Vietnam Journal of Earth Sciences, 47(3), 396–410, https://doi.org/10.15625/2615-9783/23402

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Automated dense-layer architecture search on EfficientNet: A hybrid approach for scene-based land-cover classification

Le-Tuan Pham[1], Bui Vu Vinh[2], Van-Manh Pham[1], Nguyen Thi Lan Anh[3], Quoc-Huy Nguyen[1], Huu Duy Nguyen[1], Quang-Thanh Bui[1*]

[1]*VNU University of Science, Vietnam National University, Hanoi, Vietnam*
[2]*Department of Science and Technology, Khanh Hoa province*
[3]*Hanoi Architectural University*

Received 04 March 2025; Received in revised form 26 June 2025; Accepted 05 September 2025

ABSTRACT

This article proposes a land cover classification framework based on EfficientNet-B4, a model from the EfficientNet family of convolutional neural networks developed by Google AI. EfficientNet models have been widely applied in various domains, including image classification, object detection, and medical imaging, due to their scalability and efficiency. EfficientNet-B4 and its pretrained weights were used for feature extractions, and the dense layer structure (number of layers, nodes, and dropout rates) was tuned using meta-heuristic optimisation algorithms. The model is trained and validated on the Sentinel-2 EuroSAT benchmark, which comprises 27,000 RGB (Red-Green-Blue) image tiles spanning 10 land-cover classes across Europe. The results show that this proposed model achieves an overall classification accuracy of 0.9881 for RGB images, which is higher than that of previous networks using similar datasets. This hybrid approach can be considered as an alternative solution to search for neural architectures for different applications.

Keywords: Land cover, efficient network, metaheuristic optimization.

1. Introduction

Machine learning methods have been developed to automate analysis and facilitate improvements in remote sensing observations by introducing new model network structures (Ma et al., 2019). These methods have become particularly effective with the emergence of extremely high spatial resolution data, including satellite, airborne, and drone data (Hoan et al., 2025; Nikparvar and Thill, 2021). The advent of deep learning (DL), also known as deep neural networks (DNN), has sparked a resurgence in interest in neural networks. DL, DNN has demonstrated incredible abilities that are primarily attributed to the automated extraction of critical features; this eliminates the need to identify specific cases. The driving force behind the success of DL, specifically DNNs, in image analysis can be attributed to the following three significant factors. First, more data is available for training DNN, especially in the case of supervised learning, such as classification (Bui et al., 2021; Dang Vu et al.,

*Corresponding author, Email: thanhbq@vnu.edu.vn

2025; Dinh and Vinh, 2021; Hang et al., 2017; Nguyen Thanh et al., 2025). Recently, efforts to create benchmark data have been underway, with the provision of freely available databases for tasks such as image classification, object detection, semantic segmentation, and change detection, to evaluate the performance of algorithms and techniques in remote sensing (Ma et al., 2019). These datasets typically contain various types of imagery, such as optical, radar, or multispectral images, along with ground-truth annotations (Gąsienica-Józkowy et al., 2021; Mohajerani and Saeedi, 2021; Sumbul et al., 2019; Weir et al., 2019). Ground-truth annotations provide information about objects or features in the images, such as land cover types, building footprints, or crop types. They allow researchers to compare the performance of different methods under standardised conditions and facilitate the development of new techniques. Some were designed for global scales, some for regional scales, and most for specific applications. In this regard, spatial analysts from developing countries face challenges of limited ground truth data, and models trained using European datasets might not work well in tropical countries, where land surface patterns and spatial arrangements are different. Secondly, more computational power is available, precisely the increasing number of Graphical Processing Units optimised for processing parallelisable problems like training DNNs. Third, more advanced algorithms have been developed, which have led to a significant increase in the depth and number of outputs. This increased performance was then transferred to DNNs, which resulted in their success (Adegun et al., 2023; Alzubaidi et al., 2021; Sarker, 2021; Vu et al., 2025).

Convolutional neural networks (CNN) are widely used methods that have been successfully developed for pattern recognition, natural language processing, landcover classification, and point cloud dataset processing. As they are more efficient in processing large datasets, CNNs are relevant and valuable for analyzing remotely sensed imagery. For example, CNN has been used in various applications to classify high-spatial-resolution datasets, including land cover classification and object detection (Scarpa et al., 2018; Srivastava et al., 2019; Tsagkatakis et al., 2019; Tuna et al., 2018; Wang et al., 2017; Zhou et al., 2017), and for annotation of the point cloud dataset (Hu and Yuan, 2016). For disaster studies, a systematic comparison of this method was investigated in the work of Wang et al. (2019), where a 1D, 2D, and 3D convolutional neural network (CNN) was utilized to map landslide susceptibility. CNN was also used in fire management surveillance (Muhammad et al., 2018). The choice of neural network architecture plays a crucial role in the performance of deep learning models. Different tasks or data sets may require architectures with varying depths, widths, and connectivity patterns to achieve optimal performance. The neural architecture search (NAS) aims to find architectures that maximise performance metrics such as accuracy, precision, or recall. Besides, hand-designed architectures may overfit specific datasets or tasks, limiting their generalisation ability to unseen data. NAS seeks to discover architectures that generalise well across different datasets and tasks, leading to more robust models. NAS automates this process, saving significant human effort and potentially discovering architectures more efficient or effective than hand-designed ones.

New CNN structures have been continuously developed through NAS algorithms, including reinforcement learning-based, evolutionary Algorithms, gradient-based, and metaheuristic optimisation methods to iteratively sample and evaluate architectures based on their performance on a validation set (Ahmad et al., 2020; Alzubaidi et al., 2021; Hoang and Jo, 2021; Kaveh and Mesgari, 2022). Meta-heuristic optimisation algorithms can be used to optimise the parameters of classification models, including deep learning architectures (Devikanniga et al., 2019; Muazu et al., 2022; Priyanka and Kumar, 2020). These optimisation techniques can help improve the performance of deep learning models by fine-tuning their hyperparameters or optimising the network structure to better fit the data's characteristics. This study proposes a hybrid deep learning framework that integrates EfficientNet-based convolutional feature extraction with densely connected layers optimized through a meta-heuristic algorithm to improve land cover classification accuracy. The primary objective of the research was to utilize the high representational capacity of EfficientNet to extract spatial spectral characteristics from Sentinel-2 imagery and to automatically determine the optimal configuration of dense classification layers, including the number of neurons, layers, and dropout rates, using an evolutionary optimization approach. Lastly, it aims to evaluate the effectiveness of this combined architecture on the EuroSAT benchmark dataset in comparison to conventional CNN-based methods (Helber et al., 2018).

2. Data and Methods

2.1. Datasets

In this study, we used the EuroSAT dataset, a well-known benchmark created by Helber et al. (2018), which is specifically designed for land cover classification tasks using satellite imagery. The dataset contains 27,000 labeled image patches, each with a resolution of 64 × 64 pixels, derived from Sentinel-2 satellite images. These images are available in two formats: a standard RGB version (using Sentinel-2 bands 4, 3, and 2) and a more detailed version that includes all 13 spectral bands provided by the satellite. Each image in the dataset is labelled with one of ten land cover classes, representing a range of natural and human-made environments. These include annual crops, forests, herbaceous vegetation, highways, permanent crops, residential areas, rivers, sea/lakes, industrial buildings, and pastures. Data were collected from various regions in Europe, with a significant portion originating from France and Germany, providing both diversity and consistency for training and evaluation (Fig. 1). This dataset has been utilized in numerous studies (Pham and Bui, 2021; Tsagkatakis et al., 2019), where it has enabled researchers to evaluate the performance of various machine learning and deep learning approaches, yielding strong and consistent results. In this study, we used the RGB images for our experiments, as they allow us to use pre-trained deep learning models that are commonly trained on natural image datasets such as ImageNet. This makes it easier to apply transfer learning techniques while still capturing useful spatial patterns for classification.

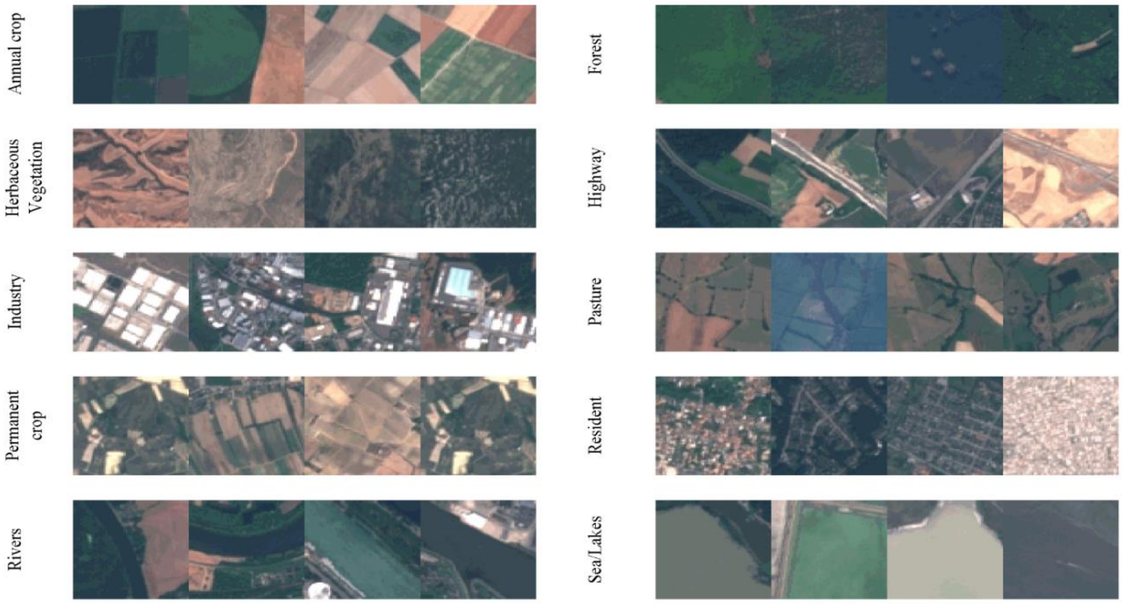

Figure 1. Samples of ten classes from the Eurosat dataset. (Helber et al., 2018). The dataset is available for download at https://www.kaggle.com/datasets/apollo2506/eurosat-dataset

2.2. Methods

2.2.1. Methodological framework

Figure illustrates the proposed EfficientNetB4-based architecture, where the feature extraction block is derived from EfficientNetB4. The weights were pre-trained using ImageNet data and downloaded from the data portal. In this study, we did not train the proposed model from scratch (using Eurosat) because the EfficientNetB4 feature extraction component was trained to learn generic features from large amounts of data. Instead of using the classification block (fully connected layer) of EfficientNet, we replaced it (while keeping other blocks frozen) with a custom block comprising layers with tunable nodes and dropout layers with tunable rates. We optimized this block and fine-tuned the pre-trained model on Eurosat to adapt the features to the land cover classification task, while keeping the earlier layers frozen to retain the generic features learned during pre-training. A similar fine-tuning of this dense layer can be found in the work of Bui et al. (2021), Nadeem et al. (2023), and Hakim et al. (2022). More justifications for EfficientNet and meta-heuristic optimisation algorithms are in the following sections.

2.2.2. EfficientNet

EfficientNet is a family of convolutional neural network architectures designed to achieve high performance with fewer parameters and computational resources than traditional CNN models, since the introduction of this network by Tan and Le (2019). Several versions of EfficientNet have been built from a base network with different compound scaling (combining three strategies: depth, width, and resolution scaling). By carefully balancing these dimensions, EfficientNet achieves better performance while being more computationally efficient than other CNN architectures (Alhichri et al., 2021; Hoang and Jo, 2021; Punuri et al., 2023). Compound scaling is controlled by a single scaling parameter, denoted phi, which scales the model's width, depth, and resolution. Since then, the networks have been compared with various state-of-the-art models, such as ResNet, AlexNet, and GoogleNet, using a reference dataset, and the performance has been robust. In this study, we used the B4 version.

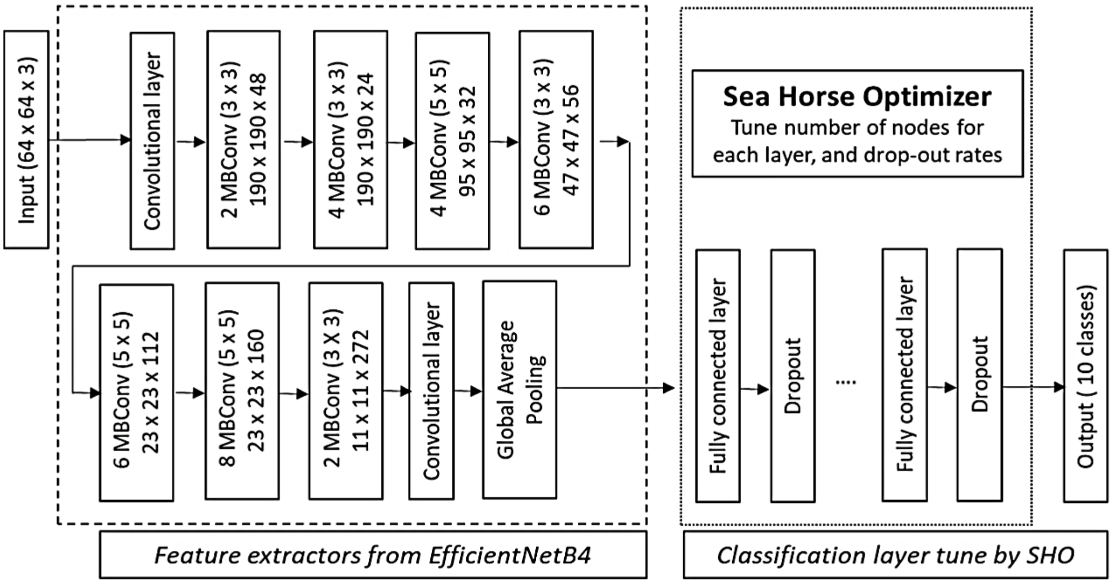

Figure 2. Tunable architecture

2.2.3. Meta-Heuristic Algorithms for Network Architecture Search

Metaheuristic optimisation algorithms are iterative methods that mimic natural processes such as evolution, swarm behaviour, or simulated annealing (Bui et al., 2018; Challapalli and Devarakonda, 2022). These algorithms do not guarantee finding the globally optimal solution, but they can often find a near-optimal solution in a reasonable time. Among all approaches, the swarm-based algorithm is considered the most robust method, as many novel algorithms have been proposed (Kaveh and Mesgari, 2022). Interactions between swarm populations during the search for food/targets form the fundamental principles of these algorithms. Individuals' positions, directions, and values are updated according to the most optimal agent (depending on how signals are propagated in the population). The process continues until the target values are reached. This mechanism helps the swarm find desirable food locations or targets in the search space. Implementing optimisation algorithms (as a replacement for gradient descent) in searching for optimal hyperparameters and weights of deep learning models is worth investigating and can be a potential alternative solution. However, it is challenging (Guo et al., 2019) and still rare (Rere et al., 2016).

Seahorse optimisation (SHO) is a nature-inspired algorithm based on seahorse behaviour and unique marine creatures known for their distinctive swimming style (Zhao et al., 2023). The algorithm mimics the movement of sea horses and social interactions to solve optimization problems in engineering, computer science, and other fields. Seahorse optimisation has shown promising results in solving constrained optimisation problems and has been compared with popular optimisation algorithms, such as genetic algorithms, particle swarm optimisation, and differential evolution. As with any optimisation algorithm, its performance may vary depending on the nature of the problem being solved. Like many other population-based optimisers, the mathematical background of this algorithm can be summarised as follows. A swarm of sea horses is initiated in an n-dimensional search space, where each dimension represents a tunable feature. Swarm movements for food search and predator escape include slow swimming and rapid darting. During this process, seahorses interact, share information, and learn from each other, thereby enhancing their overall exploration and exploitation capabilities. Each agent stores a personal best position in its memory, allowing it to recall successful movements and avoid previously visited locations. This feature is incorporated into the algorithm to prevent the search from re-visiting the exact solutions. The system has held the best global position so far.

3. Results

The SHO was started, in which the swarm population is 10 (Fig. 3), and the process is iterated 25 times (due to the memory limitation of the training computer). Each seahorse agent moves in the 16-dimensional search space, and each dimension represents a value for either layer nodes or dropout rate. The fitness function that estimates the agent's value is the model proposed in (Fig. 2), in which the agent's fitness value is updated when the agent's position changes (or when the dense layers and dropout rates are adjusted). The Eurosat data set was used for fitness estimations, and it was randomly divided into training (70%), validation (15%), and test (15%) folds. For each iteration, the training/validation data set was fed into the proposed model (with a dense layer structure defined by SHO) (Fig. 2) and trained for five epochs. In many previous studies, this training step is limited to one epoch. (Ahmad et al., 2020). However, land cover is complex, so the epoch should be higher. Subsequently, it was tested with the test data set for loss estimation and test accuracy. Technical loss was used as an objective function of the SHO to minimise loss and improve accuracy. After 25 iterations, the position of the best global agent, which results in the highest accuracy, represents the structure of dense layers and the dropout rate. After this step, we have the model with the highest precision of 25 iterations, each of which we run for five epochs (Figure 2 4).

For binary classification, losses and accuracies typically move in parallel, but in opposite directions. In some iterations, the loss decreased, but the accuracy decreased or was in reverse order. Because accuracy is essential in land cover classification, we saved structures with which the model received the global minimum loss (Model 1) or the global highest accuracy (Model 2) during the 25 iterations.

In the next step, we combined the training and validation sets into a single, more extensive training set and refitted Model 1 and Model 2 at 100 epochs (without using validation data). Model 1 and Model 2 achieved test accuracies of 0.9869 and 0.9881, respectively (Fig. 5), which are higher than those reported in previous studies using the same Eurosat data (Naushad et al., 2021). The curve variations from (Fig. 4) have strong fluctuations since they illustrate losses and accuracies of different model structures (specifically, fully connected layer structures). The pattern in (Fig. 5) looks smoother, since the chosen model was fine-tuned, and variations are minor. Figure 6 shows the proposed model structure, in which the classification layer was optimised during training.

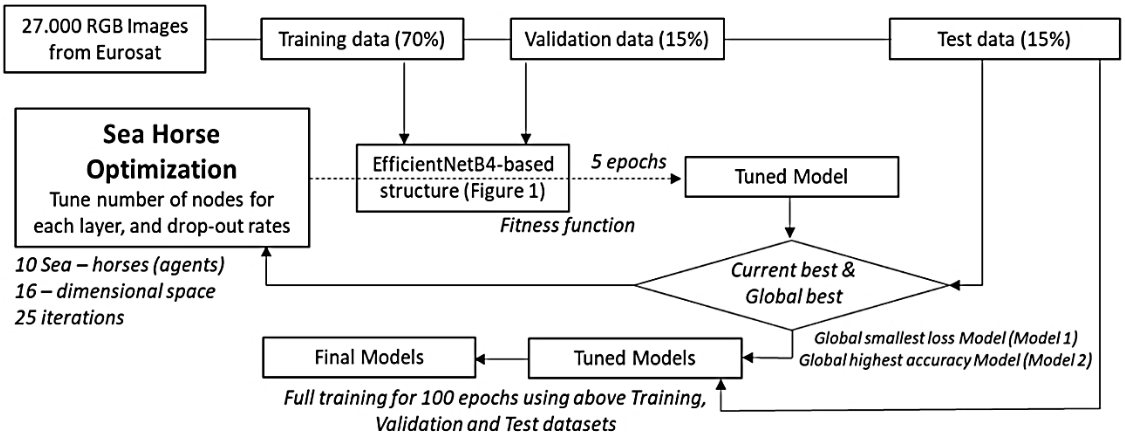

Figure 1. Proposed workflow. This process consists of two main steps. The first explores potential structures of the fully connected layer, and each candidate structure is trained for five epochs. The second step retrains the model from the previous step with more epochs (100 in this case)

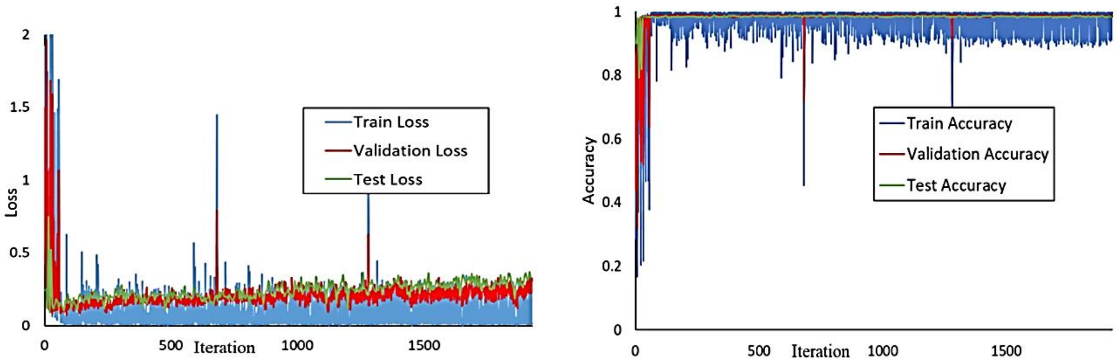

Figure 2. Variations in accuracy and loss of training and validation data sets. SHO iterates 25 times. For each iteration, the fitness of the seahorse agents was estimated 15 times. For each estimation, we run the EfficientNetB4-based model for five epochs. Including the initiation process, we have 1930 values for losses and accuracies. We saved structures, resulting in the minimal loss or the highest accuracy, as in Model 1 and Model 2, respectively

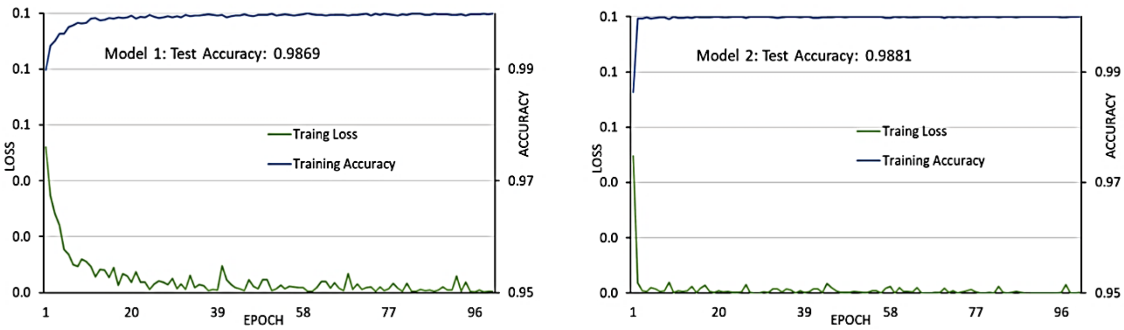

Figure 5. Training losses and accuracy of Model 1 and Model 2 during 100 epochs. Two models were fully trained using a combination of training and validation sets as a training data set. In this step, no validation was performed, and the final models yielded accuracy values of 0.9869 and 0.9881 for the test set

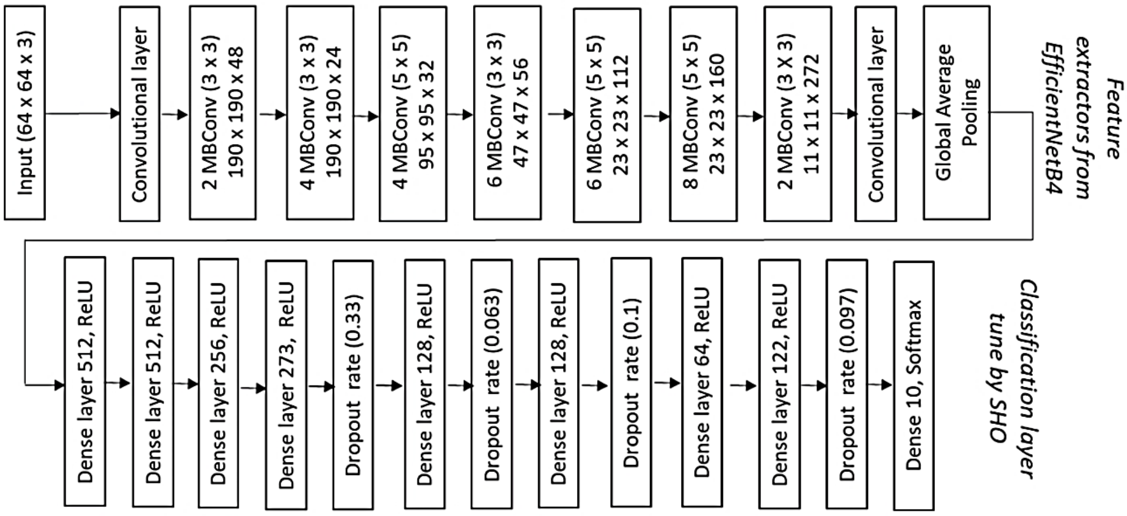

Figure 6. The final EfficientNetB4-based structure (Model 2) for the classification of landscape cover based on scenes. This model was fully trained and received a test accuracy of 0.9881

We used the EuroSAT dataset, which is specifically designed to ensure a balanced distribution across 10 land cover classes. Due to this balance, the trained models achieved exceptionally high classification accuracies of 0.9869 and 0.9881, comparable to the performance of state-of-the-art deep learning models reported in previous studies (Helber et al., 2018). Given these results, the estimated precision, recall, and F1 scores are approximately 0.986 for Model 1 and 0.988 for Model 2, indicating highly consistent and robust performance across all classes.

The proposed EfficientNetB4-SHO model showed consistent and reliable classification results when tested on Sentinel-2 images from different regions of Vietnam. We examined representative areas covering delta plains, midlands, urban centres, coastal zones, and mountainous terrains, and found that the model maintained high accuracy in all cases. It was able to correctly identify the main types present in each landscape (Fig. 7). These qualitative results align with the quantitative evaluation on the EuroSAT benchmark, where the precision, recall, and F1-scores were consistently high (0.986–0.988) across all categories. Notably, many EuroSAT classes correspond directly to land cover types common in Vietnam, such as croplands, forests, urban areas, and water bodies. In our tests, these categories were recognised with high confidence, confirming the model's ability to generalise beyond its original European training data. Although a few errors occurred due to unique local patterns (for example, tropical textures or mixed land-use mosaics not represented in EuroSAT), the predictions remained accurate for most scenes. Overall, this evidence demonstrates that the EfficientNetB4–SHO model is robust, adaptable, and well-suited to mapping across Vietnam's diverse environments.

The model's ability to accurately recognize common land-cover types-such as annual crops, plantations, urban areas, and open water is auspicious for practical applications in Vietnam. When a model performs reliably on these essential classes, it can be confidently scaled up for operational use. Our findings indicate that this hybrid CNN optimiser approach has the potential to support nationwide geospatial monitoring programmes. For example, its high accuracy in land-cover classification can provide timely updates on crop distribution and forest cover, supporting agricultural management and land resource monitoring. Urban planners could also benefit from automated mapping of residential and infrastructure areas, helping to guide development strategies and monitor urban expansion.

Furthermore, the consistent reliability across diverse Vietnamese landscapes suggests strong potential in disaster risk management. Accurate land cover maps are essential inputs to forecast flood risks in delta regions, landslide hazards in mountainous areas, and spread in forested zones. By providing reliable and consistent results, the EfficientNetB4–SHO model can become a valuable tool for sustainable land management and disaster preparedness in Vietnam, demonstrating an applicability that extends well beyond a single region or data set.

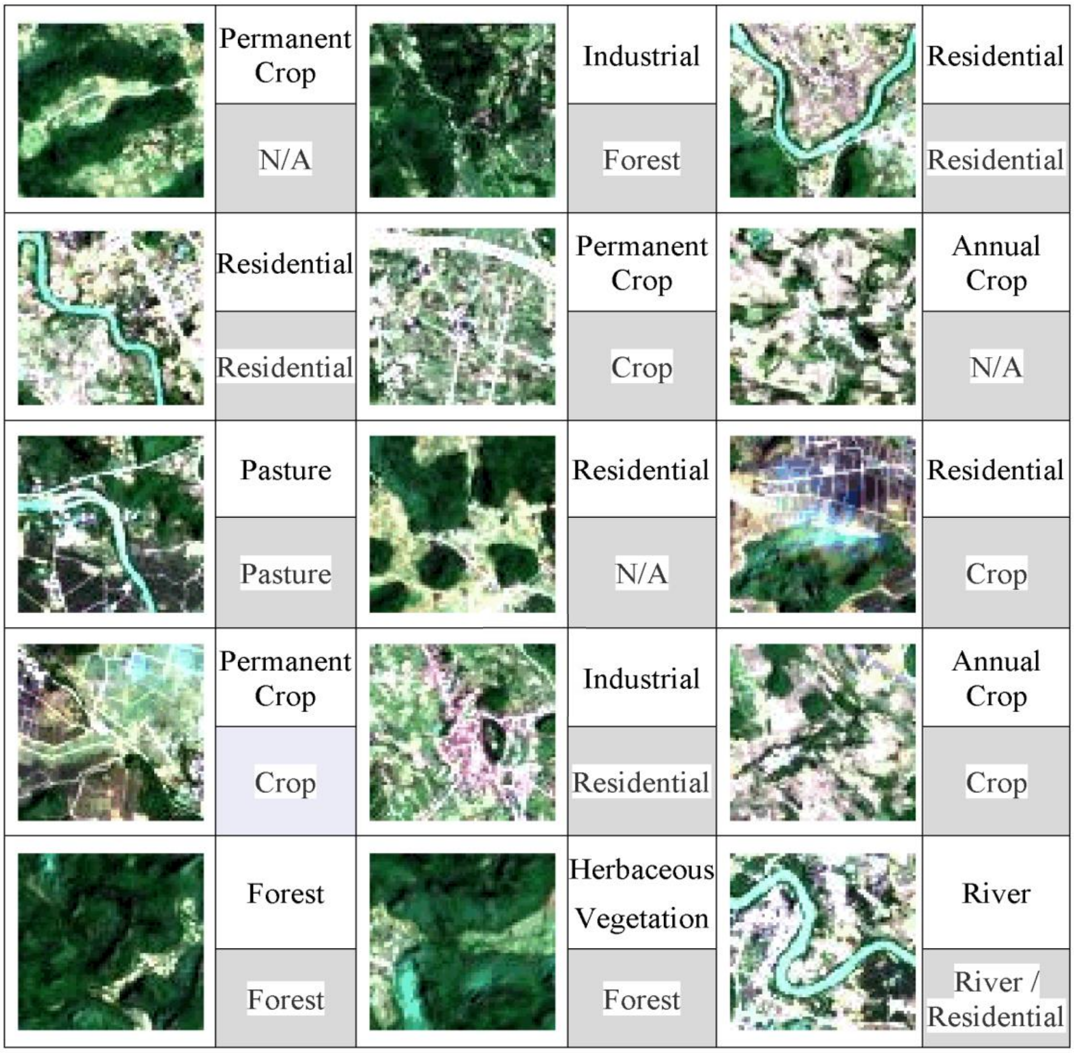

Figure 7. Some scene classifications in northeastern Vietnam use a trained model from Eurosat. The upper sections for each image represent the predicted classes, and the lower sections (shaded) represent potentially manually entered classes

4. Discussions

EfficientNets represent a significant leap in deep learning architecture, offering a range of versions optimised for various performance metrics and computational efficiencies. Kashevnik and Ali (2022) conducted a study on EfficientNet B3, affirming its ability to balance high performance and efficient runtime. This version has proven effective across different tasks where minimising computational resources while maintaining competitive accuracy is crucial. Similarly, (Hao et al., 2021) compared EfficientNet B4 directly with ResNet50 in a specialised task of identifying medicinal leaves. Their findings underscored EfficientNet B4's superior performance in complex image recognition scenarios, highlighting its potential in niche applications that require robust classification capabilities. The versatility of EfficientNets extends to land cover classification tasks that use RGB imagery, as exemplified in various studies. Although RGB inputs suffice for many applications, leveraging additional spectral bands from sources such as Sentinel 2A data (which includes 13 bands capturing diverse wavelengths) could theoretically enhance classification accuracy by providing richer data for analysis. However, this enhancement comes at the cost of increased computational complexity, necessitating careful consideration in practical deployments where resource constraints are a concern.

Hyperparameter Optimisation (HSO) techniques are pivotal in optimising EfficientNets for specific tasks. Recent studies have explored HSO's efficacy in fine-tuning the fully connected layers of EfficientNets, showcasing promising results without direct adjustments to the convolutional layers. This approach highlights the potential to achieve significant performance gains through targeted optimisation strategies tailored to the unique architecture of EfficientNets. Looking ahead, the landscape of deep learning architecture continues to evolve with the introduction of novel algorithms designed to enhance model efficiency and effectiveness. The ongoing pursuit of optimising algorithms for comprehensive network architecture searches remains critical. By continually refining these methodologies, researchers can unlock the full potential of EfficientNets across a broader spectrum of applications. This includes improving classification accuracy in image-based tasks and extending their utility to new domains and challenges. Efforts to optimise EfficientNets are particularly pertinent in light of the increasing complexity and scale of modern datasets. As datasets become more extensive and diverse, efficient processing and extracting meaningful insights become paramount. EfficientNets offers a compelling solution by prioritising computational efficiency without compromising performance metrics, thereby addressing practical concerns in real-world applications.

Moreover, as more sophisticated algorithms emerge, there is a growing need to explore their integration within the framework of EfficientNets. This involves adapting existing architectures and innovating new algorithmic enhancement and adaptation methodologies. By fostering collaboration between deep learning researchers and domain experts, novel approaches can be systematically integrated and validated, ensuring that EfficientNets remains at the forefront of technological advancement.

The recent surge in satellite data sets on different spatial scales represents a significant advancement in improving the performance and applicability of existing deep learning models across diverse geographical regions. A notable dataset, Eurosat, compiled from multiple European countries, provides labelled classes corresponding to typical land cover types specific to the area. Deep-learning models trained on Eurosat demonstrate robust performance in recognising land surface patterns and textures characteristic of Europe. However, challenges arise when applying these models to other regions, mainly tropical areas such as northern and eastern Vietnam, as observed in our study using Sentinel 2 imagery from 2020. In these regions, Model 2, as shown in Fig. 6, exhibited instances of misclassification or labelling discrepancies. Misclassification occurs when the model assigns incorrect labels to land cover types, possibly due to subtle variations in surface characteristics that differ from those in Europe. Alternatively, unmatched labels occur when certain land cover in Vietnam are either absent from the Eurosat data set or have distinct spatial arrangements not captured by the European training data. To mitigate these challenges and improve global applicability, integrating multiple data sets from various regions becomes imperative, as illustrated in Fig. 7. By combining data sets that encompass diverse geographical contexts, such as those of Europe and Southeast Asia, researchers can develop more robust deep-learning models capable of accurately identifying and classifying land cover types on a global scale. This approach improves the accuracy and reliability of land cover mapping. It fosters a deeper understanding of regional variations in land surface characteristics, which are crucial for environmental monitoring, urban planning, and worldwide biodiversity conservation efforts.

Integrating metaheuristic optimisers into pre-trained deep learning models represents a significant advancement in optimising CNNs for land cover classification and other complex tasks. Meta-heuristic algorithms, including genetic algorithms, particle swarm optimisation, and simulated annealing, offer robust solutions for hyperparameter tuning in CNNs, particularly when traditional methods face challenges in navigating large search spaces or balancing competing objectives. Researchers can effectively fine-tune hyperparameters such as learning rates, batch sizes, dropout rates, and layer configurations by leveraging metaheuristic techniques to enhance model performance and generalisation capabilities. However, successfully implementing meta-heuristic optimisers requires careful consideration of several challenges. One primary concern is selecting and optimising the metaheuristic's parameters, including population size, mutation rate, crossover probability, and convergence criteria. These parameters significantly influence the effectiveness, convergence speed, and quality of the optimisation process. Finding an optimal configuration for these meta-parameters often involves iterative experimentation and sensitivity analysis to strike a balance between the exploration and exploitation phases of the search.

Studies by Bui et al. (2024) and Chou et al. (2021) underscore the importance of parameter tuning in metaheuristic optimisation, emphasising methodologies for enhancing CNN performance across various applications. They highlight the nuanced approach required to adapt meta-heuristic algorithms effectively to the specific characteristics of deep learning models and datasets. Moreover, searching for parameter-free or self-adaptive metaheuristic optimisers remains a compelling area of research. This pursuit aims to develop algorithms capable of autonomously adjusting their parameters based on the problem's dynamics and optimisation progress, thereby reducing manual intervention and improving algorithm robustness across different scenarios.

In practical terms, integrating a metaheuristic optimiser into a pre-trained deep learning model involves several steps. Initially, the pre-trained model, typically trained on a large dataset like ImageNet, serves as a feature extractor or a starting point. The meta-heuristic optimiser then iteratively explores the hyperparameter space, evaluating different configurations to optimise the model's performance metrics. During this process, the optimizer adjusts hyperparameters based on feedback from model performance metrics, aiming to find an optimal set that maximizes classification accuracy while minimizing overfitting and computational costs. Furthermore, evaluating the efficacy of metaheuristic optimisation requires rigorous experimentation and comparison with baseline approaches, such as grid search or random search. Benchmarking against traditional methods helps quantify the improvements in convergence speed, solution quality, and computational efficiency of metaheuristic algorithms. Additionally, sensitivity analysis and convergence diagnostics provide insights into the metaheuristic's behaviour and guide further refinements in parameter tuning and algorithmic implementation.

5. Conclusions

This study aims to verify the integration of a meta-heuristic optimiser into a pre-trained deep learning model for scene-based land cover classification. We proposed a hybrid model based on EfficientNetB4 for feature extraction, combined with dense, tunable layers, for scene-based land cover classification. The Sea Horse Optimizer tuned the fully connected layer structure using the Sentinel 2-derived Eurosat dataset (RGB images). The proposed model yields an overall accuracy of 0.9881, which is slightly higher than that of previous studies using a similar dataset (OA ~ 0.987).

Nonetheless, the proposed EfficientNetB4-SHO model exhibited robust generalization and reliability when applied to Sentinel-2 imagery from Vietnam. Despite being trained exclusively on the EuroSAT benchmark, the model maintained high classification accuracy on Vietnamese test samples, underscoring its strong generalization capability. Although some minor misclassifications were observed for land cover features not represented in the EuroSAT training set, the model correctly identified the majority of typical land cover categories prevalent in Vietnam, including cropland, urban areas, and water bodies, even across diverse terrain ranging from lowland plains to hilly midlands and mountainous regions. This reliable performance across varied landscapes suggests that the EfficientNetB4-based framework can be confidently applied in different areas of Vietnam without extensive retraining, providing consistently accurate land cover classification results.

Performance can be improved if all versions of EfficientNet and different fully connected layouts are verified. Therefore, the search for NAS using the hybrid approach is worth investigating and can be used with the rapid emergence of optimisation algorithms and a variety of open pixel-based, object-based, and scene-based labelled datasets that cover more significant regions globally for the generalisation of the method.

Ground truth annotations offer insights into image features like land cover, buildings, and crops. They enable method comparison and technique advancement under standardised conditions. In this study, the trained models perform well with the benchmark dataset. Their performance using Sentinel 2 samples from Vietnam results in some misclassification or mismatching because of unknown land surface patterns and textures. It could be seen that the benchmark dataset varies globally, posing challenges for analysts in developing nations. Models trained on European datasets may struggle in tropical regions due to distinct land surface characteristics, emphasising the importance of localised data for practical analysis.

Acknowledgements

This study was funded by the Ministry of Science and Technology of Vietnam (MOST) under the Programme on the Development of Basic Sciences in chemistry, life sciences, earth sciences, and marine sciences during 2017-2025. The grant number is ĐTĐL.CN-42/23.

References

Adegun A.A., Viriri S., Tapamo J.-R., 2023. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. Journal of Big Data, 10, 93. Doi: 10.1186/s40537-023-00772-x.

Ahmad M., Abdullah M., Moon H., Yoo S.J., Han D., 2020. Image Classification Based on Automatic Neural Architecture Search Using Binary Crow Search Algorithm. IEEE Access, 8, 189891–189912. Doi: 10.1109/access.2020.3031599.

Alhichri H., Alswayed A.S., Bazi Y., Ammour N., Alajlan N.A., 2021. Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model With Attention. IEEE Access, 9, 14078–14094. Doi: 10.1109/access.2021.3051085.

Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, 53. Doi: 10.1186/s40537-021-00444-8.

Bui Q.-T., Chou T.-Y., Hoang T.-V., Fang Y.-M., Mu C.-Y., Huang P.-H., Pham V.-D., Nguyen Q.-H., Anh D.T., Pham V.-M., Meadows M.E., 2021. Gradient Boosting Machine and Object-Based CNN for Land Cover Classification. Remote Sensing, 13(14), 2709. https://doi.org/10.3390/rs13142709.

Bui Q.-T., Pham Van M., Hang N.T.T., Nguyen Q.-H., Linh N.X., Hai P.M., Tuan T.A., Van Cu P., 2018. Hybrid model to optimize object-based land cover classification by meta-heuristic algorithm: an example for supporting urban management in Ha Noi, Viet Nam. International Journal of Digital Earth, 1–15. Doi: 10.1080/17538947.2018.1542039.

Challapalli J.R., Devarakonda N., 2022. A novel approach for optimization of convolution neural network with hybrid particle swarm and grey wolf algorithm for classification of Indian classical dances. Knowl Inf Syst, 64, 2411–2434. Doi: 10.1007/s10115-022-01707-3.

Dang Vu K., Ta Duc H., Ngo Thi Hai Y., Nguyen Khac A., Pham Thanh H., Weber C., 2025. Observation of coal mining activities in Ha Long and Cam Pha cities, Quang Ninh province, Vietnam using Sentinel-2 data. Vietnam Journal of Earth Sciences, 47(2), 272–289. Doi: 10.15625/2615-9783/23010.

Devikanniga D., Vetrivel K., Badrinath N., 2019. Review of Meta-Heuristic Optimization based Artificial Neural Networks and its Applications. Journal of Physics: Conference Series, 1362. Doi: 10.1088/1742-6596/1362/1/012074.

Dinh D.N., Vinh C.L., 2021. 30-year changes of natural forests under human activities in the Indochina peninsula-case studies in Cambodia, Laos and Vietnam. Vietnam J. of Earth Sciences, 43(3), 285–300. https://doi.org/10.15625/2615-9783/16196.

Gąsienica-Józkowy J., Knapik M., Cyganek B., 2021. An ensemble deep learning method with optimized weights for drone-based water rescue and surveillance. Integrated Computer-Aided Engineering, 28, 221–235. Doi: 10.3233/ICA-210649.

Guo B., Hu J., Wu W., Peng Q., Wu F., 2019. The Tabu_Genetic Algorithm: A Novel Method for Hyper-Parameter Optimization of Learning Algorithms. Electronics, 8. Doi: 10.3390/electronics8050579.

Hoang V.T., Jo K.H., 2021. Practical Analysis on Architecture of EfficientNet. 2021 14th International Conference on Human System Interaction (HSI), 1–4.

Hu X., Yuan Y., 2016. Deep-Learning-Based Classification for DTM Extraction from ALS Point Cloud. Remote Sensing, 8. Doi: 10.3390/rs8090730.

Kashevnik A., Ali A., 2022. 3D Vehicle Detection and Segmentation Based on EfficientNetB3 and CenterNet Residual Blocks. Sensors, 8(9), 730. https://doi.org/10.3390/rs8090730.

Kaveh M., Mesgari M.S., 2022. Application of Meta-Heuristic Algorithms for Training Neural Networks and Deep Learning Architectures: A Comprehensive Review. Neural Process Lett, 1–104. Doi: 10.1007/s11063-022-11055-6.

Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Marcos D., Volpi M., Kellenberger B., Tuia D., 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. ISPRS Journal of Photogrammetry and Remote Sensing, 145, 96–107. https://doi.org/10.1016/j.isprsjprs.2018.01.021.

Mohajerani S., Saeedi P., 2021. Cloud and Cloud Shadow Segmentation for Remote Sensing Imagery Via Filtered Jaccard Loss Function and Parametric Augmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 4254–4266. Doi: 10.1109/JSTARS.2021.3070786.

Muazu A.A., Hashim A.S., Sarlan A., 2022. Review of Nature Inspired Metaheuristic Algorithm Selection for Combinatorial t-Way Testing. IEEE Access, 10, 27404–27431. Doi: 10.1109/access.2022.3157400.

Muhammad K., Ahmad J., Baik S.W., 2018. Early fire detection using convolutional neural networks during surveillance for effective disaster management. Neurocomputing, 288, 30–42. https://doi.org/10.1016/j.neucom.2017.04.083.

Nadeem M.I., Ahmed K., Li D., Zheng Z., Naheed H., Muaad A.Y., Alqarafi A., Abdel Hameed H., 2023. SHO-CNN: A Metaheuristic Optimization of a Convolutional Neural Network for Multi-Label News Classification. Electronics, 12(1), 113. https://doi.org/10.3390/electronics12010113.

Naushad R., Kaur T., Ghaderpour E., 2021. Deep Transfer Learning for Land Use and Land Cover Classification: A Comparative Study. Sensors (Basel), 21(23), 8083. https://doi.org/10.3390/s21238083

Nguyen Thanh H., Ho Le T., Nguyen Van D., Hoa Thuy Q., Nguyen Kim A., Le Duc H., Pham Van D., Phan Trong T., Tran Van P., 2025. Forest cover change mapping based on Deep Neuron Network, GIS, and High-resolution Imagery. Vietnam Journal of Earth Sciences, 47(2), 157–181. Doi: 10.15625/2615-9783/22192.

Nikparvar B., Thill J.-C., 2021. Machine Learning of Spatial Data. ISPRS International Journal of Geo-Information, 10(9), 600. https://doi.org/10.3390/ijgi10090600.

Pham V.-D., Bui Q.-T., 2021. Spatial resolution enhancement method for Landsat imagery using a Generative Adversarial Network. Remote Sensing Letters, 12, 654–665. Doi: 10.1080/2150704X.2021.1918789.

Priyanka P., Kumar D., 2020. Meta-Heuristic Optimization Based Convolutional Neural Network for Medical Image Classification: A Survey. 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 752–756.

Punuri S.B., Kuanar S.K., Kolhar M., Mishra T.K., Alameen A., Mohapatra H., Mishra S.R., 2023. Efficient Net-XGBoost: An Implementation for Facial Emotion Recognition Using Transfer Learning. Mathematics, 11(3), 776. https://doi.org/10.3390/math11030776.

Rere L.M.R., Fanany M.I., Arymurthy A.M., 2016. Metaheuristic Algorithms for Convolution Neural Network. Comput Intell Neurosci, 1537325–1537325. Doi: 10.1155/2016/1537325.

Sarker I.H., 2021. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Computer Science, 2, 420. Doi: 10.1007/s42979-021-00815-1.

Scarpa G., Gargiulo M., Mazza A., Gaetano R., 2018. A CNN-Based Fusion Method for Feature Extraction from Sentinel Data. Remote Sensing, 10(2), 236. https://doi.org/10.3390/rs10020236.

Srivastava S., Vargas-Muñoz J.E., Tuia D., 2019. Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. Remote Sensing of Environment, 228, 129–143. https://doi.org/10.1016/j.rse.2019.04.014.

Sumbul G., Charfuelan M., Demir B., Markl V., 2019. Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding. IGARSS 2019–2019 IEEE International Geoscience and Remote Sensing Symposium, 5901–5904.

Tan M., Le Q.V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv e-prints:arXiv:1905.11946. Doi: 10.48550/arXiv.1905.11946.

Tsagkatakis G., Aidini A., Fotiadou K., Giannopoulos M., Pentari A., Tsakalides P., 2019. Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement. Sensors, 19(18), 3929. https://doi.org/10.3390/s19183929.

Tuna C., Unal G., Sertel E., 2018. Single-frame super resolution of remote-sensing images by convolutional neural networks. International Journal of Remote Sensing, 39, 2463–2479. Doi: 10.1080/01431161.2018.1425561.

Dang Vu K., Ta Duc H., Ngo Thi Hai Y., Nguyen Khac A., Pham Thanh H., Weber C., 2025. Observation of coal mining activities in Ha Long and Cam Pha cities, Quang Ninh province, Vietnam using Sentinel-2 data. Vietnam Journal of Earth Sciences, 47(2), 272–289. https://doi.org/10.15625/2615-9783/23010

Wang H., Wang Y., Zhang Q., Xiang S., Pan C., 2017. Gated Convolutional Neural Network for Semantic Segmentation in High-Resolution Images. Remote Sensing, 9(5), 446. https://doi.org/10.3390/rs9050446.

Wang Y., Fang Z., Hong H., 2019. Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China. Science of the Total Environment, 666, 975–993. https://doi.org/10.1016/j.scitotenv.2019.02.263.

Weir N., Lindenbaum D., Bastidas A., Etten A., Kumar V., Mcpherson S., Shermeyer J., Tang H., 2019. SpaceNet MVOI: A Multi-View Overhead Imagery Dataset. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 992–1001.

Zhao S., Zhang T., Ma S., Wang M., 2023. Seahorse optimizer: a novel nature-inspired meta-heuristic for global optimization problems. Applied Intelligence, 53, 11833–11860.

Zhou W., Newsam S., Li C., Shao Z., 2017. Learning Low Dimensional Convolutional Neural Networks for High-Resolution Remote Sensing Image Retrieval. Remote Sensing, 9(5), 489. https://doi.org/10.3390/rs9050489.