Isolating the Primary Drivers of Fire Risk to Structures in WUI regions in California

Michael Gollner
mgollner@berkeley.edu
University of California, Berkeley https://orcid.org/0000-0002-6925-4020
Maryam Zamanialaei
University of California Berkeley
Daniel San Martin
Universidad Tecnica Federico Santa Maria
Maria Theodori
University of California Berkeley
Dwi Purnomo
University of California Berkeley
Ali Tohidi
University of Maryland College Park https://orcid.org/0000-0001-7511-9274
Chris Lautenberger
Cloudfire
Arnaud Trouve
University of Maryland
Yiren Qin
University of Maryland

Article

Keywords:

Posted Date: February 14th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5776626/v1

Additional Declarations: There is **NO** Competing Interest.

Isolating the Primary Drivers of Fire Risk to Structures in WUI regions in California

Maryam Zamanialaei[1], Daniel San Martin[2], Maria Theodori[1], Dwi Marhaendro Jati Purnomo[1], Ali Tohidi[3], Chris Lautenberger[4], Yiren Qin[5], Arnaud Trouvé[5], Michael Gollner[1]

[1] Department of Mechanical Engineering, University of California, Berkeley, USA

[2] Departamento de Informática, Universidad Técnica Federico Santa María, Valparaíso, Chile

[3] Department of Mechanical Engineering, San Jose State University, San Jose, USA

[4] CloudFire Inc., 985 Lincoln Way, Suite 207 Auburn, CA 95603, USA

[5] Department of Fire Protection Engineering, University of Maryland, College Park, MD 20742, USA

Abstract

The destructive impacts of Wildland-Urban Interface (WUI) fires on people, property and the environment have dramatically increased, especially in California. Critical factors influencing structure protection during wildfires, including home hardening (e.g., vents, siding, roof, eaves, window, construction year), defensible space (vegetation and surrounding features), exposure to flames and embers, and structure separation are well known but their interrelated impacts are not quantified. Here, we find that structure separation and exposure significantly influence the probability of loss, underscoring the role of large conflagrations in driving widespread destruction. Machine learning models combined with previously unavailable exposure data enhance the predictive accuracy of structure loss up to 82%. Home hardening and defensible space, especially closest to the home, are still vital and effective mitigation measures, cutting hypothetical structures losses by 52%. Our results offer data-driven actionable insights for prioritizing mitigation strategies to protect vulnerable communities and structures in the WUI.

Main

Globally, the frequency, severity, and size of wildland fires has been increasing, resulting in extreme events that have led to dramatic losses in terms of people, property and the environment (Moritz et al., 2014; Stephens et al., 2013). A majority of these impacts on people occur where houses and other urban development intermingle with undeveloped wildland vegetation, an area termed the WUI. This area has grown dramatically in recent years (Alexandre, Stewart, Mockrin, et al., 2016; Schug et al., 2023), with one-third of all new homes in the US built in the WUI (Schoennagel et al., 2017). The western United States has witnessed a 246% increase in structures lost to wildfires from 2010-2020 compared to the previous decade (Higuera et al., 2023). California, despite its long fire history, has experienced recent increases in the number of very large fires (over 100,000 ha) resulting in massive losses of lives and property (Keeley & Syphard, 2021). Between 2013 and 2018, approximately 47,000 structures have been damaged or destroyed and 189 fatalities have been attributed to wildfires in California (Kramer et al., 2019). This increasing risk has significant consequences that jeopardize the economic stability, well-being of local residents, and the environment in affected communities (Kearns et al., 2022).

Central to preventing future destruction has been the development of mitigation measures aimed at reducing the likelihood of ignition and spread in the WUI (Calkin et al., 2014). Both building materials (hardening) and surrounding vegetation (defensible space) play important roles facilitating fire spread into the WUI (Cary et al., 2009; Cohen, Jack D, 2000; Maranghides & Mell, 2013; Quarles et al., 2010) but differ in their characteristics because structures and vegetation have different heat release rates, durations of burning, and responses to external exposure including direct flame contact, radiation, and firebrands (Caton et al., 2017). While effective mitigation strategies have been developed based on past testing and investigations (Maranghides et al., 2022), their combined effectiveness under different exposure conditions is not yet known (Schoennagel et al., 2017). Previous geospatial studies have demonstrated the critical influence of spatial arrangement and biophysical factors (Syphard et al., 2012, 2017; Alexandre et al., 2016), with defensible space around structures playing a significant, albeit secondary, role (Syphard 2014; Mockrin et al., 2023). The role of building materials has also been examined, revealing mixed findings (Syphard & Keeley, 2019; Mockrin et al., 2023; Troy et al., 2022). For instance, Syphard & Keeley (2019) found structural features like enclosed eaves and vent screens were crucial, while others (Price et al., 2021, Metz et al., 2024, Knapp et al., 2021) identified factors such as spacing and arrangement as more significant, suggesting determinants of loss are often beyond homeowners' control. Despite these advances, many studies focus on single events or lack comprehensive data on structure features and exposure conditions.

Here, we combine the largest existing structure loss database from California with simulated fire and ember exposure conditions to structures across multiple large-loss events, compare the effects of structure hardening and nearby defensible space, and apply this to build a predictive model. Unlike past studies, fire reconstruction modeling that includes urban fire spread is used to quantitatively add the effect of flame and ember exposure on structures. Geospatial assessments of vegetation surrounding structures are added using both LiDAR and visual imagery to assess the level of defensible space (vegetation) surrounding structures. The database is then fit using a multivariate analysis similar to Mockrin et al. (2023) and Metz et al. (2024) that distinguishes between the interrelated effects of exposure, structure hardening, and defensible space. A parameter importance analysis reveals the strong role both structure separation and exposure play, distinguishing wildfire from other natural hazards that are not affected by neighboring conditions, highlighting the importance of a community approach to mitigation. The model developed is strongly predictive when incorporating all the above features and is also used to assess the impact of recommended mitigation measures on homes. It is found that it is necessary to make changes both to the structure itself and surrounding vegetation, especially that closest vegetation within 1.5 m (5 ft) of the structure (zone 0) to achieve the maximum benefit.

Results

In this study we took advantage of the Damage INSpection (DINS) Dataset collected by on-the-ground CAL FIRE crews from structures damaged, destroyed, or affected by wildfires in California during post-fire investigations between 2013 - 2022 (California Department of Forestry and Fire Protection (CAL FIRE), 2024). Figure 1 shows all fires in the DINS dataset between 2017 – 2022 as well as five of the largest loss fires in this dataset (2017 Tubbs and Thomas, 2018 Camp, 2019 Kincade, and 2020 Glass fires) selected for further analysis based on data availability and the number of structures exposed and destroyed. We combined records of damage state and building features from this dataset with remotely-sensed assessment of surrounding vegetation (akin to defensible space) and structure footprints (to assess building separation) of undamaged, damaged, and destroyed structures within the final fire perimeter (CAL FIRE Historic Fire Perimeters), including a 91 m (300 ft) buffer around any burned areas (Hakes and Theodori et al., 2022). Post-fire reconstruction modeling was then used to add local fire exposure by both flames (flame length) and embers (ember load) to the dataset resulting in a more complete picture of fire exposure and effects.

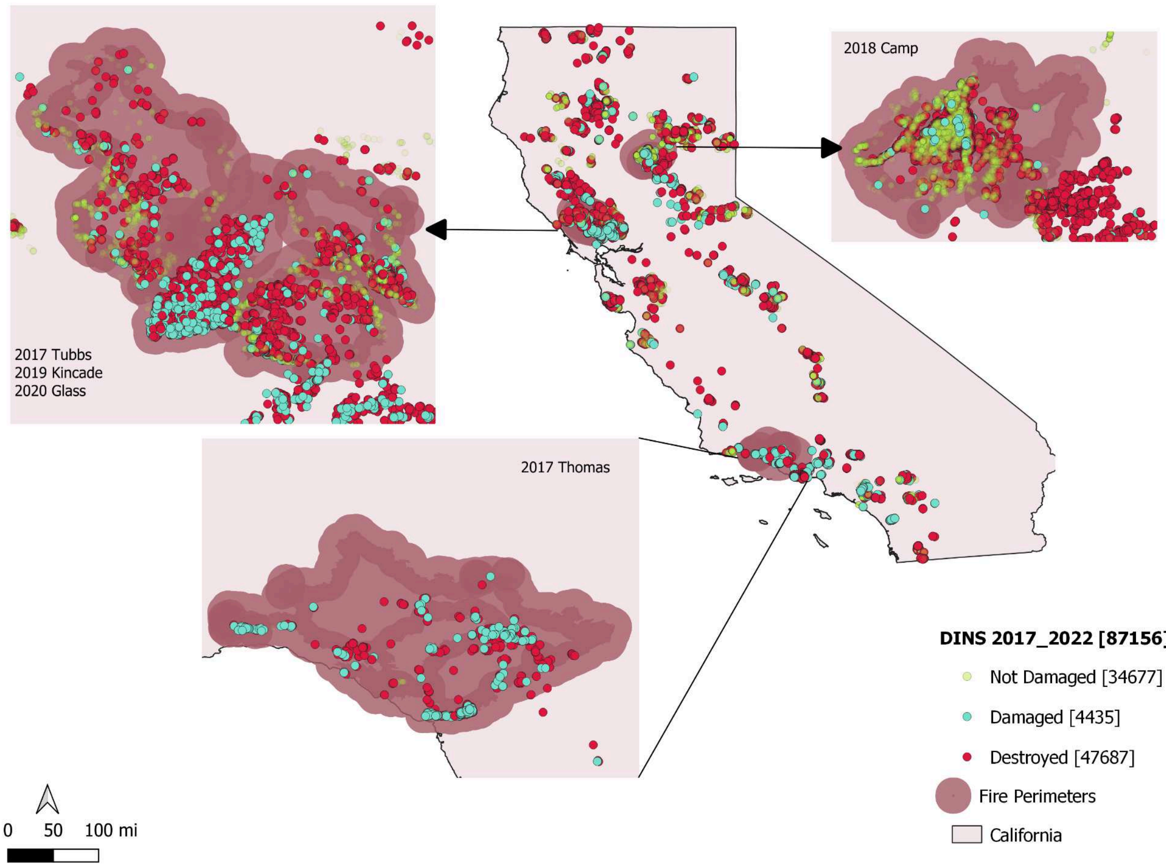

Figure 1- Spatial distribution of damage to structures in California, emphasizing five of the most destructive WUI fires.

Data from 5 selected fires were extracted from the overall DINS dataset (~90,000 structures) by combining/stacking the five fire datasets after preprocessing. Additional structures that were unburned but exposed to fire were added to the dataset (Tubbs ~14,000, Thomas ~ 6,000, Camp ~24,000, Kincade ~2,000, and Glass ~5,000 structures). We employed a resampling process to balance the samples, resulting in a total of approximately 47,000 structures and 45,947 unique data points. We simplified the damaged, non-damaged, and destroyed classifications in the original DINS to a binary classification of Survived and Damaged categories because >90% of damaged structures are destroyed.

Post-Fire Reconstruction

Five fires were reconstructed using a level-set model (ELMFIRE) that included both wildland (Lautenberger, 2013, 2017) and urban fire spread (Hamada, 1975) to re-create fire spread conditions and estimate critical missing exposure data (flame length and ember deposition) from these events. Figure 2 shows the modeled fires and resulting flame length (in meters) and ember load (in terms of number of embers deposited per meters squared). These are extracted adjacent to each of 47,000 structures in our dataset and distributions are shown in terms of flame and ember exposure as probability density functions (PDF) in Figure 3. These distributions reveal a 27% and 39% overall decrease in exposure to flames and embers, respectively for structures that survived vs. those that were damaged. The decrease in exposure, however, is small in comparison to the difference in total number of structures destroyed and suggests that other features may play a role in determining which are more or less likely to survive.

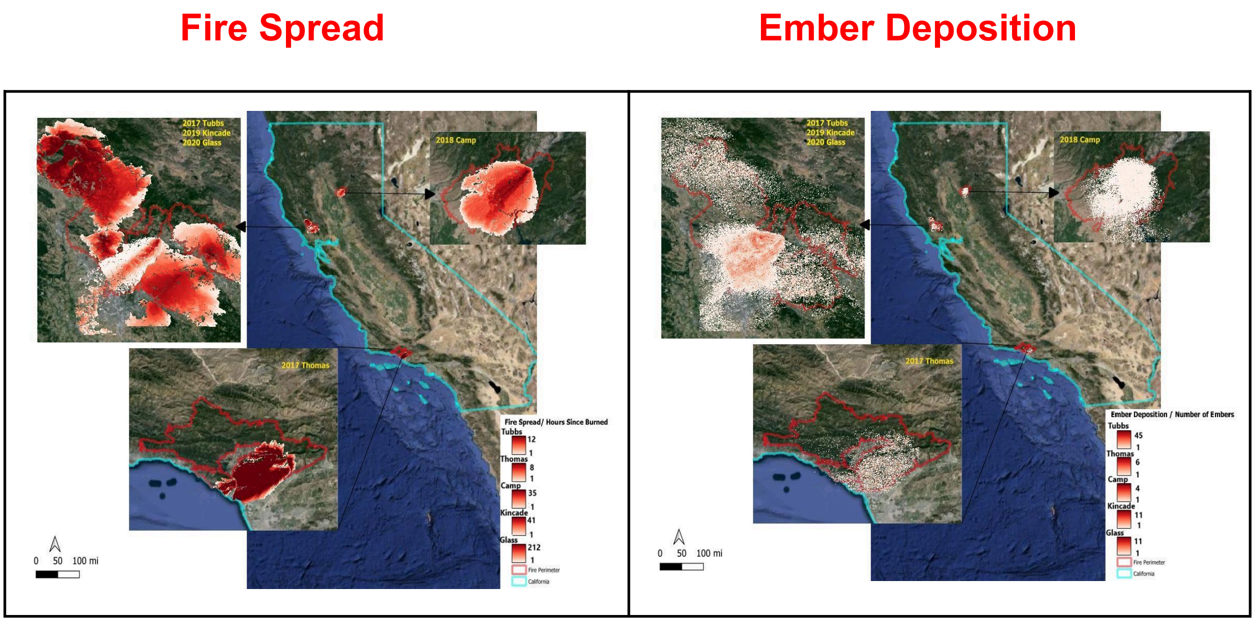

Figure 2- Ember deposition and fire spread simulation of the 5 fires

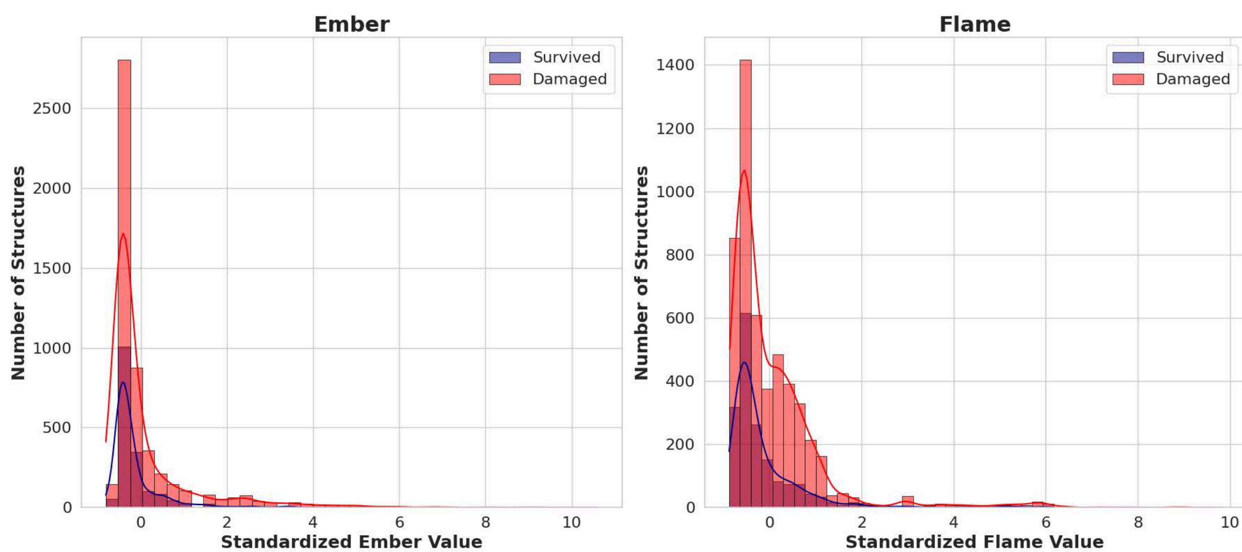

Figure 3- Number distribution showing structure damage based on standardized flame and ember values. These findings highlight the significant impact of simulated ember and flame exposure on the destruction of structures in large WUI fires.

Feature Contribution to Structure Loss

We applied an XGBoost Classifier to our dataset and then utilized a SHapley Additive exPlanations (SHAP) model to explore the importance of various features on structure destruction. By looking at the stacked results from all 5 fires (Figure 4), we found that structure density, which is determined by the distance between structures (SSD), is one of the most important features in structure destruction. The second most important contributor to the classification results from the XGBoost estimator was exterior siding, representative of the materials used in construction, followed by Year Built. Note, in the DINS database year built indicates the year that the primary structure in the parcel was constructed. Year built has therefore been identified as a confounding variable ultimately combining the effects from different parameters such as hardening (e.g., materials used for roof construction, eaves, vent screen, window pane, exterior siding), vegetation, and surrounding features (e.g., defensible space/vegetation separation distance and nearby structures/structure separation distance). Results underscore the importance of hardening structures, structure density, and building arrangements in WUI areas to mitigate fire risk and potential destruction. These results are consistent with already-established engineering knowledge (Hakes et al., 2017; Knapp et al., 2021; Maranghides et al., 2022). Furthermore, these insights are based on the available data rather than being drawn from direct experiments or detailed numerical simulations at the flame scale.

Exposure was still important in predicting damage from past WUI fires, specifically considering flame length and ember load derived from fire spread simulations. Flame length, which indicates the height and intensity of flames, can directly influence the severity of damage to structures and vegetation in its path. Ember load, representing the number and size of burning embers carried by the wind, also significantly contributes to the spread of the fire and subsequent structure loss, as these embers can ignite spot fires far beyond the main fire front. The intensity and reach of flames, as well as the quantity of embers, played significant roles in the extent of damage and structure loss observed in these fires.

SHAP values provide a unified measure of feature importance in a predictive model. Based on the evaluation metrics, the XGBoost model emerged as the most effective estimator, demonstrating superior skill in predicting losses. This finding is supported by its higher average SHAP values for key features compared to other models such as Logistic Regression and Random Forest. The average SHAP values for the XGBoost estimator revealed that certain features significantly impacted the model's predictions. For instance, Structure Density (SSD) and flame length had positive average SHAP values of 0.090 and 0.051, respectively, underscoring the importance of building arrangements and fire behavior in risk assessment. Year Built showed an average SHAP value of -0.058, suggesting that newer structures might be associated with lower predicted losses, possibly due to improved building codes and materials.

We also broke down the feature importance results for each of the 5 individual fires assessed (Figure 5), and found common features of importance: structure density (SSD), flame length, and year built, but distinct differences were also revealed between individual fires. The structure separation distance (SSD) was the most important predictive feature in the 2017 Thomas, 2018 Camp, 2019 Kincade and 2020 Glass fires while flame length was the most predictive feature for the 2017 Tubbs fire. In fires that burn through densely populated areas, wildfires can transition into urban conflagrations that become dominated by structure-to-structure spread, which is most strongly influenced by SSD. The high density of structures in the Thomas and Camp fires in the cities of Paradise and Ventura (Butte and Ventura counties) therefore emphasize this mode of spread. In the Kincade and Glass fires, the clustering of structures in Geyserville (Sonoma County) and Deer Park (Napa County) contributed to the rapid urban spread of the fire. Flame Length significantly contributes to structure destruction in the Tubbs fire and is the second most important factor for the Thomas, Camp, and Kincade fires. In the Glass fire, it ranks third in importance, emphasizing the role of nearby buildings and surrounding fuels in spreading flames to structures. Year built, in conjunction with building characteristics (eaves, roof, vent, siding, window), underscores the significance of home hardening in dense WUI areas, limiting fire spread and protecting structures from losses.

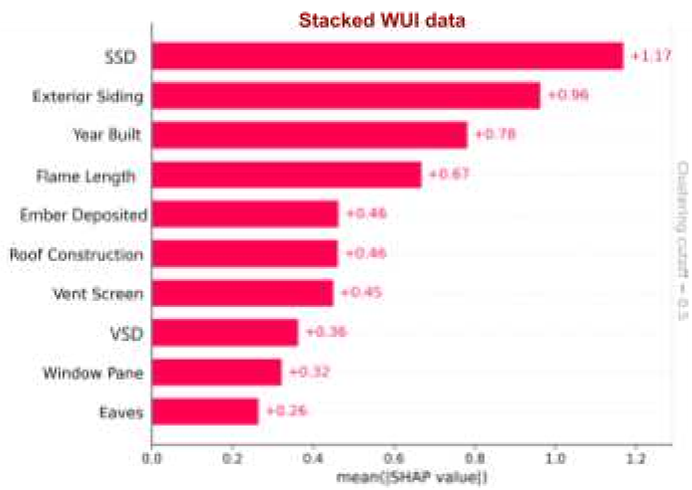

Figure 4- SHAP aggregation results for characterizing the contribution of features in estimations from an XGBoost classifier for the entire (stacked) WUI data from 5 fires.

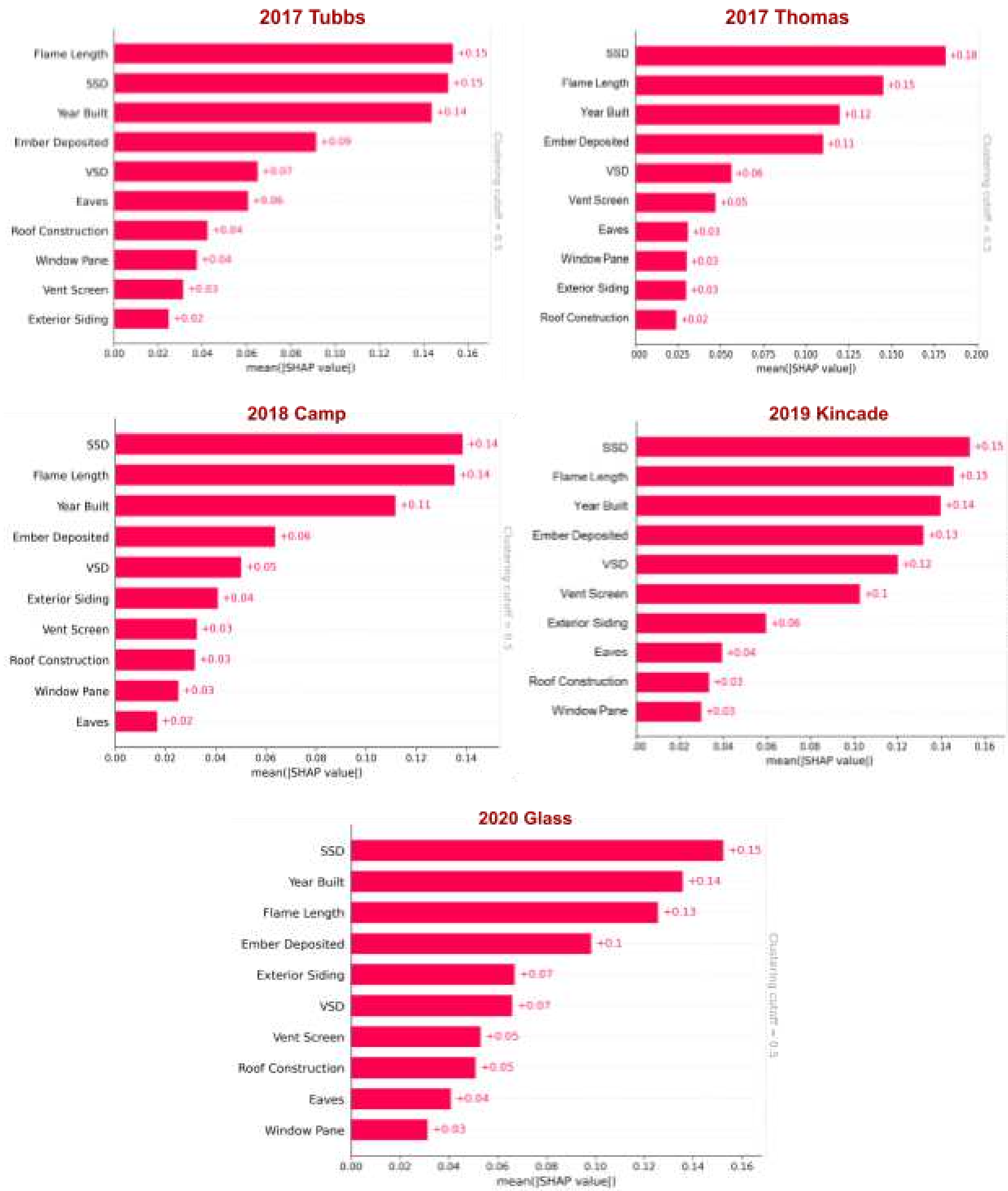

Figure 5- SHAP aggregation results for characterizing the contribution of features in estimations from an XGBoost classifier for 5 individual fires.

Figure 6 shows the distribution of four key features—SSD (structure separation distance), FLAME (flame length), YEAR BUILT (year primary structure on parcel was built), and EMBER (ember load)—across five fires: Tubbs, Camp, Glass, Kincade, and Thomas. Each panel represents one feature, displaying both the distribution of values through a violin plot (in light gray) and the mean values (in blue) overlaid with standard deviation error bars. The violin plots highlight the density of feature values, with wider sections indicating regions where values are more concentrated, allowing for a comprehensive comparison of how each feature behaves across different fires. For instance, the SSD and YEAR BUILT features show relatively wider distributions in the Glass fire, suggesting a broader range of structure separation distances and building years compared to other fires. Overlaid bar plots show the mean feature values for each fire, providing insight into the central tendencies. For example, the SSD and YEAR BUILT features have relatively higher means in the Camp and Glass fires, indicating greater separation distances and older structures on average in these regions. In contrast, the Kincade and Tubbs fires exhibit lower mean SSD values, suggesting tighter structure spacing. The combination of these two plots makes it possible to assess not only the average feature importance (through the bar plots) but also the variation within each fire (through the violin plots). This dual approach reveals both the central trend and the spread of values, offering a nuanced understanding of feature importance in the context of each fire.

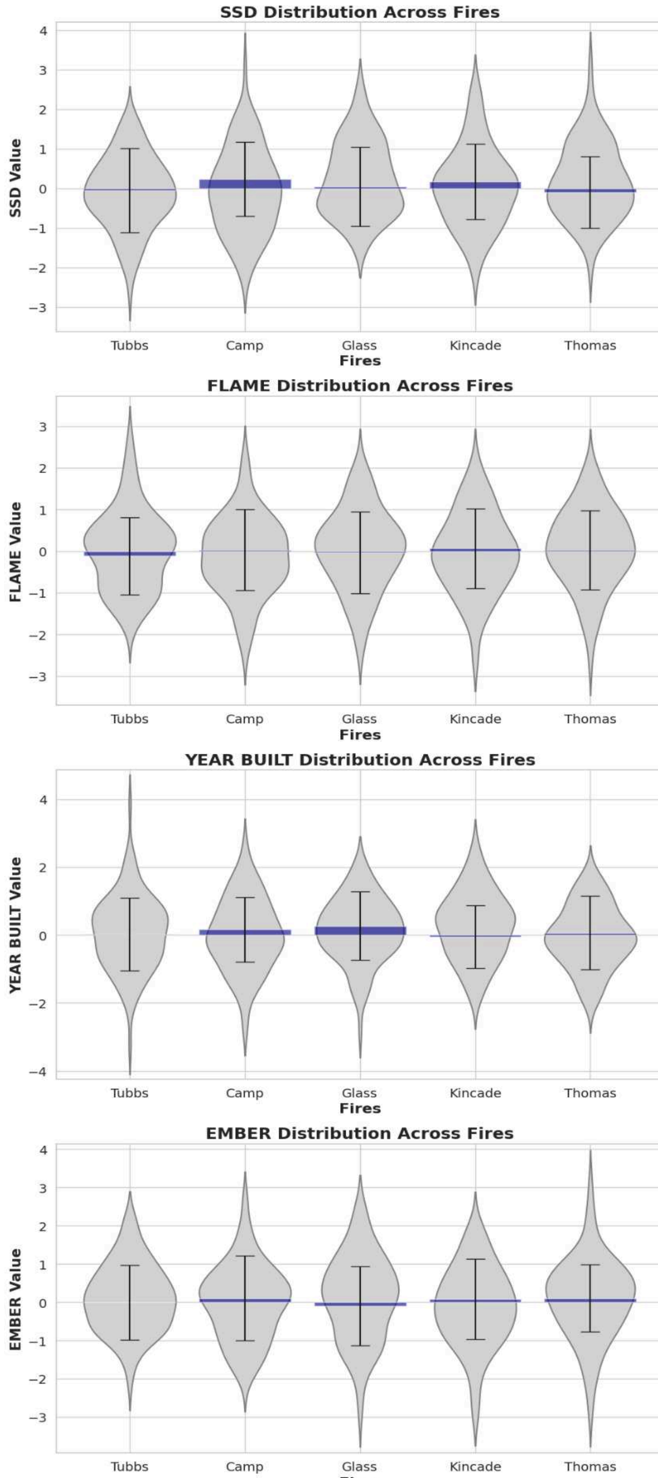

Figure 6- The distribution (gray) and mean values (blue) of important features across five fires

Damage Prediction Results

We applied a range of machine learning models, ultimately selecting an XGBoost Classifier as it was the most accurate to investigate the five large WUI fires in our dataset to predict structure survival during each fire. Linear models (logistic regression) have been used in the past (Syphard et al., 2017; Syphard & Keeley, 2019) and achieved an accuracy to predict structure losses for our 5-fire database of 78%, however the XGBoost Classifier achieved an accuracy of 82% for the 5-fire database. A confusion matrix is shown in Table 1 outlining the performance of our classification algorithm breaking down predicted outcomes against actual results, delineating true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The area under an ROC curve (AUC) is also shown as another measure to evaluate the model's overall performance alongside the accuracy (percentage of correct predictions the model makes). Overall the XGBoost Classifier has a high accuracy in predicting the occurrence of destruction for each of the 5 individual WUI fires, despite severe limitations in data coming from the aftermath of real destructive events. An accuracy threshold of 79% is achieved for each fire except for the Kincade fire (63%) which had a large number of missing values in the DINS inspection dataset and affected the results. The XGB classifier was also applied to the full DINS dataset (2017-2022) which incorporated the same preprocessing as the 5 fires but did not include exposure modeling values and defensible space, although it did incorporate all other analyses including structure spacing, and year built. An accuracy of 77% was achieved which demonstrates the flexibility and applicability of this model even when not all data can be accounted for.

A comparative analysis between the Logistic Regression, Random Forest, XGBoost and CatBoost models is included in the methods section but importantly underscores the need for selecting an appropriate algorithm based on the specific characteristics of the dataset and outcome to be achieved.

Table 1- Results of XGBoost predictions on each test set with a resulting confusion matrix displaying model performance in terms of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), area under an ROC curve (AUC), and the percentage of correct predictions the model makes (Accuracy).

WUI Fire	TP	FP	TN	FN	AUC	Accuracy
Tubbs	1041	58	0	2	0.685	0.94
Thomas	147	25	31	21	0.808	0.79
Camp	3506	686	198	110	0.784	0.82
Kincade	27	58	124	27	0.635	0.63
Glass	151	58	496	106	0.841	0.79
5 Fires Combined	4785	847	885	353	0.833	0.82

WUI Fire	TP	FP	TN	FN	AUC	Accuracy
All CA DINS (2017-22)	5198	133	2998	1073	0.84	0.77

Using our model we were able to examine various scenarios including home hardening and defensible space clearing to compare what changes in predicted structure loss and survivability might occur, in order to propose effective mitigation strategies. This is particularly important because structure density cannot be modified for existing structures (which make up more than 98% of the current housing stock (Maranghides et al., 2022; Syphard et al., 2021). We applied this to the 5-fire database we created. In the first scenario, which involved home hardening, we adjusted all hardening values in our dataset to fire-safe ones (e.g. non-flammable siding, fine mesh over vents, double paned windows, non-flammable roof, etc.) and applied the XGB model. This resulted in a 25% survival rate with 75% structure loss due to WUI fires (Fig 7, Fig S5). Next, we combined home hardening with clearing defensible space in Zone 0 (0-5 feet; 0-1.5 meters), which effectively doubled the survival rate to 40% and reduced the loss rate to 60% (Fig S7, Fig S6). Finally, we implemented an extreme mitigation scenario that included both home hardening and the clearing of defensible space in Zone 0 (0-5 feet; 0-1.5 meters) and Zone 1 (5-30 feet; 1.5-9 meters). This further increased the survival rate to 48% and reduced the structure loss to 52% (Fig 7, Fig S7). Figure 7 shows the structure loss and survivability across various mitigation scenarios.

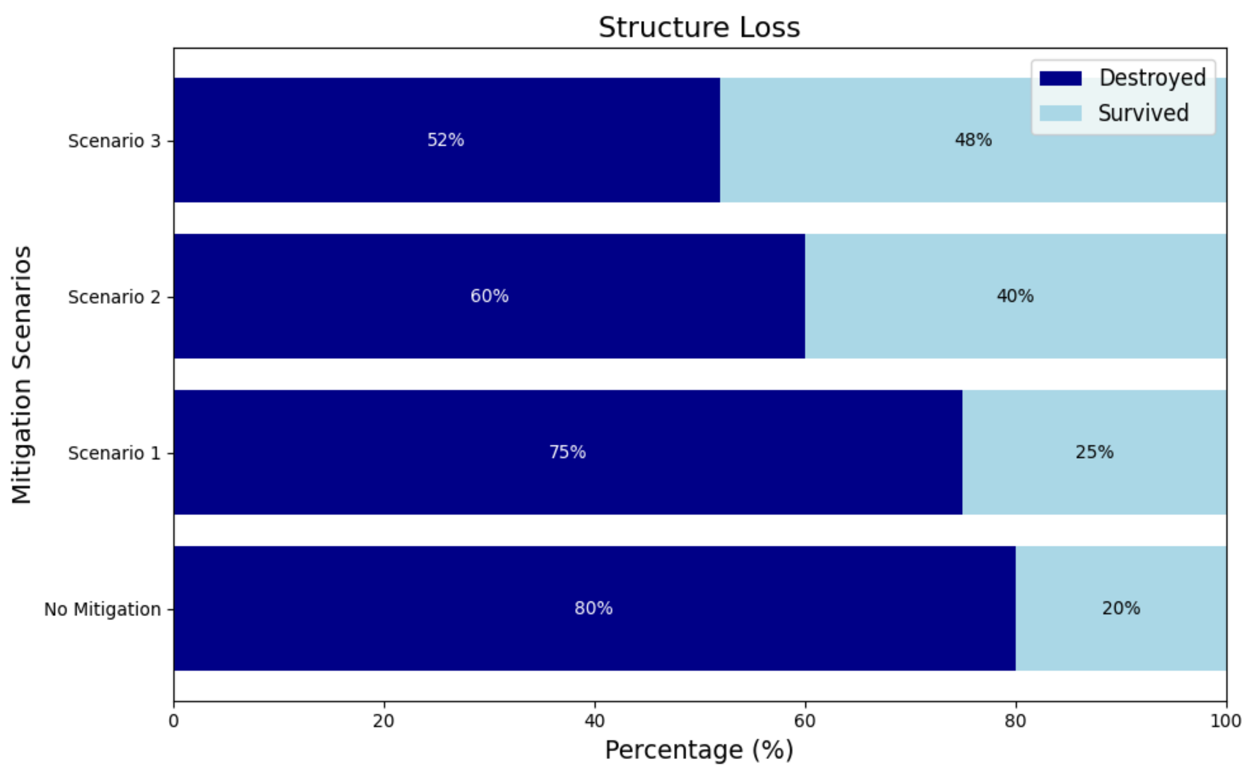

Figure 7- Probabilities of structure destruction in various mitigation scenarios. Scenario 1 (add

hardening), Scenario 2 (add hardening and clear zone 0), Scenario 3 (add hardening and clear zones 0 and 1).

Discussion

Decades of research have shown the importance of ignition-resistant construction, defensible space, and the proximity of structures to one another (Cohen, Jack D, 2000; Insurance Institute for Business & Home Safety, 2021; Knapp et al., 2021; Maranghides et al., 2022). The ranked importance and interplay between these mitigation measures has now been presented in this study, utilizing simulations to extract exposure conditions during different fires. The application of XGBoost and SHAP methods has illuminated the critical features contributing to structure destruction. Following investigations of the Camp fire by Maranghides et al. (2022) and Knapp et al. (2021), Structure Separation Distance (SSD) arose as a key metric in characterizing the likelihood of loss for any particular structure during these large WUI fires. While smaller fires may occur through sparse housing arrangements, the majority of structure losses in California have occurred in large-loss fires in moderately dense (suburban) communities (Knapp et al., 2021)(Syphard & Keeley, 2019) communities. In these fires, the structures themselves become fuel and contribute to spread. These existing structures pose a unique challenge in hazard management—they are immobile. While these structures can be hardened, they cannot be readily removed or displaced like many other WUI fuels.

The analysis revealed the role of interactions among competing factors (like SSD, flame length, ember load, year built, and exterior siding) in influencing fire dynamics, showcasing that multiple features contribute simultaneously to structure risk. Our results show that the most important factor that cannot be changed is the distance between structures, as conflagrations tend to consume a majority of houses in major fires. Nevertheless, there is still a substantial opportunity to enhance safety through effective mitigation measures such as hardening structures and establishing a defensible space. Mitigation measures on the structure (hardening) combined with removal of surrounding fuels in the area immediately adjacent to the structure (zone 0) has the potential to dramatically reduce losses in future fires. While applying these measures to any particular structure within a dense urban area makes little difference on the survivability of a single home, significant reductions in losses are achievable when community-wide actions can be applied. This has been proposed in many studies and is a major tenant of Firewise and other community risk-reduction programs, however has not been shown to be effective before because previous studies focus more on individual structures. The effect of community risk reduction may have other benefits as well, with amplified effectiveness to responding fire crews (Syphard & Keeley, 2019). When fewer homes ignite from embers or direct flame contact, less structure-to-structure spread results and the fire service is freed up to focus on those structures that are most threatened. If arriving early in fire progression, it is possible that embers can be extinguished and the “disaster sequence” posited by (Calkin et al., 2014) can be disrupted and a smaller number of homes may be lost.

While hardening and defensible space actions may not alter the fundamental risks posed by structure proximity, they can still dramatically improve the survivability of buildings during wildfires. By examining individual fire cases, we can further tailor mitigation approaches to address specific vulnerabilities and enhance resilience against future wildfires. This is apparent when we see that the factors most correlated with structure destruction change for some fires, such as the Tubbs fire where flame length played a more significant role than structure spacing overall. During this fire sparser structures could potentially have benefited from additional fuel management to reduce fire exposure, although some areas, such as the Coffey Park community were dense and still dominated by structure separation. While siding materials were not significant in any one fire, as an overall predictor they were very significant, which speaks to the fact that clustered structures on each fire often had similar construction materials, built year, etc. so that those factors do not appear as significant.

Overall we have shown the potential to combine extensive on-the-ground post-fire data collection, analysis of remotely sensed data, and fire reconstruction modeling to better understand the complicated interactions between different features and structure survivability during wildfires. Despite the improvement of these data and modeling tools, we still have a dramatic lack of information before and during fires that is desperately in need of improvement to improve future investigations. Pre-fire inspections are limited and were not available at scale to aid this study, therefore much of the data was collected after the fire and is not at the fine scale to distinguish all potential factors that play a role in destruction. For instance, year built is an inter-related term that also corresponds to materials used, construction type, building codes, etc. It is still useful because it is one of the most easily obtained factors for future analysis, but it makes it harder to distinguish between other factors. During the fires it would also have been useful to observe failure modes more directly, e.g. observation of what part of the exterior ignites by embers, structure-to-structure spread, etc. Still, this study provides a broader understanding useful to the field now and a framework for future data to be applied.

Methods

We primarily relied on a modified database from 5 selected fires that includes more than 47,000 structures with two broad damage states: “Survived” and “Destroyed”, and five detailed damage states: “Destroyed (>50%)”, Damaged (“Major (26-50%)”, “Minor (10-25%)”, “Affected (1-9%)”), “No Damage”. The CAL FIRE Damage INSpection Program (DINS) was founded with the goal to collect data on damaged, destroyed, and unburned structures during and immediately after fire events to assist in the recovery process, and to provide local governments and scientists information for analyzing why some structures burned and why some survived (Henning, Andrew et al., 2015). Through a public records request, we acquired DINS data for more than 90,000 structures that survived, were damaged, or were destroyed across all California wildfires from 2013–2022, making this potentially the largest combined dataset of its sort. We then incorporated risk factors associated with structure destruction by wildfires to the DINS data to gain a deeper understanding of WUI destruction. These factors include structure density, building materials, year built, defensible space, and exposures to structures (fire intensity and ember). We employed several Machine Learning (ML) techniques to identify and highlight the important features in our WUI data. These techniques included feature selection, feature engineering, and model interpretation methods to ensure we could pinpoint the most significant variables influencing our results. To enhance the performance of the ML model in this study, we implemented a range of data preprocessing techniques such as data cleaning, normalization, and encoding. These preprocessing steps were crucial for improving model accuracy, reducing noise, and ensuring the robustness of our findings. By meticulously preparing the data, we ensured that the ML model could effectively learn and make accurate predictions from our complex WUI dataset. We opted for the XGBoost (eXtreme Gradient Boosting) algorithm for our ML model due to its superior performance over other methods on our dataset. We also leveraged the SHAP (SHapley Additive exPlanations) model, which is providing a nuanced understanding of each column's contribution to the overall predictive outcome. This technique allowed for a comprehensive assessment of the importance of variables within the dataset, enhancing the robustness and reliability of our analysis. The results of Confusion Matrices and Receiver operating characteristics (ROC) Curves, in addition to advanced computational framework, allowed us to delve into the intricacies of the dataset, capturing complex relationships and patterns that might not be discernible through conventional methods. Our evaluation extended beyond a generalized assessment, as we calculated the accuracy and sensitivity metrics for each individual fire and aggregated the results to encompass all structures within the damage dataset. This meticulous analysis not only provided insights into the predictive performance of our model on a per-fire basis but also yielded a comprehensive understanding of its effectiveness across the entire spectrum of structures in the damage data.

Risk factors to structures from wildfires in WUI

The methodology for integrating risk factors related to structure destruction builds upon the combination of on-the-ground data with fire modeling reconstructions by Hakes and Theodori et al. (2022) for community-level risk assessment for the Tubbs fire, which includes:

Structure density which represents “Structure Separation Distance (SSD)”. We employed the Microsoft Maps dataset (available at https://github.com/microsoft/USBuildingFootprints), which encompasses open building footprints datasets for entire counties in the United States. This dataset comprises 129,591,852 computer-generated building footprints. Additionally, we utilized QGIS software to access geospatial data concerning urban infrastructure, building locations, and their spatial interconnections.

The ***year built*** refers to the year in which the primary structure on a parcel of land was constructed. In the context of analyzing the impact of WUI fires, the Year Built variable is significant because the age of a structure can influence its susceptibility to fire damage.

Furthermore, it acts as a confounding variable that can affect both the building features and the extent of damage.

Concerning fire safety in ***building construction materials***, numerous in-depth studies have been carried out through meticulously planned laboratory tests (Manzello et al., 2012; Quarles et al., 2010). Despite the solid laboratory evidence, few empirical studies have documented building characteristics associated with structure loss in real wildfire situations (Syphard & Keeley, 2019). In this study building characteristics include eaves, vent screens, exterior siding, roof construction, and window panes.

In terms of ***defensible space***, which is representing in this study as "Vegetation Separation Distance (VSD)", the state of California requires fire-exposed homeowners to create a minimum of 30 m (100 ft) of defensible space around structures, and some localities are beginning to require at least 60 m (200 ft) in certain circumstances (Syphard et al., 2014). We established three categories for the Vegetation Separation Distance (VSD): Zone0, which comprises the initial five feet from the building or "0-5"; Zone1, encompassing the area within 30 feet of the building or "5-30"; and Zone2, extending to within 100 feet of the building or "30-100" (CAL FIRE DSpace: https://www.fire.ca.gov/dspace). Remote sensing techniques were utilized to analyze the density and distribution of vegetation in the WUI regions and urban settings, extracting valuable insights from the aerial and satellite imagery and LiDAR data. The publicly available datasets (including countywide LiDAR data and a fine scale vegetation and habitat map) which produced by the Sonoma County Agricultural Preservation and Open Space District and the Sonoma County Water Agency, provide an accurate, up-to-date inventory of the county's landscape features, ecological communities and habitats (Sonoma County Vegetation Map: https://sonomavegmap.org/).

Exposures including fire intensity (flame length) and firebrand (ember load). Houses are destroyed during wildfires when exposed to flames in adjacent fuel, radiant heat from nearby fuel (≤40m) (Cohen, Jack D, 2000), or airborne embers and firebrands originating in nearby and distant fuel (typically<10 km) (Koo et al., 2010; Noble et al., 1980). In this study, we used the Eulerian Level set Model of FIRE spread, ELMFIRE, an operational fire behavior and spread simulation tool (Lautenberger, 2013) for its additional capability in simulating ember deposition of multiple embers and its implementation of Monte Carlo analysis (Lautenberger, 2017) to capture the stochasticity and uncertainty inherent in wildland fire modeling. We used and modified the semi-physical model of (Lautenberger, 2017; Purnomo et al., 2024) to include urban fire spread by using the empirical approach of HAMADA (Hamada, 1975).

Data preprocessing

To predict the damage for any of the fire datasets, the dataset was divided into the target variable or y, and all the other features as inputs or X. A stratified split was executed based on 'y' values, allocating 80% of the data for training purposes and reserving the remaining 20% for the testing set. This strategic partitioning ensured that the distribution of target variable classes remained proportionate in both the training and testing subsets, enhancing the reliability of the model's performance evaluation.

To address the noteworthy variations in the scales of the model inputs, a vital preprocessing step was implemented prior to model training. Using the scikit-learn package, we first designed imputation strategies through IterativeImputer to handle missing values. These strategies were trained on the training set and then applied to both the training and test sets. Next, we normalized the numerical variables using StandardScaler, ensuring that they were on a similar scale, which helps in the convergence and performance of various models. Additionally, we conducted OneHotEncoding and Label Encoding on categorical variables using OneHotEncoder and LabelEncoder from scikit-learn to convert them into a numerical format that can be understood by the models. Class balance is achieved through the binarization of different labels/classes with damaged and not damaged/survived. This approach is essential, particularly in scenarios where certain damage classes may be underrepresented. This preprocessing pipeline allowed us to use a variety of models on the dataset, ensuring compatibility and enhancing the overall performance of the models.

In essence, this procedure, encompassing data categorization, stratified splitting, imputation, standard scaling, OneHotEncoding/Label Encoding, and resampling, laid the foundation for a robust and unbiased evaluation of the model's predictive capabilities regarding fire damage across diverse datasets.

Machine Learning Techniques

Machine learning (ML) methods have recently been applied to wildland fire (Jain et al., 2020) and present an ideal platform for WUI fires as interactions between competing factors can be fit and modeled. In this work, we employed both regression and classification ML techniques to our combined dataset resulting in a predictive model for structure destruction based on home hardening (roof, siding, vents, eaves, window, year built), vegetation separation (defensible space and surrounding), exposure metrics (flames and embers), and structure spacing. The XGBoost (eXtreme Gradient Boosting) machine learning algorithm was chosen as it outperformed other methods on our dataset. The model hyper parameters were tuned using Grid Search and Randomized Search Cross Validation, depending on the number of the model's parameters. Hyper parameter selection is performed using the best result in terms of the following classification metrics: F-beta, accuracy, balanced accuracy and Matthews correlation coefficient. Finally, feature importance with SHAP aggregation analysis was utilized to quantify the contribution of each feature to the target variable. A higher feature importance score indicates that the feature has a greater influence on the model's prediction (Masís, 2023). The SHAP model connects optimal credit allocation with local explanations using the classic Shapley values from game theory and their related extensions (Lundberg & Lee, 2017). This was then applied to a unified framework for interpreting predictions to explain the output of any machine learning model.

Classifiers

We employed several classification models, including Logistic Regression and Random Forest (Pedregosa et al., 2012), and Gradient Boosting/XGBoost (Chen & Guestrin, 2016). Each of these models offers distinct advantages and methodologies for analyzing feature importance.

Logistic Regression is a generalized linear model used for classification problems (Bishop, 2006) and we use it as a base model to compare with more complex models. The second model used in this work is the Random Forest. Random Forests are a technique in ensemble learning utilized for tasks such as classification and regression. During the training, several decision trees are built. In classification, the random forest outputs the class chosen by the majority of trees (Ho, 1995). Finally, Gradient Boosting (GB) is a method in machine learning that employs boosting within a functional framework. The XGBoost (eXtreme Gradient Boosting) is a GB implementation that has been used as it outperformed other methods on our dataset. XGBoost is often preferable for developing predictive models for large datasets due to its accuracy, efficiency, and adaptability (Chen & Guestrin, 2016). Furthermore, XGBoost is a robust algorithm for both classification and regression problems. Due to its strengths in model prediction, XGBoost can be utilized for damage assessment to create predictive models for structure destruction.

Feature Contribution Through SHAP Analysis

While machine learning (ML) models are increasingly used due to their high predictive power, their use in understanding the data-generating process (DGP) is limited. Understanding the DGP requires insights into feature-target associations, which many ML models cannot directly provide, due to their opaque internal mechanisms. Feature importance (FI) methods provide useful insights into the DGP under certain conditions (Ewald et al., 2024). Furthermore, SHAP (SHapley Additive exPlanations) is a unified framework for interpreting machine learning models based on cooperative game theory (Lundberg & Lee, 2017). It assigns each feature an importance value for a particular prediction by computing the contribution of each feature to the prediction, averaging over all possible combinations of features. This approach ensures consistency and local accuracy, providing insights into how different features influence model predictions. SHAP values can explain individual predictions and provide a global understanding of the model's behavior, making it a valuable tool for model interpretability in research (Masís, 2023). We utilized SHAP interpretation analysis of feature importance to identify and understand the key factors driving structure destruction in WUI fires.

Confusion Matrix and ROC Curve for predictions

A confusion matrix summarizes the classification performance of a classifier with respect to some test data. It is a two-dimensional matrix, indexed in one dimension by the true class of an object and in the other by the class that the classifier assigns (Ting, 2011). Receiver operating characteristics (ROC) graphs are useful for organizing classifiers and visualizing their performance. A receiver operating characteristics (ROC) graph is a technique for visualizing, organizing and selecting classifiers based on their performance (Fawcett, 2006). We investigated the five large WUI fires in our dataset to predict structure survival during each fire by understanding the model's accuracy, and other key performance metrics. By analyzing the confusion matrices and ROC curves for each fire event, we were able to identify patterns and discrepancies in model performance, leading to a better understanding of the factors influencing structure survival in large WUI fires.

References

[1]. Alexandre, P. M., Stewart, S. I., Keuler, N. S., Clayton, M. K., Mockrin, M. H., Bar‑Massada, A., Syphard, A. D., & Radeloff, V. C. (2016). Factors related to building loss due to wildfires in the conterminous United States. *Ecological Applications*, *26*(7), 2323–2338. https://doi.org/10.1002/eap.1376

[2]. Alexandre, P. M., Stewart, S. I., Mockrin, M. H., Keuler, N. S., Syphard, A. D., Bar-Massada, A., Clayton, M. K., & Radeloff, V. C. (2016). The relative impacts of vegetation, topography and spatial arrangement on building loss to wildfires in case studies of California and Colorado. *Landscape Ecology*, *31*(2), 415–430. https://doi.org/10.1007/s10980-015-0257-6

[3]. Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 2).

[4]. California Department of Forestry and Fire Protection (CAL FIRE) (last). (2024). *CAL FIRE Damage Inspection (DINS) Data [Data set]*. https://calfire-forestry.maps.arcgis.com/home/item.html?id=994d3dc4569640caadbbc319

8d5a3da1)

[5]. Calkin, D. E., Cohen, J. D., Finney, M. A., & Thompson, M. P. (2014). How risk management can prevent future wildfire disasters in the wildland-urban interface. *Proceedings of the National Academy of Sciences*, *111*(2), 746–751. https://doi.org/10.1073/pnas.1315088111

[6]. Cary, G. J., Flannigan, M. D., Keane, R. E., Bradstock, R. A., Davies, I. D., Lenihan, J. M., Li, C., Logan, K. A., & Parsons, R. A. (2009). Relative importance of fuel management, ignition management and weather for area burned: Evidence from five landscape - fire - succession models. *International Journal of Wildland Fire*, *18*(2), 147. https://doi.org/10.1071/WF07085

[7]. Caton, S. E., Hakes, R. S. P., Gorham, D. J., Zhou, A., & Gollner, M. J. (2017). Review of Pathways for Building Fire Spread in the Wildland Urban Interface Part I: Exposure Conditions. *Fire Technology*, *53*(2), 429–473. https://doi.org/10.1007/s10694-016-0589-z

[8]. Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.48550/ARXIV.1603.02754

[9]. Cohen, Jack D. (2000). Preventing disaster: Home ignitability in the wildland-urban interface. *Journal of Forestry*, *98*(3), 15–21.

[10]. Ewald, F. K., Bothmann, L., Wright, M. N., Bischl, B., Casalicchio, G., & König, G. (2024). *A Guide to Feature Importance Methods for Scientific Inference*. https://doi.org/10.48550/ARXIV.2404.12862

[11]. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

[12]. Hakes, R. S. P., Caton, S. E., Gorham, D. J., & Gollner, M. J. (2017). A Review of

Pathways for Building Fire Spread in the Wildland Urban Interface Part II: Response of Components and Systems and Mitigation Strategies in the United States. *Fire Technology*, *53*(2), 475–515. https://doi.org/10.1007/s10694-016-0601-7

[13]. Hakes, R. S. P., Theodori, M., Lautenberger, C., Qian, L., & Gollner, M. J. (2022). Community-level risk assessment of structure vulnerability to WUI fire conditions in the 2017 Tubbs Fire. In D. X. Viegas & L. M. Ribeiro, *Advances in Forest Fire Research 2022* (1.ª Edição, pp. 552–557). Imprensa da Universidade de Coimbra. https://doi.org/10.14195/978-989-26-2298-9_86

[14]. Hamada, M. (1975). Architectural fire resistant themes. *Kenchikugaku Taikei , Shokokusha, Tokyo*. https://scholar.google.com/scholar_lookup

[15]. Henning, Andrew, Cox, Jonathan, & Shew, David. (2015). *CAL FIRE's Damage Inspection Program—Its Evolution and Implementation*. http://www.fltwood.com/perm/nfpa-2016/scripts/sessions/M26.html

[16]. Higuera, P. E., Cook, M. C., Balch, J. K., Stavros, E. N., Mahood, A. L., & St. Denis, L. A. (2023). Shifting social-ecological fire regimes explain increasing structure loss from Western wildfires. *PNAS Nexus*, *2*(3), pgad005. https://doi.org/10.1093/pnasnexus/pgad005

[17]. Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

[18]. Insurance Institute for Business & Home Safety. (2021). *Suburban wildfire adaptation roadmaps: A path to coexisting with wildfires.* IBHS. https://ibhs.org/wp-content/uploads/member_docs/ibhs-suburban-wildfire-adaptation-roa

dmaps.pdf

[19]. Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. *Environmental Reviews*, *28*(4), 478–505. https://doi.org/10.1139/er-2020-0019

[20]. Kearns, E. J., Saah, D., Levine, C. R., Lautenberger, C., Doherty, O. M., Porter, J. R., Amodeo, M., Rudeen, C., Woodward, K. D., Johnson, G. W., Markert, K., Shu, E., Freeman, N., Bauer, M., Lai, K., Hsieh, H., Wilson, B., McClenny, B., McMahon, A., & Chishtie, F. (2022). The Construction of Probabilistic Wildfire Risk Estimates for Individual Real Estate Parcels for the Contiguous United States. *Fire*, *5*(4), 117. https://doi.org/10.3390/fire5040117

[21]. Keeley, J. E., & Syphard, A. D. (2021). Large California wildfires: 2020 fires in historical context. *Fire Ecology*, *17*(1), 22. https://doi.org/10.1186/s42408-021-00110-7

[22]. Knapp, E. E., Valachovic, Y. S., Quarles, S. L., & Johnson, N. G. (2021). Housing arrangement and vegetation factors associated with single-family home survival in the 2018 Camp Fire, California. *Fire Ecology*, *17*(1), 25. https://doi.org/10.1186/s42408-021-00117-0

[23]. Koo, E., Pagni, P. J., Weise, D. R., & Woycheese, J. P. (2010). Firebrands and spotting ignition in large-scale fires. *International Journal of Wildland Fire*, *19*(7), 818. https://doi.org/10.1071/WF07119

[24]. Kramer, H. A., Mockrin, M. H., Alexandre, P. M., & Radeloff, V. C. (2019). High wildfire damage in interface communities in California. *International Journal of Wildland Fire*, *28*(9), 641. https://doi.org/10.1071/WF18108

[25]. Lautenberger, C. (2013). Wildland fire modeling with an Eulerian level set method and

automated calibration. *Fire Safety Journal*, *62*, 289–298. https://doi.org/10.1016/j.firesaf.2013.08.014

[26]. Lautenberger, C. (2017). Mapping areas at elevated risk of large-scale structure loss using Monte Carlo simulation and wildland fire modeling. *Fire Safety Journal*, *91*, 768–775. https://doi.org/10.1016/j.firesaf.2017.04.014

[27]. Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. https://doi.org/10.48550/ARXIV.1705.07874

[28]. Manzello, S. L., Suzuki, S., & Hayashi, Y. (2012). Exposing siding treatments, walls fitted with eaves, and glazing assemblies to firebrand showers. *Fire Safety Journal*, *50*, 25–34. https://doi.org/10.1016/j.firesaf.2012.01.006

[29]. Maranghides, A., Link, E. D., Hawks, S., McDougald, J., Quarles, S. L., Gorham, D. J., & Nazare, S. (2022). *WUI structureparcelcommunity fire hazard mitigation methodology* (NIST TN 2205; p. NIST TN 2205). National Institute of Standards and Technology (U.S.). https://doi.org/10.6028/NIST.TN.2205

[30]. Maranghides, A., & Mell, W. (2013). *Framework for Addressing the National Wildland Urban Interface Fire Problem – Determining Fire and Ember Exposure Zones using a WUI Hazard Scale*. NIST. https://doi.org/10.6028/NIST.TN.1748

[31]. Masís, S. (2023). *Interpretable Machine Learning with Python: Build explainable, fair, and robust high-performance models with hands-on, real-world examples* (Second edition). Packt.

[32]. Moritz, M. A., Batllori, E., Bradstock, R. A., Gill, A. M., Handmer, J., Hessburg, P. F., Leonard, J., McCaffrey, S., Odion, D. C., Schoennagel, T., & Syphard, A. D. (2014). Learning to coexist with wildfire. *Nature*, *515*(7525), 58–66.

https://doi.org/10.1038/nature13946

[33]. Noble, I. R., Gill, A. M., & Bary, G. A. V. (1980). McArthur's fire-danger meters expressed as equations. *Austral Ecology*, *5*(2), 201–203. https://doi.org/10.1111/j.1442-9993.1980.tb01243.x

[34]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). *Scikit-learn: Machine Learning in Python*. https://doi.org/10.48550/ARXIV.1201.0490

[35]. Purnomo, D. M. J., Qin, Y., Theodori, M., Zamanialaei, M., Lautenberger, C., Trouvé, A., & Gollner, M. (2024). Reconstructing modes of destruction in wildland–urban interface fires using a semi-physical level-set model. *Proceedings of the Combustion Institute*, *40*(1–4), 105755. https://doi.org/10.1016/j.proci.2024.105755

[36]. Quarles, S. L., Valachovic, Y., Nakamura, G. M., Nader, G. A., & De Lasaux, M. J. (2010). *Home Survival in Wildfire-Prone Areas: Building Materials and Design Considerations*. University of California, Agriculture and Natural Resources. https://doi.org/10.3733/ucanr.8393

[37]. Schoennagel, T., Balch, J. K., Brenkert-Smith, H., Dennison, P. E., Harvey, B. J., Krawchuk, M. A., Mietkiewicz, N., Morgan, P., Moritz, M. A., Rasker, R., Turner, M. G., & Whitlock, C. (2017). Adapt to more wildfire in western North American forests as climate changes. *Proceedings of the National Academy of Sciences*, *114*(18), 4582–4590. https://doi.org/10.1073/pnas.1617464114

[38]. Schug, F., Bar-Massada, A., Carlson, A. R., Cox, H., Hawbaker, T. J., Helmers, D., Hostert, P., Kaim, D., Kasraee, N. K., Martinuzzi, S., Mockrin, M. H., Pfoch, K. A., & Radeloff,

V. C. (2023). The global wildland–urban interface. *Nature*. https://doi.org/10.1038/s41586-023-06320-0

[39]. Stephens, S. L., Agee, J. K., Fulé, P. Z., North, M. P., Romme, W. H., Swetnam, T. W., & Turner, M. G. (2013). Managing Forests and Fire in Changing Climates. *Science*, *342*(6154), 41–42. https://doi.org/10.1126/science.1240294

[40]. Syphard, A. D., Brennan, T. J., & Keeley, J. E. (2014). The role of defensible space for residential structure protection during wildfires. *International Journal of Wildland Fire*, *23*(8), 1165. https://doi.org/10.1071/WF13158

[41]. Syphard, A. D., Keeley, J. E., Massada, A. B., Brennan, T. J., & Radeloff, V. C. (2012). Housing arrangement and location determine the likelihood of housing loss due to wildfire. *PloS One*, *7*(3), e33954–e33954. https://doi.org/10.1371/journal.pone.0033954

[42]. Syphard, A., & Keeley, J. (2019). Factors Associated with Structure Loss in the 2013–2018 California Wildfires. *Fire*, *2*(3), 49. https://doi.org/10.3390/fire2030049

[43]. Ting, K. M. (2011). Confusion Matrix. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 209–209). Springer US. https://doi.org/10.1007/978-0-387-30164-8_157

[44]. Troy, A., Moghaddas, J., Schmidt, D., Romsos, J. S., Sapsis, D. B., Brewer, W., & Moody, T. (2022). An analysis of factors influencing structure loss resulting from the 2018 Camp Fire. *International Journal of Wildland Fire*, *31*(6), 586–598. https://doi.org/10.1071/WF21176

[45]. Cary, G. J., M. D. Flannigan, R. E. Keane, R. A. Bradstock, I. D. Davies, J. M. Lenihan, C. Li, K. A. Logan, and R. A. Parsons. 2009. Relative importance of fuel management, ignition management and weather for area burned: evidence from five landscape fire succession models. International Journal of Wildland Fire 18: 147–156.

[46]. Quarles, S. L., Y. Valachovic, G. M. Nakamura, G. A. Nader, and M. J. De Lasaux. 2010. Home survival in wildfire-prone areas: building materials and design considerations. University of California, Agriculture and Natural Resources, ANR Publication 8393, Davis, CA.

[47]. 2014 Quadrennial Fire Review (QFR), Quadrennial Fire Review: Final Report (US Department of Agriculture Forest Service Fire and Aviation Management and Department of the Interior Office of Wildland Fire, Washington, DC, 2015).

[48]. Sugihara, N.G., T. Keeler-Wolf, and M.G. Barbour. 2018. Chapter 1. Introduction: Fire and California vegetation. In Fire in California's Ecosystems, ed. J.W. van Wagtendonk, N.G. Sugihara, S.L. Stephens, A.E. Thode, K.E. Shaffer, and J. Fites-Kaufman, 2nd ed., 1–8. Berkeley and Los Angeles: University of California Press.

Acknowledgements

We would like to express our deepest gratitude to Matt Lee, Dave Shew, Dave Sapsis, and Steve Hawks for their invaluable contributions, guidance, and expertise throughout this study. Funding for this project was provided by the California Department of Forestry and Fire Protection's Forest Health Program as part of the California Climate Investments Program, grant 8GG21815. Additional funding by the Gordon and Betty Moore Foundation (11999), and the National Science Foundation (NSF) (1854952).

Author Contributions

Maryam Zamanialaei: Conceptualization, Data curation, Data Analysis, Writing- Original draft preparation.
Daniel San Martin: Conceptualization, Data Analysis, Methodology
Maria Theodori: Data curation, Conceptualization, Methodology
Dwi Purnomo: Conceptualization, Reviewing and Editing
Ali Tohidi: Data Analysis, Visualization, Validation, Reviewing and Editing
Chris Lautenberger: Methodology, Reviewing and Editing
Yiren Qin: Methodology, Reviewing and Editing
Arnaud Trouvé:, Methodology, Reviewing and Editing
Michael Gollner: Conceptualization, Supervision, Writing- Reviewing and Editing

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DriversofFireRisktoStructuresinWUIsupplemental.docx