Forest Ecosystems 12 (2025) 100284

Contents lists available at ScienceDirect

Forest Ecosystems

journal homepage: www.keaipublishing.com/cn/journals/forest-ecosystems

Classification and clustering analysis of standing dead trees and associated park asset wildfire vulnerability in Yellowstone National Park

Carolyn Prescott [a], Mehmet Ozdes [b,c], Di Yang [a,d,*]

[a] School of Computing, Wyoming Geographic Information Science Center, University of Wyoming, Laramie, WY 82071, USA
[b] Environmental Engineering, Faculty of Engineering, Tekirdag Namik Kemal University, Tekirdag 59030, Turkey
[c] Civil and Environmental Engineering, Clarkson University, Potsdam, NY 13699, USA
[d] Department of Geography, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

Keywords:
Forest health
Fire risk
Rocky mountain and pacific northwest forests
Park management
Spatial autocorrelation
Random Forest

ABSTRACT

In the Rocky Mountain and Pacific Northwest regions of the United States, forests include extensive portions of standing dead trees. These regions showcase an intriguing phenomenon where the combined biomass of standing dead trees surpasses that of fallen and decomposing woody debris. This stems from a suite of factors including pest disturbances, management decisions, and a changing climate. With increasingly dry and hot conditions, dead timber on a landscape increases the probability that a fire will occur. Identifying and characterizing the presence of standing dead trees on a landscape helps with forest management efforts including reductions in the wildfire hazard presented by the trees, and vulnerability of nearby park assets should the trees burn. Using forest-based classification, exploratory data analysis, and cluster vulnerability analysis, this study characterized the occurrence and implications of standing dead trees within Yellowstone National Park. The findings show standing dead trees across the entire study area with varying densities. These clusters were cross-referenced with vulnerability parameters of distance to roads, distance to trails, distance to water, distance to buildings, and slope. These parameters inform fire ignition, propagation, and impact. The weighted sum of these parameters was used to determine the vulnerability incurred on the park assets by the clusters and showed the highest values nearest to park entrances and points of interest. High vulnerability clusters warrant priority management to reduce wildfire impact. The framework of this study can be applied to other sites and incorporate additional vulnerability variables to assess forest fuel and impact. This can provide a reference for management to prioritize areas for resource conservation and improve fire prevention and suppression efficiency.

1. Introduction

Fires are crucial components of land management and ecosystem health as regenerative agents for plant species (Baeza et al., 2007). The vegetation in North America has been shaped by recurring fires over millions of years (He et al., 2019), but climate change has increased the frequency, intensity, and extent of wildfires in the western United States by altering temperature regimes and the availability and distribution of water (Picos et al., 2019; Westerling et al., 2006). Evaluating woodland ecosystem vitality relies on understanding the relationship between forest health, climate dynamics, and management practices. Forest integrity promises biodiversity and social, economic, and ecological resources (Franklin et al., 2018). High interspecies competition or disease and pest trauma can increase the amount of dead organic matter in a forest. A forest dominated by dead organic matter also suffers in biodiversity, its ability to function as a resource, and its adaptive capacity to withstand wildfire impact (Ozsahin et al., 2022; Zeppel et al., 2013). Degraded forest conditions create environmental concerns and lead to increased risks to nearby human communities (McWethy et al., 2019). The National Center for Environmental Information (NCEI) Billion-Dollar Weather and Climate Disasters dashboard shows that from 1980 to 2023, there were 22 major wildfire events in the US exceeding one billion dollars each in damages, resulting in a total of $142.8 billion in damages and 534 deaths (NOAA, 2023). The increasing prevalence of wildfires has raised interest in improvements in prevention, preparedness, and recovery from wildfire events impacting forest ecosystems, resources, and human communities (Chuvieco et al., 2012).

Fire frequency, severity, extent, and onset and their driving variables

* Corresponding author. Department of Geography, University of Florida, Gainesville, FL 32611, USA.
E-mail addresses: cprescot@uwyo.edu (C. Prescott), mozdes@nku.edu.tr (M. Ozdes), yangdi1031@ufl.edu (D. Yang).

https://doi.org/10.1016/j.fecs.2024.100284
Received 30 May 2024; Received in revised form 9 November 2024; Accepted 28 November 2024

vary regionally and temporally (Cansler et al., 2022; Pausas and Keeley, 2021; Stevens-Rumann et al., 2016; Arsanjani and Vázquez, 2020) and have been studied using satellite imagery to observe burn scars from both before and after wildfire events (Stavrakoudis et al., 2020). Wildfire onset and severity are chiefly influenced by elevation, slope, aspect, land cover, proximity to the nearest stream, and fire recurrence (Asori et al., 2020; Fernández-Guisuraga et al., 2021; Noss et al., 2006). These driving variables dictate the fuels present to burn. Standing dead trees are a significant fuel loading component. As a potential product of disease, this can result in large homogeneous swaths of dead trees in a forest. These standing dead fuels are then a substantial fire hazard and can serve as a fuel ladder to upper crown fuels or will soon be a high volume of down woody debris. Rocky Mountain and Pacific Northwest forests have some of the highest volume and biomass of standing dead trees (Woodall et al., 2006). Landscape heterogeneity can either inhibit or promote wildfire behavior (Chen et al., 2017). Therefore, regional studies characterizing the distribution and density of standing dead trees can help with a local assessment of their role in fire hazard and impact potential. The regional variation in fire history, fire hazard, and fire impact heightens the need for regional studies to translate between analysis to indices to inform management decisions (Jones et al., 2022; Lausch et al., 2017).

Wildfire risk is the product of hazard and physical vulnerability (Scott et al., 2013). Hazard captures the likelihood of a fire burning in an area, based on the fuels available and how intensely the area is likely to burn. Vulnerability speaks to an asset's susceptibility to exposure and adverse effects from a hazard. This study addresses the physical vulnerability of the human-built infrastructure within Yellowstone National Park, encompassing structures such as buildings, trails, and roads. Vulnerability assessments can rise to a detailed risk assessment level if they capture the statistical probability of a particular intensity hazard combined with the level of vulnerability and type of assets. Vulnerability in this study is generalized simply as proximity to potential wildfire fuel (standing dead trees). Hazard posed by the likelihood of stands of dead trees to burn is simplified to a consideration of the impact of slope and waterway proximity.

Fire management requires extensive knowledge of an area's landscape, climate, and human factors. This study aims to provide a comprehensive and nuanced understanding of wildfire vulnerability in Yellowstone National Park to inform the development and implementation of effective and adaptive fire management strategies. By using remote sensing data and random forest machine learning algorithm to classify land cover and rank wildfire vulnerability of park assets with spatial data and modeling techniques this study seeks to 1) identify the pattern of standing dead trees in Yellowstone National Park, and 2)

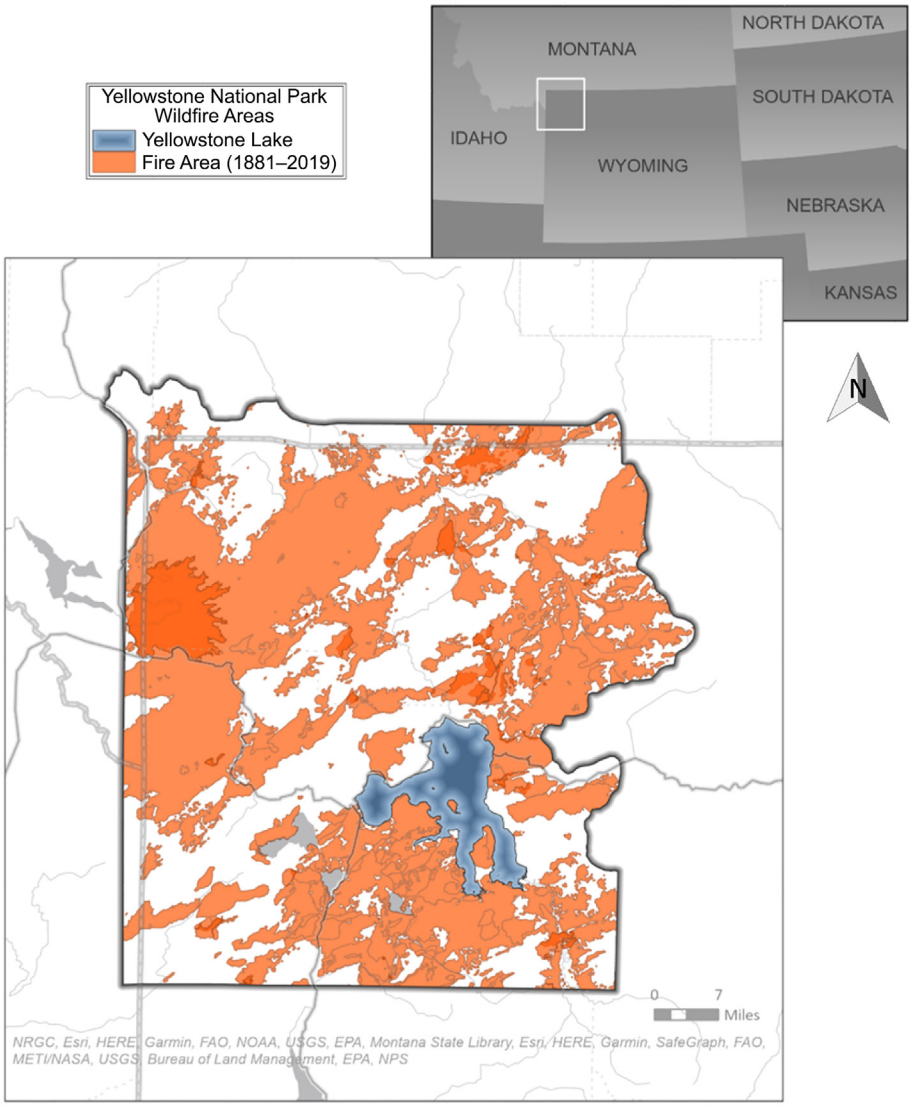


Fig. 1. The location of Yellowstone National Park with National Park Service fire perimeter history overlain to show the extent of burns across the landscape since 1881. Fire areas are symbolized with 30% transparency so areas with repeated burns appear darker.

prioritize clusters of standing dead trees for management to reduce wildfire vulnerability of park assets. Though the focus is on the Yellowstone region specifically, there is the opportunity to extend this study to other sites in Rocky Mountain and Pacific Northwest forests where standing dead tree assessments have also been completed.

2. Data and methods

2.1. Study area

The Yellowstone National Park boundary was selected to bound the area of study (Fig. 1). This includes 2.2 million acres of land in the northwest corner of Wyoming (Buskirk, 2016). This area has a history of natural fires, and the National Park Service (NPS) manages the area under a uniform strategy that has evolved over time in pace with the evolution of understanding the role of wildfire (NPS1, 2023). The current wildfire management strategy was adopted in 2009 (Botti and Nichols, 2021). These 2009 Wildland Fire Policy Guidelines allow fire personnel to manage a lightning-caused fire for multiple objectives which can include both protecting structures and allowing fire to burn for natural benefits (DellaSala et al., 2004).

Situated in the greater Yellowstone ecosystem, Yellowstone National Park hosts the intersection of species found both in the Rocky Mountains, the Great Plains, and the Intermountain region (Dersam, 2019). Vegetation communities are spatially distributed according to the underlying geology, water availability, elevation, soils, and the impact of disturbances caused by fire, floods, landslides, blowdowns, insect infestations, and competition with nonnative plants (Stegner et al., 2019). The vegetation communities present include higher and lower elevation forests and their associated understory vegetation, sagebrush-steppe, wetlands and hydrothermal.

Forest covers approximately 80% of the park (NPS1, 2023). Fire tolerance, elevation, and substrate composition dictate the species present in an area. As an example, rhyolitic soils do not have the nutrients to support fire and spruce communities, but lodgepole pines are able to sustain populations (Notaro et al., 2019). The suite of species present across the park includes lodgepole pine (*Pinus contorta*), Engelmann spruce (*Picea engelmannii*), subalpine fir (*Abies lasiocarpa*), whitebark pine (*Pinus albicaulis* Engelm.), limber pine (*Pinus flexilis*), Douglas-fir (*Pseudotsuga menziesii*) and Rocky Mountain juniper (*Juniperus scopulorum*) (NPS1, 2023).

2.2. The Yellowstone NEON field data and classification analysis

The National Ecological Observatory Network (NEON) dataset generated the classification map which is the foundation of this study. The data collected by the NEON project informs research on environmental impacts within mountain ecosystems. NEON technicians collect in-situ ecological data as part of continental-scale research. Data are collected with hand-held tools in the field, using standardized methods adopted from the forestry community (NEON, 2023).

The study area's Woody Plant Vegetation Structure data product (NEON data product ID: DP1.10098.001) contains the quality-controlled, native sampling resolution data from in-situ measurements of live and standing dead woody individuals, shrub groups, and non-herbaceous perennial plants from all terrestrial NEON sites with qualifying woody vegetation. The exact measurements collected per individual depend on growth form, and these measurements are focused on enabling biomass and productivity estimation, estimation of shrub volume and biomass, and calibration/validation of multiple NEON airborne remote-sensing data products. Measurements include taxonomic identification, stem diameter, height, crown dimensions, observations of plant health and mapped location. Among the plant health observations is a plant status variable that captures whether an individual is live, standing dead, removed, lost, or dead. For this study the focus centered on individuals with a standing dead assignment: a "standing dead tree or standing dead bole within a multi-bole tree, regardless of cause of death. The entire tree or bole must be dead, and the main bole must not be broken." (NEON, 2023).

Two methodologies were used to prepare spatial variables: 1) Remote Sensing Feature Extraction: NEON AOP-derived images trained classifiers, capturing spectral, structural, and environmental characteristics of vegetation at the site during fieldwork (Table 1). 2) Standing Dead/Living Tree Ratio Calculation: Computed the ratio of standing dead to living trees at each NEON field plot (20 m × 20 m) over the period 2019–2022. A low ratio signifies a higher living tree counts while higher ratios suggest higher standing dead tree counts.

Random Forest Classification creates models and generates predictions using the random forest algorithm based on known values provided as part of a training dataset (Breiman, 2001). The model is then applied to a prediction dataset with the same explanatory variables as the training dataset. Many decision trees, called an ensemble or a forest, each consider a random subset of features against a random data sample. While the use of a single decision tree may lead to overfitting, multiple decision trees each using a random subset of explanatory variables via resampled data from the training set reduces overfitting, enhances the accuracy and stability of the model, and avoids the need for complex assumptions about the underlying distribution of the data (Fox et al., 2017). The outcome of the model is determined by aggregating the decisions of each individual tree, where each tree "votes" for an outcome value, and the class with the majority of votes is assigned (Sage et al., 2020). This method was selected for this study with its ability to handle high-dimensional data, its low sensitivity to overfitting and the number of input variables, its intuitive interpretation of accuracy and variable importance, and its independence from requiring many user-defined parameters.

Through Google Earth Engine, the 544 NEON field plots served as training samples to model the "standing dead/living" tree ratios. The model was then applied to the prediction dataset, with independent validation set at 30%. The four resultant categories are classified based on standing dead tree percentage: 100%–70% standing dead, 70%–40% standing dead, 40%–10% standing dead, and less than 10% standing dead.

The performance of each classification model was based on internal out-of-box error estimates and the accuracy of the independent validation set predictions. Accuracy of the classification model was summarized via confusion matrices. The precision and recall, or user's and producer's accuracy, respectively, were calculated based on the confusion matrix. Precision is the proportion of correctly predicted instances in a specific class out of all instances that were predicted to be in that class, while recall is the proportion of correctly predicted instances in a specific class out of all actual instances in that class. The F1 score is the harmonic

Table 1
The data employed in the Random Forest algorithm.

Variable name	Description	Spatial resolution (m)	Data source
CHM	Height of canopy above the ground	1	NEON AOP
Vegetation Cover	Derived from NEON	1	NEON AOP
DEM	Digital Elevation Model	1	USGS.GOV
Slope	Derived from DEM	1	USGS.GOV
NDVI	Normalized Difference Vegetation Index	10	Sentinel-2
EVI	Enhanced Vegetation Index	10	Sentinel-2
PRI	Photochemical Reflectance Index (Canopy Xanthophyll)	10	Sentinel-2
SAVI	Soil-Adjusted Vegetation Index	10	Sentinel-2
NDBI	Normalized Difference Burn Index	10	Sentinel-2
ARVI	Atmospherically Resistant Vegetation Index	10	Sentinel-2

mean of precision and recall, and it provides a single metric, between zero and 1, that considers both false positives and false negatives. A F1 score of 1 indicates perfect precision and recall.

2.3. Exploratory spatial data analysis

Using ArcGIS Pro (NPS5, 2023) an Exploratory Spatial Data Analysis (ESDA) helped describe and visualize spatial distribution and identify areas of hot spots and cold spots within the classification results. Global Moran's I offers a test of global clustering by classifying spatial autocorrelation as positive, negative or none (You et al., 2017). This analysis unveils spatial patterns or trends in feature attributes, such as clustering or dispersion, that may not be readily apparent through visual inspection alone. Positive spatial autocorrelation is when similar values cluster together in a map, negative spatial autocorrelation is when dissimilar values cluster together in a map, and no spatial autocorrelation indicates a random distribution of features, with no discernible spatial pattern.

Following the assessment for statistically significant clustering in the forest data, the Hot Spot Analysis tool pinpointed areas of statistically significant high values (indicating more dead trees) and statistically significant low values (indicating more live trees). The Optimized Hot Spot Analysis tool identifies an appropriate scale of analysis, corrects for both multiple testing and spatial dependence, and generates a feature class with a calculated *z*-score, *p*-value, and a confidence level bin field. The confidence bins are derived from the *z*-scores and *p*-values of the data, offering insights into the confidence associated with the location of the hot or cold spot.

2.4. Cluster vulnerability analysis

The cluster suitability analysis process can be broken into three parts: identifying and creating cluster locations, preparing the suitability criteria, and assigning suitability to each cluster. Suitability in this context translates to areas of greater concern for wildfire probability and impact, or vulnerability. The result is to identify which clusters rank highest in concern to inform management triage.

Using the classified forest assessment data, clusters were identified as areas with greater than 500 cells of the highest dead tree ratio value (100%–70% standing dead trees). These clusters of points were then converted to convex hull polygons to serve as the areas to investigate in the vulnerability analysis. Thresholds of 250 and 1,000 cells were also tested for viability but yielded too few or too many clusters.

The following criteria were selected to serve as the inputs for the vulnerability analysis: roads, trails, building locations, slope, and water sources (Fig. 2). Water source flowline data were accessed through the WyGISC GeoHub REST service. This is the National Hydrography Dataset that has been clipped to the state of Wyoming (NHD, 2023). Road, trail, and building information was accessed through the National Park Service GIS Data Services REST server (NPS2; NPS3; NPS4, 2023). Slope data was derived from the USGS seamless 3D Elevation Program Digital Elevation Model (DEM) dataset, 2020 vintage (USGS, 2022). The resolution is approximately 10 m in the north/south direction, with variable east/west resolution due to the convergence of meridians with latitude.

All input criteria data sources were clipped to the Yellowstone Park boundary prior to analysis. From these clipped data sources distance accumulation raster data for the water source, road, trail, and building locations were generated to show the gradients of distance from each criterion throughout the study area.

The distance raster layers, and the slope raster were then each rescaled to a common scale of 1–20 using functions appropriate for each criterion. By transforming the criteria to common scales, they can be compared and combined. Because wildfire probability is higher in areas

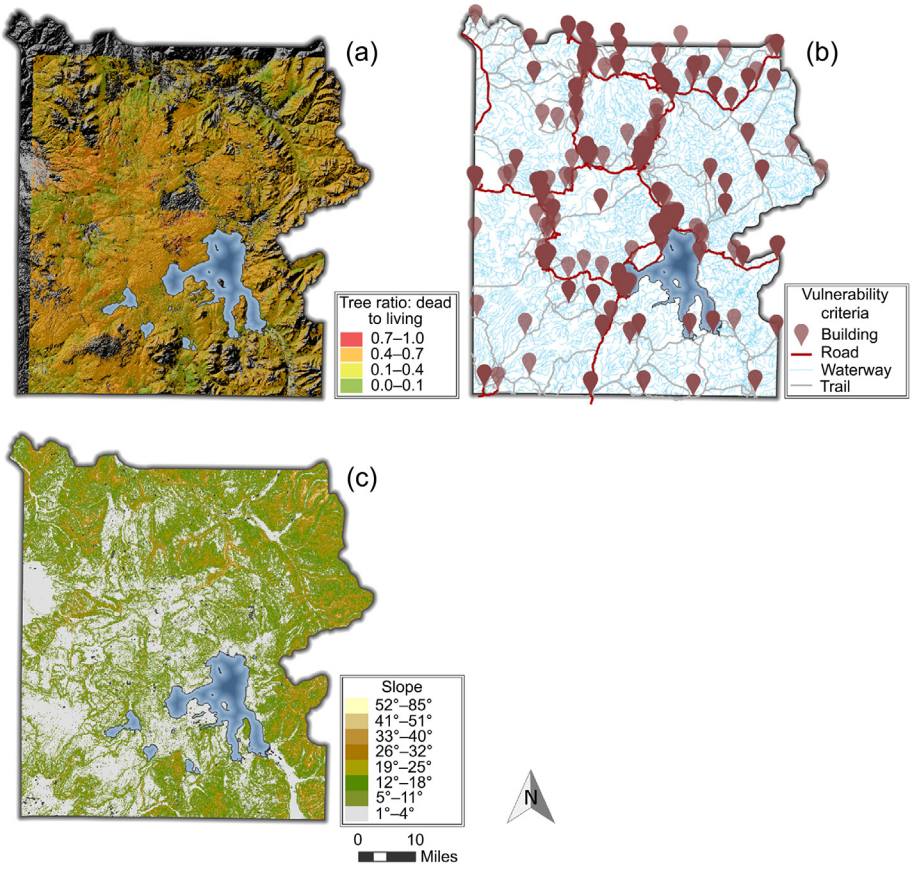


Fig. 2. **(a)** Forest assessment classification dataset used to identify clusters. **(b)** Point and vector datasets used to identify distances from fire vulnerability influences. **(c)** Slope data used to inform fire vulnerability.

of steeper slopes, the slope raster layer was rescaled using the MSLarge function – based on the mean and standard deviation where higher values of input raster have higher preference. Fire vulnerability is greatest at locations closest to park assets of roads, trails and buildings, so these criteria were rescaled using the Small function which indicates that locations with smaller values have higher preference. Locations closest to water typically have a lower probability of fire ignition so the water source raster was rescaled with the Large function indicating that locations further from water sources have higher preference. With a standardized scale these criteria layers could be combined and weighted according to impact on fire vulnerability.

Weightings were determined with a hierarchy process considering the relative importance of each variable (You et al., 2017). The vulnerability criteria consisted of asset proximity factors with the addition of the fire likelihood, or hazard, factors of slope and distance to river. Using the ArcPy weighted sum tool allows for any positive or negative decimal value to be assigned as a weight. High slope areas are prone to fast spreading fire behavior and are more difficult to manage so slope was given a weight of 1. Roads represent high use areas so distance to roads was given a weight of 1.5. Trails also represent high use areas, though less than roads, so distance to trails was given a weight of 1.1. Proximity to water has a high impact on the probability for an area to burn so distance to water was given a weight of 1.25. Because distance to buildings serves as a proxy for the highest use areas it was given a weight of 1.7.

The ArcPy Zonal Statistics tool was used to summarize the value of the combined weighted sum criteria raster within the areas of the cluster polygons. Each cluster was assigned its average value of the combined criteria raster. These values were then binned into three categories to distinguish level of concern using natural breaks classification. With this, class breaks are defined so as to best group similar values together and maximize the differences between classes. This classification is computed within ESRI ArcPro based on the Jenks Natural Breaks algorithm. This final ranking serves as the recommendation of priority for forest management efforts.

3. Results

3.1. Random forest classification

The spatial results (Fig. 4) of the random forest classification capture the distribution of dead and live trees throughout Yellowstone National Park. The classified map reveals distinct swaths of higher living tree ratios interspersed with areas dominated by higher dead tree ratios. The value distribution indicates that the highest count of cells was assigned to the 40%–70% standing dead tree category, suggesting a significant portion of the park is experiencing moderate to high levels of tree mortality. This pattern will be further explored in the ESDA results section.

The performance of the random forest model can be assessed with several evaluation metrics including an overall accuracy score. The overall accuracy of the model, which represents the proportion of correctly classified instances out of the total instances, was found to be 0.7665 or 76.65%. This indicates that the model correctly predicted the dead tree percentage category for approximately 77% of the validation samples, demonstrating a reasonable performance in capturing the spatial patterns of tree mortality.

A confusion matrix describes the model's performance between classes (Table 2). The matrix displays the counts of true positive, true negative, false positive, and false negative predictions allowing an understanding of where the classifier was confused. The diagonal cells in the matrix represent the correctly classified instances for each class, while the off-diagonal cells indicate misclassifications.

The precision values range from 0.68 to 1.00, with an average of 0.85. The highest precision of 1.00 was achieved for the 10% and 100% standing dead tree ratio classes, indicating perfect precision in the model's predictions for these extreme categories. In the confusion matrix (Table 2), the sample size for the 100% standing dead tree percentage category (representing plots with 0% live trees) is conspicuously small, comprising only 2 instances. This scarcity in samples can be attributed to the unique nature of the dataset, where each data point corresponds to a 20 m × 20 m plot. Considering the expansive study area covering approximately 400 km^2, predominantly forested, encountering patches devoid of live trees is exceptionally rare. The classification accuracy report (Table 3) shows performance metrics of the model. The lowest precision of 0.68 was observed for the 50% class, suggesting some overestimation of this category. The recall values exhibit a similar range from 0.64 to 1.00, with an average of 0.82. The model achieved perfect recall for the 10%, 90%, and 100% classes, while the lowest recall of 0.64 was observed for the 20% class, indicating some underestimation of this category. The highest F1 scores of 1.0 were obtained for the 10% and 100% classes, aligning with their perfect precision and recall values. The

Table 2
Confusion matrix of the random forest classification model for predicting standing dead tree percentage categories. Values are highlighted with a gradient where darker greens denote higher counts. The values on the diagonal are those of true value classifications. Note that while the model did produce some false classifications, the highest counts are on the true classification diagonal.

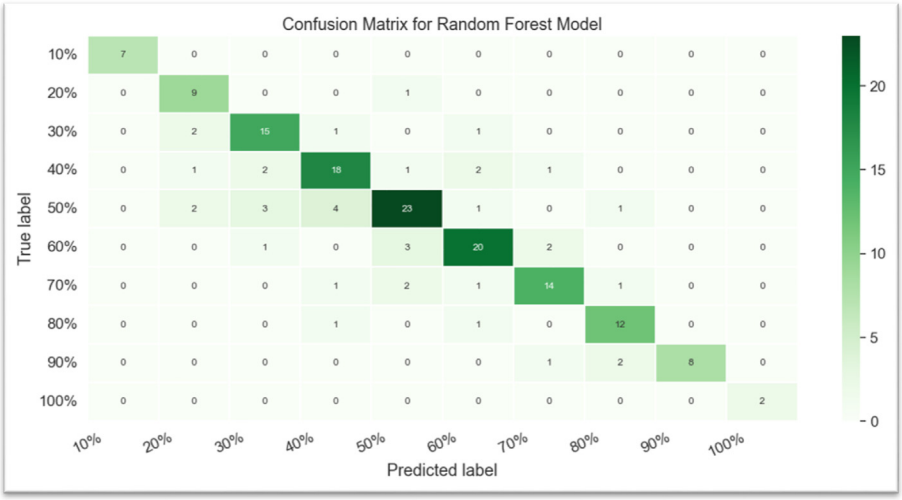
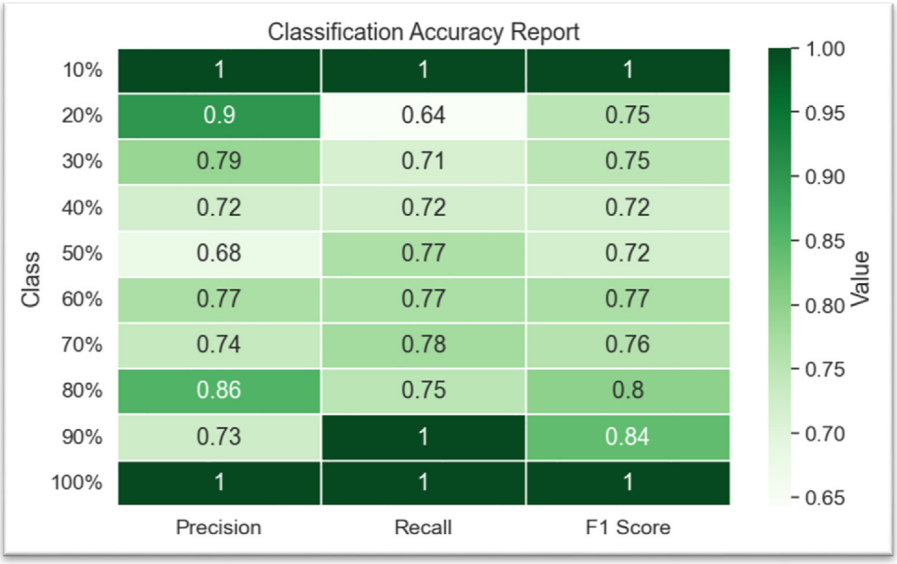


Table 3
Classification Accuracy Report of the performance metrics. Values are highlighted with a gradient where darker greens indicate better model performance and lighter greens indicate poorer model performance.



lowest F1 scores of 0.72 were observed for the 40% and 50% classes, indicating relatively lower performance for these categories compared to others.

3.2. Exploratory spatial data analysis

This exploratory spatial data analysis summarized and inspected the classified forest assessment data. The distribution of this data into the four tree ratio bins can be seen in Fig. 3. A value of 4 indicates a high count of dead trees and a value of 1 indicates a high count of live trees. Within the study area, a tree ratio bin of 3 (40%–70% dead) was most abundant, followed by 1 (0%–10% dead), then 4 (70%–100% dead), with the least abundant being 2 (10%–40% dead).

A visual inspection of the results shows distinct regions dominated by a particular tree ratio. Comparing these regions against satellite imagery confirms that they approximate the vegetation patterns observed in the imagery (Fig. 4).

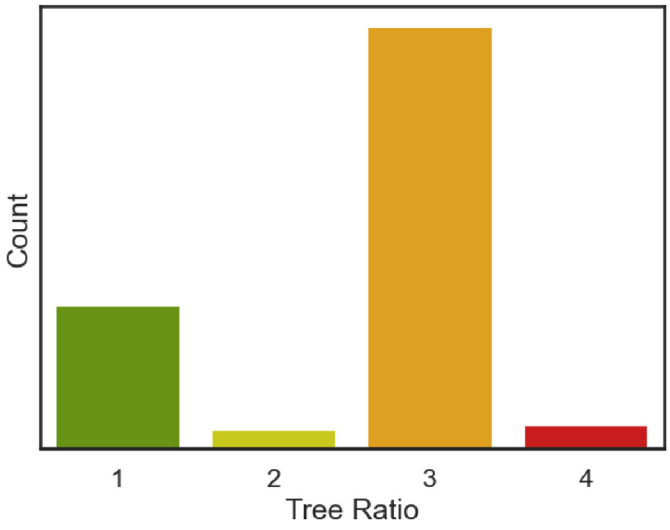


Fig. 3. Illustrates the distribution of standing dead trees across various ratio bins based on forest assessments. These bins are categorized as follows: Bin 1: 0–10%; Bin 2: 10%–40%; Bin 3: 40%–70%; and Bin 4: 70% and above.

A spatial autocorrelation analysis (Global Moran's I) described the clustering of these assigned bin values. This report indicated that the clustering of this dataset is significant (p-value <0.001) and that there is less than a 1% likelihood that this clustering is a result of random chance. The Moran's Index is 0.38 indicating a tendency towards clustering.

The Optimized Hot Spot analysis then further described the clustering of the forest assessment bin values. The spatial output of this analysis showed regions of significant hot spots (higher dead tree counts) appeared predominantly in the central west portion of the park. Significant cold spots (higher living tree counts) were indicated in the upper west and central to southern east portions of the park (Fig. 5). Statistics results show the average and majority confidence bin score to fall around zero, indicating neither a significant hot nor cold spot.

3.3. Cluster vulnerability analysis

3.3.1. Cluster identification

A threshold of 500 dead tree cells identified 53 distinct clusters (Fig. 6). A threshold of 250 dead tree cells yielded 87 clusters and a threshold of 1,000 cells yielded 17 clusters. The size distribution of the clusters showed an average of 34 km (Fig. 9). Larger clusters appear in the northeast and southern portions of the park.

3.3.2. Vulnerability criteria

The selected vulnerability criteria included road, trail, building, and waterway proximity, as well as slope. The outputs of the distance accumulation analysis run on these point and line features are shown in Fig. 7. The criteria values were transformed to a common scale so they could be compared and combined. The result of this transformation of each criterion and their combination into the weighted sum layer is shown in Fig. 8.

3.3.3. Cluster ranking

The combined weighted sum criteria raster values within the cluster polygons were averaged for each cluster. These average values were then categorized into three levels (high, medium, low) using natural breaks classification, which optimally groups similar values while maximizing differences between classes, based on the Jenks Natural Breaks

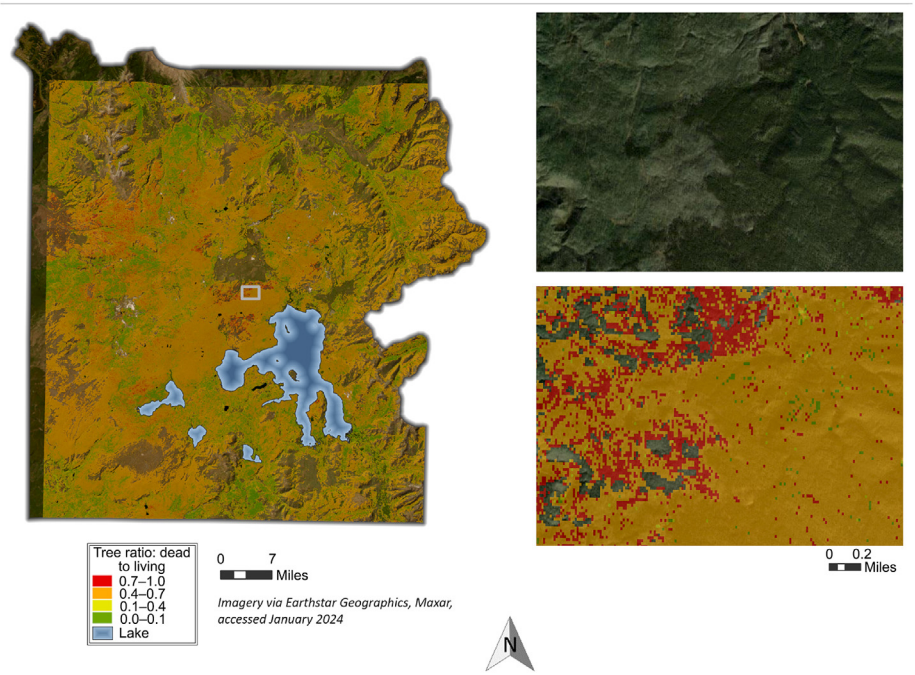


Fig. 4. Results of imagery classification into four tree ratio categories overlain on satellite imagery. The inset maps on the right show detail of the rectangular region indicated in the center of the overview map on the left. Note the alignment of the areas predicted by the model to exhibit high percentages of dead trees (shown as red), with the portion of the satellite imagery appearing less vegetated.

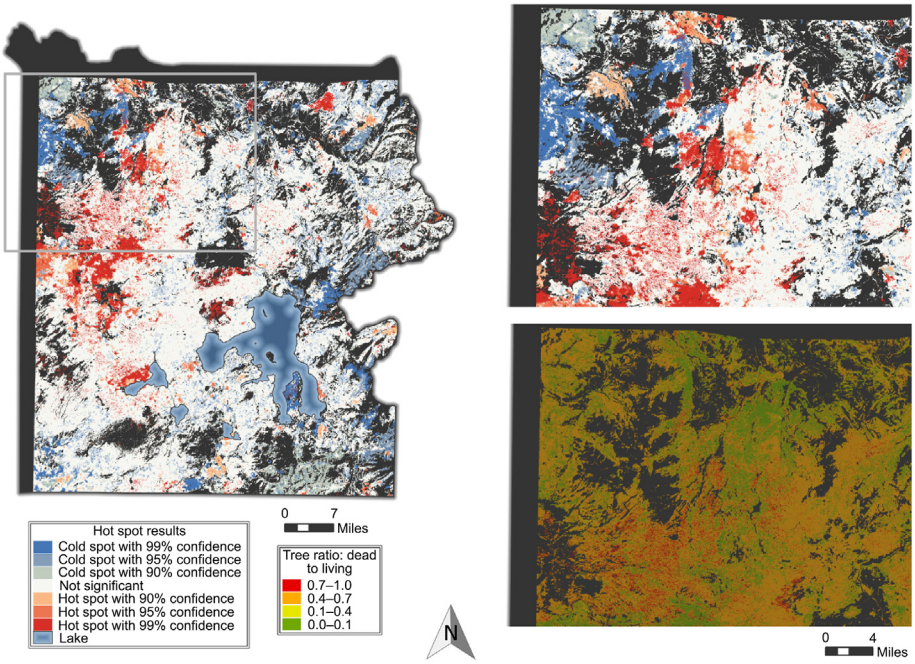


Fig. 5. Optimized hot spot results. The right figure shows greater detail to the outlined area in the upper left of the overview map on the left. The upper map on the right shows the hot spot results and the lower map on the right shows the classified tree ratio data. Note the correlation between areas classified as a high dead tree ratio (red) with a hot spot (red), and low dead tree ratio (green), with a cold spot (blue).

algorithm. This final ranking provides recommendations for prioritizing forest management efforts. Cluster vulnerability analysis indicates that the top five clusters are located in the northeastern and eastern areas of the park and have areas exceeding 10 square kilometers. A notable grouping of high vulnerability is visible in the central-western area. Clusters with lower vulnerability are primarily located in the southern and western areas of the park (Fig. 9).

4. Discussion

4.1. Random forest classification

Given the aridity of the Yellowstone National Park study area, the classification results align with the expectation that the ground cover would exhibit a patchwork of vitality trending towards more dead plant

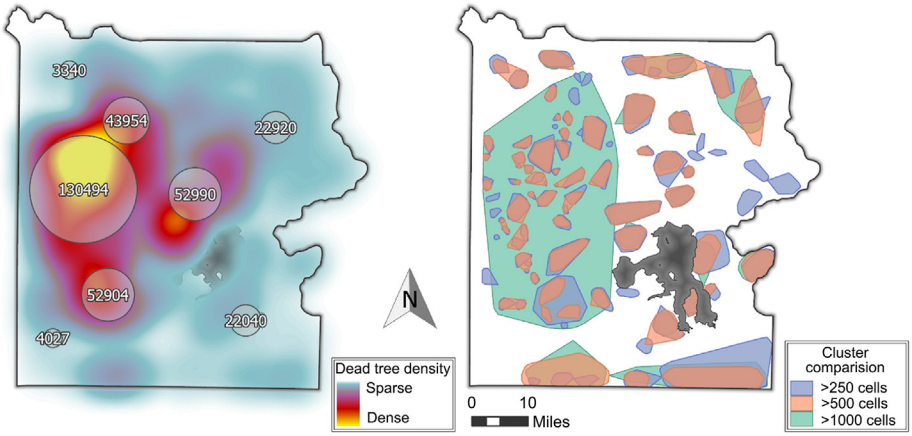


Fig. 6. (Left) Heat map labeled with count of highest dead tree class assigned cells showing the predominance of these cells in the central west portion of the park. (Right) Comparison of results of the number of clusters generated with different minimum threshold values.

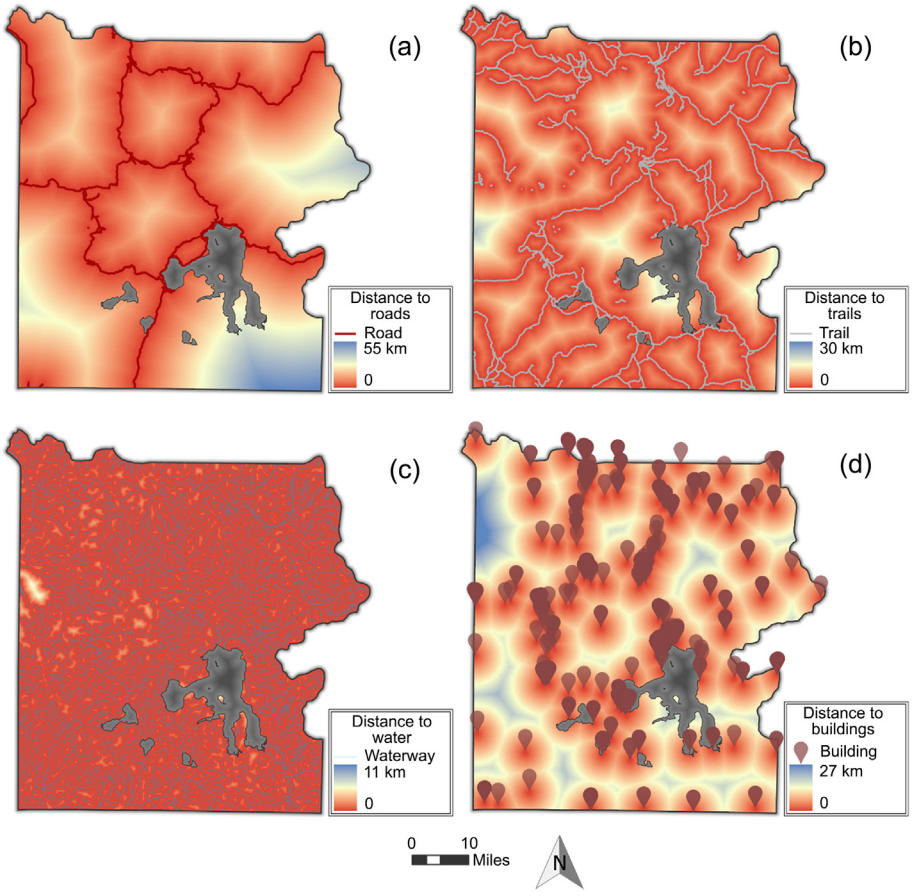


Fig. 7. Output maps of the distance accumulation calculation from the point and line vulnerability criteria. **(a)** Distance to roads, **(b)** Distance to trails, **(c)** Distance to water, **(d)** Distance to buildings.

matter given environmental stressors like heat, drought, pests, foraging, and fire. Wildfire impact on riparian areas in particular, as studied in Yellowstone National Park by Minshall et al. (1997), was shown to trigger significant changes that persist with time and hamper post fire ecological regeneration.

The random forest classification model's performance metrics indicate the success and reliability of the modeling approach. The overall model accuracy of 76.65%, obtained from a separate validation set, indicates that the model effectively captures the spatial patterns of standing dead tree percentages across Yellowstone National Park. The confusion matrix (Table 2) offers a detailed breakdown of the model's performance for each standing dead tree percentage class. An inspection of the precision, accuracy and F1 scores of the model (Table 3) demonstrates strong balance between precision and recall particularly in the endmember classes of 10% and 100% standing dead trees. These performance metrics underscore the model's ability to accurately identify areas with varying counts of standing dead trees, providing a reliable foundation for understanding the spatial distribution of dead and live trees in the park. The model's success in

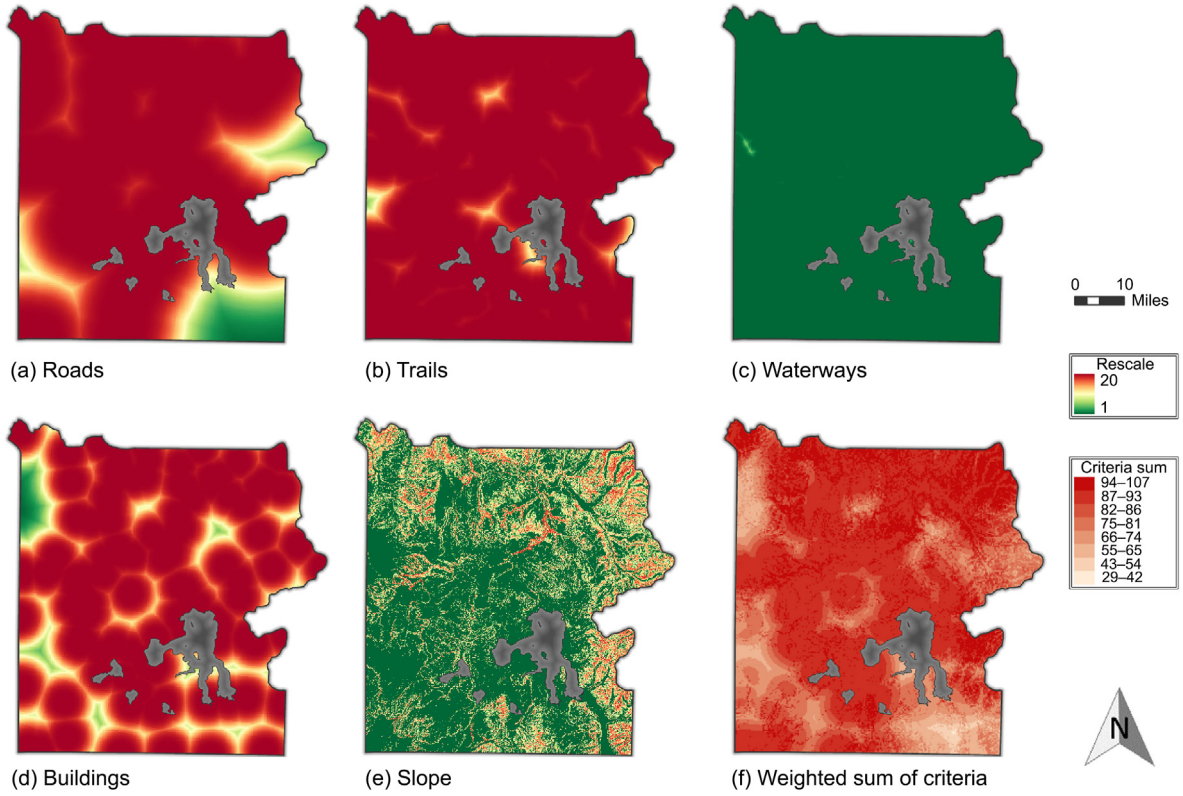


Fig. 8. The transformed criteria layers are set to a common vulnerability scale ranging from 1 to 20, where 1 is least vulnerable and 20 is very vulnerable. The most vulnerable areas are in red and the least vulnerable areas are in green. These transformed criteria layers have been weighted and combined into the Weighted sum of criteria layer (f) which serves to inform the cluster vulnerability assignment.

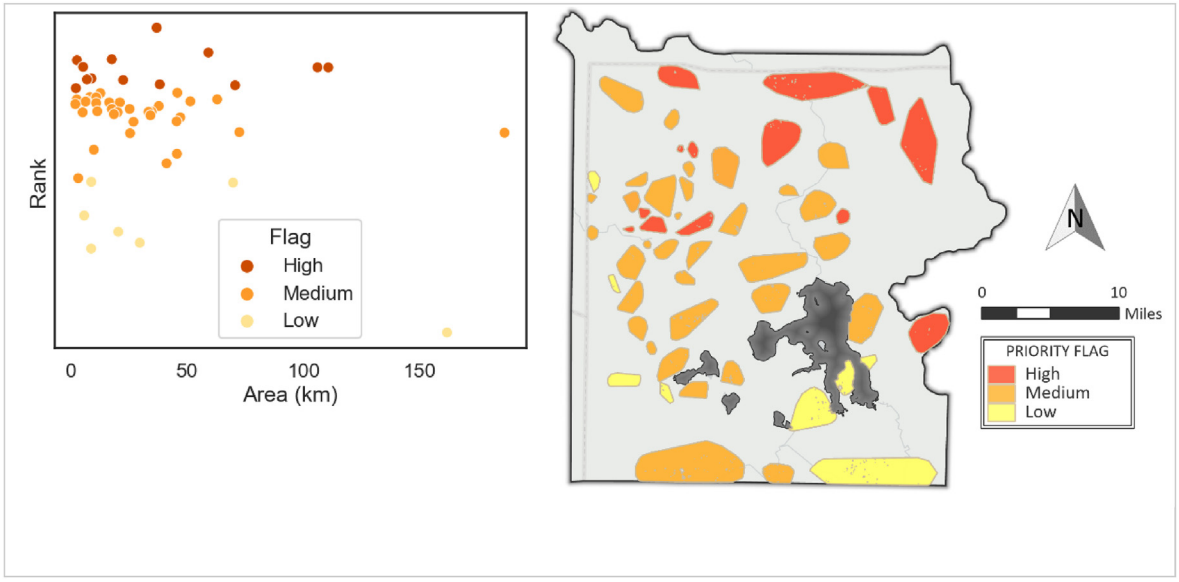


Fig. 9. (Left) Scatterplot illustrating the size distribution of the clusters in relation to their assigned vulnerability ranks, highlighting the relationship between cluster size and vulnerability assessment. (Right) Map displaying the clusters and their corresponding levels of vulnerability based on the averaged combined weighted sum criteria raster values.

capturing these patterns can be attributed to the effective integration of remote sensing features, such as spectral indices and canopy height metrics, which serve as strong predictors of tree health and mortality. Furthermore, the random forest algorithm's ability to handle complex, non-linear relationships and its robustness to overfitting contribute to the model's good performance. The high accuracy and strong agreement with ground truth data demonstrate the potential of the random forest classification approach for mapping and monitoring tree mortality across large, heterogeneous landscapes like Yellowstone National Park. These findings highlight the value of advanced machine learning techniques in extracting meaningful insights from remote sensing data to support forest health assessment and management decision-making.

4.2. Exploratory data analysis

A visual inspection of the classification results showed clear regions dominated by particular dead to live tree ratios, confirmed against satellite imagery. This aligns with expectations of vegetation distribution in this landscape. In this arid environment, vegetation is closely linked to water availability, leading to heterogeneity in vegetation distribution. North aspects in this region see less cumulative sun year-round and retain more moisture and thereby can support more vegetation. South aspects see more sun and are consequently drier and more prone to being sparsely vegetated (Noss et al., 2006). Spatial patterning is also influenced by fire disturbance with plant species, density, patch size, aspect, and burn severity all influencing post-fire recovery (Smith et al., 2021). The heterogeneous plant distribution apparent in the ESDA can be understood as a response to the terrain and the history of the area.

The optimized hot spot analysis delineated regions dominated by high dead tree counts, predominantly in the central-west sector of the park— an area impacted by both the 1988 fires and the 2016 Maple Fire. The latter, spanning from August 8, 2016 until late October, resulted in the disturbance of approximately 18,383 ha of land (Szpakowski et al., 2023). Conversely, the upper-west and central-to-southern-east portions of the park displayed statistically significant cold spots, indicative of a higher ratio of living trees. These areas do not have a history of large fire incidents and are generally at higher altitudes less suitable to forest development given thinner soils, and a reduced growing season length (NPS1, 2023).

The spatial autocorrelation analyses of the classification results indicated clustering with statistical significance. Depending on the cause of tree mortality, the observed pattern of dead trees on a landscape can be of varying homogeneity. According to the literature, a clustered pattern of standing dead trees can be a product of disease, which, if widespread enough, can result in large homogeneous swaths of dead trees in a forest (Woodall et al., 2006). This pattern of dead trees on the landscape decreases canopy continuity and alters species composition which has been shown to both inhibit and promote wildfire behavior, depending on the circumstances. Chen et al. (2017) examines the interplay between sudden oak death (*Phytophthora ramorum*) and wildfire. Using high-resolution imagery with a physical simulation model to gauge burn severity based on Landsat data, the study shows the impact of tree mortality on burn severity. Contrary to conventional expectations, landscapes impacted by the disease experience greater burn severity. This indicates that the changes in surface fuels resultant from the disease may decrease landscape resistance to fire, which can also worsen the impact of fires on adjacent forest patches. Therefore, this clustering of dead trees found in this study suggests the potential for increased fire impact.

4.3. Cluster vulnerability analysis

The spatial distribution of clusters shows larger clusters appear in the northeast and southern portions of the park. Clusters were determined by a threshold count of standing dead tree cells, meaning large cluster areas are those with less dense standing dead tree occurrence. For this reason, these clusters do not align with locations of statistically significant hot spots identified in the Hot Spot analysis of the ESDA. Examining this conclusion against the National Historic Land Cover data showing the ground cover, it is evident that these larger clusters are in areas with mixed ground cover compared to forest dominated smaller clusters located in the central portion of the park. Therefore, it should be noted that cluster size is inversely related to standing dead tree density.

Cluster vulnerability, having been examined based on distance to road, trail, building, and waterway, as well as slope, shows prioritization based on the proximity to the highest weighted vulnerability criteria: the roads, buildings and trails. These features are most dense near park entrances and points of interest and cluster prioritization aligns accordingly. An alternative prioritization method would be to rank according to density of the standing dead trees within each cluster as a means to identify clusters with higher likelihood to burn or higher hazard. However, as discussed previously, wildfire hazard is influenced by more than tree density.

The impact of fire on human park use is emphasized in this study over ecological fire impacts. This is because priority is assigned based on fire vulnerability, or the proximity to human maintained assets that may be impacted or destroyed. Vulnerability can be defined as low capacity to cope with disaster. Natural ecosystem components can be resilient to, or even regenerative alongside wildfires, but roads, trails, and buildings are not. There are also ecosystem components that are unable to regenerate post-fire given changes in climate since they were initially established, such as the aforementioned impact on riparian areas (Minshall et al., 1997), as well as the large aspen groves in Yellowstone (Romme et al., 1995). An improvement to cluster vulnerability ranking could incorporate the resiliency of additional natural ecosystem components to wildfire impact to protect the native natural and human environment (You et al., 2017).

Per the NPS's historic fire records, between 1992 and 2019 11% of fires were attributed to human ignition with the remaining 89% of fires linked to natural causes. Human caused fires are more likely within areas of high traffic: along roads, trails, or near buildings (You et al., 2017). This study examines explicitly how natural fire occurrences (89% of historical fires) interact with identified clusters of standing dead trees (the fuel hazard) to create varying levels of risk to park infrastructure and sensitive natural areas based on proximity. While climate change and multiple environmental factors influence fire probability, this prioritization study can provide Yellowstone National Park with actionable insights about which clusters of standing dead trees, if ignited, cause the greatest threat to park assets that cannot naturally recover from fire damage.

5. Conclusion

Using forest-based classification, exploratory spatial data analysis, and cluster vulnerability analysis granted an understanding of the presence, distribution, and implications of standing dead trees within Yellowstone National Park. Standing dead trees are present across the entire study area with varying degrees of density. Grouping zones of dead trees into minimum count threshold clusters and ranking these clusters by proximity to wildfire vulnerable features within the park can guide planning efforts and reduce the risk of fire. The framework of this study can be applied to other sites and incorporate additional vulnerability variables. This study offers a way to assess forest fuel and provide a reference for management to prioritize areas for resource conservation and improve fire prevention and suppression efficiency.

The results provide new insights on fire risk and add to the existing literature, offering valuable information for park management. However, it is important to acknowledge some limitations. First, the prioritization of clusters is based on proximity to human maintained park infrastructure. Incorporating a consideration of fire resiliency of natural ecosystem components would provide a more robust understanding of the ecosystem as a whole and its vulnerability to fire. Adding additional details about the wildfire hazard of the clusters of standing dead and details about the vulnerability of specific types of park infrastructure, such as loss of tourism or recovery costs, would also strengthen this study's conclusions. Finally, successful identification of the standing dead trees and therefore the prioritized clusters rely on confidence in the input data. The data used is of 2022 vintage leaving time for landcover changes between data collection and present. It is worth updating the results of this study as new data are available.

As noted by You et al. (2017), more studies of GIS-based forest fires integrating natural forest features to assess forest fire risk and map risk zones are necessary to build a better understanding of the spatial and temporal variation in the risk of forest fires. This study of standing dead trees within Yellowstone National Park offers both new results and a framework for additional study.

CRediT authorship contribution statement

Carolyn Prescott: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Mehmet Ozdes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis. **Di Yang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding and acknowledgments

We would like to express our sincere gratitude to Wyoming NASA EPSCoR Faculty Research Grant (Grant#80NSSC19M0061), Yellowstone National Park Services for their generous support and funding that made this research possible. Their commitment to advancing knowledge in fire risk management has been instrumental in the success of this project. We also extend our appreciation to all those who contributed to this research, directly or indirectly, for their valuable insights and collaboration.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Di Yang reports financial support was provided by NASA. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arsanjani, J., Vázquez, R.L., 2020. Exploring the potential socio-economic and physical factors causing historical wildfires in the Western USA. In: Vaz, E. (Ed.), Regional Intelligence: Spatial Analysis and Anthropogenic Regional Challenges in the Digital Age. Springer, Cham, pp. 95–120. https://doi.org/10.1007/978-3-030-36479-3_6.

Asori, M., Emmanuel, D., Dumedah, G., 2020. Wildfire hazard and risk modelling in the northern regions of Ghana using GIS-based multi-criteria decision-making analysis. J. Environ. Earth Sci. 10 (11). https://doi.org/10.7176/JEES/10-11-02.

Baeza, M.J., Valdecantos, A., Alloza, J.A., Vallejo, V.R., 2007. Human disturbance and environmental factors as drivers of long-term post-fire regeneration patterns in Mediterranean forests. J. Veg. Sci. 18 (2), 243–252. https://doi.org/10.1111/j.1654-1103.2007.tb02535.x.

Botti, S., Nichols, T., 2021. National Park Service fire restoration, policies versus results: what went wrong. Parks Stewardship Forum 37 (2). https://parks.berkeley.edu/psf/wp-content/uploads/2021/05/psf_372_botti_web.pdf. (Accessed 5 December 2023).

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Buskirk, S.W., 2016. Wild Mammals of Wyoming and Yellowstone National Park, first ed. University of California Press https://www.abebooks.com/servlet/BookDetailsPL?bi=31947915771. (Accessed 5 December 2023).

Cansler, C.A., Kane, V.R., Hessburg, P.F., Kane, J.T., Jeronimo, S.M., Lutz, J.A., Povak, N.A., Churchill, D.J., Larson, A.J., 2022. Previous wildfires and management treatments moderate subsequent fire severity. For. Ecol. Manag. 504, 119764. https://doi.org/10.1016/j.foreco.2021.119764.

Chen, G., He, Y., De Santis, A., Li, G., Cobb, R., Meentemeyer, R.K., 2017. Assessing the impact of emerging forest disease on wildfire using Landsat and KOMPSAT-2 data. Remote Sens. Environ. 195, 218–229. https://doi.org/10.1016/j.rse.2017.04.005.

Chuvieco, E., Aguado, I., Jurdao, S., Pettinari, M.L., Yebra, M., Salas, J., Martínez-Vega, F.J., 2012. Integrating geospatial information into fire risk assessment. Int. J. Wildland Fire 23 (5), 606–619. https://doi.org/10.1071/WF12052.

DellaSala, D.A., Williams, J.E., Williams, C.D., Franklin, J.F., 2004. Beyond smoke and mirrors: a synthesis of fire policy and science. Conserv. Biol. 18, 976–986. https://doi.org/10.1111/j.1523-1739.2004.00529.x.

Dersam, S.W., 2019. Behavioral complexities at high elevation: assessing prehistoric landscape use in the alpine regions of the Greater Yellowstone Ecosystem. UW-National Park Service Research Station Annual Reports 42, 104–112. https://doi.org/10.13001/uwnpsrc.2019.5749.

Fernández-Guisuraga, J.M., Suárez-Seoane, S., García-Llamas, P., Calvo, L., 2021. Vegetation structure parameters determine high burn severity likelihood in different ecosystem types: a case study in a burned Mediterranean landscape. J. Environ. Manag. 288, 112462. https://doi.org/10.1016/j.jenvman.2021.112462.

Fox, E.W., Hill, R.A., Leibowitz, S.G., Olsen, A.R., Thornbrugh, D.J., Weber, M.H., 2017. Assessing the accuracy and stability of variable selection methods for random forest modeling in ecology. Environ. Monit. Assess. 189, 1–20. https://doi.org/10.1007/s10661-017-6025-0.

Franklin, J.F., Johnson, K.N., Johnson, D.L., 2018. Ecological Forest Management. Waveland Press. https://www.waveland.com/browse.php?t=730. (Accessed 5 December 2023).

He, T., Lamont, B.B., Pausas, J.G., 2019. Fire as a key driver of Earth's biodiversity. Biol. Rev. 94 (6), 1983–2010. https://doi.org/10.1111/brv.12544.

Jones, M.W., Abatzoglou, J.T., Veraverbeke, S., Andela, N., Lasslop, G., Forkel, M., Smith, A.J., Burton, C., Betts, R.A., van der Werf, G.R., Sitch, S., 2022. Global and regional trends and drivers of fire under climate change. Rev. Geophys. 60 (3). https://doi.org/10.1029/2020RG000726.

Lausch, A., Erasmi, S., King, D.J., Magdon, P., Heurich, M., 2017. Understanding forest health with remote sensing-part II—a review of approaches and data models. Rem. Sens. 9 (2), 129. https://doi.org/10.3390/rs9020129.

McWethy, D.B., Schoennagel, T., Higuera, P.E., Krawchuk, M., Harvey, B.J., Metcalf, E.C., Schultz, C., Miller, C., Metcalf, A.L., Buma, B., Virapongse, A., 2019. Rethinking resilience to wildfire. Nat. Sustain. 2 (9), 797–804. https://doi.org/10.1038/s41893-019-0353-8.

Minshall, G.W., Robinson, C.T., Lawrence, D.E., 1997. Postfire responses of lotic ecosystems in Yellowstone national park, USA. Can. J. Fish. Aquat. Sci. 54 (11), 2509–2525. https://doi.org/10.1139/f97-160.

NEON, 2023. National ecological observatory Network vegetation structure. https://data.neonscience.org/data-products/DP1.10098.001/RELEASE-2023. (Accessed 5 December 2023).

NHD, 2023. The national Hydrography dataset (NHD) flowline. https://services.wygisc.org/HostGIS/rest/services/GeoHub/NHDFlowline/MapServer/0. (Accessed 1 November 2023).

NOAA, 2023. National centers for environmental information (NCEI) U.S. Billion-dollar weather and climate Disasters. https://www.ncei.noaa.gov/access/billions/. (Accessed 5 November 2023).

Noss, R.F., Franklin, J.F., Baker, W.L., Schoennagel, T., Moyle, P.B., 2006. Managing fire-prone forests in the western United States. Front. Ecol. Environ. 4 (9), 481–487. https://doi.org/10.1890/1540-9295.

Notaro, M., Emmett, K., O'Leary, D., 2019. Spatio-temporal variability in remotely sensed vegetation greenness across Yellowstone National Park. Rem. Sens. 11 (7), 798. https://doi.org/10.3390/rs11070798.

NPS1, 2023. Yellowstone national park - fire. https://www.nps.gov/yell/learn/nature/fire.htm. (Accessed 1 November 2023).

NPS2, 2023. National park Service GIS data Services directory public roads. https://mapservices.nps.gov/arcgis/rest/services/NationalDatasets/NPS_Public_Roads/FeatureServer. (Accessed 1 November 2023).

NPS3, 2023. National park Service GIS data Services directory public trails. https://mapservices.nps.gov/arcgis/rest/services/NationalDatasets/NPS_Public_Trails/FeatureServer. (Accessed 1 November 2023).

NPS4, 2023. National park Service GIS data Services directory public buildings. https://mapservices.nps.gov/arcgis/rest/services/NationalDatasets/NPS_Public_Buildings_Geographic/FeatureServer. (Accessed 1 November 2023).

NPS5, 2023. NPS national Wildland fire perimeters. https://services3.arcgis.com/T4QMspbfLg3qTGWY/arcgis/rest/services/NPS_National_Wildland_Fire_Perimeters_FLAT_view/FeatureServer. (Accessed 5 October 2023).

Ozsahin, E., Ozdes, M., Smith, A.C., Yang, D., 2022. Remote sensing and GIS-based suitability mapping of termite habitat in the African Savanna: a case study of the lowveld in Kruger National Park. Land 11 (6), 803. https://doi.org/10.3390/land11060803.

Pausas, J.G., Keeley, J.E., 2021. Wildfires and global change. Front. Ecol. Environ. 19 (7), 387–395. https://doi.org/10.1002/fee.2359.

Picos, J., Alonso, L., Bastos, G., Armesto, J., 2019. Event-based integrated assessment of environmental variables and wildfire severity through Sentinel-2 Data. Forests 10 (11), 1021. https://doi.org/10.3390/f10111021.

Romme, W.H., Turner, M.G., Wallace, L.L., Walker, J.S., 1995. Aspen, elk, and fire in northern Yellowstone Park. Ecology 76 (7), 2097–2106. https://doi.org/10.2307/1941684.

Sage, A.J., Genschel, U., Nettleton, D., 2020. Tree aggregation for random forest class probability estimation. Stat. Anal. Data Min. 13 (2), 134–150. https://doi.org/10.1002/sam.11446.

Scott, J.H., Thompson, M.P., Calkin, D.E., 2013. A Wildfire Risk Assessment Framework for Land and Resource Management. Gen. Tech. Rep. RMRS-GTR-315. Department of Agriculture, Forest Service, Rocky Mountain Research Station, U.S., p. 83. https://doi.org/10.2737/rmrs-gtr-315

Smith, C.W., Panda, S.K., Bhatt, U.S., Meyer, F.J., Badola, A., Hrobak, J.L., 2021. Assessing wildfire burn severity and its relationship with environmental factors: a case study in interior Alaska Boreal Forest. Rem. Sens. 13 (10), 1966. https://doi.org/10.3390/rs13101966.

Stavrakoudis, D., Katagis, T., Minakou, C., Gitas, I.Z., 2020. Automated burned scar mapping using sentinel-2 imagery. J. Geogr. Inf. Syst. 12 (3), 221–240. https://doi.org/10.4236/jgis.2020.123014.

Stegner, M.A., Turner, M.G., Iglesias, V., Whitlock, C., 2019. Post-fire vegetation and climate dynamics in low-elevation forests over the last three millennia in Yellowstone National Park. Ecography 42 (6), 1226–1236. https://doi.org/10.1111/ecog.04445.

Stevens-Rumann, C.S., Prichard, S.J., Strand, E.K., Morgan, P., 2016. Prior wildfires influence burn severity of subsequent large fires. Can. J. For. Res. 46 (11), 1375–1385. https://doi.org/10.1139/cjfr-2016-0185.

Szpakowski, D.M., Jensen, J.L.R., Chow, T.E., Butler, D.R., 2023. Assessing the use of burn ratios and red-edge spectral indices for detecting fire effects in the Greater Yellowstone Ecosystem. Forests 14 (7), 1508. https://doi.org/10.3390/f14071508.

USGS, 2022. U.S. Geological survey, 3D elevation Program 10-meter resolution digital elevation model. https://www.usgs.gov/3d-elevation-program/about-3dep-products-services. (Accessed 1 October 2022) (accessed October 1, 2022).

Westerling, A.L., Hidalgo, H.G., Cayan, D.R., Swetnam, T.W., 2006. Warming and earlier spring increase western US forest wildfire activity. Science 313 (5789), 940–943. https://doi/10.1126/science.1128834.

Woodall, C.W., Smith, J.E., Miles, P.D., 2006. Attributes of standing dead trees in forests of the United States. Proceedings of the Eighth Annual Forest Inventory and Analysis Symposium. US Department of Agriculture, Forest Service, Washington, DC, pp. 191–196. Monterey, CA. Gen. Tech. Report WO-79. https://www.nrs.fs.usda.gov/pubs/gtr/gtr_wo079/gtr_wo079_191.pdf. (Accessed 20 November 2023).

You, W., Lin, L., Wu, L., Ji, Z., Zhu, J., Fan, Y., He, D., 2017. Geographical information system-based forest fire risk assessment integrating national forest inventory data and analysis of its spatiotemporal variability. Ecol. Indicat. 77, 176–184. https://doi.org/10.1016/j.ecolind.2017.01.042.

Zeppel, M.J., Anderegg, W.R., Adams, H.D., 2013. Forest mortality due to drought: latest insights, evidence and unresolved questions on physiological pathways and consequences of tree death. New Phytol. 197 (2), 372–374. https://doi.org/10.1111/nph.12090.