Multishape Morphological-based Two-Stage CNN for LiDAR-DSM Classification

Behnam Asghari Beirami

Department of Photogrammetry and Remote Sensing, K. N. Toosi University of Technology, Tehran, Iran,behnam.asghari1370@gmail.com

Abstract

The classification of Digital Surface Model (DSM) images derived from LiDAR sensors is a challenging task, particularly when distinct ground classes with identical height information must be distinguished. However, DSM images contain valuable spatial information that can be utilized to enhance classification accuracy. This paper proposes a novel strategy, called Multishape Morphological Two-Stage Convolutional Neural Network (MM2CNN), for DSM classification to achieve accurate classified land-cover maps. The proposed method combines the strengths of multishape morphological profiles (MMPs) and a two-stage CNN model as a smart algorithm to effectively discriminate between different land covers from a single-band DSM image. More precisely, the CNN, as a smart method, is used to learn hierarchical rich representations of the data, while the MMPs are used to extract spatial information from the DSM imagery. The approach involves generating MMPs with three structuring elements, training three parallel CNN models, and then stacking and feeding the probability maps to a second-stage CNN to predict the final pixel labels. Experimental results on the Trento benchmark DSM image show that the suggested technique achieves an overall accuracy of 97.32% in a reasonable time, outperforming some other DSM classification methods. The success of the MM2CNN technique demonstrates the potential of integrating MMPs with CNN for precise DSM classification, which has a wide range of applications in environmental investigations.

1. Introduction

Light Detection and Ranging (LiDAR) technology represents a great improvement in remote sensing approaches, revolutionizing geospatial data collecting and processing. LiDAR systems use laser pulses to accurately record a three-dimensional representation of the scene as the point cloud data by measuring distances between the sensor and the target objects. These point clouds are processed to generate Digital Surface Models (DSMs), which serve as visual representations of the topographical features of landscapes.

The elevation details within DSMs provide essential supplementary data alongside spectral information obtained from optical remote sensing imagery, enhancing the precision and comprehensiveness of land-cover classification[1-6]. Integrating hyperspectral and LiDAR data has emerged as a promising approach for improving land-cover classification accuracy, leveraging the complementary strengths of spectral and spatial information. Recent studies have focused on developing deep learning models to effectively extract and fuse these multisource datasets, surpassing traditional methods that rely on manual feature engineering. Hong et al. proposed EndNet, a deep encoder-decoder network architecture, for the joint classification of hyperspectral and LiDAR data [7]. When compared to traditional fusion methods such as early, middle, and late fusion, EndNet performs better at fusing multimodal information because it enforces the fused characteristics to reconstruct the multimodal input. This approach highlights the potential of deep learning for capturing complex relationships between spectral and spatial information, leading to more accurate land-cover classification. Using a triple-branch CNN backbone and a hierarchical fusion technique, Wang et al. introduced the multiattentive hierarchical fusion net in another work to fuse hyperspectral and LiDAR data at the feature

level[8]. This approach effectively integrates spectral, spatial, and elevation features, demonstrating improved classification performance on real datasets. Zhao et al. further advanced the area by introducing a dual-branch network that combines a hierarchical CNN with a transformer, enabling the acquisition and learning of spectral-spatial information from hyperspectral data and elevation features from LiDAR data [9]. Their cross-token attention fusion encoder effectively fuses these features, showcasing superior performance compared to existing methods. In another study, Zhang et al. proposed structural optimization transmission networks that incorporate cross-attention modules, dual-mode propagation modules, and dynamic structure optimization modules to enhance the semantic relatedness of multisource data[10]. These efforts demonstrate the necessity for robust frameworks that efficiently utilize the complementary information included in DSM and hyperspectral data, leading to more precise and trustworthy land-cover categorization.

The precise classification of single-band DSM images has garnered increasing attention recently. Accurate DSM categorization is essential for applications such as disaster management, environmental monitoring, and urban planning. Although LiDAR-derived DSM data provides valuable spatial information, traditional machine learning techniques often struggle to distinguish between different land cover types because various classes may have comparable elevation features. Such limitations have encouraged the development of more advanced techniques which could efficiently exploit intrinsic spatial properties in DSM images. In recent years, deep learning approaches have been studied for DSM image classification. Convolutional neural networks (CNNs) as smart algorithms can extract more solid and unique characteristics from DSM images than traditional classification methods, thereby outperforming them. Utilizing CNNs' powerful feature extraction capabilities may greatly improve the accuracy of land-cover categorization using LiDAR-derived DSMs, providing valuable insights for a variety of uses. Even when employing CNN classifiers, researchers often incorporate additional features derived from DSM to enhance classification results. Morphological profiles (MPs) are among the most significant spatial characteristics used in various areas of remote sensing image classification, such as hyperspectral images. For example, Anand et al. proposed a novel approach using morphological summaries with subjective restoration and steering MPs, followed by a supervised feature extraction method to reduce dimensionality for improved image classification [11]. Similarly, Kumar et al. introduced two novel ensemble techniques based on multishape MPs (MMPs) for hyperspectral image classification [12].

In the field of single-band DSM classification, Ghamisi et al. introduced a new method based on the composite kernel Support Vector Machines (SVM) and extinction profiles derived from various LiDAR-based features [13]. Wang et al. developed a technique for DSM classification employing morphological and multiattribute profiles as input features for a CNN with a SiLU activation function[14]. Deep learning techniques have considerable ability to classify LiDAR data, as demonstrated by He et al.'s assessment of the effectiveness of spatial transformer networks coupled with morphological characteristics in classifying DSM images [15]. Wang and colleagues created a superior model for DSM data categorization by combining dense convolutional neural networks with spatial transformer networks [16]. Xie and colleagues introduced a new method for the automatic modular design of CNN architectures tailored for LiDAR data classification, making use of a searchable space to optimize building block selection through a gradient descent algorithm[17]. In a similar vein, Wang and co-authors presented ResCap-Net, a novel deep learning model that merges the strengths of ResNet and capsule networks for LiDAR data classification[18]. To derive accurate and distinguishing features from LiDAR data, Wu and colleagues proposed OctConv-CapsNet, a new DSM classification method that merges Octave Convolution with Capsule Networks[19]. Wang et al. introduced a deep learning method that combines conditional generative neural networks, residual units, and DropBlock to improve the classification of DSM data[20]. Hariyono et al. suggested a method for classifying land cover in a study region in Indonesia using a variety of DSM characteristics, including elevation and intensity, in conjunction with Support Vector Machines (SVM) [21]. In another study, Beirami and Mokhtarzade created a special feature space based on weighted local kernel matrix descriptors derived from morphological profiles (WKLM-MP) to improve DSM classification outcomes [22]. Dong et al. proposed a novel multiscale neighborhood information fusion network for LiDAR image classification, incorporating multiscale patches to capture detailed and general characteristics simultaneously while enhancing category-specific features using a spatial-aware region attention module[23].

The realm of single-band DSM classification lacks a robust body of literature, often relying on rudimentary single-band DSM or single-stage manual spatial feature engineering for classification. This study introduces MM2CNN, a novel two-stage deep learning architecture that seamlessly integrates

multishape morphological profiles (MMPs) and CNN models, addressing the limitations of existing approaches and enabling more comprehensive and accurate classification of DSM data. The hierarchical design of MM2CNN efficiently explores the information underlying the MMPs and probability maps (PMs) to classify DSM images, which have not been taken into consideration in other research. This paper's principal contributions are:

Development of a novel three-branch CNN architecture that integrates MMPs and CNN for accurate LiDAR-DSM classification. Development of two-stage CNN models that utilize PM and MMPs spatial information, two different forms of spatial information, to classify DSM images. The remainder of the paper is organized as follows: Section 2 describes the suggested MM2CNN methodology for DSM classification in detail. Section 3 presents the Trento LiDAR DSM dataset and thoroughly examines the experimental findings. Finally, Section 4 covers the study's conclusions.

2. Methodology

The proposed method, named Multishape Morphological Two-Stage CNN (MM2CNN), utilizes a two-stage CNN architecture for classifying LiDAR-DSM images. In the initial stage, a group of specialized CNN models is trained, with each CNN expert focusing on processing a specific type of MP extracted from the input DSM. The features of MMPs are generated via opening and closing by reconstruction operations using structuring elements (SEs) of various sizes and shapes (disk, square, diamond). This multiscale and multishape method allows for the capture of diverse spatial and geometric characteristics found in the DSM data, which is useful for distinguishing between land cover classes in the DSM image. The output PMs of these MMPs-specialized CNN classifiers are then combined using a second-stage CNN classifier to further refine classification results by considering the underlying information from the PMs generated from the previous stage. The final land-cover classification is produced by the second-stage CNN. The suggested MM2CNN approach for DSM classification is illustrated in Figure 1 and is divided into the following five phases:

Extraction of MMPs from the DSM image. In this paper, we use three types of MP resulting from SEs with shapes of disk, square, and diamond.

Considering three experimentally configured CNN models in parallel, each is fed with a distinct type of MP.

Calculating the PMs results of all three CNN models.

Stacking all the PMs generated from CNN models to form a datacube with the third data dimension equal to 3×number of classes.

To obtain the final class label of each pixel in the DSM image, the newly created datacube, formed by stacking the PMs, is fed to a new CNN model.
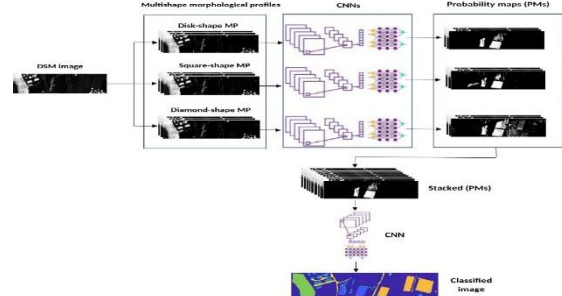

Fig. 1.   The DSM classification procedure using the MM2CNN method.

A) Multishape Morphological Profiles

MPs are widely known as a valuable tool for extracting spatial features from remote sensing imagery. These features are constructed through morphological opening (*OP*) and closing (*CL*) by reconstruction operations using SEs of varying sizes, as depicted in the equation below [14]:

$$MP(x)=[Cl_i(x),...,DSM(x),...OP_i(x)] \tag{1}$$

However, MP approaches that rely on fixed-shape or fixed-size SE might not be able to fully capture every geometric aspect found in DSM images. As a result, it has been discovered that using different SE shapes, such as disk, square, and diamond, is more successful in modeling these properties, resulting in the creation of MMP features, which are represented as [24]:

$$MMPs(x)=[MP_{Disk}(x),MP_{Square}(x),MP_{Diamond}(x)] \tag{2}$$

B) CNN

The Deep CNN is an advanced smart algorithm commonly used in image recognition and computer vision tasks. It is a smart method in the sense that it can learn to recognize patterns and features within images through training on a large dataset. CNNs are designed to recognize patterns in data by leveraging hierarchical representations. The CNN's architecture typically involves a series of blocks, including convolutions (used as the feature extractor), spatial pooling, and nonlinear activation, which ultimately feed into a classification module

[25]. Additionally, by employing more layers, the CNN can extract more abstract representations of the input data. The batch normalization technique is also employed to enhance the stability and speed of the learning process.

The training method for CNN models includes two main phases: the forward pass and the backward pass. During the forward pass, the network processes input image patches through a sequence of main blocks, such as convolution and pooling layers, generating high-level feature representations that capture complex patterns. The output of these layers is then reshaped into a vector of deep features, which is then fed into fully connected layers for classification. The network's weights are adjusted during the backward phase, where the error between predicted and actual labels is minimized [26]. Finally, after these two phases, the tuned weights of the network are used to predict the label of each pixel.

3. Dataset and experiments

C) Dataset

The Optech ALTM 3100EA sensor collected the Trento DSM dataset, which spans a rural area in Trento's southern region, Italy. As seen in Figure 2, this dataset has six different rural classes: apple trees, houses, ground, wood, vineyard, and roads. DSM image has 166×600 pixels with a spatial resolution of 1 meter.
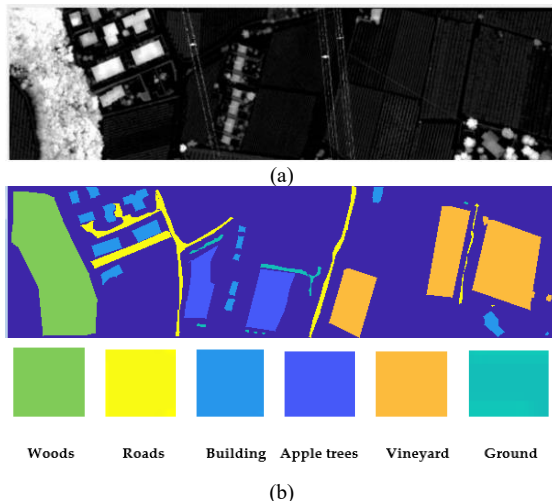

(a)

| Woods | Roads | Building | Apple trees | Vineyard | Ground |

(b)

Fig. 2.   Trento dataset, (a) DSM image, (b) Ground truth image.

D) Experiments

The proposed MM2CNN method utilizes three distinct SE shapes, namely disk, square, and diamond, with SE dimensions ranging from 2 to 25, incremented by 2. In this study, the MATLAB 2020b deep learning toolbox was utilized to implement the proposed MM2CNN method. The CNN classifier consists of two primary layers:

convolutional layers with a 2x2 kernel size, 64 filters, and batch normalization, followed by a tanh activation function and a pooling layer. Finally, the PMs or labels are determined using a single fully-connected layer. The input window size is set to 9x9, and the Adam optimizer is run with a maximum epoch of 200. All other parameters of the CNN network are at their default in MATLAB 2020b. We conducted our experiments using a personal computer equipped with a Core i5 4590 3.3GHz CPU and 8 GB RAM.

In this paper, the performance of the proposed MM2CNN method will be evaluated using the Trento DSM dataset, in which a random sampling strategy is applied to choose 40 training samples for each class from the ground truth image, while others are used as test samples in order to calculate the classification accuracy. The method's performance was statistically assessed using three measures derived from the confusion matrix [27]: overall accuracy (OA), average accuracy (AA), and kappa coefficient. These metrics provide a comprehensive understanding of the method's ability to accurately classify the DSM data.

The proposed MM2CNN method is comprehensively evaluated against several other DSM classification methods. A detailed description of these competing methods is provided below:

DSM: This method involves the direct classification of the original single-band DSM image using a CNN.

B1-disk: This approach which is proposed in [14], analyzes the results of the first branch of the parallel CNN system, which utilizes MP with a disk-shaped SE as input.

B2-square: This method examines the outcomes of the first branch of the parallel CNN system, which employs MP with a square-shaped SE as input.

B3-diamond: This approach evaluates the results of the first branch of the parallel CNN system, which utilizes MP with a diamond-shaped SE as input.

MV: This method employs the conventional majority voting (MV) strategy to combine the classified images generated by each branch of the CNN.

Stacking: This approach involves feature-level fusion by stacking, which combines the MPs and then uses a CNN to predict the labels of pixels.

WLKM-MP: As the recent approach, this method which is proposed in [22], utilizes the weighted local kernel matrix features of MP, generated from the DSM, to classify the DSM image.

Table 1 shows classification outcomes for each mentioned approach. Figure 3 displays classified images of the DSM and MM2CNN methods,

providing a visual representation of the classification results. The key findings from the experimental results of this study are as follows:

The use of MPs as spatial features significantly enhances the classification of DSM data. The OA of DSM classification increased from 87.49% to at least 94.2% when the diamond-shaped MP was used.The results of the MV, stacking, and MM2CNN methods collectively suggest that combining MPs with multishape SEs outperforms the use of MPs with single-shape SE.

A comparative analysis of the MV, stacking, and proposed MM2CNN results proved that the proposed CNN-based decision fusion strategy has more flexibility, which can effectively explore the spatial information embedded in PMs, thereby outperforming other information fusion methods. In comparison to the recently proposed WLKM-MP method for DSM classification, the proposed MM2CNN approach has the potential to achieve higher levels of classification accuracy.

Table 1 shows how the suggested MM2CNN outperforms other approaches in all of the three classification indices. This superiority is attributed to the effective ability of the MM2CNN method to incorporate the spatial information present in the DSM image. The proposed MM2CNN method demonstrates a processing time of around 80 seconds, a duration that is reasonable considering the moderate specifications of the operating system and the deep learning framework of the approach. It is notable that the MM2CNN method possesses the capacity for parallel implementation, offering the potential for further optimization of processing speeds.

Figure 3 shows that the classified image produced by the DSM method is characterized by high levels of noise. In contrast, the classified image produced by the proposed MM2CNN method exhibits a smoother appearance, which is consistent with the quantitative findings of this study.

Table.1.
The classification results

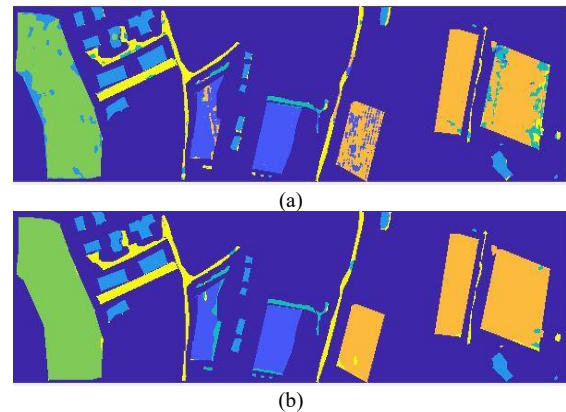| Mathods	Accuracy index		
	OA	AA	Kappa
DSM	87.49%	87.06%	0.837
B1-disk	95.96%	93.98%	0.946
B2-square	95.27%	93.99%	0.937
B3-diamond	94.2%	93.38%	0.923
MV	97%	95.51%	0.96
Stacking	96.1%	95.84%	0.948
WLKM-MP	93.42%	93.47%	0.912
Proposed MM2CNN	97.32%	96.85%	0.964

Fig. 3.   Classified images, (a) DSM method, (b) Proposed MM2CNN method.

4. Conclusions

This paper proposes MM2CNN, a new deep learning-based technique for DSM image classification. The proposed MM2CNN uses a novel deep learning strategy, comprising two stages of smart CNN models, to efficiently extract and integrate relevant features for DSM classification. In the initial phase, three parallel CNN models are employed to generate PMs from three distinct types of MPs as spatial features. Subsequently, the generated PMs are fed into a new CNN, which acts as an information fusion strategy to produce a smoother and more accurate DSM-classified image. The experiments on the Trento benchmark DSM dataset show that the MM2CNN technique achieves a high level of classification accuracy, with an OA of over 97%. This outperforms the OA achieved with CNN classification of single-band DSM image alone by approximately 10%. The suggested approach outperforms certain existing DSM classification strategies in terms of classification accuracy. The effective information fusion strategy of the proposed MM2CNN approach, based on a two-stage CNN, as a powerful smart algorithm with multishape morphological characteristics, highlights the possibility for additional developments in the classification of remote sensing data. Future research efforts could explore the application of this framework to other types of geospatial data, as well as the development of more sophisticated deep learning architectures to further enhance classification performance.

References

[1] Michelini, D., Dalponte, M., Carriero, A., Kutchartt, E., Pappalardo, S. E., De Marchi, M., & Pirotti, F, "Hyperspectral and LiDAR data for the prediction via machine learning of tree species, volume and biomass: A contribution for updating forest management plans" Italian Conference on Geomatics and Geospatial Technologies; 2022: Springer. doi: 10.1007/978-3-031-17439-1_17

[2] Liu Y, Ye Z, Xi Y, Liu H, Li W, Bai L, "Multi-Scale and Multi-Direction Feature Extraction Network for Hyperspectral and LiDAR Classification" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2024, vol. 17, pp. 9961 - 9973, doi: 10.1109/JSTARS.2024.3400872.

[3] Zhang H, Yao J, Ni L, Gao L, Huang M, "Multimodal attention-aware convolutional neural networks for classification of hyperspectral and LiDAR data" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2022, vol.16, pp. 3635-3644, doi: 10.1109/JSTARS.2022.3187730.

[4] Du X, Zheng X, Lu X, Wang X, "Hyperspectral and LiDAR Representation with Spectral-Spatial Graph Network" IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2023. vol. 16, pp. 9231 - 9245, doi: 10.1109/JSTARS.2023.3321776.

[5] Li J, Liu Y, Song R, Liu W, Li Y, Du Q, "HyperMLP: Superpixel Prior and Feature Aggregated Perceptron Networks for Hyperspectral and Lidar Hybrid Classification" IEEE Transactions on Geoscience and Remote Sensing. 2024, vol. 62, doi: 10.1109/TGRS.2024.3355037.

[6] Roy SK, Sukul A, Jamali A, Haut JM, Ghamisi P. "Cross hyperspectral and LiDAR attention transformer: An extended self-attention for land use and land cover classification" IEEE Transactions on Geoscience and Remote Sensing. 2024, vol. 62, doi: 10.1109/TGRS.2024.3374324.

[7] Hong D, Gao L, Hang R, Zhang B, Chanussot J, "Deep encoder–decoder networks for classification of hyperspectral and LiDAR data" IEEE Geoscience and Remote Sensing Letters. 2020, vol. 19, pp. 1-5, doi: 10.1109/LGRS.2020.3017414.

[8] Wang X, Feng Y, Song R, Mu Z, Song C, "Multi-attentive hierarchical dense fusion net for fusion classification of hyperspectral and LiDAR data" Information Fusion, 2022, vol. 82, pp. 1-18, doi: 10.1016/j.inffus.2021.12.008.

[9] Zhao G, Ye Q, Sun L, Wu Z, Pan C, Jeon B, "Joint classification of hyperspectral and LiDAR data using a hierarchical CNN and transformer" IEEE Transactions on Geoscience and Remote Sensing, 2022, vol.61, pp. 1-16, doi: 10.1109/TGRS.2022.3232498.

[10] Zhang M, Li W, Zhang Y, Tao R, Du Q, "Hyperspectral and LiDAR data classification based on structural optimization transmission" IEEE Transactions on Cybernetics. 2022, vol. 53, no. 5, pp. 3153 - 3164, doi: 10.1109/TCYB.2022.3169773.

[11] Anand R, Veni S, Geetha P, Subramoniam SR, "Extended morphological profiles analysis of airborne hyperspectral image classification using machine learning algorithms" International Journal of Intelligent Networks, 2021, vol. 2, pp. 1-6, doi: 10.1016/j.ijin.2020.12.006.

[12] Kumar B, Dikshit O, "Hyperspectral image classification based on morphological profiles and decision fusion" International Journal of Remote Sensing, 2017, vol.38, no.20, pp. 5830-54, doi: 10.1080/01431161.2017.1348636.

[13] Ghamisi P, Hoefle B, "LiDAR data classification using extinction profiles and a composite kernel support vector machine" IEEE Geoscience and Remote Sensing Letters, 2017, vol.14, no.5, pp. 659-63, doi: 10.1109/LGRS.2017.2669304.

[14] Wang A, He X, Ghamisi P, Chen Y, "LiDAR data classification using morphological profiles and convolutional neural networks" IEEE Geoscience and

[15] Remote Sensing Letters, 2018, vol.15, no.5, pp.774-8, doi: 10.1109/LGRS.2018.2810276.

[15] He X, Wang A, Ghamisi P, Li G, Chen Y, "LiDAR data classification using spatial transformation and CNN" IEEE Geoscience and Remote Sensing Letters, 2018, vol.16, no.1, pp.125-9, doi: 10.1109/LGRS.2018.2868378.

[16] Wang A, Wang M, Jiang K, Zhao L, Iwahori Y, "A novel lidar data classification algorithm combined densenet with STN" IEEE IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium;2019, doi: 10.1109/IGARSS.2019.8900313.

[17] Xie H, Chen Y, "LiDAR data classification based on automatic designed CNN" IEEE Geoscience and Remote Sensing Letters, 2020, vol.18, no.9, pp.1665-9, doi: 10.1109/LGRS.2020.3005209.

[18] Wang A, Wang M, Wu H, Jiang K, Iwahori Y. "A novel LiDAR data classification algorithm combined capsnet with resnet" Sensors, 2020, vol.20, no.4, pp.1151, doi: 10.3390/s20041151.

[19] Wu H, Cao M, Wang A, Wang M. "Classification of LiDAR data combined octave convolution with capsule network" IEEE Access, 2020, vol.8, pp.16155-65, doi: 10.1109/ACCESS.2020.2965278.

[20] Wang A, Xue D, Wu H, Iwahori Y, "LiDAR data classification based on improved conditional generative adversarial networks" IEEE Access. 2020 vol.8, pp.209674-86, doi: 10.1109/ACCESS.2020.3039211.

[21] Hariyono M, Tambunan M, Dewi R, "Support vector machine for land cover classification using lidar data" IOP Conference Series: Earth and Environmental Science; 2021, doi: 10.1088/1755-1315/873/1/012095.

[22] Asghari Beirami B, Mokhtarzade M, "Land Covers Classification from LiDAR-DSM Data Based on Local Kernel Matrix Features of Morphological Profiles" International Journal of Engineering, 2023, vol.36, no. 9, pp.1611-7, doi: 10.5829/IJE.2023.36.09C.04.

[23] Dong J, Liu K, Han J, Zhang M, Zhao X, Li W, X Li, M Rao, "Multi-scale Neighborhood Information Fusion Network for Classification of Remote Sensing LiDAR Images" IEEE Sensors Journal, 2024, vol.24, no.10, pp. 16601 - 16613, doi: 10.1109/JSEN.2024.3386173.

[24] Beirami BA, "Face Recognition based on Multi-shape Morphological Profiles-based Covariance Descriptors and Log-Etuclidean Kernel SVM" IEEE 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS); 2022, doi: 10.1109/CFIS54774.2022.9756491.

[25] Sharifi O, Mokhtarzadeh M, Asghari Beirami B, "A new deep learning approach for classification of hyperspectral images: Feature and decision level fusion of spectral and spatial features in multiscale CNN" Geocarto International, 2022, vol.37, no.14, pp.4208-33, doi: 10.1080/10106049.2021.1882006.

[26] Beirami BA, Mokhtarzade M, "A new deep learning approach for hyperspectral image classification based on multifeature local kernel descriptors" Advances in Space Research, 2023, vol.72, no.5, pp.1703-20, doi: 10.1016/j.asr.2023.04.025.

[27] Beirami BA, Mokhtarzade M, "SVM classification of hyperspectral images using the combination of spectral bands and Moran's I features" 2017 IEEE 10th Iranian Conference on Machine Vision and Image Processing (MVIP), 2017, doi: 10.1109/IranianMVIP.2017.8342334.