Hindawi

Research Article

Urbanization Detection Using LiDAR-Based Remote Sensing Images of Azad Kashmir Using Novel 3D CNNs

Mazhar Hameed [ID],[1] **Fengbao Yang** [ID],[1] **Sibghat Ullah Bazai** [ID],[2]
Muhammad Imran Ghafoor [ID],[3] **Ali Alshehri**,[4] **Ilyas Khan** [ID],[5] **Mehmood Baryalai** [ID],[6]
Mulugeta Andualem [ID],[7] **and Fawwad Hassan Jaskani**[8]

[1]*School of Information and Communication Engineering, North University of China, Taiyuan, China*
[2]*Department of Computer Engineering, Balochistan University of Information Technology, Engineering, and Management Sciences, Quetta, Pakistan*
[3]*Engineering Department Satellite Section Pakistan Television Corporation Lahore, Pakistan*
[4]*Department of Computer Science Applied College University of Tabuk Saudi Arabia, Saudi Arabia*
[5]*Department of Mathematics, College of Science Al-Zulfi, Majmaah University, Al-Majmaah 11952, Saudi Arabia*
[6]*Department of Information Technology, Balochistan University of Information Technology, Engineering, and Management Sciences, Quetta, Pakistan*
[7]*Department of Mathematics, Bonga University, Bonga, Ethiopia*
[8]*Department of Computer Systems Engineering, The Islamia University of Bahawalpur, Pakistan*

Correspondence should be addressed to Mazhar Hameed; mazhar_hameed@hotmail.com
and Mulugeta Andualem; mulugetaandualem4@gmail.com

An important measurable indicator of urbanization and its environmental implications has been identified as the urban impervious surface. It presents a strategy based on three-dimensional convolutional neural networks (3D CNNs) for extracting urbanization from the LiDAR datasets using deep learning technology. Various 3D CNN parameters are tested to see how they affect impervious surface extraction. For urban impervious surface delineation, this study investigates the synergistic integration of multiple remote sensing datasets of Azad Kashmir, State of Pakistan, to alleviate the restrictions imposed by single sensor data. Overall accuracy was greater than 95% and overall kappa value was greater than 90% in our suggested 3D CNN approach, which shows tremendous promise for impervious surface extraction. Because it uses multiscale convolutional processes to combine spatial and spectral information and texture and feature maps, we discovered that our proposed 3D CNN approach makes better use of urbanization than the commonly utilized pixel-based support vector machine classifier. In the fast-growing big data era, image analysis presents significant obstacles, yet our proposed 3D CNNs will effectively extract more urban impervious surfaces.

1. Introduction

Global urbanization has accelerated dramatically in the previous few decades. Around 68% of the global population will be living in cities by 2050 [1]. It can lead to various environmental problems, such as ecological issues, poor air quality, deterioration of public health, and changes in the microcli-mate that can lead to extreme weather, higher temperatures, limited water availability, and a continued vulnerability to natural disasters [2]. Due to the spectral and structural diversity and complexity of urban surfaces over a small area, these issues make sophisticated urban analysis difficult [3]. As a result, it is imperative to keep an eye on urban areas at all times. In metropolitan settings, where many items are

mobile (vehicles and temporary buildings) and where infrastructure, vegetation, and construction are continuously changing, systematic monitoring and updating of maps are vital.

With the advancements in remote sensing technologies [4], spatial and temporal analyses of metropolitan areas are now possible [5]. A strong new tool for urban analysis is airborne remote sensing, which provides fast mapping of a city for various planning [6], management operations [7], and monitoring of urban and suburban land use [8] activities. For example, social preferences, the regional ecosystem, urbanization change, and biodiversity can all be studied using this method [2]. An important component of studying three-dimensional city geometry and modelling urban morphology is to use urban remote sensing to identify various objects, heterogeneous materials, and mixes [9, 10]. However, the developing problems necessitate a cutting-edge technology solution, including sensors and analysis methodologies at the cutting edge of development. Urban land cover types can be identified using spectral, spatial, and structural aspects of remote sensing devices, constantly improving. There has been an increase in the use of LIDAR (light detection and ranging) in urban mapping, as well as hyperspectral data (HS) and synthetic aperture radar (SAR). Urban settings can be studied using various electromagnetic spectrum components, ranging from microwave radar to the reflective spectral range. Since they require oblique illumination of the scene, these high-resolution photographs are subject to occlusion and layover, making it difficult to analyze dynamic metropolitan environments [11, 12].

Single sensor urban land cover classification accuracy and interpretability in densely populated areas are frequently inadequate [13, 14]. Extensive analyses are required because of the substantial spectrum variance within a single land cover type caused by urban heterogeneity. The spectral and spatial-structural properties of urbanization (such as roofs, parking lots, highways, and pavements) differ significantly. Urban heterogeneity can be estimated using several methods, including scale and geographic resolution. Heterogeneity is defined by scale as the absence or grouping of materials into a single class, such as individual trees, forest type, or vegetation in general [15, 16]. The amount of pixel mixing depends on the spatial resolution. Nevertheless, increased spatial resolution increases the variability of the physical substance, making analysis more challenging.

Materials can be distinguished without reference to elevation context using HS data, which provides spectral information about the studied materials. The problem with pure spectral analysis is that it ignores object identification, and most objects are made of various materials, resulting in extremely high intraobject heterogeneity. As an example of how LIDAR data differs from other data types, consider asphaltic open parking lots and highways, which have varying heights of coverage [17, 18]. As a result, passive remote sensors such as HS are more sensitive to changes in air conditions and lighting than active sensors such as LIDAR are. When paired with HS data, this LIDAR characteristic allows for physical correction of shadow and lighting and intensity measurement for urban land cover mapping in shadowed areas.

Urban environments include spectral ambiguity and reduced spectral value under the shadow of terrain changes, buildings, and trees, which can be solved by adding LIDAR data as provided by [19, 20], disregarding the spatial and spectral resolution of airborne-based HS sensors. Many new urban surface classification methods use active and passive remote sensing techniques such as airborne LIDAR and hyperspectral data to overcome the limitations of individual sensor capabilities (HL-Fusion). Using an HL-Fusion for land cover classification can provide additional information on three-dimensional topography, spatial structure, and spectral data.

For this reason, it is important to combine spectral, spatial, and elevation data while studying the city environment [21, 22]. The airborne HL-Fusion has been tested for use in the classification of urban land cover. On the other hand, various combination methods based on physical or empirical approaches are put into practice at various data and product levels. Furthermore, due to the complexity of the fusion processes, no standard framework exists for fusing these sensors. As a result, a complete review of prior data fusion research may help researcher's better grasp the potential, constraints, and prevalent issues that limit categorization outcomes in cities [23–25].

Classifiers for HS data have been developed using machine learning (ML) approaches. Different mapping approaches are used to achieve the categorization goal depending on the target. Algorithms for machine learning are constantly being improved, allowing them to extract increasingly complex features in an organized manner. Deep learning, a machine learning subfield, is credited with developing this capability (DL) [26, 27]. Spatiospectral feature extraction from HS data has been accomplished using DL [1, 2]. Both ML and DL methods are useful for classifying data in remote sensing, although various algorithms are better at extracting different attributes from pixels or objects. It is necessary to understand the features that can be extracted before selecting a classification technique for HS data [28, 29]. Due to its ability to uncover unique deep characteristics at the pixel level, however, a per-pixel classification can produce noisy findings in an urban setting considering the large spatial distribution.

When the training dataset is insufficient, classification results will be limited in performance and accuracy since they heavily depend on the amount of training samples. We looked at incorporating contextual information around pixels and object-oriented classification to improve the findings and reduce heterogeneity. ML and DL algorithms [23, 30] outperform classical classifiers in the urban setting, especially when trained quickly.

The initial goal of this study is to learn more about 3D CNNs' ability to extract urban urbanization from high-resolution (HR) data. Researchers compared the results to those of support vector machine (SVM) and 2D CNN approaches to see if it worked. We've also shown how the suggested 3D CNN model's various parameters affect impervious surface extraction.

2. Literature Review

To map impermeable surfaces, scientists have relied on satellite images with a range of spatial and temporal

resolutions. Medium and low spatial resolution images, such as Landsat and MODIS data, are appropriate for large-scale urbanization mapping due to their spectrum information and high temporal resolution [26, 27]. However, in large-area mapping, issues with mixed pixels confuse when extracting impermeable surfaces. High spatial resolution images can provide a lot of information regarding land use and land cover. Still, the spectral similarities of diverse objects, as well as the shadows created by large structures or enormous trees, make it difficult to extract urbanization [28, 29]. Images captured with hyperspectral technology eliminate the spectral homogeneity of a land class's spectral similarity while also reducing the spectral heterogeneity of various things [21, 22]. Using SAR, land information previously obscured by clouds can be retrieved, and impervious surfaces under huge tree crowns can be found using SAR images [19, 20]. However, the SAR image's coherent noise poses a substantial challenge for impermeable surface extraction. As a result, urban impervious surface mapping using single-source imagery has several limitations. Several datasets derived from diverse image capture technologies have lately been deemed useful in addressing these difficulties [9, 17], including medium–low spatial resolution (MLR) photos, optical photos, SAR photos, high spatial resolution (HSR) photos, and light detection and range (LiDAR) data, to name a few [16, 19]. The height information provided by LiDAR data can greatly distinguish between objects with identical spectral characteristics, which can help with impervious surface extraction [18]. For example, even though the spectral aspects of the houses, roads, and bare land are often comparable, there is a significant height disparity. To cope with various objects with a similar spectrum, LiDAR height information is useful. In addition, buildings' roofs are flatter than trees' crowns. Buildings and trees can be distinguished using the LiDAR height variance.

However, for the most part, picture categorization was carried out using one- or two-dimensional (1D/2D) CNNs. Three-dimensional (3D) CNNs can model spatial, textural, spectral, and other information simultaneously because of their 3D convolutional function. Although 3D CNNs are already widely utilized for movies and volumetric images, their performance in extracting urban urbanization from satellite images has yet to be proved.

3. Methodology

The best performance in remote sensing comes from CNNs or multistage feed-forward neural networks. Computer advancements such as high-performance GPUs and rectified linear units are being used to create CNNs on a large scale (ReLU), and dropout or data augmentation methodologies have just recently been practical in remote sensing, a long time after they were first proposed. We propose employing 3D convolutions to calculate multiple LiDAR data features using a 3D kernel in CNNs.

The 3D CNN employed in our work is based on the notion of a convolutional neural feed-forward network with numerous hidden layers to enhance abstraction capabilities as shown in Figure 1.
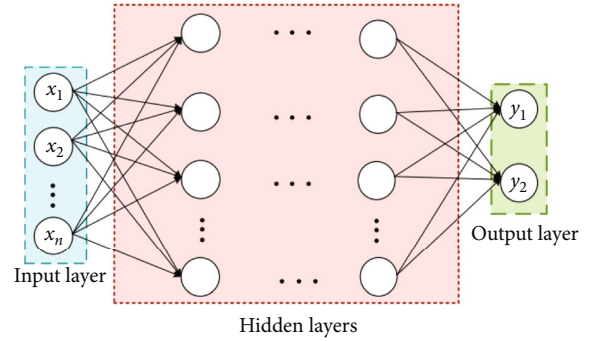

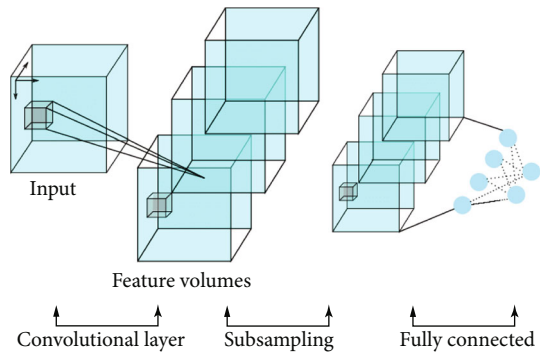Figure 1: Deep neural network model.

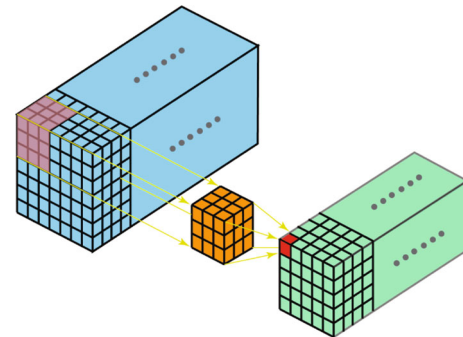Figure 2: 3DCNN image classification model.

Figure 3: The output layer employs softmax nonlinearity in multiclass logistic regression.

Table 1: Samples of sets of pixels selected in study.

Class	Training	Validation	Test
Urban	2000	200	200
Vegetation	1000	200	200
Bare soil	300	200	200

The input layer of the DNN structure employed in this study has a set of 64, 32, 16, or 8 neutrons. After that, we used four dense layers with 210, 29, 28, and 27 neurons, followed by a sigmoid classification layer with two outputs to demonstrate the abnormal and benign traffic categorization. For the experiment purpose, only five neurons with
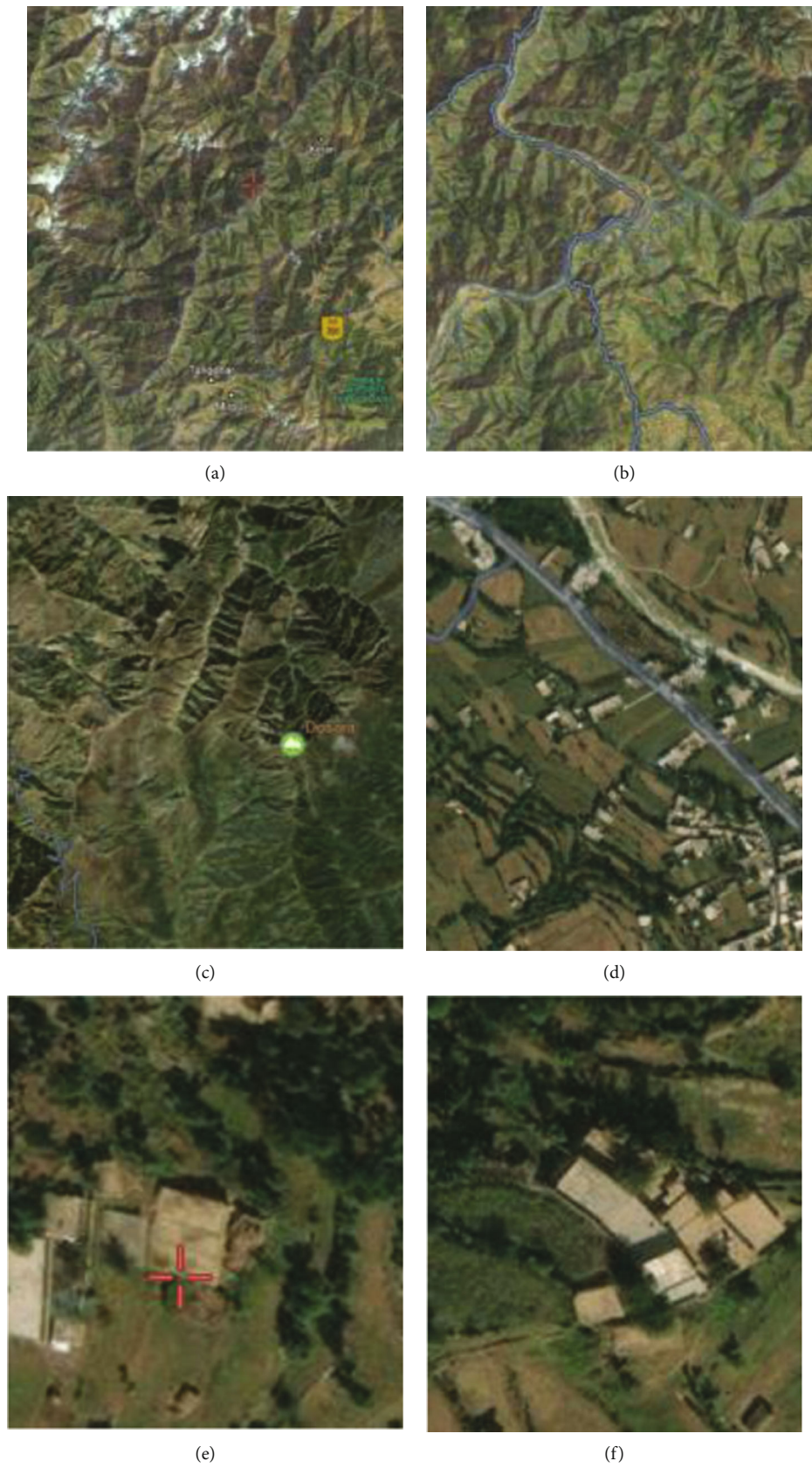
(a)

(b)

(c)

(d)

(e)

(f)

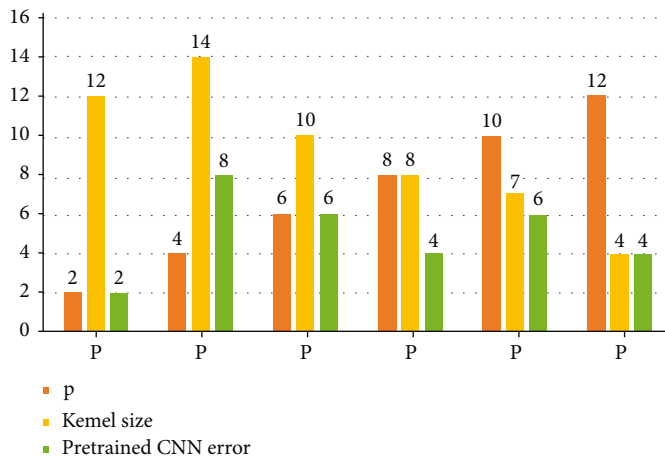FIGURE 4: Sample images taken from datasets.

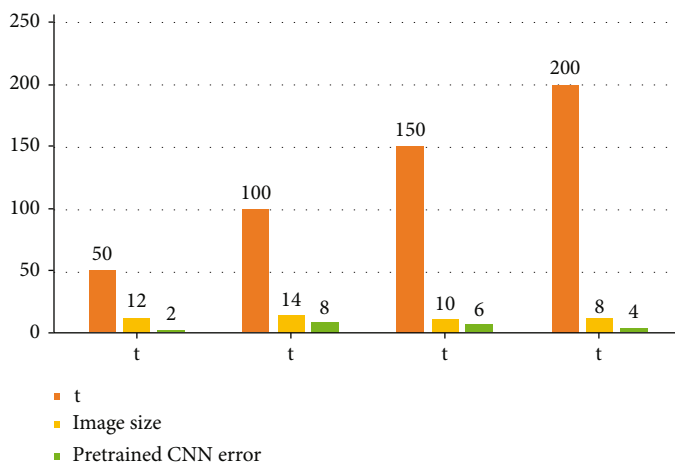FIGURE 5: Pixel sizes with pretrained kernel size and error %.

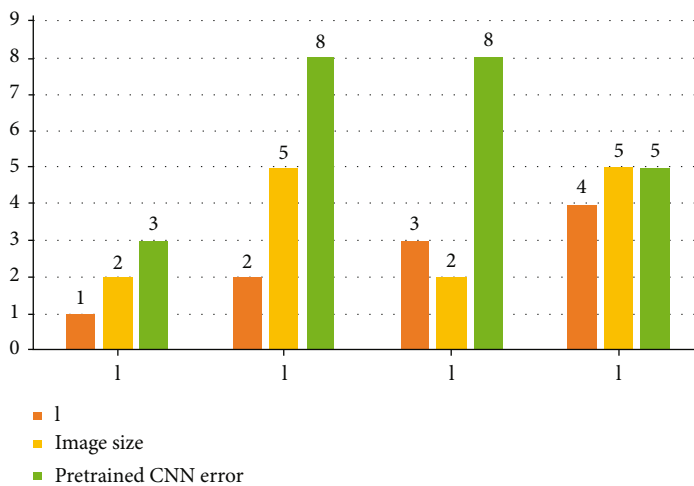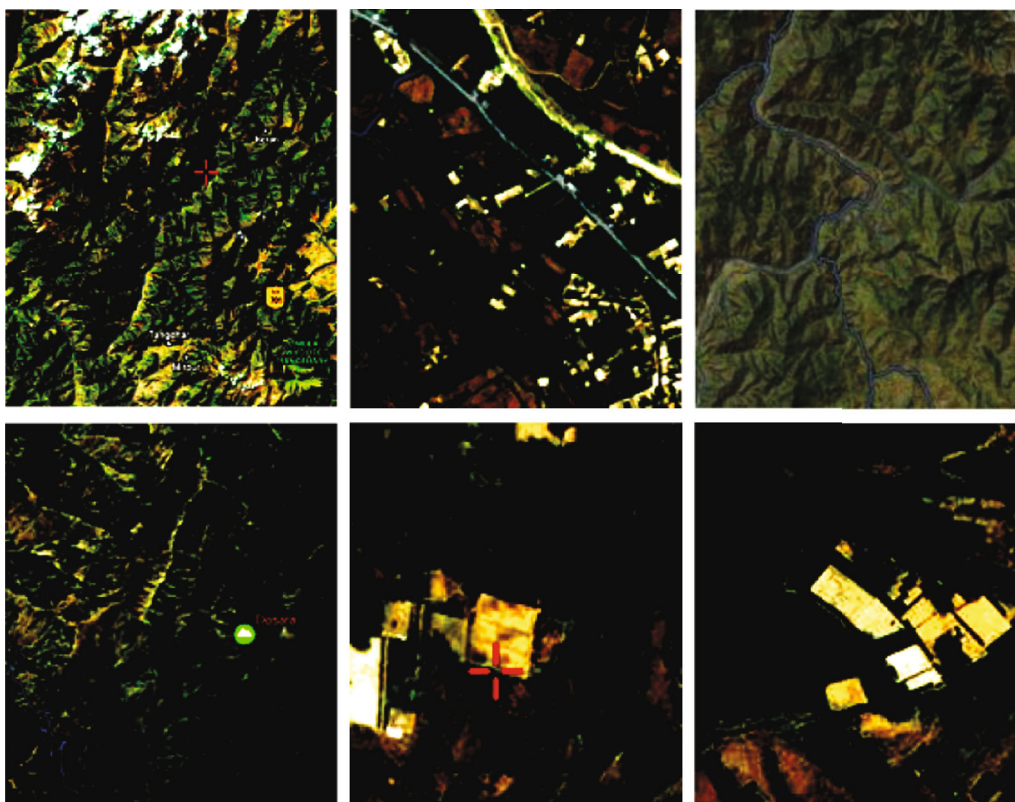FIGURE 6: Impact of t on image size and error %.

FIGURE 7: Impact of L on size of image and error %.

(a) Input image

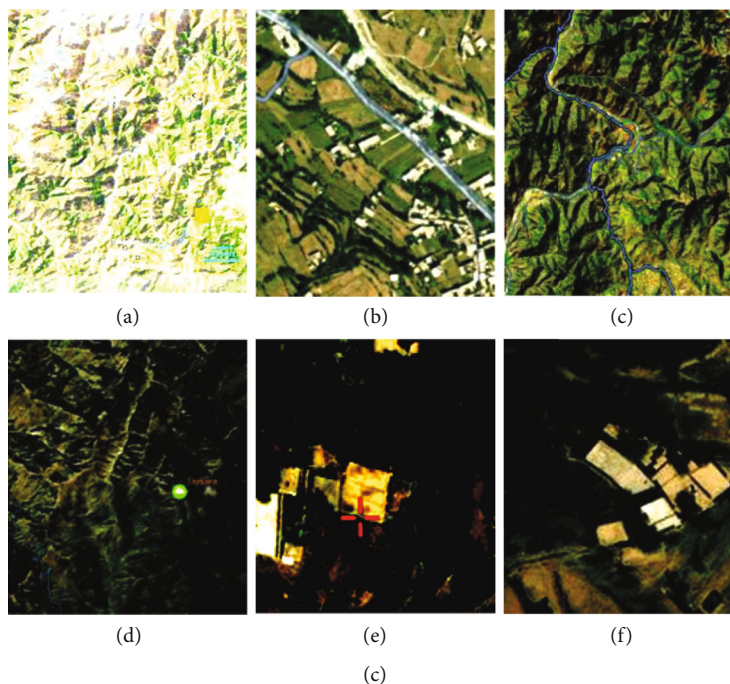(b) Feature extraction

FIGURE 8: Continued.

FIGURE 8: Urban areas classified from Azad Kashmir datasets using LiDAR to 3D CNNs.

numerical and category information are used in the input layer; after that, there are two thick layers with 28 and 27 neurons, as well as an output layer with a sigmoid activation function, which determines if the mobile activity is benign or pathological. The equation can express deep neural network model as

$$a_j^i = \sigma \left(\sum_k w_{jk}^i a_k^{l-1} + b_j^i \right) \cdots, \qquad (1)$$

where a_j^i is the jth neuron in the (l-1) th activation layer. And overall is the sum of k neurons in the l-1th layer.

3.1. 3D CNN Model Architecture. A 3D CNN is a powerful model for learning representations for volumetric data since it uses a 3D volume or a sequence of 2D frames as input. When extracting 3D characteristics or establishing a 3D association, 3D CNNs are utilized. Convolutions in three dimensions (or 3D) apply a three-directional (3-direction) filter to the dataset to calculate low-level feature representations. Typically, their output takes the form of a cube or cuboid. A 3D filter can move in all three directions when used in 3D convolution (height, width, and channel of the image). The element-by-element multiplication and addition yield one number at each place. Due to the filter's 3D movement, the output numbers are also arranged in 3D. The final result is 3D data. Figure 2 shows the 3DCNN image classification model.

3.2. Model Training. Buildings, roads/other forms of urbanization, trees, grasslands, and bare soils are the five land cover classifications in our research area. The number of

classes in our research field determines the size of the output layer. For the output layer, softmax nonlinearity is used in multiclass logistic regression. As a result, a K-dimensional vector is created, with each member representing the likelihood of a specific class occurring in the dataset. In a minibatch B, for each input sample. Figure 3 shows the output layer employs softmax nonlinearity in multiclass logistic regression.

3.3. Hypertuning of Parameters of 3D CNN Model. The 3D CNN model's starting parameters significantly impact the extracted information and classification results. To improve the training accuracy of the CNN model, it is critical to choose an ideal choice of CNN hyperparameters. The training and validation samples were used to select hyperparameters in our work. The size of the input image (m), the size of the convolution kernel (n), the pooling dimension (p), the number of feature mappings (t), and the number of CNN layers are all investigated to see how they influence classification accuracy (L). The best hyperparameters are then used to build a 3D CNN model.

First, we use different convolutional kernel sizes $n = [2, 4, 6, 8, 10, 12, 14, 16, 18, 20]$ and pooling dimensions $p = [2, 4, 6, 8, 10, 12]$ with input image layers mmcd ($c = 10$) when $m = 25$, $L = 1$, and $t = 50$ to test the accuracy of 3D CNNs. The best n and p parameter combination may be found using epoch. Second, we investigate the accuracy of various input picture sizes m and output feature counts t using ideal n and p values and a preset parameter $L = 1$. This iterative approach is used to find the best values for m and t. Finally, we examine how changing input picture sizes ($m = [3, 5, 7]$) and the number of CNN layers ($L = [1, 2, 3]$) affect the classification performance of CNNs with varied layer counts utilizing the optimal combination of n and p,

TABLE 2: Comparative study of current work with previous approaches.

Techniques	Class	Accuracy
	Urban	95%
Proposed 3DCNN+LiDAR	Rural	95.6%
	Vegetation	90%
	Urban	89%
SVM+LiDAR [18]	Rural	88%
	Vegetation	89%
	Urban	90%
2D CNN [17]	Rural	90%
	Vegetation	90.5%
	Urban	86%
2D CNN+SVM LiDAR [11]	Rural	86.43%
	Vegetation	87%

m, and t. m and L are revised iteratively until they are discovered to be optimal.

4. Results and Discussions

There is an experimental location in Azad Kashmir, State of Pakistan. The WV-2 data was collected on October 2021, whereas the aerial LiDAR data was collected in August 2021. The period of data has been selected from 2016 to 2021. Based on the 3D CNN model, we divided the study area into three land classes: urban surfaces, vegetation, and bare soil. Table 1 shows the pixels utilized in 3D CNN classification for training, validation, and testing.

Using deep learning approaches for remote sensing picture categorization, we used around the same number of training samples as other studies, with a total of about 3000–7000. 3D CNN model training uses training samples, validation samples for hyperparameter adjustment, and testing samples for ultimate accuracy assessment. Figure 4 shows sample images taken from datasets.

4.1. Hypertuning of Model. To determine the best 3D CNN parameters for model construction, randomly selected training and validation samples were used to evaluate their performance. Figure 5 shows the pixel sizes with pretrained kernel size and error %, while Figure 6 shows the impact of t on image size and error %.

Figures 5, 6, and 7 show the pretraining accuracy of 3D CNN models influenced by different 3D CNN model parameters. Parameters n and p have an effect, and parameters m and t perform well. Parameters m and L perform well as well. (A pixel is the smallest possible unit of image input size.)

Figure 8 shows the urban areas classified from Azad Kashmir datasets using LiDAR to 3D CNNs. The error matrix approach is used to conduct the accuracy assessment's error matrix.

Various accuracy metrics are calculated, including the producer's accuracy, the user's accuracy, the system as a whole, and the overall kappa coefficient (Table 2).

5. Conclusions

This study proposed and used to extract urbanization detection using LiDAR datasets by using a 3D CNN technique based on convolution. We investigate the effects of various 3D CNN settings on urbanization extraction in more detail in this study. Through the use of deep learning and 3D convolutional, ReLU, and pooling operators, our suggested 3D CNN technique can automatically extract spectral and spatial data and textural and elevation features, resulting in improved urbanization extraction performance (particularly for the construction of roofs and roads). According to the findings of this study, the urbanization extraction is significantly influenced by the 3D CNN settings. The convolutional kernel size n and the pooling dimension p are combined to form $n[2p, 3p]$, the best combination. An image size m impacts the accuracy and calculation time of the algorithm. For parameter m, a value in the 20-40 range works best. The impervious surface extraction's performance is most steady at $L = 2$, which is also true for CNN. For the 3D CNN approach, the benefit of using several sources of data is less pronounced. The 3D CNN model can extract urbanization accurately from a single-source HR picture.

Data Availability

As you are concerned with the availability of my code and practical work, it is confidential to my lab and cannot be shared to anyone until it get published. And secondly, I will publish my code and lab work according to the instructions of my supervisor.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] G. Melotti and C. Premebida, "Multimodal deep-learning for object recognition combining camera and LIDAR data," in *2020 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, Ponta Delgada, Portugal, 2020.

[2] T. Shinohara, H. Xiu, and M. Matsuoka, "FWNet: semantic segmentation for full-waveform LiDAR data using deep learning," *Sensors*, vol. 20, no. 12, p. 3568, 2020.

[3] L. Zhang, Z. Shao, J. Liu, and Q. Cheng, "Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data," *Remote Sensing*, vol. 11, no. 12, p. 1459, 2019.

[4] D. Liu, F. Yang, H. Wei, and P. Hu, "Remote sensing image fusion method based on discrete wavelet and multiscale morphological transform in the IHS color space," *Journal of Applied Remote Sensing*, vol. 14, no. 1, article 016518, 2020.

[5] Y. Wei and B. Akinci, "A vision and learning-based indoor localization and semantic mapping framework for facility operations and management," *Automation in Construction*, vol. 107, pp. 102915–102949, 2019.

[6] J. Zhang, "Multi-source remote sensing data fusion: status and trends," *International Journal of Image and Data Fusion*, vol. 1, no. 1, pp. 5–24, 2010.

[7] R. Items, W. Rose, W. Rose, T. If, and W. Rose, *Sparse Time-Frequency Representation and Deep*, 2020.

[8] B. Major, D. Fontijne, A. Ansari et al., "Vehicle detection with automotive radar using deep learning on range-azimuth-doppler tensors," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 924–932, Seoul Korea, 2019.

[9] X. X. Zhu, D. Tuia, L. Mou et al., "Deep learning in remote sensing: a comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[10] X. Sun, *Deep Learning-Based Building Extraction Using Aerial Images And Digital*, UNIVERSITY OF TWENTE STUDENT THESES, 2021, https://library.itc.utwente.nl/papers_2021/msc/gfm/sun.pdf.

[11] R. Senchuri, A. Kuras, and I. Burud, "Machine learning methods for road edge detection on fused airborne hyperspectral and lidar data," in *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, Netherlands, 2021.

[12] V. V. Gomez, *Driving Using Deep Learning* UNIVERSITAT POLITÈCNICA DE CATALUNYA.

[13] Y. James Russell III, *An evaluation of DEM Generation Methods Using a Pixel-Based Landslide Detection Algorithm*, Virginia Polytechnic Institute and State University, 2021, http://hdl.handle.net/10919/104863.

[14] T. Fyleris, A. Kriščiūnas, V. Gružauskas, and D. Čalnerytė, "Deep learning application for urban change detection from aerial images," in *Gistam*, pp. 15–24, MDPI, 2021.

[15] S. Männistö, *Mapping and Classification of Urban Green Spaces With Object-Based Image Analysis and Lidar Data Fusion*, pp. 1–46, 2020, https: //helda. helsinki.fi/handle/10138/322585.

[16] D. Huang, B. Lee, H. Nassuna, and M. Prowse, *Machine Learning and Its Potential Applications in the Independent Evaluation Unit of the Green Climate Fund: A scoping study*, 2021 Independent Evaluation Unit Green Climate Fund, 2021.

[17] M. Ahangarha, R. Shah-Hosseini, and M. Saadatseresht, "Deep learning-based change detection method for environmental change monitoring using sentinel-2 datasets," *Environmental Sciences Proceedings*, vol. 5, no. 1, p. 15, 2020.

[18] A. Kuras, M. Brell, J. Rizzi, and I. Burud, "Hyperspectral and Lidar data applied to the urban land cover machine learning and neural-network-based classification: a review," *Remote Sensing*, vol. 13, no. 17, 2021.

[19] P. H. A. Cruz, *Mapping Urban Tree Species in a Tropical Environment Using Airborne Multispectral and LiDAR Data*, NIMS - MSc Dissertations Geospatial Technologies (Erasmus-Mundus), 2021, http://hdl.handle.net/10362/113904.

[20] P. Note, "Copyright warning & restrictions," *Public Health*, p. 125, 2007, http://scholar.http://google.com/scholar?hl=en&btnG=Search&q=intitle:Some+Contributions+on+MIMO+Radar#0.

[21] P. I. S. Wang and U. N. C. Charlotte, *NCDOT Wetland Modeling Program : Development of Tidal Wetland Models using QL2 LiDAR*, pp. 11–20, University of North Carolina at Charlotte. Dept. of Computer Science NCDOT RP, 2018.

[22] Y. Li, L. Ma, Z. Zhong et al., "Deep learning for LiDAR point clouds in autonomous driving: a review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3412–3432, 2021.

[23] M. U. Ghani, T. M. Alam, and F. H. Jaskani, "Comparison of classification models for early prediction of breast cancer," in *2019 International Conference on Innovative Computing (ICIC)*, Lahore, Pakistan, 2019.

[24] F. Jaskani, S. Manzoor, M. Amin, M. Asif, and M. Irfan, "An investigation on several operating systems for Internet of Things," *EAI Endorsed Transactions on Creative Technologies*, vol. 6, no. 18, article 160386,, 2019.

[25] W. Abbas, K. B. Khan, M. Aqeel, M. A. Azam, M. H. Ghouri, and F. H. Jaskani, "Lungs nodule cancer detection using statistical techniques," in *2020 IEEE 23rd International Multitopic Conference (INMIC).*, Bahawalpur, Pakistan, 2020.

[26] X. X. Zhu, *Deep learning in remote sensing: a review*, IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE, IN PRESS, 2017, 41501462.

[27] C A S E S T UDY, *Using Geospatial Data to Track Changes in Urbanization* World Bank https://olc.worldbank.org/system/files/Using%20Geospatial%20Data%20to%20Track%20Changes%20in%20Urbanization.pdf.

[28] R. Goldblatt, M. F. Stuhlmacher, B. Tellman et al., "Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover," *Remote Sensing of Environment*, vol. 205, pp. 253–275, 2018.

[29] M. U. Freudenberg, *Master Thesis Tree Detection in Remote Sensing Imagery Baumerkennung in Fernerkundungs-Bildmaterial prepared by*, University of Göttingen, 2019.

[30] L. Mughal, K. Bhatti, F. Khuhawar, and F. Jaskani, "Comparative analysis of face detection using linear binary techniques and neural network approaches," *EAI Endorsed Transactions on Creative Technologies*, vol. 5, no. 17, article 159712, 2018.