

Domain Generalization in Deep Learning for Forest Health Monitoring Using Multispectral UAS Data

Emma Turkulainen¹, Raquel Alves de Oliveira¹, Niko Koivumäki¹, Päivi Lyytikäinen-Saarenmaa¹, Roope Näsi¹, Mikko Peltö-Arvo²,
Johanna Tuviala², Eija Honkavaara¹

¹ Department of Remote Sensing and Photogrammetry, Finnish Geospatial Research Institute in National Land Survey of Finland (FGI), 02150 Espoo, Finland – emma.turkulainen@nls.fi, raquel.alvesdeoliveira@nls.fi, niko.koivumaki@nls.fi, paivi.lyytikainen-saarenmaa@nls.fi, roope.nasi@nls.fi, eija.honkavaara@nls.fi

² School of Forest Sciences, University of Eastern Finland, Joensuu, Finland – johanna.tuviala@uef.fi, mikko.pelto-arvo@uef.fi

Keywords: Domain Generalization, Deep Learning, Bark Beetle, UAS, Multispectral Imagery

Abstract

Deep learning has significantly advanced forest health monitoring by enabling automated analysis of high-resolution aerial imagery. However, the generalization of these models across ecologically diverse regions remains limited due to domain shifts, which are variations in environmental conditions between training and testing locations. In this study, we propose a domain generalization (DG) framework that disentangles domain-invariant, task-relevant features from domain-specific environmental variations in multispectral UAS imagery. Our approach extends a baseline 2D convolutional neural network by incorporating parallel domain-specific and shared feature extractors, along with a domain classifier trained via adversarial learning. We evaluate the model using a leave-one-site-out strategy across three Finnish forest sites with diverse ecological characteristics. Results show that the DG model improves classification accuracy in previously unseen environments, with performance gains of up to 27% compared to the baseline. These findings highlight the effectiveness of feature disentanglement in enhancing the robustness and transferability of deep learning models for forest canopy health assessment, supporting more scalable and reliable forest monitoring solutions.

1. Introduction

Unmanned aerial systems (UAS) have become essential tools for forest health monitoring, providing high-resolution data needed to detect early signs of disease and pest outbreaks. The European spruce bark beetle (*Ips typographus* L.) poses a particularly serious threat to boreal forests, causing extensive damage that can lead to large-scale forest decline (Patacca et al., 2023; Hlásny et al., 2019).

Deep learning models, especially convolutional neural networks (CNNs), show promise for automating tree health classification from aerial imagery (Safonova et al., 2019; Turkulainen et al., 2023). However, these models face a significant practical limitation: they often fail when applied to new geographic areas or environmental conditions different from their training data. This problem, known as domain shift, occurs when factors like lighting, canopy structure, terrain, or sensor characteristics vary between training and deployment locations.

Domain adaptation (DA) and domain generalization (DG) techniques address this challenge by training models to focus on task-relevant features while ignoring environmental variations (Shai et al., 2006; Muandet et al., 2013). While DA methods require data from target locations during training, DG approaches work without any prior knowledge of deployment conditions. This makes DG particularly valuable for operational forest monitoring, where data from new areas may not be available during model development.

DG methods typically fall into three categories: data manipulation (such as augmentation or domain randomization), representation learning (like feature disentanglement), and learning strategies (including ensemble or meta-learning approaches) (Wang et al., 2022; Zhou et al., 2021). This study focuses on representation learning through feature disentanglement, which explicitly separates task-relevant features from domain-specific characteristics.

Recent work has shown promising results combining feature disentanglement with adversarial learning. Park et al. (2025) developed methods using gradient reversal layers to promote domain-invariant features, while Chen et al. (2024) proposed contrastive approaches that align feature distributions across domains. The gradient reversal layer (GRL) technique, which maximizes domain prediction error to encourage domain-invariant features, has proven particularly effective (Ganin et al., 2016).

Despite growing adoption in remote sensing (Zhu et al., 2021; Luo et al., 2024), DG applications in forest monitoring remain limited. Most existing bark beetle detection systems are trained and tested within single locations, severely limiting their practical applicability across diverse forest environments. This study addresses these limitations by developing a DG framework specifically for UAS-based forest health monitoring. We extend our previous CNN architecture (Turkulainen et al., 2023) by adding domain-adversarial training components. Our approach uses dual feature extractors, combined with adversarial learning to improve generalization across different environmental conditions.

Our earlier work demonstrated strong performance within individual sites but required transfer learning with local data for new locations (Turkulainen et al., 2024). Here, we investigate whether integrating DG can eliminate the need for site-specific retraining while maintaining robust performance across diverse forest environments.

We evaluate our approach using multispectral imagery from three ecologically distinct Finnish forest sites, representing different vegetation types, seasons, and imaging conditions. A leave-one-site-out strategy tests the model's ability to generalize to completely unseen environments, providing a rigorous assessment of domain generalization capabilities.

Our approach represents a novel application of domain generalization principles to the specific challenges of bark beetle detection in boreal forests, contributing both methodological innovations and practical solutions for operational forest monitoring systems. By demonstrating the effectiveness of feature disentanglement in this domain, we aim to establish a foundation for more robust and scalable deep learning approaches in forest health assessment.

2. Materials and methods

2.1 UAS data collection

Data was collected from four ecologically diverse sites across Finland: Ruokolahti (61°29'21.84"N, 29°3'0.72"E), Koli National Park (63°5'29.92"N, 29°48'36.01"E), Paloheinä in Helsinki Central Park (60°15'25.20"N, 24°55'19.20"E), and Evo National Park (61°10'19.84"N, 25°8'7.81"E). The locations of these sites are shown in Figure 1. The datasets cover a range of years, seasons, and UAS platforms, introducing natural variability to vegetation phenology, lighting, and canopy structure.



Figure 1. Location of the study sites. Background map from the National Land Survey of Finland Topographic Database.

UAS imagery from Ruokolahti, Koli, and Paloheinä has been described in prior studies (Turkulainen et al., 2023; Turkulainen et al., 2024). At Ruokolahti, data were captured in late August of 2019, 2020, and 2021 using a MicaSense Altum sensor. At Koli, a Kelluu airship equipped with a MicaSense RedEdge camera was used to collect imagery in May, June, and September of 2023. In Paloheinä, data were acquired in May, July, and September 2020 using a MicaSense Altum mounted on a quadcopter.

To support species classification, we incorporated an additional dataset from the Evo site, contributing complementary forest characteristics. The Evo site comprises mature boreal forests with mixed stands dominated by spruce, pine, and birch. Imagery was acquired in early August 2018 with a MicaSense RedEdge sensor.

2.2 Reference data collection

Ground reference data were collected across all four study sites to support supervised training and validation of the deep learning classification models. The data collection approach varied by site based on study objectives and logistical constraints, but all sites employed standardized assessment protocols for tree health evaluation.

Reference data collection procedures for the Ruokolahti and Paloheinä sites are detailed in Turkulainen et al. (2023). At Ruokolahti, systematic health assessments were conducted from 2019 to 2021 across four sub-sites using 66 circular sampling plots. At Paloheinä, individual tree assessments were performed during spring, summer, and autumn campaigns in 2020, with different trees monitored in each season. Field experts manually selected and annotated individual trees distributed throughout the area, targeting a balanced representation of crown vitality statuses.

The data collection procedures for the Koli site are described comprehensively in Turkulainen et al. (2024). Field reference data were collected across 28 circular plots, where trained experts recorded tree species and canopy condition. Plots were visited repeatedly throughout the growing season to capture temporal changes in tree health status.

At the Evo site, species information was collected in the field from four 20 × 20 meter sample plots, focusing on identifying tree species visible in aerial imagery, particularly dominant and co-dominant individuals.

Tree health was assessed using standardized symptom categories following protocols adapted from Blomqvist et al. (2018). Assessment criteria included crown discoloration, defoliation, bark damage, resin flow, and the presence of bark beetle entrance or exit holes. This standardized approach ensured consistency in health classifications across all study sites.

Tree locations were recorded using site-specific positioning methods. In Ruokolahti and Koli, plot centers were located using a Trimble Geo XT GPS device (Trimble Navigation Ltd., Sunnyvale, CA, USA) and a Topcon Hiper HR GNSS receiver (Topcon Corporation, Tokyo, Japan), respectively, with individual tree positions determined by measuring distance and azimuth from plot center. In Evo, tree positions were extracted from terrestrial laser scanning (TLS) point clouds collected using a Leica HDS6100 scanner (Leica Geosystems AG, Heerbrugg, Switzerland) and verified during field investigations. In Paloheinä, the tree positions were determined from orthophotos collected prior to field tree selection.

2.3 Data processing

2.3.1 Image preprocessing: Individual tree crowns were manually annotated by creating bounding boxes around each tree using the recorded GPS coordinates as reference points. Tree locations were cross-referenced with the generated orthophotos to ensure accurate positioning. Trees that could not be clearly identified or delineated in the orthophotos due to factors such as overlapping crowns, shadow occlusion, or poor image quality were excluded from the dataset to maintain annotation reliability.

Individual tree crown images were extracted from the orthophotos using the bounding box coordinates. To ensure consistent input dimensions for the deep learning models, all extracted images were standardized to 150 × 150 pixels. Images smaller than 150 × 150 pixels were zero-padded to maintain

aspect ratio and reach target dimensions. Images larger than 150×150 pixels were downsampled using nearest neighbour interpolation. The 150×150 pixel resolution was selected to balance computational efficiency with preservation of crown-level detail necessary for health classification.

2.3.2 Data labelling: The labels for the trees were determined based on the reference data collected in the field. The tree health classes were determined based on the colour of the crown, such that green and faded green spruces were labelled as healthy, yellow and yellowish trees as infested and reddish brown and grey trees as dead. All other trees were labelled as non-spruce. The total number of samples for each class across all study sites is presented in Table 1.

Dataset	Healthy	Infested	Dead	Non-spruce
Paloheinä	129	107	281	0
Ruokolahti	818	19	498	72
Koli	493	21	210	150
Evo	53	0	0	129

Table 1. Total numbers of tree crown images assigned to each class by study site.

For training of the DG model, we needed to define domains that capture the main sources of variation in forest imagery. We considered three approaches: geographic domains (one per site), site-temporal domains (combining location and time), and seasonal-meteorological domains (combining season and weather conditions).

We selected the seasonal-meteorological approach, creating six domains: (1) sunny spring, (2) cloudy spring, (3) sunny summer, (4) cloudy summer, (5) sunny fall, and (6) cloudy fall. This strategy was chosen for several reasons.

Firstly, seasonal and lighting conditions create the most significant visual variation in forest imagery. Spring shows trees during leaf-off conditions and early foliation, summer captures full canopy development, and fall reveals natural senescence that must be distinguished from pest damage. Weather conditions (sunny vs. cloudy) strongly affect image contrast, shadows, and color saturation. This approach also maintains adequate sample sizes across domains while capturing meaningful environmental variation. Alternative approaches like site-temporal domains would create too many small domains, some with few or no samples for certain health classes. Additionally, these domains align with practical deployment scenarios where bark beetle monitoring systems must work reliably across different seasons and weather conditions.

2.4 Model implementation

We implemented the DG model by extending the 2D-CNN architecture used by Turkulainen et al. (2023). The baseline model comprises a feature extractor with three convolutional layers and a classification head. To incorporate DG capabilities, we

augmented the architecture with two key components: (1) a domain-specific feature extractor operating in parallel with the original feature extractor, now denoted as the shared feature extractor, and (2) a domain classifier designed to mirror the task classifier's structure.

The shared feature extractor is optimized to learn domain-invariant, task-relevant features, while the domain-specific extractor captures environmental characteristics that arise from domain variation. Both shared and domain features are input to the domain classifier. A gradient reversal layer (GRL) is applied to the shared features before domain classification, promoting invariance by maximizing domain prediction error from these features.

Canopy health classification is performed exclusively using the shared features, ensuring that predictions are based on generalized, domain-independent representations. The model is trained using a composite loss function comprising three components: (i) task classification loss, (ii) domain classification loss computed on domain-specific features, and (iii) domain confusion loss applied to shared features via the GRL. This adversarial training setup encourages the shared feature space to be discriminative for the classification task while remaining invariant to domain-specific variations. The total loss function is formulated as:

$$L_{total} = \lambda_1 \times L_{task} + \lambda_2 \times L_{domain} + \lambda_3 \times L_{confusion}, \quad (1)$$

where λ denotes a loss weighing coefficient controlling how much influence each loss component has to the total loss value.

Figure 2 illustrates the architecture of the proposed DG classification model. This disentangled optimization strategy enhances the model's ability to generalize across diverse environmental contexts by effectively isolating task-relevant signals from domain-specific noise.

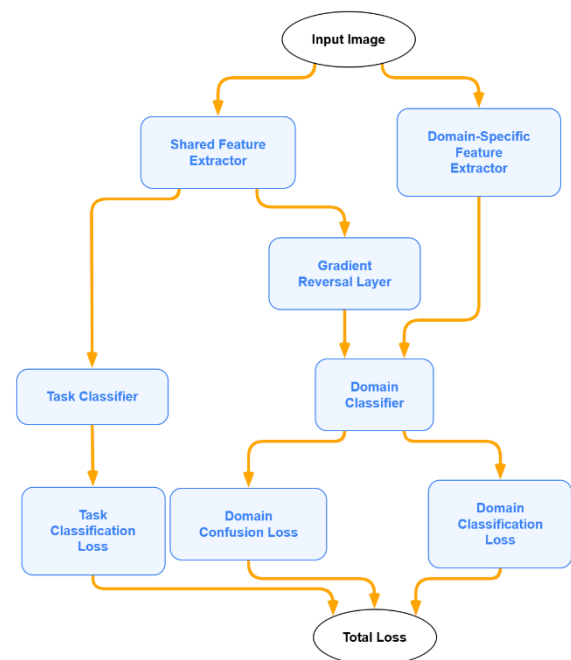


Figure 2. Architecture of the DG model.

2.5 Model training

All models were implemented using PyTorch and trained on NVIDIA GPUs with CUDA acceleration. To ensure model convergence and stability, we incorporated several regularization techniques within the network architecture. Batch normalization layers were applied after each convolutional operation to stabilize training dynamics and accelerate convergence. Dropout layers with a rate of 0.5 were inserted before the final classification layers to prevent overfitting and improve generalization performance.

The models were trained using the Adam optimizer with an initial learning rate of 0.0001, chosen based on preliminary experiments showing optimal convergence behavior for this specific task and architecture. A learning rate scheduler with exponential decay (decay rate = 0.95, decay steps = 1000) was employed to gradually reduce the learning rate during training, facilitating fine-tuning in later epochs. Early stopping was implemented based on validation loss with a patience of 20 epochs to prevent overfitting. Training was conducted in mini-batches of size 16, for a maximum of 200 epochs

To enhance model robustness to site-specific variations and increase the effective size of the training dataset, comprehensive data augmentation was applied during training. The augmentation pipeline included random horizontal and vertical flipping, random rotations, and random translations, random brightness variations, contrast adjustments, and saturation modifications to simulate different lighting and atmospheric conditions. Gaussian noise was added to improve robustness to sensor-specific characteristics. These augmentations were applied randomly during each training epoch, ensuring that the model never saw identical input images across different epochs, thereby reducing memorization and improving generalization capability.

The loss weighting coefficients for the total loss calculation were empirically determined through systematic grid search across the parameter space: $\lambda_1 \in [0.5, 1.0, 1.5]$, $\lambda_2 \in [0.1, 0.3, 0.5]$, and $\lambda_3 \in [0.05, 0.1, 0.2]$. Optimal performance was achieved with weights $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.1$, respectively. These weights ensure that the primary classification task remains the dominant objective while allowing sufficient influence from the domain generalization components.

For each leave-one-site-out experiment, the training data was further subdivided using stratified sampling to maintain class balance. The data from the source domains was divided into training and validation following an 80:20 split. Test set included 100% of data from the held-out target domain. Domain labels were assigned consistently across all splits according to the seasonal-meteorological categorization described in the data processing section. Class balancing was implemented through weighted sampling during training to address potential imbalances in the health class distribution across different domains.

3. Results and discussion

3.1 Model performance analysis

To assess the effectiveness of the proposed DG approach, we conducted a leave-one-site-out evaluation across three study areas: Ruokolahti, Koli, and Paloheinä. In each experimental round, one site was held out exclusively for testing, while the remaining sites, including Evo, were used for training. Evo was

excluded as a test site due to the absence of bark beetle symptom annotations required for canopy health classification evaluation. Model performance was evaluated using multiple metrics including overall classification accuracy, class-wise precision, recall, and F1-score to provide a comprehensive assessment of generalization capabilities across different health categories and environmental conditions.

The results, presented in Table 2 and Figure 3, demonstrate that the DG model consistently outperformed the baseline 2D-CNN across most test scenarios. The DG model achieved substantial improvements in generalization performance, with overall accuracy gains of 9.8% and 27.2% for the Ruokolahti and Koli test sites, respectively. These improvements highlight the effectiveness of feature disentanglement in learning domain-invariant representations that transfer well to unseen environments.

Test site	Class	Baseline 2D-CNN			DG 2D-CNN			Tr; Va; Te		
		Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Paloheinä	Healthy	0.89	0.80	0.85	0.86	0.82	0.84	1091;273;129		
	Infested	0.61	0.77	0.68	0.68	0.70	0.69	32;8;107		
Ruokolahti	Dead	0.97	0.97	0.97	0.94	0.94	0.94	566;142;281		
	Healthy	0.85	0.77	0.81	0.88	0.85	0.86	540;135;818		
	Infested	0.50	0.21	0.30	0.68	0.55	0.61	102;26;19		
	Dead	0.80	0.77	0.78	0.85	0.87	0.86	393;98;498		
Koli	Non-spruce	0.15	0.38	0.22	0.58	0.62	0.60	223;56;72		
	Healthy	0.82	0.56	0.66	0.89	0.84	0.86	800;200;493		
	Infested	0.04	0.48	0.08	0.45	0.57	0.50	101;25;21		
	Dead	0.68	0.68	0.68	0.83	0.85	0.84	623;156;210		
	Non-spruce	0.37	0.22	0.28	0.59	0.63	0.61	161;40;150		

Table 2. Classification performance comparison between baseline 2D-CNN and domain generalization (DG) models across test sites.

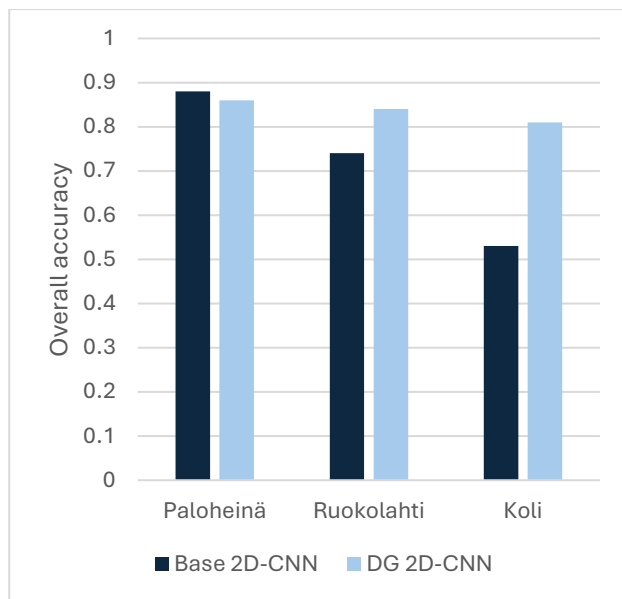


Figure 3. Overall classification accuracy of the baseline and domain generalization models across test sites.

However, the results also reveal important nuances in the DG model's performance. For the Paloheinä test set, the DG model showed a modest decrease in overall accuracy of 2.6% compared to the baseline. This performance pattern suggests that the benefits of domain generalization are most pronounced when there are significant domain shifts between training and test environments, while the approach may introduce minor performance trade-offs in scenarios where the baseline model already achieves near-optimal performance.

Examination of class-wise performance metrics reveals differential impacts of domain generalization across health categories. The DG model demonstrated particularly strong improvements in distinguishing between healthy and infested trees, which is critical for early detection of bark beetle outbreaks. For the infested class, the most challenging category in the baseline model, the DG approach achieved remarkable improvements: precision increased from 0.50 to 0.68 at Ruokolahti and from 0.04 to 0.45 at Koli, while recall improved from 0.21 to 0.55 at Ruokolahti and from 0.48 to 0.57 at Koli. These enhancements are especially valuable given the subtle visual differences that characterize early-stage infestations, where conventional models often struggle due to domain-specific variations.

The dead tree classification also benefited from domain generalization, with improved precision and recall across most test sites. At Ruokolahti, precision improved from 0.80 to 0.85 and recall from 0.77 to 0.87, while at Koli, precision increased from 0.68 to 0.83 and recall from 0.68 to 0.85. For the non-spruce class, performance improvements were substantial at both test sites where this category was present. At Ruokolahti, precision dramatically increased from 0.15 to 0.58 and recall from 0.38 to 0.62, while at Koli, precision improved from 0.37 to 0.59 and recall from 0.22 to 0.63. These improvements reflect the model's enhanced ability to distinguish between different tree species across varying environmental conditions.

The magnitude of performance improvements varied substantially across test sites, providing insights into the conditions under which domain generalization is most beneficial.

The Koli dataset showed the largest performance gains (27.2% improvement in overall accuracy), suggesting that this site presented the most significant domain shift relative to the training data. This substantial improvement can be attributed to Koli's unique environmental characteristics, including pronounced seasonal variations in vegetation phenology and distinctive topographic features that created challenging imaging conditions. The DG model's ability to achieve balanced performance across all classes at Koli demonstrates effective domain-invariant feature learning. The Ruokolahti dataset exhibited moderate but consistent improvements (9.8% increase in overall accuracy), indicating that while domain shifts were present, they were less severe than those encountered at Koli. This intermediate performance gain demonstrates the model's ability to adapt to moderately different environmental conditions while maintaining robust classification performance across all health categories.

The Paloheinä dataset's slight performance decrease (2.6% reduction in overall accuracy) warrants careful interpretation. The baseline model already achieved high accuracy on this dataset (88.6%), suggesting that the environmental conditions at Paloheinä were well-represented in the training data from other sites. In such cases, the additional complexity introduced by domain generalization mechanisms may not provide substantial benefits and could potentially introduce minor performance trade-offs due to the regularization effects of the adversarial training components.

3.2 Limitations

While the proposed DG framework demonstrates substantial improvements in generalization performance, several limitations warrant discussion. The uneven performance gains across sites, particularly the slight decrease on the Paloheinä dataset, highlight the need for more sophisticated approaches that can automatically determine when domain generalization is beneficial versus potentially counterproductive.

Class imbalance presents another significant challenge, as the distribution of health categories varies substantially across sites and domains. The infested class is severely underrepresented in some datasets, potentially biasing the model toward more prevalent conditions and limiting its ability to detect early or subtle symptoms. Future work should investigate specialized sampling strategies or loss functions that can better handle class imbalance in the context of domain generalization.

The seasonal-meteorological domain definition, while capturing important sources of variation, may not fully represent all relevant domain shifts. Geographic factors, for instance, could introduce additional sources of variation that are not explicitly modeled in the current framework. More sophisticated domain definition strategies that can automatically discover relevant domain boundaries from the data could improve generalization performance.

Finally, the current approach focuses primarily on classification accuracy and does not explicitly address other important aspects of forest health monitoring, such as the spatial distribution of symptoms or the temporal dynamics of outbreak progression. Integrating these considerations into the domain generalization framework could enhance its practical utility for operational forest management applications.

4. Conclusions

This study successfully demonstrated the potential of domain generalization techniques to address a critical challenge in operational forest health monitoring: the limited transferability of deep learning models across diverse environmental conditions. By developing a domain generalization framework that incorporates dual feature extractors and adversarial learning, we achieved substantial performance improvements of up to 27% in previously unseen environments, particularly where significant domain shifts existed between training and testing conditions.

The proposed seasonal-meteorological domain definition strategy effectively captured the primary sources of visual variation in aerial forest imagery while maintaining sufficient sample sizes for robust training. The leave-one-site-out evaluation provided compelling evidence that feature disentanglement can successfully separate task-relevant, domain-invariant features from environmental noise.

Our results reveal that the benefits of domain generalization are context-dependent, with the largest improvements observed in test environments that differed substantially from training data. At Koli, where the most significant domain shift occurred, the DG model achieved balanced performance across all health categories, with particularly notable improvements in the challenging infested class. At Ruokolahti, moderate domain shifts resulted in consistent improvements across all classes, with the infested class showing remarkable enhancement. This finding has important practical implications, suggesting that domain generalization techniques should be selectively applied based on anticipated domain shift severity.

While demonstrating clear benefits, this work also revealed important limitations. The slight performance decrease on the Paloheinä dataset highlights that domain generalization may not be universally beneficial when baseline models already achieve high performance.

Future research should focus on adaptive domain generalization approaches that can automatically determine when to apply these techniques, integration with object detection frameworks for spatially explicit symptom identification, and extension to fine-grained symptom classification for early outbreak detection. Multi-modal approaches incorporating additional data sources such as hyperspectral imagery could further enhance robustness and accuracy.

This research contributes to the broader field of remote sensing by demonstrating practical applicability of domain generalization to real-world environmental monitoring challenges. The methods and evaluation frameworks developed here can be adapted to other remote sensing applications facing similar domain shift challenges. By providing both theoretical insights and practical tools, this work represents a significant step toward more robust and scalable deep learning solutions for forest health monitoring, supporting urgent needs for reliable automated systems in the face of increasing climate-related forest disturbances.

Acknowledgements

This research was funded by the European Union within project "Network for novel remote sensing technologies in forest disturbance ecology" (decision no. 101078970) and by the Academy of Finland within project "Learning techniques for autonomous drone based hyperspectral analysis of forest vegetation" (decision no. 357380). This study has been

performed with affiliation to the Academy of Finland Flagship Forest–Human–Machine Interplay—Building Resilience, Redefining Value Networks and Enabling Meaningful Experiences (UNITE) (decision no. 357908).

References

- Blomqvist, M., Kosunen, M., Starr, M., Kantola, T., Holopainen, M., Lyytikäinen-Saarenmaa, P., 2018. Modelling the predisposition of Norway spruce to *Ips typographus* L. infestation by means of environmental factors in southern Finland. *Eur J Forest Res* 137, 675–691. <https://doi.org/10.1007/s10342-018-1133-0>
- Chen, J., Zhang, Z., Li, L., Shahrabi, B., Mishra, A., 2024. Contrastive Adversarial Training for Unsupervised Domain Adaptation. <https://doi.org/10.48550/arXiv.2407.12782>
- Dong, M., Yang, Y., Zeng, K., Wang, Q., Shen, T., 2024. Implicit Sharpness-Aware Minimization for Domain Generalization. *Remote Sensing* 16, 2877. <https://doi.org/10.3390/rs16162877>
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V., 2016. Domain-Adversarial Training of Neural Networks. *Journal of machine learning research*, 1–35.
- Hlásny, T., König, L., Krokene, P., Lindner, M., Montagné-Huck, C., Müller, J., Qin, H., Raffa, K.F., Schelhaas, M.-J., Svoboda, M., Viiri, H., Seidl, R., 2021. Bark Beetle Outbreaks in Europe: State of Knowledge and Ways Forward for Management. *Curr Forestry Rep* 7, 138–165. <https://doi.org/10.1007/s40725-021-00142-x>
- Muandet, K., Balduzzi, D., Schölkopf, B., 2013. Domain Generalization via Invariant Feature Representation. *Proceedings of the 30th International Conference on Machine Learning*, 10–18.
- Patacca, M., Lindner, M., Lucas-Borja, M.E., Cordonnier, T., Fidej, G., Gardiner, B., Hauf, Y., Jasinevičius, G., Labonne, S., Linkevicius, E., Mahnken, M., Milanovic, S., Nabuurs, G., Nagel, T.A., Nikinmaa, L., Panyatov, M., Bercak, R., Seidl, R., Ostrogović Sever, M.Z., Socha, J., Thom, D., Vuletic, D., Zudin, S., Schelhaas, M., 2023. Significant increase in natural disturbance impacts on European forests since 1950. *Global Change Biology* 29, 1359–1376. <https://doi.org/10.1111/gcb.-16531>
- Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y., Herrera, F., 2019. Detection of Fir Trees (*Abies sibirica*) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. *Remote Sensing* 11, 643. <https://doi.org/10.3390/rs11060643>
- Shai, B.-D., Blitzer, J., Crammer, K., Pereira, F., 2006. Analysis of representations for domain adaptation. *Advances in neural information processing systems*.
- Turkulainen, E., Hietala, J., Jormakka, J., Tuviala, J., De Oliveira, R.A., Koivumäki, N., Karila, K., Näsi, R., Suomalainen, J., Pelto-Arvo, M., Lyytikäinen-Saarenmaa, P., Honkavaara, E., 2025. Towards scalable wide area UAS monitoring of forest disturbance using hydrogen powered airships. *International Journal of Remote Sensing* 46, 177–204. <https://doi.org/10.1080/01431161.2024.2399327>

Turkulainen, E., Honkavaara, E., Näsi, R., Oliveira, R.A., Hakala, T., Junttila, S., Karila, K., Koivumäki, N., Peltö-Arvo, M., Tuviala, J., Östersund, M., Pölönen, I., Lyytikäinen-Saarenmaa, P., 2023. Comparison of Deep Neural Networks in the Classification of Bark Beetle-Induced Spruce Damage Using UAS Images. *Remote Sensing* 15, 4928. <https://doi.org/10.3390/rs15204928>

Wang, H., Shen, T., Zhang, W., Duan, L., Mei, T., 2020. Classes Matter: A Fine-grained Adversarial Approach to Cross-domain Semantic Segmentation. <https://doi.org/10.48550/arXiv.2007.09222>

Wu, Z.-Z., Du, C.-J., Wang, X.-Q., Zou, L., Cheng, F., Li, T., Nian, F.-D., Weise, T., Wang, X.-F., 2025. Domain Adaptation via Feature Disentanglement for cross-domain image classification. *Applied Soft Computing* 172, 112868. <https://doi.org/10.1016/j.asoc.2025.112868>

Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2022. Domain Generalization: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 1–20. <https://doi.org/10.1109/TPAMI.2022.3195549>