From Vegetation to Vulnerability: Integrating Remote Sensing and AI to Combat Cheatgrass-Induced Wildfire Hazards in California

Srikantnag A. Nagaraja[1,4], Istvan Kereszy[1,2], Chang Zhao[3], and Imre Bartos[1,2,*]

[1]Fire Neural Network, Gainesville, Florida, USA
[2]Department of Physics, University of Florida, Gainesville, Florida, USA
[3]Agronomy Department, UF/IFAS, Gainesville, Florida, USA
[4]Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA
[*]Corresponding Author: Imre Bartos (email: imrebartos@ufl.edu)

***Abstract*—Wildfire risk is on the rise around the world. In places like California, this risk is further instigated by the invasive species cheatgrass (Bromus tectorum). Cheatgrass is highly flammable and benefits from wildfires, allowing it to replace native plant communities. Through increasing both the intensity and the frequency of wildfires, it endangers not only its natural environment but also human habitats. Here, we present a novel approach to map the distribution and expansion of cheatgrass and predict potential wildfire risk zones. Utilizing the open-source CalFlora dataset, alongside data from the Sentinel–2 satellites, we created a comprehensive spatial analysis framework. We integrated temporal dynamics via Vegetation Index statistical bands that encapsulate annual vegetation information. We employed semi-supervised learning techniques to refine and filter our data labels, thereby ensuring robust model training. We utilized machine learning algorithms Random Forest and XGBoost, for model training. Our models exhibited a test accuracy of 91.1% in multiclass classification and achieved a precision rate of 91% specifically for the Cheatgrass class. Our multiclass classification model demonstrates exceptional discriminative ability and agreement with the actual classifications, with an ROC-AUC Score of 0.99 indicating near-perfect performance in distinguishing between the different classes, and a Cohen's Kappa of 0.89 signifying a strong agreement, accounting for chance. We demonstrate the effectiveness of our methodology by leveraging publicly available open-source datasets to map the spread of invasive Cheatgrass, which in turn helps identify regions potentially at high risk for wildfires across California's varied landscapes. Our analysis effectively predicts the distribution of Cheatgrass and other vegetation with data available only until June, providing insight before the peak forest fire season, which spans from mid-July to September. This capability delivers actionable intelligence for assessing fuel load and connectivity, thus laying the groundwork for targeted wildfire prevention strategies and enhanced ecological management practices in fire-prone areas.**

***Index Terms*—Land use and Land cover Mapping, Cheatgrass, Invasive Plants, Machine Learning, Semi-supervised Learning, Wildfire Prevention, Time-series Analysis.**

I. Introduction

Wildfires have become increasingly frequent across the globe, often underscored by their destructive impacts on ecosystems, infrastructure, and public health. Yet, fire also plays an essential role in maintaining the ecological balance of many forest systems. Recent concerns stem from observed shifts such as increased fire frequency, extended fire seasons, and escalating severity worldwide [1], [2].

In 2020, California witnessed record-breaking wildfire activity, with approximately 1.74 million hectares burned—more than double the previous high. The economic damage exceeded $19 billion [3], while wildfire smoke exposure led to an estimated loss of 0.7 to 2.6 million workdays due to $PM_{2.5}$-related health effects [4].

The complex dynamics of wildfire spread are influenced by a multitude of factors, including declining summer precipitation across the Western U.S. [5], topographical variation [6], fuel connectivity, and ignition sources. One invasive species—Cheatgrass (Bromus tectorum)—has been especially influential in altering fire regimes across the Intermountain West, particularly the Great Basin. Cheatgrass not only reduces biodiversity by displacing native flora and fauna [7], [8], [9], [10], but also enhances fire frequency and intensity by increasing the availability of fine fuels [11], [12]. Post-fire recovery is often hindered by Cheatgrass dominance, especially in hotter, drier regions [13].

Remote sensing has emerged as a critical tool in monitoring vegetation dynamics, assessing ecosystem structure [14], detecting individual plant species [15], and evaluating post-fire effects [16], [17]. Among various spectral cues, phenology-based detection is particularly effective for invasive species like Cheatgrass due to its early-season productivity relative to native vegetation [18], [19]. This trait shifts the timing of peak greenness, making it distinguishable in satellite time series such as Sentinel-2 [20].

Land use/land cover (LULC) classification techniques have evolved from manual interpretation and basic pixel-based methods to advanced machine learning and object-based approaches [21]. In tandem, spectral vegetation indices—such as NDVI and EVI—have become central to improving classification accuracy. For instance, Sentinel-2-derived vegetation indices have achieved nearly 80% accuracy in mapping invasive shrub species, outperforming even radar-optical data fusion [22]. The 5-day revisit time and 10–20 m spatial resolution of Sentinel-2 [23] enable multi-temporal phenological analysis through seasonal composites [24]. In this study, we extend these methods by introducing Vegetation Index Temporal Statistical Bands (Time Series Spectral Bands), leveraging monthly distributions to better capture seasonal variation in vegetation dynamics for LULC classification.

A persistent challenge in LULC mapping is the lack of high-quality, globally consistent ground truth data. Collecting reliable training and validation samples across large regions is both costly and time-consuming [25]. To address this, our study utilizes the open-source Calflora dataset [26] as a source of labeled observations for land cover classification, offering species-level data with fine

spatial and temporal resolution.

Previous work on Cheatgrass distribution has typically relied on coarse-resolution fuelscape datasets (e.g., 270-meter resolution for sagebrush ecosystems [27]) or general invasive grass mapping using NDVI-derived phenometrics [28]. While effective, such approaches often lack specificity, temporal depth, or spectral diversity. Moreover, past methods have not fully exploited the potential of vegetation index distributions across seasons. Operational fire management plans like San Bernardino County's CWPP [29] have begun integrating such data for targeting cheatgrass in fuel mitigation, underscoring the real-world need for precise mapping solutions.

In this study, we present a Sentinel-2-based methodology at 10-meter resolution, incorporating monthly phenological statistics and species-specific labels from Calflora to classify Cheatgrass distribution. By targeting the reproductive and senescence phases (May–June) as well as full-year vegetation dynamics (from July of the previous year to June of the current year), our approach supports timely and scalable mapping prior to the peak wildfire season. This enables enhanced early warning and ecosystem management strategies for high-risk areas.

The main objectives of this research are as follows:

- Leverage satellite remote sensing and the open-source Calflora dataset for land use and land cover classification, with a specific focus on detecting Cheatgrass.
- Introduce Vegetation Index Temporal Statistical Bands—derived from monthly vegetation indices—to better capture seasonal vegetation dynamics.
- Analyze spatial patterns in the generated LULC map in relation to historical wildfire events in California using zonal statistics and visual interpretation.

These objectives aim to provide a comprehensive approach to managing vegetation in fire-prone regions, with a particular emphasis on understanding and controlling Cheatgrass to reduce wildfire risk and its ecological impact.

II. Data and Methodology

A. Sentinel–2 Data Acquisition and Preprocessing

The Copernicus Sentinel-2 mission consists of two polar-orbiting satellites positioned in the same sun-synchronous orbit, with a phase difference of 180°. This configuration aims to monitor changes in land surface conditions efficiently. Equipped with a multispectral imager (MSI) that has 13 spectral bands, Sentinel-2 provides moderate resolution imagery with a swath width of 290 km. This wide swath and the high revisit time enable global coverage every five days, supporting the continuous monitoring of the Earth's surface. The primary objectives of the Sentinel-2 satellites include providing data for risk management, land use and land cover mapping, change detection, natural hazards, and water management. These capabilities are crucial for a wide range of applications in environmental monitoring and management [23]. We acquired the Sentinel-2 images using Google Earth Engine [30].

We applied a bilinear interpolation technique to enhance the resolution of specific spectral bands at 20 meters to a uniform 10 meters in order to maintain consistent resolution across all bands.

TABLE I: Sentinel–2A Bands Used

Band	Resolution	Central Wavelength	Description
B2	10m	490 nm	Blue
B3	10m	560 nm	Green
B4	10m	665 nm	Red
B5	20m	705 nm	VNIR
B6	20m	740 nm	VNIR
B7	20m	783 nm	VNIR
B8	10m	842 nm	VNIR
B8a	20m	865 nm	VNIR
B11	20m	1610 nm	SWIR
B12	20m	2190 nm	SWIR

Note: VNIR refers to Visible and Near Infrared; SWIR refers to Shortwave Infrared.

Bilinear interpolation was chosen because the bands represented continuous imagery, rather than segmentation [31]. This refinement was applied to those bands originally at 20-meter resolution, resulting in a consistent 10-meter resolution across the selected bands. Here, we used a total of ten spectral bands ranging from the visible to the shortwave infrared wavelengths (Table I) ([32]). As an initial step, Sentinel–2 Surface Reflectance (SR) images with less than 20% cloud coverage were systematically selected for the months of May and June 2022.

In our time series analysis, addressing the significance of cloud interference is essential. To mitigate this issue, we first utilized the Cloud Probability Band from Sentinel-2 Level-2A products [33], generating its inverse by multiplying the Cloud Probability values by -1, which was subsequently used as a weighting factor during the creation of temporal mosaics, effectively minimizing noise. Subsequently, we applied additional masks to filter out remaining cloud and snow pixels [33], [34]. The resulting missing data were handled by excluding the masked pixels during the computation of the time series spectral bands, as detailed in Subsection II-D. These combined techniques greatly reduced noise, providing a strong method for time series analysis in our study. During the masking process, the results were visually inspected, and all parameters were refined through multiple iterations to ensure optimal outcomes across all months.

B. Dataset

The Calflora dataset [26] was employed to assess the spatial distribution of Cheatgrass (Bromus tectorum) across California (Figure 2). As a comprehensive repository of georeferenced observations of wild plant species in the state, Calflora provides valuable data to support ecological and biogeographical research. The dataset offers detailed, species-level occurrence records that are critical for analyzing habitat preferences and monitoring the spread of invasive taxa such as Cheatgrass, which poses significant ecological threats. However, the dataset includes certain spatial inaccuracies, as illustrated in Figure 1, primarily resulting from imprecise GPS coordinates at the time of species reporting. These deviations—also evident in Figure 1—may arise from observations logged near anthropogenic features such as roads or buildings rather than from exact in-situ locations, introducing potential spatial bias into ecological analyses.

Fig. 1: Some inaccuracies in the Caflora [26] dataset with High-resolution Imagery Background, which is created by variations in collectors GPS accuracy [35].

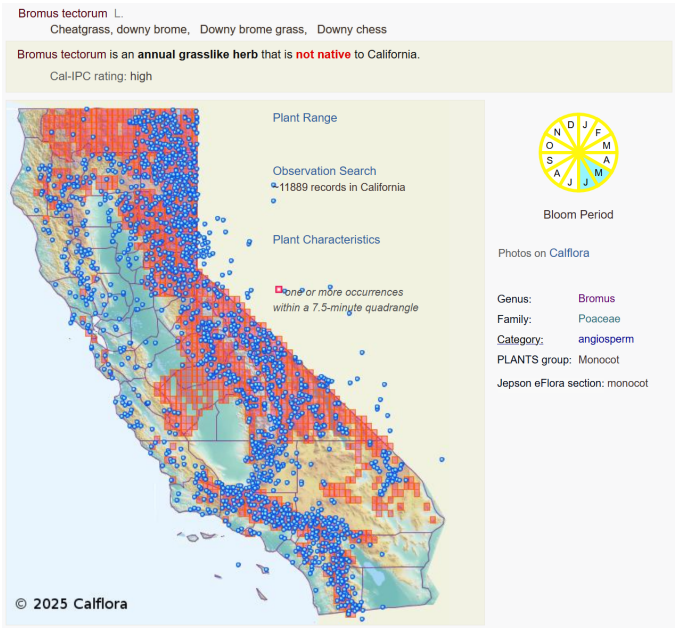

Fig. 2: A screenshot of Cheatgrass Data from the Calflora Website showing its distribution across California [26].

C. Labeling and Refining

Our research commenced with systematically gathering Cheatgrass sample locations from the CalFlora dataset [26] for Los Angeles County, spanning the period from January 1, 2016, to January 10, 2024. Los Angeles County was selected due to its ecological diversity, encompassing coastal areas, mountain ranges, valleys, forests, islands, lakes, rivers, and deserts [36], making it representative of broader Californian landscapes. To categorize regions by land cover type—such as water, grass, shrubland, trees, and bareland—the Dynamic World V1 Land Cover map was employed to extract representative samples of each class using stratified random sampling [37], [38]. Each sample was meticulously verified using Sentinel-2 imagery from April and June 2022, focusing on distinguishing color contrasts in the RGB bands within Los Angeles County. This process was further cross-checked with a high-resolution Imagery base-map in ArcGIS PRO [35]. Given that Cheatgrass often grows in patches around 300 m² [39], the 10-meter resolution of Sentinel-2 proved valuable for identifying pure pixels. Points that appeared to be mixed on high-resolution images were relocated to the nearest pixel that accurately represented the targeted vegetation type. This detailed approach produced a dataset of 650 samples across six distinct land cover types: Grassland, Shrubland, Bareland, Water, Cheatgrass, and Trees (Table II).

TABLE II: Initial 650 samples

Category	Sample Size
Trees	205
Grassland	38
Shrubland	233
Bare Soil	34
Water	30
Cheatgrass	110
Total	650

Since Data collected from Calflora in the data preparation phase, we addressed the inherent inaccuracies in some user-reported entries within the Calflora dataset(Figure 1), specifically focusing on Cheatgrass samples. For each Cheatgrass point across California from the Calflora dataset, we established a 200-meter radius buffer zone and generated four random peripheral points at least 50 meters apart from each other, resulting in a total of five distinct sampling points for each original location.

These points were then used to extract spectral values from Sentinel-2 imagery captured in April and June. The extracted data was input into a Binary Multilayer Perceptron (MLP) Classifier [40] (Table III), which was trained using 80% of the dataset, while the remaining 20% was reserved for validation (Table II). We tested a range of probability thresholds between 0.4 and 0.7 to evaluate the relationship between the number of input points labeled as Cheatgrass and the classifier's corresponding predictions. A threshold of 0.5 was ultimately selected, as it provided an optimal balance: it consistently identified one confirmed Cheatgrass point per instance, along with four additional points randomly sampled within a defined spatial buffer around each positive case. This configuration ensured that the sampling strategy neither oversampled nor under sampled the target class. The additional points functioned as a localized random search to identify spectral patterns closely resembling the confirmed Cheatgrass signature—particularly those trained in Los Angeles County—thus enhancing the model's capacity to generalize to nearby spectral analogs while preserving label integrity. This threshold effectively reduced noise, eliminating 4

out of every 5 generated samples around each point. To further validate the classifier's robustness, a subset of retained points was visually inspected in northeastern California—specifically within Nevada, Placer, and El Dorado counties. More detailed results of this MLP classifier are provided in Appendix A. Figure 3 illustrates how the MLP classifier effectively addresses inaccuracies present in the Calflora dataset. The Figure 3 clearly demonstrates the model's ability to reclassify mislocated points—originally assigned to non-grass species—by repositioning them to more plausible grass-dominated areas.

TABLE III: Binary Multilayer Perceptron (MLP) Architecture

Layer (type)	Output Shape	Param #
dense_8 (Dense)	(None, 256)	5,376
dropout_6 (Dropout)	(None, 256)	0
dense_9 (Dense)	(None, 128)	32,896
dropout_7 (Dropout)	(None, 128)	0
dense_10 (Dense)	(None, 64)	8,256
dropout_8 (Dropout)	(None, 64)	0
dense_11 (Dense)	(None, 1)	65
Total params:	46,593	
Trainable params:	46,593	
Non-trainable params:	0	

Optimizer: Adam
Loss: Binary Crossentropy
Metrics: Accuracy

This binary classification approach was replicated across all classes, leading to a refined set of labeled samples. Finally, we amalgamated the refined dataset of 650 samples to train a comprehensive multiclass classifier. This classifier aimed to effectively distinguish between the defined classes, leveraging the refined and accurately labeled dataset to enhance the predictive accuracy and reliability of the model. Except for Cheatgrass, all other class samples were sourced from Los Angeles County rather than across California. Finally we had 8855 samples across 6 classes as shown in Table IV.

TABLE IV: Final Sample Design

Category	Sample Size
Trees	1997
Grassland	1055
Shrubland	2268
Bare Soil	1498
Water	298
Cheatgrass	1739
Total	8855

This kind of label refining method would help us to integrate datasets like iNaturalist in the future [41].

D. Vegetation Indices

In our study, we employ three key vegetation indices: NDVI [42], GNDVI [43], and MSAVI2 [44](Equation 1, 2, and 3). These indices are essential for monitoring vegetation health and land surface dynamics using remote sensing.

The Normalized Difference Vegetation Index (NDVI) [42], [45] is one of the most widely used spectral indices in agriculture and environmental monitoring. It quantifies vegetation vigor and density,

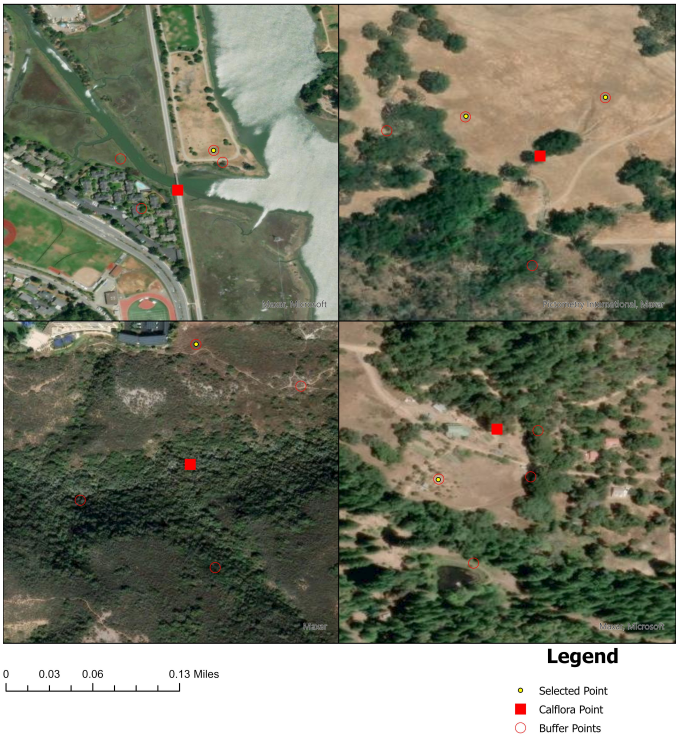

Fig. 3: Label refinement using an MLP classifier, overlaid on a high-resolution imagery background. The red squares indicate original Calflora dataset points, hollow red circles denote four randomly generated points, and yellow points represent those identified as Cheatgrass by the MLP. The alignment of yellow points with visual cues from the high-resolution map supports the classifier's accuracy [35].

supporting applications such as crop emergence analysis, growth stage tracking, weed or disease detection, and yield forecasting. NDVI is calculated using the equation 1:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{1}$$

where NIR represents the near-infrared reflectance (Band 8 for Sentinel-2) and Red denotes the red reflectance (Band 4 for Sentinel-2). Healthy vegetation absorbs most red light (0.62–0.75 μm) and reflects a large portion of NIR light (0.75–1.3 μm), enabling strong separation of vegetated and non-vegetated surfaces [45].

The Green Normalized Difference Vegetation Index (GNDVI) [43], [45] is a variation of NDVI that uses the green band instead of red, given by the equation 2:

$$\text{GNDVI} = \frac{\text{NIR} - \text{Green}}{\text{NIR} + \text{Green}} \tag{2}$$

Here, Green corresponds to Band 3 in Sentinel-2 (0.54–0.57 μm). GNDVI is more sensitive to chlorophyll content and is often used to assess nitrogen levels and moisture stress in vegetation.

The MSAVI2 index is computed as in the equation 3:

$$\text{MSAVI2} = \frac{2 \cdot \text{NIR} + 1 - \sqrt{(2 \cdot \text{NIR} + 1)^2 - 8 \cdot (\text{NIR} - \text{Red})}}{2} \tag{3}$$

MSAVI2 is particularly effective for monitoring early-stage crop development in regions with sparse vegetation cover, as it enhances the ability to differentiate vegetation from exposed soil [46].

The distinct advantages offered by each vegetation index guided their selection in this study. NDVI's capability to capture dense vegetation cover, GNDVI's sensitivity to chlorophyll concentration, and MSAVI2's strength in distinguishing vegetation from bare soil—particularly in sparsely vegetated areas—were instrumental in effectively mapping Cheatgrass, which often exhibits variable density and partial ground exposure.

E. Time Series Spectral Bands

Given the annual lifecycle of Cheatgrass, effectively predicting its proliferation requires a detailed understanding of its temporal variability. The bloom period of Cheatgrass, typically spanning May to June, serves as a key temporal indicator that differentiates it from other vegetation. To leverage this distinct seasonal behavior, we introduced an innovative set of features derived from the temporal dynamics observed in spectral data. Specifically, we utilized data from May 2022 and June 2022 to inform our analysis.

To incorporate indirect time series information for each pixel over a year, we employed a novel approach by generating statistical bands. These bands encapsulate the skewness, kurtosis, mean, and standard deviation of three key vegetation indices: NDVI [42], GNDVI [43], and MSAVI2 [44], covering the period from July of the previous year to June of the prediction year. This integration added twelve additional layers to our model, significantly enriching the dataset with temporal insights and enhancing the model's accuracy in distinguishing Cheatgrass from other vegetation types. Prior to calculating the skewness, kurtosis, mean, and standard deviation bands, monthly mosaicked Sentinel-2 images [33] were processed to derive NDVI, GNDVI, and MSAVI2, which were then used to compute the statistical bands.

Figure 4 illustrates how the NDVI distribution varies across different land cover classes over the months and highlights how metrics such as Mean, Standard Deviation, Kurtosis, and Skewness of average monthly vegetation index can aid in distinguishing between these classes.

This methodology underscores the importance of integrating time-sensitive spectral differences and statistical analyses to improve the predictive modeling of Cheatgrass, ensuring robustness across varying temporal scales.

F. Classification Algorithm

We applied two powerful ensemble machine learning algorithms, XGBoost and Random Forest (RF), to identify Cheatgrass. XGBoost, known for its efficiency in classification tasks [47], leverages an ensemble of decision trees to boost classification accuracy. Its high predictive power is due to its effective loss function and the optimization of weak learners, leading to superior model performance [47], [48], [49]. The algorithm's iterative and additive training process, combined with a stringent regularization framework, ensures robustness against overfitting [50], [7]. Detailed mathematical

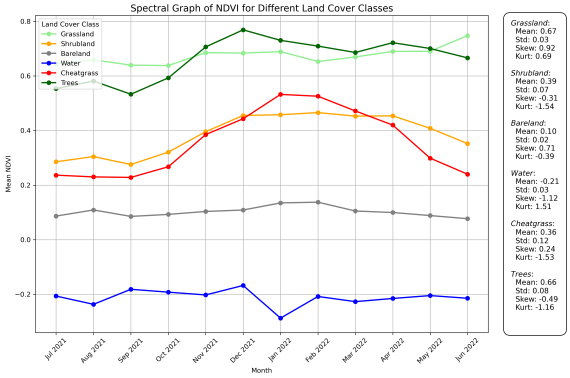

Fig. 4: NDVI variation across different classes from July 2021 to June 2022.

insights into this algorithm are thoroughly documented in Chen's works [47].

The Random Forest (RF) algorithm [51] employs bootstrapping to generate numerous decision trees from different data samples. Each decision tree utilizes the Classification and Regression Trees (CART) algorithm to split nodes by minimizing Gini Impurity, which quantifies the likelihood of a misclassification if a new random variable were assigned a random label based on the training sample distribution. The overall classification is performed through bootstrap aggregation. This method was selected due to its proven high accuracy in recent studies, particularly in land cover classification tasks [52], [48], [53]. Comprehensive details of the RF algorithm can be found in specialized literature [51], [54].

G. Example use case

Our study examines two significant wildfire cases: the El Dorado Fire of 2020 and the Corral Fire of 2024, highlighting the role of cheatgrass invasion in post-fire fuel accumulation and its contribution to fire spread. The El Dorado Fire, which burned approximately 22,744 acres (9,204 ha; 35.538 sq mi; 92.04 km²) across San Bernardino and Riverside counties, was ignited on September 5, 2020, by a pyrotechnic device in El Dorado Ranch Park and quickly expanded into the San Gorgonio Wilderness Area within the San Bernardino National Forest. The fire burned for 71 days, destroying 20 structures and resulting in a firefighter fatality [55].

To analyze vegetation dynamics pre- and post-fire, we focused on a 100-mile by 100-mile region encompassing the El Dorado fire perimeter (Table XIII), allowing us to compare vegetation changes within the burned area to surrounding areas not impacted by the fire. The land use and land cover (LULC) data for 2020 were collected up to June, just before the fire's occurrence in September. Since the West Coast's proximity introduced marine areas into this 100-mile radius—irrelevant to our study—we adjusted the centroid 20 miles northeast to focus exclusively on terrestrial changes (Figure 14).

By contrast, the Corral Fire of 2024 in San Joaquin County exemplifies the rapid advancement of grassland fires. First detected at 4:44 pm on June 1, it expanded to over 9,000 acres by 9:46 pm—within just five hours—and eventually burned 14,168 acres in total [56]. Investigations traced the ignition source to 12kV distribution pole 8009, part of feeder 8 near Substation 846. This

pole, originally installed in 1958, was identified as the ignition point[57]. According to the National Interagency Fire Center's Incident Management Situation Report, suppression efforts for this fire cost an estimated $3.4 million [58]. The fire's rapid growth emphasizes the urgent need for advanced modeling tools and real-time monitoring systems to anticipate and manage fast-moving wildfires.

Additionally, Appendix B demonstrates our model's predictive capabilities in classifying shrubland-dominated areas in the Mojave Desert, illustrating the LULC model's robustness in capturing diverse landscapes and terrain types. This example underscores the model's potential in broader applications across various ecosystems and fire-prone environments.

III. Results and Discussion

A. Classification Accuracy

We split the dataset 80:20, training our models with 80% of the data and using the remaining 20% for testing(Table IV).

Additionally, we examined the variance in band intensities between June 2022 and April 2022, as well as 12 time series statistical bands, applying a scaling factor derived from the 1st percentile minimum and maximum values within Los Angeles County.

We then trained models using the Random Forest and XGBoost algorithms, further optimizing their performance through extensive hyperparameter tuning using a 10-fold cross-validated grid search to ensure robust generalization and minimize the risk of overfitting(Table V).

TABLE V: Parameter Grid for 10-Fold Grid Search on Random Forest and XGBoost Models

Model	Parameter	Values
Random Forest	min_samples_split	{2, 3, 5, 6, 7, 8, 10}
Random Forest	n_estimators	{250, 500, 1000, 2000, 3000}
XGBoost	n_estimators	{200, 500, 1000, 3000, 5000}
XGBoost	learning_rate	{0.1, 0.05, 0.01, 0.005, 0.001}

Despite achieving an initial model accuracy of 95%, we encountered challenges in generalizing across different years and counties, partly due to the specific scaling applied using the 1st percentile minimum and maximum values from Los Angeles County. Additionally, residual snow in April led to incorrect predictions for the water class. To address these issues, we refined our dataset to include only observations from May and June 2022, corresponding to the Cheatgrass bloom period. We also incorporated twelve unaltered statistical bands from three vegetation indices to enhance the model's robustness. Data was extracted to labels from bands shown in Table VI. Although this adjustment slightly reduced the overall model accuracy to 91.1%, it significantly improved the model's reliability and performance across both temporal and spatial dimensions. As seen in Table VI, the model uses all 32 input features, and given the importance of the vegetation indices in capturing ecologically relevant patterns, we retained all input bands. Therefore, we did not apply dimensionality reduction techniques, as excluding any of these features could have compromised the interpretability and predictive strength of the model.

TABLE VI: Bands used to extract values for labels.

Data	Date	No. of Bands
June Sentinel-2	06/01/2022 to 06/30/2022	10
May Sentinel-2	05/01/2022 to 05/31/2022	10
NDVI Distribution Bands	07/01/2021 to 06/30/2022	4
GNDVI Distribution Bands	07/01/2021 to 06/30/2022	4
MSAVI-2 Distribution Bands	07/01/2021 to 06/30/2022	4

Note: Distribution bands are Mean, Standard Deviation, Skewness, Kurtosis.

TABLE VII: Classification Accuracy and Confusion Matrix for Random Forest.

True/Predicted	GS	SH	BR	WT	CG	TR
GS	189	0	0	0	**1**	21
SH	2	397	1	0	**34**	17
BR	0	1	298	0	**0**	1
WT	0	0	0	60	**0**	0
CG	**0**	**45**	**2**	**0**	**294**	**6**
TR	14	16	0	1	**1**	367
Column Total	211	450	300	60	**347**	399
Precision (UA)	0.92	0.86	0.99	0.98	**0.89**	0.89
Recall (PA)	0.90	0.88	0.99	1.00	**0.85**	0.92
Overall Accuracy	**0.908**					
Kappa	**0.885**					

Note: GS - Grassland, SH - Shrubland, BR - Bareland, WT - Water, CG - Cheatgrass, TR - Trees, UA - User accuracy, PA - Producer accuracy.

TABLE VIII: Classification Accuracy and Confusion Matrix for XGBoost.

True/Predicted	GS	SH	BR	WT	CG	TR
GS	187	0	0	0	**1**	23
SH	1	403	0	0	**29**	17
BR	0	1	297	0	**1**	1
WT	0	0	0	59	**0**	1
CG	**1**	**43**	**1**	**0**	**299**	**3**
TR	15	18	0	1	**0**	365
Column Total	211	450	300	60	**347**	399
Precision (UA)	0.92	0.87	1.00	0.98	**0.91**	0.89
Recall (PA)	0.89	0.90	0.99	0.98	**0.86**	0.91
Overall Accuracy	**0.911**					
Kappa	**0.889**					

Note: GS - Grassland, SH - Shrubland, BR - Bareland, WT - Water, CG - Cheatgrass, TR - Trees, UA - User accuracy, PA - Producer accuracy.

The ROC curves shown in Figures 5 and 6 provide a detailed evaluation of classifier performance for each class. The Random Forest model's ROC curves (Figure 5) demonstrate high AUC values for most classes, with the Grassland, Bareland, and Water classes achieving perfect scores (AUC = 1.00). The Shrubland and Cheatgrass classes show slightly lower AUC values of 0.98, indicating some misclassification errors. Similarly, the XGBoost model (Figure 6) exhibits high AUC values across all classes, with Grassland, Bareland, and Water also achieving perfect AUC scores. The Cheatgrass and Trees classes have slightly lower AUC values of 0.98 and 0.99, respectively, indicating robust performance with minor classification errors. These high AUC values across both

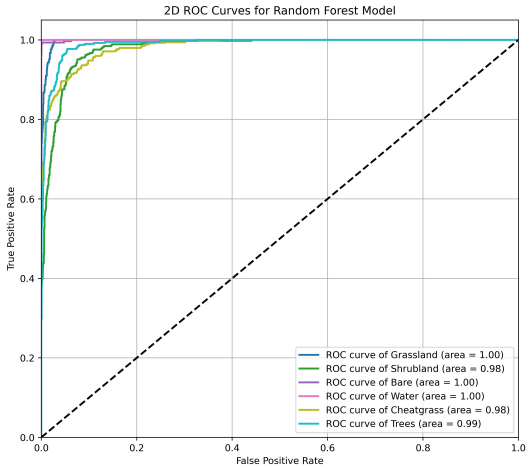

Fig. 5: Random Forest Model ROC Graphs.

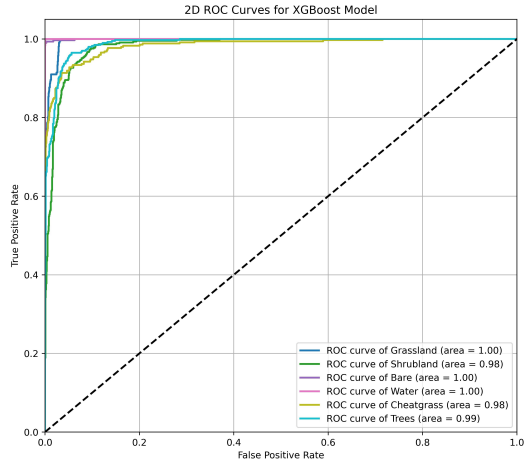

Fig. 6: XGBoost Model ROC Graphs.

models affirm the robustness of our approach in distinguishing between different land cover classes.

The confusion matrix (Table VII and Table VIII) indicates challenges in accurately classifying Shrubland within our dataset. However, the high precision in identifying Cheatgrass confirms the effectiveness of our modeling approach. Currently, our analyses rely on static 2022 data. To enhance the robustness of model performance and increase temporal diversity, raising public awareness about Cheatgrass is crucial. Encouraging the collection of Cheatgrass location data, including precise latitude and longitude coordinates, could help in obtaining high-quality data. Additionally, developing an application for the public to report Cheatgrass locations could provide a large quantity of data. This data can be filtered using a semi-supervised MLP model designed for extracting extra labels, thereby improving our dataset and enabling the development of more sophisticated models. This approach would enhance our analytical capabilities and improve predictive accuracy across diverse ecological scenarios. Furthermore, as shown in Figure 7, which is based on predictions from the XGBoost model, most of the confusion between Cheatgrass and Shrubland occurs along the west coast rather than across California as a whole. Future studies incorporating multi-year datasets and expanding spatial coverage, particularly across California and the western coast, are likely to reduce this confusion and improve classification performance.

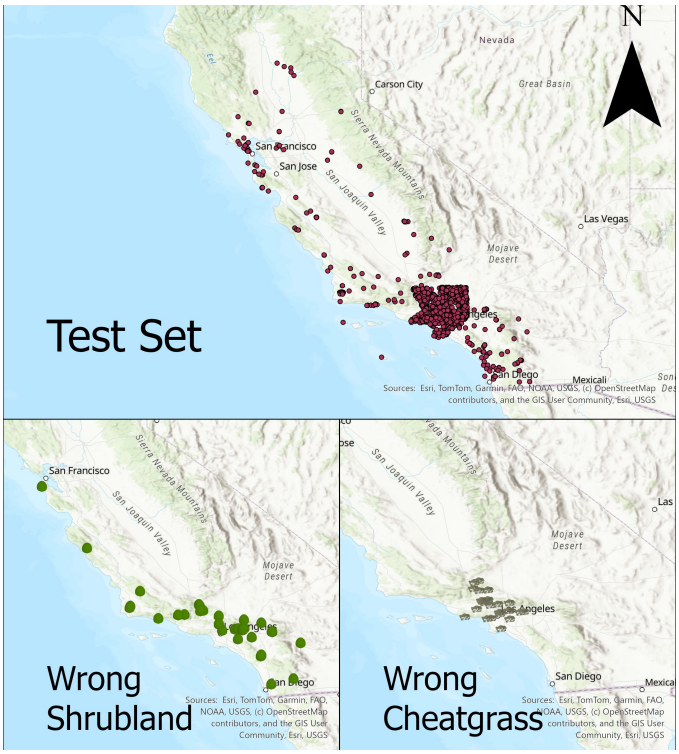

Fig. 7: Clockwise from top: (1) Test dataset distribution, (2) Shrubland misclassified as Cheatgrass, mostly along the coastal region of Los Angeles County, and (3) Cheatgrass misclassified as Shrubland, primarily along portions of the coast.

B. LULC Map

Figure 8 shows the Land Use Land Cover (LULC) map for California State 2024, generated using the trained XGBoost model with 32 bands (Table IX). Figure 9 represents the probability map of the class Cheatgrass derived using the trained XGBoost model with 32 bands (Table IX). This LULC and Cheatgrass Probability map enables the analysis of fuel connectivity, revealing areas with high Cheatgrass density. Additionally, regions where forests are surrounded by Cheatgrass pose a significant risk of large forest fires due to potential ignition sources and fuel connectivity. The analysis excluded built-up and cropland classes because of their heterogeneous landscapes and varying spectral characteristics, especially croplands whose spectral signatures change with the agricultural season. For Figure 8, for built-up areas, the GHSL: Global Settlement Characteristics (10 m) 2018 (P2023A) [59] dataset was used, which is at 10 m resolution. For cropland, the USDA NASS Cropland Data Layers [60] were used from the period between '2016-01-01' and '2017-12-31'. While mosaicking cropland and built-up areas, Cheatgrass pixels were given priority.

TABLE IX: Bands used to extract LULC Map Figure 8 and Figure 9

Data	Date	No. of Bands
June Sentinel-2	06/01/2024 to 06/30/2024	10
May Sentinel-2	05/01/2024 to 05/31/2024	10
NDVI Distribution Bands	07/01/2023 to 06/30/2024	4
GNDVI Distribution Bands	07/01/2023 to 06/30/2024	4
MSAVI-2 Distribution Bands	07/01/2023 to 06/30/2024	4

Note: Distribution bands are Mean, Standard Deviation, Skewness, Kurtosis.

Figure 8 also illustrate a strong correlation between vegetation cover and wildfire incidence patterns, significantly influenced by the classes ”Trees” and ”Cheatgrass”. These maps can be utilized to calculate historical fuel loads and assess fuel connectivity at both macro and community levels. The subsequent subsections provide examples demonstrating their potential for both preventive strategies and post-fire behavior analysis.

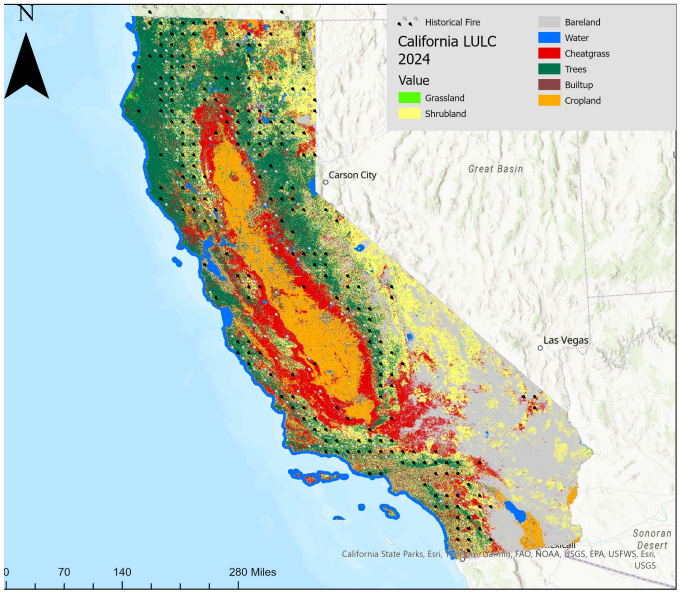

Fig. 8: LULC Map of California 2024.

C. Importance of Time Series Spectral Bands.

We evaluated XGBoost performance using different data sources as summarized in Table X. When using only June or May Sentinel-2 data combined with vegetation index (VI) bands, the recall for Cheatgrass was relatively low, with precision values of 0.85 and 0.86, respectively. Incorporating both May and June Sentinel-2 data with VI bands improved recall, but Cheatgrass precision remained at 0.85. The inclusion of time series statistical bands notably improved precision to 0.91. Figure 10 presents results for 2018, as Sentinel-2 surface reflectance data is unavailable for California prior to 2019—only top-of-atmosphere reflectance exists, which lacks the mask probability band required to compute time series features. As shown in Figure 10 and Table XI, the LULC map generated using reduced data overestimated Cheatgrass, particularly in shrubland regions and along the west coast. These findings suggest that time series spectral bands not only enhanced precision but also mitigated the overprediction of Cheatgrass.

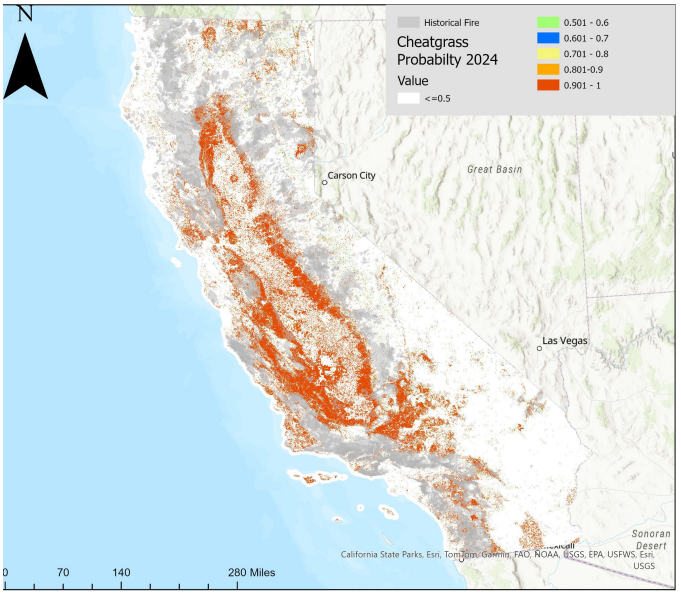

Fig. 9: Cheatgrass Probability Map of California 2024.

TABLE X: Experimentation on XGBoost Model with different sizes of data.

Data	Cheatgrass Precision / Recall	OA
June Sentinel-2 + VI	0.85 / 0.83	0.888
May Sentinel-2 + VI	0.86 / 0.82	0.882
June + May Sentinel-2 + VI (JMV)	0.85 / **0.86**	0.898
JMV + Time Series Spectral Bands	**0.91** / **0.86**	**0.911**

Note: OA = Overall Accuracy, VI = Vegetation Indices(NDVI, GNDVI, MSAVI2), Time series Spectral Bands = Distribution Bands in Table VI, JMV = June and May Sentinel-2 + VI.

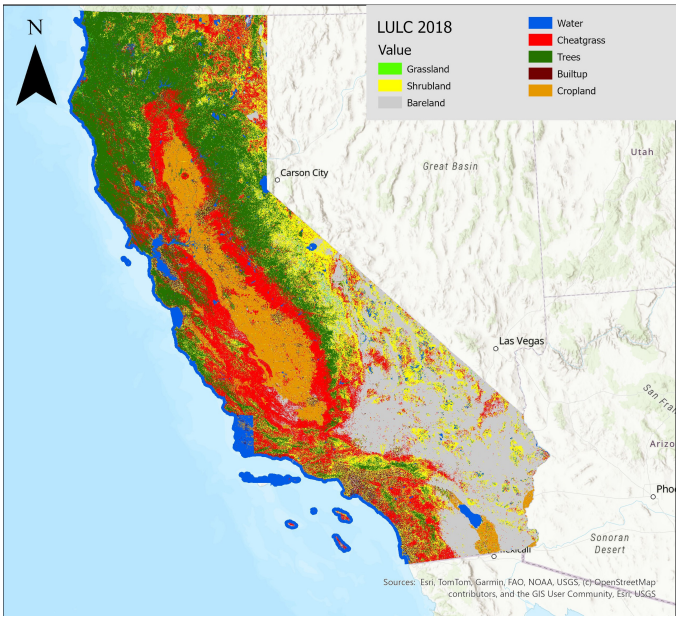

Fig. 10: LULC Map of California 2018 without Time Series Statistical Bands using data in Table XI

TABLE XI: Bands used to extract LULC Map in Figure 10

Data	Date	No. of Bands
June Sentinel-2 + VI	06/01/2018 to 06/30/2018	10
May Sentinel-2 + VI	05/01/2018 to 05/31/2018	10

Note: VI = Vegetation Indices(NDVI, GNDVI, MSAVI2).

D. Example use case: El Dorado Fire 2020 Analysis

TABLE XII: Land Cover Percentages around El Dorado Fire 2020

2020

Land Cover Type	Percentage
Grassland	1.91%
Shrubland	31.00%
Bareland	0.92%
Water/Snow	0.12%
Cheatgrass	13.55%
Trees	52.50%

2021

Land Cover Type	Percentage
Grassland	0.09%
Shrubland	38.97%
Bareland	5.46%
Water/Snow	6.01%
Cheatgrass	48.06%
Trees	1.40%

2022

Land Cover Type	Percentage
Grassland	0.17%
Shrubland	46.66%
Bareland	5.49%
Water/Snow	0.45%
Cheatgrass	45.07%
Trees	2.15%

2023

Land Cover Type	Percentage
Grassland	0.30%
Shrubland	64.97%
Bareland	5.26%
Water/Snow	0.08%
Cheatgrass	22.99%
Trees	6.40%

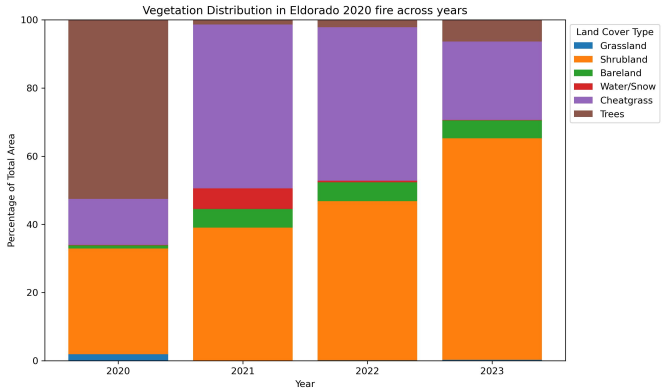

Fig. 11: Vegetation Distribution inside El Dorado Fire 2020 boundary across years.

The El Dorado Fire was a wildfire that burned 22,744 acres (9,204 ha; 35.538 sq mi; 92.04 km2) in San Bernardino and Riverside counties of California from September to November 2020. It was ignited on September 5 by a pyrotechnic device in the El Dorado Ranch Park; it quickly spread to the San Gorgonio Wilderness Area of the San Bernardino National Forest. Burning over a 71-day period, the fire destroyed 20 structures and resulted in one firefighter fatality [55].

We focused on the 2020 El Dorado wildfire to assess vegetation dynamics pre and post-wildfire within a designated 100 by 100 miles (Table XIII) area surrounding the El Dorado fire boundary. It

TABLE XIII: Land Cover Percentages around 100 by 100 miles

2020

Land Cover Type	Percentage
Grassland	0.95%
Shrubland	35.36%
Bareland	39.98%
Water/Snow	1.20%
Cheatgrass	16.79%
Trees	5.73%

2021

Land Cover Type	Percentage
Grassland	0.69%
Shrubland	29.34%
Bareland	50.81%
Water/Snow	1.64%
Cheatgrass	14.01%
Trees	3.51%

2022

Land Cover Type	Percentage
Grassland	0.60%
Shrubland	23.78%
Bareland	57.78%
Water/Snow	1.10%
Cheatgrass	13.24%
Trees	3.50%

2023

Land Cover Type	Percentage
Grassland	0.61%
Shrubland	30.05%
Bareland	49.74%
Water/Snow	1.06%
Cheatgrass	13.96%
Trees	4.58%

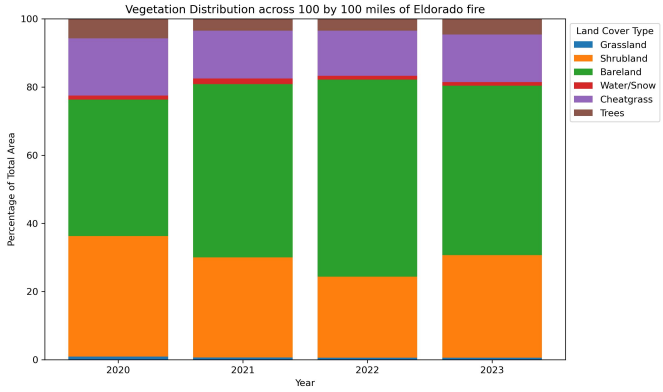

Fig. 12: Vegetation Distribution around 100 by 100 mile of El Dorado Fire 2020 across years.

is important to note that the land use and land cover (LULC) map (Figure 13) for 2020 was compiled using data collected only up until June 2020, two months prior to the actual occurrence of the El Dorado fire in September 2020. Given the proximity of the West Coast—within a 100-mile radius of the fire zone—which included parts of the ocean, and considering that marine areas are not relevant to our study, we adjusted the centroid of the mapped area 20 miles northeast to better focus on the terrestrial environment (Figure 14).

In 2020, recognized as one of the most severe wildfire seasons, the combined presence of trees and Cheatgrass, which constitute critical wildfire fuels, was observed to be 22.52% within the 100 by 100 mile (Table XIII and Figure 12) area and alarmingly higher at 66.05% within the El Dorado fire region itself (Table XII and Figure 11). Post-fire assessments indicated a predictable increase in Cheatgrass across the burnt landscape, while fuel loads began to normalize within the fire-affected region from 2020 to 2023 (Table XII and Figure 13). This analysis not only provides insight into the immediate effects of wildfires on local vegetation dynamics but also opens avenues for further research into historical wildfire patterns and the complex interactions between wildfire occurrences and the availability and connectivity of combustible materials.

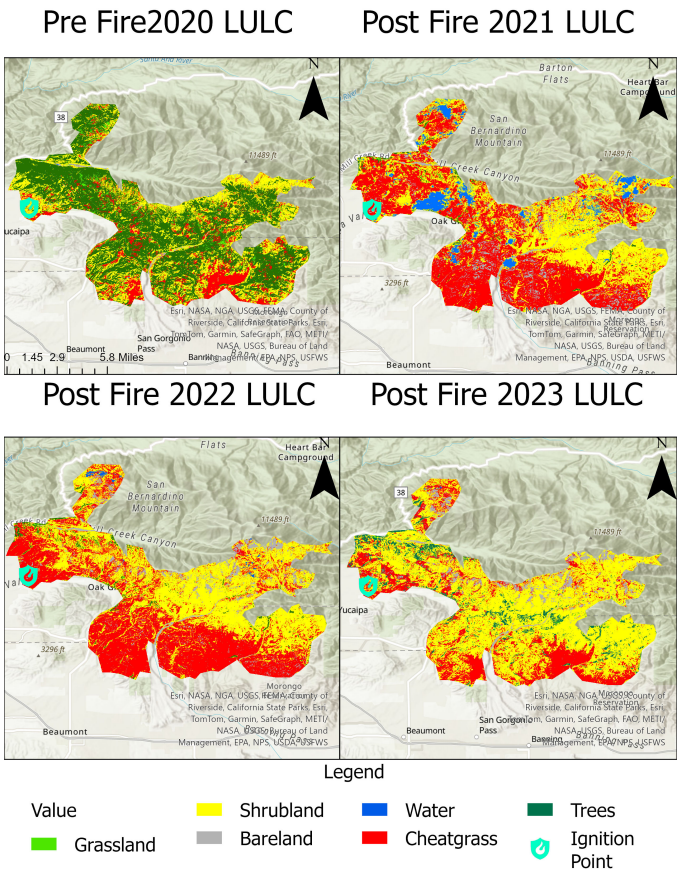

Fig. 13: Pre and Post Land Cover Change in El Dorado Fire 2020

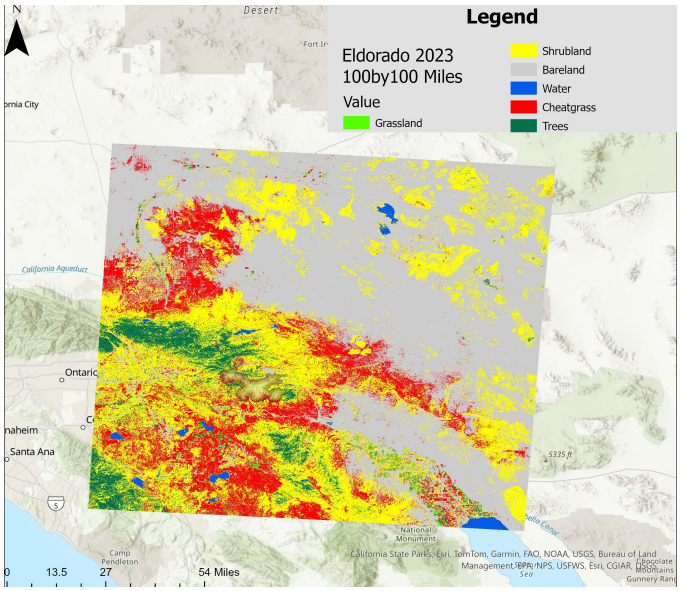

Fig. 14: LULC Map of 2023 around 100 by 100 miles of El Dorado Fire 2020.

E. Example use case 2: Corral Fire 2024

Our model identified a notable presence of Cheatgrass from 2020 through 2023 (Figure 15), suggesting a consistent accumulation of fine fuels over this period. On June 1, 2024, driven by sustained

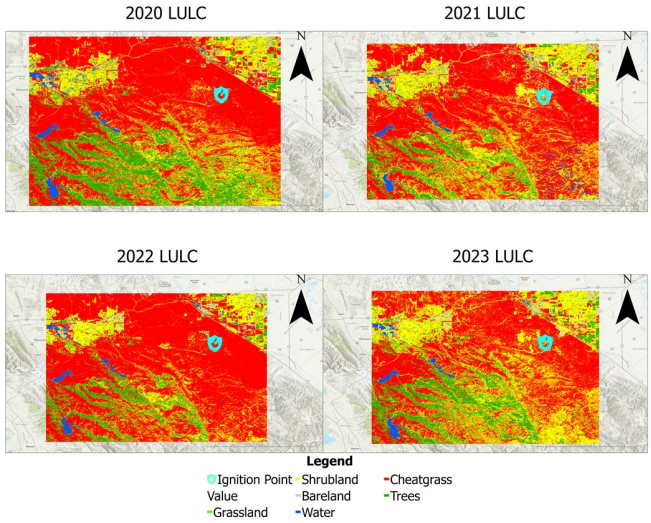

Fig. 15: Historical four-year LULC around Corral Fire 2024 area showing huge density of cheatgrass around the Ignition point.

winds averaging 32.7 mph and gusts reaching up to 43.1 mph, the Corral Fire rapidly spread across nearly 15,000 acres, causing significant damage to infrastructure, interrupting operations, and affecting offsite properties [61]. As shown in Figure 15, the dense presence of Cheatgrass near the ignition point underscores the critical value of integrating both current and historical Land Use Land Cover (LULC) maps that explicitly account for invasive grasses like Cheatgrass. Such data products are essential for pre-fire risk assessments and for implementing effective wildfire mitigation strategies.

IV. Conclusion

This study presents a comprehensive and scalable framework for detecting and mapping Cheatgrass (Bromus tectorum)—a highly flammable invasive species linked to increased wildfire frequency in California. We leveraged multi-temporal Sentinel-2 imagery, Vegetation Index Temporal Statistical Bands, and machine learning to achieve high model performance, with an overall accuracy of 91.1%, 91% precision for Cheatgrass, an ROC-AUC of 0.99, and a Cohen's Kappa of 0.89. These results affirm the robustness of our approach in distinguishing Cheatgrass from other land cover types across varied ecological settings.

One of the primary challenges in land use and land cover (LULC) classification is the acquisition of reliable ground truth data across large and diverse regions. To address this, we utilized the open-source Calflora dataset, which provides fine-scale species-level observations. Recognizing spatial inaccuracies within user-reported data, we developed a semi-supervised label refinement method using a binary MLP classifier. This approach significantly improved label quality while reducing manual labeling requirements, making the methodology more scalable and suitable for dynamic landscapes.

A key contribution of our framework is the integration of Vegetation Index statistical features—derived from monthly NDVI, GNDVI, and MSAVI2 metrics (mean, standard deviation, skewness, kurtosis)—which capture phenological patterns unique to Cheat-

grass. This enabled accurate early-season mapping using data up to June, ahead of California's peak wildfire season. Our model's ability to generate LULC maps and Cheatgrass probability layers supports pre-season fuel load assessments, identification of high-risk zones, and targeted mitigation strategies.

Through case studies such as the El Dorado and Corral Fires, we demonstrated how persistent Cheatgrass presence contributes to fuel accumulation and rapid fire spread. The predictive maps generated by our model can thus serve as valuable tools for wildfire preparedness, ecological restoration, and invasive species monitoring.

In summary, this study highlights the practical potential of combining remote sensing, semi-supervised learning, and open ecological datasets to inform proactive wildfire risk management and land-use planning in fire-prone regions.

ACKNOWLEDGMENT

The authors acknowledge the support of the University of Florida and Fire Neural Network. We also acknowledge the University of Florida for the Hipergator resources used in this study.

DATA AVAILABILITY

The data is available on request to authors.

APPENDIX

A. Binary Multilayer Perceptron (MLP) Classifier

Data for Cheatgrass (Bromus tectorum) were extracted from the Calflora database [26], covering the period from January 1, 2016, to January 10, 2023, resulting in 1,720 initial points. To expand the dataset, we generated four random buffer points within a 200-meter radius around each original point. The combined dataset was then input into a binary MLP classifier. We evaluated probability thresholds ranging from 0.4 to 0.7, with a threshold of 0.5 yielding 1,629 positive samples, closely aligning with the initial Calflora points(Table XV. These selected points underwent visual validation using high-resolution imagery in ArcGIS Pro [35]. Table XIV provides the classification report for the binary MLP model. Figure 16 represents loss and accuracy curves of train and test. For additional classes (trees, shrubland, bareland, water, and grassland), stratified random samples were drawn from a dynamic LULC map of Los Angeles County and refined through a similar binary MLP classifier.

TABLE XIV: Classification Report for Cheatgrass Detection

Class	Precision	Recall	F1-Score	Support
Non-Cheatgrass	1.00	0.99	1.00	108
Cheatgrass	0.96	1.00	0.98	22
Accuracy				0.99
Macro Average	0.98	1.00	0.99	130
Weighted Avg	0.99	0.99	0.99	130

B. Mojave Desert

The Mojave Desert landscape is primarily characterized by bare land and shrubland [62]; however, instances of Cheatgrass (Bromus tectorum) emergence have also been observed following fire events [63]. Our LULC map effectively captures this dynamic, illustrating our model's robustness in mapping diverse terrain and land cover patterns(Figure 17).

TABLE XV: Cheatgrass Sample Selection and MLP Processing Details

Description	Count
Cheatgrass points from Calflora (CA)	1,720
Random buffer points (200 m radius)	6,880
Total points used in MLP	8,600
Positive samples (probability $>$ 0.4)	2,108
Positive samples (probability $>$ 0.5)	1,629
Positive samples (probability $>$ 0.6)	1,196
Positive samples (probability $>$ 0.7)	712

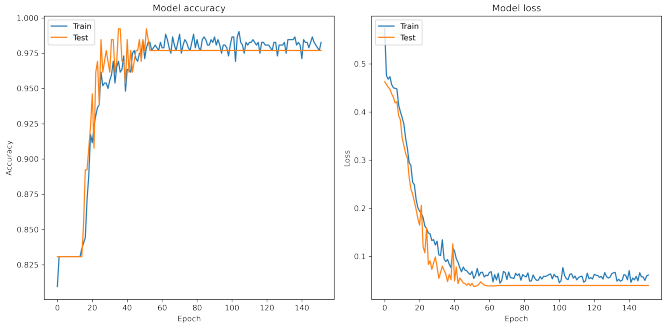

Fig. 16: Loss and Accuracy curve of MLP Classifier.

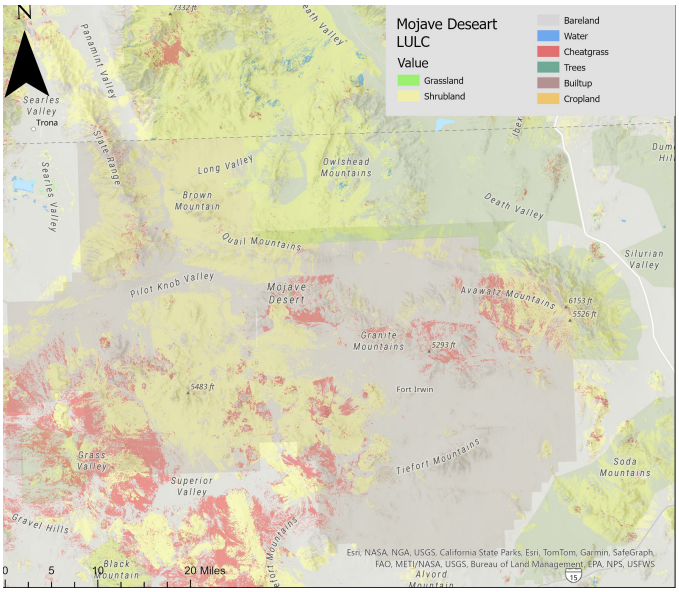

Fig. 17: Mojave Desert 2024 LULC.

REFERENCES

1 Duguy, B., Paula, S., Pausas, J. G., Alloza, J. A., Gimeno, T., and Vallejo, V. R., "Effects of climate and extreme events on wildfire regime and their ecological impacts," in *Advances in Global Change Research*. Berlin, Germany: Springer Science and Business Media LLC, 2012, vol. 51, pp. 101–134, [Google Scholar].

2 Brotons, L., Aquilué, N., De Cáceres, M., Fortin, M.-J., and Fall, A., "How fire history, fire suppression practices and climate change affect wildfire regimes in mediterranean landscapes," *PLoS ONE*, vol. 8, p. e62392, 2013, [Google Scholar] [CrossRef] [Green Version].

3 Safford, H. D., Paulson, A. K., Steel, Z. L., Young, D. J. N., Wayman, R. B., and Varner, M., "The 2020 california fire season: a year like no other, a return

to the past or a harbinger of the future?" *Global Ecology and Biogeography*, vol. 31, pp. 2005–2025, 2022.

4 Meng, Y.-Y., Yu, Y., Garcia-Gonzales, D., Al-Hamdan, M. Z., Marlier, M. E., Wilkins, J. L., Ponce, N., and Jerrett, M., "Health and economic cost estimates of short-term total and wildfire pm2.5 exposure on work loss: using the consecutive california health interview survey (chis) data 2015–2018," *BMJ Public Health*, vol. 2, p. e000491, 2024.

5 Holden, Z. A., Swanson, A., Luce, C. H., Jolly, W. M., Maneta, M., Oyler, J. W., Warren, D. A., Parsons, R., and Affleck, D., "Decreasing fire season precipitation increased recent western us forest wildfire activity," *Proceedings of the National Academy of Sciences*, vol. 115, no. 36, pp. E8349–E8357, 2018. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1802316115

6 Abouali, A., Viegas, D. X., and Raposo, J. R., "Analysis of the wind flow and fire spread dynamics over a sloped–ridgeline hill," *Combustion and Flame*, vol. 234, p. 111724, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010218021004673

7 Freeman, E. D., Sharp, T. R., Larsen, R. T., Knight, R. N., Slater, S. J., and McMillan, B. R., "Negative effects of an exotic grass invasion on small-mammal communities," *PLoS One*, vol. 9, no. 9, p. e108843, 2014.

8 Holbrook, J. D., Arkle, R. S., Rachlow, J. L., Vierling, K. T., Pilliod, D. S., and Wiest, M. M., "Occupancy and abundance of predator and prey: implications of the fire-cheatgrass cycle in sagebrush ecosystems," *Ecosphere*, vol. 7, no. 6, p. e01307, 2016.

9 Crawford, J. A., Olson, R. A., West, N. E., Mosley, J. C., Schroeder, M. A., Whitson, T. D., Miller, R. F., Gregg, M. A., and Boyd, C. S., "Ecology and management of sage-grouse and sage-grouse habitat," *Journal of Range Management*, vol. 57, no. 1, pp. 2–19, 2004.

10 Lockyer, Z. B., Coates, P. S., Casazza, M. L., Espinosa, S., and Delehanty, D. J., "Nest-site selection and reproductive success of greater sage-grouse in a fire-affected habitat of northwestern nevada," *The Journal of Wildlife Management*, vol. 79, no. 5, pp. 785–797, 2015.

11 Balch, J. K., Bradley, B. A., D'Antonio, C. M., and Gómez-Dans, J., "Introduced annual grass increases regional fire activity across the arid western usa (1980–2009)," *Global change biology*, vol. 19, no. 1, pp. 173–183, 2013.

12 Pilliod, D. S., Welty, J. L., and Arkle, R. S., "Refining the cheatgrass–fire cycle in the great basin: Precipitation timing and fine fuel composition predict wildfire trends," *Ecology and Evolution*, vol. 7, no. 19, pp. 8126–8151, 2017.

13 Davies, K. W. and Nafus, A. M., "Exotic annual grass invasion alters fuel amounts, continuity and moisture content," *International journal of wildland fire*, vol. 22, no. 3, pp. 353–358, 2012.

14 Langford, Z. L., Kumar, J., Hoffman, F. M., Breen, A. L., and Iversen, C. M., "Arctic vegetation mapping using unsupervised training datasets and convolutional neural networks," *Remote Sensing*, vol. 11, no. 1, p. 69, 2019.

15 Cao, J., Leng, W., Liu, K., Liu, L., He, Z., and Zhu, Y., "Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models," *Remote Sensing*, vol. 10, no. 1, p. 89, 2018.

16 Pérez-Rodríguez, L. A., Quintano, C., Marcos, E., Suarez-Seoane, S., Calvo, L., and Fernández-Manso, A., "Evaluation of prescribed fires from unmanned aerial vehicles (uavs) imagery and machine learning algorithms," *Remote Sensing*, vol. 12, no. 8, p. 1295, 2020.

17 Sankey, T., Donager, J., McVay, J., and Sankey, J. B., "Uav lidar and hyperspectral fusion for forest monitoring in the southwestern usa," *Remote Sensing of Environment*, vol. 195, pp. 30–43, 2017.

18 Wolkovich, E. M. and Cleland, E. E., "The phenology of plant invasions: a community ecology perspective," *Frontiers in Ecology and the Environment*, vol. 9, no. 5, pp. 287–294, 2011.

19 Bradley, B. A., "Remote detection of invasive plants: a review of spectral, textural and phenological approaches," *Biological invasions*, vol. 16, pp. 1411–1425, 2014.

20 Peterson, E., "Estimating cover of an invasive grass (bromus tectorum) using tobit regression and phenology derived from two dates of landsat etm+ data," *International Journal of Remote Sensing*, vol. 26, no. 12, pp. 2491–2507, 2005.

21 Phiri, D. and Morgenroth, J., "Developments in landsat land cover classification methods: A review," *Remote Sensing*, vol. 9, no. 9, p. 967, 2017. [Online]. Available: https://doi.org/10.3390/rs9090967

22 Veisimae, S., Kask, K., Poska, A., Rivis, R., Sepp, M., and Reisner, T., "Detecting invasive shrub species using sentinel-2 time-series data: case study of bramble (rubus fruticosus)," *Nature Conservation*, vol. 47, pp. 37–52, 2022. [Online]. Available: https://natureconservation.pensoft.net/article/83624/

23 European Space Agency (ESA), "Sentinel-2 data," 2024, accessed: 2024-07-05. [Online]. Available: https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2

24 Inglada, J., Vincent, A., Arias, M., and Tardy, D., "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sensing*, vol. 9, no. 1, p. 95, 2017. [Online]. Available: https://www.mdpi.com/2072-4292/9/1/95

25 Stanimirova, R., Tarrio, K., Turlej, K., McAvoy, K., Stonebrook, S., and et al., "A global land cover training dataset from 1984 to 2020," *Scientific Data*, vol. 10, p. 879, 2023. [Online]. Available: https://doi.org/10.1038/s41597-023-02798-5

26 Calflora, "Calflora: Information on california plants for education, research and conservation," https://www.calflora.org/, 2024, accessed: 2024-01-10.

27 United States Department of Agriculture, "Fuelscape data for wildfire risk assessment in the sagebrush biome of the western united states," 2024, accessed: 2024-05-31. [Online]. Available: https://www.fs.usda.gov/rds/archive/catalog/RDS-2024-0004

28 Pastick, N. J., Dahal, D., Wylie, B. K., Parajuli, S., Boyte, S. P., and Wu, Z., "Characterizing land surface phenology and exotic annual grasses in dryland ecosystems using landsat and sentinel-2 data in harmony," *Remote Sensing*, vol. 12, no. 4, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/4/725

29 Department, S. B. C. F., "Santa barbara county fire department official website," 2024, accessed: October 22, 2024. [Online]. Available: https://sbcfire.com/

30 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R., "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sensing of Environment*, vol. 202, pp. 18–27, 2017.

31 Dodgson, N. A., "Quadratic interpolation for image resampling," *IEEE transactions on image processing*, vol. 6, no. 9, pp. 1322–1326, 1997.

32 names as per the article (replace this placeholder), A., "Title of the article (replace this placeholder)," *Remote Sensing*, vol. 13, no. 8, p. 1494, 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/8/1494

33 Google Earth Engine Developers, "COPERNICUS_S2_SR Bands," Google Earth Engine Dataset Catalog, 2024, accessed: 2024-03-22. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR\#bands

34 Spatial Thoughts, "Extracting time series using google earth engine," Spatial Thoughts, 4 2020, accessed: 2024-03-22. [Online]. Available: https://spatialthoughts.com/2020/04/13/extracting-time-series-ee/

35 Esri, "Arcgis pro," https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview, 2024, accessed: 2024-06-13.

36 Britannica, E., "Los angeles: Landscape," n.d., accessed: 2024-10-09. [Online]. Available: https://www.britannica.com/place/Los-Angeles-California/Landscape

37 Engine, G. E., "Dynamic world v1 land cover," https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_DYNAMICWORLD_V1, 2022, accessed: 2024-06-13.

38 Lohr, S. L., *Sampling: Design and Analysis*, 2nd ed. Brooks/Cole, Cengage Learning, 2010.

39 Bureau of Land Management, "The great basin restoration initiative, a hand to nature: progress to date," 2001, available online at: www.fire.blm.gov/gbri/docs/gbri_progress_901.pdf (accessed 24 September 2007).

40 Bishop, C. M., *Pattern Recognition and Machine Learning*. Springer, 2006.

41 iNaturalist, "Observations: Cheatgrass (bromus tectorum)," 2024, accessed: 2024-07-06. [Online]. Available: https://www.inaturalist.org/observations?place_id=any&subview=map&taxon_id=164056

42 Rouse, J., Haas, R., Schell, J., and Deering, D., "Monitoring vegetation systems in the great plains with erts," *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, vol. 1, pp. 309–317, 1974.

43 Gitelson, A. and Merzlyak, M., "Signature analysis of leaf reflectance spectra: Algorithm development for remote sensing of chlorophyll," *Journal of Plant Physiology*, vol. 148, no. 3-4, pp. 494–500, 1996.

44 Qi, J., Chehbouni, A., Huete, A., Kerr, Y., and Sorooshian, S., "A modified soil adjusted vegetation index (msavi)," *Remote Sensing of Environment*, vol. 48, no. 2, pp. 119–126, 1994.

45 Soft.Farm, "Vegetation indices – ndvi, evi, gndvi, cvi, true color," Online blog, 2025. [Online]. Available: https://www.soft.farm/en/blog/vegetation-indices-ndvi-evi-gndvi-cvi-true-color-140

46 Farmonaut, "Mastering satellite vegetation indices: Mndwi, msavi, and ndmi formulas for sentinel-2 imagery in precision agriculture," https://farmonaut.com/remote-sensing/mastering-satellite-vegetation-indices-mndwi-msavi-and-ndmi-formulas-for-sentinel-2-i 2025, [Online; accessed 23-June-2025].

47 Chen, T. and Guestrin, C., "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

48 Abdi, A. M., "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using sentinel-2 data," *GIScience & Remote Sensing*, vol. 57, no. 1, pp. 1–20, 2020.

49 Shiferaw, H., Bewket, W., and Eckert, S., "Performances of machine learning algorithms for mapping fractional cover of an invasive plant species in a dryland ecosystem," *Ecology and evolution*, vol. 9, no. 5, pp. 2562–2574, 2019.

50 Man, C. D., Nguyen, T. T., Bui, H. Q., Lasko, K., and Nguyen, T. N. T., "Improvement of land-cover classification over frequently cloud-covered areas using landsat 8 time-series composites and an ensemble of supervised classifiers," *International Journal of Remote Sensing*, vol. 39, no. 4, pp. 1243–1255, 2018.

51 Breiman, L., "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

52 Hunter, F., Mitchard, E., Tyrrell, P., and Russell, S., "Inter-seasonal time series imagery enhances classification accuracy of grazing resource and land degradation maps in a savanna ecosystem," *Remote Sens.*, vol. 12, p. 198, 2020, [CrossRef] [Green Version]. [Online]. Available: https://www.mdpi.com/2072-4292/12/1/198

53 Bradter, U., O'Connell, J., Kunin, W., Boffey, C., Ellis, R., and Benton, T., "Classifying grass-dominated habitats from remotely sensed data: The influence of spectral resolution, acquisition time and the vegetation classification system on accuracy and thematic resolution," *Science of The Total Environment*, vol. 711, p. 134584, 2020, [CrossRef]. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0048969719345759

54 Strobl, C., Malley, J., and Tutz, G., "An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological methods*, vol. 14, no. 4, p. 323, 2009.

55 Wikipedia contributors, "El dorado fire," 2024, accessed: 2024-05-31. [Online]. Available: https://en.wikipedia.org/wiki/El_Dorado_Fire

56 Cal Fire, "Corral fire incident updates," https://www.fire.ca.gov/incidents/2024/6/1/corral-fire/updates, 2024, accessed: 2024-06-13.

57 Tri-Valley CAREs (with LLNL Causal Analysis Team), "Root cause analysis and emd timeline for the corral fire (site 300)," Online PDF report, Mar. 2025, "Causal Report–Corral Fire and EMD Timeline 9-6-2024.". [Online]. Available: https://trivalleycares.org/wp-content/uploads/2025/03/Causal_Report-Corral-Fire-and-EMD-TImeline-9-6-2024.pdf

58 National Interagency Fire Center, "Incident management situation report," June 2024, accessed: 2024-06-14. [Online]. Available: https://www.nifc.gov/sites/default/files/NICC/1-Incident%20Information/IMSR/2024/June/IMSR_CY24_06062024_0.pdf

59 European Commission, Joint Research Centre, "Jrc ghsl p2023a ghs built-up grid," https://developers.google.com/earth-engine/datasets/catalog/JRC_GHSL_P2023A_GHS_BUILT_C, accessed: 2024-07-07.

60 U.S. Department of Agriculture, National Agricultural Statistics Service, "Usda nass cropland data layer," https://developers.google.com/earth-engine/datasets/catalog/USDA_NASS_CDL, accessed: 2024-07-07.

61 Raj, A. (2025, Mar. 19) Tri-valley cares uncovers cause of corral fire. Tri-Valley CAREs. Published March 19, 2025; root-cause analysis by LLNL attributes the fire to electrical failure at high-voltage pole 8009 :contentReference[oaicite:1]index=1. [Online]. Available: https://trivalleycares.org/2025/tri-valley-cares-uncovers-cause-of-corral-fire

62 Author(s), "Title of the document," Year, retrieved from CiteSeerX. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6eb2aebd61f83cd3aaff8e68878af721913ef911

63 United States Geological Survey (USGS), "Invasive plant cover in the mojave desert, 2009-2013, ver 2.0, april 2021," 2021, accessed: [insert date you accessed this]. [Online]. Available: https://www.usgs.gov/data/invasive-plant-cover-mojave-desert-2009-2013-ver-20-april-2021