University of Tehran

Earth Observation and Geomatics Engineering

Online ISSN: 2588-4360

Homepage: https://eoge.ut.ac.ir/

Deep Learning-Based Prediction of Land Use and Land Cover Dynamics in the Urmia Lake Basin

(Selected Paper in the 8th ISPRS Geospatial Conference 2025, University of Tehran, Iran)

Amirarsalan Ghafoori[1], **Parham Pahlavani**[2, ✉], **Behnaz Bigdeli**[3], **Omid Ghorbanzadeh**[4]

1. School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: arsalanghafoori@ut.ac.ir
2. Corresponding author, School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: pahlavani@ut.ac.ir
3. School of Civil Engineering, Shahrood University of Technology, Shahrood, Iran. Email: bigdeli@shahroodut.ac.ir
4. University of Natural Resources and Life Sciences, Vienna. Email: omid.ghorbanzadeh@boku.ac.at

Article Info

Article type:
Research Article

Article history:
Received 2026-01-18
Received in revised form 2026-02-10
Accepted 2026-02-14
Available online 2026-06-02

Keywords:
Land cover and land use,
Deep learning,
Classification,
Convolutional neural networks,
Recurrent neural networks

ABSTRACT

Monitoring and analyzing land use and land cover (LULC) is crucial for understanding environmental transformation, urban planning, deforestation, and water resource management. The Urmia Lake basin-a critical water body in northwestern Iran- has vital LULC changes over the past two decades.

The study aims to predict future LULC in the Urmia Lake basin and generate LULC maps using deep learning models applied to satellite imagery. by comparing the performance of diverse deep learning models, the study seeks to identify most accurate approach for modeling spatio-temporal LULC changes. To achieve this, four models are implemented: a Multilayer Perceptron (MLP) which captures complex nonlinear relationships in data, two-dimensional and three-dimensional Convolutional Neural Network (2D CNN and 3D CNN) for effective extraction of spatial features from satellite imagery, and a Recurrent Neural Network (RNN) to model temporal changes over time. The input data consist of 64 features derived from six optical Moderate Resolution Imaging Spectroradiometer (MODIS) bands, Land Surface Temperature (LST), vegetation and surface indices (Normalized Difference Vegetation Index (NDVI), Normalized Difference Snow Index (NDSI), and Normalized Difference Water Index (NDWI)), topographical layers (Digital Elevation Model (DEM), slope, and aspect), 50 ERA5-Land monthly climate variables, and precipitation data from the PERSIANN Climate Data Record (PERSIANN-CDR). The dataset spans 22 years and includes eight land cover classes as output labels. Among the evaluated models, the 3D CNN reached the highest performance with a test accuracy of 92.82%, mean Intersection over Union (IoU) of 0.6351, and the lowest number of mismatched pixels. Test accuracy of 3D CNN indicates that 92.82% of the pixels in the independent test dataset for the year 2022 were correctly classified into their respective land use/land cover categories. These results confirm its superior ability to capture complex spatio-temporal patterns for accurate LULC prediction in the Urmia Lake basin.

Cite this article: Ghafoori, A., Pahlavani, P., Bigdeli, B. & Ghorbanzadeh, O. (2025). Deep Learning-Based Prediction of Land Use and Land Cover Dynamics in the Urmia Lake Basin, *Earth Observation and Geomatics Engineering*, Volume 9, Issue 2, Pages 82-92. http//doi.org/ 10.22059/eoge.2026.409793.1205

Publisher: University of Tehran.

DOI: http//doi.org/ 10.22059/eoge.2026.409793.1205

1. Introduction

Land use and land cover change is influenced by both human activities and natural environmental changes Hasan et al. (2020). LULC changes have become increasingly significant at regional and global scales, driven by urban expansion, agricultural intensification, deforestation, and climate variability. Monitoring and predicting these changes using multitemporal, multisource satellite data is essential for sustainable natural resource management, ecosystem preservation, and informed policy development Mashala et al. (2023); Winkler et al. (2021). These changes directly influence hydrological cycles, soil health, and habitat integrity, and have profound implications for food security and climate resilience.

Climate and land use changes are among the most critical factors influencing the environment and agricultural systems of a region CALANCA (2007). Traditional remote sensing and GIS approaches offer valuable insights into LULC dynamics, but often face challenges in accurately modeling complex spatial–temporal patterns Cheng et al. (2024).Deep learning models such as Multi-Layer Perceptron, Convolutional Neural Networks, and Recurrent Neural Networks have demonstrated strong performance in extracting spatial and temporal features from satellite imagery for land cover mapping and change detection Sefrin et al. (2020).

MLP is a type of feedforward neural network that learns complex nonlinear relationships between input and output variables, making it effective for LULC change prediction when spatial and thematic complexity is involved. The results of the MLP network in Ahmed and Ahmed (2012) indicate that it achieved the most accurate modeling performance for the Dhaka region in Bangladesh. One of the advantages of this method is its ability to simultaneously model all types of land use changes. CNNs are among the most widely used architectures in remote sensing image processing. A typical CNN consists of an input layer, multiple convolutional and pooling layers, and a final output layer responsible for image classification. CNNs are particularly effective in learning from the spatial context of the data, enabling accurate extraction of spatial features Minar and Naher (2018). RNNs are well-suited for modeling sequential data, making them particularly effective for capturing dependencies in both spatial and temporal sequences. Their architecture allows information to persist across time steps, which is essential for analyzing dynamic patterns in time-series remote sensing data Liu et al. (2018).

Monitoring and predicting LULC dynamics in the Lake Urmia Basin are particularly urgent, given the profound environmental changes observed in recent decades. Lake Urmia, once the largest hypersaline lake in the Middle East, has experienced dramatic shrinkage—over 80% reduction in surface area—primarily due to unsustainable agricultural water use, dam construction on inflow rivers, and prolonged drought conditions AghaKouchak et al. (2015); Feizizadeh et al. (2022). This has led to substantial vegetation loss, increased soil salinity, and rising risks of desertification, with cascading effects on biodiversity, local climate, and agricultural livelihoods Alizade Govarchin Ghale et al. (2017). As such, accurate LULC change modeling is essential not only for ecological restoration but also for informed land and water resource management policies across the basin.

In this study, we employ a deep learning structure—comprising MLP, CNN, and RNN models—to monitor and predict LULC changes in the Lake Urmia Basin using a 22-year time series of satellite data. The goal is to evaluate the capacity of these models to capture spatial and temporal patterns of land transformation and provide insights that support sustainable regional planning and environmental recovery.

The primary problem addressed in this study is the accurate modeling and prediction of long-term land use and land cover dynamics in the Urmia Lake basin using multi-source and multi-temporal geospatial data. The major contributions of this research are threefold: (1) the development and comparison of multiple deep learning architectures (MLP, 2D CNN, 3D CNN, and RNN) within a unified pixel-based framework for long-term LULC prediction; (2) the integration of optical, climatic, topographic, and land surface variables over a 22-year period to capture complex spatio-temporal patterns; and (3) a comprehensive evaluation of model performance using both quantitative metrics and spatially explicit map comparisons. These contributions provide new insights into the suitability of different deep learning paradigms for LULC change modeling in environmentally sensitive basins.

The remainder of this paper is organized as follows. Section 2 describes the study area, datasets, and preprocessing steps, including feature construction and sampling strategy. Section 3 presents the deep learning architectures, training configuration, and evaluation metrics. Section 4 reports the experimental results and comparative analysis of the models. Section 5 discusses the findings and their implications for LULC monitoring.

2. Study area and Dataset

2.1. Study Area

The study area surrounds the basin of Lake Urmia, located in the northwest of Iran. This region includes parts of the key cities such as Urmia, Tabriz, Ardebil, Erbil, and Zanjan. The total area spans approximately 102,690 km², selected for its ecological importance and diverse land cover dynamics, making it suitable for evaluating and comparing deep learning algorithms for LULC prediction.

The Urmia Lake basin was defined as the hydrological catchment area contributing surface runoff to Lake Urmia. The basin boundary was delineated based on existing hydrological and administrative datasets and includes the entire lake surface as well as surrounding upstream regions influencing land use and land cover dynamics. This definition ensures that all land cover changes with potential

impact on the lake's hydrological and ecological conditions are captured.

Geographically, the region lies between 44°10′41″E, 35°40′55″N (southwest) and 47°56′13″E, 38°32′12″N (northeast), as shown by the red rectangle in Figure 1. It includes the entirety of Lake Urmia, one of the largest saltwater lakes in the Middle East, and is mostly covered by grasslands, with additional areas of cropland, and other land cover types.

Figure 1. study area (Urmia Lake Basin)

In recent decades, the dramatic shrinkage of Lake Urmia and the reduction in vegetation cover have severely affected the region's ecosystem, biodiversity, and agricultural productivity. These changes emphasis the critical need for robust LULC monitoring and future prediction.

2.2. Data

The dataset comprises 22 GeoTIFF files, each representing one year from 2001 to 2022. All data were collected in June of each respective year to maintain seasonal consistency. Each annual file contains 65 bands, including: six optical bands from MODIS; three topographical layers (Digital Elevation Model, slope, and aspect); three spectral indices—Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Normalized Difference Snow Index (NDSI); 50 bands from ERA5-Land monthly averaged climate reanalysis (ECMWF); precipitation data from NOAA's Climate Data Record (CDR) of PERSIANN-CDR; and land cover information from the MCD12Q1.066 MODIS Land Cover Type Yearly Global 500m product.

The selection of the 64 input features was guided by their relevance to land use and land cover dynamics and their proven effectiveness in previous remote sensing and environmental modeling studies. Optical MODIS bands and land surface temperature provide information on surface reflectance and thermal characteristics, while vegetation and surface indices capture vegetation cover, water presence, and surface conditions. Topographic variables influence land use suitability and hydrological processes. Climate variables from ERA5-Land and precipitation data from PERSIANN-CDR represent long-term atmospheric forcing that drives land cover transitions. Together, these features comprehensively characterize the spectral, topographic, and climatic factors controlling LULC changes in the Urmia Lake basin.

MODIS optical bands, land surface temperature, and the MCD12Q1 land cover product have an original spatial resolution of 500 m. ERA5-Land climate variables are originally provided at approximately 9 km spatial resolution. Topographic data derived from the SRTM Digital Elevation Model have an original resolution of 30 m. To ensure spatial consistency, all layers were resampled to match the 500-meter resolution of the land cover product. Per-cell means were used for topographical data, while kriging interpolation was applied to the climate variables. Each raster file covers a region of 630 × 652 pixels.

To ensure spatial completeness and consistency among input variables, spatial interpolation was applied to the climate-related datasets. Specifically, ordinary kriging interpolation was employed to resample the ERA5-Land monthly climate variables to the target spatial resolution. Kriging was selected due to its ability to model spatial autocorrelation and provide statistically optimal estimates. the combination of mean-based temporal interpolation and kriging-based spatial interpolation ensures data completeness while maintaining the physical and statistical characteristics of the environmental variables.

In the base year (2001), the region included eight land cover classes: Class 1 (Grasslands): dominated by herbaceous annuals (<2m); Class 2 (Croplands): areas used for agriculture; Class 3 (Water Bodies): permanent water coverage of at least 60%; Class 4 (Barren Lands): areas with sand, rock, or soil and less than 10% vegetation; Class 5 (Open Shrublands): woody perennials (1–2m height) with 10–60% cover; Class 6 (Urban and Built-up Lands): at least 30% impervious surface; Class 7 (Savannas): tree cover of 10–30% with canopy height >2m; Class 8 (Permanent Wetlands): permanently inundated lands with 30–60% water cover and >10% vegetation cover.

Table 1. LULC classes in 2001

Class No	Class Name	Pixels	Percentage (%)
1	Grasslands	300507	73.15
2	Croplands	66631	16.22
3	Water Bodies	17731	4.31
4	Barren lands	16725	4.07
5	Open shrublands	6150	1.49

6	Urban and Built-up lands	2958	0.73
7	Savannas	51	0.01
8	Permanent Wetlands	7	0.01
-	Total	410760	100.00

2.3. Data Preprocessing

The quality of the employed satellite and auxiliary datasets was assessed through a combination of preprocessing steps, including outlier detection, missing-value handling, and normalization. Extreme and physically implausible values were identified using broad thresholding and a temporal 3-sigma test and subsequently corrected. Temporal consistency was preserved through interpolation across time series, while spatial consistency was ensured using neighborhood-based filling. These procedures helped mitigate noise, sensor artifacts, and data gaps, ensuring reliable input quality for deep learning model training.

2.3.1. Outlier Detection and Removal

For input bands, due to diverse value ranges across the bands a broad threshold of $\pm 10^{10}$ was applied for outlier detection. Any pixel value outside this range replaced with a placeholder value that does not occur in any band. After this step, a 3-sigma test was conducted across the time series for each spatial location. Pixel values beyond 3-sigma test were considered as outliers and replaced with the placeholder. for the output (LULC) layer, any pixel with a class value outside the valid range of 1-17 was considered as an outlier and set to 0. This step helped eliminate corrupted or missing data values.

2.3.2. Filling Missing Values

For input bands, outliers beyond the range of $\pm 10^{10}$ were replaced with the local mean value from a 3×3 neighborhood. Padding was applied to handle edge pixels. For outliers detected by the 3-sigma test, temporal interpolation was performed: each pixel's values were replaced by the mean across its temporal observation over the 22 years. This approach preserved both spatial and temporal continuity.

For the output band, each outlier pixel was replaced with the most frequent valid value in its 3×3 neighborhood, preserving local spatial patterns.

2.3.3. Normalization

Finally, both input feature bands and output labels were normalized to the [0,1] range using min-max scaling. This normalization was applied independently to each band due to their differences in value ranges, and to ensure greater stability and faster convergence during deep learning model training, as shown in Equation (1).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}, Y_{norm} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \quad (1)$$

where X_{norm}, Y_{norm} *= normalized input and output values*

X, Y = original input and output values

X_{min}, Y_{min} *= minimum values in each band*

X_{max}, Y_{max} *= maximum values in each band*

3. Model Architecture and Training

The selection of deep learning models in this study was motivated by the complex, nonlinear, and spatio-temporal nature of land use and land cover changes in the Urmia Lake basin. MLP models are effective for capturing nonlinear relationships among high-dimensional environmental variables in pixel-based classification tasks. CNN architectures are well suited for learning spatial and spatio-temporal patterns from satellite imagery, enabling the identification of heterogeneous land cover structures. RNN-based models are designed to model temporal dependencies and long-term trends, making them suitable for analyzing multi-year satellite time series. Evaluating these complementary deep learning approaches allows a comprehensive assessment of their relative strengths for LULC change prediction in a dynamic and environmentally sensitive region.

3.1. Multi-Layer Perceptron

In recent years, the application of MLP—has gained increasing attention for LULC classification due to their ability to model complex, nonlinear relationships in high-dimensional remote sensing data—particularly when spatial context is less critical. MLPs have been successfully used in several studies for this purpose. For instance, Shafizadeh-Moghadam et al. (2015) applied MLPs to LULC mapping in Tehran using Landsat imagery and reported reliable classification accuracy compared to traditional classifiers. Jamali et al. (2024) demonstrated that their Spatial Gated MLP (SGU-MLP) surpassed several CNN and Vision Transformer models in mapping urban and agricultural land in both Houston (USA) and Berlin (Germany), achieving over 15% higher accuracy in some cases. The SGU-MLP is a purely MLP-based architecture that captures spatial dependencies through a Spatial Gating Unit (SGU), which splits feature representations into two branches and applies spatially varying gating operations to model long-range spatial interactions without convolution or self-attention mechanisms. In addition, Kumar and Agrawal (2023) reported that integrating MLPs with Markov-chain transitions achieved over 75% prediction accuracy in forecasting future LULC dynamics in India, demonstrating their merit for change prediction.

In this study, we applied an MLP-based deep learning model to classify LULC classes in the Lake Urmia basin using a time-series of multi-band satellite imagery spanning 21-time steps, with 64 spectral bands at each step. The temporal stack was reshaped and flattened into a feature vector of dimension 1,344 per pixel, which was then used to

train the MLP classifier. The dataset was split into training (70%), validation (15%), and test (15%) subsets using a stratified random sampling strategy based on land cover classes. This approach ensured representative class distributions across all subsets and mitigated the effects of class imbalance during model training and evaluation. The MLP architecture shown in Figure 2 consisted of five dense layers with Rectified linear unit (ReLU) activations, L2 regularization, dropout for overfitting control, and batch normalization to stabilize learning and output layer with softmax activation corresponding to the number of LULC classes.

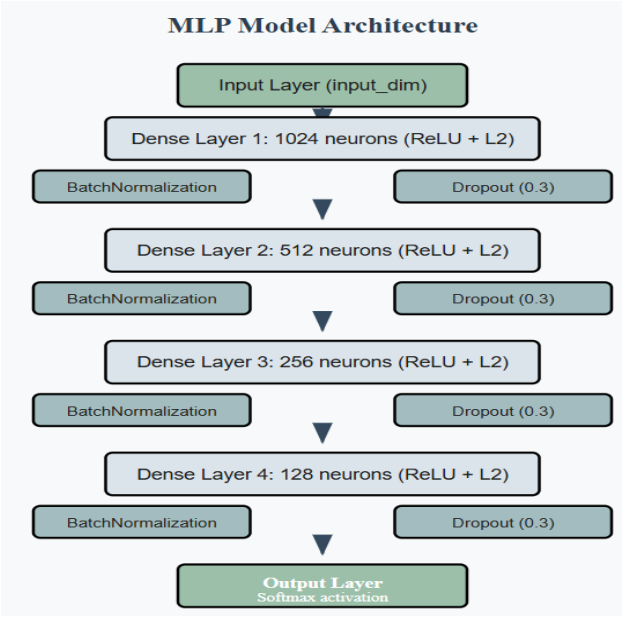

Figure 2. MLP Model Architecture

The model was trained using a categorical cross-entropy loss with the Adam optimizer. A batch size of 128 was used over a maximum 100 epochs and to improve generalization and prevent overfitting, early stopping was employed with patience of 10 epochs.

3.2. 2D Convolutional Neural Networks

2D CNNs have emerged as a powerful tool for LULC classification, particularly due to their ability to extract feature from high-resolution remote sensing images. Unlike pixels-based classifier, 2D CNNS can learn hierarchical spatial patterns by applying convolutional filters across local neighbourhoods, making them especially effective in distinguishing complex land cover classes. Recent studies have showed strong performance of 2D CNNs in this domain. For instance, Simón Sánchez et al. (2023) applied a 2D CNN to multi-temporal Sentinel-2 imagery for agricultural land use classification in Europe, achieving a weighted accuracy of over 91% across 19 classes, where class-wise accuracies were weighted by the proportion of samples in each class to account for class imbalance. Their results showed that 2D CNNs consistently outperformed traditional classifiers, including MLPs and Random Forest, especially in heterogeneous landscapes. Similarly, Abidi et al. (2023) proposed encoding time-series data into 2D satellite images using Gramian Angular technique before applying a CNN, and reported average F1-scores of 89-90% across multiple regions in France and Burkina Faso, highlighting the power of indirect CNN-based time-series modelling.

In this study, to capture the complex spectral-temporal relationships in the LULC classification task, a deep 2D CNN with residual connections was developed. The input dataset consists of 21 time -steps, each containing 64 features. After flattening spatially, the dataset was split into training (70%), validation (15%), and test (15%) subsets. The model was trained using the Adam optimizer and using categorical cross-entropy as the loss function. A batch size of 1024 and a maximum of 100 epochs were used.

The model as shown in Figure 3, employed multiple convolutional layers with residual blocks to deepen the network while mitigating vanishing gradients. Residual connections help preserve the input information and enhance gradient flow. Each block consists of two convolutional layers followed by batch normalization and ReLU activations. Dropout layers were added to prevent overfitting, and L2 regularization was applied to the convolutional and dense layers.

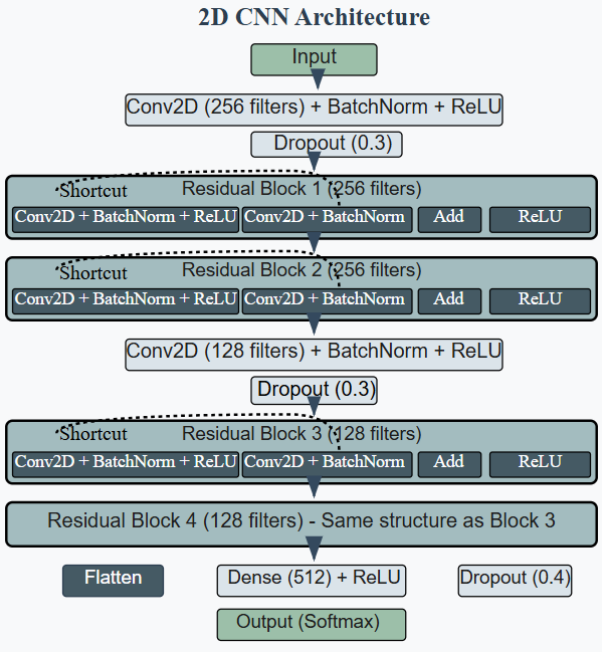

Figure 3. 2D CNN Model Architecture

3.3. 3D Convolutional Neural Networks

The use of 3D Convolutional Neural Networks has gained substantial attraction in recent years for LULC classification, particularly in applications involving multi-temporal and multi-spectral satellite data. Unlike 2D CNNs, which operate on spatial dimensions alone, 3D CNNs extend the convolution operation into the temporal or spectral

dimension, enabling the model to learn joint spatiotemporal and spectral features directly from image sequences. This characteristic makes 3D CNNs particularly well-suited for remote sensing tasks where both temporal evolution and spatial context are critical.

Voelsen et al. (2022) conducted a comparative study using Sentinel-2 time series and demonstrated that incorporating 3D convolution yields a mean F1-score increase of 4% and overall accuracy gain of 1.3% over equivalent 2D CNNs. Carneiro et al. (2025) introduced an efficient 3D CNN architecture tailored for Sentinel-2 land cover classification under limited ground truth constraints. Their model, which utilized 3×3×3 spatial-spectral kernels on 5×5 pixel patches, achieved an F1-score of 0.84—outperforming both 2D CNN variants and a baseline Random Forest classifier. This study underscores the capacity of 3D CNNs to leverage spatial-spectral information even when training data is scarce, making them suitable for operational use in large-area LULC assessments aligned with international frameworks like Land use, land-use change and forestry.

To effectively capture the spectral and temporal dependencies inherent in multi-temporal remote sensing data, a custom 3D CNN architecture was developed for pixel-based LULC classification. The input data consisted of 21 time-steps with 64 features per timestamp, forming a 4D input tensor of shape (H, W, T, C), where H and W represent spatial dimensions (630×652), T = 21 is the temporal depth, and C = 64 is the number of channels per time step.

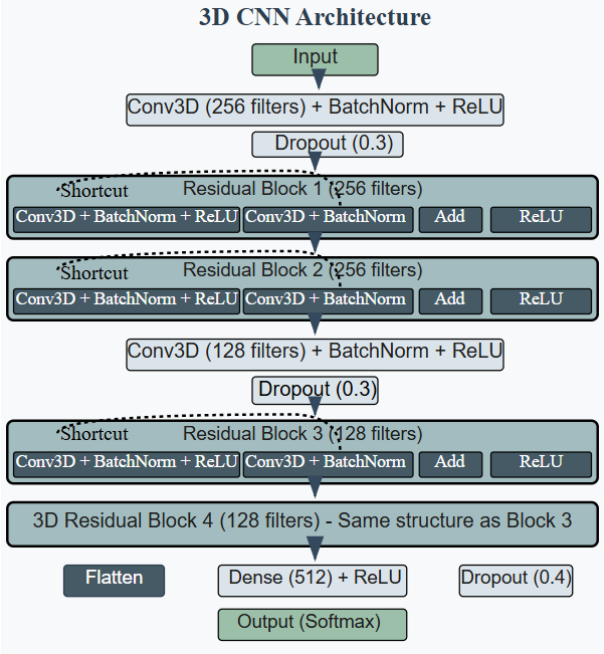

Figure 4. 3D CNN Model Architecture

The model architecture shown in Figure 4 begins with a 3D convolution layer capturing local patterns across adjacent time steps and spectral bands. Batch normalization and ReLU activation follow to stabilize and non-linearize the learning process. A dropout layer with a rate of 0.3 is applied to reduce overfitting. To enhance representational capacity, two residual blocks are used, each consisting of two 3D convolutional layers with skip connections and ReLU activations. These blocks preserve input information while allowing deeper transformations of the spectral-temporal signal. The model is trained using the Adam optimizer and categorical cross-entropy loss. Early stopping and learning rate reduction callbacks are used to prevent overfitting and optimize convergence. The data is split into training (70%), validation (15%), and testing (15%) sets, and the model is trained for up to 75 epochs with a batch size of 1024.

3.4. Recurrent Neural Networks

RNNs and their advanced variants, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures, have emerged as powerful tools for LULC classification, particularly when working with multi-temporal remote sensing data. Unlike traditional neural networks that process data independently, RNNs are specifically designed to capture dependencies across time by retaining information from previous steps in a sequence. This characteristic makes them especially suitable for handling time-series satellite imagery, where the spectral and phenological patterns evolve over months or years and are critical for distinguishing between land cover types such as cropland, forest, urban areas, and water bodies.

In Tejasree and Agilandeeswari (2024) have used networks on multi-temporal Sentinel-2 data. Their model significantly improved classification accuracy in regions with seasonal agricultural and ecological changes, outperforming traditional pixel-based methods. Alonso-Sarria et al, (2024) demonstrated that RNN-based models outperform Random Forests for cloud removal and land cover classification, achieving over 92% test accuracy on multi-temporal Sentinel-2 data.

To exploit the temporal dynamics of remote sensing data, we implemented a deep RNN-based model using Bidirectional Long Short-Term Memory (BiLSTM) layers, designed to classify LULC at the pixel level. The input data consisted of 21 time-steps and 64 features per time steps. The model architecture as shown in Fig 5 begins with two stacked BiLSTM layers. The first BiLSTM layer contains 128 hidden units and outputs full sequences, allowing the second BiLSTM layer (with 64 units) to process the entire sequence again in both forward and backward directions. Each LSTM layer is followed by Layer Normalization and Dropout to stabilize training and reduce overfitting.

A multi-head self-attention mechanism is applied after the second BiLSTM layer. This layer enables the model to focus on important time steps by learning weighted dependencies across the entire temporal sequence. The attention output is added back to the LSTM output via a residual connection, followed by another normalization step. The third BiLSTM layer has 128 units and outputs a fixed-length vector

summarizing the entire time series for each pixel. This output is then passed through a residual feedforward block, where two dense layers (each with 256 units and ReLU activation) are added to a projected shortcut connection to allow gradient flow and deeper representation learning. Finally, a dense layer with 128 units and another dropout layer lead into a softmax output layer for multi-class classification.

The model was compiled using the Adam optimizer and trained with categorical cross-entropy loss. Early stopping and learning rate reduction callbacks were used to improve convergence and prevent overfitting. The training was conducted over 100 epochs with a batch size of 1024, using stratified random splits based on land cover classes for training (70%), validation (15%), and test (15%) sets, ensuring representative class distributions across all subsets.

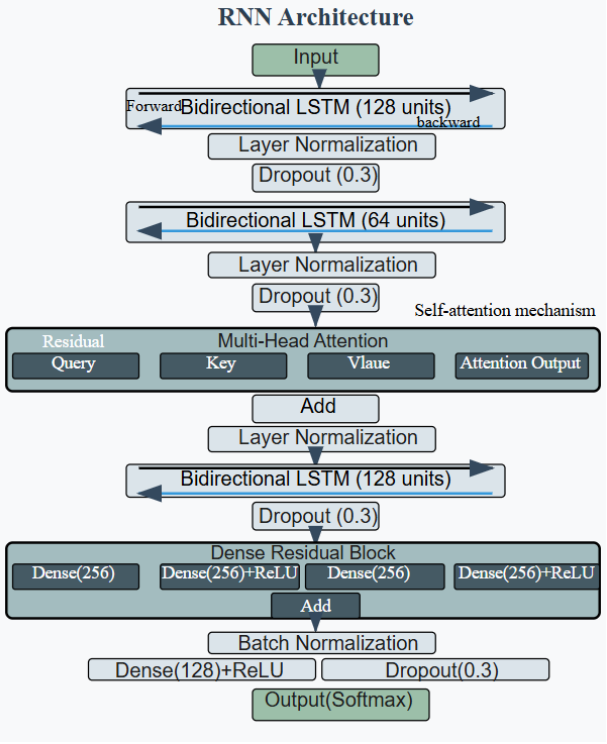

Figure 5. RNN Model Architecture

3.5. Evaluation Metrics

To evaluate the performance of the developed models, we employed a comprehensive set of quantitative and visual metrics. These includes test accuracy, learning curves (accuracy and loss), classification report (precision, recall, F1-score, and support), IoU per class, mean IoU and pixel-wise difference between predicted and ground truth land cover maps for last year (2022).

These metrics are widely used and recommended in remote sensing and land use/land cover classification studies, as they provide class-specific and overall performance information and are robust to class imbalance Congalton and Green (2015); Pontius and Millones (2011)

In this study, thematic accuracy evaluation is reported through these indices rather than the confusion matrix itself, as all metrics are directly computed from it.

Test accuracy reflects the proportion of correctly classified pixels on the unseen test dataset. Precision quantifies the model's accuracy in labelling a particular class. Recall measures the ability to identify all instance of that class. F1-score balances precision and recall. Support indicates the number of ground truth samples for each class IoU is a widely used metrics in semantic segmentation tasks, calculated as the overlap between predicted and ground truth areas divided by their union. Based on these metrics, we compared the performance of the MLP, 2D CNN, 3D CNN, and RNN models.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (3)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (4)$$

$$F1 - score_i = 2 * \frac{Recall_i * Precision_i}{Recall_{i_i} + Precision_{i_i}} \quad (5)$$

$$Support_i = Total\ number\ of\ class\ i \quad (6)$$

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (7)$$

where TP_i = True positives for class i
TN_i = True negatives for class i
FP = False positives for class i
FN_i = False negatives for class i

4. Results

To assess the learning behavior and convergence of the models during training, we plotted accuracy and loss curves for both the training and validation sets (Figures 6,7,8,9). These curves help visualize how quickly and effectively each model learned, as well as potential overfitting or underfitting patterns. All models showed stable convergence, with the 3D CNN exhibiting smoother loss reduction and consistently high validation accuracy across epochs.

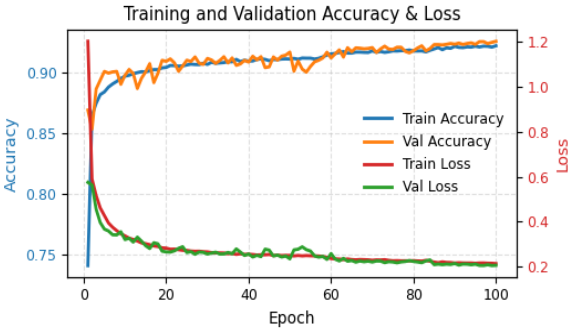

Figure 6. MLP model accuracy/loss curves

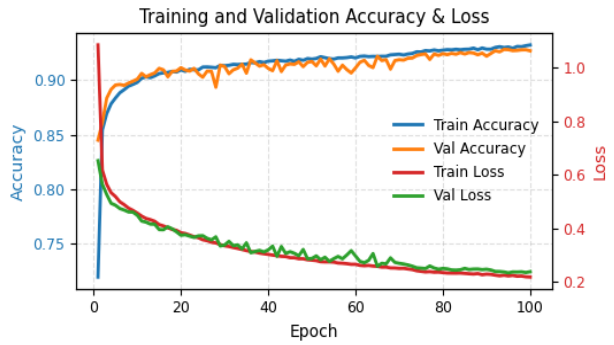

Figure 7. 2D CNN model accuracy/loss curves

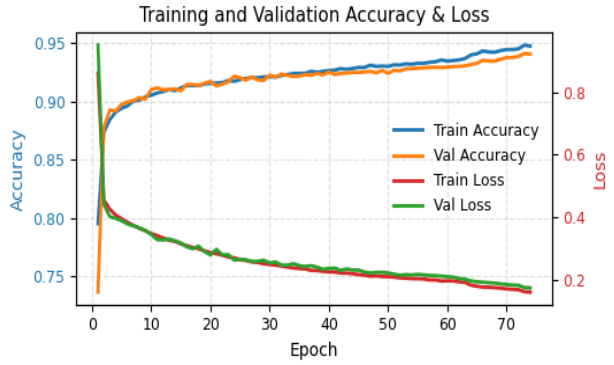

Figure 8. 3D CNN model accuracy/loss curves

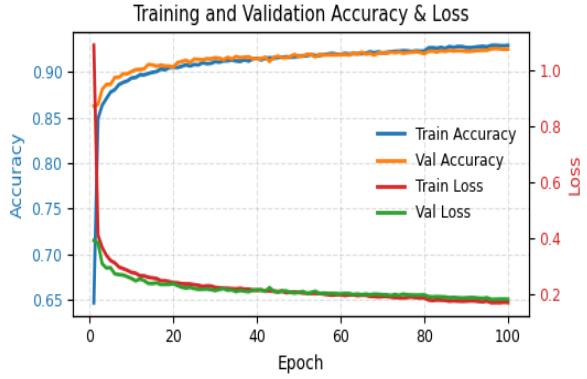

Figure 9. RNN model accuracy/loss curves

As shown in Table 2, the 3D CNN model achieved the highest test accuracy (92.82%) and mean IoU (0.6351), while also exhibiting the least number of differing pixels compared to the MLP and RNN models performed slightly worse, while the 2D CNN model achieved competitive results, particularly in spatial classification.

Table 2. Overall performances metrics of models

Model	Test	Mean IoU	Pixel difference
MLP	0.9248	0.5764	29114
2D CNN	0.9268	0.5908	26510
3D CNN	0.9282	0.6351	22896
RNN	0.9232	0.5887	27652

Table 3 provides the Intersection over Union (IoU) scores for each land cover class across the four models. The 3D CNN consistently outperformed others in most classes, especially in stable land covers such as class1-4. The models generally achieved higher IoU scores for land cover classes that occupied larger spatial extents and had greater representation in the training data. This suggests that model performance is positively correlated with class prevalence.

Table 3. Per class IoU comparison for models

Class/ Model	MLP	2D CNN	3D CNN	RNN
Class 1	0.9158	0.5764	0.8303	29114
Class 2	0.7326	0.5908	0.8456	26510
Class 3	0.9231	0.6351	0.8677	22896
Class 4	0.7958	0.5887	0.8383	27652
Class 5	0.5029	0.4769	0.5596	0.5245
Class 6	0.5450	0.5953	0.6429	0.5657
Class 7	0.0430	0.0971	0.1970	0.0611
Class 8	0.1533	0.1423	0.1947	0.1730

Tables 4 to 7 show the precision, recall, and F1-score per class for each model. The 3D CNN again shows superior balance across all metrics, particularly for dominant land cover types.

Table 4. Evaluation metrics for MLP model

Class/ Model	precision	recall	F1-score
Class 1	0.95	0.96	0.96
Class 2	0.86	0.83	0.85
Class 3	0.95	0.97	0.96
Class 4	0.90	0.88	0.89
Class 5	0.72	0.63	0.67
Class 6	0.82	0.62	0.71
Class 7	0.48	0.05	0.08
Class 8	0.58	0.17	0.27
Accuracy	----------	---------	0.93
Macro avg	0.78	0.64	0.67
Weighted avg	0.93	0.93	0.93

Table 5. Evaluation metrics for 2D CNN model

Class/ Model	precision	recall	F1-score

Class 1	0.95	0.97	0.96
Class 2	0.87	0.86	0.86
Class 3	0.97	0.96	0.96
Class 4	0.91	0.87	0.89
Class 5	0.78	0.55	0.65
Class 6	0.80	0.70	0.75
Class 7	0.44	0.11	0.18
Class 8	0.52	0.16	0.25
Accuracy	----------	---------	0.94
Macro avg	0.78	0.65	0.69
Weighted avg	0.93	0.94	0.93

Table 6. Evaluation metrics for 3D CNN model

Class/ Model	precision	recall	F1-score
Class 1	0.96	0.97	0.97
Class 2	0.89	0.88	0.88
Class 3	0.97	0.96	0.97
Class 4	0.91	0.90	0.90
Class 5	0.72	0.71	0.72
Class 6	0.84	0.74	0.78
Class 7	0.68	0.22	0.33
Class 8	0.74	0.21	0.33
Accuracy	----------	---------	0.94
Macro avg	0.84	0.70	0.73
Weighted avg	0.94	0.94	0.94

Table 7. Evaluation metrics for RNN model

Class/ Model	precision	recall	F1-score
Class 1	0.95	0.97	0.96
Class 2	0.87	0.84	0.86
Class 3	0.95	0.97	0.96
Class 4	0.92	0.86	0.89
Class 5	0.73	0.65	0.69
Class 6	0.80	0.66	0.72
Class 7	0.47	0.07	0.12
Class 8	0.43	0.23	0.29
Accuracy	----------	---------	0.93
Macro avg	0.76	0.65	0.69
Weighted avg	0.93	0.93	0.93

Figure 10,11 compares the predicted land cover map for the year 2022, generated by the best-performing 3D CNN model, with the actual ground truth. The ground truth land cover map shown in Figure 10 corresponds to the reference LULC data for the year 2022, which was obtained from the MODIS MCD12Q1.006 land cover product and used as the target labels for model training and evaluation. This dataset provides annually updated land cover information and is widely used in regional- and global-scale LULC studies. The predicted map demonstrates high visual similarity and spatial coherence, particularly in dominant land covers. Both Figures 10 and 11 follow the same classification.

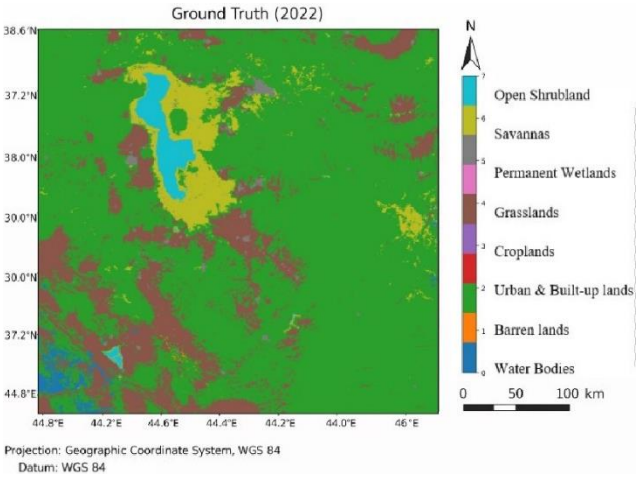

Figure 10. Ground truth map of 2022

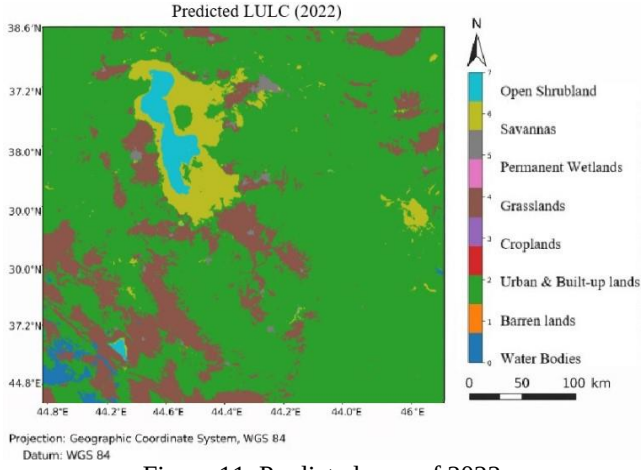

Figure 11. Predicted map of 2022

A comparison between the ground truth land cover map (Figure 10) and the predicted land cover map (Figure 11) reveals a high level of spatial agreement across the study area. Major land cover classes, particularly dominant vegetation and water bodies associated with Lake Urmia, are consistently identified in both maps with similar spatial extent and boundaries. The lake shoreline and surrounding transitional zones are well preserved in the predicted map, indicating the model's ability to capture fine-scale spatial patterns. Minor discrepancies are mainly observed in heterogeneous and fragmented regions, where gradual transitions between land cover types occur. These spatial differences are reflected quantitatively by an overall accuracy of 92.82% and a mean Intersection over Union (IoU) of 0.6351, confirming that the predicted map closely

matches the reference data while maintaining spatial coherence across the basin.

5. Conclusion

In this study, we explored the application of deep learning techniques for land use and land cover (LULC) prediction in the Urmia Lake basin over a 22-year period using multi-source geospatial data. Four deep learning models were developed and systematically evaluated based on their ability to capture complex spatial and temporal patterns in satellite imagery and climate variables. The results demonstrated that all models achieved strong performance, with the 3D Convolutional Neural Network (3D CNN) outperforming the other architectures in terms of both classification accuracy and spatial consistency.

Notably, the 3D CNN model achieved the highest mean Intersection over Union (IoU), overall accuracy, and the lowest pixel-wise mismatch, indicating robust generalization across diverse land cover types. Class-wise analysis revealed that predictive performance was strongest for dominant land cover categories, highlighting the influence of class prevalence on classification accuracy. Compared with previous LULC studies in the Urmia Lake basin, which primarily relied on conventional machine learning or single-source data, the proposed deep learning framework demonstrates improved capability in modeling nonlinear relationships and long-term spatio-temporal dynamics. These findings confirm the research hypothesis that deep learning models incorporating explicit spatial and temporal context provide more accurate and reliable LULC predictions.

The research objectives were successfully fulfilled by generating spatially explicit LULC maps, quantitatively comparing multiple deep learning architectures, and identifying the most effective model for the study area. The close agreement between the predicted and reference LULC maps for the year 2022 supports the generality and robustness of the proposed framework, suggesting its potential applicability to other regions experiencing complex land cover dynamics.

The main contributions of this research include the integration of a multi-dimensional Earth observation data cube combining optical, climatic, topographic, and precipitation variables, as well as a comprehensive and fair comparison of multiple pixel-based deep learning models under identical experimental conditions. The demonstrated effectiveness of the 3D CNN highlights its advantage in jointly learning spatial and temporal features for environmentally sensitive basins.

Despite these advantages, some limitations should be acknowledged. The use of medium-resolution input data may limit the detection of fine-scale land cover transitions, and the availability of reference data constrained validation to recent years. Additionally, the computational complexity of deep spatio-temporal models imposes higher processing demands.

Future research may address these limitations by integrating higher-resolution datasets, extending temporal validation with additional reference data, and incorporating explainable artificial intelligence techniques to enhance model interpretability. Further application of the proposed framework to other lake basins and arid or semi-arid regions would also help assess its broader transferability and operational potential for long-term environmental monitoring and decision-making.

References

Abidi, A., Ienco, D., Abbes, A. Ben, and Farah, I. R. (2023). Combining 2D encoding and convolutional neural network to enhance land cover mapping from Satellite Image Time Series. Engineering Applications of Artificial Intelligence,122,106152.https://doi.org/10.1016/j.engappai.2023.106152

AghaKouchak, A., Norouzi, H., Madani, K., Mirchi, A., Azarderakhsh, M., Nazemi, A., Nasrollahi, N., Farahmand, A., Mehran, A., and Hasanzadeh, E. (2015). Aral Sea syndrome desiccates Lake Urmia: Call for action. Journal of Great Lakes Research, 41(1), 307–311. https://doi.org/10.1016/j.jglr.2014.12.007

Ahmed, B., and Ahmed, R. (2012). Modeling urban land cover growth dynamics using multioral satellite images: A case study of Dhaka, Bangladesh. ISPRS International Journal of Geo-Information, 1(1), 3–31. https://doi.org/10.3390/ijgi1010003

Alizade Govarchin Ghale, Y., Baykara, M., and Unal, A. (2017). Analysis of decadal land cover changes and salinization in Urmia Lake Basin using remote sensing techniques. https://doi.org/10.5194/nhess-2017-212

Alonso-Sarria, F., Valdivieso-Ros, C., and Gomariz-Castillo, F. (2024). Imagery Time Series Cloud Removal and Classification Using Long Short Term Memory Neural Networks. Remote Sensing, 16(12), 2150. https://doi.org/10.3390/rs16122150

CALANCA, P. (2007). Climate change and drought occurrence in the Alpine region: How severe are becoming the extremes? Global and Planetary Change, 57(1–2), 151–160. https://doi.org/10.1016/j.gloplacha.2006.11.001

Carneiro, G. A., Svoboda, J., Cunha, A., Sousa, J. J., and Štych, P. (2025). Efficient 3D CNN for Sentinel-2 land cover classification with limited ground truth data. European Journal of Remote Sensing, 58(1). https://doi.org/10.1080/22797254.2025.2512531

Cheng, G., Huang, Y., Li, X., Lyu, S., Xu, Z., Zhao, H., Zhao, Q., and Xiang, S. (2024). Change Detection Methods for Remote Sensing in the Last Decade: A Comprehensive Review. Remote Sensing, 16(13), 2355. https://doi.org/10.3390/rs1613235

Congalton, R.G., & Green, K. (2019). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices, Third

Edition (3rd ed.). CRC Press. https://doi.org/10.1201/9780429052729

Feizizadeh, B., Omarzadeh, D., Mohammadzadeh Alajujeh, K., Blaschke, T., and Makki, M. (2022). Impacts of the Urmia Lake Drought on Soil Salinity and Degradation Risk: An Integrated Geoinformatics Analysis and Monitoring Approach. Remote Sensing, 14(14), 3407. https://doi.org/10.3390/rs14143407

Hasan, S. S., Zhen, L., Miah, Md. G., Ahamed, T., and Samie, A. (2020). Impact of land use change on ecosystem services: A review. Environmental Development, 34, 100527. https://doi.org/10.1016/j.envdev.2020.100527

Shafizadeh-Moghadam, H., Hagenauer, J., Helbich, M., Farajzadeh, M. (2015). Performance analysis of radial basis function networks and multi-layer perceptron networks in modeling urban change International Journal of Geographical Information Science, 29(4), 606–623.

Jamali, A., Roy, S. K., Hong, D., Atkinson, P. M., and Ghamisi, P. (2024). Spatial-Gated Multilayer Perceptron for Land Use and Land Cover Mapping. IEEE Geoscience and Remote Sensing Letters, 21, 1–5. https://doi.org/10.1109/LGRS.2024.3354175

Kumar, V., and Agrawal, S. (2023). A multi-layer perceptron–Markov chain based LULC change analysis and prediction using remote sensing data in Prayagraj district, India. Environmental Monitoring and Assessment, 195(5), 619. https://doi.org/10.1007/s10661-023-11205-w

Liu, J., Shahroudy, A., Xu, D., Kot, A. C., and Wang, G. (2018). Skeleton-Based Action Recognition Using Spatio-Temporal LSTM Network with Trust Gates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(12), 3007–3021. https://doi.org/10.1109/TPAMI.2017.2771306

Mashala, M. J., Dube, T., Mudereri, B. T., Ayisi, K. K., and Ramudzuli, M. R. (2023). A Systematic Review on Advancements in Remote Sensing for Assessing and Monitoring Land Use and Land Cover Changes Impacts on Surface Water Resources in Semi-Arid Tropical Environments. Remote Sensing, 15(16), 3926. https://doi.org/10.3390/rs15163926

Minar, M. R., and Naher, J. (2018). Recent Advances in Deep Learning: An Overview. https://doi.org/10.13140/RG.2.2.24831.10403

Pontius, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement. *International Journal of Remote Sensing*, 32(15), 4407–4429.

Se frin, O., Riese, F. M., and Keller, S. (2020). Deep Learning for Land Cover Change Detection. Remote Sensing, 13(1), 78. https://doi.org/10.3390/rs13010078

Simón Sánchez, A.-M., González-Piqueras, J., de la Ossa, L., and Calera, A. (2022). Convolutional Neural Networks for Agricultural Land Use Classification from Sentinel-2 Image Time Series. Remote Sensing, 14(21), 5373. https://doi.org/10.3390/rs14215373

Tejasree, G., and Agilandeeswari, L. (2024). Land use/land cover classification using deep-LSTM for hyperspectral images. The Egyptian Journal of Remote Sensing and Space Sciences, 27(1), 52–68. https://doi.org/10.1016/j.ejrs.2024.01.004

Voelsen, M., Teimouri, M., Rottensteiner, F., and Heipke, C. (2022). INVESTIGATING 2D AND 3D CONVOLUTIONS FOR MULTITEMPORAL LAND COVER CLASSIFICATION USING REMOTE SENSING IMAGES. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, V-3–2022, 271–279.https://doi.org/10.5194/isprs-annals-V-3-2022-271-2022

Winkler, K., Fuchs, R., Rounsevell, M., and Herold M. (2021). Global land use changes are four times greater than previously estimated. Nature Communications,12(1),2501.https:/doi.org/10.1038/s41467-021-22702-2