Mapping Burned Forests Using a k-Nearest Neighbors Classifier in Complex Land Cover

Lee, Hanna* , Yun, Konghyun** , Kim, Gihong***

이한나* · 윤공현** · 김기홍***

k-Nearest Neighbors 분류기를 이용한 복합 지표 산불피해 영역 탐지

ABSTRACT
As human activities in Korea are spread throughout the mountains, forest fires often affect residential areas, infrastructure, and other facilities. Hence, it is necessary to detect fire-damaged areas quickly to enable support and recovery. Remote sensing is the most efficient tool for this purpose. Fire damage detection experiments were conducted on the east coast of Korea. Because this area comprises a mixture of forest and artificial land cover, data with low resolution are not suitable. We used Sentinel-2 multispectral instrument (MSI) data, which provide adequate temporal and spatial resolution, and the k-nearest neighbor (kNN) algorithm in this study. Six bands of Sentinel-2 MSI and two indices of normalized difference vegetation index (NDVI) and normalized burn ratio (NBR) were used as features for kNN classification. The kNN classifier was trained using 2,000 randomly selected samples in the fire-damaged and undamaged areas. Outliers were removed and a forest type map was used to improve classification performance. Numerous experiments for various neighbors for kNN and feature combinations have been conducted using bi-temporal and uni-temporal approaches. The bi-temporal classification performed better than the uni-temporal classification. However, the uni-temporal classification was able to detect severely damaged areas.

Keywords : Forest fire, Damage detection, k-Nearest neighbor, Classification, Sentinel-2

초 록
인간 활동 영역이 산지 곳곳에 퍼져 있는 한국에서는 산불이 주거지역이나 각종 시설물들을 위협하는 경우가 많다. 따라서 산불 이후 대책 마련과 피해 복구를 위한 방안을 마련하기에 도움이 필요하게 되었으며, 이러한 경우 원격탐사가 유용한 도구가 될 수 있다. 본 연구에서는 2019년 4월에 발생한 고성·속초 산불 피해지역에 k-nearest neighbor (kNN) 알고리즘을 적용하여 피해 범위를 탐지하는 실험을 수행하였다. 다양한 인공지질을 포 함하는 지표와 숲이 혼재된 지역 특성을 고려하여 적절한 공간 해상도와 시간 해상도를 제공하는 Sentinel-2 multispectral instrument (MSI) 자료를 사용하였다. Sentinel-2 MSI의 여섯 밴드와 정규식생지수(NDVI), 정규탄화지수(NBR)를 분류 특성으로 사용하였다. 산불 피해지역과 비피해 지역에서 무작위로 추출된 2,000개 지점 정보를 이용하여 kNN 분류기를 훈련시켰다. 분류 성능을 높이기 위해 데이터에서 특이값을 제거하고 임상도를 병용하였다. 다양한 이웃(neighbors) 수와 분류 특성 조합을 적용하여 산불 후 데이터를 이용한 실험과 산불 전 데이터를 이용한 실험을 수행하였다. 산불 전후 데이터를 이용하였을 때 더 우수한 분류 성과를 얻을 수 있었지만, 산불 후 데이터만을 이용한 경우에도 피해지역의 범위를 파악할 수 있었다.

검색어 : 산불, 피해 탐지, k-Nearest Neighbor, 분류, Sentinel-2

Received July 27, 2023 / revised August 19, 2023 / accepted August 19, 2023

Copyright © 2023 by the Korean Society of Civil Engineers
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
1. Introduction

Climate change is a global problem, and the world must be unified in its response. At the same time, a regional system designed to suit local characteristics is needed to respond to disasters caused by climate change. The frequency and intensity of forest fires are increasing owing to climate change (Dennison et al., 2014). This study details a technique that can quickly and easily map forest fire damage in an area where various land cover types are mixed with forests, that is, with “complex land cover” using satellite data.

The spring in Korea from March to May is very dry. Every spring, the wind blows hard in the east coast region, where a large mountain range is located, causing numerous forest fires. There were 2,211 forest fires in 2019 and 1,619 in 2020, of which more than 50% occurred in spring (National Fire Agency, 2021). Korea has few plains; two-thirds of its land is mountainous and is where most forests are located. For this reason, the term “mountain fire” is used synonymously with “forest fire” in Korea. Moreover, due to the high population density, human activities are conducted throughout the mountain regions. Forest fires are therefore dangerous in Korea, and any damage they cause needs to be investigated and remediated immediately. This situation complicates the detection of fire-damaged areas as regions contain various types of land cover, including forests.

In Korea, damage from forest fires was measured using field surveys until research on forest fire damage detection using remote sensing was introduced in the mid-2000s. At that time, Landsat Thermal Mapper (TM) data were generally used for research in Korea. Studies focused on the differences in spectral indices pre- and post-fire and the digital numbers of each band immediately after a forest fire (Choi et al., 2006; Won et al., 2007).

Open source access and proper spectral resolution are the major advantages of Sentinel-2 and Landsat-8; therefore, they have been commonly used in recent forest fire damage detection (Hawbaker et al., 2017; Lee et al., 2017; Roy et al., 2019; Bar et al., 2020; Knopp et al., 2020; Sim et al., 2020). PlanetScope, which has high temporal and spatial resolution, and Korea Multi-Purpose Satellite-3 (KOPMSAT-3) were also employed for forest fire damage detection in Korea (Chung and Kim 2020; Won et al., 2019). With the development of data processing technology, many researchers have imported machine learning or deep learning methods for forest fire damage detection (Mithal et al., 2018; Pinto et al., 2020). Techniques such as decision tree, random forest (RF), logistic regression (LR), gradient boosting, support vector machine (SVM), and convolutional neural networks have been tested and validated (Hawbaker et al., 2017; Weaver et al., 2018; Roy et al., 2019; Bar et al., 2020; Knopp et al., 2020; Sim et al., 2020). Most of recent studies have focused on determining the burned region over a large area. The burned area mapping in large mountainous (Fornacca et al., 2018; Bar et al., 2020) or continental-scale areas (Hawbaker et al., 2017; Mithal et al., 2018; Roteta et al., 2019; Roy et al., 2019), or studies conducted on several fire spots across the globe (Knopp et al., 2020; Pinto et al., 2020) are different from mapping in an area composed of small and diverse patches of land cover, as in our case. As an exception, Sim et al. (2020) evaluated the performance of RF, LR, and SVM models for wildfires that occurred on the east coast of Korea and found that the RF model performed the best; however, its accuracy was poor in lightly damaged areas.

In this study, the k-nearest neighbors (kNN) algorithm was applied to Sentinel-2 multispectral instrument (MSI) data to detect fire-damaged areas on the east coast of Korea. Owing to the aforementioned characteristics of mountainous regions in Korea, high temporal resolution and adequate spatial resolution are required for satellite data. A low spatial resolution may result in the poor accuracy and impractical classification results. The Sentinel-2 data satisfied the requirements in terms of both spatial and temporal resolution. These data have a temporal resolution of 2-3 d in the study area and a spatial resolution of 10 m (visible spectra and near infrared) or 20 m (short-wave infrared (SWIR)). We used six Sentinel-2 MSI bands (bands 2, 3, 4, 8, 11, and 12) and two spectral indices (normalized differential vegetation index (NDVI) and normalized burn ratio (NBR)) calculated using them. The kNN algorithm is a simple machine learning classification technique that is used in various fields (Nigsch et al., 2006; Kara et al., 2017). The main reason for the popularity of kNN is that it makes no assumptions about the distribution of the underlying data (Fix and Hodges, 1989). This is a considerable advantage...
in this study because we used spectral reflectance and indices with various distributions.

2. Materials and Methods

The kNN algorithm is a classification method that assigns unclassified samples to the same category as the nearest classified samples (Cover and Hart, 1967). Areas affected by forest fires are expected to show similar spectral characteristics; therefore, burned area mapping is a field well suited for the application of the kNN algorithm. Assume that a kNN classifier uses n features to classify the degree of damage to each pixel in an image. A feature is a distinct trait that is used to describe each pixel. First, the classifier learns the positions of the previously classified pixels (training dataset) in an n-dimensional space consisting of n axes corresponding to n features. When the coordinates of an unclassified pixel are determined in this space, the classifier finds k preclassified pixels nearest to the unclassified pixel. The unclassified pixel is classified into the category with the most pixels out of k preclassified pixels. A uniform weight was applied to the training dataset and the Euclidian distance metric was used.

We conducted numerous classification experiments, as shown in Fig. 1. While comparing the performance of the bi- and uni-temporal approaches, we observed the performances according to the number of neighbors applied to the kNN classifier and the feature combination. In the case of the uni-temporal approach, we also tested how the classifier, which was trained on a dataset on a specific date, performed on other date images. Each experiment was performed using the process illustrated in Fig. 2.

2.1 Study Area

The study area is a 10.5 km × 11.7 km (latitude 128.48 °N-128.60 °N, longitude 38.17 °S-38.27 °S) region on the east coast of South Korea (Fig. 3(a)). The elevation ranges from

![Fig. 1. Flowchart of Experiments. The Date of Fire was April 4. Differenced Data (April 3-8) was used for Bi-temporal Experiments](image1)

![Fig. 2. Experimental Procedure](image2)
Mapping Burned Forests Using a k-Nearest Neighbors Classifier in Complex Land Cover

10 m to 670 m above sea level. According to the forest type map (Section 2.3), the forest in this area comprises 47% pine, 18% deciduous broad-leaved trees, and 14% coniferous mixed trees, in addition to other types. Artificial surfaces are also present among the forested areas. Fig. 3(b) shows the complex land cover of the study area. Approximately 49% of the area is forested, and cultivated land and rural villages account for a large proportion. Resorts, golf courses, and other facilities are mixed with the forests.

At approximately 7 pm on April 4, 2019, a fire broke out in a mountain in this area and spread rapidly due to strong winds, threatening the city. It was extinguished at approximately 8 am the next day. The damaged area was estimated to be approximately 13 km² (Won et al., 2019), and more than 900 people lost their homes.

2.2 Multispectral Satellite Data and Indices
Sentinel-2 MSI level-2A data, collected on April 3, 5, 8, 15,

---

**Table 1. Features for kNN Classifier**

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>Data Source</th>
<th>Spatial resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>Sentinel-2 MSI band 4</td>
<td>10m</td>
</tr>
<tr>
<td>Green</td>
<td>Sentinel-2 MSI band 3</td>
<td>10m</td>
</tr>
<tr>
<td>Blue</td>
<td>Sentinel-2 MSI band 2</td>
<td>10m</td>
</tr>
<tr>
<td>NIR</td>
<td>Sentinel-2 MSI band 8</td>
<td>10m</td>
</tr>
<tr>
<td>SWIR1</td>
<td>Sentinel-2 MSI band 11</td>
<td>Resampled to 10m</td>
</tr>
<tr>
<td>SWIR2</td>
<td>Sentinel-2 MSI band 12</td>
<td>Resampled to 10m</td>
</tr>
<tr>
<td>NDVI</td>
<td>Calculated from band 4 and 8</td>
<td>10m</td>
</tr>
<tr>
<td>NBR</td>
<td>Calculated from band 12 and 8</td>
<td>10m</td>
</tr>
<tr>
<td>dRed</td>
<td>$R_{April3} - R_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dGreen</td>
<td>$G_{April3} - G_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dBlue</td>
<td>$B_{April3} - B_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dNIR</td>
<td>$NIR_{April3} - NIR_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dSWIR1</td>
<td>$SWIR1_{April3} - SWIR1_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dSWIR2</td>
<td>$SWIR2_{April3} - SWIR2_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dNDVI</td>
<td>$NDVI_{April3} - NDVI_{April8}$</td>
<td>10m</td>
</tr>
<tr>
<td>dNBR</td>
<td>$NBR_{April3} - NBR_{April8}$</td>
<td>10m</td>
</tr>
</tbody>
</table>
and 20, 2019, covering the study area were downloaded from the Copernicus Open Access Hub. This is a list of all available April data for the study area, excluding those with clouds. The Sentinel-2 level-2A main output is an orthoimage Bottom-Of-Atmosphere (BOA)-corrected reflectance product (European Space Agency, 2015).

Among the 13 spectral bands of the Sentinel-2 product, three visible spectra (bands 2, 3, and 4), band 8 (near-infrared, NIR), band 11 (short wave infrared1, SWIR1), and band 12 (short wave infrared2, SWIR2) were used as features (Table 1) for the kNN classifier. The Sentinel-2 MSI level2A product contains 10 m resolution data for bands 2, 3, 4, and 8, while the resolution of bands 11 and 12 is only up to 20 m. Band 11 and 12 data were resampled to a 10 m resolution in this study.

2.3 Forest Type Map
During the spring farming season, land cover often changes significantly over a short period, especially in areas with rice paddies and dry fields. For the bi-temporal approach, arable land far from forest fires can frequently be classified as damaged. To prevent this error, a forest type map (Fig. 4) was used to exclude non-forest areas from the damaged area. The forest type map, which was produced by the National Institute of Forest Science, is updated approximately every ten years, but land cover in Korea changes on a much shorter time scale. Therefore, the forest type map alone is insufficient, and the classifier must learn various land cover types to accurately map burned areas.

2.4 Reference Map
The burn severity map, derived from KOMPSAT-3 2 m resolution multi spectral bands (Won et al., 2019), was used as a reference (Fig. 5). Won et al. (2019), with the support of the National Institute of Forest Science and the Korea Forest Service, investigated the 2019 east coast forest fires to map the fire-damaged area and burn severity. This reference map classifies damage grades as extreme, high, or low. ‘Extreme’ means that the tree canopies are completely burned, and ‘high’ means that more than 60% of the canopy has withered due to heat. ‘Low’ refers to a surface fire in which the canopies mostly survived. In this study, for convenience, we renamed these damage levels as DL3, DL2, and DL1. The undamaged area corresponds to the DL0.

2.5 Methods
The experimental process is shown in Fig. 2 and details are described below.

2.5.1 Spectral Indices and Differenced Images
For April 3rd, 5th, 8th, 15th, and 20th, a dataset consisting of eight features—six band reflectance and two indices (Table 1) was prepared for uni-temporal classification. NDVI and NBR were calculated from the Sentinel-2 MSI band reflectance as follows:

\[
\text{NDVI} = \frac{\text{band8} - \text{band4}}{\text{band8} + \text{band4}}, \quad \text{NBR} = \frac{\text{band8} - \text{band12}}{\text{band8} + \text{band12}} \tag{1}
\]

Next, six differenced reflectance and two differenced indices (Table 1) were prepared for bi-temporal classification.
2.5.2. Preprocessing
As discussed in the introduction, the data used for the kNN algorithm do not require any mathematical assumptions. However, because kNN depends on the distance between samples, it is affected by the feature scales. A feature with a wide range dominates the classification results.

Level-2A data derive from corresponding Level-1C products. Level-1C products provide TOA-normalized reflectances in which the physical values range from $10^{-4}$ to 1, but values higher than 1 can be observed in some cases.

Fig. 6. Histograms of “Red” and “NIR” Features. The Left Two Show the Distributions of Single Date Data (April 8) and the Right Two Show the Distributions of Differenced Data (April 3-April 8): (a) Raw Data, (b) Preprocessed (Outliers Removed and Normalized) Data

Fig. 7. Outliers on a Differenced Image (April 3-April 8, Band 4). Red Pixels are the Outliers which are Defined as 0.1% at Both Ends of Tails of the Distribution (0.05% each Side)
due to specific angular reflectivity effects (Gascon et al., 2017). Each pixel value of Level-1C and Level-2A is an integer multiplied by 10,000 by the reflectance. In this study, this integer was regarded as reflectance. As a result, the reflectance ranges from 1 to 10,000, but may have a value higher than 10,000 in some cases.

In our dataset, the reflectance of each band was mostly distributed below 5,000, but it increased to approximately 20,000 in a few pixels, namely outliers. NDVI and NBR ranged from -1 to 1. In the case of differenced reflectance, the range extends to approximately 25,000. Fig. 6(a) shows the distributions of reflectance of “Red” and “NIR” features. In particular, long tails on both sides make the distribution of the differenced reflectance very sharp. The wide range of reflectance is due to the diversity of land cover in the study area. Most outliers appear on artificial structures such as resorts, buildings, barns, and markets (Fig. 7). The characteristic that outliers mainly appear on the artificial surface is also seen in single date images.

Outliers should be removed to improve classification performance. Outliers do not have a mathematical definition and may be defined in different ways for various models (Klebanov, 2016). In this study 0.05% on the upper side in the single-date reflectance distribution and 0.05% on both sides (total 0.1%) in differenced reflectance distributions were defined as outliers and removed. No outliers were removed from the indices. These outlier boundaries were determined by analyzing the histograms and images (Fig. 7 and 8). The outlier-removed reflectance data and indices were normalized using a minimum-maximum scaler in the range of 0 to 1. The histograms of the preprocessed features are shown in Fig. 6(b).

2.5.3. Training Dataset
A total of 2,000 sample points, 1,000 from fire-damaged areas, and 1,000 from undamaged areas, were randomly extracted to build a training dataset (Fig. 9). A uni-temporal training dataset was constructed by extracting the feature (Table 1) values corresponding to the coordinates of these sample points from the April 8 image. These feature values are stored as attribute information for the corresponding fields of the points. The bi-temporal training dataset was extracted from the differenced images, which were obtained by subtracting the image pixel values collected on April 8th from the image pixel values collected on April 3rd. In addition to the 16 feature fields, the attribute table also has two more fields for damage level and forest type map.

3. Results
Fig. 10 shows the classification result of bi-temporal classification in which all eight features (Table 1) were used,
Mapping Burned Forests Using a k-Nearest Neighbors Classifier in Complex Land Cover

Fig. 9. Training Dataset

Fig. 10. Comparison of the Reference and the Result Classified Through Bi-temporal, 8-feature, 9-neighbor Classification. DL0 (Undamaged) was Set to be Transparent

* DL: Damage Level (see Section 2.4)
and nine neighbors were applied to the kNN classifier. It can be seen that a number of isolated pixels were classified as DL1 or DL2 in the undamaged area (Fig. 10[A]). These commission errors contribute significantly to the deterioration of classification performance. Although the majority of commission errors occurring in non-forest areas have been eliminated using the forest type map (section 2.3), some are still found, as shown in Fig. 10[A]. This is because the forest type map has a 10-year update cycle and does not reflect changes of land cover within that 10-year period. A large number of these errors can be removed later using image processing techniques, which are not covered in this study.

The less damaged the area, the lower the match rate between the classified result and the reference image. This is clearly shown in the error matrix (Tables 2 and 3). The producer and user accuracies of DL1 and DL2 were lower than those of DL3. While overall accuracy is the most representative measure that summarizes classification performance, it is also a very ambiguous measure that does not reveal any inertial information (Alberg et al., 2004; Story and Congalton, 1986). In this study, the overall accuracy was not considered because the classification accuracy of DL0, which occupies a large proportion of the study area, determines the overall accuracy. Instead, classification performance was evaluated based on Cohen's kappa coefficient $\kappa$, which measures the agreement between two classifiers (Cohen, 1960; Landis and Koch, 1977; Nichols et al., 2010).

### 3.1 Number of Neighbors for kNN Classifier

Various numbers of neighbors (N) were applied to the kNN classifier to examine the variation in performance. Ns were chosen as odd numbers to prevent a tie. The more N is, the better the performance. However, when N exceeds 5, the increase rate of Cohen's Kappa decreases significantly, and when N exceeds 9, the increase becomes smaller than 0.01 (Fig. 11).

<table>
<thead>
<tr>
<th>Classified</th>
<th>Reference</th>
<th>Burn severity</th>
<th>Sum</th>
<th>User's accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL0</td>
<td>1,067,267</td>
<td>10,316</td>
<td>1,096,809</td>
<td>0.9731</td>
</tr>
<tr>
<td>DL1</td>
<td>6,424</td>
<td>5,329</td>
<td>18,332</td>
<td>0.2907</td>
</tr>
<tr>
<td>DL2</td>
<td>9,066</td>
<td>4,857</td>
<td>42,047</td>
<td>0.4257</td>
</tr>
<tr>
<td>DL3</td>
<td>5,860</td>
<td>1,026</td>
<td>57,431</td>
<td>0.7119</td>
</tr>
<tr>
<td>Sum</td>
<td>1,088,617</td>
<td>21,528</td>
<td>1,214,619</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Error Matrix of Bi-temporal, 8-feature, and 9-neighbor Classification

<table>
<thead>
<tr>
<th>Classified</th>
<th>Reference</th>
<th>Burn severity</th>
<th>Sum</th>
<th>User's accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DL0</td>
<td>1,062,369</td>
<td>12,532</td>
<td>1,099,052</td>
<td>0.9666</td>
</tr>
<tr>
<td>DL1</td>
<td>4,334</td>
<td>2,376</td>
<td>11,428</td>
<td>0.2079</td>
</tr>
<tr>
<td>DL2</td>
<td>17,557</td>
<td>4,820</td>
<td>47,829</td>
<td>0.3436</td>
</tr>
<tr>
<td>DL3</td>
<td>5,905</td>
<td>1,810</td>
<td>58,119</td>
<td>0.7091</td>
</tr>
<tr>
<td>Sum</td>
<td>1,090,165</td>
<td>21,538</td>
<td>1,216,428</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3. Error Matrix of Uni-temporal, 8-feature, and 9-neighbor Classification
3.2 Feature Combination

Feature combinations were designed as the table at the bottom of Fig. 12. Fig. 12 shows the performance of the 9-neighbor classification for each feature combination.

The bi-temporal approach showed the best performance when all eight features were used (8f in Fig. 12). There was no noticeable change in performance if one of the three visible spectral features was removed (7f-b-d). Even if all R, G, and B are removed (5f), there is only a slight deterioration in performance. In contrast, removing SWIR1(7f-a) degraded the performance. For the uni-temporal approach, removing all three visible spectra results in the best performance (5f).

3.3 Application Dates

A classifier trained with the April 8th dataset was applied on the datasets taken from April 5th, 8th, 15th, and 20th for classification. As expected, the classification performance on April 8th was the best, but the other date classification also identified the rough extent of damage area well (Fig. 13). In Korea, vegetation grows rapidly in April, and it is observed that the area classified as DL1 or DL2 on April 5th, because it had not yet sprouted, was later excluded from the damaged area. Fig. 14 shows the classification performance quantitatively. The DL1 and DL2 regions have low producer accuracies. In the case of DL3, unlike other groups, the producer’s accuracy was highest on April 5th, immediately after the fire, and then gradually decreased.
4. Discussion

The kNN classifier was applied to an area with complex land cover and was able to map fire damaged areas. Depending on the classification conditions, the kappa values ranged from 0.58 to 0.62 in the bi-temporal classification and 0.54 to 0.58 in the uni-temporal classification (Fig. 11 and 12). The performance of the uni-temporal classifier deteriorated when applied to other images not included in the training data (Fig. 14), but it was able to identify the boundary of the damaged area in the classified image (Fig. 13).

In a previous study (Sim et al., 2020), the user’s and producer’s accuracies for RF, LR, and SVM were 85-95% in undamaged areas and 65-85% in extremely damaged areas,
while the range in accuracy for lightly damaged areas, from 15-60%, was much larger. Our results were not significantly different with 97-98% in undamaged areas, 68-71% in extremely damaged areas, and 24-29% in lightly damaged areas (Table 2). Slightly different study conditions complicate a direct comparison, but the kNN classifier seems to have performance similar to RF, LR, and SVM. Poor classification accuracy in lightly damaged areas was a common problem in both studies because of the difficulty distinguishing between lightly damaged and undamaged areas. However, lightly damaged areas are rare and it may be the case that a sufficient number of high-quality samples were not included in the training dataset to allow for a full characterization of this class. A possible solution is to consider using a 2-class classification scheme that only distinguishes between damaged/undamaged areas, or to divide the damaged area into two classes instead of three. The high accuracy of the kNN classifier for undamaged areas is presumably because relatively more samples were allocated to the DL0 area than to DL1, DL2, and DL3 areas. This imbalance was intended to account for the large amount of variation present in multiple land cover types. For an accurate comparison of different approaches, more experimental data will be needed.

As expected, the bi-temporal classification showed better performance than the un-temporal classification (Fig. 11 and 12). However, with un-temporal classification, it is possible to determine the outline of the damaged areas of DL2 and DL3. DL1, which suffered minor fire damage, was roughly detected. There is no disagreement that a bi-temporal approach incurs more costs than a un-temporal approach. In some early studies, the bi-temporal approach’s superiority of performance did not cover these costs; thus, a un-temporal approach was suggested (Weber et al., 2008). Now, as data accessibility and data processing technology are more advanced, the difference in cost is greatly reduced, and hence, it is less burdensome to choose a bi-temporal approach for a small improvement. In the case of Korean forest fires, there is a high probability that the satellite data immediately before and after the fire are available, and the cost of data processing is relatively low as the burned areas are not large. Therefore, if sufficient satellite data are available, there is no reason to avoid the bi-temporal approach. However, if it is difficult to collect data due to weather conditions, a un-temporal approach can also provide useful information about forest fires.

As a result of the experiment on the number of neighbors for the kNN classifier, it was observed that the improvement rate decreased when the number of neighbors exceeded 9 (Fig. 11). The optimal number of neighbors according to the number of samples and the number of classes for a region with specific geographical and social characteristics seems to be another interesting area of research.

According to the experiments on feature combinations (Fig. 12), the use of Red, Green, and Blue features slightly reduced the un-temporal classification performance. In the bi-temporal classification, these features help a little, and there is no big loss even if all are excluded. In contrast, when SWIR1 was excluded, the classification performance was significantly reduced.

The kNN algorithm is a machine learning technique that has been used for a long time. Compared to newer machine learning techniques, it has the advantage of being easy to understand and simple to apply, and it is worth examining how it performs in detecting areas affected by forest fires within complex land cover. Although experiments in which the classifier was applied to imagery acquired on various dates (test IDs P805 and P815 in Fig. 1) were included, the training and application of the classifier were performed within the same area in our experiment, and it is highly likely that performance will deteriorate when it is applied to other regions. In future studies, the generality of the classifier can be improved by including multiple sites and the above discussion regarding feature combinations, appropriate number of neighbors, and outliers may be helpful in guiding the experimental setup.

5. Conclusions

In this study, experiments were conducted under the assumption that the area damaged by forest fire in a complex area comprising a mixture of forest and artificial land cover could be quickly identified using the k-nearest neighbors (kNN) algorithm. The two requirements for the kNN algorithm are the number of neighbors and the feature combination for the classifier. A combination of nine neighbors and eight features (Red, Green, Blue, NIR, SWIR1, SWIR2, and two spectral
indices, NDVI and NBR) showed appropriate performance. Although this combination did not always exhibit the best performance, it ensured consistently good performance throughout all our experiments. When the 9-neighbor 8-feature kNN classifier was applied over the images of different days following the forest fire, the Cohen’s kappas of the classification results ranged between 0.52 and 0.64. Although this quantitative figure seems insufficient, the resulting classification images sufficiently recognized the boundaries of the fire-damaged area.

Forest fires are a persistent phenomenon on the east coast of Korea. This study demonstrated the possibility of establishing a stable automated system that can quickly detect the boundaries of fire-damaged areas by applying a prepared kNN classifier when a forest fire occurs in the future. Furthermore, owing to the removal of outliers and the use of forest type map, this classification system is expected to be effective even in complex areas with artificial land covers. Information on the boundaries of burned areas, that is, the burned area map, can be used as a basic reference for on-the-spot investigations and countermeasures immediately after a forest fire, and can be provided to residents or related parties in the surrounding area who need information.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2021R1A6A1A03044326) and the Ministry of Science and ICT (No. RS-2023-0025262).

ORCID

Hanna Lee https://orcid.org/0000-0001-7862-8306
Konghyun Yun https://orcid.org/0000-0001-8999-0204
Gihong Kim https://orcid.org/0000-0003-3280-5340

References


Mapping Burned Forests Using a k-Nearest Neighbors Classifier in Complex Land Cover


