

FUELVISION: A multimodal data fusion and multimodel ensemble algorithm for wildfire fuels mapping[☆]



Riyaaz Uddien Shaik ^a, Mohamad Alipour ^b, Eric Rowell ^c, Bharathan Balaji ^d, Adam Watts ^e, Ertugrul Taciroglu ^a,*

^a Department of Civil & Environmental Engineering, University of California Los Angeles, Los Angeles, 90095, CA, USA

^b Department of Civil & Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, 61801, IL, USA

^c Division of Atmospheric Sciences, Desert Research Institute, Reno, 89512, NV, USA

^d Amazon, Seattle, 98109, WA, USA

^e Pacific Wildland Fire Sciences Laboratory, United States Forest Service, Wenatchee, 98801, WA, USA

ARTICLE INFO

Keywords:

Wildfires
Fuels mapping
Machine learning
Ensemble model
Synthetic data generation

ABSTRACT

Accurate assessment of fuel conditions is a prerequisite for fire ignition and behavior prediction, and risk management. The method proposed herein leverages diverse data sources – including L8 optical imagery, S1 (C-band) Synthetic Aperture Radar (SAR) imagery, PL (L-band) SAR imagery, and terrain features – to capture comprehensive information about fuel types and distributions. An ensemble model was trained to predict landscape-scale fuels – such as the 'Scott and Burgan 40' – using the as-received Forest Inventory and Analysis (FIA) field survey plot data obtained from the USDA Forest Service. However, this basic approach yielded relatively poor results due to the inadequate amount of training data. Pseudo-labeled and fully synthetic datasets were developed using generative AI approaches to address the limitations of ground truth data availability. These synthetic datasets were used for augmenting the FIA data from California to enhance the robustness and coverage of model training. The use of an ensemble of methods – including deep learning neural networks, decision trees, and gradient boosting – offered a fuel mapping accuracy of nearly 80%. Through extensive experimentation and evaluation, the effectiveness of the proposed approach was validated for regions of the 2021 Dixie and Caldor fires. Comparative analyses against high-resolution data from the National Agriculture Imagery Program (NAIP) and timber harvest maps affirmed the robustness and reliability of the proposed approach, which is capable of near-real-time fuel mapping.

1. Introduction

Recent studies indicate an unparalleled rise in the magnitude, severity, and impact of wildfire occurrences (Burke et al., 2021; Iglesias et al., 2022). In 2018, California witnessed the deadliest fire in its history, the Camp Fire, resulting in the loss of 85 lives, and the destruction of nearly 14,000 homes and over 500 commercial structures (Iglesias et al., 2022). With the exacerbation of these incidents due to climate change, the United Nations Environment Program projects a further global increase of approximately 30% by 2050 and 50% by the end of the century (Sullivan et al., 2022). Despite advances in fire science, both technologically and theoretically, wildfires persist as a significant and escalating threat to communities, infrastructure, and the environment. The unprecedented scale and complexity of this issue necessitate interdisciplinary and data-informed research on wildfire

risk management, encompassing assessment, mitigation, and response strategies.

Efficient wildfire risk management relies on accurate simulations of wildfire spread, as these simulations can significantly enhance the effectiveness of pre-event mitigation, evacuation, rescue, and fire suppression efforts (Kalabokidis et al., 2016; Sakellarou et al., 2017). An essential component of wildfire simulations is obtaining reliable estimates of the fuels that contribute to the spread of wildfires. Fuels are typically classified into three categories: ground fuels (including litter, duff, and coarse woody debris), surface fuels (such as grass, forb, shrubs, and large logs), and canopy fuels (consisting of trees and snags) (Keane et al., 2015). While surface fuels play a primary role in initiating and propagating forest fires, in this research, we are considering mapping the '40 Scott and Burgan' standard fuel models (Scott, 2005),

[☆] Note: Work is unrelated to Amazon.

* Corresponding author.

E-mail address: etacir@ucla.edu (E. Taciroglu).

which were the primary input for point-based and spread simulations and were derived from the ‘Anderson 13’ categorization standard fuel models (Anderson, 1981).

Methods for characterizing surface fuels have been developed in general, failing to capture the full range of temporal variability and spatial non-conformity inherent to the surface fuel beds (Keane et al., 2015). Consequently, the input data for modern fire behavior models contain uncertainties in describing the dynamic processes that traditional fuel inventories miss (Rowell et al., 2016). The review of the current state of surface fuel mapping research reveals that past efforts have predominantly focused on site-specific semi-manual expert systems or traditional machine learning methods (such as decision trees and random forests) at regional scales (approximately 30 km × 30 km) (Rollins, 2009; Shaik et al., 2022; Alipour et al., 2023; Shaik et al., 2023). However, these systems have limited capabilities in harnessing big data analytics, which could be leveraged to extract knowledge from spatial and spectral consistencies and ensure consistent vegetation and fuel assessment across a given landscape. Consequently, such systems experience a decrease in prediction accuracy when attempting to generalize their results to larger problem domains such as state or nation-wide.

LANDFIRE (Rollins, 2009) provides comprehensive and consistent geospatial fuel map products at the national level (CONUS), which were initially released in 2009 and remapped in 2016 by integrating remote sensing, machine learning, expert-driven rule sets, and quality control. These products are updated annually years to incorporate data on disturbances such as deforestation and fires. Although these products have provided a valuable foundation for fire spread simulation efforts, there is a need for large-scale modeling techniques that can generate georeferenced fuel mapping in near-real-time (within the same season) and without relying solely on experience-driven rule sets and localized vegetation models (Keane and Reeves, 2011). Implementing such models enhance the frequency and reduce the time delay of fuel data, which currently takes an year to update. Furthermore, new techniques could facilitate a comprehensive and systematic accuracy assessment using independent validation datasets.

Taking into account all above-mentioned limitations, this paper presents an AI-based framework that incorporates multiple modalities of data, including multi-spectral satellite imagery, C-band SAR data, L-band SAR data, and terrain data. The framework relies on a combination of ensemble and stacked machine learning models, which are trained using state-wide georeferenced labeled data. The trained ensemble model can be utilized to identify on-demand near-real-time fuels (within the same season).

1.1. Background

Fuel mapping research is underway worldwide, but the primary focus has been on developing fuel map products. Efforts to develop algorithms that can generate real-time, on-demand fuel maps have been sparse. Notable examples include the study by Pickell et al. (2020), who have developed FuelNet, which is an artificial neural network-based algorithm for updating existing fuel maps in Canada (ca. 2016). They utilized remotely sensed satellite imagery to create updated fuel maps and achieved an overall accuracy of approximately 63.1%. Shaik et al. (2022) have developed a semi-supervised algorithm using PRISMA hyperspectral imagery to map fuel types across Europe, achieving an overall accuracy of 87%. However, this algorithm does not directly map fuels; instead, it maps vegetation types and correlates them with the ‘Anderson 13’ categorization of standard fuel models (Anderson, 1981). Furthermore, its effectiveness is limited by the availability of PRISMA hyperspectral imagery.

Before undertaking the present effort, a preliminary study was carried out to develop a deep learning framework capable of processing multimodal data for large-scale surface fuel mapping (Alipour et al., 2023). A multi-layer neural network was employed to incorporate both

spectral and biophysical data, while a convolutional neural network backbone was utilized to extract visual features from high-resolution imagery. A Monte Carlo dropout mechanism was also developed to generate a stochastic ensemble of models, which effectively captured classification uncertainties while enhancing prediction performance. To demonstrate the system’s efficacy, fuel pseudo-labels were generated by randomly sampling the LANDFIRE fuel map across California as a proof-of-concept. This prior effort paved the way for the current study by demonstrating the feasibility of fuel mapping for the entire state of California and indicated that the method can produce an operationalizable fuel mapping tool by incorporating field data derived from in-situ surveys, such as the FIA database (Alipour et al., 2023).

1.2. Research significance

In this work, emerging machine learning techniques are leveraged to develop an algorithm for real-time, on-demand fuel maps for any selected domain. To that end, a data fusion scheme is devised to integrate optical, SAR, and terrain data and to identify fuels using a single end-to-end model for the state of California. To create a training dataset, fuel labels from FIA plots (Elizabeth et al., 2023) are coupled with multimodal input data sourced from various data repositories and geospatial data products, including multispectral satellite data based time-series Normalized Difference Vegetation Index (NDVI), C-band SAR data (Vertical-Vertical (VV) co-polarized and Vertical-Horizontal (VH) cross-polarized bands), L-band SAR data (Horizontal-Horizontal (HH) co-polarized and Horizontal-Vertical (HV) cross-polarized bands), SAR-based spectral indices and topography and terrain data (from the U.S. Geological Survey (USGS) Digital Elevation model). The dataset size has been improved by pseudo-labeling and synthetic data augmentation. We trained the machine learning model using the created dataset and named it as the FuelVision. The proposed approach presents the following technical contributions and benefits with respect to the existing literature: 1. Creating fuel identification models that are applicable to any selected domain with a spatial resolution of 30 m while integrating multispectral, two types of SAR and terrain information and providing a measure of model uncertainty. 2. Creating a fuel identification model that can generate on-demand near real-time fuel maps.

The methodology and results present a detailed analysis of the effect of the individual components of the FuelVision model, the multi-model ensemble approach, pseudo-labeling, and synthetic data augmentation on the training dataset. Pseudo-labels and synthetic data augmentation demonstrate proof-of-concept, and examine the feasibility of fuel identification models in California.

2. Study area and data used

This section provides an overview of the geographical scope and data sources utilized in this study. The analysis conducted in this research focuses on a specific region, referred to as the Region of Analysis. Since we obtained fuel labels from FIA plots, this section further explores the valuable insights gained from the analysis of FIA plots within the study area, allowing for a comprehensive understanding of the plots and their distribution across the region of interest. Additionally, the section delves into the utilization of remote sensing data and spectral information extracted from the remote sensing data.

2.1. Region of analysis

The region selected for data sampling and modeling in the present study is California. Fig. 1 and Table SI-1 depict the distribution of fuel labels we had for California and fuel models explanation, respectively. California offers a diverse range of elevations, ranging from -86 m in Death Valley to 4421 m at the summit of Mt. Whitney. Approximately 40% of California’s total area, which amounts to around 13.35

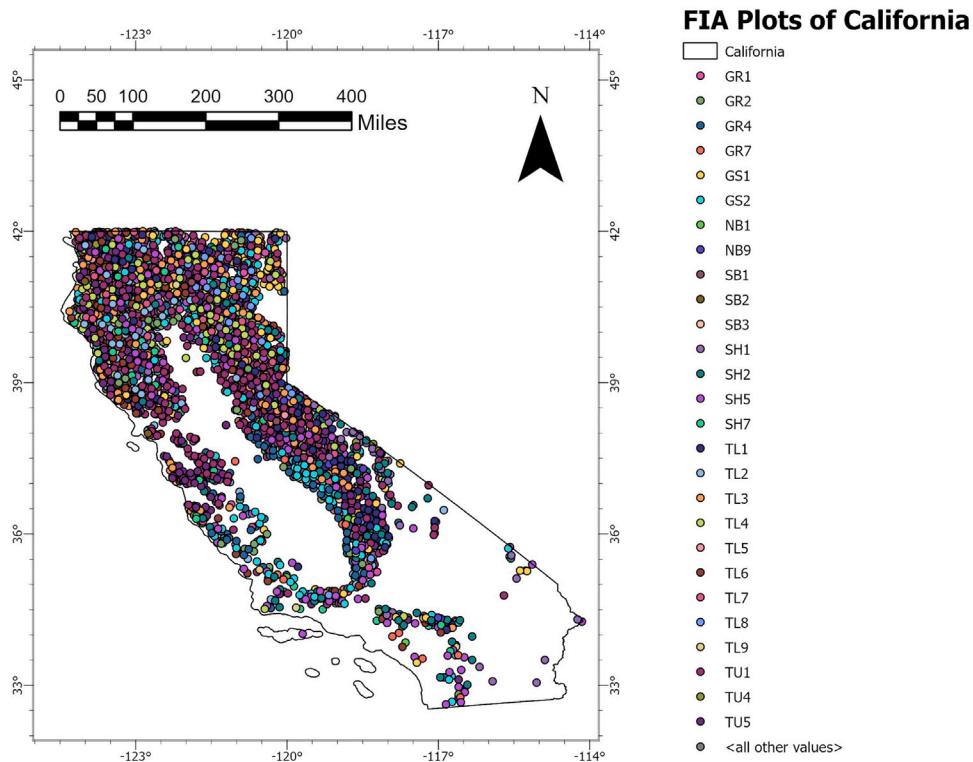


Fig. 1. FIA Plots with assigned fuel models in the study area of California (locations here are approximate to maintain FIA spatial confidentiality).

million hectares, is covered by forests (Bytnerowicz and Fenn, 1996). Moreover, the state boasts a rich assortment of ecosystems, including alpine, montane, and subalpine forests, coastal forests, mixed conifer-deciduous forests, chaparral, pinyon-juniper woodlands, and desert scrub. Due to this wide variety of forest ecotypes and the fact that California is frequently affected by wildfires, it serves as an ideal case study for this research.

2.2. FIA plots analysis

The USDA Forest Service's Forest Inventory and Analysis program conducts a comprehensive national inventory of forests in the United States. FIA stands as the sole program responsible for collecting, publishing, and analyzing data concerning forest land ownership across the nation. The spatial sampling intensity roughly equates to one plot per 6,000 acres. Each plot consists of four fixed-radius subplots, each measuring 24 ft, arranged in a clustered configuration throughout forest land. The FIA program defines forest lands as land with at least 10 percent canopy cover by live tree tally of any size, excluding tree-covered areas in agricultural production settings, such as fruit orchards, or tree-covered areas in urban settings, such as city parks. On each subplot, information about the stand and site is collected, including metrics like standing live/dead tree height/diameter and physiographic class/ownership. A subset of FIA's permanent inventory plots is sampled annually to assess indicators of forest health, including soils, understory vegetation, and down woody materials (DWM). The DWM indicator offers estimates regarding down and deceased woody materials within forest ecosystems. These DWM estimates play a crucial role in evaluating forest ecosystem attributes like fuel loadings, carbon stocks, and structural diversity. As defined by the FIA program, DWM encompasses both fine and coarse woody debris, slash piles, duff, litter, and cover and height of shrubs/herbs.

The bar graph presented in Fig. 2 displays the number of FIA plots with fuel type information collected from the years 2013 to 2019 (Elizabeth et al., 2023). The purpose of this graph is to provide an

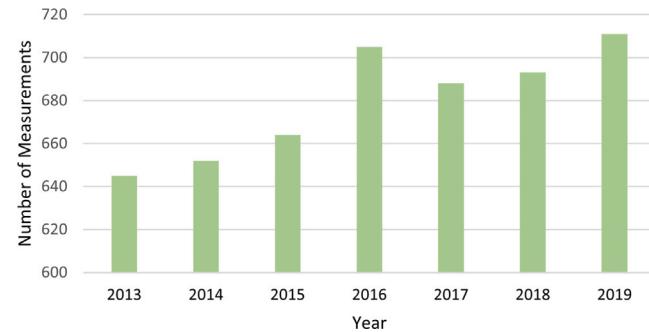


Fig. 2. Number of FIA plots with fuel data assignment per year.

overview of the temporal distribution and sampling effort of FIA plots during the specified time period. However, since the multi-modalities of satellite based observations required for this study is only available from 2015, only FIA plots from 2015 to 2019, totaling 3461, were utilized.

Fig. 3 illustrates the distribution of different fuel model types in FIA plots. Among the fuel models, the category SB3 represents the minimum count, while TU1 exhibits the maximum count. One of the reasons for this disparity could be the sole consideration of forested lands (i.e., those with at least 10 percent canopy cover by live tallied trees of any size) for inventory analysis under FIA. This disparity indicates that the ground truth data we possess is imbalanced, highlighting the uneven representation of fuel labels in the dataset.

2.3. Remote sensing data

The input variables for the machine learning model architecture are derived from a combination of four distinct open-source remote sensing datasets accessible in the Google Earth Engine (GEE). These datasets, namely Landsat –8 (L8) optical imagery, S1 SAR data, Phased

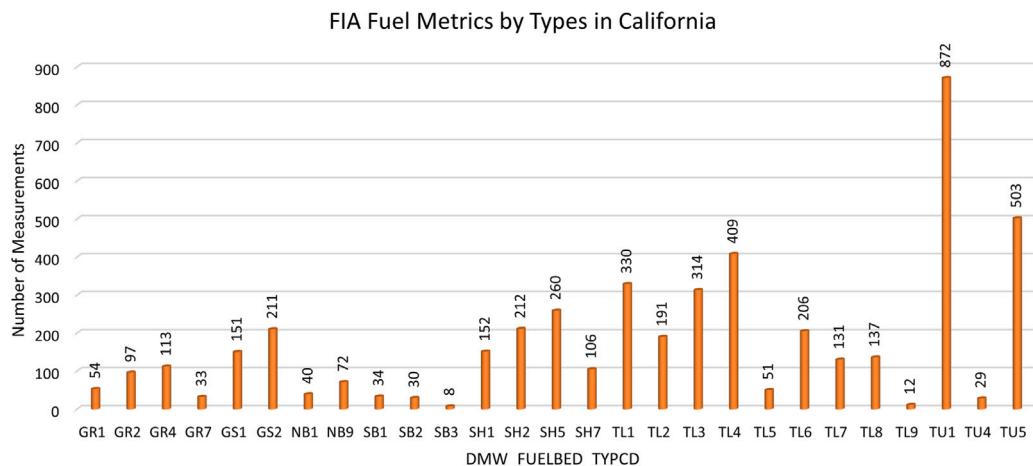


Fig. 3. Number of fuel models assignment in FIA plots for California.

Table 1
Remote sensing datasets used in this study.

Data type	Spectral data	Spatial resolution (m)
L8	Two-years time-series NDVI	30
S1	VV and VH Polarization	10
PL	HH and HV Polarization	25
SRTM	Elevation and Slope	30

Array type L-band Synthetic Aperture Radar -PALSAR (PL) data, and Shuttle Radar Topography Mission (SRTM) elevation and slope data, are merged at various spatial and temporal resolutions to generate individual arrays, each approximating a spatial resolution of 30×30 m.

The table displays various types of imagery, including L8, S1, PL, and SRTM, along with their corresponding spectral data and resolution. The worldwide utilization of these data for land use land cover classification (LULC) applications can be attributed to three key characteristics: free availability (Wang et al., 2017), interoperability (Wulder et al., 2015), and the ability to monitor expansive regions (Piedelobo et al., 2019). These factors contribute to their high demand as crucial inputs for LULC analysis on a global scale.

L8 Operational Land Imager and Thermal Infrared Sensor (OLI/TIRS) imagery provides multispectral data with a 30-meter resolution, enabling analysis across various spectral bands ranging from 442 nm to 1373 nm. With a temporal resolution of 16 days and a radiometric resolution of 16 bits, it offers valuable insights (E.D. Chaves et al., 2020). Each scene covers an area of 185×180 km and is captured from sun-synchronous orbits, making it comparable in terms of spectral, spatial, and angular characteristics to the referenced source (E.D. Chaves et al., 2020). A two-year seasonal time series of the NDVI was generated, covering eight seasons prior to the field survey date, with features named as NDVI_1, NDVI_2, up to NDVI_8. Although the main period considered is between 2015 and 2019, seasonal NDVI data from previous years were included for each respective period. This time series captures the variations in NDVI throughout different seasons over the course of two years and NDVI created using overstory reflectance will have indirect relationship with surface fuels.

The S1 SAR GRD collection available in GEE was utilized because it comprises radiometrically calibrated and terrain-corrected scenes (Mullissa et al., 2021). Initially, an image collection encompassing all accessible S1 scenes from 2015 to 2019 was gathered in the Interferometric Wide Swath mode, encompassing VV and VH polarizations at a resolution of 30 m from cross orbits (Ban et al., 2020). Subsequently, biannual mosaics were generated using medians. Six effective polarimetric features, namely SR-1 (Sayedain et al., 2020),

SR-2 (Sayedain et al., 2020), Power Ratio (PR) (Sayedain et al., 2020), Total Scattering Power (SPAN) (Sayedain et al., 2020), Difference Intensity (DI) (Sayedain et al., 2020), Radar Vegetation Index (RVI) (Periasamy, 2018) and C-band Normalized Polarized Difference Index (C-NPDI) (Afifi and Dusséaux, 2019), were identified as appropriate features for consideration as shown in Table 2. These features were selected due to their relevance and potential impact on vegetation classification.

The Japan Aerospace Exploration Agency (JAXA) provides the ALOS/PALSAR yearly mosaic at a resolution of 25 m (Chen et al., 2021). This mosaic is created by merging SAR images obtained from either PALSAR-1 or PALSAR-2, available for each year (Xu et al., 2022). This SAR imagery underwent orthorectification and slope correction using the 90 m SRTM Digital Elevation Model (DEM). To address intensity differences caused by variations in surface moisture conditions, a despeckling process was employed. This process equalized the intensities between neighboring strips, which often arise from seasonal and daily fluctuations in surface moisture (Huang et al., 2019). Initially, the data were in digital number (DN) format and were later converted to gamma-naught (γ^0) values using Eq. (1) (Huang et al., 2019) within the GEE platform.

$$\gamma^0 = 10 \times \log_{10}(DN^2) - 83 \quad (1)$$

Four effective polarimetric features based on PL, namely L-band Normalized Polarized Difference Index (L-NPDI) (Afifi and Dusséaux, 2019), Estimation of Signal Parameters via Rotational Invariance Techniques (ESPRIT) (Liu and Wang, 2021), L-band Difference (L-DIFF) (Chen et al., 2021), and C-band Ratio (C-Ratio) (Chen et al., 2021), were identified as suitable candidates for analysis, as depicted in Table 2.

SAR data has the ability to penetrate through moderate density canopy cover and provide information regardless of cloud cover or lighting conditions. The C-band frequency band is sensitive to the moisture content within vegetation. In comparison, the L-band has a longer wavelength than the C-band, allowing it to penetrate deeper into the canopy and even the forest understory. This characteristic makes the L-band particularly useful for estimating surface forest biomass. In the literature, several indices have been identified as relevant for mapping surface fuels. These include SR-2, RVI, C-NPDI, L-NPDI, ESPRIT, L_DIFF and C-Ratio. These indices have applications in assessing surface vegetation growth, vegetation optical depth, and sensitivity to vegetation height, which are indirectly related to surface fuel mapping.

The SRTM originally produced the digital elevation dataset to offer reliable, high-quality elevation data on a near-global scale with a spatial resolution of 30 m (Uuemaa et al., 2020). GEE has processed

Table 2
Spectral indices used as training features.

Index	Formula	Source	Application	Ref
NDVI	$\frac{NIR-R}{NIR+R}$	L8	Vegetation dynamics over time	Chang et al. (2019)
SR-1	$\frac{VH}{VV}$	S1	Vegetation separability	Blaes et al. (2006)
SR-2	$\frac{VV}{VH}$	S1	Surface vegetation growth	Koley and Chockalingam (2022)
PR	$\frac{ VV _{dB}^2}{ VH _{dB}^2}$	S1	land cover classification	Sayedain et al. (2020)
SPAN	$\frac{1}{2}(VV ^2 + VH ^2)$	S1	land cover classification	Sayedain et al. (2020)
DI	$\frac{1}{2}(VV ^2 - VH ^2)$	S1	land cover classification	Sayedain et al. (2020)
RVI	$\frac{4 \times VH}{VV + VH}$	S1	Vegetation and bare soil separability	Periasamy (2018)
C-NPDI	$\frac{VV-VH}{VV+VH}$	S1	Vegetation optical depth	Li et al. (2022)
L-NPDI	$\frac{VV-VH}{VV+VH}$	PL	Vegetation optical depth	Li et al. (2022)
ESPRIT	$\frac{HH+HV}{2}$	PL	Sensitive to vegetation height	Liu and Wang (2021)
L-DIFF	$HH - HV$	PL	Sensitive to vegetation height	Chen et al. (2021)
C-Ratio	$\frac{HH}{HV}$	PL	Sensitive to vegetation height	Chen et al. (2021)

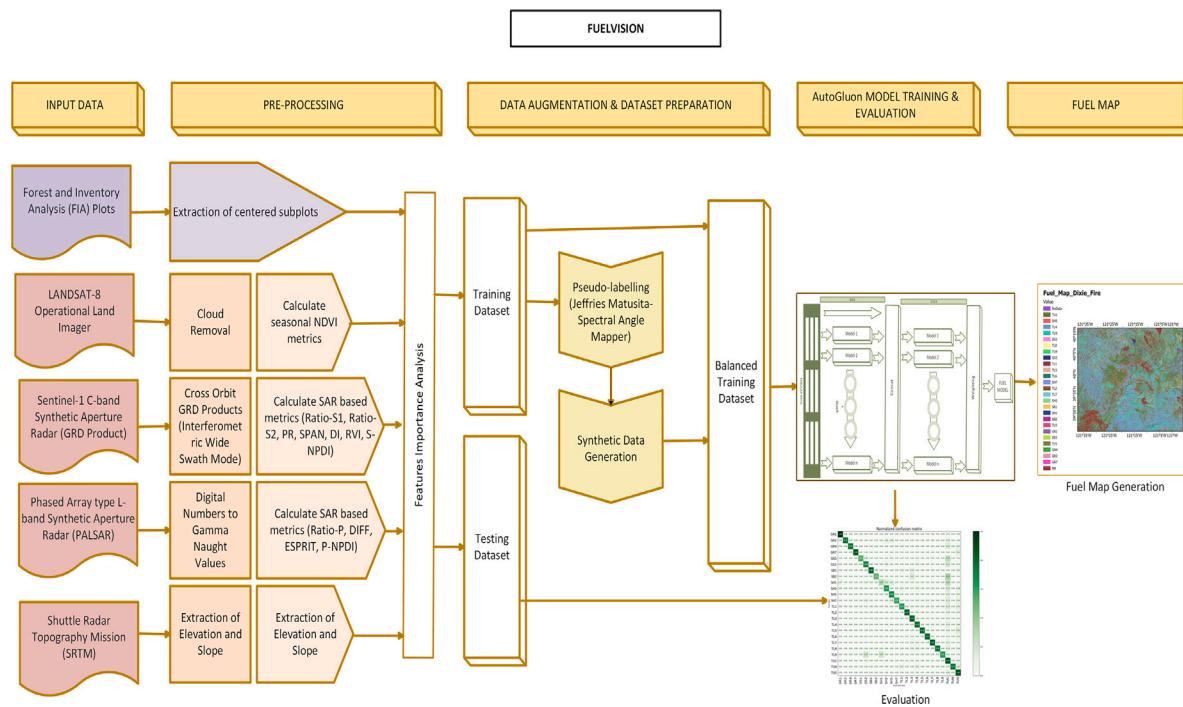


Fig. 4. FuelVision framework.

the SRTM digital elevation data to fill in data gaps and enhance its usability (Jarvis et al., 2008). By utilizing this elevation data, the slope is calculated on GEE and added as a predictor variable.

3. Methodology

In this section, Fig. 4 presents a flowchart that illustrates the step-by-step procedures of the proposed fuel mapping method. The flowchart consists of six main stages: (1) Input data is collected and filtered to ensure relevance and quality. (2) A pre-processing step cleans and standardizes the data, removing inconsistencies or noise that could affect the analysis. (3) Additionally, various indices are calculated, as shown in Table 2, and a cross-orbit data merging step is conducted for S1 data, as explained earlier. (4) Data augmentation and dataset preparation techniques expand the dataset, increasing diversity and enhancing the model's generalization. (5) An ensemble model is trained on the augmented dataset, using automated machine-learning capabilities to optimize performance. (6) The trained model

is evaluated to assess its accuracy and effectiveness in capturing data patterns. (7) Finally, the trained model generates comprehensive fuel maps, exemplified for the Dixie and Caldor Fire impact regions.

3.1. Data preprocessing & dataset preparation

For each plot in the FIA data, an array of input features was extracted. Table 2 summarizes these input features, which were selected based on a review of relevant literature across various fields. L8 (NIR & Red bands), S1(VV & VH bands), PL (HH & HV bands) and SRTM (DEM) datasets were accessed through the GEE platform and merged at varying spatial and temporal resolutions to produce individual arrays, each approximating a spatial resolution of 30×30 m. A seasonal composite of L8 data that has been corrected for atmospheric effects and geometrically rectified was calculated for each plot location. This was done using the medoid compositing criterion (Flood, 2013), which minimizes the sum of Euclidean distances in the multispectral space across all observations over the time period of interest (i.e., seasons).

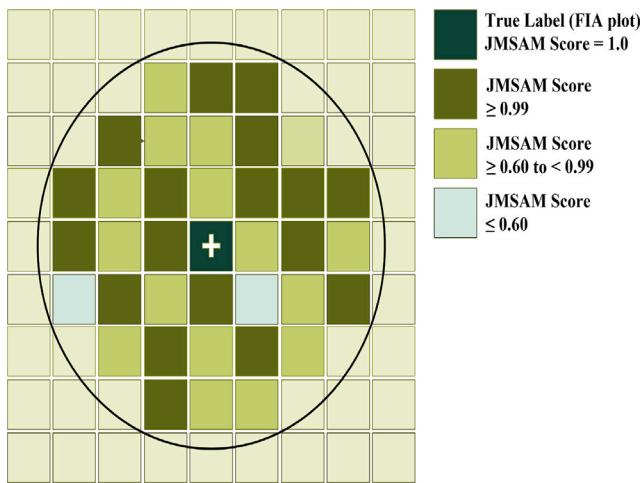


Fig. 5. Pseudo-labeling using Jeffries-Matusita-spectral angle mapping scores.

This method selects seasonal representative values while maintaining the relationships between the spectral bands, and has been demonstrated to produce radiometrically consistent composites (Van doninck and Tuomisto, 2018). Quality assessment (QA) band codes were used to mask pixels contaminated with clouds and shadows. S1 images were acquired in ascending and descending orbits around the dates of the FIA plot field surveys in GEE. The ESA S1 observation strategy uses the Interferometric Wide swath mode, which provides dual-polarization (VV and VH) imagery. Each S1 tile has high geometric accuracy, generated using the S1 Toolbox (Mahdianpari et al., 2018). The PL data was converted to gamma-naught values as described in the previous section, and it also provides dual-polarization (HH and HV) images. We prepared a dataset comprising of features extracted from remote sensing data co-located with fuel labels from FIA which is then divided into training (80%) and testing (20%) datasets. To combat spatial autocorrelation, we prevent the training and testing samples from being too closely located or spatially correlated, which could impact the generalizability of the testing outcomes. To that end, a minimum distance of 1 mile was enforced between the training and testing samples. This mechanism effectively eliminates the chance of spatially autocorrelated data points being included in both the training and testing sets, thus preventing overly optimistic performance estimates.

3.2. Label propagation

Label propagation of satellite imagery is a technique employed to assign labels to unlabeled pixels or regions within satellite images, utilizing the available labeled samples. The process involves disseminating the known labels to neighboring or similar pixels in the image, thereby expanding the labeled dataset. To carry out label propagation in this work, we analyze the spectral characteristics of both labeled and unlabeled pixels, taking into account the similarity or proximity between pixels using the Jeffries-Matusita Spectral Angle Mapper (JMSAM) technique (Shaik et al., 2022). In this approach, the deterministic spectral angle mapping method is combined with the stochastic Jeffries-Matusita distance. The tangent function is used to project the target spectrum and reference spectrum perpendicularly, which takes into account both the geometric aspects (angle and distance) as well as the band information between the spectral vectors (Padma and Sanjeevi, 2014). Ultimately, this method identifies the best match by considering the least separable distance between the spectral vectors at each band and the smallest spectral angle between the vectors yielding a score map ranging from 0 to 1. Leveraging this similarity information, the algorithm propagates labels from labeled pixels to neighboring or similar unlabeled pixels, as illustrated in Fig. 5. The propagation is

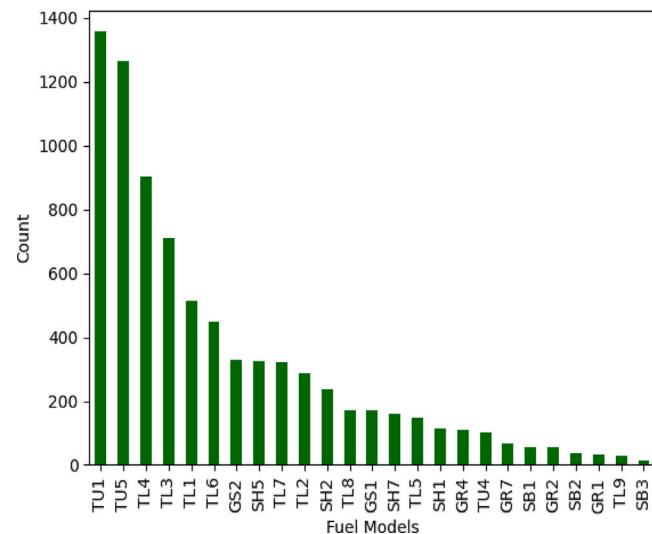


Fig. 6. Distribution of samples across different fuel models in the dataset after label propagation.

typically executed iteratively on each pixel within a 1 km radius buffer circle, gradually refining the labels of unlabeled pixels based on their neighboring pixels' labels. This iterative process was performed on all FIA plots, pseudo-labeling the pixels with a JMSAM score exceeding 0.99 and dropping same scores if obtained more than once in order to avoid spatial autocorrelation. In summary, the utilization of label propagation with satellite imagery and FIA plots increased the training dataset size by 3.5 times.

The bar graph in Fig. 6 shows the number of samples available for each fuel model in the dataset before label propagation. It can be observed from the figure that this dataset is imbalanced, which is known to cause difficulties in modeling with a significant disparity in the number of samples across different fuel models. Using an imbalanced dataset can have several implications for predictive modeling. First, it can introduce a bias towards the majority class, as the model may be more inclined to predict the dominant fuel model due to its higher representation. Second, the model's performance on the minority classes may be severely compromised, leading to lower accuracy and recall for those classes. Therefore, it is crucial to address the class imbalance issue to ensure more reliable and balanced predictions.

3.3. Synthetic data generation

We attempted to balance the dataset by augmenting it with synthetic data. Based on the existing literature (Endres et al., 2022; Pathare et al., 2023; Abufadda and Mansour, 2021), we selected five methods for comparative analysis with our dataset—namely, Tabular Variational Autoencoders (TVAE), Conditional Tabular General Adversarial Networks (CTGAN), Synthetic Minority Oversampling Technique (SMOTE), Copula General Adversarial Networks (CGAN), and Gaussian Copula Synthesizer (GCS).

TVAE — Auto-encoders are deep neural networks that learn the underlying structure of input data by passing it through a constricted or ‘bottleneck’ layer, then attempting to reconstruct the original input. This process aims to minimize the reconstruction error, effectively encoding the key features of the data (Alain and Bengio, 2014). VAE’s build on this concept, adding a regularization step that helps prevent overfitting (Wan et al., 2017). The regularization in VAEs ensures the latent space has desirable properties, enabling the generation of high-utility synthetic data.

CTGAN — Generative Adversarial Networks (GANs) (Aggarwal et al., 2021) have emerged as powerful models for synthetic data generation.

These deep learning systems consist of two key components: a generator that creates synthetic data, and a discriminator that distinguishes real from generated data. The discriminator's feedback is then used to guide and improve the generator's training (Gui et al., 2021). CTGAN is a modified GAN architecture specifically designed to generate synthetic tabular data (Xu et al., 2019). It addresses challenges like non-Gaussian and multimodal distributions by employing mode-specific normalization techniques. Additionally, CTGAN utilizes conditional generators and training-by-sampling approaches to handle imbalanced discrete columns in the dataset.

SMOTE (Chawla et al., 2002) is an oversampling technique that generates synthetic data to address imbalanced datasets. It works by taking samples from the minority classes, identifying their nearest neighbors in the feature space, and creating new synthetic samples. This is done by calculating the vector difference between the original sample and a randomly selected nearest neighbor, multiplying that difference by a random value between 0 and 1, and adding the result back to the original sample. This process helps to generalize the decision boundaries for the minority classes.

CGAN — Copula GAN is a variant of CTGAN proposed by Masarotto and Varin (2017) in 2012, utilizes Gaussian copulas to transform the data using Cumulative Distribution Functions (CDFs). This approach aims to facilitate the learning of patterns in the real data. Additionally, the Gaussian copulas model the dependence structure as multivariate normal errors and perform inference using a likelihood-based method. **GCS** — Gaussian Copulas are a popular choice for modeling the dependence structure in multiple imputation techniques used to handle missing data (Hollenbach et al., 2021). This approach can also be extended to generate synthetic data from scratch by modeling the joint distribution of the data attributes in the training set. The resulting Copula function can then be used to produce new, synthetic data attributes. The paper by Meyer et al. (2021) provides a detailed explanation of the Gaussian Copula. Readers are encouraged to refer to the cited paper for more information.

Prior research suggested that these methods had demonstrable accuracy in other application fields (Abufadda and Mansour, 2021).

We trained the above-mentioned models using the Synthetic Data Vault library (Patki et al., 2016; Lemaître et al., 2017; Montanez et al., 2018) in Python and assessed them using various evaluation metrics, including overall quality score, column shapes, column pair trends, pairwise correlation distance, and proximity level. The definitions of the former three synthetic data evaluation metrics can be found in the paper by Endres et al. (2022). These metrics evaluate the affinity of synthetic data to real data (Zhang et al., 2022) and enable a quantitative assessment of the quality of the generated/synthetic data. Values close to 0 indicate poor data quality, while values close to 1 indicate good data quality.

Among these metrics, pairwise correlation distance concept is to ascertain whether the relationships among variables in the real data are maintained in the generated synthetic data. In order to do so, Pearson correlation coefficients for real data and synthetic data were computed and stored in matrices $df_{real,corr}$ and $df_{synth,corr}$ respectively. Then the pairwise correlation distance is calculated as element-wise difference between two stored matrices as follows:

$$diff_{corr} = df_{real,corr} - df_{synth,corr} \quad (2)$$

Then, heatmaps are generated utilizing the correlation distances of real and synthetic data to visualize and understand the correlation structure. When the correlation between two items is zero, it indicates that they are equivalent to each other, and vice versa (Endres et al., 2022). This analysis was carried to assess the fidelity of the correlation structure in synthetic data generated using different techniques.

To ensure a high-quality generated dataset, the value of 'diff' should be close to zero, indicating high proximity. Conversely, low-quality datasets will have proximity values significantly different from zero. Proximity, in this context, refers to a measure of similarity or dissimilarity between data points. The proximity value mentioned in this case refers to the average of pairwise correlation distances between real data and synthetic data (Bock, 2005).

3.4. Importance of inclusive features

A feature importance study can be carried out based on permutation importance. This involves permuting the column values of a single predictor feature and then passing all test samples back through the Random Forest to recompute the accuracy. The importance of that feature is determined by the difference between the baseline accuracy and the decrease in overall accuracy caused by permuting the column. While the permutation mechanism is considerably more computationally expensive than the mean decrease in impurity mechanism, the results are more reliable. The feature importance Table 3 provides insightful information regarding the significance of various features. This table encompasses essential metrics, including importance score, standard deviation, *p*-value, p99-high, and p99-low.

The importance score of a feature signifies the decrease in model performance based on perturbed data, where the values of this specific feature have been randomly shuffled across rows (Song et al., 2022). For instance, a feature score of 0.01 indicates a predictive performance drop of 0.01 when the feature was randomly rearranged. The higher the score, the more critical the feature is for the model's performance. Conversely, a negative score implies that the feature is potentially detrimental to the final model, and removing it could enhance predictive performance. The standard deviation reflects the variability in a feature's importance across different model runs, where a low standard deviation indicates a consistent impact on predictions. The *p*-value helps determine the statistical significance of a feature's importance. A low *p*-value suggests that the feature significantly influences the target variable. For example, a *p*-value of 0.01 indicates a 1% chance of the feature being useless or harmful and a 99% chance of it being useful. Additionally, the p99-high and p99-low values specify the upper and lower bounds, respectively, for a feature's importance at the 99-th percentile. This range provides an estimate of the feature's likely impact. Overall, the feature importance table obtained is a valuable resource for comprehending the role and significance of different features within the predictive model.

Negative feature importance scores may be obtained, as we observe here for S1_VHVV, S1_VVVH, and S1_RVI. Although these scores are negative up to the fourth decimal digit, we do not dismiss them solely based on this observation and proceed to retrain the model by excluding these features. If the results remain unchanged – as they do in the present example – then, these features are kept within the model.

3.5. Ensemble model training

Our proposed machine learning architecture for fuel mapping is a heterogeneous ensemble model that combines multiple techniques, including bagging (Breiman, 1996), boosting (Freund et al., 1996), and stacking (Džeroski and Ženko, 2004). We performed bagging in combination with cross-validation, as illustrated in Fig. 7. Specifically, we implemented k-fold bagging (or cross-validation) for each of the selected models in the ensemble (Parmanto et al., 1996). This k-fold bagging technique reduces the variance of predictions by training multiple copies of the models on different randomly created subsets of the data. In our case, we randomly divided the data into five distinct subsets (ensuring that class labels were evenly distributed across the subsets) and then trained the same model five times, each time holding out a different subset, as shown in Fig. 7. The final prediction, referred to as the bagged model, was obtained by averaging the predictions from the five models. To indicate that a model underwent this procedure, we appended _BAG to its name. All the models in our study were trained using this bagging technique.

The ensemble is structured into two layers (L1 and L2), as shown in Fig. 8. We selected seven machine learning (ML) models based on a bibliometric analysis of the fuel mapping literature (Shaik et al., 2023), (Shaik et al., 2025). These models include Neural Network (McElfresh et al., 2024), Light Gradient Boosting Machine (LightGBM) (Ke et al.,

Table 3

Importance score of inclusive features in terms of overall accuracy.

No.	Variance	Importance	StdDev	p-value	p99_high	p99_low
1	Elevation	0.190246	0.012607	2.30072e-06	0.216205	0.164288
2	NDVI_3	0.111513	0.003718	1.48138e-07	0.119170	0.103857
3	NDVI_1	0.087381	0.001928	2.84011e-08	0.091350	0.083412
4	NDVI_4	0.080040	0.006472	5.08484e-06	0.093366	0.066715
5	NDVI_8	0.079638	0.005422	2.56236e-06	0.090802	0.068474
6	NDVI_5	0.074912	0.003695	7.07647e-07	0.082519	0.067305
7	NDVI_2	0.065058	0.003512	1.01531e-06	0.072289	0.057826
8	NDVI_6	0.060935	0.005431	7.49421e-06	0.072118	0.049752
9	NDVI_7	0.058019	0.002835	6.82073e-07	0.063857	0.052181
10	Slope	0.022624	0.005710	4.48223e-04	0.034382	0.010867
11	PL_ESPRIT	0.022524	0.001301	1.33134e-06	0.025203	0.019844
12	PL_HV	0.014882	0.002675	1.19981e-04	0.020389	0.009375
13	S1_VH	0.012971	0.003544	6.07255e-04	0.020269	0.005673
14	PL_HH	0.011262	0.002091	1.36325e-04	0.015568	0.006956
15	S1_SPAN	0.004525	0.001629	1.71010e-03	0.007879	0.001170
16	S1_VV	0.004324	0.002091	4.92799e-03	0.008630	0.000018
17	PL_NPDI	0.003419	0.003419	9.24982e-03	0.007521	-0.000683
18	PL_DIFF	0.002614	0.002115	2.53267e-02	0.006970	-0.001741
19	PL_Ratio	0.002614	0.002145	2.63424e-02	0.007031	-0.001802
20	S1_PRatio	0.001508	0.000941	1.15252e-02	0.003445	-0.000428
21	S1_D1	0.000101	0.001440	4.41731e-01	0.003065	-0.002864
22	S1_VHVV	-0.000402	0.001252	7.43887e-01	0.002175	-0.002980
23	S1_VVHV	-0.000804	0.001690	8.26413e-01	0.002675	-0.004284
24	S1_RVI	-0.001106	0.001440	9.19529e-01	0.001858	-0.004070

2017), XGBoost (Chen and Guestrin, 2016), CatBoost (Prokhorenkova et al., 2018), Random Forest (Pal, 2005), Extra Trees Classifier (Geurts et al., 2006), and k-Nearest Neighbors (Zhang, 2016), to include in the ensemble. All the models in L1 were trained using the bagging approach, leveraging the features of the training dataset. The hyperparameters of the models trained in L1 are presented in the Table 4. In total, 13 models were trained in L1. This includes the seven aforementioned models, as well as additional variations: neural networks implemented in PyTorch and FastAI (differing in their implementation), Random Forest trained with two different criteria (Gini and Entropy), LightGBM configured as an Extra Trees model, LightGBM with a learning rate of 0.005, Extra Trees Classifier with Gini and Entropy criteria, and k-Nearest Neighbors with two different weighting schemes (uniform and distance). The total of 13 ML models trained in L1 which are considered as base models.

In the first layer, all the base models are trained independently, and their predictions are concatenated with the original training dataset features to create a new dataset with an extended set of features. In the second layer (L2), instead of using a simpler model as the stacker, we reused all the base models as stacker models—except for k-nearest neighbors, which yielded low accuracy in L1. These models retrained with the same hyperparameter values and on the new dataset, which includes the predictions from the base models as additional features. In total, 11 models were retrained in L2.

Finally, the stacking layer applies ensemble selection (Caruana et al., 2004) to aggregate the stacker models' predictions in a weighted manner. The ensemble selection method is applied after every layer, although the weighted ensemble created after the second stacking layer is considered final. Ensemble selection is recommended for combining base models as it is resistant to overfitting. To further enhance the stacking performance, we implemented the ensemble model using the Python-based Gluon library doi:https://auto.gluon.ai/stable/api/autogluon.tabular.TabularPredictor.feature_importance.html, (Erickson et al., 2020). This library supports multi-layer stacking strategies and can automatically tune model hyperparameters, often outperforming manual adjustments doi:<https://auto.gluon.ai/stable/index.html>, doi:https://auto.gluon.ai/stable/api/autogluon.tabular.TabularPredictor.feature_importance.html. It has gained widespread use across various applications (Gao et al., 2024; Qi et al., 2021; An et al., 2024) due to its ability to fine-tune hyperparameters across the full range of each machine learning model without human intervention. We trained our ensemble model on an NVIDIA GeForce RTX 3070 GPU for 13,828 s

(approximately 4 h), utilizing aforementioned base models across the two stacking layers (see Table 5).

Table 6 presents evaluation results of the FuelVision model. Models appended with _BAG are bagged models, _L1 indicates models trained in stacking layer 1, and _L2 indicates models trained in stacking layer 2. After each layer, a weighted ensemble selection method is applied for evaluation. The ensemble model after layer 1, *Weighted Ensemble L2*, achieved a testing accuracy of 0.74 and a validation accuracy of 0.73. The final ensemble model, *Weighted Ensemble L3*, achieved the highest validation accuracy of 0.77. This approach not only aids in minimizing overfitting but also enhances the overall system performance. The table reveals that the maximum difference between the testing and validation scores does not exceed 1.5%, indicating minimal overfitting. By leveraging the strengths of multiple models, this architecture effectively generalizes and accurately predicts outcomes on both the validation and testing datasets.

3.6. Non-burnables mapping: Post-processing step

The FIA field survey data only focuses on forest lands and does not include information on non-burnable land cover types. As a post-processing step, we map and label four types of non-burnable fuel models (NB1 (urban/developed), NB2 (snow/ice), NB8 (open water), and NB9 (barren ground)) as NB. To map barren lands, snow, and water in the region of interest, we create NDVI and Normalized Difference Water Index (NDWI) (Wang et al., 2020) images using the GEE. Additionally, we generate a built-up index (BUI) (Rendana et al., 2023), which is the difference between NDVI and Normalized Difference Build-up Index (NDBI), to map urban components. Utilizing the NDVI, NDWI, and BUI, we perform a post-processing step on the fuel map generated by the FuelVision model. Pixels with $NDVI < 0$, $NDWI > 0.5$, and $BUI > 0.5$ are labeled as NB. We then decide on these values by conducting experiments with a simple trial-and-error approach and visually checking the map. The equations for calculating NDWI and NDBI are provided in Eqs. (3) and (4).

$$NDWI = \frac{G - NIR}{G + NIR} \quad (3)$$

$$NDBI = \frac{MIR - NIR}{MIR + NIR} \quad (4)$$

$$BUI = NDVI - NDBI \quad (5)$$

where G = Green, NIR = Near-infrared and MIR = Mid-infrared

Table 4
Machine learning models and hyperparameters.

No	ML model	Hyperparameters
1	Neural Networks	hidden layer sizes=100, activation=relu, solver=adam, alpha=0.0001, batch size=200, learning rate init=0.001, power t=0.5, max iter=200, shuffle=True, tol=0.0001, beta 1=0.9, beta 2=0.999, epsilon=1e-08, n iter no change=10, max fun=15000
2	LightGBM	num boost round=100, early stopping rounds=10, metric=multi logloss, learning rate=0.1, num leaves=64, feature fraction=0.9, bagging fraction=0.9, max depth=6, min data in leaf=3, boosting=gbdt, tree learner=serial, feature fraction bynode=1, max bin=255
3	XGBoost	alpha=0, base score=0.5, booster=gbtree, colsample bylevel=1, colsample bynode=1, colsample bytree=1, eta=0.3, eval metric=error, gamma=0, grow policy=depthwise, lambda=1, lambda bias=0, max bin=256, max delta step=0, max depth=6, min child weight=1, monotone constraints=(0,0), normalize type=tree, objective=multi:softmax, rate drop=0, refresh leaf=1, sample type=uniform, sketch eps=0.03, tweedie variance power=1.5
4	CatBoost	iterations=500, early stopping rounds=5, eval metric=MultiClass, learning rate=0.009, depth=6, l2 leaf reg=3, random strength=3, max leaves=31, rsm=1, sampling frequency=PerTreeLevel, min data in leaf=1, bagging temperature=1, boosting type=Ordered, grow policy=SymmetricTree, random seed=1.0
5	Random Forests	n estimators=100, criterion=gini & entropy, min samples split=2, min samples leaf=1, max features=sqrt, bootstrap=True
6	Extra Trees Classifier	n estimators=100, criterion=gini & entropy, min samples split=2, min samples leaf=1, max features=sqrt
7	k-Nearest Neighbors	n neighbors=5, weights=uniform & distance, algorithm=auto, leaf size=30, p=2, metric=minkowski

Table 5
Validation and accuracy testing of the FuelVision model.

No.	Model	Test Acc	Validation Acc
1	NeuralNetFastAI_BAG_L2	0.772750	0.761689
2	LightGBMXT_BAG_L2	0.769734	0.775989
3	WeightedEnsemble_L3	0.769735	0.776899
4	LightGBM_BAG_L2	0.767722	0.774007
5	XGBoost_BAG_L2	0.763700	0.773002
6	CatBoost_BAG_L2	0.762695	0.769985
7	LightGBMLarge_BAG_L2	0.761187	0.769356
8	RandomForestGini_BAG_L2	0.753142	0.756662
9	ExtraTreesGini_BAG_L2	0.753142	0.746481
10	ExtraTreesEntr_BAG_L2	0.749120	0.738185
11	WeightedEnsemble_L2	0.748115	0.749246
12	RandomForestEntr_BAG_L2	0.745601	0.748366
13	NeuralNetTorch_BAG_L2	0.730015	0.710659
14	LightGBMLarge_BAG_L1	0.725993	0.717320
15	CatBoost_BAG_L1	0.720463	0.720714
16	LightGBMXT_BAG_L1	0.720463	0.713298
17	LightGBM_BAG_L1	0.715938	0.710533
18	RandomForestGini_BAG_L1	0.714932	0.699723
19	ExtraTreesGini_BAG_L1	0.713424	0.716817
20	ExtraTreesEntr_BAG_L1	0.712921	0.713801
21	NeuralNetTorch_BAG_L1	0.707391	0.691554
22	XGBoost_BAG_L1	0.703871	0.689542
23	RandomForestEntr_BAG_L1	0.701357	0.704500
24	NeuralNetFastAI_BAG_L1	0.689794	0.649070
25	KNeighborsDist_BAG_L1	0.359980	0.340372
26	KNeighborsUnif_BAG_L1	0.317245	0.298140

4. Results

In this section, we carry out a comprehensive evaluation of the data augmentation techniques employed in our study, followed by a detailed model evaluation. We analyze the impact of pseudo-labeling and data augmentation techniques on the overall performance of our fire fuel mapping models. Additionally, fuel maps for select case studies are presented, showcasing the effectiveness of FuelVision in accurately classifying fuel types. Furthermore, we delve into the analysis of prediction uncertainties, leveraging prediction probabilities to assess the

reliability and confidence of our model's predictions. The following subsections offer insights and discussions on each of these aspects and examine the robustness and effectiveness of the proposed methodology.

4.1. Evaluation of data augmentation techniques

We assessed all implemented data augmentation techniques using SD metrics and testing dataset. The obtained values from the SD metrics are displayed in Table 6, while the pairwise correlation distance heatmaps are presented in Fig. 9. We examined the metric values and

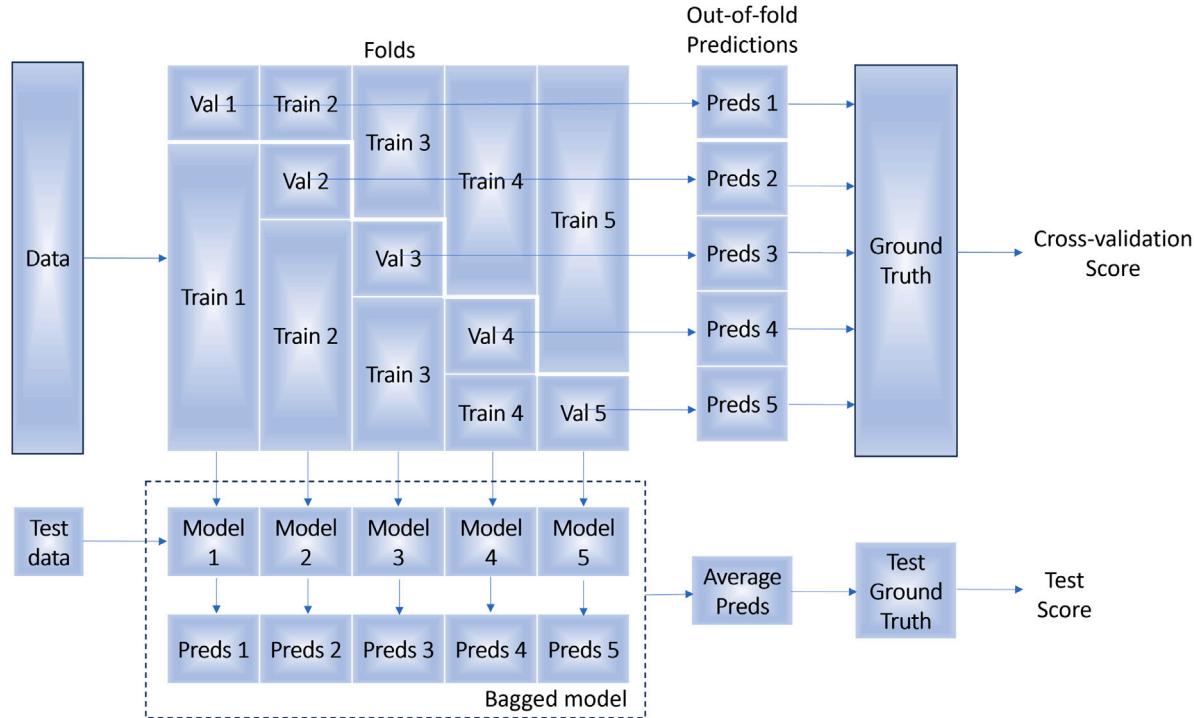


Fig. 7. The concept of bagging applied in our framework.

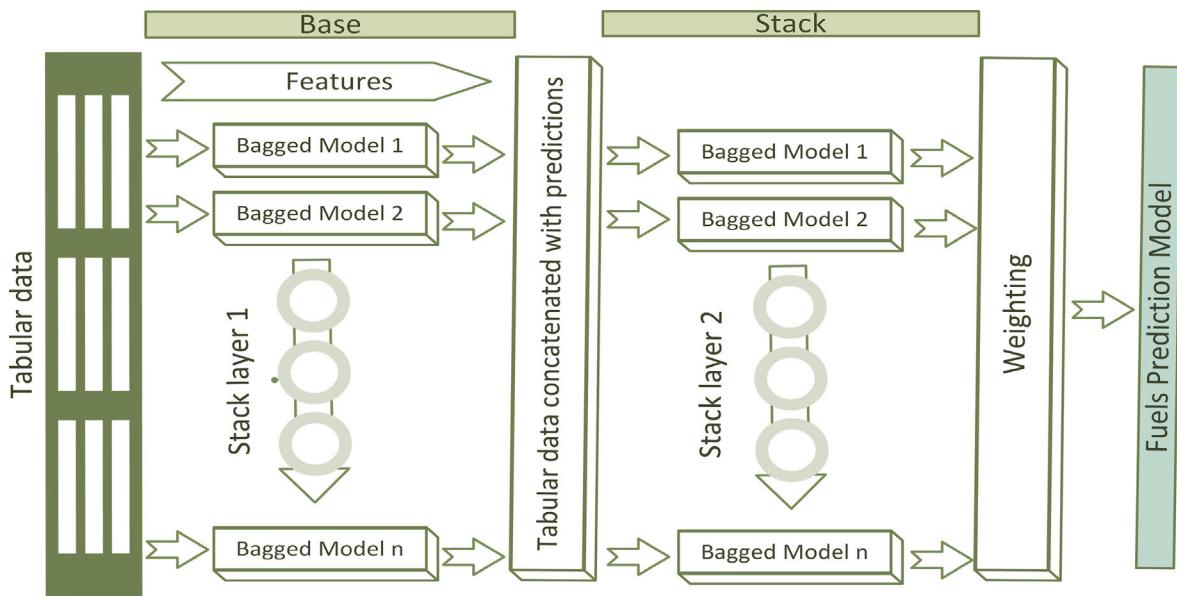


Fig. 8. Ensemble model with multi-layer stacking strategy, shown here using two stacking layers and multiple base learners.

compared them with non-SD evaluation metrics. The pairwise correlation distance heatmaps offer a visual representation of the relationships among features in the dataset, providing insights into both similarities and differences within the correlation patterns.

Upon analyzing Table 6, it becomes evident that TVAE outperforms other models in terms of SD metrics, while GCS excels at proximity. However, upon visualizing the heatmaps of these two models in Fig. 9, we discover that GCS is missing a feature, namely PL_HV, which may be attributed to the data complexity—specifically, the non-Gaussian distribution of PL_HV data.

Additionally, TVAE's heatmap appears less smooth in comparison. Consequently, we proceed to evaluate the CTGAN model, which yields the highest scores for both SD metrics (overall quality score of 91.39, column shapes value of 89.55, and column pair trends value of 93.23) and non-SD metrics (proximity level of -0.023). Furthermore, the heatmap generated by CTGAN exhibits noticeable improvements. As a result, CTGAN is selected as the optimal model for generating synthetic data to augment our training dataset. This approach allows us to increase the training dataset by 10times to a total of 26,000 samples.

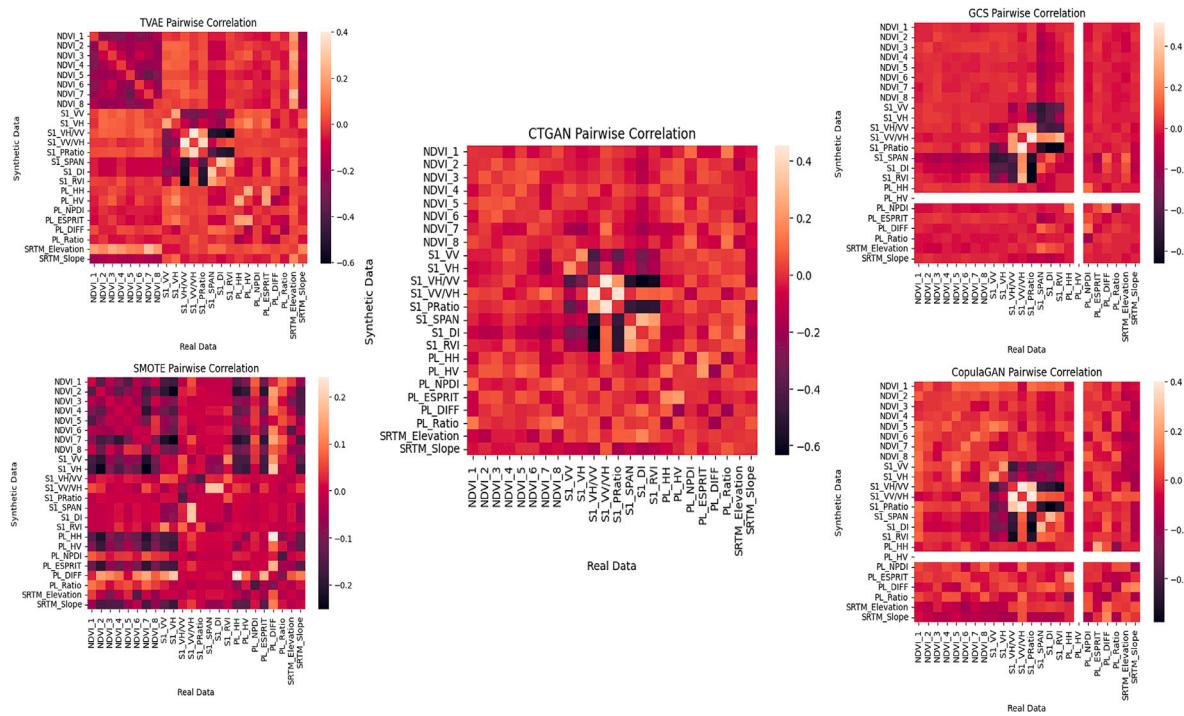


Fig. 9. Pairwise Correlation distance for real and synthetic data.

Table 6
Comparative analysis of Synthetic models.

No.	Models	Overall quality score	Column shapes	Column pair trends	Proximity level
1	TVAE	93.88	94.62	93.14	-0.039
2	CTGAN	91.39	89.55	93.23	-0.023
3	SMOTE	90.81	88.75	92.88	-0.026
4	CGAN	89.28	86.72	91.85	-0.033
5	GCS	89.22	84.43	94.02	-0.021

4.2. Impact of pseudo-labeling & data augmentation on accuracy

The application of pseudo-labeling using the JM-SAM method and augmenting synthetic data generating using the CTGAN method enhances the accuracy of our FuelVision model, as shown in figures SI-1, SI-2 and SI-3.

Figure SI-1 depicts the evaluation results of the ensemble model trained with raw data, and it can be seen that there is no diagonal; instead, most of the fuel models are classified as TU1 as it has the higher number of plots, and can be correlated with FIA plots. However, after implementing pseudo-labeling and increasing the size of the training dataset, the diagonal accuracies experience a substantial boost (20%) as evidenced in figure SI-2.

This increase is further amplified (20%) by incorporating data augmentation techniques, resulting in an even higher overall accuracy of 0.77 with strong diagonal accuracies. The confusion matrices shown in figure SI-3 indicate remarkable improvements in the classification accuracy of each fuel model. The initial training dataset exhibits a significant class imbalance, with TU1 being the majority class. Consequently, the trained model displays a bias towards classifying instances as TU1, resulting in an inflated number of misclassifications. However, after incorporating the pseudo-labels and augmenting the dataset with synthetic data, we observe a substantial reduction in this bias. The model's ability to generalize improves remarkably, as evidenced by a more balanced distribution of classifications across different classes. These advancements highlight the contributions of pseudo-labeling and augmentation with synthetic data, underscoring their effectiveness in refining our fuel mapping model.

4.3. Model evaluation

The FuelVision model was evaluated with a testing dataset of 1,989 samples with the performance metrics including precision, recall, and F1-score, as shown in Table 7. Therefore, these scores take false positives and false negatives into account together. F1-score is usually more beneficial than accuracy, especially when there is an uneven class distribution (Shaik et al., 2022). Accuracy works best when false positives and false negatives have similar costs. The recall score, also known as sensitivity or true positive rate, measures the model's ability to correctly identify positive instances. In cases with different false positives and false negatives, it is better to consider precision and recall. The "Support" column in the table depicts the number of samples in the testing dataset. Due to insufficient ground truth data, an equal number of samples could not be considered. These performance metrics, along with the confusion matrix in figure SI-3, allowed for a comprehensive evaluation of FuelVision's accuracy, precision, recall, and overall effectiveness in predicting fuel models.

The f1-score for most of the fuel models is around 0.70, except for GS1 and SH1. Upon examining the confusion matrix in figure SI-3, we see that GS1 and SH1 were misclassified as TU1. According to the definition, TU1 represents a fuel bed with a low load of grass and/or shrubs with litter, while GS1 and SH1 refer to shrubs that are 1 foot high, with the former having a low load of grass. There is a possibility that the low load of litter is insensitive to reflectance profiles or may appear as mixed pixels with a majority of shrubs and grasses. This could be the reason for misclassifying the GS1 and SH1 fuel models as TU1. Considering the presented performance metrics and confusion matrix, an overall accuracy of 0.77 is observed.

4.4. Fuel map generation for case studies

The severity of wildfires in the Western United States has been steadily increasing since the mid-1980s, posing greater risks to human lives, properties, carbon storage, biodiversity, and other vital ecosystem services. In the past decade, the expansion of wildfire incidents has accelerated, culminating in an unprecedented fire season

Table 7
Performance metrics on test dataset.

No.	Fuel model	Precision	Recall	f1-Score	Support
1	GR1	0.50	1.0	0.67	4
2	GR2	0.90	0.75	0.82	12
3	GR4	1.0	0.75	0.86	28
4	GR7	0.78	0.88	0.82	8
5	GS1	0.96	0.52	0.68	44
6	GS2	0.79	0.68	0.73	76
7	SB1	1.0	0.80	0.89	20
8	SB2	1.0	0.50	0.67	6
9	SH1	0.62	0.30	0.40	27
10	SH2	0.75	0.63	0.68	65
11	SH5	0.85	0.68	0.75	68
12	SH7	0.90	0.58	0.71	31
13	TL1	0.84	0.61	0.71	141
14	TL2	0.92	0.79	0.85	92
15	TL3	0.80	0.82	0.81	173
16	TL4	0.78	0.78	0.78	215
17	TL5	0.90	0.79	0.84	33
18	TL6	0.93	0.81	0.86	113
19	TL7	0.77	0.82	0.80	80
20	TL8	0.83	0.76	0.79	50
21	TL9	0.60	0.60	0.60	5
22	TU1	0.65	0.85	0.74	358
23	TU4	1.00	0.73	0.85	15
24	TU5	0.73	0.86	0.79	325
	accuracy			0.77	1989
	macro avg	0.82	0.72	0.75	1989
	weighted avg	0.79	0.77	0.77	1989

in 2020. This particular year witnessed over 2.5 million acres burning in the Western United States, with California accounting for 38% of that total (McWethy et al., 2019). Since 2018, California has endured a series of fire seasons that broke previous records in burned area and losses. More than 27,000 homes and commercial structures have been destroyed, and the costs associated with fire suppression have soared (Palinkas and Palinkas, 2020). In this study, we considered two representative California wildfire events for the development of fuel maps, which will be described below and will be incorporated into wildfire simulations in the near future.

The *Dixie fire*, which began on July 13, 2021, holds the record as the largest documented wildfire to date, scorching an area of 374,000 hectares. The cost of suppressing this particular wildfire surpassed the \$500 million mark for the first time, reaching \$637 million in 2021 (Taylor et al., 2022). This fire's rapid spread and exceptionally intense behavior have piqued the interest of numerous fire scientists, prompting numerous studies on fire simulation. Fuel mapping plays a vital role in understanding and effectively managing wildfire risks. By accurately evaluating fuel conditions, fire behavior can be better predicted, strategies for fire suppression can be optimized, and overall fire management endeavors can be improved. The region where the Dixie fire occurred is considered a significant case for the development of fuel maps.

The second test case, known as the *Caldor fire*, continued on for 67 days, burning through approximately 221,835 acres of land (including 9,885 acres within the Lake Tahoe Basin), destroying 1,003 structures and necessitating the evacuation of over 50,000 residents (Sion et al., 2023; East et al., 2022). The confluence of regional drought, intense heat, and powerful winds resulted in highly active fire behavior, giving rise to additional wildfires known as spot fires. These spot fires exacerbated the fire's expansion and posed additional challenges for firefighters in their containment efforts (Wadhwan et al., 2022).

The process employed to generate these fuel maps for the Dixie and Caldor fires' regions of interest utilizing FuelVision is outlined as follows:

1. The input data for the region of interest is collected as outlined in Table 1 using Google Earth Engine.
2. The remote sensing data is pre-processed, and spectral indices, as specified in Table 2, are calculated using Python.

3. The satellite imagery and spectral indices are converted into a tabular format and saved as a CSV file, serving as the training dataset.
4. Employing the generated CSV file and FuelVision, fuel models are predicted for each pixel parallelly.
5. A post-processing algorithm is applied to map the non-burnables as described in §3.5 and transforms the CSV file into a GTiff file.

The proposed FuelVision model successfully maps all the fuel models in the Dixie and Caldor fire areas, which are presented in Figs. 11(a) and 12(a), respectively. Additionally, Figs. 11(b) and 12(b) display the high-resolution imagery from NAIP for visual comparison. These high-resolution images allow us to compare the predicted fuel maps, specifically in terms of vegetation cover and vegetation patterns. These case studies can serve as a reference guide for creating fuel maps for any region of interest in California.

4.5. Model uncertainty analysis

Model uncertainty plays a crucial role in evaluating the reliability and confidence of predictions made by machine learning models. Analyzing the prediction probabilities provided by the model allows us to assess the level of uncertainty associated with each prediction. Our study generated prediction probabilities for the Dixie fire of 2021, as illustrated in Fig. 10. These probabilities range from 0.01 to 0.99. They represent the model's estimation of the likelihood of a given instance belonging to a particular class. Higher prediction probabilities indicate a greater level of certainty, suggesting that the model has more confidence in its predictions. Conversely, lower prediction probabilities imply a higher degree of uncertainty, indicating that the model may be less confident or more conflicted in its classification decision.

In our case, we start the analyses by comparing non-burnables (NBs) since the model was not trained on NBs, as mentioned in the previous section, but they were mapped in the post-processing steps. Consequently, the model predicted these areas with lower probabilities, which shows the potential value of the uncertainty analysis, which may reveal hidden issues in the mapping process. Referring to the NAIP image of 2018 in Fig. 12, we observe that the southwest region of interest contains non-burnables such as rocks, barren lands, and roadways passing through the center. We notice a similar pattern in the probability map, where these areas showed lower probabilities since they were mapped as fuels in the fuel map.

Furthermore, we compared the vegetated parts, which are primarily located on the south and west sides. Correspondingly, the probabilities in these regions are high (> 0.9). Towards the northeast, we observe slightly lower probabilities ranging from 0.7 to 0.9. Further analysis reveals that these areas roughly corresponded to the timber harvest locations indicated on the timber harvest map of 2022, as shown in Fig. 12 and presence of timber changes with harvesting period. It should be noted that the timber harvest and NAIP images were used here only to analyze the potential reasons behind the observed uncertainty and were not used in the mapping process. This shows that the analysis of uncertainty enabled quantified interpretations of the varying degrees of confidence in downstream predictions, accounting for post-processing mapping effects, and revealing correlations between probability patterns and landscape features.

5. Discussion

We developed the FuelVision model using data from FIA plots and multiple sources of satellite data including optical, SAR and topographical data, employing a set of machine learning models in an ensemble approach. One of the challenges in creating this framework was unbalanced data, which was overcome by generating synthetic data using CTGAN. Our approach of directly mapping surface fuels yielded an overall accuracy of 0.77. In recent years, researchers have been integrating field survey data, remote sensing data, and AI to produce state-of-the-art fuel maps. Carbone et al. (2023) proposed

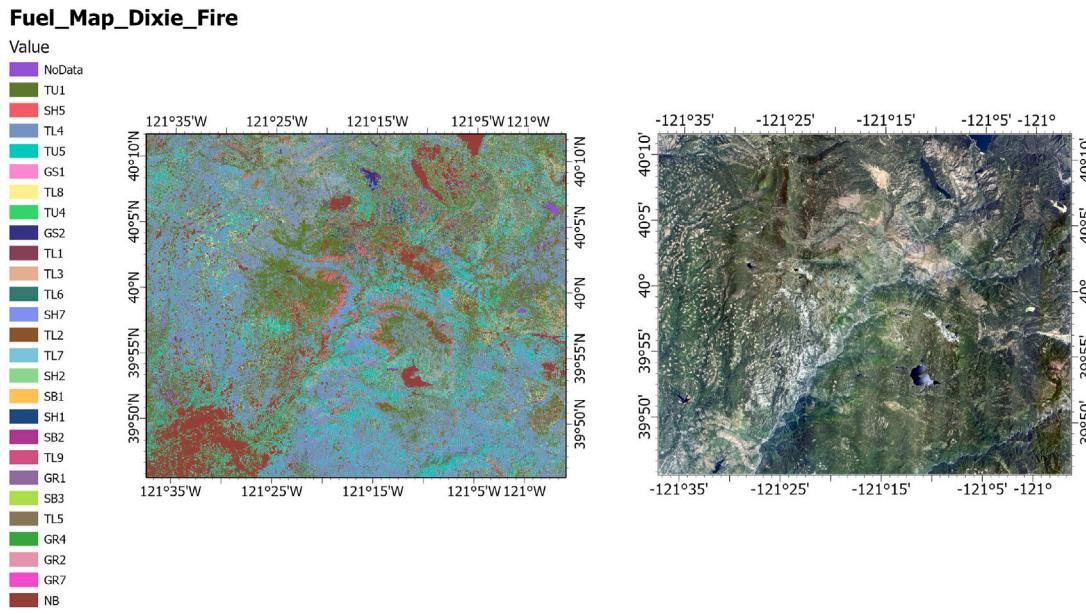


Fig. 10. (a) Fuel Map and (b) NAIP Image for the region of Dixie Fire.

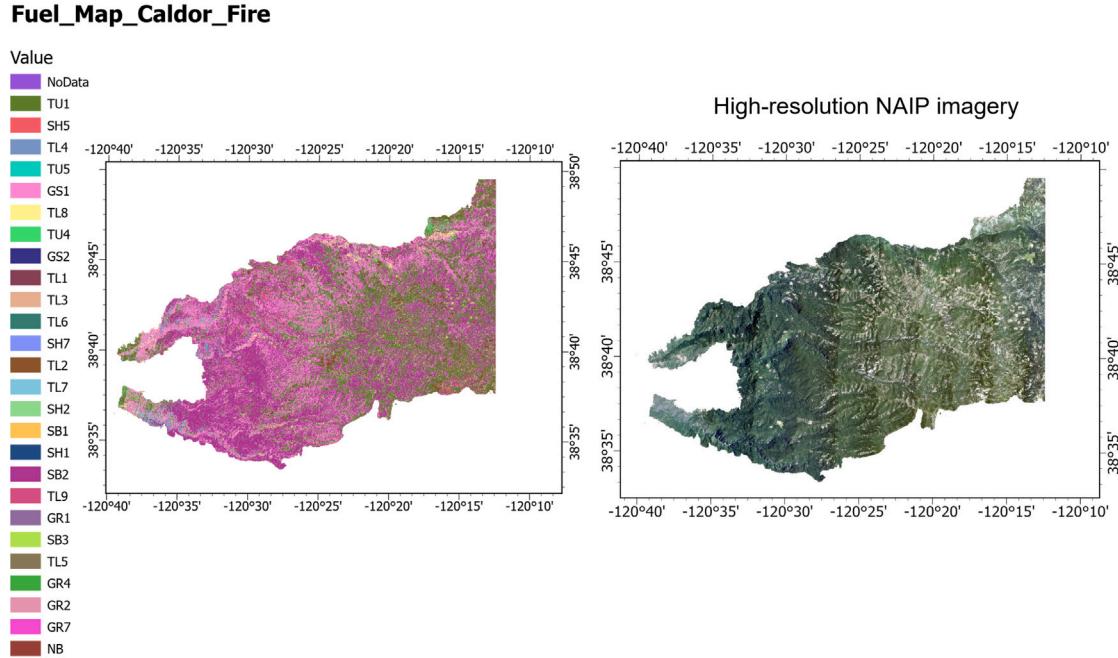


Fig. 11. (a) Fuel Map and (b) NAIP Image for the region of Caldor Fire.

a fuel mapping approach using a convolutional neural network and remote sensing data, in which they mapped broadleaf, grass, shrub, and conifer vegetation, and then used an above-ground biomass map to further categorize fuel models. However, their work was limited by its dependence on the above-ground biomass maps which may not be readily available to everyone. Another study developed an automatic algorithm with PRISMA hyperspectral data and an ML model, which used the CORINE land cover (CLC) map provided by the EU Copernicus program as input (Shaik et al., 2022). Their study concluded that their algorithm could map vegetation types across Europe using the CLC input, with an overall accuracy of 80%, but the mapped vegetation needed to be subsequently categorized into fuel models based on their characteristics, requiring human intervention (Shaik et al., 2022). In

the most recent effort, Alipour et al. (2023), Shaik et al. (2025) developed a regional-scale, AI-based fuel mapping algorithm using fuel pseudo-labels from existing fuel maps. Their study produced a fuel map for California with an accuracy of around 70%. However, the authors mentioned that in future work, the method could be retrained using FIA plot data instead of pseudo-labels. Many other studies on fuel mapping have been reported in the literature, but they were limited to specific regions or classifications of selected fuel types (Solares-Canal et al., 2023; Aragoneses et al., 2023) or classification on only vegetation types (Sismanis et al., 2023), or they used non-spaceborne data (Labenski et al., 2022) that is not globally available. To the best of the authors' knowledge, the presented approach is the first to directly map surface fuels at the regional scale by fusing multi-modal remote sensing data and large-scale in situ measurements.

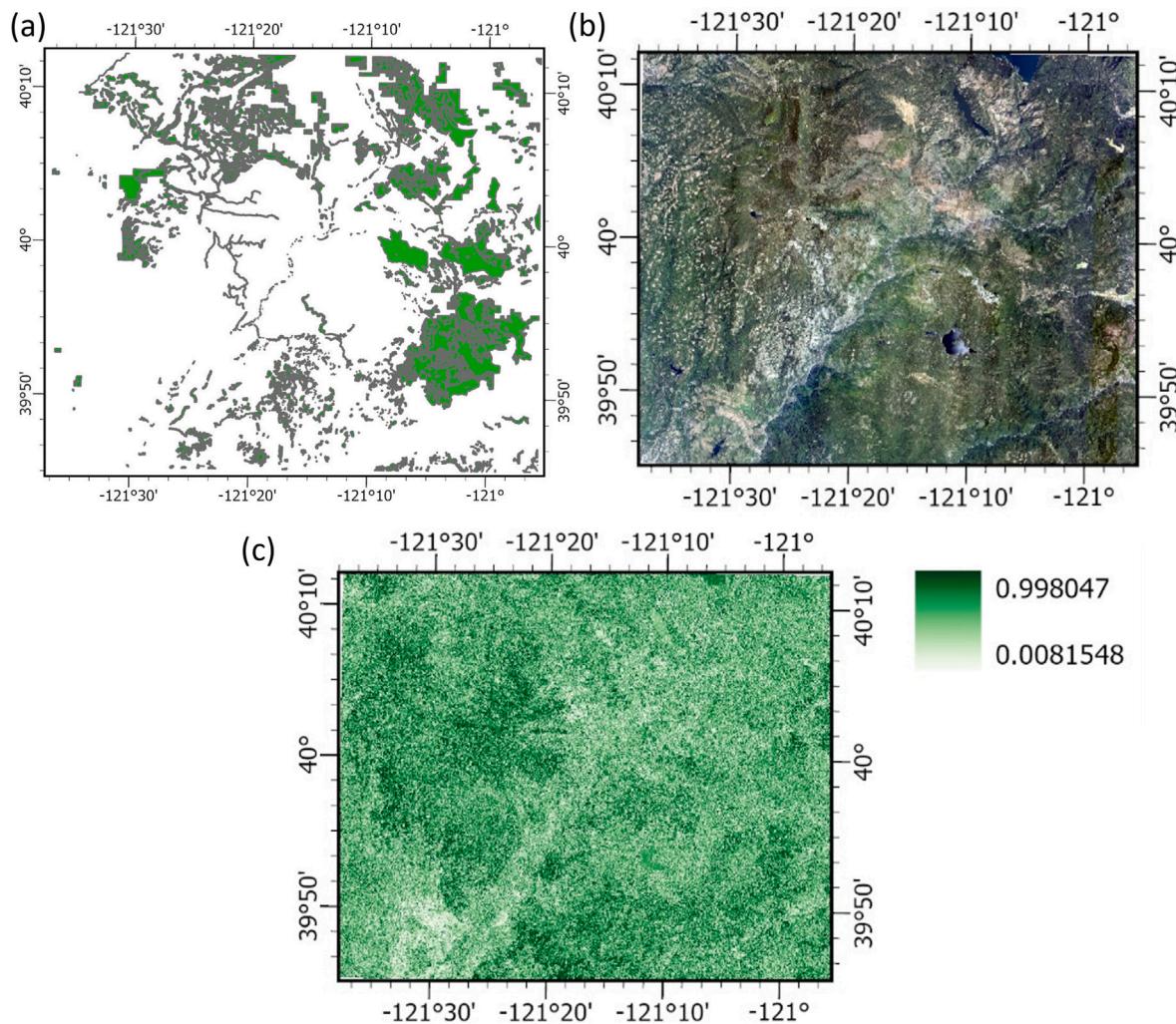


Fig. 12. Analysis of model uncertainty using (a) Timber harvest map for Dixie fires region, (b) NAIP high-resolution imagery for the same region and (c) Dixie fire's fuel prediction probability map.

The proposed fuel identification system in our paper was created using data from the FIA program. The FIA program focuses on ‘forested land’ with at least 10% tree cover. In other words, while data from non-forested lands is not directly used in model training, the model is expected to approximate results in burnable land cover with tree cover smaller than the FIA definition for forest land. Given that forest lands comprise the most important areas subject to wildland fires, this work did not focus on performance evaluation in burnable areas that fall outside FIA-designated forest lands. Such performance evaluation requires extensive in situ data collection in non-forested but burnable lands, which can be the subject of future studies. For non-burnable areas, a separate post-processing step was devised and explained in the corresponding section. Overall, the proposed system presents a successful methodology to leverage the FIA database as the largest existing repository of data on in situ measurements of understory for wildfire fuel mapping.

While the FIA data used in this study has proven exceptionally helpful for model training, [Woodall and Lutes \(2005\)](#) conducted a sensitivity analysis on the adequacy of fine woody debris representation in this dataset. They argued that the uncertainty in estimating the occurrence of small 1-hour fuel particles, combined with assumptions about quadratic mean diameter ([Woodall and Monleon, 2010](#)) significantly influence the resulting fuel loading calculations for any given forest, more so than many other aspects of the DDW inventory program. While a study by [Yanai et al. \(2022\)](#) reported approximately

a 1% discrepancy in tree species identification using FIA data from the northern United States, there is no existing literature addressing the uncertainty in understory fuels in FIA data. This highlights the importance of accounting for uncertainty in practical applications, particularly when quantitative accuracy is critical. Nevertheless, FIA data remains a valuable resource for mapping and qualitatively characterizing fuel loads ([Shaik et al., 2024a](#)) and fine wood carbon ([Shaik et al., 2024b](#)). Furthermore, uncertainty in the training input data is expected due to factors such as temporal and spatial resolution differences and potential mismatches between the FIA plot location coordinates and spaceborne data ([Schrader-Patton et al., 2015](#)). For instance, if the multispectral imagery for a plot is obscured by clouds on the FIA measurement date, we used the nearest available imagery, which could introduce some uncertainty. Uncertainty is a well-recognized challenge in any identification task, arising from various sources like randomness in the data, models, and sensors, as well as environmental noise. Knowledge of the uncertainty in identified fuels is crucial, as it allows us to account for uncertainties in wildfire simulations, which can inform risk assessment and decision-making processes ([Riley and Thompson, 2016](#)). Furthermore, understanding the confidence in fuel predictions can serve as a useful tool for model diagnostics and quality control. High uncertainty in identification can point to underlying issues in the data, guiding the users towards actions to improve the accuracy of the fuel identification system.

The FuelVision model potentially improves pipelines for a range of environmental, hazard, and risk modeling. The adaptability of this

framework provides a mechanism to produce more rapid updates that better reflect changes on the landscape due to disturbance. This adaptability is crucial not only to fuels for fire behavior modeling, but improving smoke and emissions estimates, tracking carbon pools, and fuels treatments. Improved characterizations of fuels that eventually lead to specific predictions of duff, litter, and fuel class sizes will be critical as computational fluid dynamics and hybrid models become more prevalent in operational fire decision support and prescribed fire planning. The FuelVision model framework is a starting point to begin refining accuracy and potentially scale to meet a multitude of needs for natural resource managers nationwide.

6. Conclusions

Most existing studies on wildfire fuel mapping focus on developing models that are trained and applicable to small areas of interest. In contrast, this paper presents a model for mapping wildfire fuels in real-time at any selected domain, including large scales such as regional or state levels. The proposed model leverages AutoML to create a predictive model that combines information from optical, SAR, and terrain data. Specifically, we opted to devise an ensemble model with a multilayer stacking strategy that is known for its ability to classify multiple classes. This model utilizes a multi-model stacked ensemble approach, enhancing model performance while also providing a measure of uncertainty for the predicted fuels.

To evaluate the proposed system, we applied it to a dataset labeled with FIA plots and then improved by pseudo-labeling and augmenting synthetic data. The fuel labels for the state of California were based on the Scott and Burgan 40 fuel models. Synthetic data was generated using the CTGAN technique, and we conducted a comparative analysis using TVAE, GCS, CGAN, and SMOTE to select the most suitable method. Additionally, we performed a feature importance analysis to assess the impact of each feature on accuracy. The final results demonstrated the feasibility of the proposed approach, achieving an overall fuel classification accuracy of 77% on an independent test set. Furthermore, analyzing the properties of the system revealed that fusing different modalities of data improves identification accuracy compared to using each data source individually.

To demonstrate the effectiveness of the proposed approach, we considered two cases: the Dixie and Caldor fires. Fuel maps were developed for these cases, and uncertainty analysis was conducted using the Dixie fire fuel map. The probabilities obtained from the model were found to correspond well with the landscape features. This procedure indicates that the proposed approach can be employed for large-scale real-time wildfire fuel mapping.

CRediT authorship contribution statement

Riyaaz Uddien Shaik: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mohamad Alipour:** Writing – review & editing, Supervision, Methodology. **Eric Rowell:** Writing – review & editing, Supervision. **Bharathan Balaji:** Supervision, Resources. **Adam Watts:** Writing – review & editing, Supervision, Resources. **Ertugrul Taciroglu:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The author is an Guest Editor for International Journal of Applied Earth Observation and Geoinformation and was not involved in the editorial review or the decision to publish this article.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Ertugrul Taciroglu reports financial support was provided by National Science Foundation. Ertugrul Taciroglu reports financial support

was provided by Amazon. Ertugrul Taciroglu reports financial support was provided by Edison International. 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.

Acknowledgments

This work was supported through the National Science Foundation grant CMMI-1953333. The authors Shaik and Taciroglu also acknowledge the additional support provided through UCLA's Amazon Science Hub Program and Edison International. The authors would like to thank Hans Anderson and John Chase from USDA Forest Services for providing the FIA data and assisting in its interpretation. Also, the authors gratefully acknowledge the support of the NVIDIA Corporation for this research. The opinions and perspectives expressed in this study are those of the authors and do not necessarily reflect the views of the sponsor.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.jag.2025.104436>.

Data availability

The ground truth data supporting this study are not publicly available due to confidentiality agreements with U.S. Development Agriculture. Fuel maps for the reader's area of interest can also be generated upon request by contacting the corresponding author or Riyaaz Shaik at riyaaz@risksciences.ucla.edu.

References

- Abufadda, M., Mansour, K., 2021. A survey of synthetic data generation for machine learning. In: 2021 22nd International Arab Conference on Information Technology. ACIT, pp. 1–7. <http://dx.doi.org/10.1109/ACIT53391.2021.9677302>.
- Afifi, S., Dusséaux, R., 2019. Statistical distribution of the normalized difference polarization index for 3-D layered structures with slightly rough interfaces. *IEEE Trans. Antennas and Propagation* 67 (6), 4291–4295.
- Aggarwal, A., Mittal, M., Battineni, G., 2021. Generative adversarial network: An overview of theory and applications. *Int. J. Inf. Manag. Data Insights* 1 (1), 100004.
- Alain, G., Bengio, Y., 2014. What regularized auto-encoders learn from the data-generating distribution. *J. Mach. Learn. Res.* 15 (1), 3563–3593.
- Alipour, M., La Puma, I., Picotte, J., Shamsaei, K., Rowell, E., Watts, A., Kosovic, B., Ebrahimi, H., Taciroglu, E., 2023. A multimodal data fusion and deep learning framework for large-scale wildfire surface fuel mapping. *Fire* 6 (2), 36.
- An, J., Kim, I.S., Kim, K.-J., Park, J.H., Kang, H., Kim, H.J., Kim, Y.S., Ahn, J.H., 2024. Efficacy of automated machine learning models and feature engineering for diagnosis of equivocal appendicitis using clinical and computed tomography findings. *Sci. Rep.* 14 (1), 22658.
- Anderson, H.E., 1981. Aids to Determining Fuel Models for Estimating Fire Behavior, vol. 122, US Department of Agriculture, Forest Service, Intermountain Forest and Range
- Aragonese, E., García, M., Salis, M., Ribeiro, L.M., Chuvieco, E., 2023. Classification and mapping of European fuels using a hierarchical, multipurpose fuel classification system. *Earth Syst. Sci. Data* 15 (3), 1287–1315.
- Ban, Y., Zhang, P., Nascenti, A., Bevington, A.R., Wulder, M.A., 2020. Near real-time wildfire progression monitoring with sentinel-1 SAR time series and deep learning. *Sci. Rep.* 10 (1), 1322.
- Blaes, X., Defourny, P., Wegmuller, U., Della Vecchia, A., Guerrero, L., Ferrazzoli, P., 2006. C-band polarimetric indexes for maize monitoring based on a validated radiative transfer model. *IEEE Trans. Geosci. Remote Sens.* 44 (4), 791–800. <http://dx.doi.org/10.1109/TGRS.2005.860969>.
- Bock, H.-H., 2005. Proximity measures. *Encycl. Stat. Behav. Sci.*
- Breiman, L., 1996. Bagging predictors. *Mach. Learn.* 24, 123–140.
- Burke, M., Driscoll, A., Heft-Neal, S., Xue, J., Burney, J., Wara, M., 2021. The changing risk and burden of wildfire in the United States. *Proc. Natl. Acad. Sci.* 118 (2), e2011048118.
- Bytnarowicz, A., Fenn, M.E., 1996. Nitrogen deposition in California forests: a review. *Environ. Pollut.* 92 (2), 127–146.

- Carbone, A., Spiller, D., Laneve, G., 2023. Fuel type mapping using a CNN-based remote sensing approach: A case study in Sardinia. *Fire* 6 (10), 395.
- Caruana, R., Niculescu-Mizil, A., Crew, G., Ksikes, A., 2004. Ensemble selection from libraries of models. In: Proceedings of the Twenty-First International Conference on Machine Learning. p. 18.
- Chang, T., Rasmussen, B.P., Dickson, B.G., Zachmann, L.J., 2019. Chimera: A multi-task recurrent convolutional neural network for forest classification and structural estimation. *Remote. Sens.* 11 (7), 768.
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. *J. Artificial Intelligence Res.* 16, 321–357.
- Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 785–794.
- Chen, W., Zheng, Q., Xiang, H., Chen, X., Sakai, T., 2021. Forest canopy height estimation using polarimetric interferometric synthetic aperture radar (PolInSAR) technology based on full-polarized ALOS/PALSAR data. *Remote. Sens.* 13 (2), 174.
- Džeroski, S., Ženko, B., 2004. Is combining classifiers with stacking better than selecting the best one? *Mach. Learn.* 54, 255–273.
- East, A.E., Dartnell, P., Logan, J.B., Lindsay, D., 2022. Analysis of post-fire sediment yield in a western sierra nevada watershed burned by the 2021 caldor fire. In: AGU Fall Meeting Abstracts, Vol. 2022. pp. GC55H–0338.
- ED. Chaves, M., C.A. Picoli, M., D. Sanches, I., 2020. Recent applications of landsat 8/OLI and sentinel-2/MSI for land use and land cover mapping: A systematic review. *Remote. Sens.* 12 (18), 3062.
- Elizabeth, A., Glenn, A., Barbara, L., Andrea, M., Lucie, L., Carol, J., Scott, A., Jeffery, A., David, W., Mary, A., 2023. The Forest Inventory and Analysis Database: Database description and users manual version 9.0 for Phase 2. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, <http://dx.doi.org/10.2737/rmrs-gtr-245>.
- Endres, M., Mannarapotta Venugopal, A., Tran, T.S., 2022. Synthetic data generation: A comparative study. In: Proceedings of the 26th International Database Engineered Applications Symposium. pp. 94–102.
- Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A., 2020. Autogluon-tabular: Robust and accurate automl for structured data. arXiv preprint arXiv:2003.06505.
- Flood, N., 2013. Seasonal composite landsat TM/ETM+ images using the medoid (a multi-dimensional median). *Remote. Sens.* 5 (12), 6481–6500.
- Freund, Y., Schapire, R.E., et al., 1996. Experiments with a new boosting algorithm. In: Icm, Vol. 96. Citeseer, pp. 148–156.
- Gao, X., Lin, J., Qu, C., Wang, C., Wu, A., Zhu, J., Xu, C., 2024. Computer-aided diagnostic system with automated deep learning method based on the AutoGluon framework improved the diagnostic accuracy of early esophageal cancer. *J. Gastrointest. Oncol.* 15 (2), 535.
- Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Mach. Learn.* 63, 3–42.
- Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J., 2021. A review on generative adversarial networks: Algorithms, theory, and applications. *IEEE Trans. Knowl. Data Eng.* 35 (4), 3313–3332.
- Hollenbach, F.M., Bojinov, I., Minhas, S., Metternich, N.W., Ward, M.D., Volfovsky, A., 2021. Multiple imputation using Gaussian copulas. *Sociol. Methods Res.* 50 (3), 1259–1283.
- Huang, H., Liu, C., Wang, X., 2019. Constructing a finer-resolution forest height in China using icesat/glas, landsat and Alos Palsar data and height patterns of natural forests and plantations. *Remote. Sens.* 11 (15), 1740.
- Iglesias, V., Balch, J.K., Travis, W.R., 2022. US fires became larger, more frequent, and more widespread in the 2000s. *Sci. Adv.* 8 (11), eabc0020.
- Jarvis, A., Guevara, E., Reuter, H., Nelson, A., 2008. Hole-Filled SRTM for the Globe : Version 4 : Data Grid. CGIAR Consortium for Spatial Information, Published by CGIAR-CSI on 19 August 2008.
- Kalabokidis, K., Ager, A., Finney, M., Athanasis, N., Palaiologou, P., Vasilakos, C., 2016. AEGIS: a wildfire prevention and management information system. *Nat. Hazards Earth Syst. Sci.* 16 (3), 643–661.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* 30.
- Keane, R.E., Reeves, M., 2011. Use of expert knowledge to develop fuel maps for wildland fire management. In: Expert Knowledge and Its Application in Landscape Ecology. Springer, pp. 211–228.
- Keane, R.E., et al., 2015. Wildland Fuel Fundamentals and Applications. Tech. Rep., Springer.
- Koley, S., Chockalingam, J., 2022. Sentinel 1 and sentinel 2 for cropland mapping with special emphasis on the usability of textural and vegetation indices. *Adv. Space Res.* 69 (4), 1768–1785.
- Labenski, P., Ewald, M., Schmidlein, S., Fassnacht, F.E., 2022. Classifying surface fuel types based on forest stand photographs and satellite time series using deep learning. *Int. J. Appl. Earth Obs. Geoinf.* 109, 102799.
- Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* 18 (1), 559–563.
- Li, Z., Ding, L., Xu, D., 2022. Exploring the potential role of environmental and multi-source satellite data in crop yield prediction across Northeast China. *Sci. Total Environ.* 815, 152880.
- Liu, D., Wang, Y., 2021. Estimation of tree height using the ESPRIT algorithm with ALOS/PALSAR dual-pol InSAR datasets. In: EUSAR 2021; 13th European Conference on Synthetic Aperture Radar. VDE, pp. 1–4.
- Mahdianpari, M., Salehi, B., Mohammadiresh, F., Homayouni, S., Gill, E., 2018. The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Remote. Sens.* 11 (1), 43.
- Masarotto, G., Varin, C., 2017. Gaussian copula regression in R. *J. Stat. Softw.* 77, 1–26.
- McElfresh, D., Khandagale, S., Valverde, J., Prasad C, V., Ramakrishnan, G., Goldblum, M., White, C., 2024. When do neural nets outperform boosted trees on tabular data? *Adv. Neural Inf. Process. Syst.* 36.
- 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., et al., 2019. Rethinking resilience to wildfire. *Nat. Sustain.* 2 (9), 797–804.
- Meyer, D., Nagler, T., Hogan, R.J., 2021. Copula-based synthetic data generation for machine learning emulators in weather and climate: application to a simple radiation model. *Geosci. Model. Dev. Discuss.* 2021, 1–21.
- Montanez, A., et al., 2018. SDV: An Open Source Library for Synthetic Data Generation (Ph.D. thesis). Massachusetts Institute of Technology.
- Mullissa, A., Vollrath, A., Odongo-Braun, C., Slagter, B., Balling, J., Gou, Y., Gorelick, N., Reiche, J., 2021. Sentinel-1 sar backscatter analysis ready data preparation in google earth engine. *Remote. Sens.* 13 (10), 1954.
- Padma, S., Sanjeevi, S., 2014. Jeffries Matusita based mixed-measure for improved spectral matching in hyperspectral image analysis. *Int. J. Appl. Earth Obs. Geoinf.* 32, 138–151.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26 (1), 217–222.
- Palinkas, L.A., Palinkas, L.A., 2020. The California wildfires. *Glob. Clim. Chang. Popul. Displacl. Public Health: Next Wave Migr.* 53–67.
- Parmanto, B., Munro, P.W., Doyle, H.R., 1996. Reducing variance of committee prediction with resampling techniques. *Connect. Sci.* 8 (3–4), 405–426.
- Pathare, A., Mangrulkar, R., Suvarna, K., Parekh, A., Thakur, G., Gawade, A., 2023. Comparison of tabular synthetic data generation techniques using propensity and cluster log metric. *Int. J. Inf. Manag. Data Insights* 3 (2), 100177.
- Patki, N., Wedge, R., Veeramachaneni, K., 2016. The synthetic data vault. In: 2016 IEEE International Conference on Data Science and Advanced Analytics. DSAA, IEEE, pp. 399–410.
- Periasamy, S., 2018. Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on sentinel-1. *Remote Sens. Environ.* 217, 537–549.
- Pickell, P.D., Chavardès, R.D., Li, S., Daniels, L.D., 2020. FuelNet: An artificial neural network for learning and updating fuel types for fire research. *IEEE Trans. Geosci. Remote Sens.* 59 (9), 7338–7352.
- Piedelobo, L., Hernández-López, D., Ballesteros, R., Chakhar, A., Del Pozo, S., González-Aguilar, D., Moreno, M.A., 2019. Scalable pixel-based crop classification combining sentinel-2 and landsat-8 data time series: Case study of the Duero river basin. *Agricult. Syst.* 171, 36–50.
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. *Adv. Neural Inf. Process. Syst.* 31.
- Qi, W., Xu, C., Xu, X., 2021. AutoGluon: A revolutionary framework for landslide hazard analysis. *Nat. Hazards Res.* 1 (3), 103–108.
- Rendana, M., Razi Idris, W.M., Abdul Rahim, S., Ghassan Abdo, H., Almohamad, H., Abdullah Al Dughairi, A., Albanai, J.A., 2023. Effects of the built-up index and land surface temperature on the mangrove area change along the southern Sumatra coast. *For. Sci. Technol.* 1–11.
- Riley, K., Thompson, M., 2016. An uncertainty analysis of wildfire modeling. *Nat. Hazard Uncertain. Assessment: Model. Decis. Support.* 191–213.
- Rollins, M.G., 2009. LANDFIRE: a nationally consistent vegetation, wildland fire, and fuel assessment. *Int. J. Wildland Fire* 18 (3), 235–249.
- Rowell, E., Loudermilk, E.L., Seielstad, C., O'Brien, J.J., 2016. Using simulated 3D surface fuelbeds and terrestrial laser scan data to develop inputs to fire behavior models. *Can. J. Remote Sens.* 42 (5), 443–459.
- Sakellarou, S., Tampeki, S., Samara, F., Sfougaris, A., Christopoulou, O., 2017. Review of state-of-the-art decision support systems (DSSs) for prevention and suppression of forest fires. *J. For. Res.* 28 (6), 1107–1117.
- Sayedain, S.A., Maghsoudi, Y., Eini-Zinab, S., 2020. Assessing the use of cross-orbit sentinel-1 images in land cover classification. *Int. J. Remote Sens.* 41 (20), 7801–7819.
- Schrader-Patton, C., Liknes, G.C., Gatziolis, D., Wing, B.M., Nelson, M.D., Miles, P.D., Bixby, J., Wendt, D.G., Kepler, D., Schaaf, A., 2015. Refining FIA plot locations using LiDAR point clouds. In: Proceedings of the Forest Inventory and Analysis (FIA) Symposium.
- Scott, J.H., 2005. Standard Fire Behavior Fuel Models: A Comprehensive Set for Use with Rothermel's Surface Fire Spread Model. US Department of Agriculture, Forest Service, Rocky Mountain Research Station.

- Shaik, R., Alipour, M., Rowell, E., Balaji, B., Watts, A., Taciroglu, E., 2024a. Estimation of downed woody time-lag fuel loadings with multimodal remote sensing data and ensemble machine learning regression model. In: IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium. pp. 3453–3456. <http://dx.doi.org/10.1109/IGARSS53475.2024.10641641>.
- Shaik, R.U., Alipour, M., Rowell, E., Watts, A., Woodall, C., Taciroglu, E., 2024b. Remote sensing and mapping of fine woody carbon with satellite imagery and super learner. *IEEE Geosci. Remote. Sens. Lett.* 1. <http://dx.doi.org/10.1109/LGRS.2024.3503585>.
- Shaik, R.U., Alipour, M., Shamsaei, K., Rowell, E., Balaji, B., Watts, A., Kosovic, B., Ebrahimiyan, H., Taciroglu, E., 2025. Wildfire fuels mapping through artificial intelligence-based methods: a review. *Earth-Science Reviews* (ISSN: 0012-8252) 262, 105064. <http://dx.doi.org/10.1016/j.earscirev.2025.105064>, <https://www.sciencedirect.com/science/article/pii/S001282522500025X>.
- Shaik, R.U., Alipour, M., Taciroglu, E., 2023. A bibliometric analysis of artificial intelligence-based solutions to challenges in wildfire fuel mapping. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. pp. 1610–1613. <http://dx.doi.org/10.1109/IGARSS52108.2023.10282462>.
- Shaik, R.U., Laneve, G., Fusilli, L., 2022. An automatic procedure for forest fire fuel mapping using hyperspectral (PRISMA) imagery: A semi-supervised classification approach. *Remote. Sens.* 14 (5), 1264.
- Sion, B., Samburova, V., Berli, M., Baish, C., Bustarde, J., Houseman, S., 2023. Assessment of the effects of the 2021 calder megafire on soil physical properties, Eastern Sierra Nevada, USA. *Fire* 6 (2), 66.
- Sismanis, M., Stefanidou, A., Stavrakoudis, D., Gitas, I.Z., 2023. Wildland fuel type mapping in attica using sentinel-2 time-series. In: 2023 8th International Conference on Smart and Sustainable Technologies. SpliTech, IEEE, pp. 1–5.
- Solares-Canal, A., Alonso, L., Rincón, T., Picos, J., Molina-Terrén, D.M., Becerra, C., Armesto, J., 2023. Operational fuel model map for Atlantic landscapes using ALS and sentinel-2 images. *Fire Ecol.* 19 (1), 61.
- Song, Y., Xu, Y., Chen, B., He, Q., Tu, Y., Wang, F., Cai, J., 2022. Dynamic population mapping with AutoGluon. *Urban Informatics* 1 (1), 13.
- Sullivan, A., Baker, E., Kurvits, T., 2022. Spreading Like Wildfire: The Rising Threat of Extraordinary Landscape Fires. UNEP: United Nations Environment Programme.
- Taylor, A.H., Harris, L.B., Skinner, C.N., 2022. Severity patterns of the 2021 dixie fire exemplify the need to increase low-severity fire treatments in California's forests. *Environ. Res. Lett.* 17 (7), 071002.
- Uuemaa, E., Ahi, S., Montbeller, B., Muru, M., Kmoch, A., 2020. Vertical accuracy of freely available global digital elevation models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote. Sens.* 12 (21), 3482.
- Van doninck, J., Tuomisto, H., 2018. A landsat composite covering all Amazonia for applications in ecology and conservation. *Remote. Sens. Ecol. Conserv.* 4 (3), 197–210.
- Wadhwan, R., Sullivan, C., Wickramasinghe, A., Kyng, M., Khan, N., Moinuddin, K., 2022. A review of firebrand studies on generation and transport. *Fire Saf. J.* 103674.
- Wan, Z., Zhang, Y., He, H., 2017. Variational autoencoder based synthetic data generation for imbalanced learning. In: 2017 IEEE Symposium Series on Computational Intelligence. SSCI, IEEE, pp. 1–7.
- Wang, Q., Blackburn, G.A., Onojehuo, A.O., Dash, J., Zhou, L., Zhang, Y., Atkinson, P.M., 2017. Fusion of landsat 8 OLI and sentinel-2 MSI data. *IEEE Trans. Geosci. Remote Sens.* 55 (7), 3885–3899.
- Wang, Y., Li, Z., Zeng, C., Xia, G.-S., Shen, H., 2020. An urban water extraction method combining deep learning and Google Earth engine. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 13, 769–782.
- Woodall, C.W., Lutes, D.C., 2005. Sensitivity analysis of down woody material data processing routines. In: United States Department Agriculture Forest Service General Technical Report NC, vol. 352, US DEPARTMENT OF AGRICULTURE, p. 199.
- Woodall, C.W., Monleon, V.J., 2010. Estimating the quadratic mean diameters of fine woody debris in forests of the United States. *Forest Ecol. Manag.* 260 (6), 1088–1093.
- Wulder, M.A., Hilker, T., White, J.C., Coops, N.C., Masek, J.G., Pflugmacher, D., Crevier, Y., 2015. Virtual constellations for global terrestrial monitoring. *Remote Sens. Environ.* 170, 62–76.
- Xu, L., Skouliaridou, M., Cuesta-Infante, A., Veeramachaneni, K., 2019. Modeling tabular data using conditional gan. *Adv. Neural Inf. Process. Syst.* 32.
- Xu, P., Tsendsazar, N.-E., Herold, M., Clevers, J.G., Li, L., 2022. Improving the characterization of global aquatic land cover types using multi-source earth observation data. *Remote Sens. Environ.* 278, 113103.
- Yanai, R.D., Young, A.R., Campbell, J.L., Westfall, J.A., Barnett, C.J., Dillon, G.A., Green, M.B., Woodall, C.W., 2022. Measurement uncertainty in a national forest inventory: results from the northern region of the USA. *Can. J. Forest Res.* 53 (3), 163–177.
- Zhang, Z., 2016. Introduction to machine learning: k-nearest neighbors. *Ann. Transl. Med.* 4 (11).
- Zhang, K., Patki, N., Veeramachaneni, K., 2022. Sequential models in the synthetic data vault. arXiv preprint [arXiv:2207.14406](https://arxiv.org/abs/2207.14406).