Integrating satellite imagery and Machine Learning for infrastructure risk forecasting in Honduras

Gefer Haciel Hernandez Romero

Gefer Haciel Hernandez Romero, Óbuda Egyetem, Bécsi út 94-96, 1034, Budapest, Hungary, gefer.hernandez@stud.uni-obuda.hu

Abstract: Honduras and neighboring countries are known vulnerable territories to extreme weathers in terms of rainfalls and irregular areas with dense vegetaion, these factors along with the scarce access to studies, lack of economic resources and tabulated data makes it of particular interest to develop a method that helps assess risk infrastructure in areas of difficult access. This study integrates satellite imagery, AI and machine learning into training a model that helps establish an index of failure for given regions in Honduras, based on weather features and terrain characteristics that include vegetation indexes compared to real data about outages in a given period of time. With AUC = 0.857 this study shows a promising correlation between irregularity of terrain, temperature among other factors and power outages, however its Cross-validation AUC for this model 0.638 ± 0.229 highlights the need for further research with different training methods, more samples and potentially more features from the data sources.

Keywords: Machine learning; AI; Honduras; Power outage

1 Introduction

1.1 Background and problem

Honduras is a latin american country located in Central America, in an area that is well known to be vulnerable to extreme weather conditions that include prolonged precipitations and hurricanes [1, 2], including extreme temperatures. All these features combined with limited resources, makes the area particularly prone to outages in services that can't be addressed right away, affecting the economy and the quality of life of its inhabitants. This study is intended to serve as a basis for power outage analysis based on weather and terrain data that can help the authorities and any interested parties to address areas with potential for power outages before they happen, and this analysis can be extended to many other hazarous events in terms of infrastructure that can potentially pose a risk on the lives of local communities. For countries like Honduras or any other entity involved, this information would allow a more efficient allocation of resources that would result in less outages, and in general a better quality of life.

In Honduras, nearly 60% of outages correspond to scheduled maintenance reports versus the total gross number of outages, representing around 70% of the time offline, leaving a gap of 40% of outages that are unexpected which account for around 30% of the time offline of the electric grid [3, 4]. This highlights the importance of determining a helper predictor which could serve the interests of the industry in terms of better planning and the government for better allocation of resources.

1.2 Data sources

In order to conduct the study and develop a trained model, real outages data was needed, which was obtained from the official website of the National Company of Electrical Energy (ENEE, acronym in Spanish) of Honduras with the National Dispatch Center (CND, acronym in Spanish) [3]. Reports are available to be downloaded from the public, and this data was compared against certain features that were obtained from freely available data from the mission Sentinel-2A and Sentinel 2B with Copernicus for terrain data, which provides up to 10 meters in spatial resolution [5]. It allows the extration of certain bands that can help determine the physical composition of the terrain, which include: moisture, urban infrastructure, vegetation density. Elevations are also important in the sense that they allow us to determine the irregularities of the terrain as these could also represent a contributing factor for a failure index, and this feature was extracted from the Shuttle Radar Topography Mission (SRTM), a jointly project that creates a digital topographic map of the Earth’s land surface [6]. Additionally, weather data was used as well, and ERA5 from ECMWF reanalysis, altogether with TerraClimate was the source of this information. Google Earth Engine was the platform used to gather and download all the data from the different services already mentioned.

1.3 Scope

The model presented in this study was trained using features from Honduras specifically for the power outages index determination, with completely freely available data for academic purposes. The power lines of study was the medium voltage lines of 34.5 kV. Due to the similarities with neighboring regions, the same study could be conducted in surrounding areas with potentially similar results, and potentially for different magnitudes of voltages, or even different types of infrastructure that are under the stress of weather or terrain features and that could be subject to outages. Additionally, this type of study has potential to be used on any country that be interested in establishing risk indexes for transportation network or to evaluate any other service subject to different conditions too.

2 Literature review

2.1 Remote Sensing for infrastructure monitoring

Risk forecasting through remote sensing has been used already by extracting data from the Sentinel-2 mission along with additional imaginery such as from Landsat and MODIS, where different bands are used to determine reflectance indexes that are associated to the humidy of the soil [5, 7]. Sentinel-2a and Landsat8 are „the most advanced satellites with freely available data for long-term high-frequency remote sensing applications" [8]. Some of the bands that can be collected are shown in the following table.

Table 1. Characteristics of ESA Sentinel-2A and 2B [9]

		Sentinel-2A		Sentinel-2B	
Spatial Resolution (m)	**Bands**	**Central wavelength (nm)**	**Bandwidth (nm)**	**Central Wavelength (nm)**	**Bandwidth (nm)**
10	Band 2—Blue	492.4	66	492.1	66
	Band 3—Green	559.8	36	559	36
	Band 4—Red	664.6	31	664.9	31
	Band 8—NIR	832.8	106	832.9	106

A major example in the implementation of remote sensing for infrastructure risk and monitoring of satellite data to monitor and evaluate the September 2019 floods in Eastern Spain from the 9th to the 3th, causing severe floodings in populated areas, where four different prediction systems were available to the forecasters in the actual, real time management for the storm: „two regional models for the short range (HARMONIE-AROME and γ-SREPS) and the HRES-IFS and the ENS-IFS ensemble for the medium range" [10], where the systems accurately forecasted the severe floodings episodes and the affected areas.

2.2 Machine Learning models for this study

For this study data was tabulated, and due to the low amount of samples available, models such as XGBoost are more appropriate. Other models such as eXtreme, Random Forest were used as well and evaluated against each other to determine the best model for our data, using the AUC (Area under the curve) parameter.

In general, AUC can be described as follows, in terms of the False Positive Rate (FPR) or also known as alfa (α), or the probability that the algorithm classifies a negative instance as positive:

$$\alpha = P(FP) = \frac{FP}{FP + TN} \tag{1}$$

Where:

FP is the number of False Positives and TN is the number of True Negatives. Alfa is, in simpler wording, „out of all the actual negative cases, what fraction of them is incorrect?".

3 Methodology

3.1 Data acquisition and pre-processing

3.1.1 Primary raw data and pre-procesing for outages

The data used to train the model was collected from January to May of 2025, and it corresponded to the circuit CYG L325, which is a primary line of medium-low voltage line of 34.5 kV and a valid point of reference is the main substation from which it derives, it corresponds to, in decimal coordinates, in format latitude, longitude: 14.446609, -87.636291. The available data contained the starting and ending date and time of the documented outages and names of areas, not the coordinates of the device that was triggered. Therefore, an algorithm that would request such coordinates from AI services was created, and for this study Google Geocoding API was used and the exported KML files were visually inspected and corrected. Therefore an algorithm that helped determine the closest point to the substation that also belonged to the circuit CYG L325 was chosen, simulating a close point to the failure point. In this ocassion, Microsoft Excel VBA[15] was used to develop and run such script.

[15] VBA is Visual Basic for Applications. Microsoft Excel is a registered trademark of Microsoft Corporation.

3.1.2 Primary raw data and pre-procesing for outages

The use of a multi-sensor approach was necessary to consider as many bands as the sample allowed it. With a buffer or radious of 500 meters around the specified point of failure, considered for outage and non-outage, the information was extracted from three separate sensors:

Optical images: Mission Sentinel-2A from Copernicus. used to dentify the type of vegetation or features of the soil. For each point and period, there was a consideration of ±45 days in case that in the previous 7 days there was no usable imaginery found (prolonged cloudy regions).

Topographic data: DEM model: Stands for Digital Elevation Model, with a resolution of 30 meters, it obtains the elevation in absolute values above the sea level (meters) and slope of the place.

Meteorological Data: Parameters are extracted from a meteorological standpoint to study a potential correlation between the occurrence of meteorological adverse events and the disruption of power service derived from failure of structures, using ERA5. Alternatively, Global Forecast System (GLS) was used when ERA5 data was not available.

Data was created of non-outages following the same locations but in different dates and times that were not considered outages.

3.2 Features selected

Based on the amount of data, the following features were selected: B11_mean, B11_max, B12_stdDev, B2_stdDev, B4_max, NDBI_mean, NBR_max, NDWI_stdDev, NDWI_max, temperature_max_max, temperature_min_max and temperature_min_stdDev.

3.3 Results and analysis

A comparison was established between five models and the one that had the highest AUC value was the choosen one. With the highest AUC-ROC score of 0.8571, XGBoost was selected as the best model. However, from cross-validation we obtain a AUC-Test value of 0.6375 (±0.229) which would be a more realistic number for this model, based on the limited dataset (n=39). The model shows some discriminative ability, but more training data is required to improve the stability of it. The following Table 2 shows the performance of the model for the test dataset.

The model achieved a test accuracy of 70%. Even when the AUC score is high, the precision 0.5 highlights a tendency to false positives, given that half of the predicted values were false alarms. The recall of 0.667 for outages demonstrates that the model identified 66.7% of the actual outage events, which is reasonably valuable for risk management.

The significant difference between AUC and cross validation AUC represents a weak predictive power in some scenarios, which is compatible with the precision being 0.5, therefore there his a high dependency upon the dataset used for training, and in the future more training data would be advisable.

Metric	Value	Interpretation
AUC-ROC	0.857	Good discriminative ability
Accuracy	0.7	70% overall correct classification
Precision	0.5	50% of predicted outages were correct
Recall	0.667	67% of actual outages detected
F1-Score	0.571	Balanced measure of precision and recall
Cross-validation AUC	0.638 ± 0.229	Moderate performance with high variability

Table 2. XGBoost Model Performance on Test Dataset.

B12_stdDev accounts for the most influential factor in predicting outages accounting for 20.4%, which is the variability of the SWIR-2 band, and it encloses features such as terrain heteroegenity and moisture patterns. Extreme temperatures follow accounting for 17.2% of predictive power.

Summary

Honduras is located in a region prone to extreme and uncertain weather conditions, where irregular terrain is predominant. Due to lack of resources and difficult access to locations, outages are of common occurrence, causing disruptions in the economoy and local communities. Therefore, this study aims to establishing a risk index model to improve the allocation of resources, helping transition from a reactive model to a more proactive, smart one, using freely available data, Machine Learning and AI to automate and build such index. This model is a proof of concept that can be extended to many other regions to assess risk infrastructure.

Acknowledgments

This work was supported by Miss Borsos Döníz & Mr. Ady László, for whom no words could make any justice on my gratitude to their continuous support. I thank Obuda University for providing the opportunity. Tudományos Diákköri Konferencián

References

[1] N. CEPAL, "Evaluación de los efectos e impactos causados por la tormenta tropical Eta y el huracán Iota en Honduras," 2021.

[2] D. H. V. Rojas and S. M. R. Corea, "Desastres y sufrimiento: Temporada de huracanes en Colombia y Honduras," Revista de Estudios Latinoamericanos sobre Reducción del Riesgo de Desastres REDER, vol. 9, no. 1, pp. 97-110, 2025.

[3] C. N. d. Despacho. "Informe de fallas." Centro Nacional de Despacho - ENEE. https://cnd.enee.hn/informe-de-fallas/ (accessed 2025, 2025).

[4] U. t. d. C. d. Distribución. "Mantenimientos Programados." ENEE. https://www.eneeutcd.hn/es/mantenimientos-programados-NTC249 (accessed November 5, 2025).

[5] F. Spoto et al., "Overview of sentinel-2," in 2012 IEEE international geoscience and remote sensing symposium, 2012: IEEE, pp. 1707-1710.

[6] T. G. Farr et al., "The shuttle radar topography mission," Reviews of geophysics, vol. 45, no. 2, 2007.

[7] G. Misra, F. Cawkwell, and A. Wingler, "Status of phenological research using Sentinel-2 data: A review," Remote Sensing, vol. 12, no. 17, p. 2760, 2020.

[8] T. Zhang, J. Su, C. Liu, W.-H. Chen, H. Liu, and G. Liu, "Band selection in sentinel-2 satellite for agriculture applications," in 2017 23rd international conference on automation and computing (ICAC), 2017: IEEE, pp. 1-6.

[9] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," Remote sensing, vol. 12, no. 14, p. 2291, 2020.

[10] F. J. Tapiador et al., "The September 2019 floods in Spain: An example of the utility of satellite data for the analysis of extreme hydrometeorological events," Atmospheric Research, vol. 257, p. 105588, 2021/08/10/ 2021, doi: https://doi.org/10.1016/j.atmosres.2021.105588.