Spectral-Spatial Ensemble Learning for Invasive Robinia Pseudoacacia Detection Using UAV-Based Hyperspectral Imaging

San Gwon[1], Ayano Aida[2], Chan Park[3], Sejong Yu[4], Jaeyong Lee[5], Choongsik Kim[6], Hosik Choi[1*]

[1] Department of Urban Big Data Convergence, University of Seoul, Seoul, Republic of Korea – (jmt30269, choi.hosik)@uos.ac.kr
[2] Department of Urban Planning & Design, University of Seoul, Seoul, Republic of Korea – ayano91@uos.ac.kr
[3] Department of Landscape Architecture, University of Seoul, Seoul, Republic of Korea – chaneparkmomo7@uos.ac.kr
[4] Geostory, Seoul, Republic of South Korea – sjongyu@geostory.co.kr
[5] Department of Traditional Landscape Architecture, Korea National University of Heritage, Buyeo, Republic of Korea – leejaeyong82@knuh.ac.kr
[6] Department of Heritage Science and Technology Studies, Korea National University of Heritage, Buyeo, Republic of Korea – kimch@knuh.ac.kr

Keywords: Hyperspectral Image, Image Classification, Deep Learning, Ensemble.

Abstract

The National Heritage Administration of Korea designates and manages particular non-native species, as well as highly reproductive native plants with strong environmental adaptability, as *invasive plants* to preserve the unique natural landscapes within cultural heritage sites. However, investigating and managing large-scale cultural heritage areas—such as palaces and fortresses—requires considerable time and labor, and these efforts are further hindered by challenging terrain. This study investigates the use of hyperspectral imaging (HSI), acquired via unmanned aerial vehicles (UAVs), as an efficient approach for monitoring the distribution of black locust (*R. pseudoacacia*), a representative invasive species. HSI provides rich spectral information by continuously measuring reflectance across a wide range of wavelength bands. We utilize HSI data comprising 150 spectral bands to detect the presence of black locust in the Gongsanseong and Busosanseong fortress areas. To address the limitations of benchmark-based models, such as poor generalizability and overfitting to dataset-specific features, we propose an ensemble approach that integrates the strengths of multiple learning models. This includes neural networks designed to capture both spectral and spatial features, allowing for complementary processing of complex spectral patterns and spatial contextual information. From a numerical study, the proposed method achieves robust detection performance for target species, even in heterogeneous environments.

1. Introduction

Cultural heritage, as a tangible legacy of human history and tradition, embodies cultural uniqueness, identity, and the evolution of societal practices. Among these, large-scale cultural heritage sites are characterized by extensive spatial coverage or the presence of structures with considerable vertical elevation or facade area. These sites often include diverse vegetative elements, which, while contributing to unique natural landscapes, may also include invasive plant species. Non-native and certain native species with high reproductive capacity and broad environmental adaptability are classified and managed as invasive due to their potential ecological impact.

In South Korea, the conservation and management of state-designated cultural heritage sites are periodically assessed through visual inspections, as stipulated in Article 44 of the Cultural Heritage Protection Act. However, monitoring large-scale cultural heritage sites imposes significant financial and human resource burdens, with accessibility further constrained by topographical complexity. To address these issues, we require the utility of drone-based HSI as a viable remote sensing approach. HSI enables the simultaneous acquisition of reflectance data across a wide spectral range, encompassing visible to near-infrared wavelengths.

Unlike traditional RGB imaging, which captures information in three broad, discrete bands, HSI records reflectance across numerous contiguous, narrow spectral bands. This high spectral resolution facilitates the differentiation of materials with similar spectral characteristics, a capability beyond the scope of RGB or conventional multispectral systems (Landgrebe et al., 2002). Despite its advantages, the use of aerial HSI poses several practical challenges. Accurate spectral acquisition is highly dependent on optimal solar illumination geometry and consistent lighting conditions, given the sensitivity of spectral reflectance to ambient variations. Additional complications arise from atmospheric interference, sensor misregistration, and spectral mixing caused by scattered incident/reflected light. These factors contribute to intra-class variability and may degrade classification accuracy. Furthermore, the labelling process—crucial for supervised learning—is hindered by the inaccessibility of ground-truthing locations due to rugged terrain. In such cases, label assignment for R. pseudoacacia must rely on expert interpretation of HSI data, supplemented by auxiliary RGB imagery. The spatial distribution of R. pseudoacacia, while often clustered, can also appear random, introducing further uncertainty and noise into the labelling process (Wang et al., 2023).

Conventional approaches typically assess model performance on well-structured benchmark datasets. However, such evaluations may result in overfitting to dataset-specific characteristics, leading to performance degradation when models are deployed in real-world scenarios. To address this issue, the present study proposes an ensemble learning framework that integrates predictions from multiple heterogeneous models to enhance generalization and robustness.

This study focuses on detecting R. pseudoacacia—an invasive species—within two large-scale cultural heritage sites in South Korea: Gongsanseong Fortress and Busosanseong Fortress. High-resolution HSI data acquired via drone are used to formulate the detection problem as a binary classification task, distinguishing between pixels corresponding to R. pseudoacacia and those of the surrounding environment.

This paper is organized as follows: Section 2 details the dataset used, the data preprocessing steps, and the proposed methodology. Section 3 describes the experimental results and a

https://doi.org/10.5194/isprs-annals-X-M-2-2025-107-2025 | © Author(s) 2025. CC BY 4.0 License.

comparative analysis between baseline models and the proposed approach. Finally, Section 4 presents the conclusions and limitations of this paper.

2. Methodology

2.1 Target Region

Gongsanseong fortress and Busosanseong fortress are representative mountain fortresses dating from the Baekje period, constructed utilizing natural topographical features. As shown in Figure 1-2, both fortresses are situated within forested regions and host diverse plant communities, conferring significant ecological and landscape value. Notably, the intra-fortress vegetation is considered a critical element that both harmonizes with the natural environment and influences the conditions for preserving cultural heritage.

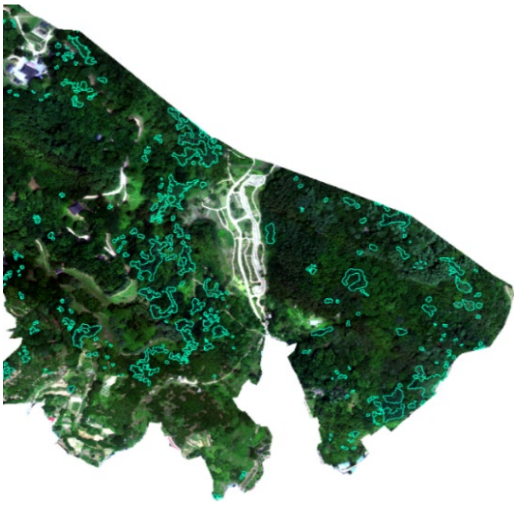


Figure 1. View of Gongsanseong Fortress with Polygon

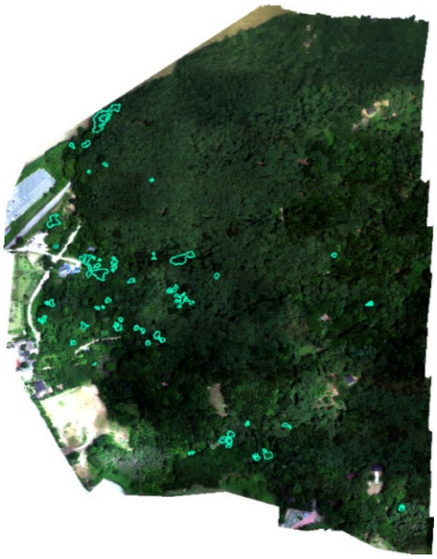


Figure 2 Busosanseong Fortress with Polygon Labelling

The dominant indigenous arboreal species within Gongsanseong and Busosanseong predominantly include Quercus spp. (e.g., Quercus acutissima, Quercus serrata, Quercus aliena), Pinus spp., Prunus spp., and Acer spp. These species constitute the principal canopy layer of the forest structure within the sites. The understory layer comprises various shrub species such as Rhododendron mucronulatum, Cornus walteri, and Weigela spp., contributing to a naturally stratified vegetation structure. R. pseudoacacia, an alien tree species, has established extensive colonies within the fortress precincts, facilitated by its rapid growth rate and high reproductive capacity. However, this species is officially designated as a noxious invasive plant due to its detrimental ecosystem impacts, including suppressing the growth of indigenous flora and reducing local biodiversity. Furthermore, its aggressive root system expansion and the accumulation of litterfall and organic matter pose potential risks to the structural integrity and preservation environment of cultural heritage elements. Consequently, intensive and sustained monitoring and management protocols specifically targeting R. pseudoacacia are deemed essential.

2.2 Dataset

2.2.1 Original Dataset

This study utilized HSI acquired over the Gongsanseong and Busosanseong sites in July 2023. Owing to the extensive spatial extent of these areas, data acquisition was conducted using an Unmanned Aerial Vehicle (UAV). Multiple overlapping image strips were captured during sequential flight lines. Subsequently, these individual strips were processed and synthesized into a single composite mosaic image for each respective site, generating the final datasets.

The acquired hyperspectral data encompass the spectral range from 400 nm to 1000 nm, comprising a total of 150 spectral bands. The imagery has a ground sampling distance (GSD) of 0.3 meters, corresponding to its spatial resolution. The resulting hyperspectral data cube for the Gongsanseong fortress has dimensions of 2,799 × 2,563 × 150 (rows × columns × bands), and for the Busosanseong fortress, the dimensions are 2,715 × 1,843 × 150.

Site	Spatial Resolution	Number of Spectral Bands	Spectral Range (nm)	Image Dimensions (Rows × Columns × Bands)
Gongsanseong	0.3m	150	400 – 1000	2799 * 2563 * 150
Busosanseong				2715 * 1843 *150

Table 1. Specifications of Hyperspectral Images

2.2.2 Data Preprocessing

Following the generation of the integrated mosaic imagery, non-forested areas were masked out and excluded from the primary analysis domain. Polygonal regions of interest were subsequently delineated based on a combination of previously utilized polygon datasets and ground-truth data obtained through dedicated *in situ* field surveys. The finalized ROI polygons for each study site are presented in Figure 1-2. A total of 251 polygons were delineated for the Gongsanseong fortress, and 49 polygons for the Busosanseong fortress. The labeled data in this study was not solely reliant on visual annotations by experts; instead, it was constructed based on ground truth information acquired through twelve direct field surveys using a high-precision GPS (Trimble R10), thereby minimizing noise and bias in the annotation process.

This dataset suffers from severe class imbalance problem. Pixels corresponding to the target class, *R. pseudoacacia*, represent only a small fraction of the total pixel population within the analysis areas, accounting for just 6.02% of all pixels in the Gongsanseong fortress and only 0.8% in the Busosanseong fortress. To mitigate potential performance degradation and bias resulting from this severe imbalance between the target and background classes, several remedial strategies were considered, including oversampling, undersampling, and data augmentation.

Due to XGBoost’s requirement for tabular input, direct application to 3D HSI data is not feasible. Accordingly, HSI was processed on a per-pixel basis, using the spectral bands as the primary input features. To incorporate spatial context, additional features were generated by computing the mean reflectance values across 5×5-pixel neighborhoods centered on each pixel. In contrast, the deep learning models directly utilized 7×7-pixel patches extracted from the HSI data, allowing them to inherently learn both spectral and spatial features in a unified spatio-spectral representation.

In this study, the dataset was partitioned by allocating 3% of the total data for the training set, with the remaining 97% reserved as the test set. This partitioning strategy was adopted primarily considering the severe class imbalance inherent in the dataset, as previously detailed. Allocating a significantly large proportion of the data for training under such conditions of pronounced imbalance presents a considerable risk: specifically, utilizing a larger training subset could result in insufficient exemplars for the minority class (R. pseudoacacia), potentially hindering the model's ability to effectively learn its distinguishing features during the training phase.

2.3 Proposed Method

2.3.1 XGBoost

Extreme Gradient Boosting (XGBoost) is an ensemble learning technique belonging to the boosting family, which constructs a robust predictive model by combining multiple weak learners. This method is founded on the boosting principle, wherein new models are iteratively trained based on the errors remaining from preceding models in the sequence, thereby incrementally improving the overall model performance (Chen et al., 2016). XGBoost operates within the generalized gradient boosting framework. This framework utilizes the gradient descent algorithm to minimize a specified differentiable loss function, optimizing the model at each iteration by fitting a new learner to the negative gradient of the loss function concerning the current ensemble's predictions. XGBoost incorporates several algorithmic enhancements and optimization techniques (including sub-optimal heuristics) designed to implement this gradient boosting process with high computational efficiency, often by reducing algorithmic complexity. It is engineered with the dual objectives of achieving both enhanced predictive performance and efficient training speed, incorporating advanced features such as regularization and the capability to handle missing values during the model training process intrinsically.

2.3.2 SpectralFormer

SpectralFormer (SF) model leverages a state-of-the-art Transformer-based architecture specifically adapted for HSI analysis. It is designed to effectively capture subtle variations within spectral sequences, a task often considered challenging for conventional CNN or RNN approaches. Key innovations include the Group-wise Spectral Embedding module, which embeds adjacent spectral bands collectively on a group basis, and the Cross-layer Adaptive Fusion module, which facilitates the effective integration of feature information propagated across different network layers. This design enables the model to simultaneously learn both local spectral differences and long-range dependencies within the HSI data cube (Hong et al., 2021).

2.3.3 DSNet

DSNet (Dual-Branch Subpixel-Guided Network) is a deep learning framework specifically engineered to enhance HSI classification performance. It features a dual-branch architecture comprising a deep autoencoder-based unmixing network and a CNN-based classifier network. The former extracts subpixel-level information, while the latter derives pixel-level class features. These distinct feature sets are subsequently integrated via a subpixel fusion module. This module is designed to ensure high-fidelity information fusion between pixel and subpixel representations, thereby enabling more robust and accurate classification outcomes. DSNet surpassed the performance of several existing state-of-the-art HSI classification methodologies across three standard benchmark datasets (Han et al., 2024).

2.3.4 Group-Aware-Hierarchical-Transformer

Group-Aware Hierarchical Transformer (GAHT) is a Transformer-based model specifically designed for HSI classification. It introduces a novel Grouped Pixel Embedding (GPE) module designed to constrain the scope of the Multi-Head Self-Attention mechanism to local spatio-spectral contexts. The GPE module emphasizes local relationships within the spectral channels of the HSI data, facilitating feature extraction that captures both global and local characteristics within the spatio-spectral domain. Furthermore, GAHT employs a hierarchical architecture, which aims to minimize the number of model parameters while concurrently improving classification accuracy (Mei et al., 2022).

2.3.5 SSFTT

Spectral–Spatial Feature Tokenization Transformer (SSFTT) is a Transformer-based model proposed for the effective extraction and integration of spectral and spatial information from high-dimensional HSI data. The core concept involves partitioning the input data into distinct 'Spectral Tokens' and 'Spatial Tokens'. Features associated with each token type are learned independently before being integrated for the final classification task.

Specifically, the Spectral Tokenization module focuses on extracting features along the spectral dimension, while the Spatial Tokenization module captures information about the spatial structure. This separation allows the Transformer architecture to process these different, yet crucial, information types in a balanced manner. Subsequently, a Feature Tokenization module integrates these spectral and spatial tokens, generating a unified spectral-spatial representation that is fed into the classifier. This architectural design aims to mitigate potential information loss issues inherent in high-dimensional data processing and contribute to enhanced classification accuracy (Sun et al., 2023).

2.3.6 RSSAN

RSSAN (Residual Spectral-Spatial Attention Network) is a deep learning model specifically designed for HSI classification, which uniquely combines residual learning principles with attention mechanisms to extract and fuse spectral and spatial information effectively. The model typically learns spectral and spatial features through separate pathways initially, followed by an integration step that effectively represents the complex data structure inherent in HSIs. The incorporation of residual connections facilitates the training of deeper networks while mitigating the gradient vanishing problem. Concurrently, attention mechanisms are employed to selectively emphasize salient spectral and spatial features, thereby enhancing the overall classification performance (Zhu et al., 2021).

2.3.7 Ensemble

Ensemble learning combines multiple individual base models to enhance overall predictive accuracy and robustness. Rather than relying on the output of a single model, ensemble methods aggregate predictions from a diverse set of learners. This process typically leads to reduced variance, mitigates overfitting, and improves generalization performance compared to standalone models.

Representative ensemble methods include bagging (Bootstrap Aggregating), boosting, and stacking (Stacked Generalization). Bagging techniques, exemplified by Random Forests, generally involve training multiple base learners independently on different bootstrap samples drawn from the original dataset and combining their predictions, often through averaging or majority voting. Boosting methods, such as XGBoost, sequentially construct an ensemble by training weak learners. Each new model focuses on correcting the errors or residuals of the preceding models in the sequence. Stacking utilizes the predictions generated by multiple diverse base models as input features for a secondary model termed a meta-learner, which is trained to produce the final output prediction.

In the context of analyzing high-dimensional data like HSI, ensemble learning provides a potent framework. It allows for the effective utilization of both spectral and spatial information by integrating the complementary strengths inherent in diverse model architectures to improve overall predictive performance. We use a soft voting method. This performs the final prediction by averaging the probability-based outputs produced by each individual constituent model:

$$\mathrm{F(x)}=\sum_{\mathrm{m=1}}^{\mathrm{M}} \mathrm{w_m\, f_m(x)} \qquad (1)$$

Here, w_m represents the weight assigned to the m-th constituent model in the ensemble ($\sum_{\mathrm{m=1}}^{\mathrm{M}} \mathrm{w_m}=1, \mathrm{w_m} \geq 0, \forall m$) . These weights can typically be determined based on the performance of individual models evaluated on a separate validation dataset, or alternatively, they can be assigned uniformly (e.g., $w_m = \frac{1}{M}$ for an ensemble of M models) to implement a simple averaging scheme.

2.3.8 Process of Proposed Model

Fig. 3 illustrates the overall processing pipeline of the proposed methodology designed for R. pseudoacacia cluster detection utilizing HSI data. The input to the pipeline is the HSI data cube, which inherently contains both spatial and spectral information. From this input, various relevant features are extracted. Subsequently, classification predictions are performed independently using a suite of six distinct models: XGBoost, DSNet, GAHT, RSSAN, SF, and SSFTT.

Each of these models yields an output representing the estimated probability of presence for the target (R. pseudoacacia). These individual probabilistic predictions are then aggregated using the previously mentioned soft voting ensemble strategy. This probability averaging approach is adopted because combining outputs from multiple diverse models is generally effective in reducing prediction variance and mitigating the risk of overfitting compared to deploying any single model. The final classification output is thus derived based on the averaged probability scores across all participating models in the ensemble.

3. Results

3.1 Experimental Environment and Parameter Settings

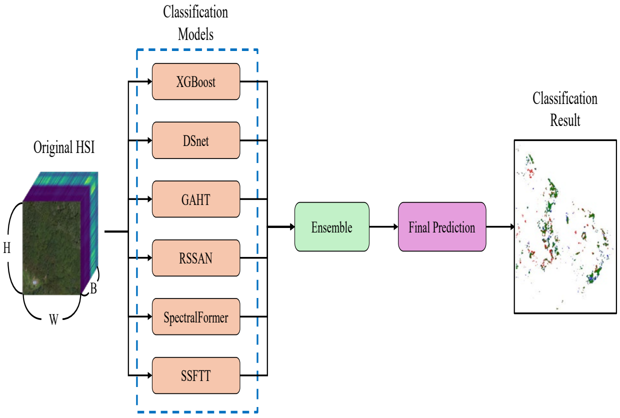


Figure 3. Pipeline of Proposed Model

All experiments were conducted on a system equipped with an Intel(R) Xeon(R) Gold 6348 CPU, 256GB RAM, and an NVIDIA A10 24GB GPU. We set 200 training epochs. For the XGBoost model, the learning rate was set to 0.3. For the deep learning-based models, excluding RSSAN and GAHT, we used a 0.001 learning rate for the Adam optimizer. The RSSAN model was trained using a learning rate of 0.003 with the RMSprop optimizer, while the GAHT model was trained using a learning rate of 0.001 with the SGD optimizer.

3.2 Performance Evaluation Metrics

In this study, the performance of the models is evaluated using three key metrics: precision, recall, and the F1-score. Precision is defined as the ratio of correctly predicted positive samples to the total number of samples predicted as positive. Specifically, true positives (TP) represent the number of actual positive instances that are correctly identified, while false positives (FP) refer to the number of negative cases that are incorrectly predicted as positive. Recall is the ratio of correctly predicted positive samples to the total number of actual positive samples. In this context, false negatives (FN) are the number of positive instances that are incorrectly classified as negative. The F1-score, which is the harmonic mean of precision and recall, provides a single measure that balances the trade-off between the two. The formulas used to compute these performance metrics are as follows:

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

3.3 Experimental Results

Table 2 and 3 present the comparative classification performance results for the Gongsanseong and Busosanseong datasets, respectively, evaluated using precision, recall, and F1-score metrics for the proposed ensemble model versus the individual baseline models. Overall, the proposed ensemble model demonstrates superior performance across all three metrics, thereby substantiating the effectiveness of the proposed method over reliance on any single baseline model.

Models	Precision	Recall	F1-score
XGBoost	0.5565	0.7970	0.6553
SF	0.7937	0.6928	0.7397
DSNet	0.7437	0.6236	0.6688
GAHT	0.7939	0.7481	0.7702
SSFTT	**0.7992**	0.6834	0.7361
RSSAN	0.7972	0.7116	0.7514
Proposed	0.7901	**0.7973**	**0.7937**

Table 2. Classification Result of Gongsanseong Fortress

Focusing on the results for the Gongsanseong dataset (Table 2), specific observations regarding the baseline models are as follows: XGBoost achieved a high recall of 0.7970, but its precision was markedly low at 0.5565. This resulted in an F1-score of 0.6553, the lowest recorded among all evaluated models, suggesting a potential tendency for over-detection leading to a high incidence of FPs. SF exhibited relatively stable performance, recording precision of 0.7937, recall of 0.6928, and F1-score of 0.7397, thus maintaining a reasonably high recall. DSNet showed comparatively lower performance, with a recall of 0.7437 and notably the lowest Precision (0.6236), yielding a correspondingly low F1-score of 0.6688. The GAHT, SSFTT, and RSSAN models all demonstrated generally stable performance, each achieving recalls exceeding 0.79. Among these, GAHT displayed balanced and excellent results with a precision of 0.7481 and an F1-score of 0.7702. RSSAN also yielded a satisfactory F1-score of 0.7514. Although SSFTT attained the highest recall (0.7992), its lower precision of 0.6834 constrained its F1-score to 0.7361.

In contrast, the proposed ensemble model, which integrates the prediction probabilities derived from these diverse classifiers, achieved the highest performance across all three-evaluation metrics for the Gongsanseong dataset. It recorded a precision of 0.7901, a recall of 0.7973, and an F1-score of 0.7937. Regarding the results for the Busosanseong dataset (Table 3), the XGBoost model recorded the highest recall score at 0.8835. However, its relatively low precision (0.4746) and consequent F1-score (0.6174) indicate a tendency towards high detection sensitivity coupled with numerous false positives. This characteristic, suggesting frequent misclassification of non-target vegetation, limits its practical utility for standalone application. Conversely, the GAHT model demonstrated generally balanced and strong performance, exhibiting particularly excellent results in terms of harmonizing high precision and recall. RSSAN and SSFTT also delivered comparable top-tier results, recording F1-scores of 0.7995 and 0.7938, respectively.

Models	Precision	Recall	F1-score
XGBoost	0.4746	0.8835	0.6174
SF	0.8293	0.7318	0.7774
DSNet	0.7034	0.5624	0.6074
GAHT	0.8537	**0.8095**	0.8308
SSFTT	0.8269	0.7638	0.7938
RSSAN	0.8417	0.7619	0.7995
Proposed	**0.8782**	0.8053	**0.8400**

Table 3. Classification Result of Busosanseong Fortress

The proposed ensemble model yields an F1-score of 0.8400 for the Busosanseong dataset. It achieved the highest F1-score among all models and concurrently secured the highest precision, indicating outstanding performance capabilities in both accurate detection of the target species and the minimization of FPs. These results further underscore the efficacy of the proposed method in synergistically combining the complementary strengths of diverse constituent models to maximize overall vegetation detection performance.

In summary, the proposed ensemble model demonstrates the ability to achieve both high recall and strong precision, confirming its robustness and reliability across the diverse study sites.

3.3.1 Ablation Study

To evaluate stability and generalization capability of the proposed ensemble model, we conduct experiments. For both the Gongsanseong and Busosanseong datasets, the evaluation was repeated utilizing 10 random seed initializations. For each random seed, optimal ensemble weights for combining base models are derived based on the classification performance achieved on the corresponding test dataset. The precision, recall, and F1-score metrics were calculated for the ensemble model configured with this optimal weight. The six base models used in the ensemble, in the order corresponding to the weight sequence, are GAHT, RSSAN, SpectralFormer (SF), SSFTT, XGBoost, and DSNet.

Seed	Best weights	Precision	Recall	F1-score
20250401	(0.3, 0.1, 0.2, 0.1, 0.3, 0.0)	0.7908	0.7855	0.7881
20250402	(0.3, 0.1, 0.1, 0.1, 0.4, 0.0)	0.7829	0.8104	0.7964

20250403	(0.4, 0.1, 0.1, 0.0, 0.4, 0.0)	0.7783	0.8045	0.7912
20250404	(0.3, 0.2, 0.2, 0.0, 0.3, 0.0)	0.7947	0.7929	0.7938
20250405	(0.4, 0.0, 0.1, 0.2, 0.3, 0.0)	0.7907	0.7962	0.7935
20250406	(0.3, 0.2, 0.1, 0.1, 0.3, 0.0)	0.7917	0.8036	0.7976
20250407	(0.2, 0.3, 0.1, 0.1, 0.3, 0.0)	0.7897	0.7935	0.7916
20250408	(0.3, 0.2, 0.1, 0.1, 0.3, 0.0)	0.8021	0.7936	0.7978
20250409	(0.3, 0.1, 0.1, 0.2, 0.3, 0.0)	0.7883	0.7978	0.7930
20250410	(0.3, 0.2, 0.2, 0.0, 0.3, 0.0)	0.7907	0.7937	0.7922

Table 4. Result of Gongsanseong Fortress

Table 4 summarizes the results obtained from the seed-specific experiments conducted on the Gongsanseong dataset. The performance metrics exhibited the following ranges: precision varied from 0.7738 to 0.8021, recall ranged from 0.7855 to 0.8134, and the F1-score spanned from 0.7881 to 0.7978. The average F1-score calculated over these 10 seeds was approximately 0.7935.

A consistent trend observed across most trials was the concentration of optimized ensemble weights on the GAHT, XGBoost, and RSSAN models. In contrast, the SSFTT and DSNet models were frequently assigned low weights, and in some cases, weights of zero. This implies that the proposed method effectively employs a selective combination strategy—allocating higher weights to models that demonstrate greater reliability, as inferred from their performance contributions on the test set during the weight optimization process, while automatically down-weighting or excluding models with comparatively limited predictive value.

The experiment conducted using Seed 20250408 yielded the highest performance among trials with a precision of 0.8021, recall of 0.7936, and an F1-score of 0.7978. The corresponding optimal weights for this seed, following the order GAHT, RSSAN, SF, SSFTT, XGBoost, and DSNet, were 0.3, 0.2, 0.1, 0.1, 0.3, and 0.0, respectively.

Table 5 presents the results for the Busosanseong dataset, which generally exhibited higher performance levels compared to those observed for Gongsanseong. Across all trials, the performance metrics for Busosanseong were observed within the following ranges: Precision from 0.8603 to 0.8982, recall from 0.7824 to 0.8247, and F1-score from 0.8363 to 0.8442. The average F1-score across these trials was calculated as 0.84.

For the Busosanseong dataset, weight distribution concentrated on the GAHT, SSFTT, and RSSAN models was particularly prominent. In most cases, the optimization process assigned DSNet and XGBoost zero weight and thus effectively removed them from the ensemble process. It implies that the relative contribution of specific constituent models was critically influenced by the data characteristics inherent to the Busosanseong region. The most outstanding performance for this dataset was achieved with Seed 20250401, recording a precision of 0.8647, a recall of 0.8247, and an F1-score of 0.8442. This optimal result corresponds to an ensemble combination that assigned a high weight (0.5) to the GAHT model. The corresponding optimal weights, where the orders are GAHT, RSSAN, SF, SSFTT, XGBoost, and DSNet, were 0.5, 0.3, 0.2, 0.0, 0.0, and 0.0, respectively.

Seed	Best weights	Precision	Recall	F1-score
20250401	(0.5, 0.3, 0.2, 0.0, 0.0, 0.0)	0.8647	0.8247	0.8442
20250402	(0.5, 0.0, 0.1, 0.4, 0.0, 0.0)	0.8603	0.8217	0.8406
20250403	(0.5, 0.1, 0.1, 0.3, 0.0, 0.0)	0.8881	0.8043	0.8442
20250404	(0.5, 0.0, 0.2, 0.3, 0.0, 0.0)	0.8982	0.7824	0.8363
20250405	(0.5, 0.1, 0.4, 0.0, 0.0, 0.0)	0.8673	0.8098	0.8376
20250406	(0.5, 0.2, 0.2, 0.1, 0.0, 0.0)	0.8826	0.8068	0.8430
20250407	(0.4, 0.2, 0., 0.1, 0.0, 0.0)	0.8907	0.7882	0.8363
20250408	(0.5, 0.2, 0.1, 0.2, 0.0, 0.0)	0.8850	0.8028	0.8419
20250409	(0.5, 0.2, 0.2, 0.1, 0.0, 0.0)	0.8717	0.8064	0.8378
20250410	(0.5, 0.0, 0.3, 0.2, 0.0, 0.0)	0.8733	0.8062	0.8384

Table 5. Result of Busosanseong Fortress

The inter-site comparison revealed that the Busosanseong dataset generally yielded superior performance compared to Gongsanseong, with an observed difference of approximately 0.05 based on the average F1-score. This performance discrepancy could potentially be attributed to variations between the study sites in factors such as vegetation density, inherent spectral characteristics of the landscape elements, or the quality of the original acquired HSI data.

Furthermore, the proposed ensemble method exhibited low performance variance and consistently stable results across different random seed initializations for both sites. This consistency highlights the model’s inherent reproducibility and robustness. Additionally, the observation that the optimal weight combinations differed between the two regions suggests that the ensemble framework possesses a dynamic, weight-adaptive structure capable of adjusting to site-specific data characteristics.

Overall, the multi-seed experiments confirm that the proposed ensemble model delivers robust and consistent performance irrespective of initialization conditions, while also demonstrating a valuable ability to adaptively adjust weighting schemes based on the contextual characteristics of the input data. This finding is particularly significant in terms of the model’s reliability and scalability for real world field.

4. Conclusion

We propose an ensemble-based methodology that integrates HSI with machine learning and deep learning techniques to detect *R. pseudoacacia* in large-scale cultural heritage sites effectively. Specifically, we use an ensemble framework to aggregate prediction probabilities from a diverse set of models, including XGBoost, SpectralFormer, DSNet, GAHT, SSFTT, and RSSAN. By leveraging this ensemble approach, the predictive limitations of individual models are mitigated, thereby enhancing overall classification accuracy.

While the present study addressed the data imbalance issue through undersampling, this method has an inherent limitation of potential information loss. Future research should therefore apply data augmentation techniques, such as Thin Plate Spline (TPS) transformation, to better preserve the structural characteristics of vegetation like Robinia pseudoacacia. Beyond this data-level consideration, the study's scope itself presents a limitation. As the research is centered on Korean fortresses, the generalizability of its results to other geographical locations or different categories of cultural heritage is inherently limited. Consequently, further studies should also aim to improve the model's universality by introducing domain adaptation methods.

The proposed method demonstrates robust detection capabilities for target species, even in heterogeneous environments. Moreover, it offers significant practical utility for large-scale vegetation monitoring in cultural heritage areas, potentially reducing the need for extensive manual fieldwork. Future work will focus lightweight model optimization. The primary goal is to improve computational efficiency by reducing inference time and minimizing model complexity, thereby enhancing the system's feasibility in operational settings.

Acknowledgments

This research was supported by the Korea Heritage Service and the National Research Institute of Cultural Heritage (RS-2021-100402) and the National Research Foundation of Korea grant funded by the Korea government (RS-2022-NR068754). The authors acknowledge the Urban Big Data and AI Institute of the University of Seoul supercomputing resources (http://ubai.uos.ac.kr) made available for conducting the research reported in this paper.

References

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, 226–231.

Han, Z., Yang, J., Gao, L., Zeng, Z., Zhang, B., Chanussot, J., 2024. Dual-branch subpixel-guided network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 62, 1–13, Art. no. 5521813. https://doi.org/10.1109/TGRS.2024.3418583.

Hong, D., Han, Z., Yao, J., Gao, L., Zhang, B., Plaza, A., Chanussot, J., 2022. SpectralFormer: Rethinking hyperspectral image classification with transformers. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–15. https://doi.org/10.1109/TGRS.2021.3130716.

Kim, D., Na, M., 2023. Rice yield prediction and self-attention visualization using a Video Vision Transformer. *J. Korean Data Anal. Soc.*, 25(4), 1249–1259 (in Korean). https://doi.org/10.37727/jkdas.2023.25.4.1249.

Kim, H., 2020. The prediction of PM2.5 in Seoul through XGBoost ensemble. *J. Korean Data Anal. Soc.*, 22(4), 1661–1671 (in Korean). https://doi.org/10.37727/jkdas.2020.22.4.1661.

Kuo, B., Ho, H., Li, C., Hung, C., Taur, J., 2013. A kernel-based feature selection method for SVM with RBF kernel for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 7(1). https://doi.org/10.1109/JSTARS.2013.2262926.

Landgrebe, D., 2002. Hyperspectral image data analysis. *IEEE Signal Process. Mag.*, 19(1), 17–28. https://doi.org/10.1109/79.974718.

Lee, D., Kim, K., 2023. Analysis of the effect of surface temperature in accordance with the composition of land cover based on XAI SHAP. *J. Korean Data Anal. Soc.*, 25(5), 1735–1748 (in Korean). https://doi.org/10.37727/jkdas.2023.25.5.1735.

Mei, S., Song, C., Ma, M., Xu, F., 2022. Hyperspectral image classification using group-aware hierarchical transformer. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–14, Art. no. 5539014. https://doi.org/10.1109/TGRS.2022.3207933.

Mercier, G., Lennon, M., 2003. Support vector machines for hyperspectral image classification with spectral-based kernels. *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, 1, 288–290. https://doi.org/10.1109/IGARSS.2003.1293752.

Shin, J., Lee, K., 2012. Comparative analysis of target detection algorithms in hyperspectral image. *Korean J. Remote Sens.*, 28(4), 369–392. https://doi.org/10.7780/kjrs.2012.28.4.3.

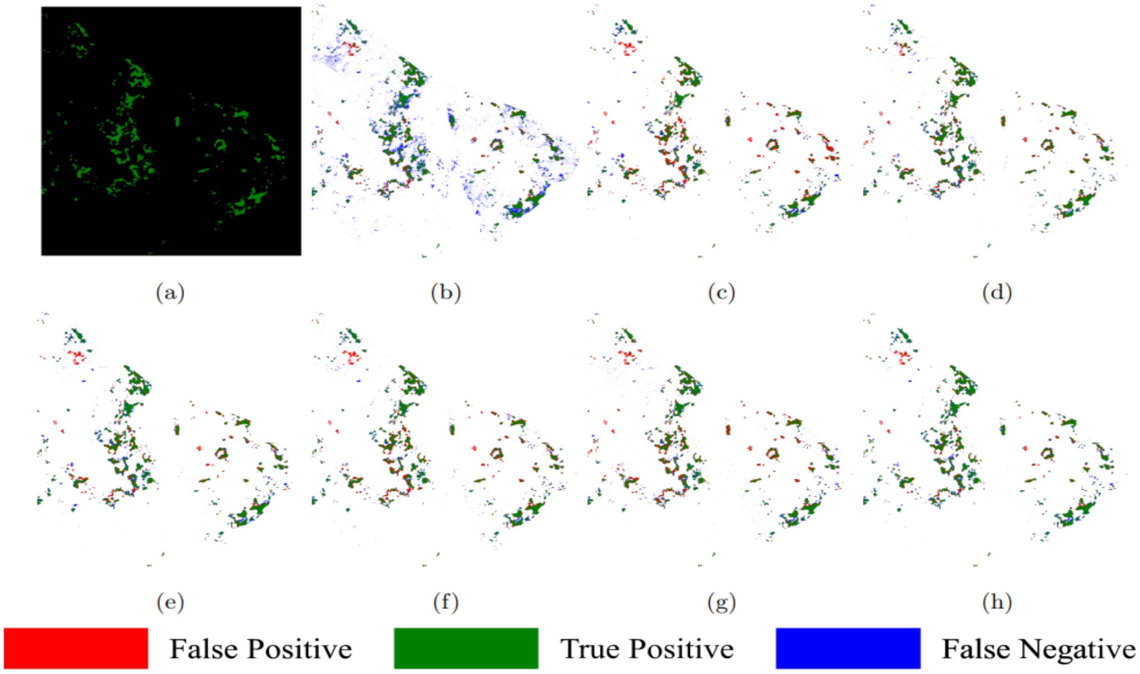
Sun, L., Zhao, G., Zheng, Y., Wu, Z., 2022. Spectral–spatial feature tokenization transformer for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–14, Art. no. 5522214. https://doi.org/10.1109/TGRS.2022.3144158.

Wang, X., Liu, J., Chi, W., Wang, W., Ni, Y., 2023. Advances in hyperspectral image classification methods with small samples: A review. *Remote Sens.*, 15, 3795. https://doi.org/10.3390/rs15153795.

Yang, X., Ye, Y., Li, X., Lau, R., Zhang, X., Huang, X., 2018. Hyperspectral image classification with deep learning models. *IEEE Trans. Geosci. Remote Sens.*, 56(9), 5408–5423. https://doi.org/10.1109/TGRS.2018.2815613.

Zhong, Z., Li, J., Luo, Z., Chapman, M., 2018. Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework. *IEEE Trans. Geosci. Remote Sens.*, 56(2), 847–858. https://doi.org/10.1109/TGRS.2017.2755542.

Appendix A. Result of Gongsanseong Fortress and Busosanseong Fortress



(i)

Figure A1. Classification Map for Gongsanseong Fortress, (a) Ground Truth, (b) XGBoost, (c) DSNet, (d) GAHT, (e) RSSAN, (f) SF, (g) SSFTT, (h) Proposed Model, (i) Color Labels

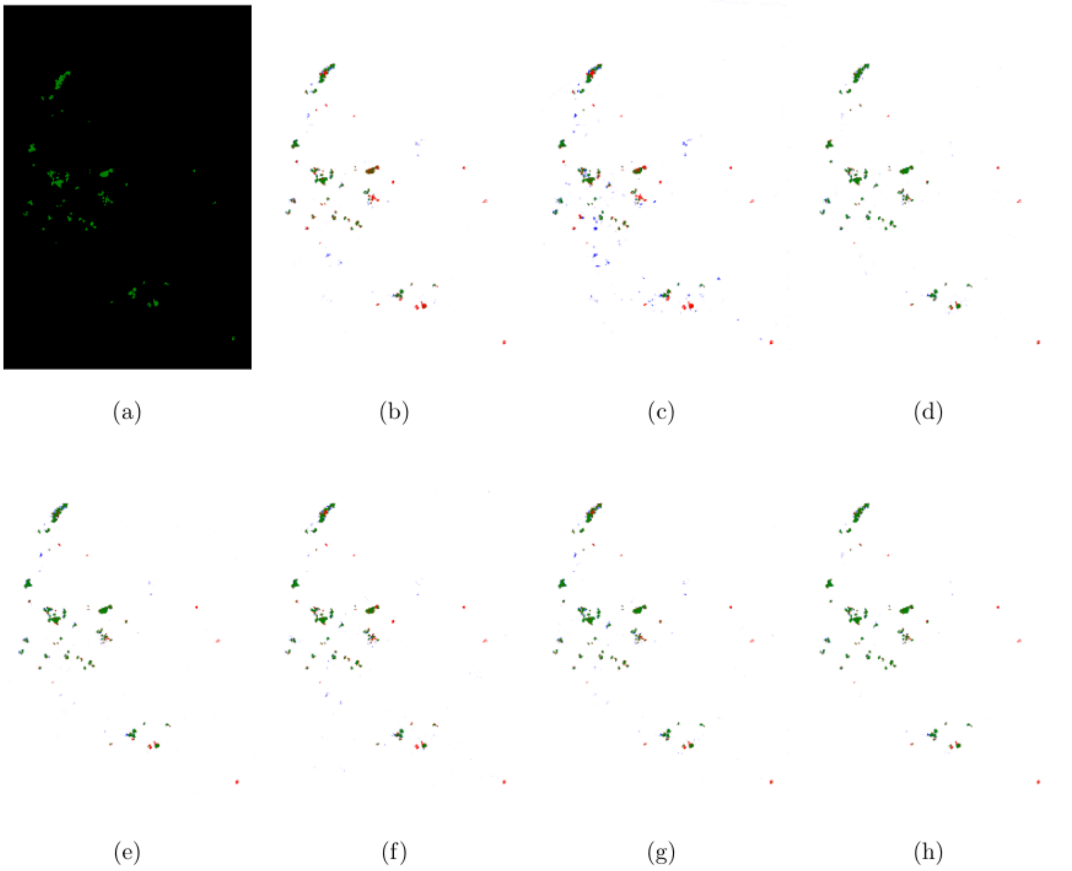


Figure A2. Classification Map for Busosanseong Fortress, (a) Ground Truth, (b) XGBoost, (c) DSNet, (d) GAHT, (e) RSSAN, (f) SF, (g) SSFTT, (h) Proposed Model