections (HSS), and (2) concrete slab roofs. While CGI roofs are more common in low-income residential settings, concrete slab roofs are used in urban areas or in better-off households since they allow for future vertical expansions.

**Rural and Vernacular Architecture** Many rural buildings incorporate a mix of CMU-load bearing walls with vernacular designs, including those using wattle and daub construction. Vernacular designs often have wood frames that are infilled with materials locally available, such as stone bound in place by lime and earth mixtures. Such vernacular designs-even

historical wooden "Gingerbread" houses-have done well during past earthquake events because of the lesser seismic mass.

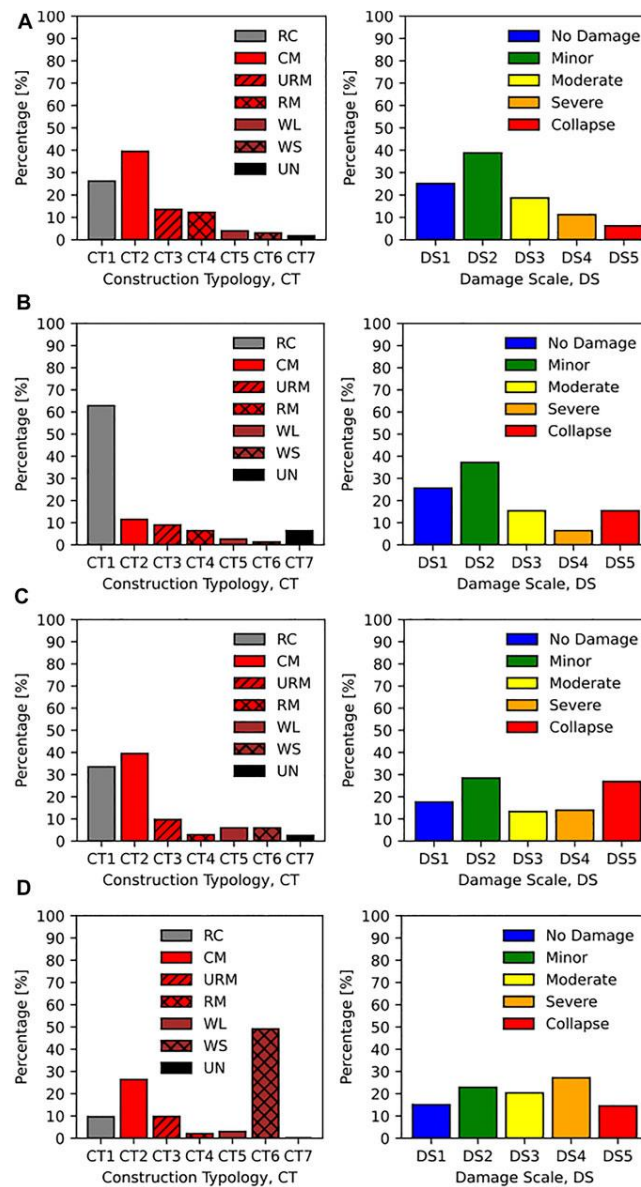
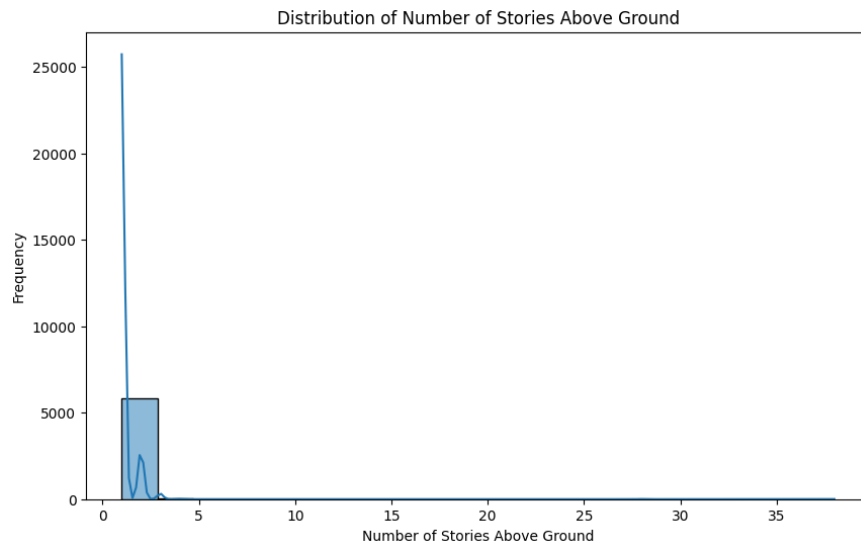


Figure 1. (Left column) construction typology and (right column) overall damage distribution for (A) school, (B) hospitals, (C) churches and (D) residential buildings.

## 4. Post-Earthquake Building Screening Data

### a. Input Parameters: Image Data

In this study, the primary input parameters utilized in the damage classification model are images captured after a disaster—specifically, images of buildings after an earthquake. The dataset consists of 14,920 training images and 3,731 validation images, which are divided into five classes according to the degree of damage: 'Minor,' 'Moderate,' 'Severe,' 'Partial or Total Collapse,' and 'No Visible Damage.'

Each image is in RGB format, normalized to a resolution of 128x128 pixels for all images to have the same size within the dataset. Before passing them to the model, the images are preprocessed: resized to that size and normalized to have pixel values within the [0, 1] range, as this is important for practical neural network training.

Data augmentation is applied to increase the diversity of the training set by means of random rotations and horizontal flips. The proposed method is very effective in consideration of potential variability in real-world scenarios regarding angles and illumination conditions under which images are captured. Careful selection and preprocessing of image data are very important to achieve better performance of the model, since they directly enter the model to determine the accuracy in damage classification. Preliminary results show that integrating diverse images, which are well-preprocessed, significantly improves generalizability over unseen examples. In this study, the primary source of input parameters for the damage classification model is images of buildings damaged due to disaster, more precisely due to earthquakes. The dataset contains 14,920 training images and 3,731 validation images where damage in each image is labeled as one of five classes based on the damage level witnessed:

**No Visible Damage (Grade 1):** This is the class of minimum damage, characterized by hairline cracks in masonry walls, sprinkles of small pieces of plaster falling, and loose stones from some structures. In the case of reinforced concrete buildings, this grade indicates merely fine cracks in plaster, walls, columns, beams, and partitions. The pictures showing this class depict undamaged structures except for slight imperfections.



No Visible Damage

**Minor (Grade 2):** This category corresponds to moderate non-structural damage but slight structural damage. For masonry buildings, this includes widespread wall cracking, partial chimney collapses and larger plaster pieces falling off the walls. For the reinforced concrete building, this represents cracking of columns, beams and walls as well as brittle cladding and falling plaster. The photos in this category represent buildings where damage is visible but can easily be dealt with.



Minor

**Moderate (Grade 3):** This damage grade indicates moderate structural and severe non-structural damage. Images of masonry buildings in this category may be exemplified by widespread cracking in walls, fallen-off roof tiles, shattered chimneys, and failures of single non-structural elements. In a building with a reinforced concrete structure, there are commonly large cracks in columns and beam-column joints and spalling of concrete covers. The images show more serious dilapidation compared with the previous grades.



Moderate

**Severe (Grade 4):** This class represents heavy structural and very heavy non-structural damage, including images showing serious wall failures, partial structural failures of roofs, and tilting of columns in buildings with reinforced concrete. Compression failures, fractured rebars, and the collapse of upper floors may also be noted. The severity of the damage in these images is significantly pronounced, showing buildings likely to collapse.

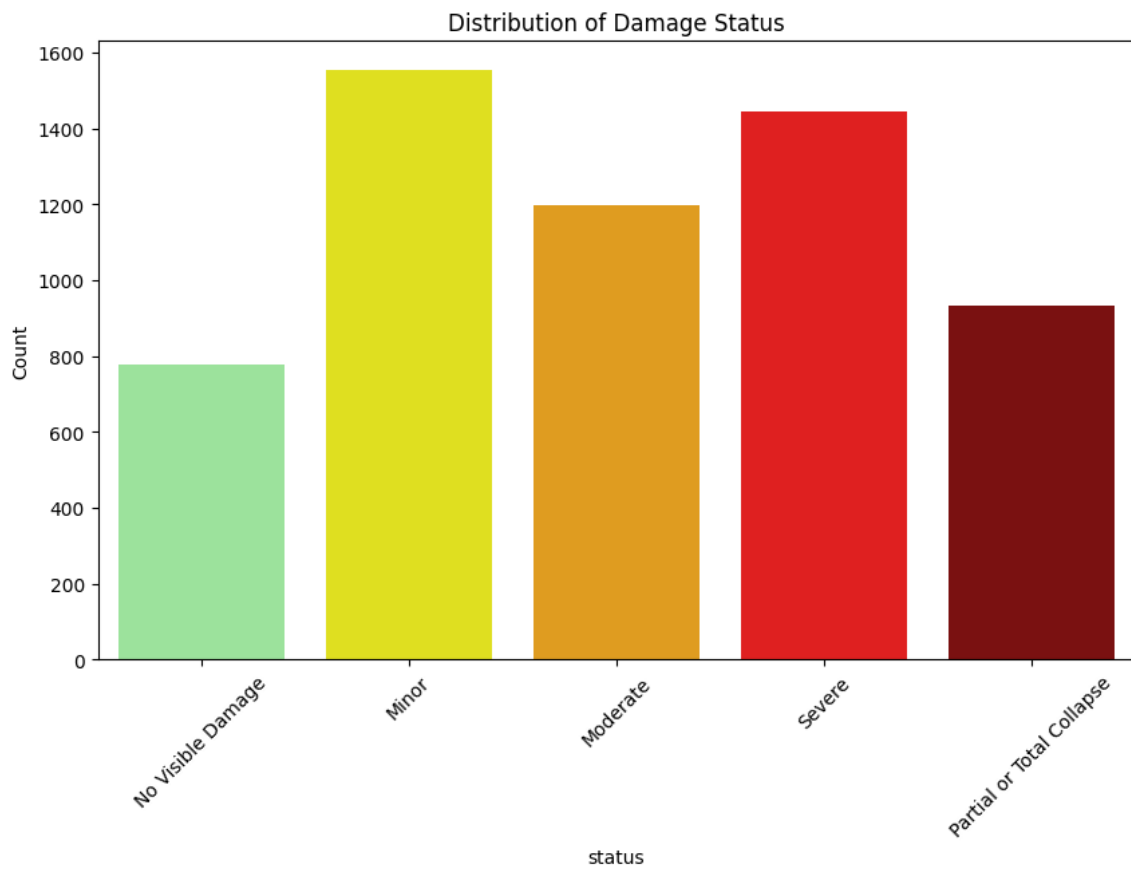


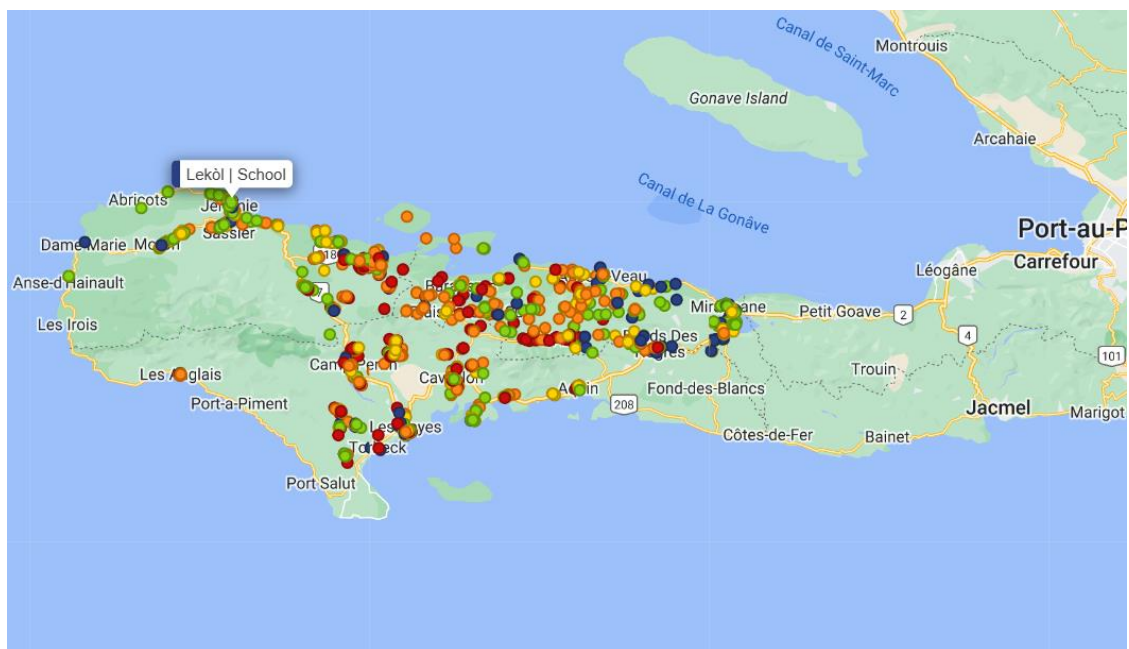
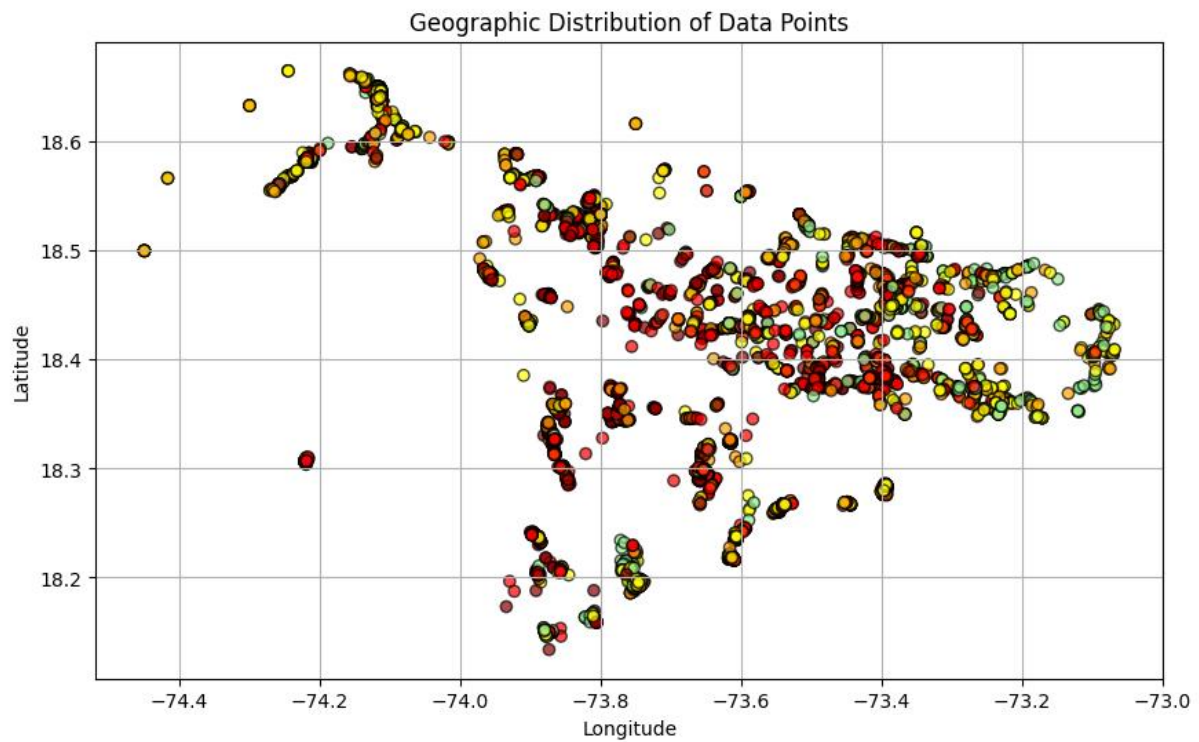
Severe

**Partial or Total Collapse (Grade 5):** The highest grade of damage represents very heavy structural damage, possibly a total or near-total collapse, which would require demolition and complete reconstruction. A building that sustained catastrophic damage, as shown in the attached images classified under this category, depicts the worst-case scenarios for the buildings [7].



Partial or Total Collapse





## 5. Methodology

This section presents the methodology that has been followed for this work, describing the dataset, outlining the pre-processing done, listing the models tried, and explaining the setup used for training together with performance metrics.

## **a. Dataset**

The dataset used in this study was taken from the Structural Extreme Events Reconnaissance (StEER) Network. The organization issued a Preliminary Virtual Reconnaissance Report (PVRR) after an M7.2 Nippes, Haiti Earthquake that happened on August 14, 2021. The dataset contains a total of 18,651 labeled images, which portray a wide extent of damages the earthquake had on different buildings within the affected areas [13].

Most of the pictures were taken outdoors, and most of them cover different angles of each building, which gives a holistic view of the damage. Each house has two or more images from different perspectives, enhancing the ability of the dataset to convey the general structural integrity of the buildings. The pictures are then carefully labeled by the experts from the StEER Network, providing trustworthy categorization into five distinct classes: minor, moderate, severe, partial or total collapse, and no visible damage. These categories show progressive levels of structural damage severity, hence allowing precise classification and analysis in post-disaster assessment.

The distribution of images per damage class is as follows:

- Minor: 4,875 images
- Severe: 4,601 images
- Moderate: 3,845 images
- Partial or Total Collapse: 2,951 images
- No Visible Damage: 2,379 images

## **b. Preprocessing and Data Augmentation**

### **Data split:**

In doing so, we first split the dataset into two subsets: a training set and a test set, using a stratified data-splitting approach. This ensures that each damage class is proportionally represented in both subsets. We used an 70:30 split ratio, with 70% of the images going into the training set and the rest into the test set. Specifically, the split yielded:

- Training Set: 13,055 images across all five classes.
- Test Set: 5,596 images across all five classes.

The stratified splitting was performed with the `train_test_split` function from `sklearn.model_selection`. The advantage of using stratified splitting is that it preserves the original class distribution in both the training and the test sets. It would thus prevent the model from becoming biased toward any particular class during training and hence give correct performance metrics on the test set, reflecting real classification ability. This also uses a fixed random state so we can reproduce the same split in different runs, hence ensuring the results are consistent and reliable across different runs of the experiment.

### **Class Distribution and Balancing Techniques :**

In many classification tasks, imbalanced datasets can result in models that perform poorly on less represented classes, because the model tends to be biased toward the more frequent classes. The following are common data balancing techniques applied to overcome this problem:

**Random Oversampling:** It involves duplicating examples from the underrepresented class until all classes are balanced. While simplistic in nature, it may lead to overfitting on duplicated samples, where the model may learn redundant information.

**Random Undersampling:** This approach balances the dataset by removing examples of the overrepresented classes; some useful data could potentially be lost in this process. This approach is applicable when the size of the dataset is very large, and oversampling would lead to excessive data redundancy. **Synthetic Data Generation:** The SMOTE-like techniques generate synthetic samples for the underrepresented classes through interpolation between existing examples. It is good in that it enriches the dataset without making a mere copy, although sometimes it may generate unrealistic samples.

**Cost-Sensitive Learning:** Instead of balancing the dataset, cost-sensitive learning penalizes misclassifications of minority classes with higher costs, which helps the model learn to accurately predict those classes. In our case, we have selected Random Under sampling because of the large size of datasets and the possibility of redundancy when using oversampling. We picked the minimum class size as a baseline and then randomly sampled from each class to make sure all of them have the same number of images. The result was a balanced dataset with 2,379 images for each damage class: Minor Damage, Moderate Damage, Severe Damage, Partial or Total Collapse, and No Visible Damage. The balancing strategy guaranteed that the model would have an equal representation of each class and potentially less bias and better generalization across classes. This is a known approach for image classification tasks with an imbalanced dataset, as described by [8] in their work on balancing techniques for deep learning

### **Label Preprocessing and One-Hot Encoding:**

In this study, these damage-state labels—Minor, Moderate, Severe, Partial or Total Collapse, and No Visible Damage—were pre-processed using one-hot encoding, which is a common method to transform categorical labels into a binary format. One-hot encoding is the way of transforming a label into a vector where each position corresponds to a class. For instance, an image with a label of "Severe" would be encoded as [0, 0, 1, 0, 0] in a five-class system, where the "1" corresponds to the class and the rest is "0s".

This approach has several advantages:

1. **Better Model Compatibility:** Most machine learning models, especially neural networks, tend to perform better if the inputs are numerical. One-hot encoding ensures labels are in a format conducive to learning without introducing class hierarchies.
2. **Supports multiclass classification:** The model, by keeping the classes independent, can support complex multiclass structures without imposing ordinal or ranked relationships.
3. **Efficient Loss Calculations:** This structure allows neural networks to easily compute loss functions such as categorical cross-entropy, thereby optimizing the learning process.

One-hot encoding has proven effective in a wide range of classification tasks where categorical labels have represented independent classes [9].

### **Image Resizing:**

In preparing images for a convolutional neural network (CNN), consistency in input dimensions is crucial for efficient model training. In this study, all images were resized to 128x128 pixels, ensuring uniformity across the dataset. This resizing step minimizes potential irregularities during training that could arise from varying image dimensions, as CNNs typically require fixed-size inputs for batch processing.

By standardizing the image size, we streamline the data flow through the network and reduce computational complexity. Additionally, resizing allows the model to more efficiently capture essential features without handling unnecessary data, thus optimizing training speed and memory usage.

### **Data Augmentation:**

Data augmentation is an important technique in deep learning model training, especially for small or imbalanced datasets. It artificially inflates the diversity of the training data, helping the model generalize better and reducing overfitting. Overfitting occurs when the model becomes too complex to learn only the specific details of the training set rather than capturing the general patterns, which results in poor performance on unseen data. Augmentation helps simulate real-world variations, making the model robust to changes in orientation, position, and lighting; this will improve the resilience of the model and its accuracy on test data.

Data augmentation techniques were applied to the training images in order to introduce variability, including:

- **Normalization:** Each pixel value was scaled between 0 and 1 by dividing by 255. This normalization step helps the model converge faster and improves numerical stability during training.
- **Rotation Range:** Images are rotated in the range of  $\pm 20$  degrees. This rotation creates variability in building orientation so that the model can recognize damaged buildings from slightly different angles.
- **Width and Height Shifts:** up to 20% of random shifts in the width and height of the image. These shifts help the model be more robust when the building is not centered in an image, showing some tolerance to framing and positioning differences. Horizontal Flip: Horizontal flipping was performed to achieve greater image diversity. While the orientation of buildings sometimes affects damage assessment, horizontal flipping was applied to improve generalization with consideration for similar structural patterns.

These augmentations proved effective in improving model performance on our damage classification task by making the model invariant to minor spatial and orientation variations [10].

## **c. Model structure**

Convolutional Neural Networks (CNNs) have released a new revolution of image classification tasks and become one of the most widely used applications in medical imaging, facial recognition, and disaster damage assessment [1]. CNNs use a series of convolutional and

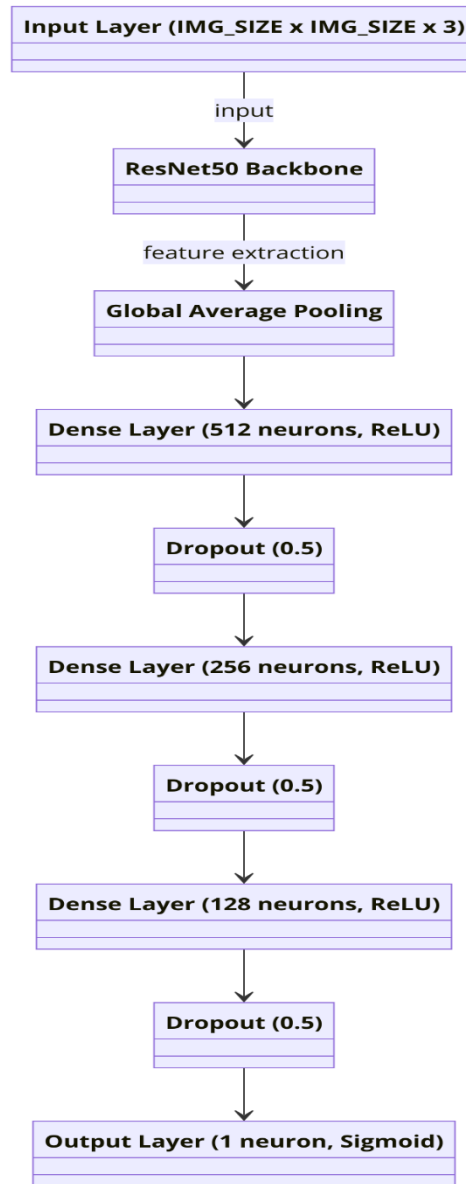
pooling layers to capture spatial and hierarchical patterns in images, which are critical for accurately identifying and classifying complex visual features. In this context, the building damage classification can efficiently learn discriminative features from images of damaged structures that include cracks, debris, and collapse patterns to assist post-disaster assessment [11].

For this work, we have used the ResNet50 as the backbone of our CNN architecture. ResNet50 is a 50-layer deep network with wide application in image classification tasks. It was motivated by the ability of residual connections to overcome the vanishing gradient problem and to represent deep features. This was achieved by the skip of layers permitted by residual connections, which helped preserve major information and resulted in the possibility of training much deeper networks without degradation in performance [12].

This architecture has shown good accuracy and robustness and therefore would be the best choice for tasks involving complex image data, such as the classification of various levels of structural damage.

Our model's architecture is structured as follows:

- **Input Layer:** We begin with an input layer that takes in images resized to 128×128 pixels with three color channels (RGB). This helps to maintain uniform dimensions for all images, which is a requirement for batch processing and uniform feature extraction.
- **Base Model (ResNet50 Backbone):** We start with the pre-trained ResNet50 model, excluding its top classification layer, trained on the ImageNet dataset. A pre-trained model extracts general features learned from a large-scale dataset, which would yield a good starting point for our task. In order to fine-tune, all the layers in ResNet50 are set to be trainable so that the model can fully adapt to our building damage classification task.
- **Feature Extraction:** The output from the ResNet50 backbone is fed into a global average pooling (GAP) layer to reduce dimensionality and aggregate important features while retaining spatial information.
- **MLP Block:** The MLP block contains fully connected (Dense) layers interleaved with Dropout layers. Dropout regularization with a rate of 0.5 helps prevent overfitting by randomly deactivating neurons during training, enhancing generalization.
  - Dense layer with 512 units and ReLU activation.
  - Dropout layer with a rate of 0.5.
  - Dense layer with 256 units and ReLU activation.
  - Dropout layer with a rate of 0.5.
  - Dense layer with 128 units and ReLU activation.
  - Dropout layer with a rate of 0.5.
- **Output Layer:** The final output layer uses a single neuron with a sigmoid activation function for binary classification (damaged vs. undamaged), which outputs a probability score for each class. By taking advantage of this architecture, our model balances the ability to learn intricate structural features with the stability needed to prevent overfitting. The ResNet50 backbone enhances our model in extracting features, and the MLP layers provide further layers of abstraction, making the network even more apt at handling the post-disaster damage assessment complexities.



Our model structure

#### d. Training Setup

The model is trained in a supervised learning fashion, where, of course, the dataset is divided into training, validation, and test sets. The training set consists of those images used for optimizing the parameters of the model, while the validation set is used in fine-tuning the hyperparameters to avoid overfitting of the model. The test set was kept aside for the model's performance on unseen data.

Training was performed with the following parameters:

- **Batch Size:** The selected batch size was 32 to balance the computational load and memory usage.
- **Epochs:** The model will be trained for a total of 30 epochs; however, early stopping is activated to stop the training if there is no improvement in the validation performance for a given patience period.

- **Optimizer and Loss Function:** Adam optimizer is selected because it is efficient at converging to optimal weights. On the other hand, categorical crossentropy is selected as the loss function because this network performs multiclass classification tasks.

## e. Evaluation Metrics

### **Accuracy:**

Accuracy is the overall percentage of images classified correctly by the model.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

- Where:
  - TP = True Positives
  - TN = True Negatives
  - FP = False Positives
  - FN = False Negatives

### **Precision:**

Precision is the ratio of true positive predictions to all positive predictions, indicating how well the model avoids false positives.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

### **Recall (Sensitivity)**

Recall measures the percentage of true positive predictions out of all actual positive instances, showing how well the model identifies all relevant cases.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

### **F1 Score:**

The F1 score balances precision and recall, providing an overall metric for model performance, especially useful for imbalanced classes.

$$\text{F1 Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

### **Cross-Validation**

Cross-validation techniques help evaluate the model's performance on different subsets of the dataset, ensuring that the results are not dependent on a single train/test split.

The most common method is k-fold cross-validation, where the dataset is divided into k subsets (or folds). The model is trained on k-1 folds and validated on the remaining fold, repeated k times. The overall performance is the average of the k validation scores.

These metrics provide a comprehensive evaluation of the model's performance, allowing for a more nuanced understanding of its strengths and weaknesses, particularly in the context of class imbalance.

## 6. Results

### a. Results for Binary Classification (Damaged vs. Undamaged) with Data Augmentation Applied

For this task, we want our model to classify pictures into two classes: damaged and undamaged. To make the model more generalizable and less prone to overfitting, we used data augmentation with transformations like rotation, zoom, and horizontal flipping. The model was trained for a total of 25 epochs using a binary cross-entropy loss function.

Performance Overview:

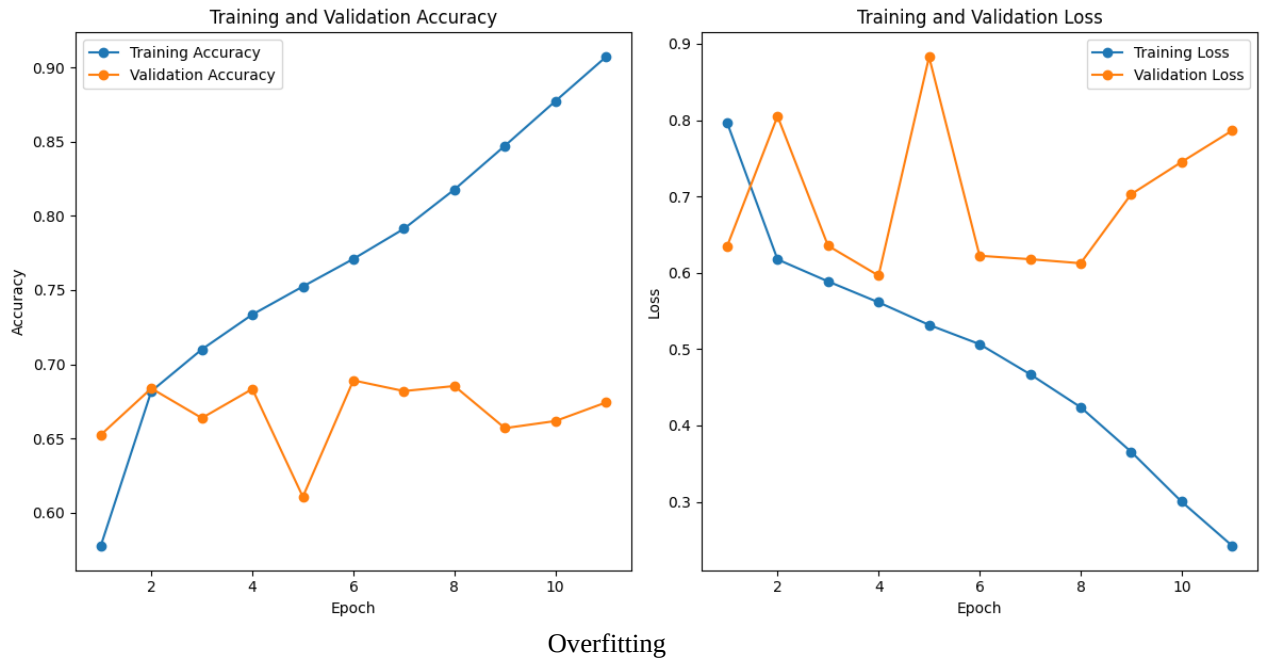
- Best Training Accuracy: 79.6%
- Best Validation Accuracy: 70%
- Precision: 76%
- Recall: 70%
- F1 Score: 72%

The above metrics show that such data augmentation helped to generalize better the model, although the validation accuracy is lower than the training accuracy. This proves that the model was not overfitted since there was a relatively close performance between the training and validation sets.

### b. Results for Binary Classification (Damaged vs. Undamaged) without Data Augmentation Applied

We experimented with not using data augmentation. In this case, the model reached high accuracy of 90% after only 12 epochs. However, it was accompanied by a validation accuracy of only 67%, meaning it was again overfitting the training data. Overfitting is when a model learns the training data too well, including its noise and outliers, at the cost of generalization ability on unseen new data.

Given the large difference in training and validation accuracies, we decided to stop training to avoid overfitting of the model. Such a finding supports the data augmentation strategy for increasing the robustness and generalization ability of the model. Data augmentation actually decreases the initial training accuracy but plays a vital role in increasing the model's performance on the validation data, which is important for increasing overall generalization abilities.



### c. Results for three-class classification task, distinguishing between "Damaged," "Moderate," and "Undamaged"

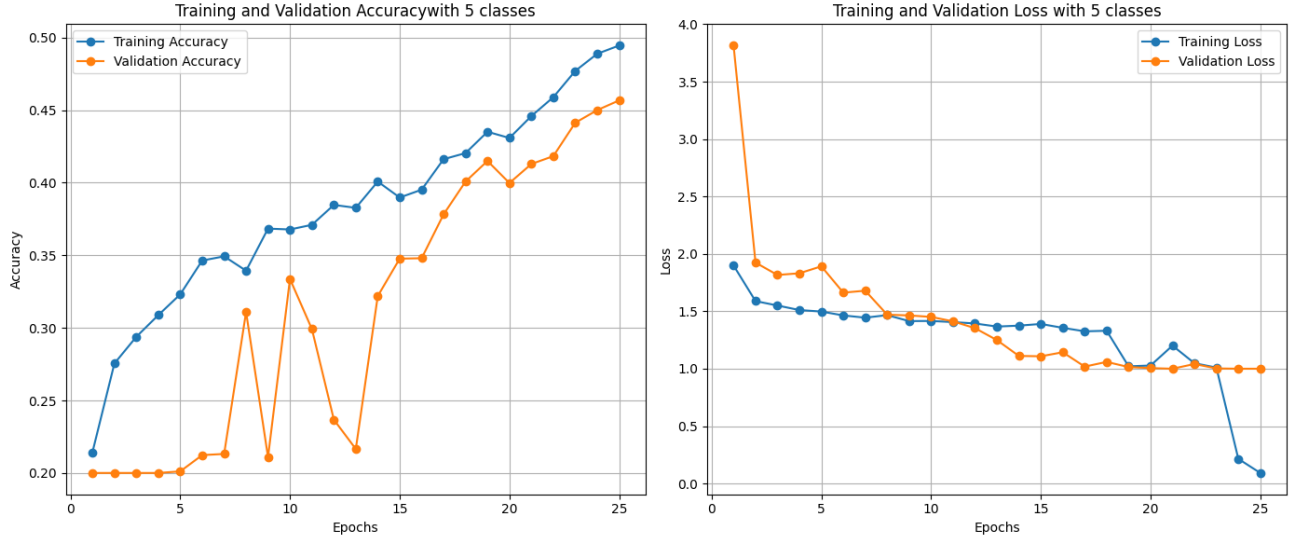
We examined the model's performance on a three-class classification task, distinguishing between "Damaged," "Moderate," and "Undamaged" categories. This training utilized data augmentation over 20 epochs with a batch size of 64 and a learning rate of 0.0005.

| Class     | Precision | Recall | F1-Score | Support |
|-----------|-----------|--------|----------|---------|
| Damaged   | 0.69      | 0.6    | 0.62     | 2265    |
| Moderate  | 0.6       | 0.58   | 0.59     | 1154    |
| Undamaged | 0.61      | 0.85   | 0.63     | 2177    |
| Accuracy  |           |        | 0.67     | 5596    |

The overall accuracy of the model was 67%, indicating that while the model could effectively categorize images and the validation accuracy is 65%.

### d. Results for training on a model designed to classify images into five distinct classes: "Severe," "Partial or Total Collapse," "Moderate," "Minor," and "No Visible Damage."

This training utilized a batch size of 64, ran for 25 epochs, and employed a learning rate of 0.0005. The best accuracy achieved during this training phase was 49%, with a corresponding validation accuracy of 45%. These results indicate significant challenges in accurately classifying the images into the five categories, reflecting the complexities inherent in distinguishing between classes that may have subtle differences in visual features.



## 7. Conclusion

In our study, we examined the influence of data augmentation on the performance of our convolutional neural network (CNN) in classifying building damage from post-disaster images. We first trained our model without any kind of data augmentation and achieved an impressive 90% accuracy on the training set. That came at the cost of overfitting, as evidenced by a much lower validation accuracy of only 70%. By incorporating data augmentation strategies, we found the training accuracy dropped to 80%. While the training accuracy did decrease, the validation accuracy remained high and more consistent with training performance, showing that data augmentation is key to model generalization and avoiding overfitting. We also tried a multiclass classification, but the best accuracy we could obtain was 50%. We then concluded that binary classification results are more dependable for this particular task. By experiments on pre-trained models, such as VGG16, we found the best performance with ResNet50 as our CNN backbone. The residual learning framework of this architecture allowed it to catch complex features from the images, which definitely improved our classification results. These findings, in general, provide the importance of using data augmentation and transfer learning in building an effective image classification model in practical disaster management [3]. Our results also show that future work would refine the classification strategies and also explore more augmentation techniques to further improve the performance of the models.

## 8. Future Work

Looking ahead, we aim to enhance our damage classification model by incorporating multi-perspective features for each house. By utilizing images captured from various angles, we believe we can provide a more comprehensive understanding of the structural conditions and damages sustained during disasters. This multi-view approach has the potential to improve classification accuracy by enabling the model to learn diverse characteristics and context about each building, which a single perspective may not adequately convey.

Integrating multi-perspective images could also facilitate a richer feature extraction process, allowing the model to capture nuanced details that may be critical for accurate damage assessment. We plan to explore advanced techniques such as 3D reconstruction and depth estimation to further enrich the dataset and improve classification performance. Additionally, we will investigate the use of ensemble methods, combining predictions from multiple models trained on different views, to optimize our approach.

In summary, our future work will focus on leveraging multi-perspective features to build a more robust and accurate classification system for building damage assessment, ultimately contributing to more effective disaster response and recovery efforts.

## References

- [1] Z. Hong, H. Zhong, H. Pan, J. Liu, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, C. Zhong, Classification of Building Damage Using a Novel Convolutional Neural Network Based on Post-Disaster Aerial Images, *Sensors* 22 (2022) 5920. <https://doi.org/10.3390/s22155920>.
- [2] A. Azizi, M. Yaghoobi, S.R. Kamel, Intelligent Detection and Assessment of Damaged Buildings using UAV Imagery and YOLOv8, (2023). <https://doi.org/10.21203/rs.3.rs-3143110/v1>.
- [3] P.D. Ogunjinmi, S.-S. Park, B. Kim, D.-E. Lee, Rapid Post-Earthquake Structural Damage Assessment Using Convolutional Neural Networks and Transfer Learning, *Sensors* 22 (2022) 3471. <https://doi.org/10.3390/s22093471>.
- [4] Q. Yu, C. Wang, F. McKenna, S.X. Yu, E. Taciroglu, B. Cetiner, K.H. Law, Rapid visual screening of soft-story buildings from street view images using deep learning classification, *Earthq. Eng. Eng. Vib.* 19 (2020) 827–838. <https://doi.org/10.1007/s11803-020-0598-2>.
- [5] H. K. Miyamoto, G. J. Sechi, G. Victor, B. St Come, M. Broughton, A. S. J. Gilani, and A. Singh, "Haiti earthquake 2021: Findings from the repair and damage assessment of 179,800 buildings," *\*International Journal of Disaster Risk Reduction\**, vol. 73, p. 102846, 2022. DOI: 10.1016/j.ijdrr.2022.102846.
- [6] United States Geological Survey (USGS), M 7.2 - Nippes, Haiti, (n.d.). <https://earthquake.usgs.gov/earthquakes/eventpage/us6000f65h/pager>(accessed September 20, 2024).
- [7] European Macroseismic Scale 1998 (EMS-98), 1998.
- [8] A systematic study of the class imbalance problem in convolutional neural networks Mateusz Buda, Atsuto Maki, Maciej A. Mazurowski
- [9] Goodfellow, I., et al. (2016) Deep Learning. MIT Press, Cambridge, MA.
- [10] A survey on Image Data Augmentation for Deep Learning Connor Shorten & Taghi M. Khoshgohar
- [11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep learning." *Nature*, 521(7553), 436-444. This paper provides an overview of deep learning advancements, including CNNs and their applications in image classification.

- [12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep residual learning for image recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778). This paper introduces ResNet, detailing the residual connections that allow for training deep networks and presenting results that showcase its effectiveness in image recognition tasks.
- [13] StEER, "Preliminary Virtual Reconnaissance Report (PVRR): Nippes, Haiti Earthquake (M7.2), August 14, 2021," 2021. [Online]. Available: <https://www.steer.network/>