Backup Generator Use Prediction using Geo-Spatial ML

Muhammad Abdul Rahman

University of Pennsylvania, School of Social Policy & Practice

MUSA 6950: AI for Urban Sustainability

Instructor: Dr. Xiaojiang Li

Final Project Report

Introduction

Natural disasters are known to wreak havoc in the regions where they occur (Botzen et al., 2019). Beyond the immense infrastructure damage and tragic loss of life, they often lead to widespread power outages, and because electricity is so crucial to daily life, its disruption severely affects everyday functioning (FEMA, 2025). Some localities, however, are noticeably better prepared for these events than others (Chandra et al., 2011). With the help of backup power generators, these communities are partially shielded from some of the negative consequences of outages.

Using geospatial analysis and machine learning prediction algorithms, we seek:

1. Test the predictive power of socioeconomic and other variables for identifying locations with more generator use.
2. Profile the sociodemographic characteristics of more resilient areas.

Power Loss Methodology

For this case study, we examined Hurricane Ida, which hit the New Orleans region in August 2021, and where Moran et al. (2022) report considerable post-storm backup generator activity. Using NASA's VIIRS Black Marble imagery's nighttime light intensity band as a proxy for power availability within each grid cell, we first estimated the change in light intensity before and after the storm. Because the imagery is available at approximately 15-arc-second resolution (Román et al., 2021), our analysis was conducted entirely at the gridded level, both for measuring outage impacts and for building predictive models.

Power loss in each grid cell is defined as follows, where the pre-event data in Fig. 1 was derived from the VNP46A3 monthly composite for August 9, 2021, and the post-event data in Fig. 2 was derived from the VNP46A2 daily product for August 31, 2021, with lighter shades indicating higher values and darker shades indicating lower values across all images:

$$Power\ loss = \frac{Pre\text{-}event\ intensity - Post\text{-}event\ intensity}{Pre\text{-}event\ intensity}$$

Fig. 1 Pre-event data (Aug 9, $nW \cdot cm^{-2} \cdot sr^{-1}$)

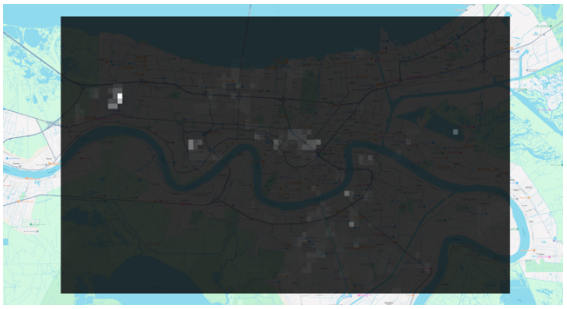

Fig. 2 Post-event data (Aug 31, $nW \cdot cm^{-2} \cdot sr^{-1}$)

Prediction Methodology

We used a Random Forest (RF) classifier to predict whether each grid cell contained a backup generator. The dependent variable was derived from the *Power loss* measure: cells with *Power loss* < 1 were coded as $Y = 0$ (backup generator present), while all cells with *Power loss* $= 1$ were coded as $Y = 1$ (no backup generator).

To assess the relative importance of each predictor, we used permutation-based importance, computed over 100 RF replicates, and summarized the resulting variable inclusion proportions for the following grid-level features:

1. Total building area
2. Proportion of nonresidential buildings
3. Social Vulnerability Index (SVI)
4. Proportion of the black population
5. Light intensity pre-event
6. Elevation
7. August precipitation average

Our datasets were available at different spatial resolutions, so we rescaled all layers to a common grid. Specifically, we interpolated the rasters to an approximately 1 km × 1 km resolution, using bilinear or sum interpolation depending on the dataset type. We then filtered the data using a built-up mask and removed pixels with missing (NaN) values. The final

dataset contained 727 observations with 8 variables: 7 predictors and 1 outcome, with an 80 percent training and 20 percent test split.

Data Sources

The building area and nonresidential building proportion dataset were derived from the work of F. de Arruda et al. (2024), which used building polygons and OpenStreetMap (OSM) data to classify structures as residential or non-residential. We replicated their code for our study area and computed, for each grid cell, the total built-up area and proportion of non-residential buildings, as seen in Fig. 3 and Fig. 4 below, respectively.

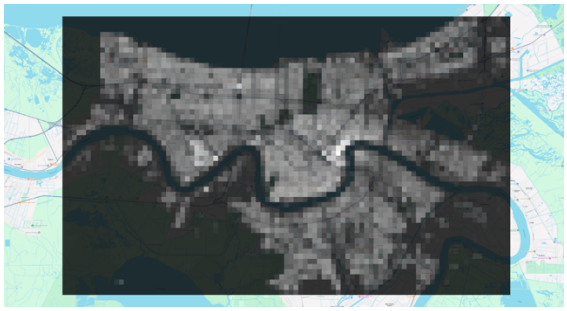

Fig. 3 Total Built Up Area (m^2)

Fig. 4 Proportion of nonresidential buildings (0 – 1)

The SVI variable, which can be seen in Fig. 5, was obtained from NASA's U.S. Social Vulnerability Index Grids, Revision 01 dataset. This product provides gridded layers for the Centers for Disease Control and Prevention (CDC) SVI using four thematic components (Socioeconomic, Household Composition & Disability, Minority Status & Language, and Housing Type & Transportation), constructed from 15 tract-level input variables for the years 2000, 2010, 2014, 2016, 2018, and 2020 (CIESIN-Columbia University, 2023).

The proportion of the Black population in Fig. 6 was derived from NASA's U.S. Census Grids (Summary File 1), 2010 dataset, which contains gridded demographic and socioeconomic variables derived from census 2010 block geography (TIGER/Line Files) and associated census counts of population, households, and housing characteristics (CIESIN-Columbia University, 2017).

Fig. 5 SVI (0 least, 1 most)

Fig. 6 Proportion Black Population (0-1)

Pre-event nighttime light intensity, as described above, was measured using VIIRS Black Marble imagery, while elevation (Fig. 7) was derived from U.S. Geological Survey (USGS) elevation data (U.S. Geological Survey, 2025). The August 2021 precipitation average shown in Fig. 8 was obtained from the CHELSA Monthly precipitation dataset, "a global, kilometer-scale climate dataset generated using the CHELSA downscaling model" (Karger et al., 2025).

Fig. 7 Elevation (m)

Fig. 8 Mean Precipitation Aug (kg/m^2)

Results

The confusion matrix in Fig. 9 shows that the model correctly classified 76 "No Use" pixels and 40 "Use" pixels in the test dataset, indicating a broadly balanced performance across the two classes. Overall, the model achieved an accuracy of 80% and an F1 score of 84%. Given these strong metrics, we are confident that the power loss signal and the subsequently derived dependent variable captured meaningful patterns rather than being driven primarily by noise.

The variable inclusion proportion scores in Fig. 10 are also informative. They fall into three broad tiers: August precipitation average, pre-event light intensity, and proportion of

Black population form the highest-inclusion group (0.98–1.00); total building area, elevation, and SVI comprise a second tier with consistently high inclusion (0.90–0.94); and nonresidential building proportion appears as a lower-inclusion feature (0.61). Overall, these scores suggest that, with the exception of nonresidential building proportion, most predictors contribute meaningfully to the model's performance.

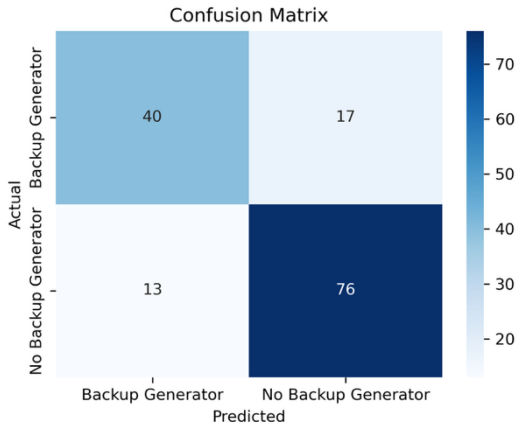

Fig. 9 Confusion Matrix

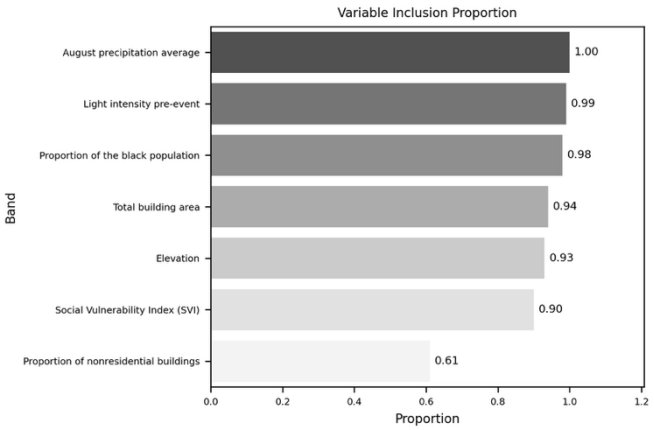

Fig. 10 Variable Inclusion proportion

The correlation graph in Fig. 11 is particularly informative because it shows how each feature relates to no backup generator use. The nonresidential building proportion has a very small correlation (0.03), which suggests a weak relationship with the outcome, consistent with Fig. 10. SVI shows a slightly higher positive correlation (0.09), which is intuitive: more socially vulnerable communities may be less able to invest in backup generators.

Elevation (0.06) is also plausible, since higher-elevation areas may face lower storm-related disruption from reduced water accumulation, which could translate into decreased reliance on backup generators. Total building area (-0.24) aligns with expectations, since larger built-up areas likely reflect greater population and more opportunities for generator installation. Similarly, pre-event light intensity (-0.25) may proxy stronger local economic conditions and thus higher capacity to adopt backup power.

August precipitation (-0.07) appears slightly counterintuitive at first but may reflect better preparedness in regions more accustomed to heavy rainfall and storm risks. The most

striking result is the positive correlation for the proportion of Black population (0.12), implying that areas with higher Black population shares are less likely to have backup generators; taken together with this variable's high inclusion in the model, this pattern suggests meaningful social and economic disparities in backup generator access and use.

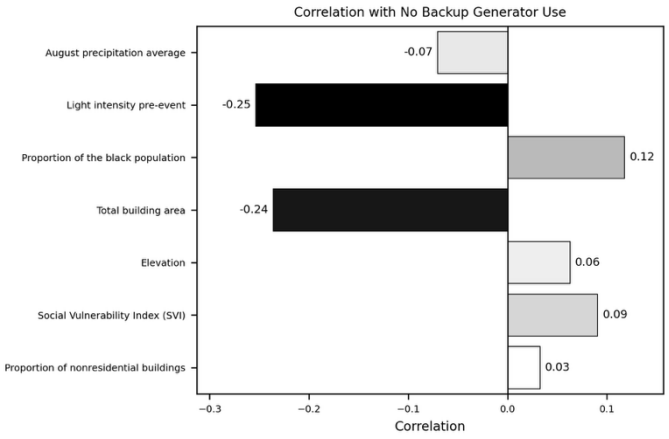

Fig. 11 Correlation values

Conclusion

Natural disasters such as hurricanes can severely damage the electric grid, producing widespread power loss that disrupts daily life and recovery efforts. In this study of Hurricane Ida, we used VIIRS Black Marble-derived power loss as a proxy for outage impact and applied a Random Forest classifier to predict backup generator presence at the grid-cell level using a mix of socioeconomic, built-environment, and environmental predictors.

Our results suggest that post-storm backup generator use is not random; rather, it is systematically associated with local conditions and can be meaningfully predicted from the features included in our model.

Most importantly, the prominence of the Social Vulnerability Index (SVI) and the proportion of the Black population among high-inclusion predictors, coupled with their positive correlations with no backup generator use, points to a clear equity dimension in backup generator access, indicating that resilience to extreme events is shaped not only by hazard exposure but also by underlying social and economic disparities. These findings underscore the need for targeted preparedness and resilience policies that prioritize communities facing structural disadvantages in energy security and disaster recovery.

References

Botzen, W. J., Deschenes, O., & Sanders, M. (2019). The economic impacts of natural disasters: A review of Models and empirical studies. Review of Environmental Economics and Policy, 13(2), 167–188. https://doi.org/10.1093/reep/rez004

Chandra, A., Acosta, J., Howard, S., Uscher-Pines, L., Williams, M., Yeung, D., Garnett, J., & Meredith, L. S. (2011, March 1). Building Community Resilience to disasters: A way forward to enhance National Health Security. Rand health quarterly. https://pmc.ncbi.nlm.nih.gov/articles/PMC4945213

CIESIN-Columbia University. (2017). U.S. Census Grids (Summary File 1), 2010 (Version 1.00) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H40Z716C

CIESIN-Columbia University. (2023). U.S. Social Vulnerability Index Grids, Revision 01 (Version 1.01) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/2EV1-G103

F. de Arruda, H., Reia, S. M., Ruan, S., Atwal, K. S., Kavak, H., Anderson, T., & Pfoser, D. (2024). An OpenStreetMap derived building classification dataset for the United States. Scientific Data, 11(1). https://doi.org/10.1038/s41597-024-04046-w

FEMA. (2025). Power outage incident annex: Managing the cascading impacts from a long-term power outage (Response and Recovery Federal Interagency Operational Plans). U.S. Department of Homeland Security. https://www.fema.gov/sites/default/files/documents/fema_incident-annex_power-outage.pdf

Karger, D. N., Brun, P., & Zilker, F. (2025). CHELSA-monthly climate data at high resolution [Data set]. EnviDat. https://doi.org/10.16904/envidat.686

Moran, A., Stevens, J., & Carlowicz, M. (2022b, December 8). Suomi NPP Satellite Observes Power Outages in New Orleans. NASA. https://svs.gsfc.nasa.gov/31211/

Román, M. O., Wang, Z., Shrestha, R., Yao, T., & Kalb, V. (2021). Black Marble user guide (Version 1.2). NASA Goddard Space Flight Center. https://viirsland.gsfc.nasa.gov/PDF/BlackMarbleUserGuide_v1.2_20210421.pdf

U.S. Geological Survey. (2025). Seamless 1 meter Digital Elevation Models (DEMs) - USGS National Map 3DEP Downloadable Data Collection [Data set]. U.S. Geological Survey. https://doi.org/10.5066/P13LJKFS