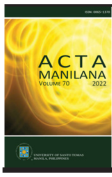
Acta Manilana 70 (2022), 13-28

ISSN: 0065-1370



Acta Manilana

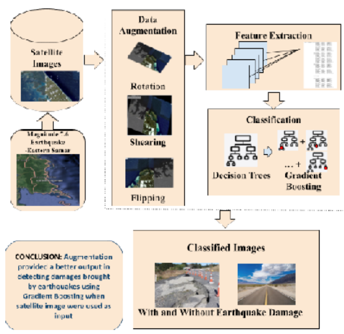
journal homepage: https://actamanilana.ust.edu.ph/

Detection of Earthquake Damages from Satellite Images using Gradient Boosting Algorithm with Decision Trees as Base Estimator

Donata D. Acula[1]

[1]College of Information and Computing Sciences; Research Center for Natural and Applied Sciences, University of Santo Tomas, Manila Philippines

Graphical Abstract



Abstract

The Philippines experiences an average of twenty (20) earthquakes every day and 100 to 150 felts every year. With the aim to help the government in mitigation of the potential impact of the earthquake in the country, this research explored the detection of earthquake damages from the satellite images. Since Gradient Boosting Algorithm is considered one of the efficient and powerful predictive models, this method was employed in this research to classify if the satellite images before and after the earthquake brought damages in the infrastructure of affected areas. The satellite images used in the research were downloaded from Landsat 8 via Google Earth pro, which focused on a Magnitude 7.6 earthquake in Eastern Samar on August 31, 2012. The research used image augmentation such as rotation, shearing and flipping before the image extraction. The extracted features were used as classifiers of the models and were separated into 80:20, 70:30 and 60:20 ratios for training and testing sets respectively. The detection of damages was evaluated five (5) times using different n-estimators or numbers of trees {10, 20, 30, 40, 50}. The experiment concluded that Gradient Boosting Algorithm is an efficient model for classification and detection of earthquake damages using satellite image data, with an optimal detection accuracy of 85.71% and 100% without and with image augmentation.

Keywords: image augmentation; satellite imagery, artificial intelligence; ensemble machine learning

Corresponding author: ddacula@ust.edu.ph
DOI: https://doi.org/10.53603/actamanil.70.2022.wpyl1791

Introduction

The Philippines is considered vulnerable to hazards brought by various natural calamities such as tropical cyclones, earthquakes, volcanic eruptions, and other natural disasters. This is due to the geographical location of the country in Southeast Asia and in the western edge of the so-called "Ring of Fire" [1], [2]. Among the natural calamities, earthquakes brought a high number of casualties and damages in different infrastructures. Earthquakes have two types, tectonic and volcanic. The tectonic earthquake was produced by sudden movement along faults and plate boundaries, while the volcanic earthquake is caused by rising lava or magma beneath active volcanoes. This can be measured by magnitude and intensity. Magnitude is proportional to the energy released by an earthquake and calculated from earthquakes recorded by seismograph while the intensity is the strength of an earthquake as perceived and felt by the people in a certain location which was determined using the PHIVOLCS Earthquake Intensity Scales [3]. On average, there are 20 earthquakes every day in the Philippines and 100 to 150 felts every year [1], [4]. This is because the country has five (5) active fault lines known as Western Philippine Fault, Eastern Philippine Fault, South of Mindanao Fault, Central Philippine Fault, Marikina/ Valley Fault [4]. As of 2020, earthquakes have the highest risk index (10.0), followed by tropical cyclones (9.5), tsunami (9.3), floods (7.2) and drought (4.0) [4]. Among the earthquakes happened in the Philippines, the following is considered to be strongest and deadliest: Moro Gulf Earthquake and Tsunami (1976) with more than 8,000 casualties and deaths, Luzon Earthquake (1645) with more than 600 casualties and deaths and Luzon Earthquake (1990) with more than 1621 casualties and deaths [4], [5].

To prepare for the different natural calamities, the Philippines government led by the National Disaster Coordinating Council (NDRRMC) continuously plans and improves their system to ensure that the country is prepared for these phenomena in accordance with the Republic Act 10121 of 2010. For the safer, adaptive and disaster resilient Filipino communities towards sustainable development the NDRRMC have four (4) thematic areas: (1) Disaster Prevention and Mitigation which aims to avoid hazards and mitigate their potential impacts by reducing vulnerabilities and exposure and enhancing capacities of communities, (2) Disaster Preparedness which aims to establish and strengthen capacities of communities to anticipate, cope and recover from the negative impacts of emergency occurrences and disasters, (3) Disaster Response which provide life preservation and meet the basic subsistence needs of affected population based on acceptable standards during or immediately after a disaster and (4) Disaster Rehabilitation and Recovery which aims to restore and improve facilities, livelihood and living conditions and organizational capacities of affected communities, and reduced disaster risks in accordance with the "building back better" principle [6]. The aim of this proposed research coincides with the mentioned thematic areas.

At present, the Philippine government through the DOST detects the damages of earthquakes using satellite images by measuring if there is change or no change in the location of the earthquake. There is no existing tool to categorize the level of earthquake damages based on the satellite imagery. Instead, they relied on the actual occurrence of damages on the specific property such as buildings, roads and other establishments. Thus, satellite imagery is not required in their existing method of earthquake damage classification.

Data Science, the newest and the fastest-growing area in the field of information technology education, has a bigger role and impact in the policymaking of different private and public institutions. One of the major roles of this area of study is to help the government or private institutions to develop a policy and strategic plan by applying various algorithmic models. As highlighted in the study of Kolman [7], these algorithms and algorithmic models may dictate what policy is perceived feasible. Hagen et al. [8] also concluded that visual analytics, a major area of Data Science, has potentially positive impacts on policy making practices. Artificial Intelligence (AI) is one of the main tools needed in exploring the Data Science area, specifically in this type of research that involves data analysis using predictions or classifications. Artificial Intelligence is sometimes denoted as Machine Learning (ML) techniques [9].This is because ML is an application of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed, and focuses on the development of computer programs that can access data and use it to learn from themselves [10].

In this study, the satellite images gathered from Google Earth pro were used. The study focused on the magnitude 7.6 earthquake with epicenter in Eastern Samar, Philippines (latitude: 10.811°N and longitude: 126.638°E) as shown in Figure 1. The estimated damage to the said earthquake is more than PHP 26M which includes the schools, bridges and roads with 5,839 affected families (28,869 persons) [11].

The satellite image data underwent preprocessing methods and then employed the gradient boosting algorithm with decision trees as base estimator to detect the earthquake damages. A total of 69 images were used after applying data augmentation (rotation, flipping and shearing) with 14,212,800 features.To optimize the classification model, different data partitioning and fine-tuning of parameters were employed.

To summarize, this paper would like to address the problem of image detection using a reliable model known as Gradient Boosting, which was discussed in the next section. It also explores various sub-problems such as the effect of applying the augmentation techniques in the dataset, exploring different distributions of training and testing data and experimentation in the parameter of the base estimator.

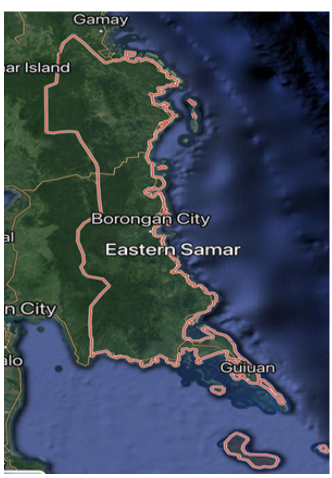


Figure 1. Location of Epicenter of Magnitude 7.6 Earthquake on June 31, 2012 (Source: Google Maps)

This study is just an initial step in the ultimate goal of classifying the level of damage of earthquakes in the country through satellite images using the most efficient artificial intelligence which was not implemented and used by the Philippine government.

The remaining parts of this paper were organized as follows: Section 2 discusses the related research outputs or articles, Section 3 discusses the methodology used in the research and Section 4 and 5 showcase the results of the experiment and the conclusion based on these results.

Related Works

Satellite image data or satellite remote sensing is very useful in analyzing the disaster risks such as prediction, classification/detection. It is already the current trend or practice of different disaster management units to make use of the extracted data from the images. In the proposed study, satellite images will also be used to detect the disaster risks in the country through the use of Artificial Intelligence. Some research outputs that involve the satellite imagery used in this study are discussed in [12], [13], [14], [15], [16 - 19].

Zhang et al. [19] proposed a method of automatically extracting house damage information from post-quake high resolution optical remote sensing imagery using the multiscale fusion of spectral and textural features. In their experiment, they first enhanced the textural and spectral features of the images at the pixel level. Then, the resulting feature images are fused at the feature level and the fused feature images are segmented using superpixels. Finally, they constructed the post-quake house damage index model. The implemented model garnered an overall accuracy of 76.75%, 75.35% and 83.25% using the three different types of imagery. Based on the performance of the model, they claimed that the model is useful in extracting the damage information from multi source remote sensing data and this will help the government agencies for post-disaster rescue and earthquake damage assessment. In the paper of Niu et al. [18] to respond to natural disasters, a key problem is how to efficiently schedule multiple earth observation satellites to acquire image data of a large stricken area by coordinating multiple different even conflicting needs of disaster relief, such as the extent of coverage over the stricken area, timeliness, and the spatial resolution. In this paper, considering two typical application scenarios during the response phase, the authors proposed a multi-objective optimization method to solve the problem of satellite scheduling of a large area target. The team believed that earth satellite observations are very useful during the response phase of disaster management, since satellites could provide accurate, frequent and almost instantaneous data for large areas anywhere in the world. Chen et al. [17] also claimed that landslide barrier lakes usually form quickly after disasters and require very timely remote sensing images to monitor the land-cover change. However, cloud-free images are not always available in emergencies.

This paper provided a method to fuse multitemporal cloud-covered images for change detection, based on the evidential fusion framework.Their team proposed a method that detects the landslide barrier lake in a real case study, using a series of cloud-covered images from the GF-1 satellite.

In the article published by University of Iowa and the U.S. Geological Survey [16], they concluded that the data obtained from orbiting satellites provided more accurate information on the impact of large earthquakes, which will provide more effective emergency response. The satellite imagery provides detailed information about where the earthquakes occurred. This information was then incorporated into a set of operational response guides managed by the USGS National Earthquake Information Center (NEIC) that is distributed to decision makers, search and rescue operations, and other groups.Satellite remote sensing is one of the primary support tools for disaster management. They regarded the Sentinel Asia as an empirical research project to study how satellite remote sensing can support disaster management, in collaboration with users. This paper derives requirements for applying satellite remote sensing to disaster management support via a holistic (including human factors) and staged approach based on case studies in SA. The paper of Li et al. [13] also used the high-resolution remote sensing technology in disaster Reduction and Relief. Specifically, the authors focused on the magnitude 7.0 earthquake that occurred in Lushan County, Sichuan Province last April 20th, 2013. This phenomenon has caused extensive damage to housing and roads, and led to seismic secondary disasters such as mountain collapse and landslides. The paper interpreted images based on disaster characteristics, extracted and analyzed targets such as landslide bodies and roads, and monitored and assessed disaster conditions based on the results of extraction. The utilization of the satellite SAR image for monitoring the occurrence of a natural disaster has been tried by Umemura et al. [15] by obtaining the satellite images using Synthetic aperture radar (SAR). Aside from the fact that SAR is flexible because it can obtain the data anytime (night/day). It also has a feature of all-weather-observation because it emits an L-band microwave, which is less affected by clouds and rain. The PALSAR-2 aboard the ALOS-2, which achieves high-resolution (3m) observation, is now in operation as the satellite L-band SAR in Japan. Finally, the study of Hassan et al. [12] presented the pre- and post-event of Sarpol-Zaha Earthquake using the Pleiades-1 satellite optical image and two Sentinel-2 satellite images. Their team employed the Support Vector Machine and Neural Networks classification methods to detect the damages of the aforementioned earthquake.They concluded that the accuracy for detecting collapsed buildings in the SVM classification method is 78.34% and 72.43% for neural network classification. They claimed that the results of their experiment show the change map of the pre- and post-earthquake medium-resolution satellite images such as Sentinel-2 can reveal the collapsed buildings caused by the earthquake successfully.

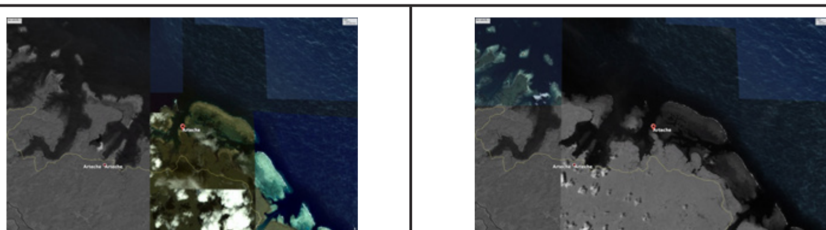
The role of artificial intelligence is vital in dealing with satellite image data. Various researchers tried and explored this area from preprocessing up to the detection, classification or prediction. Among the existing machine learning algorithms, this study focused on Gradient Boosting Algorithms as a model for detection. Since numerous researchers considered this method as a powerful machine learning technique that showed considerable success in a wide range of practical applications and at the same time this method is highly customizable to the particular needs of the application, like being learned regarding different loss functions [14].

Gradient boosting algorithm (GBA) was used in drought classification and prediction and compared to decision tree and genetic programming by Danandeh [20]. These models were trained and validated using the first 70% and last 30% of the datasets, respectively. The author concluded that the gradient boosting algorithm provided a better performance than DT and GP and also exhibited that it can effortlessly classify and predict the number of dry, normal, and wet events in both case studies. Pigeon and Duval [21] also used the gradient boosting algorithm along with the generalized linear models in predicting the total paid amount of each claim in insurance. It is also claimed that GBA is an efficient method on computing the predicted values such as total amount paid for each claim and payment schedule. Alcolea and Resano [22] also implemented the GBA in an accelerator design to optimize the execution of the decision trees in reducing the energy consumption. It was implemented in a Field Programmable Gate Array (FPGA) for embedded systems, and tested it with a relevant case-study: pixel classification of hyperspectral images. The authors concluded that with the execution of LightGBM in a high-performance CPU, the model is capable of achieving 2x performance while consuming 72x less energy and reaches a 30x performance improvement while maintaining 23x less energy consumption during the execution. Thus, GBA is suitable to provide high-performance for embedded systems.

Körner et al. [23] also used the GBA as an adaptable machine learning method to fill gaps caused by missing or erroneous data in meteorological time series. The method was applied on a large data set of hourly time series of the measurements: air temperature, wind speed and relative humidity for station-based observations in Germany covering the period from 1951 to 2015. It was claimed that the GBA outperforms the other techniques such as neural network in calculation time, performance and the handling of missing predictor values. Finally, Kim et al. [24] implemented the GBA along with ridge regression and Support Vector Machines in predicting the electricity demand by designing optimal combinations of weather data, stations, and years. After the experiment, the team concluded that GBA has the best forecasting accuracy as compared to the other methods.

Materials and methods

This research used the satellite images as an input of the system gathered from the Google Earth pro, which relied on from NASA and USG's Landsat 8. A total of 23 images from the different municipalities in Eastern Samar were used with an average eye altitude of 17.78 km and an image resolution of 2961 by 4800 pixels. The sample input images before and after the earthquake were shown in Figure 2.



(a) Input Image Before Earthquake | (b) Input Image After Earthquake

Figure 2. Sample Input Satellite Images Before and After the Earthquake

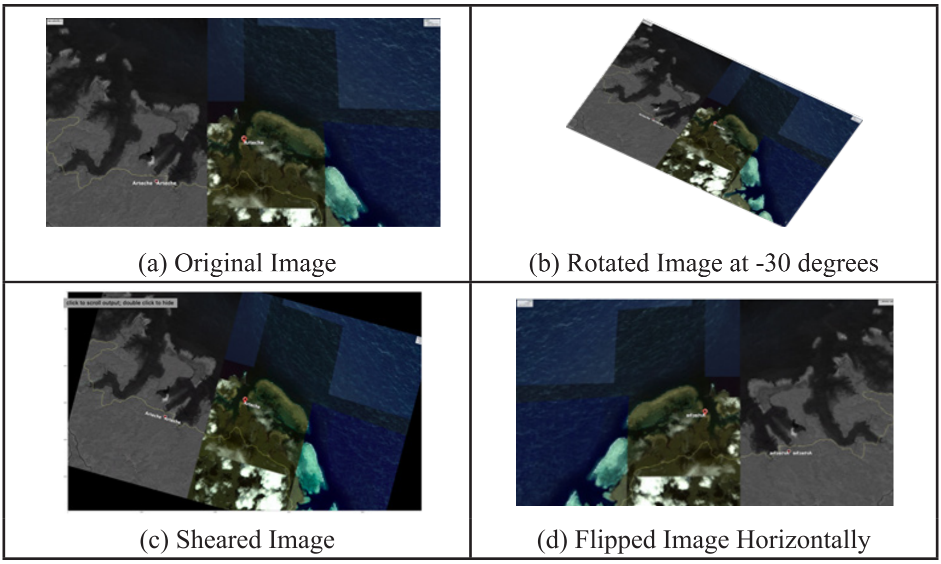


Figure 3. Sample Input Satellite Images Before and After the Data Augmentation

```
array([[[25, 23, 28],
        [25, 23, 26],
        [26, 24, 27],
        ...,
        [17, 31, 58],
        [16, 30, 57],
        [15, 31, 57]],

       [[25, 23, 26],
        [25, 23, 26],
        [26, 24, 27],
        ...,
        [16, 30, 57],
        [15, 29, 56],
        [16, 30, 57]],

       [[25, 23, 26],
        [25, 23, 26],
        [26, 24, 27],
        ...,
        [17, 31, 58],
        [16, 30, 57],
        [16, 30, 57]],

       ...,

       [[74, 76, 73],
        [81, 83, 80],
        [86, 86, 84],
        ...,
        [11, 19, 55],
        [12, 20, 56],
        [13, 21, 57]],
```

Figure 4. Extracted Features from the Satellite Image

	0	1	2	3	4	5	6	7	8	9	...	14212790	142127
0	0.093276	0.092711	0.096633	0.100554	0.100554	0.100554	0.108397	0.112319	0.096633	0.108397	...	0.113607	0.1136

1 rows × 14212800 columns

Figure 5. Dataframe of Extracted Features from the Satellite Image

In this research, the effect of image augmentation was also explored. The first group of dataset used in gradient boosting is the collection of satellite images without augmentation, as shown in Figure 3a. While the second group is the collection of all the images including the segmented images through rotation, shearing and flipping as shown in Figure 3b, 3c and 3d.

After data augmentation, the features of the images were extracted. The extracted features per image are 14,212,800 which will serve as the classifier or the predictor of the gradient boosting classification algorithm. This study employed 20-80, 30-70 and 40-80 for testing and training data respectively. Figure 4 and Figure 5 shows the extracted features before and after binarization of image.

In the classification stage, fine-tuning of hyperparameters focusing with the n-estimators or depth of decision tree estimators was also experimented by using 10, 20, 30, 40 and 50. The classified images will be labeled as with or without damage. The complete flow can be viewed in Figure 6.

After completing the classification stage, the classification metrics of algorithm and models will be evaluated by determining the confusion matrix where accuracy, precision, recall and f1-score will be showcased. To determine if there is no significant difference between the actual and classified output, the data was normalized and used the paired t-test with 5% level of significance. This hypothesis will be ignored in case the experiment obtained a 100% accuracy rate in detecting the damages brought by earthquake.

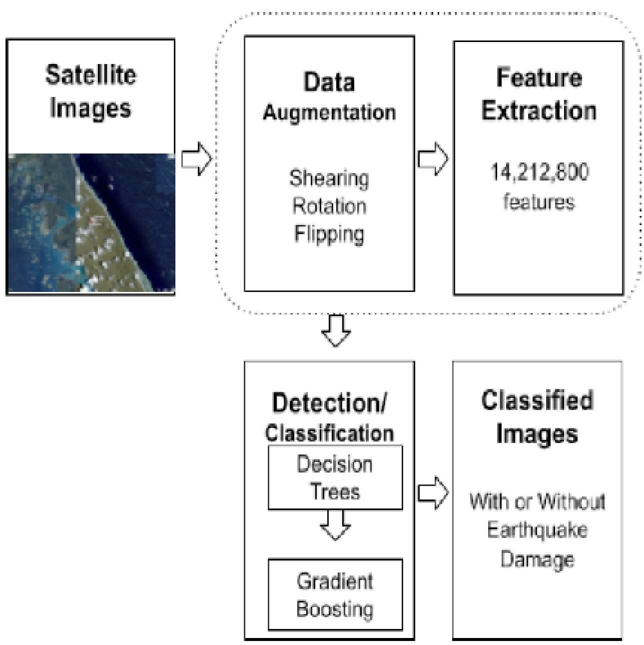


Figure 6. Framework of the study

A. Decision Trees. Decision trees used in this study, since it does not require any prior assumptions about the probability distributions that govern the class and attributes of the data. This model can also be applied to both categorical and continuous data without requiring the attributes to be transformed into a common representation via binarization, normalization, or standardization [25]. The Pseudocode below shows the steps of Decision Tree Classifiers when used in extracted features of the satellite images. It will select the best attributes recursively and split the data and expand the nodes of the tree until the required condition is satisfied.

Decision Tree Induction

```
TreeGrowth(I, A)
if stopping_cond(I, A) = true then
    leaf = createNode()
    leaf.label = Classify (I).
    return leaf.
else
    root = createNode()
    root.test_cond = find_best_split(I, A)
    let O = {o|o is the possible outcome of root.test_cond}.
    for each o element of O do
      Io = {i|root.test_cond(i) = o and i is an element of I}
      child = TreeGrowth(Io, A).
      add child as descendant of root and label the edge as o.
   end for
end if
return root.
```

B. Gradient Boosting Algorithm. Gradient Boosting Algorithm is one of the most efficient and powerful predictive models [26], [14]. This method involves loss function, weak learner that will serve as predictor and the additive which will be added in the weak learners to minimize the loss function. Since this research implemented the binary classification, the appropriate loss function is the logarithmic.In terms of weak learner, this study utilized the decision trees and the depth of the trees used is limited only to 10, 20, 30 and 40. The additive refers to the trees which were added one at a time. To minimize the loss when adding the trees, the gradient descent was used [26]. The pseudocode of the gradient boosting is shown below [27]:

Gradient Boosting Algorithm

1. Initialize $f_o(x) = argmin_{\gamma} \sum_{i=1}^{N} L(y_i, \gamma)$.

2. For $m = 1$ to M:

   i. For $i = 1, 2, \ldots, N$ compute the pseudo-residuals

   $$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}$$

   ii. Fit the classification tree to the target r_{im} giving terminal regions where R_{jm}, $j = 1, 2, \ldots, J_m$.

   iii. For $j = 1, 2, \ldots, J_m$ compute the multiplier γ_{jm} by solving the 1-dimensional optimization problem

   $$\gamma_{jm} = argmin_{\gamma} \sum_{x_i \epsilon R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma) .$$

   iv. Update the model

   $$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \epsilon R_{jm}).$$

3. Output $\widehat{f}(x) = f_M(x)$.

C. Classification Metrics. Classification Metrics was used to determine the performance of different classification models presented in this paper. Confusion Matrix, Accuracy, Precision, Recall or Sensitivity and F1 Score will be provided and computed.

C.1 Confusion Matrix. Confusion matrix, utilized for classification problems that can generate an output with two or more types of classes. It is presented in two-dimensional matrices containing actual and prediction classes whose elements were represented by True Positive (TP), True Negatives (TN), False Positives (FP) and False Negative (FN) The said matrix is presented in the table below.

Table 1. Confusion Matrix.

		Actual	
Predicted	Positives	TP	FP
	Negatives	FN	TN

C.2. Accuracy. Accuracy will be used in order to determine the correct classification. The numerator will contain the correctly classified risk level while the denominator will be the total classified risk levels. Using the confusion matrix, the accuracy can be calculated using TP and TN using (1).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

C.3. Precision. The precision will be used to determine the proportion of the results which are relevant and correctly classified images.

$$Precision = \frac{TP}{TP+TN} \tag{2}$$

C.4. Recall or Sensitivity. Recall or sensitivity will be used to determine the proportion of True Positive and the sum of True Positive and False Positive.

$$Recall = \frac{TP}{TP+FP} \tag{3}$$

C.5. F1-Score. F1- score is used to maximize both precision and recall using the harmonic mean as shown in (4).

$$F1 - Score = \frac{2*Precision*Recall}{Precision + Recall} \tag{4}$$

Presentation and Analysis of Data

This section showcased the results of the experiment in detecting the earthquake damages from satellite image data by employing the gradient boosting algorithm. Table 2 to Table 6 shows the results of classification without the image augmentation and Table 7 to Table 11 deals with the classification with the image.

Table 2. Results of Classification using Gradient Boosting Algorithm without Augmentation (n_estimator 10)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	80	71.43	50
Precision	16	51	48
Recall	14	71	50
F1-Score	23	60	48

Table 3. Results of Classification using Gradient Boosting Algorithm without Augmentation (n_estimator 20)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	40	42.86	70
Precision	16	43	80
Recall	40	43	70
F1-Score	23	43	64

Table 4. Results of Classification using Gradient Boosting Algorithm without Augmentation (n_estimator 30)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	60	85.71	70
Precision	87	88	81
Recall	60	86	70
F1-Score	63	84	67

Table 5. Results of Classification using Gradient Boosting Algorithm without Augmentation (n_estimator 40)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	40	42.86	60
Precision	16	43	60
Recall	14	71	60
F1-Score	23	43	60

Table 6. Results of Classification using Gradient Boosting Algorithm without Augmentation (n_estimator 50)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	60	57.14	30
Precision	36	48	35
Recall	60	57	30
F1-Score	45	52	32

For the first group of data, it was shown that the best model for detecting the earthquake damage is the gradient boosting partitioned in 30:70 for testing and training respectively and best n_estimator is 30 with an accuracy, precision, recall and f1-score of 85.71%, 88%, 86% and 84% respectively as shown in Table 4. Among the models, the less efficient combination is the gradient boosting with 40:60 partition for testing and training and n_estimator of 50 with just 30% accuracy.

Table 7. Results of Classification using Gradient Boosting Algorithm with Augmentation (n_estimator 10)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	85.71	76.19	82.14
Precision	88	82	88
Recall	86	76	86
F1-Score	85	69	85

Table 8. Results of Classification using Gradient Boosting Algorithm with Augmentation (n_estimator 20)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	85.71	100	100
Precision	89	100	100
Recall	86	100	100
F1-Score	85	100	100

Table 9. Results of Classification using Gradient Boosting Algorithm with Augmentation (n_estimator 30)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	100	85.71	78.57
Precision	100	89	79
Recall	100	86	79
F1-Score	100	85	79

Table 10. Results of Classification using Gradient Boosting Algorithm with Augmentation (n_estimator 40)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	100	85.71	100
Precision	100	89	100
Recall	100	86	100
F1-Score	100	86	100

Table 11. Results of Classification using Gradient Boosting Algorithm with Augmentation (n_estimator 50)

Metrics	Partitioning		
	20-80	30-70	40-60
Accuracy	100	71.43	89.29
Precision	100	71	91
Recall	100	71	89
F1-Score	100	71	89

After employing the data augmentation in the satellite images, a significant improvement in the models compared to the first group of data was observed.The accuracy of all the models reached more than 70% in terms of accuracy. It was also recorded that the gradient boosting algorithm with n_estimators of 20 partitioned 30:70 and 40:60 as shown in Table 8, n_estimators of 30 partitioned 20:80 shown in Table 9, n_estimators of 40 partitioned 20:80 and 40:60 as shown in Table 10, and n_estimators of 50 partitioned 20:80 as shown in Table 11 garnered the best classification with accuracy, precision, recall and f1-score of 100%. Thus, based on these results, the best partitioning is 20:80.

Conclusion

This paper employed the Gradient Boosting algorithm to detect the damages brought by earthquakes using the satellite images.The satellite images extracted from Landsat 8 via Google Earth pro before and after the earthquake in Eastern Samar with 7.6 magnitude on August 31, 2012, were used as the first group of dataset and the same images were augmented to create the second dataset. The features extracted from the two datasets were implemented in a gradient boosting algorithm to detect the earthquake damages. Both datasets were partitioned into 20:80, 30:70 and 40:60 for testing and training data and all categories fed in the GBA with n_estimators of 10, 20, 30, 40 and 50. The result shows that data augmentation provided a better output when the models have n_estimators of 20 partitioned in 30:70 and 40:60, n_estimators of 30 partitioned in 2:80, n_etimators of 40 partitioned in 20:80 and 40:60 and n_etimators of 50 partitioned in 20:80 with and accuracy, precision, recall and f1-score of 100%. This study concluded that Gradient Boosting is an efficient and powerful classification algorithm in detecting the damages brought by earthquakes when satellite image data were used as input.

For future works, the preprocessing and classification techniques implemented in this study can also be used in other fields such as handwriting recognition, medical images classification and other fields involving image analysis.

Acknowledgements

The author would like to thank the University of Santo Tomas for funding this research.

Conflict of Interest

The author declares that no potential conflict of interest exists related to this article.

References

[1] Kavanagh, J., & Leung, R. (2018). Volcanoes: A Folding Pocket Guide to Volcanoes, Earthquakes, Hot Springs, Geysers & More (Earth, Space and Culture) (Second ed.). Waterford Press.

[2] Philippines: Disaster Management Reference Handbook (March 2018) - Philippines. (2018, March 30). ReliefWeb. https://webcache.googleusercontent.com/search?q=cache:pvzngTylj1sJ:https://reliefweb.int/report/philippines/philippines-disaster-management-reference-handbook-march-2018+&cd=1&hl=en&ct=clnk&gl=ph

[3] Introduction to Earthquake. (2021). https://www.phivolcs.dost.gov.ph/index.php/earthquake/introduction-to-earthquake

[4] Hilotin, J., & Llata, S. D. (2021, February 7). Deadliest earthquakes in the Philippines: What you need to know. Special-Reports – Gulf News. https://gulfnews.com/special-reports/deadliest-earthquakes-in-the-philippines-what-you-need-to-know-1.1612705914449

[5] Maso, M. S. (2017). Catalogue of Violent and Destructive Earthquakes in the Philippines. CreateSpace Independent Publishing Platform.

[6] National Disaster Risk Reduction and Management Plan (NDRRMP) 2020□2030. (2020). Https://Ndrrmc.Gov.Ph. https://ndrrmc.gov.ph/attachments/article/4147/NDRRMP-Pre-Publication-Copy-v2.pdf

[7] Kolkman, D. (2020). The usefulness of algorithmic models in policy making. Government Information Quarterly, 37(3), 101488. https://doi.org/10.1016/j.giq.2020.101488

[8] Hagen, L., Keller, T. E., Yerden, X., & Luna-Reyes, L. F. (2019). Open data visualizations and analytics as tools for policy-making. Government Information Quarterly, 36(4), 101387. https://doi.org/10.1016/j.giq.2019.06.004

[9] Toh, T. S., Dondelinger, F., & Wang, D. (2019). Looking beyond the hype: Applied AI and machine learning in translational medicine. EBioMedicine, 47, 607–615. https://doi.org/10.1016/j.ebiom.2019.08.027

[10] Statista. (2020, December 21). Risk index for natural disasters Philippines 2020 by type. https://www.statista.com/statistics/921036/philippines-risk-index-for-natural-disasters

[11] NDRRMC (2012). https://ndrrmc.gov.ph/attachments/article/2081/SitRep_No_08_re_Effects_of_Earthquake_in_Guiuan_Eastern_Samar_as_of_06SEP2012_0600H.pdf

[12] Aali, H., Sharifi, A., & Malian, A. (2019). EARTHQUAKE DAMAGE DETECTION USING SATELLITE IMAGES (CASE STUDY: SARPOL-ZAHAB EARTHQUAKE). ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W18, 1–5. https://doi.org/10.5194/isprs-archives-xlii-4-w18-1-2019

[13] C. Li, X. Yuan, J. Zhang, P. Du, L. Mi and Z. Li, “Earthquake Damage Monitoring and Assessment Based on High-Resolution Remote Sensing Images-Take Lushan Earthquake as an Example,” 2018 26th International Conference on Geoinformatics, Kunming, 2018, pp. 1-4, doi: 10.1109/GEOINFORMATICS.2018.8557152.

[14] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7. https://doi.org/10.3389/fnbot.2013.00021

[15] R. Umemura, T. Samura and K. Tadamura, “An efficient orthorectification of a satellite SAR image used for monitoring occurrence of disaster,” 2018 International Workshop on Advanced Image Technology (IWAIT), Chiang Mai, 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369771.

[16] University of Iowa. (2019, June 14). Satellite observations improve earthquake monitoring, response. ScienceDaily. Retrieved May 20, 2021 from www.sciencedaily.com/releases/2019/06/190614125848.html

[17] X. Chen, J. Li, Y. Zhang, W. Jiang, L. Tao and W. Shen, “Evidential Fusion Based Technique for Detecting Landslide Barrier Lakes From Cloud-Covered Remote Sensing Images,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 5, pp. 1742-1757, May 2017, doi: 10.1109/JSTARS.2017.2665529.

[18] Xiaonan Niu, Hong Tang, Lixin Wu, “Satellite scheduling of large areal tasks for rapid response to natural disaster using a multi-objective genetic algorithm”, International Journal of Disaster Risk Reduction, Volume 28, 2018, Pages 813-825, ISSN 2212-4209,https://doi.org/10.1016/j.ijdrr.2018.02.013.

[19] Zhang, R., Duan, K., You, S., Wang, F., & Tan, S. (2020). A novel remote sensing detection method for buildings damaged by earthquake based on multiscale adaptive multiple feature fusion. Geomatics, Natural Hazards and Risk, 11(1), 1912–1938. https://doi.org/10.1080/19475705.2020.1818637

[20] Danandeh Mehr, A. (2021). Drought classification using gradient boosting decision tree. Acta Geophysica. Published. https://doi.org/10.1007/s11600-021-00584-8

[21] Pigeon, & Duval. (2019). Individual Loss Reserving Using a Gradient Boosting-Based Approach. Risks, 7(3), 79. https://doi.org/10.3390/risks7030079

[22] Alcolea, A., & Resano, J. (2021). FPGA Accelerator for Gradient Boosting Decision Trees. Electronics, 10(3), 314. https://doi.org/10.3390/electronics10030314

[23] Körner, P., Kronenberg, R., Genzel, S., & Bernhofer, C. (2018). Introducing Gradient Boosting as a universal gap filling tool for meteorological time series. Meteorologische Zeitschrift, 27(5), 369–376. https://doi.org/10.1127/metz/2018/0908

[24] Kim, S., Son, M., Kwon, K., & Lee, D. (2019). Electricity Demand Forecasting Algorithm Development Based on the Gradient Boosting Machine by Selecting the Optimal Combination of Weather Data and Stations. The Transactions of The Korean Institute of Electrical Engineers, 68(7), 834–843. https://doi.org/10.5370/kiee.2019.68.7.834

[25] Tan, P. N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Introduction to Data Mining. Pearson.

[26] Brownlee, J. (2020, August 15). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. Machine Learning Mastery. https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/

[27] Johnston, B., & Mathur, I. (2019). Applied Supervised Learning with Python: Use scikit-learn to build predictive models from real-world datasets and prepare yourself for the future of machine learning. Packt Publishing.