Land Zoning Using Satellite Imagery and Machine Learning

Surya Vardhan Karella, R Reddy Neeraj and M Aruna

November 25, 2024

Land Zoning using Satellite Imagery and Machine Learning

Karella Surya Vardhan
Computing Technologies,College of Engineering and Technologies.
SRM Institute of Science and Technology
SRM Nagar,Kattankulathur,Chennai 603203,TN,India
kk6916@srmist.edu.in

Reddy Neeraj.R
Computing Technologies,College of Engineering and Technologies.
SRM Institute of Science and Technology
SRM Nagar,Kattankulathur,Chennai 603203,TN,India
rr5276@srmist.edu.in

Dr. Aruna M
Associate Professor
Dept. of Computing Technologies
College of Engineering and Technology
SRM Nagar, Kattankulathur, Chennai 603203, TN, India
arunam@srimst.edu.in

***Abstract*— Appropriate land use planning is one of the new frontiers of sustainable development, particularly in the context of fast developing cities. Most land use classification systems, especially those that rely on visual interpretation and rudimentary image processing, continue to face difficulties accommodating the sophisticated, extensive datasets produced by contemporary remote sensors. This study investigates the new techniques of land zoning utilizing space images in conjunction with machine learning algorithms, in this case, neural networks, for the automatic classification of various types of land and fitting them for the purposes of agriculture, urbanization and industrialisation.**

In the project, a specific land cover type classification model has been created, trained, and tested with Sentinel-2 and location centered satellite data. By the use of high resolution images this approach allows classification of some vital land cover use categories i.e. barren land, vegetation, and water. In order to increase comprehensibility and guide further analysis, the classifications are superimposed on the respective images with every land type annotated and discussed. It shows how remote sensing data integrated with machine learning can solve land zoning issues efficiently, reliably and in a scalable manner helping to make economy and environment friendly land use decisions.

Keywords—Land zoning, satellite imagery, machine learning, neural networks , sustainable development, remote sensing, urban planning, environmental sustainability.

I. INTRODUCTION

One of the most important aspects that features in the urban planning process is Land zoning, this makes it very paramount for environmental management as well as the sustainable development of urban areas since different land areas are planned for specific uses e.g. Agricultural land, industrial land, residential land, conservation land, etc. How people utilize land has evolved with time, moving from manual maps and local historians to historical records that take too much time and effort as well as increasing chances of error or inconsistency with each new development. Rather, when zoning became increasingly needed and the need for effective land management policies became apparent, these primitive practices could not satisfy anyone's needs or objectives.

We depict high expectations on recent breakthroughs in high-resolution satellite earth observation images combined with cutting-edge machine learning technologies to revolutionize the process within the current marketplace. For many years now, satellite images have facilitated the acquisition of vast areas of land, providing relevant data on various aspects of the area, including vegetation, water areas, settlement infrastructure, and soil types, among hundreds of different attributes.

Using high resolution satellite images in conjunction with neural networks to automatically classify and zone lands is the main focus of this paper. Considering the accessibility of the open-source software and publicly available datasets such as Sentinel 2, this model applies remote sensing, decision rule and image interpretation in determining land cover including these targets: vegetation, barren land and water bodies. This methodology has the advantage of making less cumbersome the conventional practice of zoning and offers an environment friendly scalable solution to policy makers and land planners. Related work, the neural networks based approach, outcomes of the study as well as the scope and relevance of the idea for sustainable land use management are presented in the subsequent sections of the paper.

.

II. LITERATURE REVIEW

Land-use change and degradation on the Mongolian Plateau from 1975 to 2015 have been analyzed by Batunacun and colleagues [1], highlighting significant environmental impacts in the Xilingol region. Their findings underscore the dynamic interactions between human activities and land degradation, offering valuable insights into sustainable land management strategies.

Belgiu and Drăguț [2] conducted a review of random forest applications in remote sensing, demonstrating its versatility and effectiveness in various land cover classification tasks. They emphasized the algorithm's ability to handle high-dimensional data and its relevance in environmental monitoring.

Breiman [3] explore the foundational concepts of Classification and Regression Trees (CART), providing a comprehensive framework for decision tree algorithms that have become pivotal in machine learning. This work lays the groundwork for understanding model complexities and is essential for researchers employing tree-based methods in various applications.

Breiman [4] introduces the concept of bagging predictors, a technique aimed at enhancing the stability and accuracy of machine learning algorithms. This method significantly mitigates overfitting and improves predictive performance, making it a vital reference for those working with ensemble methods in data classification.

Chen and colleagues [5] enhanced land cover mapping by integrating pixel-based and object-based classifications from remotely sensed imagery. Their approach reveals the benefits of combining methodologies to improve classification outcomes, indicating that such integration can lead to more accurate and reliable land cover maps.

Ojwang and team [6] presented an integrated hierarchical classification approach for mapping land use in complex social-ecological systems. Their research highlights the need for multi-layered analytical frameworks that account for diverse ecological factors and socio-economic influences, showcasing the potential of machine learning in understanding intricate land use dynamics.

Bayas and team [9] investigated land use classification using Sentinel-2 imagery, applying various machine learning algorithms. Their study demonstrates the effectiveness of different approaches in achieving high accuracy in land cover mapping.

Gómez and colleagues [10] reviewed optical remotely sensed time series data, emphasizing its significance in tracking land cover dynamics over time. Their study provides insights into the methodologies and technologies that facilitate effective land monitoring.

Weng [11] explored the use of transfer learning with deep convolutional neural networks for land-use classification, showcasing its potential to improve accuracy in scenarios with limited training data. The study illustrates how leveraging pre-trained models can enhance classification performance significantly.

Feng’s [12] research on urban zoning employed higher-order Markov random fields, revealing the complexity of urban land use. The study integrates multi-view imagery data, emphasizing the advantages of statistical models in understanding urban dynamics.

Kerins and colleagues [13] analysed urban land use in India and Mexico, emphasizing the interplay of socio-economic factors with remote sensing techniques. Their research provides a comprehensive overview of urban expansion and its implications for regional planning.

Mitra and Basu [14] conducted a survey on machine learning and deep learning techniques for land cover classification. They highlighted the evolution of methodologies and the challenges faced in achieving accurate land cover maps, advocating for the continued development of robust classification algorithms.

Tung and team [15] examined urban expansion in Hanoi using machine learning and multi-temporal satellite imagery. Their findings shed light on urban growth patterns and the effectiveness of remote sensing in monitoring changes over time.

Ouchra and colleagues [16] examined the application of machine learning algorithms for land cover classification using Google Earth Engine, emphasizing the platform's efficiency in processing large datasets. Their study highlights how cloud computing technologies can facilitate advanced analysis and improve the scalability of remote sensing applications.

Kussul and collaborators [18] explored deep learning techniques for crop type classification, revealing the high accuracy potential of convolutional neural networks in agricultural assessments. Their research highlights the transformative impact of deep learning on remote sensing applications.

III. PROPOSED METHODOLOGY

This paper uses Neural Network (NN) methods to discriminate between barren land, water, and vegetation land cover in a satellite image. The authors use sentinel-Satellite images and prepare three folders labeled barrens, water and vegetation. Image preprocessing involves normalizing and resizing images to 256 by 256 pixels so as to ensure uniformity in the size of the input dimension. The model neural network is built in sequential order and starts with the Flatten layer that restructures image size of 256 by 256 by 3 into one large array. A Dense layer with 128 units and ReLU activation follows to enable feature capturing. The last Dense layer employs softmax and its activation to saturate the number of land cover classes, enabling multi-class classification.

Two callbacks are included in order to make the training more efficient. When the validation loss has flatlined for three epochs, the ReduceLROnPlateau callback reduces the learning rate by a factor of 0.5, while the EarlyStopping callback stops the training and resets to the best model weights if the validation loss has not improved over the three most recent epochs. With the Adam optimizer and categorical cross-entropy loss, the accuracy is set as the primary metric within the model. For increasing generalization, random rotations, random shifts, and random flips are performed during training. Each training session lasts for 10 epochs with an 80-20 training-validation splitting.

After training, the model is assessed using the test set and test loss and test accuracy as performance metrics. To draw attention to the barren land, vegetation, and water areas, overlays classifying the respective parts of the original images are placed for every type of land cover. Such ways of presenting the results give useful information and make the interpretation of the model more convenient for its use in urban planning and resource management.

IV. Data Collection

The data collection phase of this study involves high-resolution satellite imagery, sourced from publicly accessible platforms such as Sentinel Hub, Earth Explorer and Kaggle's web maps were used for satellite image data collection in this study.

Satellite images available on these platforms include spectral bands that help distinguish between three major types of land coverage, which are barren land, vegetation, and water bodies. In this case, the images depict areas quite different in land cover and therefore can be employed for computing through patch-based analysis, which is one of the most helpful methods of identifying various regions. Such an index and a range of other spectral data incorporated into modeling the neural network are also expected to assist in its training and validation. By using the images of the target area from the Kaggle website, we present a broad spectrum of the samples in the study making it easier for the classification model to fit a wide array of the different terrain and land coverage characteristic of the world's regions.

Fig 1: Sample Satellite Image from Sentinel-2 Dataset (Water Region)

V. Data Preprocessing

The Data preprocessing is the most important part when developing a Neural Network for image classification. First, the dataset was loaded, which contained the images organized in specific directories. Every image was resized to a standard dimension of 256x256 pixels. This is an important step because the model would have consistent input and therefore the Neural Networks could treat images uniformly across batches.

After the resizing, the normalization process is followed by scaling down pixel values lying between 0 and 255 by division with 255.0. This step plays a very important role in adjusting the pixel values in the range of [0, 1], which also assists in enhancing the rate of convergence of the optimization algorithms employed during the training process. The reshaping of images gives it a new dimension and, therefore, changes the data format to match the expected input shape of the model. This reshape is critical since it enables the model to handle batches of images in an effective manner both during the training as well as the evaluation phases.

These preprocessing techniques-loading, resizing, normalization, and reshaping preprocess the dataset appropriately for further processes in training a model. Thus, a perfect feature extraction and learning of the Neural Networks from the input data will be assured, hence leading to better accuracy of classification and generalization capacity of the model.

VI. Model Development

For classification of the land cover types in the current study, we adopted the use of a neural network model. The first stage was data preprocessing which was classified into subdirectories for various land cover categories. We set the path to the dataset, and obtained the class labels from the names of the folders. Next, using more formal manners, we obviously gathered paths to images as well as their labels so that the dataset would be as structurally correct as possible for further processing. After verifying the total number of images collected, we split the dataset into three subsets: The proposed partition of sets is as follows: training—70%, validation—10%, testing—20% In this work, classes are divided into sets strategically.

For this purpose, we utilized the ImageDataGenerator class from Keras, which rescales pixel values and creates sets of images in batches directly at the directory level. This made it easier in terms of preprocessing and batch creation of every subset of the data we had starting with the training, our validation subset and the final test subset. The Q-learning automata neural network architecture was designed in Keras with a set of input layers along with several dense layers.

These structures allowed features elucidated from the input images to be learned by the model. The output layer used the softmax activation function of estimating the probabilities of classes of different land covers. Following this model's definition, we used the Adam optimizer and categorical cross-entropy as the loss function. To improve the training's stability and prevent overfitting, we regularized callback features as EarlyStopping and ReduceLROnPlateau. The model was trained for 10 iterations with the validation data used in performance check during the training process.

```
In [8]: cnn_model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 196608)	0
dense (Dense)	(None, 128)	25,165,952
dense_1 (Dense)	(None, 3)	387

Total params: 75,499,019 (288.01 MB)

Trainable params: 25,166,339 (96.00 MB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 50,332,680 (192.00 MB)

Fig 2: Model Summary

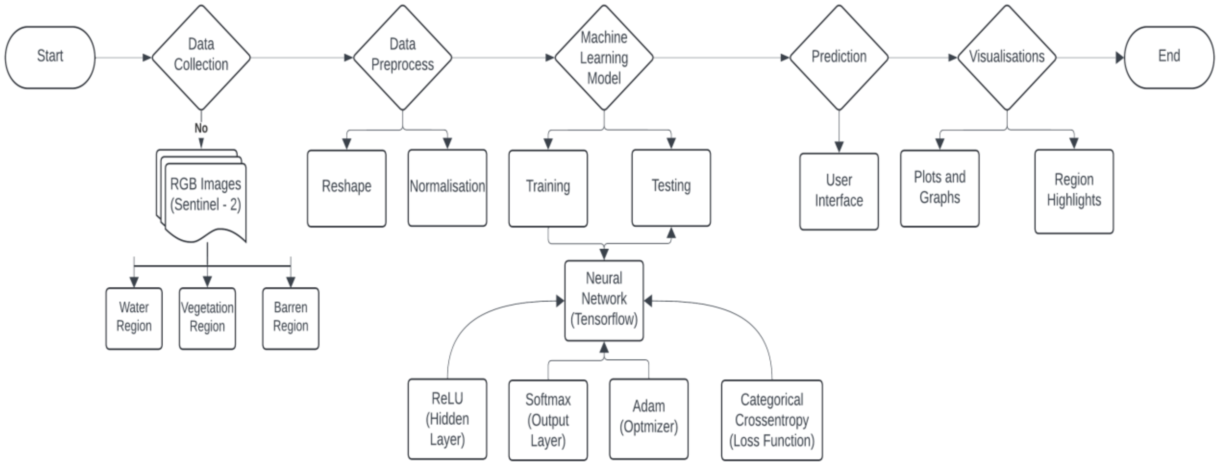


Fig 3: Model Architecture

VII. RESULTS AND VISUALIZATION

The objective of this study was to design a neural network model that successfully gives good classification results of Barren, Vegetation, and Water categories from satellite images. The model also thus went through stringent training and validation and the accuracy of the model was checked using the validation dataset and it was found to be 97%. This performance shows the effectiveness of neural networks for the monitoring and forecasting of the environment in order to make use of and plan for land correctly.

To confirm the assessment of the model, area coverage percentages were computed using the patches classified from the test image “Water and Vegetation.jpg”. The analysis yielded the following area coverage percentages, for Barren land only 2.6 %; for Vegetation only 46.75%; and for water only 50.65%. These results suggest that Vegetation and Water forms a significant part of the observed region and hence a broad understanding of the land cover base. The forecasted results were in fact depicted using color overlays on top of the raw pictures in the form of masks.

The highlighted predictions are shown in Figure 8 which represents the red areas as Barren land, yellow as Vegetation and blue as Water cover. From this visualization one can easily harvest that the model is capable of providing the right classification of the given land cover. However, there is some variability in the results associated with the patch-based prediction approach. The restricted analysis was determined only on the identified land cover classes and without making any consideration to the potential modifying factors in the imagery.

Visual representations of the predictions were created by overlaying color-coded masks on the original images. Figure 8 illustrates these highlighted predictions, with red regions signifying Barren land, yellow regions denoting Vegetation, and blue areas indicating Water coverage. This visualization clearly demonstrates the model's capability to accurately identify and differentiate between various land cover types. The application of color-coding enhances the interpretability of the results, allowing stakeholders to quickly assess land cover conditions.

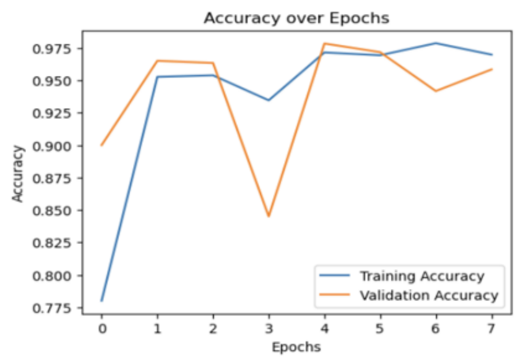


Fig 4: Accuracy over Epochs

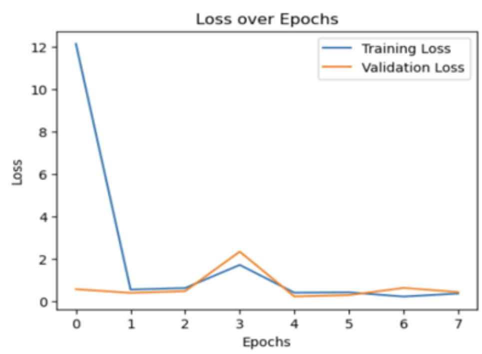


Fig 5: Loss over Epochs

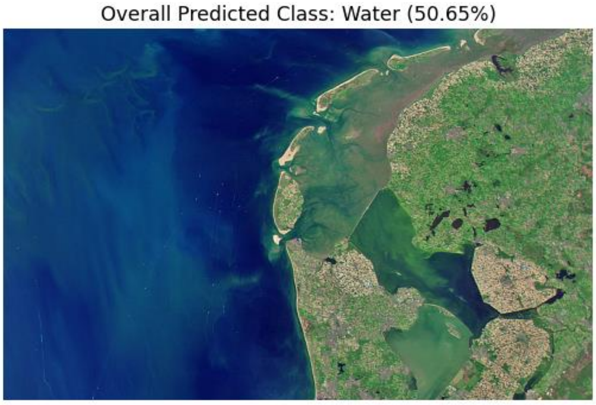
```
Epoch 1/10
C:\Users\surya\anaconda3\lib\site-packages\keras\src\trainers\data_adapters\py_dataset_adapter.py:121: UserWarning: Your `PyDat
aset` class should call `super().__init__(**kwargs)` in its constructor. `**kwargs` can include `workers`, `use_multiprocessing
`, `max_queue_size`. Do not pass these arguments to `fit()`, as they will be ignored.
  self._warn_if_super_not_called()
132/132 ━━━━━━━━━━━━━━━━━━━━ 48s 347ms/step - accuracy: 0.6019 - loss: 33.5255 - val_accuracy: 0.9000 - val_loss: 0.5792 - lear
ning_rate: 0.0010
Epoch 2/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 47s 338ms/step - accuracy: 0.9614 - loss: 0.4372 - val_accuracy: 0.9650 - val_loss: 0.4055 - learn
ing_rate: 0.0010
Epoch 3/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 80s 332ms/step - accuracy: 0.9252 - loss: 1.1474 - val_accuracy: 0.9633 - val_loss: 0.4871 - learn
ing_rate: 0.0010
Epoch 4/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 45s 331ms/step - accuracy: 0.9541 - loss: 0.7950 - val_accuracy: 0.8450 - val_loss: 2.3487 - learn
ing_rate: 0.0010
Epoch 5/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 46s 331ms/step - accuracy: 0.9621 - loss: 0.6466 - val_accuracy: 0.9783 - val_loss: 0.2371 - learn
ing_rate: 0.0010
Epoch 6/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 45s 333ms/step - accuracy: 0.9803 - loss: 0.2275 - val_accuracy: 0.9717 - val_loss: 0.3013 - learn
ing_rate: 0.0010
Epoch 7/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 49s 357ms/step - accuracy: 0.9781 - loss: 0.2134 - val_accuracy: 0.9417 - val_loss: 0.6391 - learn
ing_rate: 0.0010
Epoch 8/10
132/132 ━━━━━━━━━━━━━━━━━━━━ 0s 312ms/step - accuracy: 0.9660 - loss: 0.4317
Epoch 8: ReduceLROnPlateau reducing learning rate to 0.0005000000237487257.
132/132 ━━━━━━━━━━━━━━━━━━━━ 49s 353ms/step - accuracy: 0.9661 - loss: 0.4313 - val_accuracy: 0.9583 - val_loss: 0.4447 - learn
ing_rate: 0.0010
38/38 ━━━━━━━━━━━━━━━━━━━━ 12s 308ms/step - accuracy: 0.9809 - loss: 0.3946
Test accuracy: 97.92%
```

Fig 6: Epochs Showcase

VIII. CONCLUSION

Neural network models for classification of land cover data from satellite images can be advocated as a promising approach to the efficient management of resources on the territory. This research further highlights that high client satellite imagery enables precise demarcation between Barren, Vegetation, and Water classes of change in land cover. This work has the advantage of learning from multiple images and the model ability to refine its predictions from the new data base that may be advanced in the future.

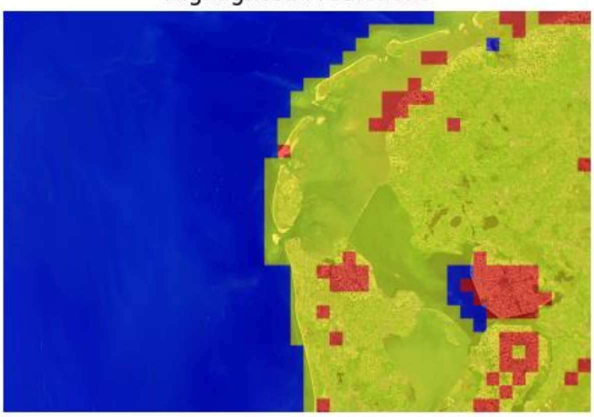
The normalized area coverage percentages of the various land cover types help the stakeholders to make the right decisions while planning the use of the land as well as conversing the natural resources. The combination of scoring metrics for suitability assessments also provides the tool's end-users with a framework for examining strategies relating to land management to ensure sustainable usage in relation to development objectives.




```
Area coverage percentages by class (based on patches):
Barren: 2.60%
Vegetation: 46.75%
Water: 50.65%
```

Fig 7: Metrics Coverage

using images to highlight the extracted classifications makes it easy for users to understand sophisticated results and enable the transition from ideas to solutions. This work demonstrates how neural networks and remote sensing techniques can improve the practices of land management and support the improvement of other activities aimed at the development of sustainable processes.



a.

Fig 8: Highlighted Predictions

References

[1] B. Batunacun, C. Nendel, Y. Hu, and T. Lakes, "Land-use change and land degradation on the Mongolian Plateau from 1975 to 2015—A case study from Xilingol, China," Land Degrad. Dev., 2019. https://doi.org/10.1002/ldr.2948

[2] M. Belgiu and L. Drăguț, "Random Forest in remote sensing: A review of applications and future directions," ISPRS J. Photogramm. Remote Sens., vol. 114, pp. 24–31, 2016. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[3] L. Breiman, J. Friedman, R.A. Olshen, and C.J. Stone, Classification and Regression Trees, Routledge, 1984. Available at: https://doi.org/10.1201/9781315139470

[4] L. Breiman, "Bagging predictors," Mach. Learn., vol. 24, pp. 123–140, 1996. https://doi.org/10.1007/BF00058655

[5] Y. Chen, Y. Zhou, Y. Ge, R. An, and Y. Chen, "Enhancing land cover mapping through integration of pixel-based and object-based classifications from remotely sensed imagery," Remote Sens., vol. 10, no. 1, 2018. https://doi.org/10.3390/rs10010077

[6] G.O. Ojwang, J.O. Ogutu, M.Y. Said, M.A. Ojwala, S.C. Kifugo, F. Verones, B.J. Graae, R. Buitenwerf, and H. Olff, "An integrated hierarchical classification and machine learning approach for mapping land use and land cover in complex social-ecological systems," Front. Remote Sens., 2023. Available at: https://doi.org/10.3389/frsen.2023.1188635

[7] R. Mahmoud, M. Hassanin, H. Al Feel, and R.M. Badry, "Machine learning-based land use and land cover mapping using multi-spectral satellite imagery: A case study in Egypt," Sustainability, 2023. https://doi.org/10.3390/su15129467

[8] B.S. Jasim, O.Z. Jasim, and A.N. Al-Hameedawi, "Evaluating land use land cover classification based on machine learning algorithms," Eng. Technol. J., 2024. https://doi.org/10.30684/etj.2024.144585.1638

[9] S. Bayas, S. Sawant, I. Dhondge, P. Kankal, and A. Joshi, "Land use land cover classification using different ML algorithms on Sentinel-2 imagery," Int. J. Res. Appl. Sci. Eng. Technol., 2023. https://doi.org/10.22214/ijraset.2023.57427

[10] A. Nigar, Y. Li, M.Y.J. Baloch, A.F. Alrefaei, and M.H. Almutairi, "Comparison of machine and deep learning algorithms using Google Earth Engine and Python for land classifications," Int. J. Geo-Inf., 2023. https://doi.org/10.1080/23729333.2023.2166252

[11] C.-Y. Weng, "Land-use classification via transfer learning with a deep convolutional neural network," Front. Environ. Sci., 2024. https://doi.org/10.3389/fenvs.2024.137844

[12] T. Feng, Q.-T. Truong, D.T. Nguyen, J.Y. Koh, L.-F. Yu, A. Binder, and S.-K. Yeung, "Urban zoning using higher-order Markov random fields on multi-view imagery data," in Proc. 2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2023. https://ieeexplore.ieee.org/document/10177754

[13] P. Kerins, B. Guzder-Williams, E. Mackres, T. Rashid, and E. Pietraszkiewicz, "Mapping urban land use in India and Mexico using remote sensing and machine learning," World Resour. Inst., 2023. https://www.wri.org/research/mapping-urban-land-use-india-and-mexico-using-remote-sensing-and-machine-learning

[14] S. Mitra and S. Basu, "Remote sensing-based land cover classification using machine learning and deep learning: A comprehensive survey," Front. Remote Sens., vol. 14, no. 2, 2023. Available at: https://doi.org/10.47164/ijngc.v14i2.1137

[15] D.T. Tung, N.T. Tung, H.T. Thuy, T.M. Ngoc, D.T.T. Huyen, and P.C. Linh, "Analyzing urban expansion in Hanoi using machine learning and multi-temporal satellite imagery," Remote Sens., 2023. Available at: https://link.springer.com/chapter/10.1007/978-3-030-01237-3_38

[16] H. Ouchra, A. Belangour, and A. Erraissi, "Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study," IEEE Access, 2023. Available at: https://ieeexplore.ieee.org/document/7891032

[17] S. Bayas, S. Sawant, I. Dhondge, P. Kankal, and A. Joshi, "Land use land cover classification using different ML algorithms on Sentinel-2 imagery," Arab. J. Geosci., 2019. Available at: https://doi.org/10.1007/s42452-019-1527-8

[18] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," arXiv, 2015. https://doi.org/10.48550/arXiv.1508.00092

[19] M. Castelluccio, et al., "CNNs for land use classification in remote sensing," ISPRS J. Photogramm. Remote Sens., 2016. https://doi.org/10.1016/j.isprsjprs.2016.01.011

[20] S. Hussain, et al., "Comparison of ML models for land use classification," Remote Sens. Appl., 2019. https://doi.org/10.1016/j.rsase.2019.100100

[21] P. Sharma and R.K. Gupta, "Land-use mapping with Google Earth Engine," Remote Sens. Appl., 2022. https://doi.org/10.1016/j.rsase.2022.100301

[22] X. Zhang, et al., "Urban expansion in mega-cities using Landsat and ML," Remote Sens., 2021. https://doi.org/10.3390/rs13020294

[23] R. Gonzalez, et al., "Assessment of urban heat island effect using ML and satellite data," Int. J. Remote Sens., 2022. https://doi.org/10.1080/01431161.2022.2021123

[24] M. Taufik, et al., "Land cover classification of deforested land with deep learning," Comput. Geosci., 2020. https://doi.org/10.1016/j.cageo.2019.104429

[25] Z. Chen, et al., "Urban vegetation and land cover mapping with Sentinel-2 and ML," Ecol. Ind., 2021. https://doi.org/10.1016/j.ecolind.2020.106245

[26] Satellite Images of Regions- Sentinel-2(Kaggle)
https://www.kaggle.com/datasets/requiemonk/sentinel12-image-pairs-segregated-by-terrain

[27] Satellite Images of Water Bodies – Sentinel-2 (Kaggle)
https://www.kaggle.com/datasets/franciscoescobar/satellite-images-of-water-bodies