

## GEOSPATIAL MODELING OF ROADSIDE VEGETATION RISK ON DISTRIBUTION POWER LINES IN CONNECTICUT

Harshana Wedgedara<sup>a</sup>, Chandi Witharana<sup>a</sup>, Durga Joshi<sup>a</sup>, Diego Cerrai<sup>bc</sup>, Robert Fahey<sup>ac</sup>

<sup>a</sup>Department of Natural Resources and the Environment, University of Connecticut, Storrs, CT, USA – (harshana.wedagedara, durga.joshi, chandi.witharana, robert.fahey)@uconn.edu

<sup>b</sup>Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT, USA – diego.cerrai@uconn.edu

<sup>c</sup>Eversource Energy Center, University of Connecticut, Storrs, CT, USA.

**KEY WORDS:** Power outages, Roadside trees, LiDAR, Vegetation management, Geospatial modeling

### ABSTRACT:

Roadside trees cause almost 90% of the power outages in the forested Northeastern US. Management of roadside vegetation risk on electrical infrastructure demands timely and accurate information on forest conditions. Tasking conventional ground-based scouting methods along thousands of kilometers of powerlines in a repeated fashion are labor-/cost-/time-intense. Geospatial and earth observation (EO) technologies serve as cost-effective tools in monitoring, inspecting, and managing utility corridors. EO technologies, from drones, aircraft, to satellites can efficiently acquire information over large areas at regular intervals while probing forest physical structure and health conditions. LiDAR is a useful data stream for modeling terrain conditions and estimation of multiple forest inventory variables that explain the physical structure of the forest. Various EO imagery provides information on bio-physical characteristics of trees that affect forest health at finer granularity. The goal of this study is to combine multiple environmental variables to develop a spatially-explicit vegetation risk model using machine learning algorithms. Some of the key inputs used in our analysis include LiDAR-derived tree-related variables (e.g., tree height, proximity pixels, canopy cover), LiDAR-derived terrain data (slope, aspect, topographic index), soil characteristics, vegetation management data (tree trimming methods), infrastructure data (wire type), and power outages reported from 2005 to 2017 in Connecticut. Findings of this research will be vital in informing vegetation management decision-making processes, which eventually reduce power outages and the cost of utility corridor maintenance.

### 1. INTRODUCTION

Electricity is an essential basic need of modern life (EIA, 2021). Power interruptions cause many compromises in transport, healthcare, communication, national and economical security (NIAC, 2018). Power failures are caused by many factors, such as weather, vegetation patterns, and utility practices (Lindstrom and Hoff, 2020). Most of the tree failures that cause power outages are weather related (Campbell, 2012) and cause non-defective trees to fall on power lines (Finch and Allen, 2001). Kenward and Raja (2014) reported that more than 80% of outages are caused during weather events. The Congressional Research Service (CRS) Report for Congress (Campbell, 2012) states that high winds with precipitation from seasonal storms cause massive disruptions to the electric grid. Hines and others (2009) analyzed the events causing outages from 1984 to 2006 using data available from North American Electric Reliability council (NERC) and showed that almost 44% of the events are weather related. In late August 2011, Hurricane Irene caused power loss for more than 6.5 million customers across 14 states in the US. According to electric utilities in the Northeastern US, roadside trees cause up to 90% of outages during storms (Eversource, 2019). Tropical storms, such as Isaias, Sandy and Irene resulted in massive damage in this region which eventually led to prolonged outages and more than one million customers were affected. Winter storm Alfred and hurricane Irene in 2011 which affected 1.2 million customers, caused extensive damage in the northeastern U.S and the complete restoration of power outages took 9 to 12 days (McGee et al., 2012). In 2012 over 225,000 power outages were caused following hurricane Sandy in Connecticut. Hurricane Isaias caused power outages for more than 2 million customers across states in the Northeast in August 2020. In the same month, Hurricane Laura caused power outages for nearly 400,000 customers (Climate central, 2020). When

focusing on Connecticut only, Isaias caused massive destruction to the power grid in Connecticut. Nearly 675,000 customers lost power and over 21,000 damage locations were reported throughout the state.

Electric utilities make substantial financial investments annually on roadside vegetation management programs to ensure reliable electrical services and to improve the resiliency of the power grid in extreme weather events. According to PURA (2013), Vegetation management is the removal of trees, shrubs, and other vegetation that pose a risk to the utility infrastructure and the retention of trees and shrubs that are compatible with utility infrastructure. In Connecticut, vegetation management is restricted within the utility protection zone (UPZ), which is within 2.5 lateral meters of the power lines from ground to sky (PURA, 2013). The monetary cost associated with tree trimming is considered as the largest cost factor of the total money spent on power distribution system maintenance (Radmer et al., 2002 and Lovlace et al., 1996). Utility companies manage tens of thousands of kilometers of power lines, and their biggest challenge is to determine when and where vegetation may become a risk to power lines. Accurate identification of vegetation risk areas and ideal use of vegetation management interventions alongside other grid hardening programs are crucial for maximizing grid resilience and reliability. Data-driven vegetation management decision making demands spatially explicit and temporally conversant information on roadside forest conditions. Conventional vegetation scouting/inspection methods along utility corridors are labor-/time-intense and cost prohibitive

Geospatial and earth observation (EO) technologies serve as cost-effective tools in monitoring, inspecting, and managing utility corridors. EO technologies, from drones, aircraft, to

satellites can efficiently acquire information over large areas at regular intervals while probing forest physical structure and health conditions. The Light Detection and Ranging (LiDAR) is known as one of the emerging remote sensing technologies which can be used in obtaining terrain models, estimation of multiple resource inventory variables through active sensing of three-dimensional (3D) forest vegetation (Reutebuch, 2005). Using LiDAR, numerous tree variables have been derived such as height and size of individual trees, canopy closure, volume, and biomass of forest stands (Andersen et al., 2005, Hinsley et al., 2006, Means et al., 2000, Naesset, 2002, Persson et al., 2002, Popescu and Zhao, 2007). Moreover, tree height in deciduous forests during leaf-off conditions (Parent and Volin, 2015), canopy rugosity (Parker and Russ, 2004), mapping stem locations and tree crowns (Holopainen et al., 2013), and stem diameter (Kankare et al., 2015; Tanhuanpaa et al., 2014) have been derived using this technology. On top of that, LiDAR can store and review or utilize forest data over time. This will allow LiDAR to monitor changes in tree height and crown radius over a five-year period (Duncanson and Dubayah, 2018).

Understanding what environmental factors contribute to tree failures is critical for developing spatially explicit vegetation risk

models. Confounding factor analysis can primarily be performed at the circuit level, which is at coarser granularity, as well as at the device exposure zone (DEZ) level (the outage locations correspond to isolating devices each of which protect a section of power line called a device exposure zone). Most of the studies have been conducted at the circuit level (Taylor et al., 2022, Zhu et al., 2007). A DEZ level modeling approach will account, in addition to forests related factors, multitudes of other environmental factors, soil and terrain (soil, slope and aspect), utility infrastructure related variables (overhead length, overhead to underground powerline length), and vegetation management practices (e.g., tree trimming and hazard tree removal). From the perspectives of utility industry and as well as the regulatory bodies, the envisioned outcome would be to identify and localize where the “HOT” spots are for tree failures during storms. A finer granularity DEZ-level vegetation risk model is vital in informing vegetation management decision making processes. This will eventually reduce power outages and the cost of utility corridor maintenance. The goal of this study is to combine multiple environmental variables to develop a spatially-explicit vegetation risk model using machine learning algorithms. Also, we are interested in finding the major environmental variables that contribute to tree failures along the power lines.

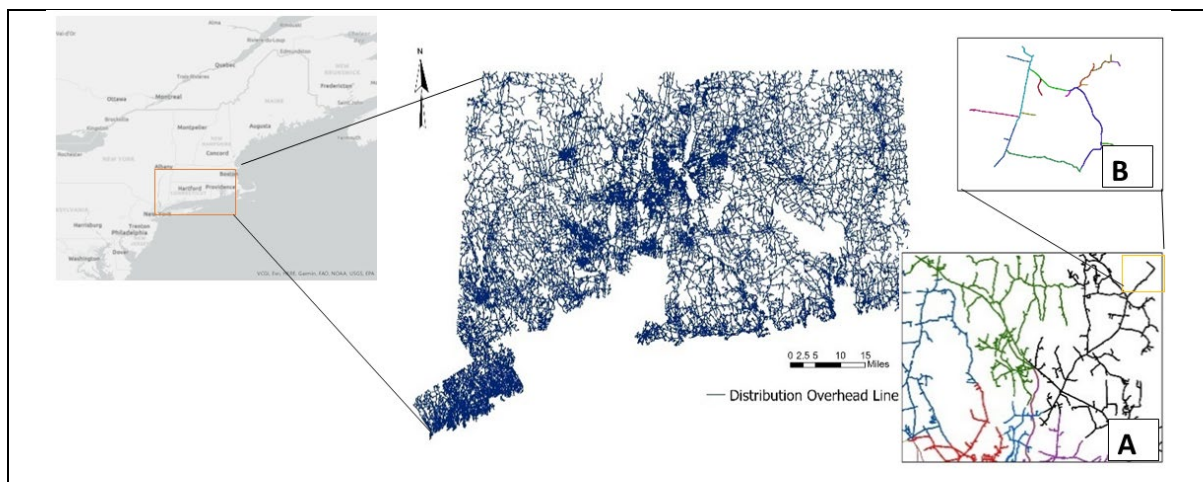


Figure 1. Map showing the distribution power line network (Eversource Energy) for the state of Connecticut. (A) Zoomed-in view of a circuit, (B) Zoomed-in view of a device exposure zone

## 2. METHODS

### 2.1 Study Area

We conducted our analysis in the state of Connecticut. Eversource Energy, the main utility company in the state, manages nearly 27,000 km of distribution lines (Figure 1). Eversource delivers electricity to almost every town in Connecticut, serving ~1.2 million customers (Eversource, 2021). The topography is generally hilly with elevation ranging from sea level in the south to nearly 750 m in the northwest. All the variables and data are derived/reported at the DEZ level. The location of an outage is reported at the associated DEZ isolating device regardless of where the actual cause (tree failure) of the outage occurred within the DEZ. The distribution powerline data set contains approximately 49,000 DEZs within 900 circuits. The length of DEZs range from 100s to 1000s of meters.

### 2.2 Data

**2.2.1 LiDAR-derived tree heights:** We used publicly-available LiDAR data that was acquired in early spring 2016 during leaf-

off condition, (point density 2.2 pts/m<sup>2</sup>) to derive canopy height model (CHM) and successive vegetation parameters along the power lines (Parent and Volin, 2015).

Previous literature has shown that minimum of 4–8 pts/m<sup>2</sup> spatial resolution is required to map individual tree crowns (Evans et al., 2009 and Laes et al., 2008). Therefore, we used the concept of “Proximity tree pixels (pPix)” (Figure 2) explained by “Wanik et al., 2017” to create the LiDAR derived tree parameters described in Table 1. pPix refers to the 1 m pixels in a canopy height model (CHM) that are tall enough and close enough to contact a power line in the event of a whole or partial tree failure. Proximity pixels are identified from the CHM as pixels with a height larger than the distance from the pixel at ground-level of the tree to the point at the location of the nearest power line (Wanik et al., 2017).

### 2.2.2 Power outages, Utility infrastructure, and vegetation data:

We obtained power outage data for the period of 2005–2017 from Eversource Energy. The outage database allowed us to identify the approximate locations of tree-related outages occurred during windstorms. Database includes comprehensive information for each outage (trouble spots) recorded in its

corresponding isolating device, such as date of occurrence, restored date and time, device exposure zone ID, circuit id, number of customers affected, affected pole id, cause of damage, actions taken to repair and geographic locations.

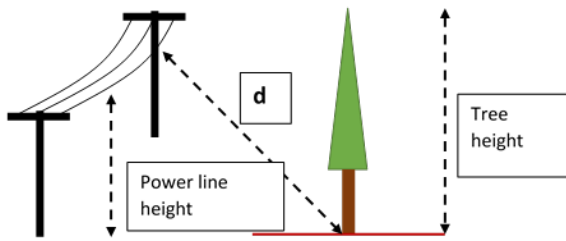


Figure 2. Proximity pixel modelling using LIDAR-based canopy height model (Proximity pixel if tree height > d)

wire type of the conductor (e.g.; aerial cable, large bare wire, large-covered wire, large spacer cable, etc.) and length of underground powerline. We also obtained geospatial layers corresponding to Enhance Tree Trimming (ETT) from Eversource Energy. ETT refers to the removal of above and below that are 8 feet (2.4m) to the side of the power lines and removal of dead, dying, or diseased trees (Eversource, 2019). The data layer consists of several information such as date of trimming year, work type, and object id.

**2.2.3 Secondary products:** We used previously described LiDAR and NLCD (2016) data to develop different potential indicators (as shown in Table 1), such as the percent of proximity pixels (pPix) on steep slopes and percent of pPix with percent canopy cover and topographic indexes. Also, the percentage of pPix on different soil types (such as wetland soils, rocky soils, and shallow soils) were calculated using the USDA Natural Resources Conservation Service (NRCS) soil data.

Table 1. List of explanatory and response variables included in the machine learning model development

Type	Abbreviation	Description	Source	Variable Type
Vegetation	medianH	pPix of median height	LiDAR	Continuous
	pTreeCov	Canopy Cover percentage	NLCD	Continuous
	h50_t	Percentage of pPix with heights higher than 15m	LiDAR	Continuous
	exp15_t	Percentage of pPix that are 4.5m ft taller than surrounding 8nn	LiDAR	Continuous
	Clsr_tot	Percent canopy area around DEZ with tree height higher than 15m	LiDAR	Continuous
Soil and Terrain	wlnd_tot	Percent of pPix on wetland soils	LiDAR	Continuous
	rcky_tot	Percent of pPix on extremely rocky soils	CT Deep	Continuous
	shlw_tot	Percent of pPix on shallow soils (< 40cm to bedrock)	CT Deep	Continuous
	stp_tot	Percent of pPix on steep slopes (>50pct) and with aspect toward DEZ point	LiDAR	Continuous
	avgGrndAz	Average ground azimuth	NLCD	Continuous
	gSlp_mean	Percent of pPix on ground slope (<50pct) and with aspect toward DEZ point	NLCD	Continuous
	TPI150_mn	TPI (z-medainZ) for window approx. 150m in radius, to determine landform	NLCD	Continuous
	TPI450_mn	TPI (z-meanZ) for window approx. 450m in radius, to determine landform	NLCD	Continuous
Infrastructure	coverWire	Overhead lines are covered or not	Eversource	Binary
	UG_ratio	Ratio between overhead and underground power lines	Eversource	Continuous
	length_km	Primary OH Length	Eversource	Continuous
	Class	Presence of outages (Yes/No)	Eversource	Binary
Vegetation Management	pctETT	Percentage of ETT treated in the DEZ	Eversource	Continuous
	SUM_ETTLen	Total Enhanced tree trimming length in the DEZ	Eversource	Continuous

For the modeling, we only included outages, which were caused by trees either as broken stem or windthrow (i.e., Entire tree uproots). Outages were only included if they were tree-related and associated with a primary isolating device, and occurred during weather conditions: blizzard, hurricane, ice storm, rain, snow, thunderstorm, tornado, and high winds. The outage data includes nearly 53,000 tree-related incidents for the time period of interest. All utility infrastructure geospatial layers were acquired from Eversource Energy. The DEZs are represented as polyline geometry, which includes attributes of; conductor segment length, associated isolating device information, major

### 2.3 Correlation Analysis

We used Spearman's rank test to determine the relationship between each explanatory variable and response variable (Equation 1). The Spearman's rank correlation coefficient is generally denoted as  $\rho_s$  for a population parameter and as  $r_s$  for a sample statistic. This method is more appropriate for a dataset with skewed or ordinal data and is robust when extreme values are present. (Mukaka, 2012)

$$r_s = 1 - \frac{6 \sum_{i=2}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

## 2.4 Machine Learning (Random Forest (RF)) Algorithm

The RF classification algorithm is an extension of the classification and regression tree (CART) developed by Breiman and others (1984). This algorithm recursively partitions the training data set into groups of records with similar values for the target variable. The tree is grown by conducting for each decision node, an exhaustive search of all variables and all possible splitting values, selecting the optimal split (LaRose, 2015 and Kennedy, 1995).

RF produces multiple CART-like tree classifiers using random subsets of the training data records and explanatory variables to fit multiple decision trees and this improves the classification performance of a single tree classifier by combining bootstrap aggregating method and randomization during the construction of a decision tree. The predictions from all the decision trees are referred to as the “forest”, and the average of the individual tree predictions in the forest are used as the final prediction of the model (Breiman, 2001). A decision tree with M leaves divides the feature space into M regions  $R_m$ ,  $1 \leq m \leq M$ . For each tree, the prediction function  $f(x)$  can be defined as (Equation 2).

$$f(x) = \sum_{m=1}^M C_m \pi(x, R_m) \quad (2)$$

Where  $M$  = the number of regions in the feature space  
 $R_m$  = a region appropriate to  $m$   
 $c_m$  = a constant suitable to  $m$  (Equation 3)

$$\pi(x, R_m) = \begin{cases} 1, & \text{if } x \in R_m \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

RF is a nonparametric integrated data mining algorithm based on trees. Unlike a single regression tree with high variance and low bias, RF overcomes the problem of high variance by using model average. In addition, when the number of input variables is large, RF has better precision than other classical machine learning algorithms (Hui, 2019). The RF model is used for our modelling effort because previous studies have shown its efficiency and satisfactory performance in outage prediction models and has been widely used in outage prediction models (Nateghi et al., 2013, Wanik et al., 2017, Li et al., 2021a)

## 2.5 Performance estimation

We used repeated stratified 10-fold cross-validation to estimate the performance of the classification model. The repeated stratified k-fold cross-validation procedure is known as a standard method for estimating the performance of machine learning algorithm and provides the improved version of the estimated performance of a machine learning model. The performance was measured using area under receiver operating characteristic curve (AUC-ROC) (Bradley AP., 1997 & Fawcett T., 2006). We also used the confusion matrix of classification model and the accuracy score to summarize and visualize the results. The formula for calculating the accuracy is shown in (Equation 4)

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \quad (4)$$

Where, true-positives (TP) are the number of outcomes where the model correctly predicts the positive class, true-negatives (TN) are the outcomes where the model correctly predicts the negative class, false-positives (FP) are the outcomes where the model incorrectly predicts the positive class, and false negatives (FN) are the outcomes where the model incorrectly predicts the negative classes.

## 2.6 Evaluation of variable importance

Shapley Addictive exPlanations (SHAP) proposed by Lundberg and Lee (2017) was used to evaluate the importance of each predictor variable. The SHAP method is based on the game theoretic approach (Shapley, 2016; Lundberg et al., 2018) and allocates payouts (i.e., importance) among all pairs of features, not only limited to the individual features. This helps to explain the modelling of the local interaction effects (Li et al., 2020b). Partial dependence plots (PDP) (Goldstein et al., 2015) were generated to visualize the average relationship between the response variable and highest contributed important variables.

## 3. RESULTS

### 3.1 Correlation Analysis

Spearman’s correlation among 18 continuous explanatory variables is shown in Figure 3. Tree pPix variables such as Percentage of pPix with heights higher than 15m (h50\_t), Percentage of pPix that are 4.5 ft taller than surrounding (exp15\_t), and Percent canopy area around DEZ with tree height higher than 9m (Clsr\_tot) reported highest correlation between these three variables. Moreover, slightly higher correlation could be observed between percent of pPix on extremely rocky soils (rcky\_tot) and total percent of rocky soils/wetland soils/shallow soils (badSoils). In contrast, other explanatory variables reported less correlation values (Figure 3).

### 3.2 Performance estimation

A confusion matrix for random forest classifier can be used to summarize and visualize the prediction values (Figure 4). It shows the true predicted and false predicted labels for each outage class. Based on test results, the overall accuracy of our RF classifier is 72%. As seen in Figure 4, class 0 is predicted to be better than class 1. Moreover, FP and FN values have shown slightly higher values even though they are less than TP and TN values. The mean area under receiver operating characteristic curve (AUC-ROC) is shown in Figure 5. AUC-ROC score is found to be higher according to the RS classifier which is 0.78.

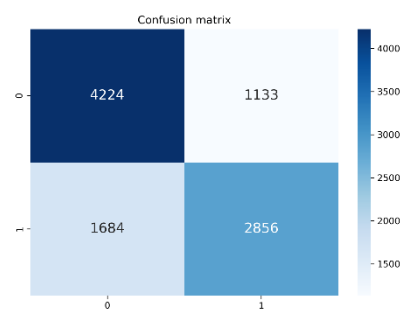


Figure 4. Confusion matrix for RF classifier

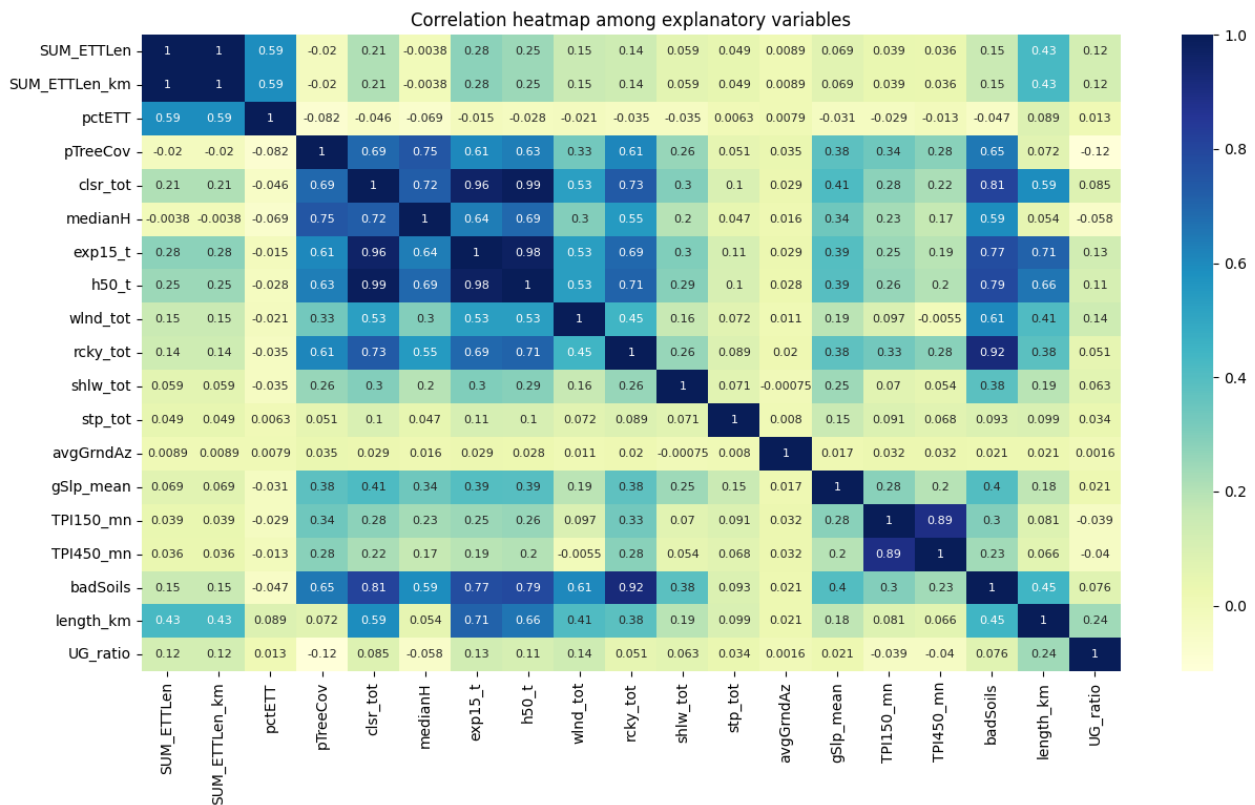


Figure 3. Correlation heatmap among explanatory variables

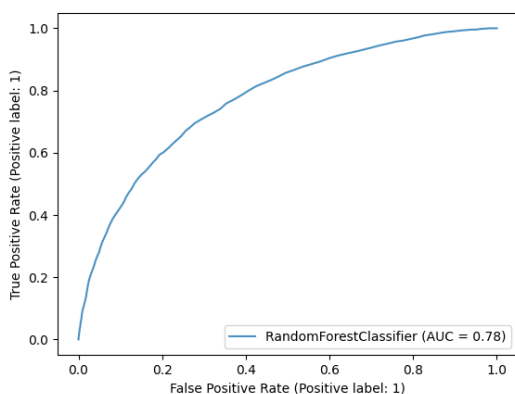


Figure 5. Mean area under receiver operating characteristic curve (AUC-ROC)

Most of the soil and terrain variables (e.g., Topographic index (z-medainZ) for window approx. 150/450m in radius, to determine landform, Average ground azimuth, Percent of pPix on shallow soils, and Percent of pPix on steep slopes (>50pct) and with aspect toward DEZ point) showed very little contribution to the overall model performance. The first eight explanatory variables have shown approximately 90% contribution to the model. Figure 7. Shows the partial dependence plots, which elucidate how the predictions partially depend on the values of the input variables of interest. The explanatory variable of interest was plotted on the X axis and y axis shows the change in response variable. Primary OH Length,

median height of pPix, Canopy Cover percentage, and total Enhanced tree trimming length in the DEZ contributed to increased predicted outages. The percentage of pPix with heights higher than 50 ft (15m) had a nonlinear pattern. Percent of pPix on extremely rocky soils, and percent of pPix on wetland soils showed slight impact on increasing predicted outages.

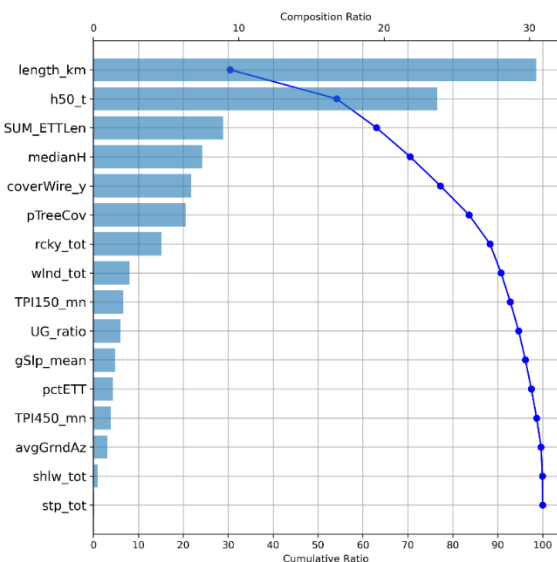


Figure 6. Variable importance with composition ratio and cumulative ratio

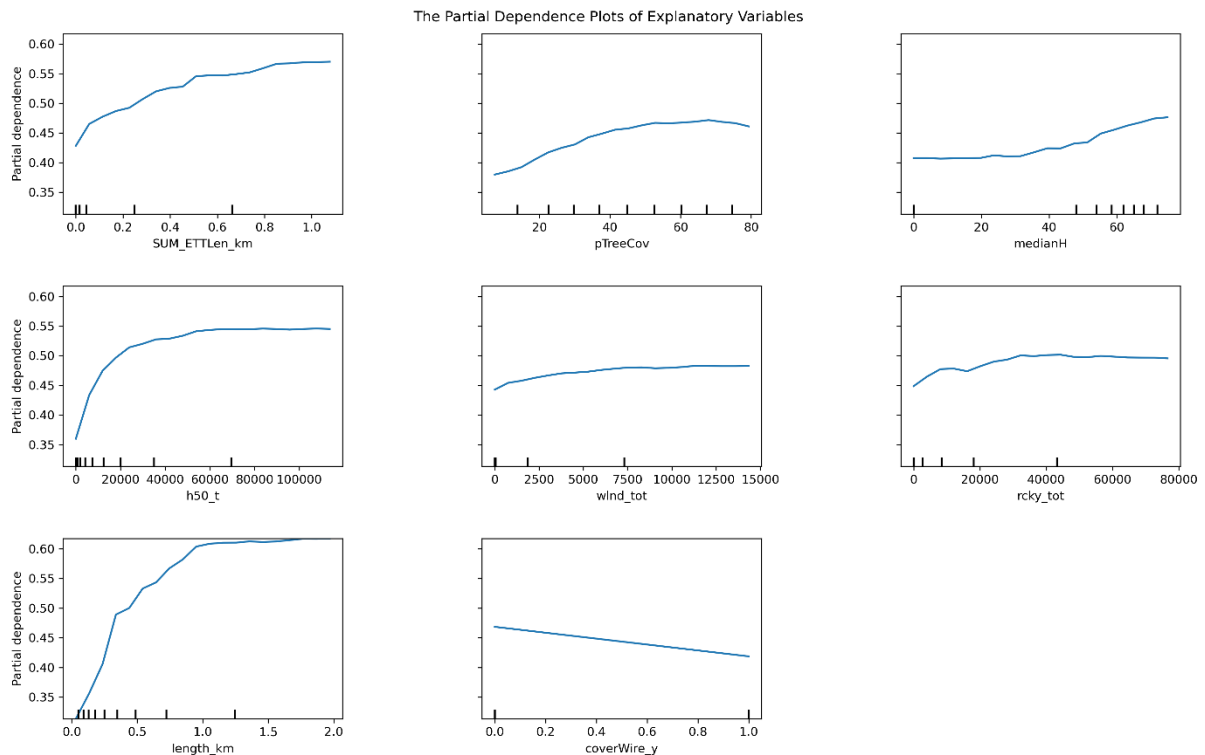


Figure 7. The partial dependence plots for highest importance variables

#### 4. DISCUSSION

The purpose of this study was to develop a geospatial model that can localize the vegetation risk areas along the power lines and to identify confounding environmental variables contributing to tree failures. We used Random Forest (RF) classification algorithm for our analysis. Because RF has been widely used in outage prediction models (Nateghi, 2014 and Wanik et al., 2017) and previous studies have shown the applicability of Random Forest in outage prediction models (Li et al., 2021a). Parr and others (2019) have shown that presence of multicollinearity in the data provides the least meaningful importance values. Therefore, we attempted to find the correlated explanatory variables in the beginning of the analysis and kept one input variable from the highly correlated cluster. Bradley (1997) and Fawcett (2006) have shown the use of area under the receiver operating characteristic curve (AUC-ROC) to evaluate the model performance in machine learning. AUC-ROC can measure the ability of a classifier to distinguish between classes. According to the results of RF model, we can observe that FN and FP numbers show somewhat higher values even though their values do not exceed the TP and TN values. This can affect the accuracy of the overall model prediction. AUC-ROC curve showed plausible values. When AUC-ROC is between 0.5 and 1, there is a high chance that random forest classifier can distinguish the positive classes labels from negative class labels. According to the results of variable importance analysis, we can find that 90% of the model has been explained by 8 variables. Our results are in alignment with the previous studies, such as Wanik et al., (2017) who found utility infrastructure (total overhead length and sum of assets) and proximity pixel variables as highest importance variables. This implies the importance of including both tree related and utility infrastructure data to the outage prediction modeling process. A similar study conducted by Li et al., (2021a) found that geographical/terrain factors, such as

altitude, slope, slope direction, and longitude greatly contributed to the accuracy of the prediction model. However, most soil and terrain variables showed the least significance to the overall model prediction in our study. A possible reason could be, we included power outages occurred due to both tree and limb/branches failures in our analysis. In the effect of tree limbs and branches failures, we can assume that soil and terrain variables have the least impact on determining outages. Another potential reason could be the spatial granularity of the soil data. Soil data layers are generally reported at much coarser resolution which could overlook the local variabilities. Our results elucidated the importance of vegetation management and grid hardening programs in outage prediction and provided motivation to keep investing in such measures to increase the grid reliability and resilience. Tree decay is one of the natural phenomena occurring in trees and 76% of the trunk failures involve a decay (Kane, 2008) and leads to loss in moment capacity of tree branches and stems (Ciftci et al., 2014). There is a higher probability of these unhealthy trees falling onto powerlines during the events of storm and normal weather conditions. Therefore, it is vital to include the tree's health condition as an input variable to the vegetation risk modelling process.

#### 5. CONCLUSION

In this study, we investigated the prediction of power outages on distribution lines using multiple environmental variables in Connecticut. Our analysis was conducted at the device exposure zone level. Utility infrastructure data (Primary overhead length, overhead lines are covered or not), tree variables (Percentage of pPix with heights higher than 15m, canopy cover percentage, median height of proximity pixels) and vegetation management variables (total enhanced tree trimming length in each DEZ) were found to be highly important predictor variables.

Importantly, LiDAR derived variables were shown to be more important than vegetation management and soil/terrain variables.

## ACKNOWLEDGEMENTS

The authors would like to thank the Eversource Energy center at University of Connecticut for all the data and funding.

## REFERENCES

- Andersen, H.E., Mcgaughey, R.I., Reutebuch, S.E., 2005. Estimating canopy fuel parameters using airborne LIDAR data. *Rem. Sens. Environ*, 94, pp. 441-449
- Bradley, A.P., 1997. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recogn*, 30(7), pp. 1145–1159
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45, pp. 5-32.
- Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 2017. Classification and regression trees. Routledge.
- Campbell, R.J., 2012. Weather-related power outages and electric system resilience. Congressional Research Service Available at <http://docs.jalite-group.com/data/docs/pdf/R42696.pdf>, (7 Apr. 2022)
- Ciftci, C., Arwade, S.R., Kane, B., Brena, S.F., 2014. Analysis of the probability of failure for open-grown trees during windstorms. *Probab Eng Mech*, 37, pp. 41–50.
- Climate central, 2020. Power off, Extreme weather and power outages, Climate Central <https://medialibrary.climatecentral.org/resources/power-outages> (04 Apr. 2022)
- Duncanson, L., Dubayah, R., 2018. Monitoring individual tree-based change with airborne lidar. *Ecology and Evolution*. 18, pp. 5079-5089.
- EIA., 2021. Electricity Consumption in the United States was about 3.8 trillion kilowatthours (kWh) in 2020. Electricity explained, Use of electricity, U.S. Energy Information Administration, Washington, DC, USA. <https://www.eia.gov/energyexplained/electricity/use-of-electricity.php> (04 Apr. 2022)
- Evans, J.S., Hudak, A.T., Faux, R., Smith, A.M.S, 2009. Discrete return lidar in natural resources: recommendations for project planning, data processing, and deliverables. *Remote Sens*. 1(4), pp. 776–794.
- Eversource, 2019. Understanding Vegetation Management, Evercourse, CT <https://www.eversource.com/content/ct-c> (04 Apr. 2022)
- Fawcett, T., 2006. An introduction to roc analysis. *Pattern Recogn Lett*, 27(8), pp. 861–874
- Finch, K.E., Allen, C., 2001. Understanding Tree-Caused Outages. EEI Natural Resource Conference. Apr. 2001. Palm Springs, CA. <https://www.eci-consulting.com/wp-content/uploads/2017/10/Understanding-Tree-Caused-Outages.pdf> (04 Apr. 2022)
- Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, 24(1), pp. 44–65.
- Hines, P., Jay, A., Talukdar, S., 2009. Large blackouts in North America: Historical trends and policy implications. *Energy Policy*, 37, pp. 5249-5259.
- Hinsley, S.A., Hill, R.A., Bellamy, P.E., Baltzer, H., 2006. The application of LiDAR in woodland bird ecology: climate, canopy structure and habitat quality. *Photogramm. Eng. Rem. Sens.* 72, pp. 1399-1406.
- Holopainen, M., Kankare, V., Vastaranta, M., Liang, X., Lin, Y., Vaaja, M., Yu, X., Hyyppä, J., Hyyppä, H., Kaartinen, H., 2013. Tree mapping using airborne, terrestrial and mobile laser scanning-A case study in a heterogeneous urban forest. *Urban Forestry & Urban Greening*. 12(4), pp. 546-553
- Hui, J., (2019) Machine Learning-Algorithms. <https://jonathan-hui.medium.com/machine-learning-summary-algorithm-d75c64963800> (5 Apr.2022).
- Kane, B., 2008. Tree Failure following a windstorm in Brewster, Massachusetts, USA. *Urban Forestry and Urban Greening*, 7, pp. 15-23.
- Kankare, V., Liang, X., Vastaranta, M., Holopainen, Yu, M., Hyyppä, J., 2015. Diameter distribution estimation with laser scanning based multisource single tree inventory. *ISPRS J. Photogramm. Remote Sens.*, 108, pp. 161-171.
- Kennedy, R., 1995. *Solving Data Mining Problems Through Pattern Recognition*. Pearson, Upper River Saddle, New Jersey,
- Kenward, A., Raja, U., 2014. Blackout: Extreme weather, climate change and power outages. [https://www.eenews.net/assets/2014/04/14/document\\_ew\\_01.pdf](https://www.eenews.net/assets/2014/04/14/document_ew_01.pdf) (7 Apr. 2022)
- Laes, D., Reutebuch, S., McGaughey, B., Maus, P., Mellin, T., Wilcox, C., Anhold, J., Finco, M., Brewer, K., 2008. Practical Lidar Acquisition Considerations for Forestry Applications, RSAC-0111-BRIEF1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center. Pp. 32.
- Larose, D.T., 2015. *Data mining and predictive analytics*. John Wiley & Sons.
- Li, M., Hou, H., Yu, J., Geng, H., Zhu, L., Huang, Yong., Li, X., 2021a. Prediction of Power Outage Quantity of Distribution Network Users under Typhoon Disaster Based on Random Forest and Important Variables. *Mathematical Problems in Engineering*, 2021, pp. 1-14.
- Li, R., Shinde, A., Liu, A., Glaser, S., Lyou, Y., Yuh, B., Wong, J., Amini, A., 2020b. Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. *JCO Clin. Cancer Inform*. 4, pp. 637–646.
- Lindstrom, A., Hoff, S., 2020. U.S. customers experienced an average of nearly six hours of power interruptions in 2018,

- Today in Energy, U.S. Energy Information Administration, Washington, DC, USA <https://www.eia.gov/todayinenergy/detail.php?id=43915> (04 Apr. 2022)
- Lovlace, W.R., 1996. Vegetation management of distribution line right-of-way: Are you getting top value for your money?, in Proc. 39th Annu. Rural Electric Power Conf., pp. B5-1–B5-3.
- Lundberg, S. M., Lee, S.I., 2017. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 20, pp. 4765–4774.
- Lundberg, S.M., Nair, B., Vavilala, M.S., Horibe, M., Eisses, M.J., Adams, T., Liston, D.E., Low, D.K., Newman, S.F., Kim, J., Lee, S.I., 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng.* 2(10), pp. 749-760.
- Mcgee, J., 2012. Report of the Two Storm Panel, Presented To: Governor Dannel P. Malloy, [http://www.governor.ct.gov/malloy/lib/malloy/two\\_storm\\_panel\\_final\\_report.pdf](http://www.governor.ct.gov/malloy/lib/malloy/two_storm_panel_final_report.pdf) (10 Apr. 2022).
- Means, J., Acker, S., Fitt, B., Renslow, M., Emerson, L., Hendrix, C., 2000. Predicting forest stand characteristics with airborne scanning LIDAR. *Photogramm. Eng. Rem. Sens.* 66, pp. 1367-1371.
- Mukaka M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* 24(3). Pp. 69-71.
- Naeset, E., 2002. Predicting forest stand characteristics with airborne laser using a practical two-stage procedure and field data. *Rem. Sens. Envir.* 80, pp. 88-99.
- Nateghi, R., Guikema, S., Quiring, S.M., 2014. Power outage estimation for tropical cyclones: Improved accuracy with simpler models. *Risk Anal.* 34(6). Pp. 1069–10788.
- NIAC, 2018. The President’s National Infrastructure Advisory Council, Surviving a Catastrophic Power Outage, How to Strengthen the Capabilities of the Nation. USA [https://www.cisa.gov/sites/default/files/publications/NIAC%20Catastrophic%20Power%20Outage%20Study\\_FINAL.pdf](https://www.cisa.gov/sites/default/files/publications/NIAC%20Catastrophic%20Power%20Outage%20Study_FINAL.pdf) (04 Apr. 2022)
- Parent, J.R., Volin, J.C., 2015. Assessing species-level biases in tree heights estimated from terrain-optimized leaf-off airborne laser scanner (ALS) data. *International Journal of Remote Sensing*, 36(10), pp. 2697-2712
- Parker, G.G., Russ, M.E., 2004. The canopy surface and stand development: Assessing forest canopy structure and complexity with near-surface altimetry. *Forest Ecology and Management.* 189, pp. 307-315.
- Parr, T., Turgutlu, K., Csiszar, C., Howard, J., 2018. Beware Default Random Forest Importances. <https://explained.ai/rf-importance/index.html> (06 Apr. 2022).
- Persson, A., Holmgren, J., Soderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogramm. Eng. Rem. Sens.* 68(9): pp. 925-932
- Popescu, S., Zhao, K., 2007. A voxel-based LIDAR method for estimating crown base height for deciduous and pine trees. *Rem. Sens. Environ.* 112(3): pp. 767-781.
- PURA, 2013. PURA Regulations (Public Utility Regulatory Authority) (Title 16 – Public Service Companies). Connecticut Department of Energy and Environmental Protection, <https://law.justia.com/codes/connecticut/2013/title16/chapter-283/section-16-245> (04 Apr. 2022)
- Radmer, D.T., Kuntz, P.A., Christie, R.D., Venkata, S.S., Fletcher, R.H., 2002. Predicting vegetation-related failure rates for overhead distribution feeders. *IEEE Trans. Power Deliv.* 17 (4), pp. 1170–1175.
- Reutebuch, S.E., Andersen, H., McGaughey, R. J., 2005. Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. *Journal of Forestry*, pp. 286-292
- Shapley, L. S., 2016. A value for n-person games. *Contributions to the Theory of Games (AM-28)*, 2, L. S. Shapley, Ed., Princeton University Press, pp. 307–318
- Tanhuanpää, T., Vastaranta, M., Kankare, V., Holopainen, M., Hyypää, J., Hyypää, H., Alho, P., Raisio, J., 2014. Mapping of urban roadside trees – a case study in the tree register update process in Helsinki City. *Urban For. Urban Green*, 13, pp. 562–570.
- Taylor, W.O., Watson, P.L., Cerrai, D., Anagnostou, E.N., 2022. Dynamic modeling of the effects of vegetation management on weather-related power outages. *Electric Power Systems Research*, 207, 107840, <https://doi.org/10.1016/j.epsr.2022.107840>.
- Wanik, D.W., Parent, J.R., Anagnostou, E.N., Hartman, B.M., 2017. Using vegetation management and LiDAR-derived tree height data to improve outage predictions for electric utilities. *Electric Power Systems Research*, 146, pp. 236-245.
- Zhu, D., Cheng, D., Broadwater, R. P., Scirbona, C., 2007. Storm modeling for prediction of power distribution system outages. *Electric Power Systems Research*, 77(8), pp. 973–979.