Sensors and Materials, Vol. 34, No. 12 (2022) 4599–4614
MYU Tokyo

S & M 3131

Forest Burn Severity Mapping Using Multispectral Unmanned Aerial Vehicle Images and Light Detection and Ranging (LiDAR) Data: Comparison of Maximum Likelihood, Spectral Angle Mapper, and U-Net Classifiers

Boknam Lee,[1] Bomi Kim,[2] Choongshik Woo,[3]
Geonhwi Jung,[4] Gyeongwon Kwon,[4] and Joowon Park[4*]

[1]Human Resources Development Center for Big Data-based Glocal Forest Science 4.0 Professionals, Kyungpook National University, Daegu 41566, Rep. of Korea
[2]Chungnam Forest Environment Research Institute, Sejong 30085, Rep. of Korea
[3]Division of Forest Disaster Management, National Institute of Forest Science, Seoul 02455, Rep. of Korea
[4]Department of Forestry, Kyungpook National University, Daegu 41566, Rep. of Korea

(Received September 30, 2022; accepted November 24, 2022)

Keywords: forest burn severity mapping, carbon emission, maximum likelihood, spectral angle mapper, U-Net

The automated mapping of forest burn severity using remote sensing imagery has been popular over the last decade. However, there is a lack of studies examining the performance of a range of classifiers for forest burn severity mapping for different burn severity classes. In this study, the performance of three supervised classifiers, maximum likelihood (ML), spectral angle mapper (SAM), and deep learning (U-Net), was evaluated for mapping forest burn severity under different burn severity class settings (two-level burn severity classes: burned and unburned; four-level burn severity classes: crown fire, heat-damaged, ground fire, and unburned). Multispectral unmanned aerial vehicle (UAV) images and light detection and ranging (LiDAR) points obtained from forest fire areas of Andong in South Korea were used to evaluate burn severity. The results show that all classifiers were capable of mapping the two-level burn severity with high overall accuracy (OA) (SAM: OA = 92.05%, kappa coefficient (K) = 0.84; U-Net: OA = 91.83%, K = 0.83; ML: OA = 90.92%, K = 0.82). For four-level burn severity mapping, U-Net (OA = 79.23%, K = 0.64) outperformed the conventional classifiers of SAM (OA = 50.61%, K = 0.38) and ML (OA = 46.85%, K = 0.34). Regarding class separability, SAM and U-Net showed good performance in detecting the severe burn severity class (crown fire areas), whereas a high rate of misclassification occurred in identifying the moderate burn severity classes (heat-damaged, ground fire) for all classifiers. In particular, ML and SAM showed a low capability in identifying unburned areas, while U-Net showed the lowest capability in mapping heat-damaged and ground fire areas. Overall, our study demonstrated that the reliable mapping of burn severity for Korea's forest fires largely depends on the number of levels of burn severity classes as well as the classifier's capability in discriminating moderate burn severity classes.

*Corresponding author: e-mail: joowon72@knu.ac.kr
https://doi.org/10.18494/SAM4159

ISSN 0914-4935
https://myukk.org/

1. Introduction

Widespread and recurrent severe wildfires have been a severe problem around the world, ranging from western North America and Europe to the Amazon and most recently in eastern Australia.[1] Wildfires cause severe and chronic damage to ecosystems and human properties. For example, wildfires are a major driver of greenhouse gas emissions, causing changes in terrestrial carbon stocks and altering ecological processes.[2–4] The property damage and economic loss from forest fires have also greatly increased over the past decades. The identification of burn areas and burn severity via burn severity mapping is thus important to investigate the impacts of fire on forests and to implement timely restoration and mitigation activities.[5–7]

To assist broad-scale burn severity estimation, remote sensing techniques such as satellite and unmanned aerial vehicle (UAV) platforms have been widely used to map the extent and variability of burned areas through the use of different spectral bands or derived reflectance indices.[8–10] The remote sensing-based identification and mapping of burn severity is based on unique spectral variations of vegetation and burn residuals caused by fires. Once vegetation is burned, the decreases in the chlorophyll and moisture contents of plant species lead to decreases in visible and near-infrared (NIR) reflectances and increases in mid-infrared (SWIR) reflectance.[11,12] In this manner, the degree of burn severity can be captured as a variation in the spectral index of a remote-based optical sensor. The widely used spectral indices are the normalized difference vegetation index (NDVI), normalized burn ratio (NBR), and burned area index (BAI) since NIR and SWIR bands are sensitive to forest fire damage.[13–19]

Supervised classification algorithms have been widely used to map burn severity.[20,21] In supervised classification, the algorithm is trained using ground-sampled data to predict the defined burn severity classes such as high burned, low burned, and unburned. For example, Amos *et al.* used the spectral angle mapper (SAM) classifier to assess burn severity levels using a collection of spectra from selected regions of interest using a Sentinel-2A image.[13] Anggraeni and Lin used the SAM and support vector machine (SVM) algorithms for burned area mapping from a Landsat TM image and obtained accuracies of 83 and 97% for the classifiers, respectively.[22] Since various supervised classification methods identify image pixels and classify them in categories or classes according to their spectral information, several burn severity classes are comprehensively required to represent the variations of the images. However, the large variations of classification across supervised classification methods and levels of burn severity classes result in inconsistent classification accuracies of burn severity mapping.[23] It is thus important to find an appropriate classifier for a more accurate burn severity mapping to prevent further misclassification of areas or blind spots for post-fire forest restoration and mitigation practices.

Furthermore, while the classical supervised classifiers provide a reasonable accuracy, it is still difficult to characterize subtle levels of burn severity owing to spectral similarities between burn severity classes.[24–26] For example, Woo *et al.* and Shin *et al.* observed limited discrimination between slightly burned and unburned vegetation in burn severity mapping with the maximum likelihood (ML) and SAM classifiers.[27,28] Gibson *et al.* found a relatively low

accuracy in identifying a moderate burn severity class compared with unburned and extreme burn severity classes using a random forest classifier.[21] Therefore, it is important to extend our understanding to a more advanced approach, such as a deep learning classifier (i.e., U-Net), and to identify the feasibility, capability, and accuracy of a deep learning classifier compared with conventional supervised classifiers for mapping burn severity levels.[29]

In Korea, there have been few studies examining the performance of a deep learning approach to map burn severity in comparison with conventional supervised classifiers. The majority of burn severity studies were limited to classical supervised classifiers, such as ML and SAM, without the consideration of the effect of the number of burn severity classes on classification accuracy.[27,28] Therefore, it is necessary to broaden our understanding and identify a better combination of the classifier and burn severity level for a more reliable mapping of burn severity in Korea's forest fires.

To address this issue, the main objective of this study was to compare the capability of the conventional (i.e., ML and SAM) and deep learning (i.e., U-Net) classifiers for mapping burn severity. We also examined the effect of different burn severity classes on the performance of each classifier by testing two different burn severity class settings: 1) two-level burn severity classes (burned and unburned) and 2) four-level burn severity classes (crown fire, heat-damaged, ground fire, and unburned). To implement the above-mentioned tasks, we collected field-based data for each burn severity class combined with LiDAR-derived canopy boundary data to extract more pure training data within the forest fire areas of Andong in Korea, trained the classifiers with data from unmanned aerial vehicle (UAV) imagery, and evaluated the performance of the classifiers through accuracy assessments.

2. Materials and Methods

2.1 Study area

We investigated a forest fire that occurred in Namhu-myeon near the city of Andong in South Korea. The fire broke out on April 24, 2020 and damaged 2,025 ha out of the 3,797 ha forest in three days, which was the biggest forest fire since 2010 in South Korea (Fig. 1). The study area is located in the temperate forest zone (Lat: 36.31–36.33˚, Lon: 128.32–128.39˚). The forest vegetation in the study area mainly consisted of coniferous (92%), mixed conifer (5%), and deciduous (1%) trees. The dominant species was *Pinus densiflora*, occupying 74% of the area. The main classes of tree age were the fourth class (46%), third class (45%), fifth class (6%), and second class (2%). The average annual rainfall in this area was 101 mm and the average temperature was 12.6 ℃ in 2020.

2.2 Data

2.2.1 Multispectral UAV image

The image of forest fires was collected on April 28, 2020 with a NIFos-2 UAV, which has a mounted multispectral sensor consisting of five bands (Blue, Green, Red, Red Edge, and NIR).

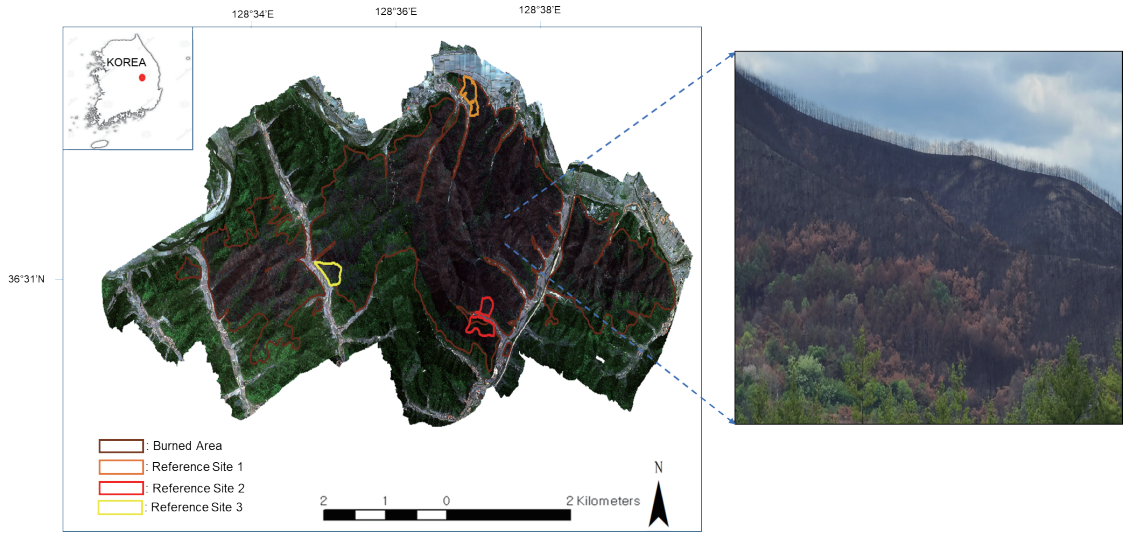

Fig. 1. (Color online) RedEdge multispectral UAV image and picture of area burned by forest fire in Namhu-myeon near the city of Andong in Korea. The brown line in the UAV image is the perimeter of the burned area and ground reference sites are presented with orange (site 1), red (site 2), and yellow (site 3) boundaries.

The acquired images consisted of 300–400 scenes for each band with 80% frontal overlap and 60% side overlap. The acquired UAV images were converted into orthographic images using Pix4D software (Pix4 S.A., Prilly, Switzerland). Atmospheric and radiating corrections were applied to the images using ENVI 5.5.3.[30] Figure 1 presents the RedEdge multispectral UAV images of the burned area in Andong. The specifications of the NIFos-2 UAV are an overall width of 1 m, a total length of 1 m, and a total height of 0.7 m. The spatial resolution is 30 cm at an altitude of 500 m. The maximum take-off weight, flight time, speed, and altitude are 10 kg, 30 min, 15 m/s, and 1000 m, respectively.

2.2.2 Forest fire reference data

The ground reference map of forest fires was provided by the National Institute of Forest Science (NIFS). The reference map was produced by the combination of four spectral indices obtained from UAV images and a field survey. Since there was no comprehensive forest fire information available, NIFS utilized the mean threshold values of four spectral indices, NDVI, RedEdge (RE) NDVI, visible-band difference vegetation index (VDVI), and modified chlorophyll absorption in reflectance index (MCARI), obtained from past forest fires in Korea to classify the burn severity of the forest fire area in Andong. Burn severity maps were then updated by matching the degree of burn severity through a field survey.

2.3 Definition of burn severity class

Burn severity is defined as the impact of fire on vegetation and soil, and it is usually expressed as low, medium, and high.[31] In Korea, burn severity is defined with five levels of

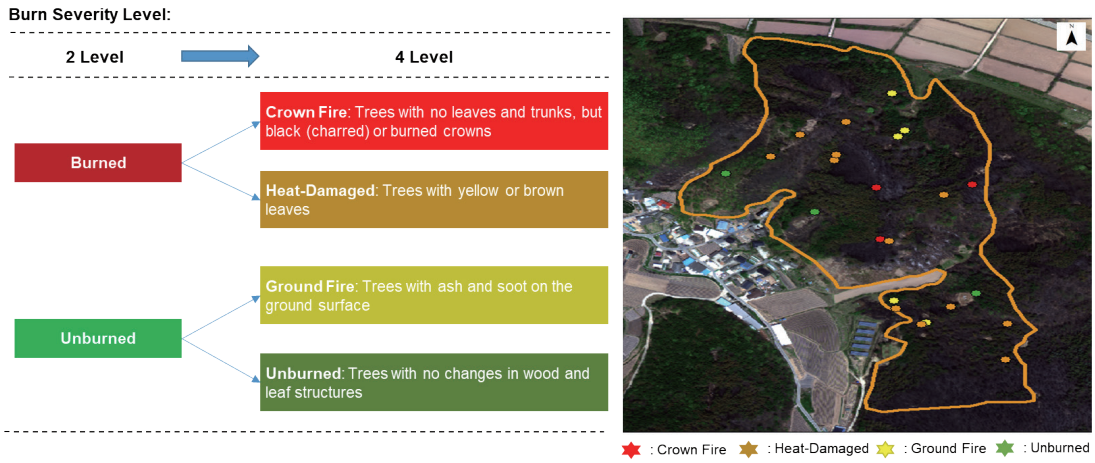

Fig. 2. (Color online) Definition of two different levels of burn severity classes used in this study. The photograph is a UAV image giving examples of the burn severity classes in the ground reference site (site 2).

burn severity according to the Korean composite burn index (KCBI): unburned, low, moderate, high, and extreme. Since there are ambiguous definitions among KCBI's burn severity classes, in this study, two different levels of burn severity classes were defined with the consideration of KCBI and field surveys. As shown in Fig. 2, two-level burn severity classes were categorized as unburned and burned, which were expanded to four-level burn severity classes, unburned, ground fire, heat-damaged, and crown fire, to represent changes in surface reflectance following the fire.

2.4 Spectral analysis

The main goal of this study was to evaluate and compare the performance characteristics of three classifiers, ML, SAM, and U-Net, for different burn severity levels. The workflow to analyze forest burn severity using UAV images comprised three main steps: input data preparation (pre-processing), burn severity mapping, and burn severity assessment (Fig. 3). In the first step, geometric, radiometric, and atmospheric corrections were applied to the UAV images of forest-burned areas using ENVI 5.5.3.[30] The three reference sites marked with orange, red, and yellow boundaries were then combined with a LiDAR-derived crown boundary to generate the training data (labeled images) using FUSION software.[32] In the second step, the burn severity classification was implemented by applying the ML, SAM, and U-Net classifiers under two different burn severity levels. The labeled images generated from the three reference sites of the burned area were converted to the corresponding burn severity classes and used to train classifiers. In the last step, an accuracy assessment was conducted to evaluate the performance of the three classifiers in terms of overall accuracy (OA), kappa coefficient (K), producer accuracy (PA), and user accuracy (UA) based on an error matrix.

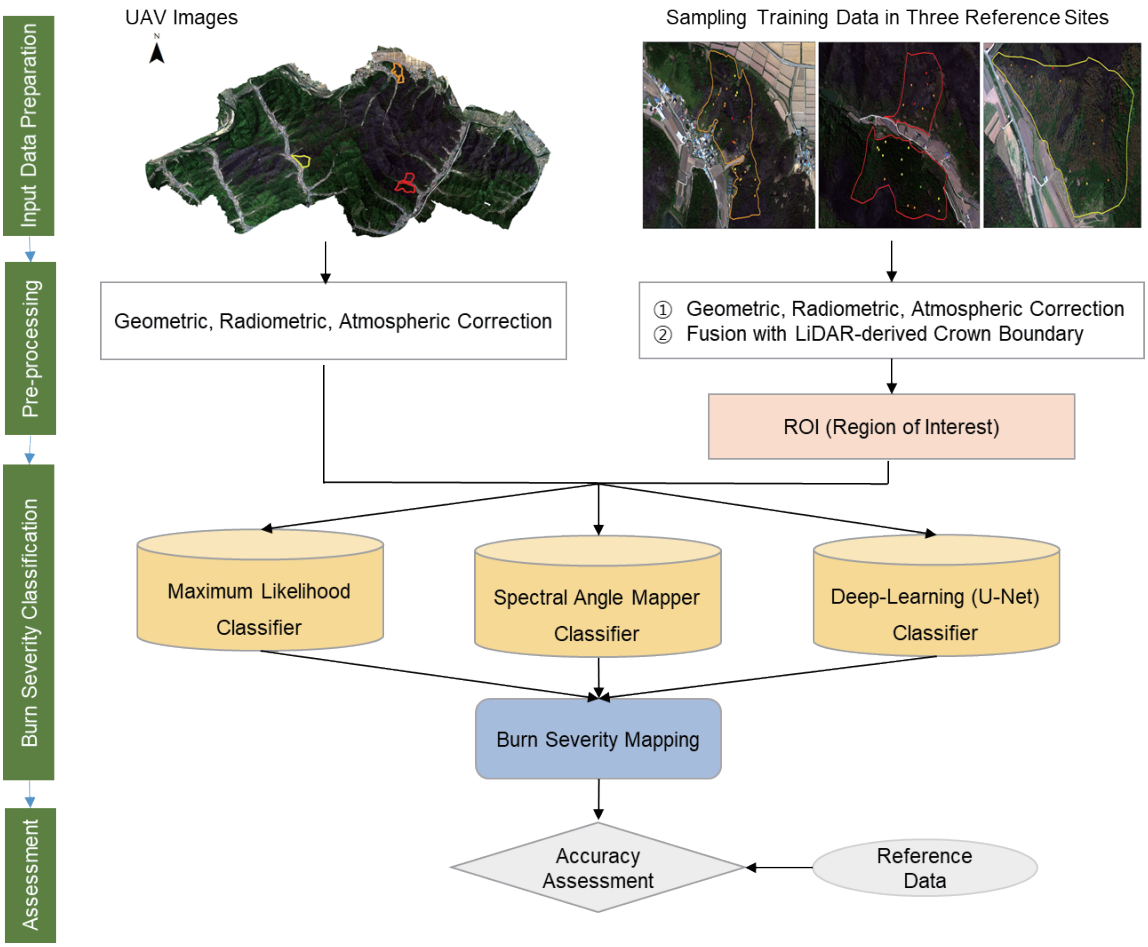

Fig. 3. (Color online) Overall flowchart of forest burn severity analysis.

2.4.1 Preparation of training data

In supervised classification, training samples are required for each burn severity class to classify the UAV image of a forest fire. In this study, the training data were prepared by combining the field survey data with LiDAR-derived crown boundary data. The field survey was conducted to collect the burn severity of individual trees within the circular sample plots in three reference sites (Fig. 1). The random sampling method was used to select the locations of the circular plots. The circular plot size was 0.01 ha with a radius of 5.65 m. The degrees of burn severity of 264, 169, and 79 individual trees were investigated from the 23, 20, and 12 circular plots located in reference sites 1, 2, and 3, respectively (Fig. 3). Next, the LiDAR-derived crown boundary was obtained to match the field-based burn severity class with individual trees. To extract the crown boundary of individual trees, first, high-density point cloud data (average point densities of 1231.98, 414.66, and 1123.98 pts/m^2 from sites 1, 2, and 3, respectively) were obtained by UAV-based LiDAR from July to August 2020. From the LiDAR point clouds, a digital terrain model (DTM) and a digital surface model (DSM) with a spatial resolution of 30

cm were produced by the GridSurfaceCreate algorithm in FUSION software.[32] The height of the canopy was then derived by a canopy height model (CHM), estimated as the difference between the top canopy surface (DSM) and the underlying ground topography (DTM). The canopy boundary was finally obtained by detecting treetops using local maxima and delineating individual tree crowns with the watershed algorithm scheme.[33,34]

For the ML and SAM classifications, a group of labeled regions of interest (ROIs), referring to the burn severity samples for each class, were obtained for the corresponding burn severity classes by matching the LiDAR-derived canopy boundary with the field survey data of individual trees. When different burn severity classes were located within a single canopy boundary, the severity data were excluded from the training data to obtain the most representative ROIs for each burn severity class (Fig. 4). The numbers of ROIs used for the two-level burn severity classification were 197 (26656 pixels) and 84 (11218 pixels) for the burned and unburned classes, respectively. For the four-level burn severity, 56 (6154 pixels), 141 (20502 pixels), 68 (9038 pixels), and 16 (2180 pixels) ROIs were used for the crown fire, heat-damaged, ground fire, and unburned classes, respectively.

For the deep learning (U-Net), 30 training samples were derived for each burn severity class using a group of ROIs used for the ML and SAM classifiers. Since the U-Net classifier requires at least 30 training samples for each burn severity class to train the model, 30 labeled images of 10 m circles were generated for the respective burn severity classes from the UAV images of each reference site through the visual assessment of burn severity intensity based on field surveys and were converted into the burn severity training raster.

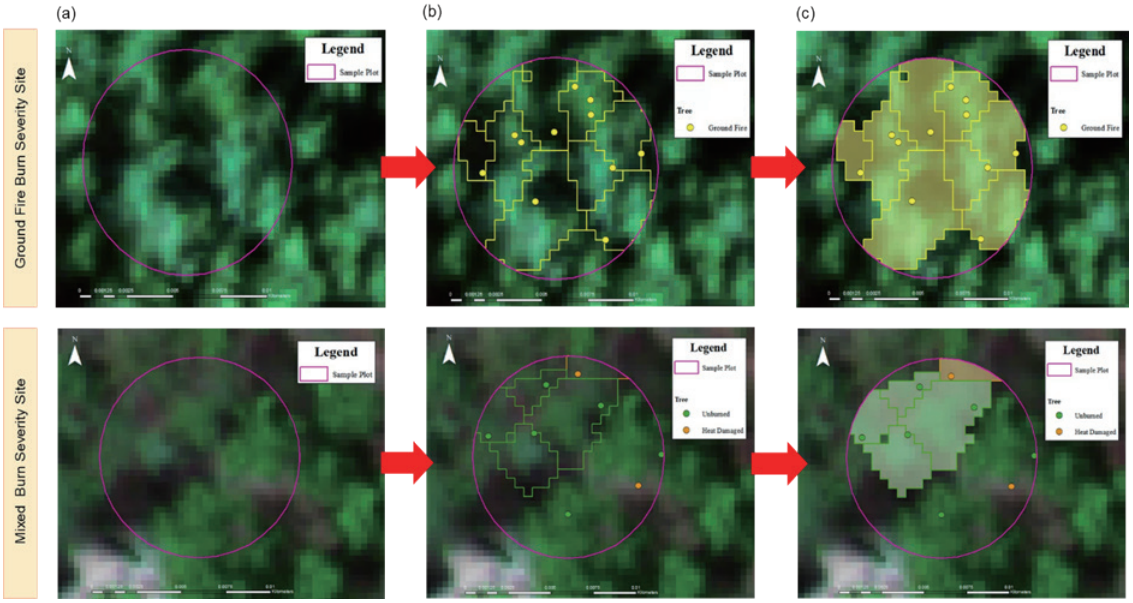

Fig. 4. (Color online) Process of generating the training data (ROIs) for burn severity classes (unburned, ground fire, heat-damaged, and crown fire) by combining the field survey data from reference sites with LiDAR-derived crown boundary data: (a) placing the boundary of a circular sampling plot on the UAV image, (b) extracting the LiDAR-derived canopy boundaries within the boundary of the circular sampling plot, and (c) generating ROI data for each burn severity class by matching the ground sampling data with LiDAR-derived crown boundaries.

2.4.2 Burn severity classification

The ML, SAM, and U-Net classifiers are supervised classification techniques based on the pixel-based classification algorithm. Descriptions of the ML, SAM, and U-N*et alg*orithms are given in the next section.

2.4.2.1 Maximum likelihood classifier

The ML classifier assumes that the statistics for each class in each band are represented by the normal probability distribution and allocates each pixel to a specific class according to the probability (i.e., ML). The probability of each pixel is analyzed by considering the normal density function from the mean, variance, and covariance of training samples. Once the signatures of each class are estimated using the available training data (ROIs), the class with the highest probability is allocated for each pixel by comparing the class probabilities.[35]

In this study, ENVI software was used to calculate the ML for each pixel using the following equation:

$$g_i(x) = 1np(\omega_i) - \tfrac{1}{2}1n|\Sigma_i| - \tfrac{1}{2}(x-m_i)^T \Sigma_i^{-1}(x-m_i), \quad (1)$$

where i is the class, x is the n-dimensional data, n is the number of bands, $p(w_i)$ is the probability that class w_i occurs in the image and is assumed the same for all classes, $|\Sigma_i|$ is the determinant of the covariance matrix of the data in class w_i, Σ_i^{-1} is its inverse matrix, and mi is the mean vector of each training class.[30]

2.4.2.2 SAM classifier

The SAM is a physically based spectral classification that utilizes the n-dimensional angle from the reference spectra to allocate each pixel.[36] The algorithm determines the spectral similarity by calculating the spectral angle between each pixel and the reference spectra in an n-dimensional space, where n is the number of bands. A pixel is assigned to a specific class according to the smallest spectral angle that represents a closer match to the reference class.

In this study, the mean of each burn severity class was considered as the reference spectra, estimated from the spectral reflectance values of the training data (ROIs) from the UAV images. The SAM function in ENVI software was used to estimate the cosine similarity for every pixel in each image relative to the reference spectra using the following equation:

$$\alpha = \cos^{-1}\left(\frac{\Sigma_{i=1}^{n} t_i r_i}{\left(\Sigma_{i=1}^{n} {t_i}^2\right)^{\frac{1}{2}} \left(\Sigma_{i=1}^{n} {r_i}^2\right)^{\frac{1}{2}}}\right), \quad (2)$$

where n is the number of multispectral bands, t_i is the unknown measurement vector, r_i is the reference spectrum vector, and α is an angle between vectors r and t.

2.4.2.3 Deep learning (U-Net) classifier

U-Net is convolutional network architecture that uses a semantic segmentation algorithm to classify each pixel of an image as belonging to a particular class. The U-Net semantic segmentation is accomplished using an improved fully convolutional network (FCN) structure, which is characterized by the encoder–decoder network structure. The encoder network extracts potential features from hierarchically given labeled input data through the processes of convolutional max-pooling, cropping, and concatenate operations, while the decoding network predicts a probability map at the pixel level by an upsampling operation.[37,38]

In this study, the U-Net classifier was operated using the ENVINet5 deep learning model in ENVI software, which adopts a TensorFlow model to perform deep learning tasks.[30] The TensorFlow model is an open-source library defined by an underlying set of convolutional neural network (CNN) parameters.[39] To predict the burn severity class for each pixel, the TensorFlow model is trained for specific burn severity features using a set of training label rasters, which are samples with different types of burn severity [e.g., two-level: 0 (unburned), 1 (burned); four-level: 0 (unburned), 1 (ground fire), 2 (heat-damaged), 3 (crown fire)], by providing a sliding-window patch around the pixel. Model training is performed to expose the training raster to the model repeatedly. As the training progresses, the model learns to convert the spatial and spectral information in the training raster into a class activation map, and it classifies the target burn severity in a pixel-by-pixel manner by looking for a similar match with a trained model.

ENVINet5 uses the binary cross-entropy loss function and adjusts the internal parameters or weights of the model to fit a U-Net model.[38] In this work, the U-Net model was trained using small patches of imagery from three reference sites of forest fires and then was used to predict the burn severity in the whole study area. The training dataset included three images, each with five channels (i.e., five bands), and the corresponding labels of objects. The patch size was set to 464 by 464 pixels. The numbers of epochs and patches per epoch were set to 30 and 4, respectively. The other parameters were the default values in ENVINet5.

2.4.3 Accuracy assessment

The accuracy of the three classifiers was evaluated using a confusion matrix (called an error matrix). The error matrix compares the relationship between the reference data (ground truth) and the corresponding results of a classification in terms of descriptive and statistical metrics, such as OA, K, PA, and UA.[40–42] In this study, error matrices were produced by assessing the correspondence between the reference data and the burn severity classification maps derived by the three classifiers. OA was calculated from the error matrices using the following equation and represents the total proportion of correctly mapped pixels among the reference data:

$$OA = \frac{\Sigma_{i=1}^{k} x_{ii}}{N}, \tag{3}$$

where N is the total number of reference pixels and x_{ii} is the number of correctly classified pixels.

K was obtained using the following equation and used to evaluate the viability of the classification:

$$K = \frac{N\Sigma_{i=1}^{r} x_{ii} - \Sigma_{i=1}^{r}\left(x_{i+} \times x_{+j}\right)}{N^2 - \Sigma_{i=1}^{r}\left(x_{i+} \times x_{+j}\right)}, \tag{4}$$

where N is the number of pixels in the error matrix, r is the number of rows and columns in the error matrix, x_{ii} is the major diagonal element for class i, x_{i+} is the total number of pixels in row i, and x_{+j} is the total number of pixels in column j.

PA (i.e., the error of omission) was obtained as the proportion of the reference data correctly classified for a certain burn severity class, while UA (i.e., the error of commission) was derived as the proportion of pixels assigned to an incorrect burn severity class.

3. Results and Discussion

3.1 Characteristics of spectral reflectance variation

The spectral reflectance curves of the training samples are presented for the two-level and four-level burn severity classes in Fig. 5. For the two-level burn severity classes, unburned showed a lower reflectance than burned, but both curves showed an increasing reflectance in the order of the Blue, Green, Red, RedEdge, and NIR bands. The large difference in reflectance between unburned and burned in the Green, Red, RedEdge, and NIR bands can be attributed to the high accuracy of discrimination of the two-level burn severity classes. For the four-level burn severity classes, unburned showed the highest reflectance in the Green, Red, RedEdge, and NIR bands, while ground fire had the lowest reflectance in the Blue, Green, and Red bands. In particular, the NIR band showed an increasing reflectance in the order of crown fire, heat-damaged, ground fire, and unburned classes. This can be attributed to the loss of chlorophyll content of the tree species in burned areas. Moreover, the largest spread of spectral reflectance among the burn severity classes was in the NIR band, suggesting that the NIR band is the most useful spectral wavelength for discriminating burn severity classes. This pattern is consistent with the findings of other studies that demonstrated the suitability of the NIR band for mapping burn severity.[43,44]

3.2 Comparison of burn severity classification performance characteristics

The performance of the three classifiers (ML, SAM, and U-Net) was examined by mapping burn severity under different burn severity class settings (two-level and four-level). Figure 6 displays the spatial distribution of burn severity class outputs. Compared with the ground truth label image, the three classifiers performed well in mapping the overall pattern of two-level burn

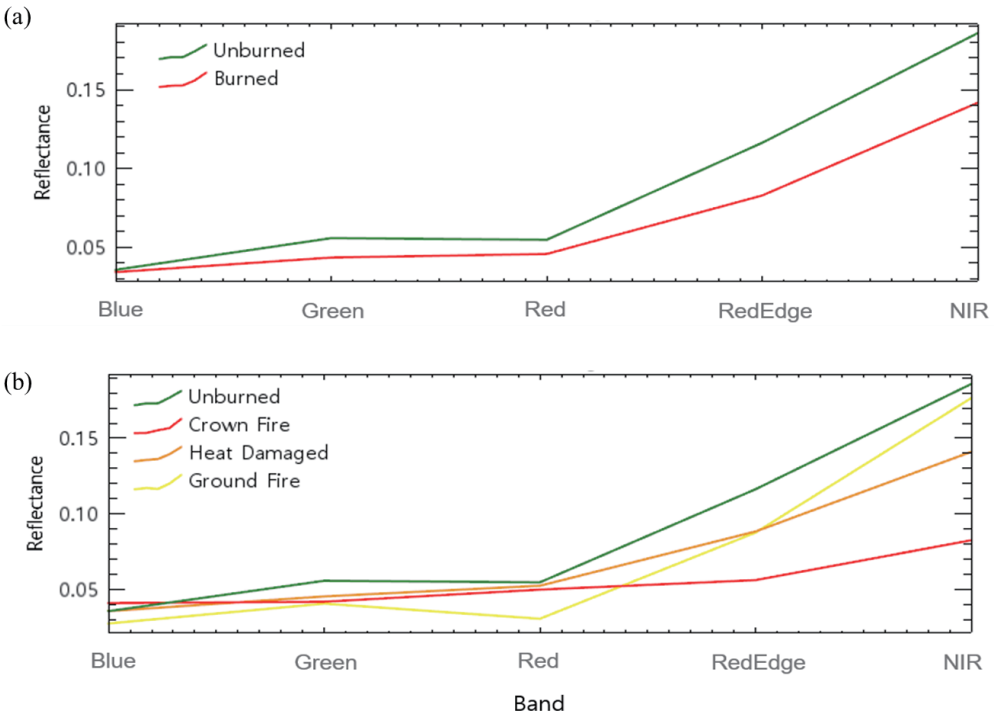

Fig. 5. (Color online) Mean spectral reflectance curves for burn severity classes: (a) reflectance curves for two-level burn severity classes and (b) reflectance curves for four-level burn severity classes.

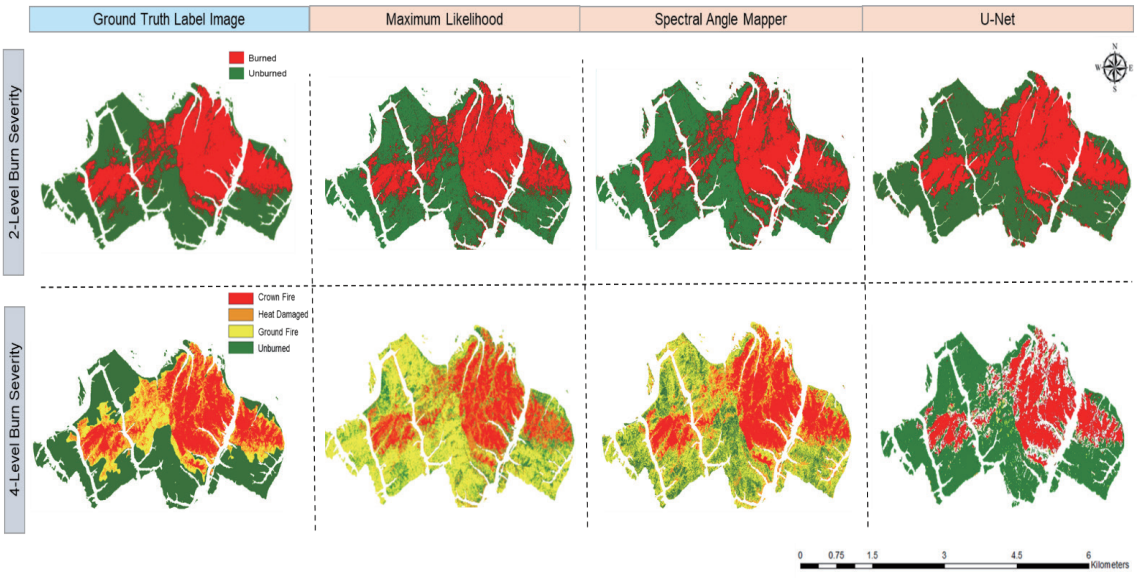

Fig. 6. (Color online) Burn severity maps produced by ML, SAM, and U-Net classifiers.

severity. However, it was found that all classifiers showed similar patterns of minor misclassification along the boundary of burned areas as well as at the boundary between forest and urban areas, where the ML classifier produced the highest rate of misclassification. It was also notable that U-Net produced the clearest and most seamless boundary of burned areas. This

result shows that deep learning-based boundary semantic segmentation could be more effective than the conventional classifiers for extracting the irregular pattern of burned areas.[37,45]

Regarding the four-level burn severity classification, U-Net showed a higher performance than ML and SAM. However, the overall patterns of four-level burn severity mapping showed a poor match with the ground truth image. Among the four-level burn severity classes, all classifiers produced a relatively accurate mapping for the high-severity class (i.e., crown fire), but there was poor discrimination between the low burn severity (i.e., ground fire) and unburned classes by all classifiers. In particular, ML and SAM misclassified most of the unburned areas as ground fire areas, while U-Net showed poor performance in matching ground fire areas, resulting in the over-prediction of the unburned areas. The high rates of misclassification of unburned areas with ML and SAM may be due to the low proportion of unburned training samples compared with other burn severity classes, which were insufficient to provide a full and representative distribution of the unburned class. For heat-damaged areas, ML and SAM showed a higher performance than U-Net where U-Net identified most of the heat-damaged areas as unclassified areas. This may be due to a particular type of texture or color of heat-damaged training samples being confused in U-Net owing to its low occurrence within the fire areas.

Overall, this result suggests the difficulty in discriminating the moderate and low burn severity classes from the other severity classes. This finding corresponds to that of Collins *et al.*[46] In particular, the higher rate of misclassification between the low burn severity (i.e., ground fire) and unburned classes likely reflects the necessity of improving the training datasets for these classes. Since the UAV multispectral image of a ground fire class appears as a mixture of green, yellow, or light black in the tree canopies, which may be mixed with unburned areas, there is a need for more representative training data that cover the diverse range of ground fire conditions. This would improve the quality of the training data, resulting in reduced errors in the algorithms of ML and SAM as well as strengthening the deep learning algorithm in U-Net. Moreover, an increased volume of training data for the unburned class could enhance the accuracy of ML and SAM. Similarly, learning with sufficient training data for the heat-damaged class could prevent unclassified pixels in the U-Net classification.

3.3 Comparison of burn severity classification accuracies

In the evaluation of the mapping accuracy of the three classifiers, confusion matrices were used to assess OA, K, PA, and UA for each classifier (Fig. 7 and Tables S1 and S2). Regarding the two-level burn severity mapping, all classifiers had OA of more than 90%, with SAM appearing to be the most accurate (OA = 92.05% and K = 0.84), followed by U-Net (OA = 91.83% and K = 0.83) and ML (OA = 90.92% and K = 0.82) [Fig. 7(a)]. This result implies that all classifiers are capable of mapping the binary burn severity classes.

For the four-level burn severity mapping, U-Net (OA = 79.23% and K = 0.64) outperformed SAM (OA = 50.61% and K = 0.38) and ML (OA = 46.85% and K = 0.34) [Fig. 7(a)]. In particular, U-Net showed the highest accuracy in matching crown fire (PA = 99.19% and UA = 88.26%) and unburned (PA = 97.71% and UA = 76.32%) but extremely poor performance in identifying both heat-damaged (PA = 3.96% and UA = 61.85%) and ground fire (PA = 5.3% and UA = 32.54%).

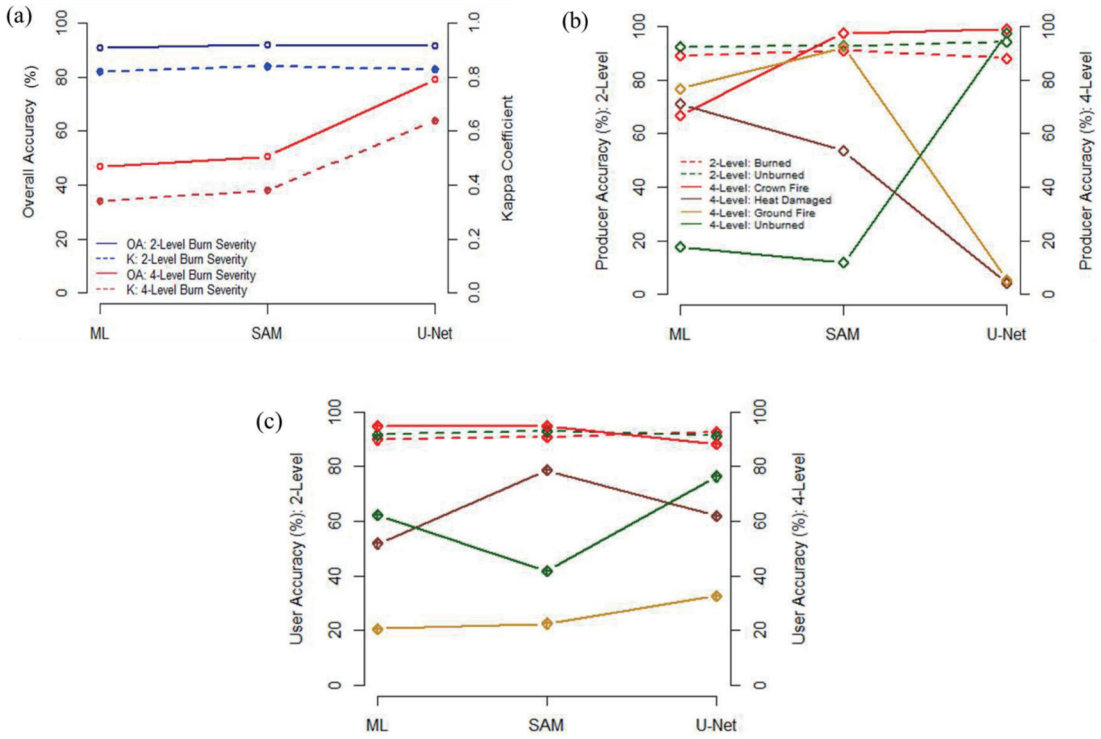

Fig. 7. (Color online) Comparison of burn severity classification accuracies for ML, SAM, and U-Net classifiers: (a) OA and K, (b) PA, and (c) UA for the two- and four-level burn severity classifications.

SAM also showed a high accuracy in identifying crown fire (PA = 97.65% and UA = 94.73%) but often misclassified heat-damaged (PA = 53.72% and UA = 78.56%) and unburned (PA = 11.62% and UA = 41.60%). ML showed poor performance across all burn severity classes, with PA values of less than 71%. In particular, the false negative mapping (omission error) for unburned was 82.26% [Figs. 7(b) and 7(c)]. The low classification accuracies of individual burn severity classes with SAM and ML are mainly due to their close spectral similarity in the UAV imagery. This indicates that the conventional pixel-based algorithms that only use spectral information of individual pixels without considering the correlation between adjacent pixels are insufficient for complex burn severity mapping.

Overall, the results of the accuracy assessment indicate that the accuracy of binary burn severity mapping is not significantly affected by the type of classifier used, whereas for a higher level of burn severity mapping, U-Net demonstrates the potential to overcome the limitation of conventional pixel-based approaches (ML and SAM) based on OA values. This finding indicates that U-Net, which uses a deep learning-based (CNN) semantic segmentation algorithm, is more capable of identifying complex patterns of burn severity classes than other classifiers. However, note that the U-Net's OA could be misleading as it is dominated by the false negative detection of heat-damaged and ground fire areas. This indicates that more reliable labeled data of heat-damaged and ground fire classes should be provided to enhance the performance of U-Net since the quality and authenticity of data are critical factors affecting the performance of deep learning-based classification.[47] Furthermore, note that all classifiers showed limited

performance in discriminating the moderate and low burn severity areas, in which the texture has very little contrast. This suggests that additional representative training data for these classes will improve the accuracy of all classifiers.

4. Conclusions

We examined the performance and accuracy of the ML, SAM, and U-Net classifiers for forest burn severity mapping using multispectral UAV imagery. The accuracy of burn severity mapping varied across the three classifiers as well as across the burn severity classes. ML showed the worst performance in terms of OA regardless of the burn severity level, while U-Net outperformed ML and SAM for high-level burn severity mapping. Concerning class separability, the high burn severity class (crown fire) attained a higher accuracy than the low burn severity classes (heat-damaged, ground fire) regardless of the classifier. Overall, our study revealed that the stability of burn severity mapping for Korea's forest fires is dependent on the level of burn severity classes as well as the classifier's capability in discriminating the moderate burn severity classes.

To further improve the accuracy of high-level burn severity mapping, it is recommended to provide more representative training data covering a diverse range of burn severity conditions to fit each classifier and reduce misclassification errors. Adopting longer wavelengths such as SWIR for UAV sensors could also enhance the accuracy of extracting complex burn severity classes by providing more diverse spectral information.

Acknowledgments

This study was supported by Korea Forest Service through the R&D Programs of the National Institute of Forest Science (Project No. FM0103-2022-02) and the Korea Forestry Promotion Institute (Project No. 2019149B10-2223-0301).

References

1 D. B. Lindenmayer and C. Taylor: PNAS **117** (2020) 12481. https://doi.org/10.1073/pnas.2002269117
2 A. Datta and R. Krishnamoorti: Front. Clim. **4** (2022) 1. https://doi.org/10.3389/fclim.2022.799632
3 H. Keith, D. B. Lindenmayer, B. G. Mackey, D. Blair, L. Carter, L. McBurney, S. Okada, and T. Konishi-Nagano: PLoS One **9** (2014) 1. https://doi.org/10.1371/journal.pone.0107126
4 S. G. Conard and A. M. Solomon: Dev. Environ. Sci. **8** (2008) 109. https://doi.org/10.1016/S1474-8177(08)00005-3
5 J. Franklin, J. M. Serra-Diaz, A. D. Syphard, and H. M. Regan: PNAS **113** (2016) 3725. https://doi.org/10.1073/pnas.1519911113
6 E. Sertel and U. Alganci: Geomatics, Nat. Hazards Risk **7** (2016) 1198. https://doi.org/10.1080/19475705.2015.1050608
7 A. Polychronaki and I. Z. Gitas: Remote Sens. **4** (2012) 424. https://doi.org/10.3390/rs4020424
8 A. A. Pereira, J. M. C. Pereira, R. Libonati, D. Oom, A. W. Setzer,F. Morelli, F. Machado-Silva, and L. M. D. Carvalho: Remote Sens. **9** (2017) 1. https://doi.org/10.3390/rs9111161
9 N. Pettorelli, W. F. Laurance, T. G. O'Brien, M. Wegmann, H. Nagendra, and W. Turner: J. Appl. Ecol. **51** (2014) 839. https://doi.org/10.1111/1365-2664.12261
10 J. E. Keeley: Int. J. Wildland Fire **18** (2009) 116. https://doi.org/10.1071/WF07049
11 S. Escuin, R. Navarro, and P. Fernández: Int. J. Remote Sens. **29** (2008) 1053. https://doi.org/10.1080/01431160701281072

12 J. D. White, K. C. Ryan, C. C. Key, and S. W. Running: Int. J. Wildland Fire **6** (1996) 125. https://doi.org/10.1071/WF9960125
13 C. Amos, G. P. Petropoulos, and K. P. Ferentinos: Int. J. Remote Sens. **40** (2019) 905. https://doi.org/10.1080/01431161.2018.1519284
14 F. Filipponi: Proc. 2018 2nd Int. Electronic Conf. Remote Sensing. (2018) 1–7.
15 B. N. Tran, M. A. Tanase, L. T. Bennett, and C. Aponte: Remote Sens. **10** (2018) 1. https://doi.org/10.3390/rs10111680
16 B. M. Collins, M. Kelly, J. W. van Wagtendonk, and S. L. Stephens: Landscape Ecol. **22** (2007) 545. https://doi.org/10.1007/s10980-006-9047-5
17 J. Epting, D. Verbyla, and B. Sorbel: Remote Sens. Environ. **96** (2005) 328. https://doi.org/10.1016/j.rse.2005.03.002
18 C. J. Chafer, M. Noonan, and E. Macnaught: Int. J. Wildland Fire **13** (2004) 227. https://doi.org/10.1071/WF03041
19 E. Chuvieco, M. P. Martín, and A. Palacios: Int. J. Remote Sens. **23** (2002) 5103. https://doi.org/10.1080/01431160210153129
20 D. Hamilton, K. Brothers, C. McCall, B. Gautier, and T. Shea: Remote Sens. **13** (2021) 1. https://doi.org/10.3390/rs13193843
21 R. Gibson, T. Danaher, W. Hehir, and L. Collins: Remote Sens. Environ. **240** (2020) 1. https://doi.org/10.1016/j.rse.2020.111702
22 A. Anggraeni and C. Lin: 2011 2nd International Conference on Environmental Science and Technology (IPCBEE, 2011) V2-160.
23 L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu: ISPRS J. Photogramm. Remote Sens. **130** (2017) 277. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2017.06.001
24 A. I. R. Cabral, S. Silva, P. C. Silva, L. Vanneschi, and M. J. Vasconcelos: ISPRS J. Photogramm. Remote Sens. **142** (2018) 94. https://doi.org/10.1016/j.isprsjprs.2018.05.007
25 A. C. Edwards, S. W. Maier, L. B. Hutley, R. J. Williams, and J. Russell-Smith: Remote Sens. Environ. **136** (2013) 56. https://doi.org/10.1016/j.rse.2013.04.013
26 J. W. van Wagtendonk, R. R. Root, and C. H. Key: Remote Sens. Environ. **92** (2004) 397. https://doi.org/10.1016/j.rse.2003.12.015
27 H. Woo, M. Acuna, B. Madurapperuma, G. Jung, C. Woo, and J. Park: Sens. Mater. **33** (2021) 3745. https://doi.org/10.18494/SAM.2021.3365
28 J. I. Shin, W. W. Seo, T. Kim, J. Park, and C. S. Woo: Forests **10** (2019) 1. https://doi.org/10.3390/f10111025
29 L. Giglio, J. T. Randerson, G. R. van der Werf, P. S. Kasibhatla, G. J. Collatz, D. C. Morton, and R. S. DeFries: Biogeosciences **7** (2010) 1171. https://bg.copernicus.org/articles/7/1171/2010
30 L3Harris Geospatial Solutions, Inc: https://www.l3harrisgeospatial.com (accessed May 2022).
31 J. D. Miller and A. E. Thode: Remote Sens. Environ. **109** (2007) 66. https://doi.org/10.1016/j.rse.2006.12.006
32 FUSION/LDV: Software for LIDAR Data Analysis and Visualization: https://w3.ual.es/GruposInv/ProyectoCostas/FUSION_manual.pdf (accessed June 2022).
33 J. Yang, Z. Kang, S. Cheng, Z. Yang, and P. H. Akwensi: IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **13** (2020) 1055. https://doi.org/10.1109/JSTARS.2020.2979369
34 L. Vincent and P. Soille: IEEE Trans. Pattern Anal. Mach. Intell. **13** (1991) 583. https://doi.org/10.1109/34.87344
35 R. O. Duda, P. E. Hart, and D. G. Stork: Pattern Classification. (Wiley-Interscience, NY, USA, 2000) 2nd ed., Chap. 2.
36 J. Weyermann, D. Schläpfer, A. Hueni, M. Kneubühler, and M. Schaepman: Proc. SPIE 2009 Imaging Spectrometry XIV (2009) 74570B.
37 A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez: arXiv. 1704.06857 (2017) 1. https://doi.org/10.48550/arXiv.1704.06857
38 O. Ronneberger, P. Fischer, and T. Brox: Proc. Int. Conf. Medical Image Computing and Computer Assisted Intervention (Springer, 2015) 234–241.
39 B. Ramsundar and R. B. Zadeh: TensorFlow for Deep Learning (O`Reilly Media, Sebastopol, 2018) 1st ed., p. 256.
40 P. F. Rozario, B. D. Madurapperuma, and Y. Wang: Remote Sens. **10** (2018) 1. https://doi.org/10.3390/rs10091427
41 J. R. Jensen: Introductory Digital Image Processing: A Remote Sensing Perspective (Prentice-Hall Inc., Hoboken, 2015) 4th ed., p. 656.
42 M. Story and R. G. Congalton: Photogramm. Eng. Remote Sens. **52** (1986) 397. https://www.asprs.org/wp-content/uploads/pers/1986journal/mar/1986_mar_397-399.pdf

43 B. N. Tran, M. A. Tanase, L. T. Bennett, and C. Aponte: Remote Sens. **10** (2018) 1. https://doi.org/10.3390/rs10111680
44 E. Chuvieco, M. P. Martín, and A. Palacios: Int. J. Remote Sens. **23** (2002) 5103. https://doi.org/10.1080/01431160210153129
45 X. X. Zhu, D. Tuia, L. Mou, G. S. Xia, L. Zhang, F. Xu, and F. Fraundorfer: IEEE Geosci. Remote Sens. Mag. **5** (2017) 8. https://doi.org/10.1109/MGRS.2017.2762307
46 L. Collins, P. Griffioen, G. Newell, and A. Mellor: Remote Sens. Environ. **216** (2018) 374. https://doi.org/10.1016/j.rse.2018.07.005
47 Z. Pan, J. Xu, Y. Guo, Y. Hu, and G. Wang: Remote Sens. **12** (2020) 1. https://doi.org/10.3390/rs12101574

Supplementary Materials

Table S1
Confusion matrices of ML, SAM, and U-Net classifiers for the two-level burn severity classification.

ML	Unburned	Burned	Total
Unclassified	0.00	0.00	0.00
Unburned	1886.63	170.85	2057.49
Burned	156.75	1395.13	1551.88
Total	2043.38	1565.98	3609.36
SAM	Unburned	Burned	Total
Unclassified	0.00	0.00	0.00
Unburned	1897.82	141.46	2039.28
Burned	145.56	1424.52	1570.08
Total	2043.38	1565.98	3609.36
U-Net	Unburned	Burned	Total
Unclassified	3.57	0.70	4.27
Unburned	1930.25	185.08	2115.33
Burned	109.56	1380.20	1489.76
Total	2039.81	1565.28	3605.09

Table S2
Confusion matrices of ML, SAM, and U-Net classifiers for the four-level burn severity classification.

ML	Unburned	Ground Fire	Heat-Damaged	Crown Fire	Total
Unclassified	0.00	0.00	0.00	0.00	0.00
Unburned	286.25	68.31	83.05	22.13	459.74
Ground fire	1195.28	329.85	73.97	9.25	1608.35
Heat-damaged	129.93	31.63	464.66	271.57	897.79
Crown fire	2.13	0.00	31.10	610.26	643.49
Total	1613.58	429.80	652.78	913.20	3609.36
SAM	Unburned	Ground Fire	Heat-Damaged	Crown Fire	Total
Unclassified	0.00	0.00	0.00	0.00	0.00
Unburned	187.52	32.71	229.73	0.84	450.79
Ground fire	1335.31	396.79	38.75	0.00	1770.86
Heat-damaged	74.79	0.29	350.65	20.63	446.36
Crown fire	15.96	0.00	33.65	891.74	941.35
Total	1613.58	429.80	652.78	913.20	3609.36
U-Net	Unburned	Ground Fire	Heat-Damaged	Crown Fire	Total
Unclassified	90.24	58.25	414.69	49.72	612.90
Unburned	1488.49	350.08	106.74	4.94	1950.25
Ground fire	18.18	19.69	21.63	1.02	60.53
Heat-damaged	3.75	1.02	9.43	1.05	15.24
Crown fire	12.92	0.76	100.29	856.47	970.44
Total	1523.34	371.55	238.09	863.48	2996.46