DOI: https://doi.org/10.15276/hait.09.2026.38
UDC 681.3.07: 004.8

Improving accuracy of land cover classification through ensemble classification of lightweight deep learning models on EuroSAT dataset

Oleg M. Galchonkov[1)]
ORCID: http://orcid.org/0000-0001-5468-7299; o.n.galchenkov@gmail.com. Scopus Author Id 56081377900
Illia O. Baskov[1)]
ORCID: https://orcid.org/0000-0002-3517-6773; illyabaskov@gmail.com. Scopus Author ID 57788059000
Mykola I. Babych[1)]
ORCID: https://orcid.org/0000-0002-3946-9880; babich.tiger@gmail.com. Scopus Author ID 57253653800
Sergiy Y. Smyk[1)]
ORCID: https://orcid.org/0000-0001-7020-1826; smyk@op.edu.ua. Scopus Author ID 56996011100
[1)]Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

The object of this research is the classification of land surface types in multispectral images from the EuroSAT dataset. The subject of the study is high-precision algorithms for classifying images with thirteen channels, characterized by a small amount of computation. The use of such algorithms for processing images by onboard computers of the small satellites of the Sentinel-2 constellation, launched as part of the Copernicus program of the European Space Agency for Environment and Security, will improve classification accuracy and reduce the required power consumption, memory capacity, and the load on communication channels. Given the exponential growth in the required number of parameters of known neural networks to improve classification quality, an alternative approach is proposed in this paper. It involves the use of ensemble classifiers with stacking. In the first stage, it is proposed to use the most lightweight versions of neural networks with a relatively high level of classification quality. In the second stage, a multilayer perceptron is used. In this paper, lightweight versions of well-known neural networks are constructed: ResNet, EfficientNet, Compact SS-GhostNet, FastViT, and CCT. Using them within a stacked ensemble classifier yielded a classification quality of ninety-nine point fifty-nine percent on the EuroSAT dataset. This is zero point four percent higher than the best classification quality achieved by the CALCNet neural network. Moreover, the number of parameters in the proposed ensemble classifier is five times smaller than that of CALCNet. Versions of all programs are available on GitHub.

Keywords: Compact convolutional transformer; relative positional bias; relative attention; multilayer perceptron; remote sensing; ensemble classifier; tokenizer; hybrid architecture; depthwise convolution; inverted residual blocks

***For citation*: Galchonkov O. M., Baskov I. O., Babych M. I., Smyk S. Y. "Improving accuracy of land cover classification through ensemble classification of lightweight deep learning models on EuroSAT Dataset". *Herald of Advanced Information Technology*. 2026; Vol.9 No.4: 620–635. DOI: https://doi.org/10.15276/hait.09.2026.38**

1. INTRODUCTION

Earth remote sensing (ERS) plays a vital role in natural resource management, providing valuable insights into the dynamics of the Earth's environment and emerging environmental risks. The Sentinel-2 satellites, launched as part of the European Space Agency's Copernicus program, provide access to multispectral imagery with high spatial and temporal resolution. However, high-precision processing of 13-channel Sentinel-2 data requires a large amount of computation, which complicates the deployment of such systems on onboard computers of small satellites (CubeSats) and makes it difficult to process large volumes of data on regional servers. In addition to the fact that the onboard computers of small satellites are subject to strict requirements for energy consumption and the amount of memory used, an additional requirement is to reduce the load on communication channels.

Onboard image processing allows for the transmission of more compressed results (land use maps, emergency zones, segmented areas) rather than raw multispectral data. It's worth noting that older satellites transmit terabytes of raw multispectral data to regional servers. Therefore, there's also a need to reduce computational effort and improve the quality of the algorithms used. This, in turn, should lead to increased energy efficiency and environmental friendliness for regional computing centers.

Traditionally, convolutional neural networks (CNNs) have been used in remote sensing image classification. Recently, a large number of transformer-based neural networks have been developed for EuroSAT, demonstrating higher accuracy rates of approximately 99 % [1]. However, the number of parameters in such neural networks ranges from several million to several hundred

© Galchonkov O., Baskov I., Babych M., Smyk S., 2026

This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/deed.uk)

million. This significantly complicates their use not only on small satellites but also on regional servers. Therefore, the development of new architectures that provide higher classification quality with significantly lower computational requirements is urgently needed.

2. ANALYSIS OF LITERATURE DATA AND PROBLEM STATEMENT

Traditionally, in the field of land remote sensing, improvements to convolutional neural networks (CNNs) optimized for land cover classification have been proposed. In [2], a self-attention block was introduced to select the best parameters. However, the proposed neural network has 134 layers and 18.6 million parameters, while providing a classification accuracy of only 90.3% on the EuroSAT dataset [3]. Alternatively, the Res-Net-50 neural network [4] without additional modifications has 25.6 million parameters and provides an accuracy of 98.57 % on this dataset [3]. In [5], a hybrid spectral convolutional neural network was proposed that uses sequentially connected three-dimensional and two-dimensional convolutional networks. Despite the reduced spatial window size (25×25), the number of parameters in such a neural network is more than 5 million. In [6], compression-excitation and global averaging pooling blocks were used in the convolutional network. This made it possible to increase the classification accuracy to 98.44% and reduce the number of parameters to 2.2 million. In [7] and [8], it was shown that modification of the convolutional network architecture can significantly improve the ratio of classification quality to the volume of computations. The ConvNeXt neural network proposed there initially extracts feature maps from satellite and drone images, which are then fed to an aggregation module for multi-level fusion and subsequent multi-branch classification. In [9], it was found that carefully balancing the depth, width, and resolution of the network can lead to improved performance. Based on this, a new method for scaling neural networks was proposed and its effectiveness was demonstrated for MobileNets and ResNets. Furthermore, the EfficientNets family of neural networks was developed. The EfficientNet-B7 model from this family, with a high accuracy of 84.7 % on the ImageNet dataset, has 8.4 times fewer parameters than the best existing ConvNet model. A new Ghost module for generating more feature maps using inexpensive operations is proposed in [10], which can be used to modernize existing convolutional neural networks. Additionally, Ghost bottlenecks are developed in [10] for combining Ghost modules, allowing for the easy creation of lightweight networks.

Additional opportunities for improving classification quality in remote sensing emerged with the development of transformers [11]. Transformer-based neural networks have outperformed convolutional networks in many image processing tasks. However, their implementation on small satellites faces the same challenges as convolutional neural networks—their computational overhead is excessive [1]. This necessitates the development of specialized architectures. A hybrid architecture (CMT) was proposed in [12] that combines the advantages of convolutional networks (modeling local connections, inductive bias) and transformers (capturing global context and long-term dependencies). To achieve this, a convolutional input block with overlapping filters, a local perception block using depthwise convolution in the residual connection, a lightweight self-attention module to reduce computational complexity, an expansion-compression block to enhance the exchange between adjacent tokens, and global feature averaging over spatial coordinates were introduced into the architecture. An alternative approach (Swin Transformer) to modifying the transformer architecture is proposed in [13]. This approach uses local shifted windows, and self-attention is calculated within these non-overlapping windows. In [14], the FastViT family of neural networks was proposed, which allows for flexible selection of the required neural network parameters for an onboard computer. To ensure high classification quality and reduce computational effort, FastViT combines transformer and convolutional architectures. Furthermore, a building block, RepMixer, was added to the network, which uses structural reparameterization to remove redundant connections in the network after applying excessive parameterization during training and convolution with a large kernel. On the ImageNet dataset, this family demonstrated significantly better performance compared to CMT, EfficientNet, ConvNeXt, and Swin Transformer.

It should be noted that these neural networks were primarily developed for large input images and large training datasets. EuroSAT images are 64x64 in size, and each pixel has 13 channels. For such small images, high results can be achieved using neural networks with a hybrid architecture. One such network is CALCNet (Cross-Module Attention Network for Land Cover Classification) [15]. Its input contains convolutional layers that extract local spatial and channel features. A transformer block

then extracts global connections between distant parts of the image. Cross-attention modules combine local feature maps with the global features extracted by the transformer layers. CALCNet achieved the best result of 99.19% with 3.86 million parameters. Another family of neural networks with hybrid architecture, the compact convolutional transformer (CCT) family [16], also demonstrates high performance. Its architecture includes blocks such as a convolutional tokenizer, a transformer, and an additional attention mechanism (sequence pooling). Furthermore, the architecture can be easily supplemented with blocks proposed in other neural networks. For example, the relative attention blocks (Relative Attention) from CoAtNet, inverted residual blocks (Mobile Inverted Bottleneck Convolution - MBConv) from CoAtNet, EfficientNet, and MobileNetV2 [17].

Although the number of parameters in all of these neural networks ranges from several million to hundreds of millions, the proposed architectural solutions are promising for constructing neural networks with fewer than one million parameters. High classification quality in all of these neural networks is achieved through a disproportionately large increase in the number of adjustable parameters. An alternative option for improving classification quality is combining heterogeneous neural networks with a small amount of computation into stacked ensembles [18].

3. THE PURPOSE AND OBJECTIVES OF THE STUDY

The aim of the work is to develop an approach that provides algorithms for classifying the type of earth's surface for the Sentinel-2 satellites with increased classification accuracy and a reduction in the required number of computations.

To achieve this goal, the following objectives were set:

– develop a data preparation pipeline and calculate spectral indices for the EuroSAT dataset;

– develop modifications to the land surface classification algorithms for the first stage of the ensemble classifier with stacking;

– investigate the effectiveness of using ensemble classification with stacking for the EuroSAT dataset.

4. MATERIALS AND METHODS OF RESEARCH

The object of research in this paper is the classification of land surface types in multispectral images of high spatial and temporal resolution processed by onboard computers of the Sentinel-2 small satellites. The subject of the research is an ensemble classifier with stacking. The dataset for the study is EuroSAT [3]. This dataset contains images of 64 × 64 pixels, each of which contains 13 channels (Table 1). The number of images is 27,000. The images belong to 10 classes (Table 2).

Fig. 1 shows example images from all classes (only RGB colors were used) [3]. In this work, the data was divided as follows: 80% for training, 10 % for validation, and 10% for testing. Training was performed over 50 epochs. The best weighting coefficients from the validation data were saved.

Table 1. **Spectral channels of EuroSAT MS (Multispectral)**

Channel number	Name of the channel in the dataset	Purpose/Description	Central wavelength
B01	Coastal Aerosol	Coastal aerosol	~443 nm
B02	Blue	Blue (RGB)	~490 nm
B03	Green	Green (RGB)	~560 nm
B04	Red	Red (RGB)	~665 nm
B05	Vegetation Red Edge 1	Red edge of vegetation (VRE 1)	~705 nm
B06	Vegetation Red Edge 2	Red edge of vegetation (VRE 2)	~740 nm
B07	Vegetation Red Edge 3	Red edge of vegetation (VRE 3)	~783 nm
B08	NIR	Near infrared	~842 nm
B08A	Narrow NIR	Narrow near infrared	~865 nm
B09	Water Vapour	Water vapor	~945 nm
B10	SWIR – Cirrus	Shortwave IR (cirrus clouds)	~1375 nm
B11	SWIR 1	Shortwave IR 1	~1610 nm
B12	SWIR 2	Shortwave IR 2	~2190 nm

***Source*: adapted by the author based on [3]**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Table 2. **Data classes in the EurSAT MS (Multispectral) set**

No.	Name	Number of images
1	AnnualCrop	3000
2	Forest	3000
3	Herbaceous Vegetation	3000
4	Highway	2500
5	Industrial	2500
6	Pasture	2000
7	PermanentCrop	2500
8	Residential	3000
9	River	2500
10	SeaLake	3000

***Source*: adapted by the author based on Helber et al [3]**

Classification accuracy was estimated with these weighting coefficients on the test data as the ratio of the number of images with correctly recognized surface types to the total number of test images (2700). The programs for conducting the study were written in Python using the Keras library and are available on GitHub [19].

To achieve high classification quality, the following methods were used:

– AdamW optimizer with a damping coefficient of 10^{-4};

– Label Smoothing (0.005 is used instead of 0, 0.955 is used instead of 1);

– cosine-law damping of the training step with an initial warm-up;

– combinations of random rotations, brightness and contrast variations, and simulating seasonal and atmospheric noise are used for training augmentation.

5. RESULTS OF A STUDY ON THE EFFECTIVENESS OF USING ENSEMBLE CLASSIFICATION TO IMPROVE THE QUALITY OF SURFASE TYPE CLASSIFICATION

5.1. Development of a data preparation pipeline

5.1.1. Primary normalization. The EurSAT MS source file contains data in 13×64×64 format

$$x_{raw} \in R^{13\times64\times64} .$$

For further use in programs, the data is transposed

$$x_0 = (x_{raw})^T \in R^{64\times64\times13} .$$

The channel values of each pixel in the images are presented as indexes of reflected radiation energy in units of surface spectral reflectance. For normalization, the channel values are divided by 10,000 and, if necessary, limited to fall within the range [0.0, 1.0]

$$x_{norm} = clip(\frac{x_0}{10000}, 0.0, 1.0) \in R^{64\times64\times13} .$$

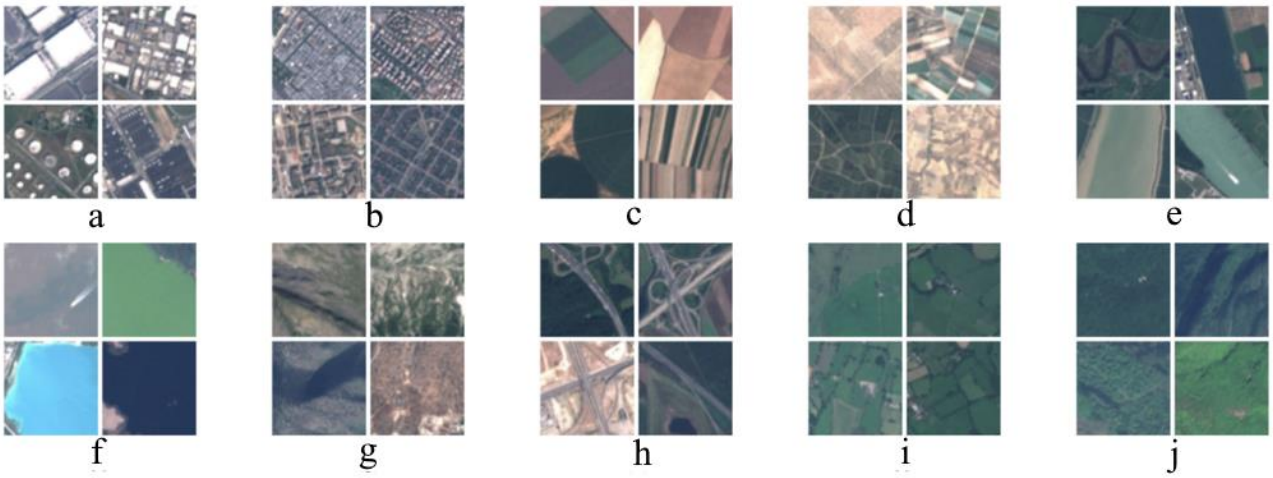


Fig. 1. **Examples of images from the EurSAT dataset (only RGB colors used): a – Industrial buildings; b – Residential buildings; c – Annual crop; d – Permanent crop; e – River; f – Sea and lake; g – Herbaceous vegetation; h – Highway, I – Pasture; j – Forest**

***Source*: adapted by the author based on Helber et al [3]**

5.1.2. Calculation of physical-geographical spectral indices. In accordance with common practice [20], [21], three more spectral indices are added to the existing 13 channels, explicitly reflecting fundamental physical properties: light absorption by chlorophyll, reflectance from leaf structure, and water absorption in the shortwave infrared range. This facilitates neural network training, increases the robustness of features to variations in solar illumination angle and atmospheric haze, improves the separability of spectrally similar classes (e.g., River and Highway, Pasture and Permanent Crop, or Forest), and increases the spectral information content of each pixel.

Normalized Vegetation Index

$$NDVI = \frac{B_{08} - B_{04}}{B_{08} + B_{04} + \varepsilon} = \frac{x_{norm}[:,:,7] - x_{norm}[:,:,3]}{x_{norm}[:,:,7] + x_{norm}[:,:,3] + \varepsilon} .$$

Normalized water index

$$NDWI = \frac{B_{03} - B_{08}}{B_{03} + B_{08} + \varepsilon} = \frac{x_{norm}[:,:,2] - x_{norm}[:,:,7]}{x_{norm}[:,:,2] + x_{norm}[:,:,7] + \varepsilon} .$$

Normalized Development Index

$$NDBI = \frac{B_{11} - B_{08}}{B_{11} + B_{08} + \varepsilon} = \frac{x_{norm}[:,:,11] - x_{norm}[:,:,7]}{x_{norm}[:,:,11] + x_{norm}[:,:,7] + \varepsilon} .$$

In all expressions for indices $\varepsilon=10^{-8}$ to protect against division by zero.

5.1.3. Formation of the final input tensor. The final input tensor for further processing is formed as a concatenation of the 13 original channels and the calculated indices

$$X = Concat([x_{norm}, NDVI, NDWI, NDBI], axis = -1),$$
$$X \in R^{64\times64\times16} .$$

5.2. Development of modifications of classification algorithms for the first stage of an ensemble classifier with stacking

5.2.1. ResNet UltraLight. The base ResNet-18 architecture was modified to reduce the number of parameters to 196,000 with a classification performance of 98.41 % by using the scaling method from [9] (width factor of 0.707) and replacing conventional convolutional blocks with spatially separable convolution blocks (DepthwiseSeparableBlocks) [17].

The learning curves for ResNet UltraLight are shown in Fig. 2, and the confusion matrix is shown in Fig. 3.

ResNet UltraLite Architecture Description:

1. Architectural type: A modification of ResNet-18.
2. Instead of the classic 3x3 Conv2D, it uses `DepthwiseSeparableBlock`.
3. The channel pruning factor is 0.707, which reduces the number of filters in all layers and cuts the number of parameters exactly in half.
4. The number of filters has been reduced from the traditional ones (64, 128, 256, 512) to 23 (Stem/Stage 1), 45 (Stage 2), 91 (Stage 3), and 181 (Stage 4).
5. The model consists of an initial Stem layer and 4 stages, each containing 2 lightweight residual blocks.
6. Spatial resolution downsampling (strides=2) is performed on the first blocks of the 2nd, 3rd, and 4th stages using 1x1 projections.
7. Feature aggregation is completed by a GlobalAveragePooling2D layer, which collapses the maps into a 181-dimensional vector.
8. Output head: a fully connected Dense layer with a softmax function corresponding to the number of classes in the EuroSAT dataset.

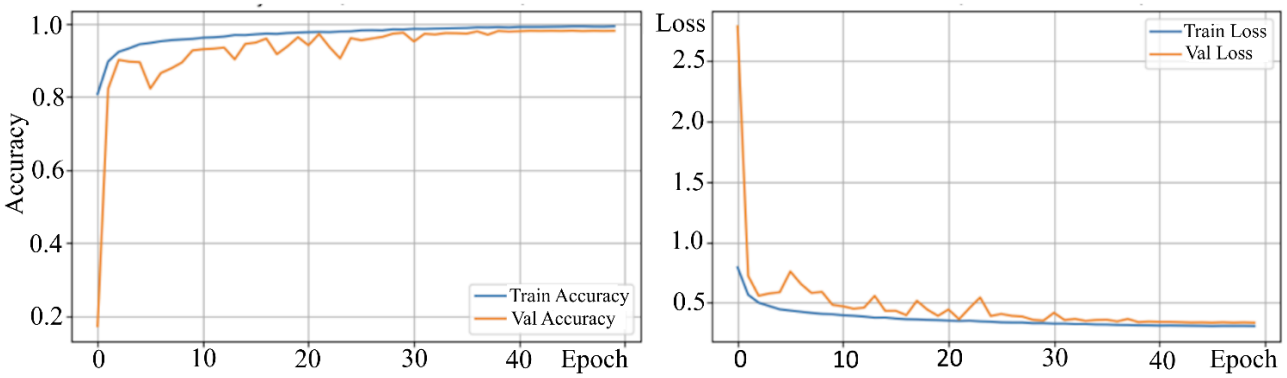


Fig. 2. **Learning curves for ResNet UltraLight**
***Source*: compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

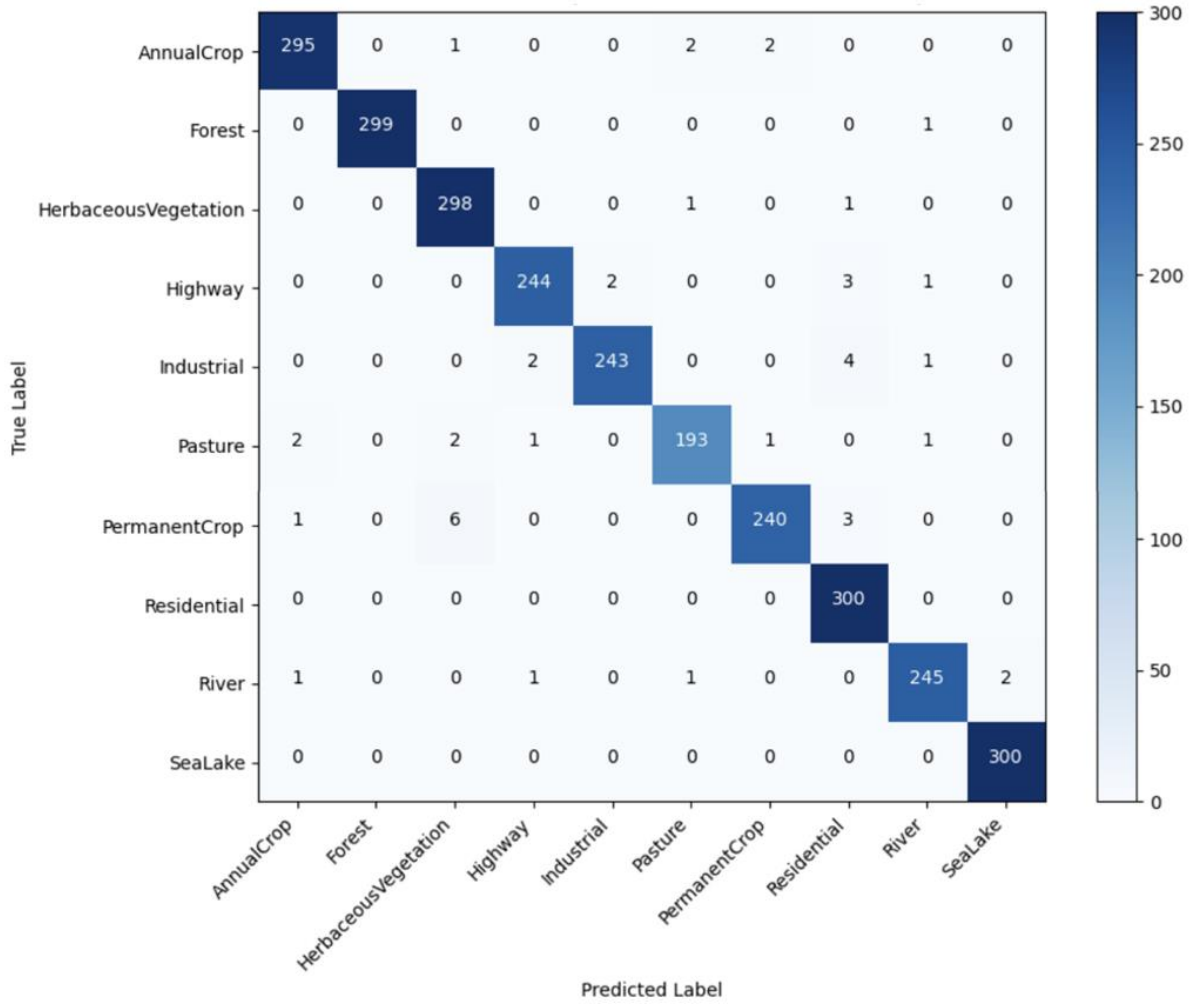


Fig. 3. **Confusion matrix for ResNet UltraLight**

***Source*: compiled by the authors**

5.2.2. EfficientNet UltraLight. Reducing the number of parameters to 148,000 while maintaining classification quality at 98.81 % in this neural network was achieved through a combination of the following four architectural and mathematical principles.

1. Scaling Down Width & Depth

Depth is reduced by reducing the number of MBConv blocks used. Width is compressed by reducing the number of channels in the tensors: the input contains only 16 channels, while the maximum number of channels in the tensors is 160.

2. Modification of MBConv blocks (Inverted Residuals)

The standard kernel convolution is replaced with Depthwise Separable Convolution, which splits the process into two steps: Depthwise Convolution (spatial filtering of each channel separately) and a subsequent Pointwise Convolution (1x1 convolution for mixing channels).

Additionally, the Expansion Ratio in MBConv has been reduced from the standard 6x to 2x and 3x.

3. Efficient Attention Blocks (SEBlock2D)

Fully connected layers within the Squeeze-and-Excitation (SE) block compress the channel vector by a factor of 4 before inverse reconstruction.

4. Replacing the heavyweight classifier with Global Average Pooling

Instead of fully connected layers, the output feature map in EfficientNet UltraLite is compressed to a dimension of (8, 8, 160). Global Average Pooling2D averages each 8x8 spatial window into a single value, and a single fully connected layer immediately maps 160 features into 10 EuroSAT classes.

The learning curves for EfficientNet UltraLight are shown in Figure 4, and the confusion matrix in Fig. 5.

5.2.3. Compact SS-GhostNet UltraLight. The learning curves for Compact SS-GhostNet UltraLite are shown in Fig. 6, and the Cofusion matrix in Fig. 7.

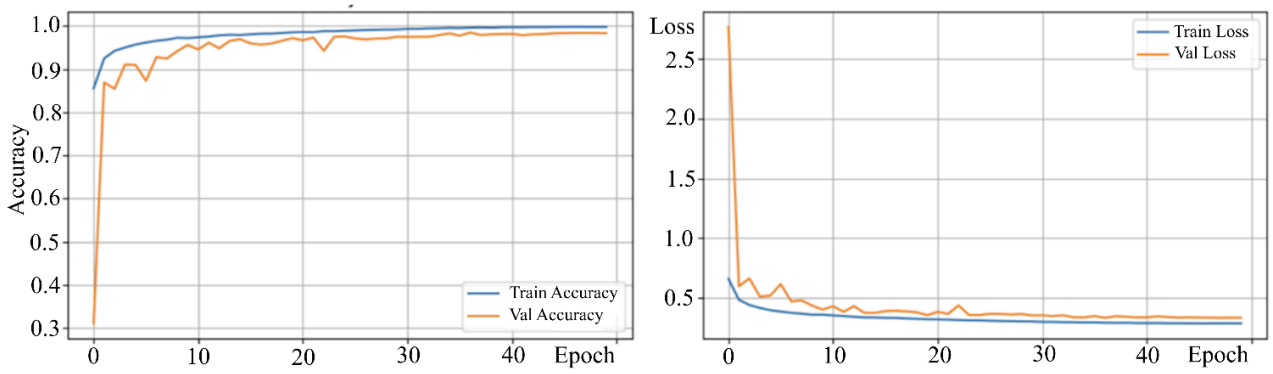


Fig. 4. **EfficientNet UltraLight learning curves**
***Source*: compiled by the authors**

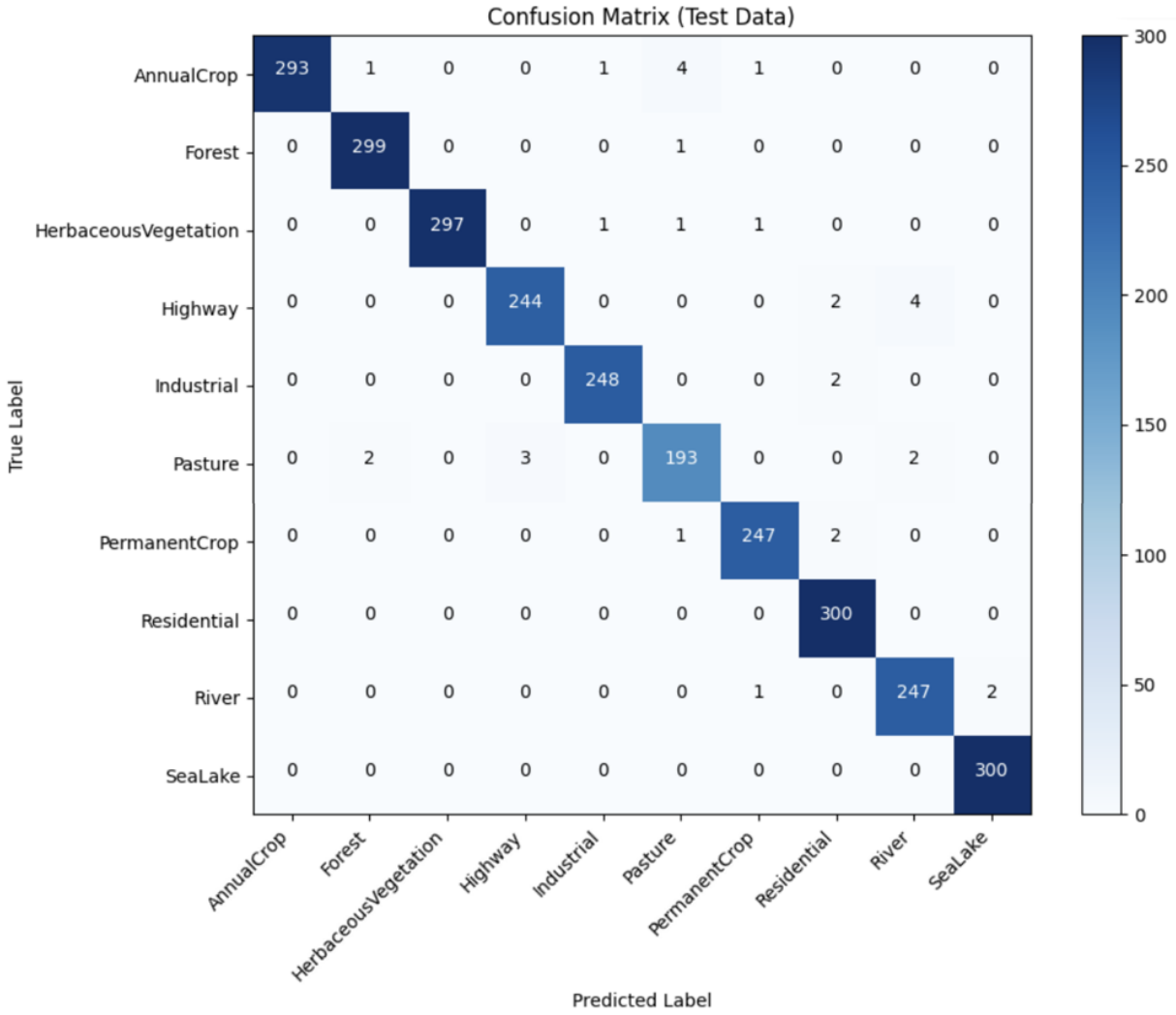


Fig. 5. **Confusion matrix for EfficientNet UltraLight**
***Source*: compiled by the authors**

Reducing the number of parameters to 101,000 while maintaining classification quality at 97.89% in this neural network was achieved through a combination of the following coordinated solutions.

1. Linear decomposition of convolutions: the "ghost" map principle.

To reduce the number of parameters, only a subset of the main feature maps is created, and the remaining feature maps are generated from these maps using element-by-element spatial processing.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

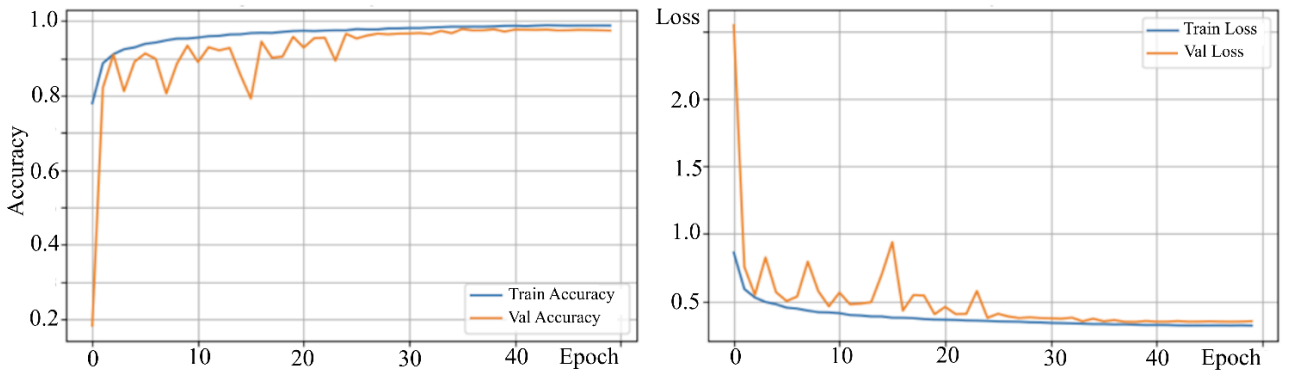


Fig. 6. **Compact SS-GhostNet UltraLight learning curves**

***Source*: compiled by the authors**

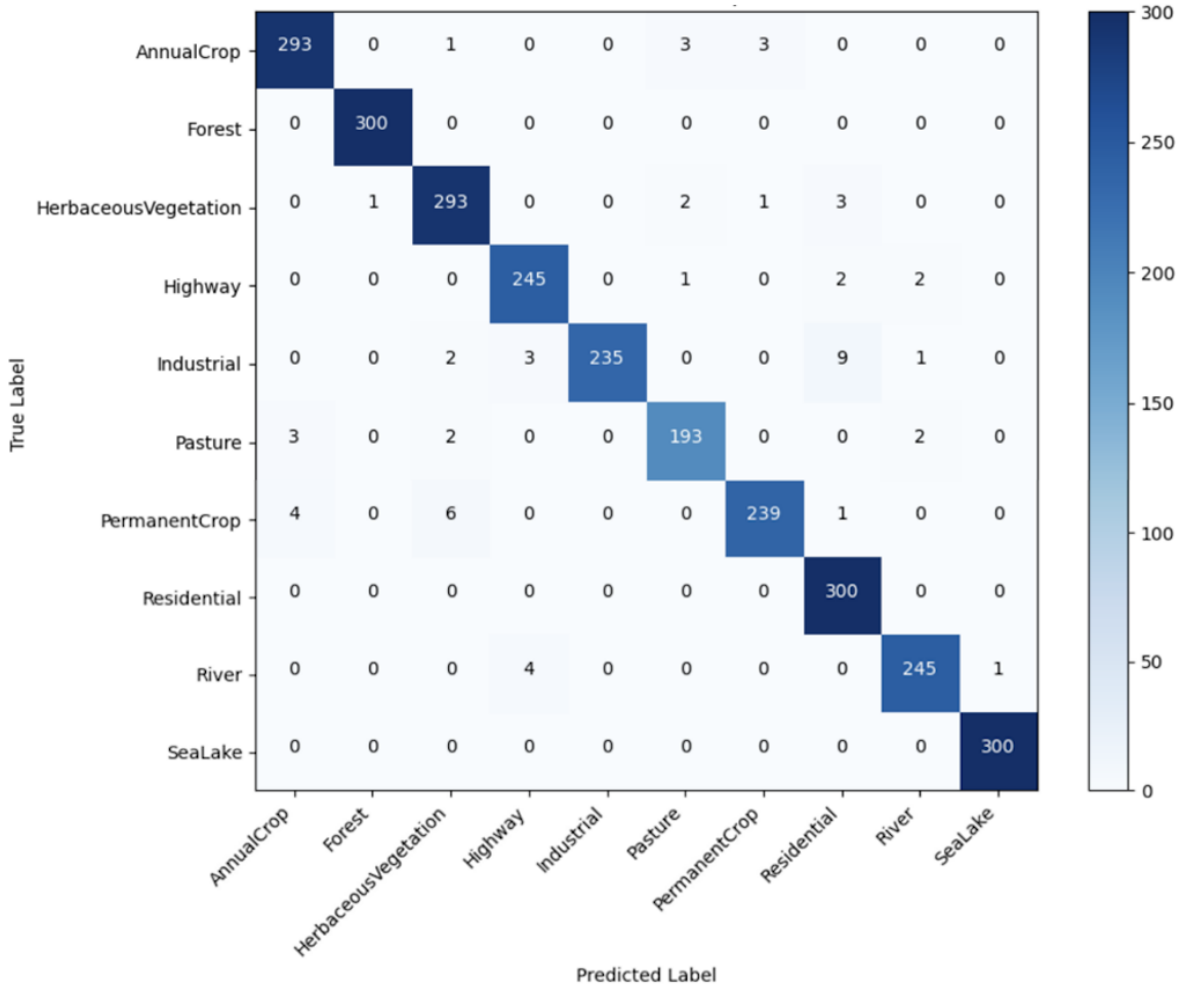


Fig. 7. **Compact SS-GhostNet UltraLight Confusion matrix**

***Source*: compiled by the authors**

2. Separation of spatial and spectral processing (Spectral-Spatial Bottleneck).

Instead of using 3D kernels for convolution, 2D convolutions (DepthwiseConv2D) are initially used for each channel without interlayer summation, and then the coupling coefficients between spectral channels are calculated using an attention mechanism (Spectral Processing - SEBlock2D).

3. Using Spectral Attention Blocks via Squeeze-and-Excitation (SEBlock2D)

Spectral attention blocks evaluate the importance of each channel for each class, and the most important channels are further processed with larger weights.

4. Global Channel Compression

Instead of the standard increase in the number of channels from 16 to 960, a truncated range of 8 to 240 is used in intermediate stages.

5.2.4. FastViT UltraLight. Reducing the number of parameters to 137,000 while maintaining

classification quality at 98.19% in this neural network was achieved through a combination of the following architectural solutions.

1. Drastic reduction in channel width and depth

In typical FastViT architectures, the number of channels sequentially increases to 512 or more at different stages. In this version, the maximum number of channels is 96. Furthermore, the number of blocks of each type is minimal: 1, 1, 2, 1.

2. Group and Depthwise Convolutions (Group & Depthwise Convolutions)

As in other lightweight neural network versions, instead of 3-D convolutions, depthwise convolutions are used for each channel individually in tokens. For mixing between channels, pointwise convolutions (1×1) are used, not across the full channel depth, but separately for each of the two channel groups.

3. A hybrid attention block is present only at the final stage.

In the first three stages, a simple convolutional shuffle, requiring a minimal number of weights, is used instead of an attention block. The fourth stage includes an attention block, but only with one head.

4. Using ultra-light attention blocks per channel (UltraLightSEBlock2D)

In the channel attention blocks, the compression ratio has been increased from the standard 2-4 to 8.

5. Downsampling structure (PatchEmbedDownsample)

A smooth and efficient transition from the original frames, with dimensions (64×64), to (4×4) in the following sequence: 64→32→16→8→4.

6. Minimalistic block for the resulting classification.

While conventional networks use several fully connected layers at the output after Global Average Pooling, this FastViT GlobalAveragePooling2D implementation directly convolves the 4×4×96 feature maps into a vector of 96 values, followed by only one fully connected layer.

The learning curves for FastViT UltraLight are shown in Fig. 8, and the confusion matrix in Fig. 9.

5.2.5. CCT UltraLight. The learning curves for the CCT UltraLight are shown in Fig. 10, and the confusion matrix in Fig. 11.

Reducing the number of parameters to 183,000 while maintaining classification performance at 98.81 % in this neural network was achieved through the following six key principles.

1. Replacing standard convolutions in the tokenizer with GhostModule

The GhostModule generates only half of the feature maps using standard pointwise convolutions (1x1). The other half of the feature maps ("ghost features") are generated using Depthwise Conv2D (a 3x3 channel-wise convolution).

2. Using Mobile Inverted Bottleneck (GhostMBConvBlock)

The GhostConvTokenizer is built on the MBConv architecture, but with integrated Ghost modules. Spatial convolutions (3x3) are performed not on the full number of channels, but in a compressed intermediate space. Filtering separation: spatial processing (3x3) is completely separated from inter-channel mixing (1x1).

3. Abandoning the classic [CLS] token and using Sequence Pooling

This is a standard feature of all CCTs. Instead of extracting a separate trainable vector, the layer compresses the token sequence using a compact attention mechanism (just one fully connected layer). This minimizes the number of parameters in the output classification block.

4. Relative Position Bias instead of Absolute Position Embeddings.

With Absolute Position Embeddings, each token is assigned its own unique vector (sinusoidal or trainable). The model sees exactly where the token is located in the image, but it is forced to infer information about the distance between tokens i and j indirectly through the dot product of their embeddings.

With Relative Position Embeddings, the model is only told the relative distance between tokens, and the relative bias information vector is embedded within the attention mechanism. As a result, instead of operating with the full absolute positional offset coordinate matrix for relative positional encoding, only the Toeplitz offset matrix is used to calculate the attention matrix, significantly reducing the computational effort.

5. Compact latent space dimensionality

The neural network uses a reduced token size of 64, a reduced number of heads of 4, and a reduced number of transformer layers of 4.

6. Lightweight channel and spatial attention blocks (SEBlock)

Using a channel compression factor of 4 allows for the weighting of the importance of channels and spatial features with a small amount of computation.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

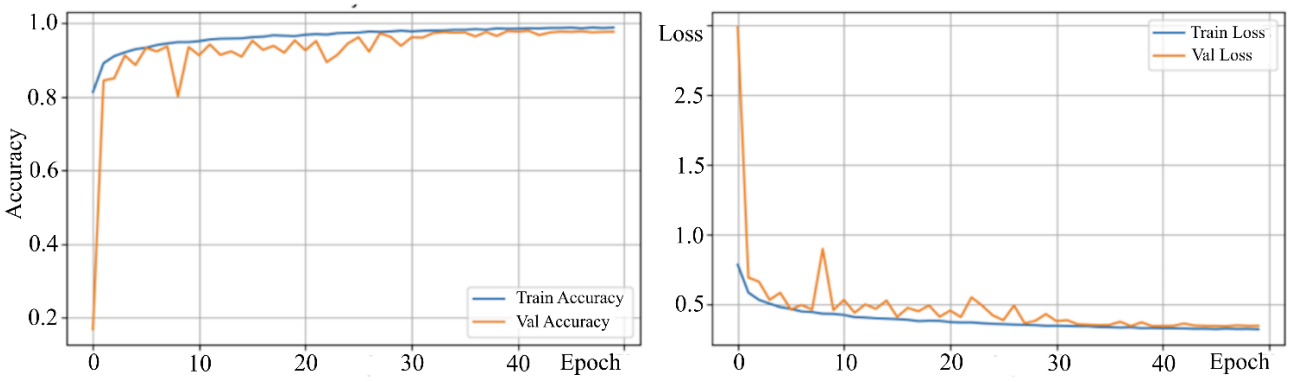


Fig. 8. **FastViT UltraLight learning curves**
***Source*: compiled by the authors**

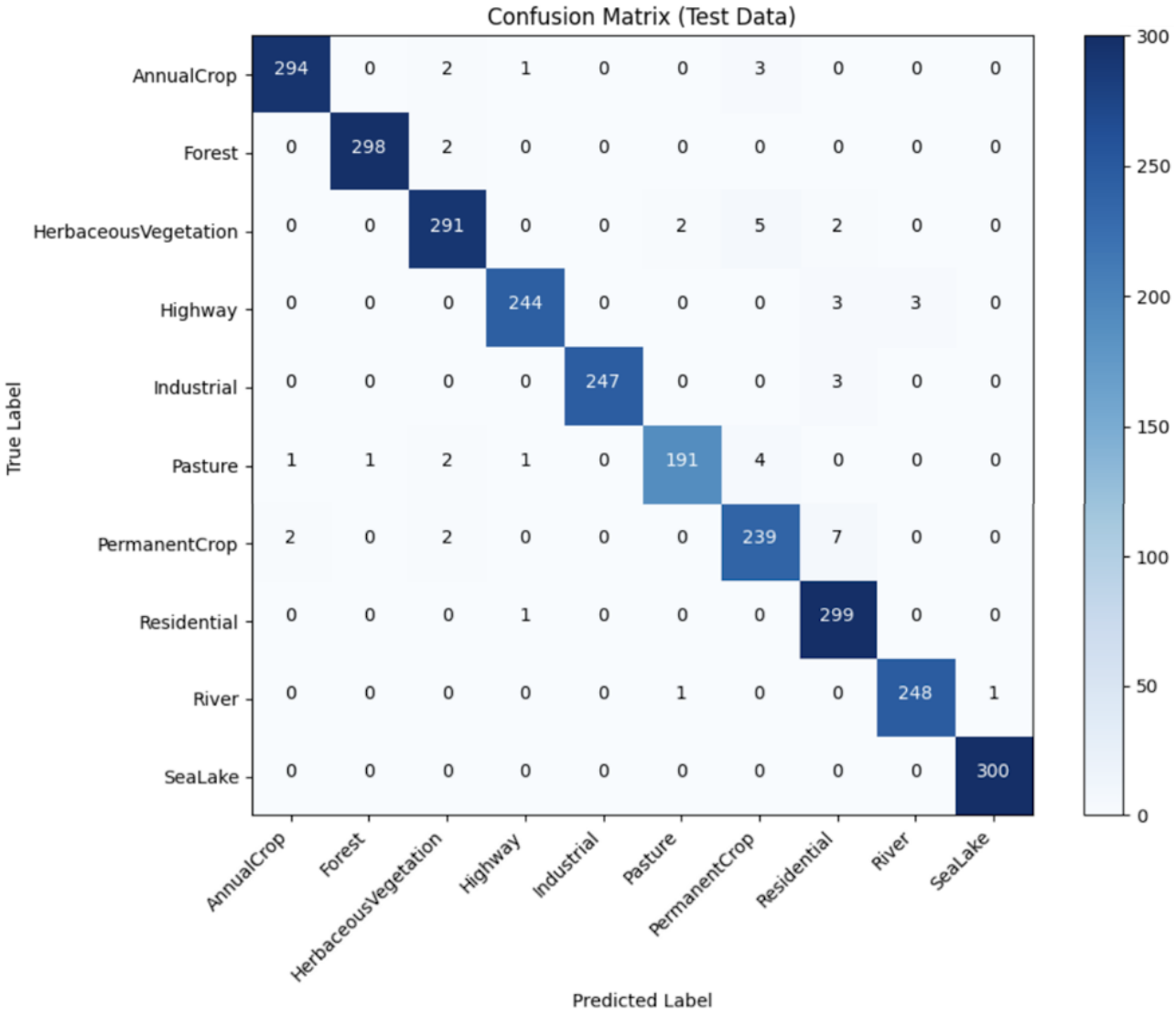


Fig. 9. **FastViT UltraLight Confusion Matrix**
***Source*: compiled by the authors**

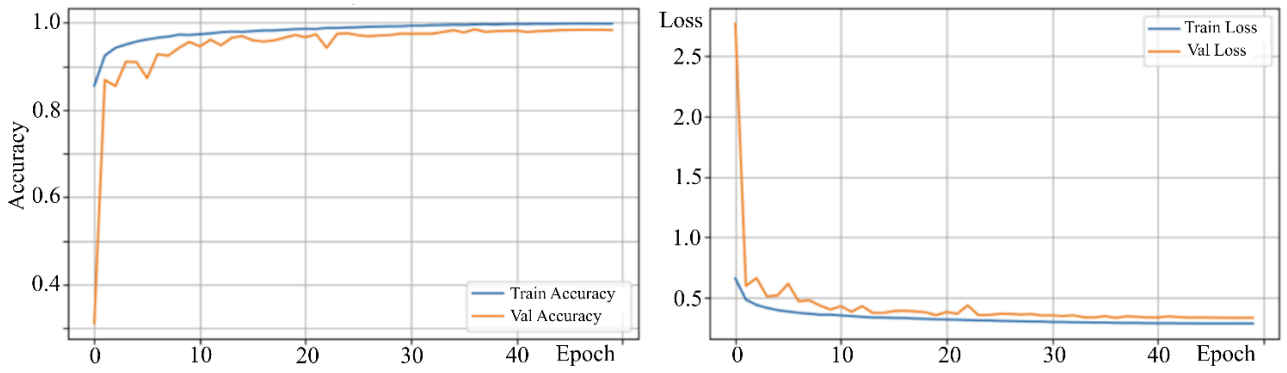


Fig. 10. **Learning curves for CCT UltraLight**
***Source*: compiled by the authors**

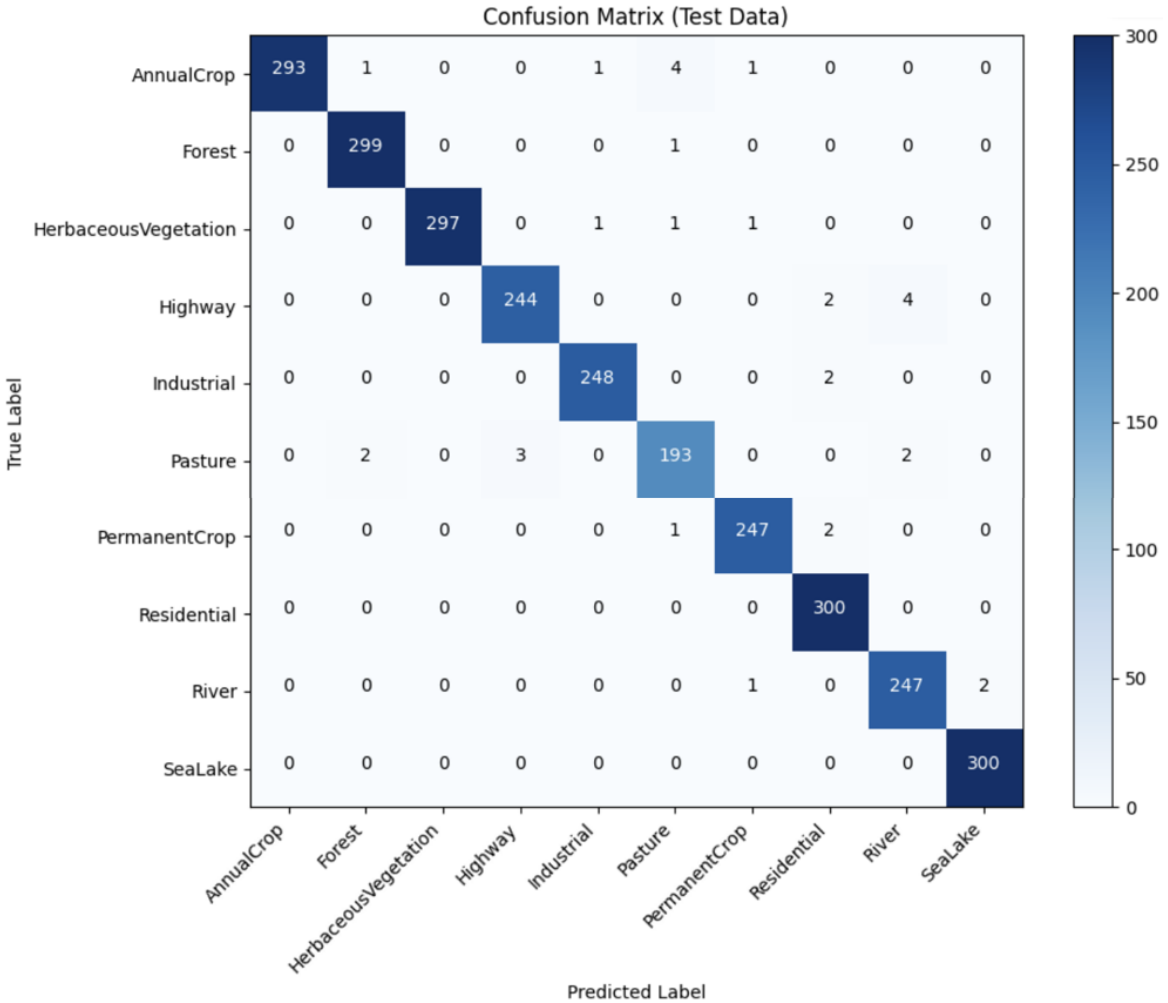


Fig. 11. **Confusion matrix for CCT UltraLight**
***Source*: compiled by the authors**

5.3. Study of the effectiveness of using ensemble classification with stacking for the EuroSAT dataset

A comparison of the error matrices of the developed algorithm modifications (Fig. 3, Fig. 5, Fig. 7, Fig. 9 and Fig. 11) shows that they produce different amounts of errors when classifying images with the same surface type. This can be exploited by building an ensemble classifier with stacking based on them. All algorithm modifications process the same images in the first stage, and in the second stage, their results are processed and the final result is produced (Fig. 12).

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

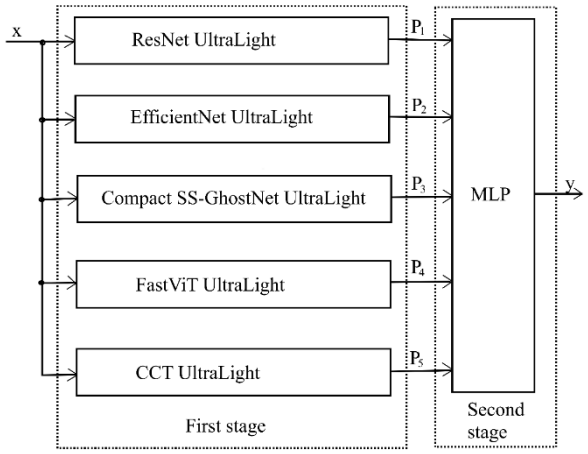


Fig. 12. **Ensemble classifier with stacking**
***Source*: compiled by the authors**

The results of all first-stage neural networks are combined into one vector by concatenation

$$P = concat(P_1, P_2, P_3, P_4, P_5) ,$$

where $P_1,\ldots,P_5$ are the output vectors of the first stage neural networks, each of dimension k=10, and represent estimates of the probability of the input image belonging to each of the classes of the earth's surface.

Vector P is fed to the input of the second stage, which contains a multilayer perceptron (MLP). The MLP structure is as follows:

– input layer, dimension - 50,

– first hidden layer, dimension - 40, swish activation function,

– second hidden layer, dimension - 20, swish activation function,

– output layer, dimension - 10, sigmoid activation function.

The class of an image x is determined as the index of the maximum element in vector y (the MLP output vector)

$$class(x) = \arg\left\{\max_j (y_j)\right\} ,$$

where y_j are the elements of the vector y, j=0,…,9.

Each of the first-stage neural networks is trained independently, as shown in Section 5.2. After this, the best values of their weights are fixed, and the second-stage MLP is trained. The MLP learning curves are shown in Fig. 13. The overall results for the entire ensemble classifier are presented in Table 3.

For comparison, a 3-out-of-5 majority vote was implemented in the second stage of the ensemble classifier. This ensured a classification accuracy of 99.26 %.

According to [1], the best result of 99.09% classification accuracy on the EuroSAT dataset was achieved by the pretrained MobileViT V2 neural network (Apple), which has 4,393,971 parameters and requires converting the image from 64x64 to 256x256. Compared to this network, the ensemble classifier under consideration provides 0.5 % higher classification quality and has 5.7 times fewer parameters.

The CALCNet neural network proposed in [15] provides 99.19 % accuracy with 3.86 million parameters. Compared to this network, the ensemble classifier considered provides 0.4 % higher classification quality and has 5 times fewer.

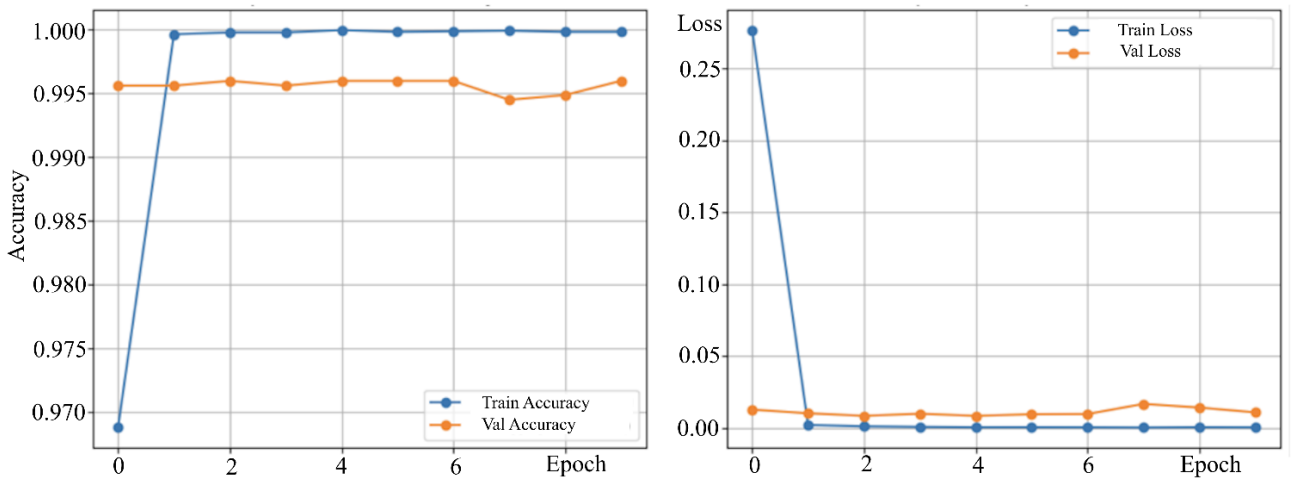


Fig. 13. **MLP learning curves**
***Source*: compiled by the authors**

Table 3. **Overall results for the entire ensemble classifier**

Neural network	Number of parameters (thousands)	Classification accuracy
ResNet UltraLight	196	98,41 %
EfficientNet UltraLite	148	98,81 %
Compact SS-GhostNet UltraLite	101	97.89 %
FastViT UltraLite	137	98.19 %
CCT UltraLite	183	98.81%
MLP	3	–
Ensemble classifier	768	99.59 %

***Source*: compiled by the authors**

6. DISCUSSION OF THE PROPOSED APPROACH TO INCREASING CLASSIFICATION ACCURACY AND REDUCING THE REQUIRED AMOUNT OF COMPUTATION

As shown in the study conducted in [1], increasing the depth of the neural network does not significantly improve the classification accuracy on the EuroSAT dataset, but significantly increases the number of computations. In this paper, we propose building processing based on an ensemble classifier with stacking. In this case, neural networks with a high, but not maximum, classification coefficient can be used at the first stage. The main requirements for them are: a minimum number of parameters and differences in processing algorithms, which leads to the fact that different neural networks make errors on different images. In this paper, 5 neural networks that meet these requirements were used. Comparison of their confusion matrices (Fig. 3, Fig. 5, Fig. 7, Fig. 9 and Fig. 11) shows that they make errors on different images. Even simple majority voting at the second stage of the ensemble classifier allows for a significant increase in the classification quality to 99.26%, compared to known results. However, using MLP at the second stage allows for achieving 99.59%. This is explained by the fact that the input of MLP is full vectors from the first-stage neural networks, and not just the indices of the maximum values of the components of these vectors.

An additional advantage of the proposed approach is that simple first-stage neural networks with a relatively small number of parameters can be trained individually, and their training is signifycantly simpler than training neural networks with a larger number of parameters. The learning curves shown in Fig. 2, Fig. 4, Fig. 6, Fig. 8 and Fig. 10 show that, within 50 epochs, all first-stage neural networks were successfully trained to sufficiently high classification quality values. It should also be noted that parallel first-stage neural networks can be implemented on separate processors onboard the aircraft, significantly increasing the speed of image processing.

The limitations of the proposed approach include the need for additional research and selection of those neural networks whose errors differ as much as possible from each other.

Directions for further development of the proposed approach include the development of modifications of neural networks for the first stage with a higher ratio of classification quality and the number of parameters, compared to those used in this work.

7. CONCLUSIONS

1. A data processing pipeline was developed that supplements the number of spectral channels in the EuroSAT dataset with spectral index values.

2. Modified land surface classification algorithms for the first stage of the stacked ensemble classifier (ResNet, EfficientNet, Compact SS-GhostNet, FastViT, and CCT) were developed and trained. These algorithms are characterized by a small number of parameters and relatively high classification quality.

3. The proposed approach enables the construction of classifiers for the EuroSAT dataset that are characterized by high classification quality and a small number of parameters. The stacked ensemble classifier built based on the proposed approach demonstrated 0.4% higher classification quality compared to the best known result, with a fivefold reduction in the number of parameters and higher parallelizability.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

Conflict of interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other nature, which could influence the study and its results presented in this article.

Financing: The study was conducted without financial support

Accessibility data: The manuscript has linked data in the data repository. Links are provided in the text of the article

Use of artificial intelligence tools: The authors used artificial intelligence technologies within the permissible framework to provide their own verified data, which is described in the research methodology section

REFERENCES

1. Le, T.-D., Ha, V., Nguyen, T., et al. “Onboard satellite image classification for Earth observation: A comparative study of ViT models”. *arXiv*. 2025. DOI: https://arxiv.org/pdf/2409.03901.

2. Albarakati, H. M., Khan, M. A., Hamza, A., et al. “A novel deep learning architecture for agriculture land cover and land use classification from remote sensing images based on network-level fusion of self-attention architecture”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024; 17: 6338–6353, https://www.scopus.com/pages/publications/85186968522. DOI: https://doi.org/10.1109/JSTARS.2024.3369950.

3. Helber, P., Bischke, B., Dengel, A. & Borth, D. “EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification”. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*”. July 2019; 12 (7): 2217–2226, https://www.scopus.com/pages/publications/85070509108. DOI: https://doi.org/10.1109/JSTARS.2019.2918242.

4. He, K., Zhang, X., Ren, S. & Sun, J. “Deep residual learning for image recognition”. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. 2016. p. 770–778, https://www.scopus.com/pages/publications/84986274465.
DOI: https://doi.org/10.1109/CVPR.2016.90.

5. Roy, S. K., Krishna G., Dubey, S. R. & Chaudhuri, B. “HybridSN: Exploring 3D-2D CNN feature hierarchy for hyperspectral image classification”. *arXiv*. 2019. DOI: https://arxiv.org/pdf/1902.06701.

6. Alikhani, M., Ebadati, B. & Attarzadeh, R. “Lightweight and efficient deep learning models for remote sensing-based land use and land cover classification: A case study on EuroSAT Dataset, Earth observation and geomatics engineering”. 2025; 8 (2): 106–117. DOI: https//doi.org/10.22059/eoge.2025.396331.1179.

7. Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T. & Xie, S. “A ConvNet for the 2020s”. *arXiv*. 2022. DOI: https://arxiv.org/pdf/2201.03545.

8. Guan, F., Zhao, N., Fang, Z., Jiang, L., Zhang, J., Yu, Y. & Huang, H. “Multi-level representation learning via ConvNeXt-based network for unaligned cross-view matching”. *Geo-spatial Information Science*. 2025; 28 (5): 2344–2357, https://www.scopus.com/pages/publications/85215119583. DOI: https//doi.org/10.1080/10095020.2024.2439385.

9. Tan, M. & Le, Q. V. “EfficientNet: Rethinking model scaling for convolutional neural networks”. *arXiv*. 2020. DOI: https://arxiv.org/pdf/1905.11946.

10. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C. & Xu, C. “GhostNet: More features from cheap operations”. *arXiv*. 2020. DOI: https://arxiv.org/pdf/1911.11907.

11. Bazi, Y., Bashmal, L., Rahhal, M. M. A., Dayil, R. A. & Ajlan, N. A. ”Vision transformers for remote sensing image classification”. *Remote Sens.* 2021; 13: 516, https://www.scopus.com/pages/publications/85100460069. DOI: https://doi.org/10.3390/rs13030516.

12. Guo, J., Han, K., Wu, H., Tang, Y., Chen, X., Wang, Y. & Xu, C. “CMT: Convolutional neural networks meet vision transformers”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA. 2022. p. 12165–12175, https://www.scopus.com/pages/publications/85141758309.
DOI: https//doi.org/10.1109/CVPR52688.2022.01186.

13. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. “Swin transformer: Hierarchical vision transformer using shifted windows”. In: *Proceedings of the IEEE/CVF International*

Conference on Computer Vision (ICCV). Montreal, QC, Canada. 2021. p. 10012–10022, https://www.scopus.com/pages/publications/85127818917.
DOI: https//doi.org/10.1109/ICCV48922.2021.00986.

14. Vasu, P., Gabriel, J., Zhu, J., Tuzel, O., & Ranjan, A. "FastViT: A fast hybrid vision transformer using structural reparameterization". *arXiv*. 2023. DOI: https://arxiv.org/pdf/2303.14189.

15. Fayaz, M., Yar, H., Jiang, W., Ibrahim, A.H., Islam, M. & Dang, L. M. "CALCNet: A novel cross-module attention network for efficient land cover classification". *Special Issue. "Hyperspectral Remote Sensing Image Analysis via Advanced Deep Learning and Computer Vision". Edited by Peixian Zhuang, Xiangyong Cao, Xueyang Fu &. Remote Sens*. 2026; 18: 1218, https://www.scopus.com/pages/publications/105037063149. DOI: https://doi.org/10.3390/rs18081218.

16. Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J. & Shi, H. "Escaping the big data paradigm with compact transformers". *ArXiv*. 2021. DOI: https://arxiv.org/pdf/2104.05704.

17. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. C. "MobileNetV2: Inverted residuals and linear bottlenecks". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. p. 4510–4520, https://www.scopus.com/pages/publications/8506279951 DOI: https//doi.org/10.1109/CVPR.2018.00474.

18. Galchonkov, O., Babych, M., Zasidko, A. & Poberezhnyi, S. "Using a neural network in the second stage of the ensemble classifier to improve the quality of classification of objects in images". *Eastern-European Journal of Enterprise Technologies*. 2022; 3 (9(117)): 15–21, https://www.scopus.com/pages/publications/85133776625. DOI: https://doi.org/10.15587/1729-4061.2022.258187.

19. Galchonkov O. M. "EuroSAT-1-**". –** Available from: https://github.com/Galchonkov/EuroSAT-1-.

20. Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V., Murayama, Y. & Ranagalage, M. "Sentinel-2 data for land cover/use mapping: A review". *Remote Sensing*. 2020; 12 (14): 2291, https://www.scopus.com/pages/publications/85088629563. DOI: https//doi.org/ 10.3390/rs12142291.

21. Tzepkenlis, A., Marthoglou, K. & Grammalidis, N. "Efficient deep semantic segmentation for land cover classification using sentinel imagery". *Remote Sensing*, 2023; 15 (8): 2027, https://www.scopus.com/pages/publications/85156166842. DOI: https//doi.org/10.3390/rs15082027.

Received 02.09.2026
Received after revision 28.10.2026
Accepted 02.11.2026

DOI: https://doi.org/10.15276/hait.09.2026.38
УДК 681.3.07: 004.8

Підвищення точності класифікації типу земної поверхні за допомогою ансамблевої класифікації з використанням полегшених моделей глибокого навчання на наборі даних EuroSAT

Галчонков Олег Миколайович[1)]
ORCID: https://orcid.org/0000-0001-5468-7299; o.n.galchenkov@gmail.com. Scopus Author Id 56081377900
Баськов Илля Олександрович[1)]
ORCID: https://orcid.org/0000-0002-3517-6773; illyabaskov@gmail.com. Scopus Author ID 57788059000
Бабич Микола Іванович[1)]
ORCID: https://orcid.org/0000-0002-3946-9880; babich.tiger@gmail.com. Scopus Author ID 57253653800

Смик Сергій Юрійович[1)]
ORCID: https://orcid.org/0000-0001-7020-1826; smyk@op.edu.ua. Scopus Author ID 56996011100
[1)] Національний університет «Одеська політехніка», пр.Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

Об'єкт дослідження – класифікація типів земної поверхні на мультиспектральних зображеннях набору даних EuroSAT. Предмет досліджень – високоточні алгоритми класифікації зображень з кількістю каналів тринадцять, які відрізняються невеликим обсягом обчислень. Використання таких алгоритмів для обробки знімків бортовими обчислювачами малих супутників угруповання Sentinel-2, що запускаються в рамках програми Copernicus Європейського космічного агентства з моніторингу довкілля та безпеки, підвищить точність класифікації та зменшить необхідне енергоспоживання, обсяг пам'яті та навантаження каналів зв'язку. Враховуючи експоненційне зростання необхідної кількості параметрів відомих нейронних мереж для підвищення якості класифікації, у роботі запропоновано альтернативний підхід. Він передбачає використання ансамблевих класифікаторів зі стекінгом. На першому ступені пропонується використовувати максимально полегшені варіанти нейронних мереж, що мають відносно високий рівень якості класифікації. На другому ступені – багатошаровий персептрон. У роботі побудовано полегшені варіанти відомих нейронних мереж ResNet, EfficientNet, Compact SS-GhostNet, FastViT, CCT. Використання їх у рамках ансамблевого класифікатора зі стекінгом дозволило отримати якість класифікації на наборі даних EuroSAT дев'яносто дев'ять і п'ятдесят дев'ять сотих відсотка. Це на нуль цілих і чотири десяті відсотки більше за найкращий показник якості класифікації, що забезпечується нейронною мережею CALCNet. При цьому кількість параметрів запропонованого варіанта ансамблевого класифікатора в п'ять разів менше, ніж у CALCNet. Варіанти всіх програм розміщені у GitHub.

Ключові слова: компактний згортковий трансформер; відносне позиційне зміщення; відносну увагу; багатошаровий персептрон; дистанційне зондування; ансамблевий класифікатор; токенізатор; гібридна архітектура; пошарова згортка; інвертовані залишкові блоки

ABOUT THE AUTHORS



Oleg M. Galchonkov - PhD, Assistent Professor, Information Systems Department. Odesa Polytechnic National University, 1, Shevchenko Ave, Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-5468-7299; o.n.galchenkov@gmail.com. Scopus Author ID: 56081377900
***Research field*:** Artificial intelligence; evolution systems; machine learning

Галчонков Олег Миколайвич - кандидат технічних наук, доцент кафедри Інформаційних систем. Національний університет "Одеська політехніка", пр. Шевченка, 1. Одеса, 65044,Україна
Галузь досліджень: Штучний інтелект; еволюційні системи; машинне навчання



Illia O. Baskov - Senior Lecturer, Information Systems Department. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-3517-6773; illyabaskov@gmail.com. Scopus Author ID 57788059000
***Research field*:** Artificial intelligence; evolution systems; machine learning

Баськов Ілля Олександрович - старший викладач кафедри Інформаційних систем. Національний університет "Одеська політехніка", пр. Шевченка, 1. Одеса, 65044,Україна
Галузь досліджень: Штучний інтелект; еволюційні системи; машинне навчання



Mykola I. Babych - PhD, Assistent Professor, Information Systems Department. Odesa Polytechnic National University, 1, Shevchenko Ave, Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-3946-9880; babich.tiger@gmail.com; Scopus Author ID 57253653800
***Research field*:** Artificial intelligence; evolution systems; machine learning

Бабич Микола Іванович - кандидат технічних наук, доцент кафедри Інформаційних систем. Національний університет "Одеська політехніка", пр. Шевченка, 1. Одеса, 65044,Україна
Галузь досліджень: Штучний інтелект; еволюційні системи; машинне навчання



Sergiy Y. Smyk - PhD, Assistent Professor, Information Systems Department. Odesa Polytechnic National University, 1, Shevchenko Ave, Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-7020-1826; smyk@op.edu.ua; Scopus Author ID 56996011100
***Research field*:** Artificial intelligence; evolution systems; machine learning

Смик Сергій Юрійович - кандидат технічних наук, доцент кафедри Інформаційних систем. Національний університет "Одеська політехніка", пр. Шевченка, 1. Одеса, 65044,Україна
Галузь досліджень: Штучний інтелект; еволюційні системи; машинне навчання