IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

An Intelligent Dilated Convolutional Attention Network and LSTM for Enhanced Land Use and Land Cover Mapping

Anushka Dilip Game, Pallavi Sitaram Gaikwad, Pratik Nanasaheb Jugrut, Suraj Bharat Gitkhane, Prof. Y. M. Wadghule
SND College of Engineering & Research Centre, Yeola, Maharashtra, India

Abstract: *Land Use and Land Cover (LULC) mapping is an important remote-sensing task used to identify and monitor the spatial distribution of forests, agricultural land, urban areas, water bodies, barren land, and transportation corridors. Manual interpretation of satellite imagery is time-consuming and can be affected by spectral similarity, mixed pixels, image noise, seasonal variation, and unclear class boundaries. This paper presents a proposed deep-learning framework that combines dilated convolutional feature extraction, attention mechanisms, and Long Short-Term Memory (LSTM) modelling for enhanced LULC mapping. The proposed workflow receives satellite imagery, performs radiometric correction, normalization, noise handling, augmentation, and patch extraction, and then extracts multi-scale spatial features using dilated convolution blocks. Channel and spatial attention are used to emphasize informative features, while the LSTM module is intended to learn temporal relationships when aligned multi-temporal observations are available. A segmentation head generates pixel-wise class predictions, followed by post-processing, boundary extraction, and class-wise area statistics. The supplied project material reports comparative values for Random Forest, U-Net, SegNet, DeepLabV3+, and the proposed model. These values are reproduced as supplied and are not treated as a controlled cross-paper benchmark because dataset, split, training configuration, and evaluation protocol details were not included in the presentation. The framework is intended to support more detailed land-cover classification, boundary delineation, and optional year-to-year change analysis.*

Keywords: Land Use and Land Cover (LULC), Remote Sensing, Semantic Segmentation, Dilated Convolution, Attention Mechanism, LSTM, Multi-Temporal Satellite Images, Change Detection, Area Estimation.

I. INTRODUCTION

Land Use and Land Cover (LULC) information describes how the Earth's surface is occupied and covered by natural and human-made features. Land-cover classes such as forest, agriculture, urban surfaces, water bodies, barren land, and roads provide essential information for urban planning, agricultural assessment, environmental monitoring, resource management, and disaster response. Satellite remote sensing enables repeated observation of large geographic regions and provides imagery at different spatial, spectral, and temporal resolutions. The conversion of these observations into accurate thematic maps is therefore a key task in geospatial analysis.

Traditional image interpretation and pixel-based classification approaches may struggle when different land-cover categories have similar spectral signatures or when a single pixel contains multiple surface materials. High-resolution remote-sensing imagery also contains complex spatial patterns, irregular object shapes, shadows, and fine boundaries. Deep-learning-based semantic segmentation has improved the extraction of spatial context from remote-sensing data, but performance remains dependent on training data, class balance, image quality, and the ability to represent objects at

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

multiple scales. Reviews of high-spatial-resolution land-use mapping identify these issues as important research challenges [1].

Convolutional neural networks (CNNs) are effective at extracting local spatial features. However, conventional convolution layers have a limited receptive field unless the network becomes deeper or uses additional pooling. Dilated convolution expands the receptive field without requiring equivalent increases in kernel size, making it useful for capturing context at multiple spatial scales. Liu et al. proposed a dense dilated convolution merging network for land-cover classification and demonstrated the value of combining local and broader contextual information [2]. Recent remote-sensing segmentation work has also explored dilated convolution together with attention-based decoding to improve feature selection and segmentation efficiency [3].

Attention mechanisms provide a complementary capability by weighting informative channels or spatial regions. Global-local attention approaches have been investigated for remote-sensing semantic segmentation because both broad scene context and local details are needed to distinguish complex land-cover patterns [4]. Other attention-guided approaches have combined multi-scale features and boundary supervision to improve the delineation of small or irregular objects [5], [6]. These studies motivate the use of channel attention and spatial attention in the proposed architecture.

Land-cover mapping may also involve observations collected at multiple dates. Temporal information can help distinguish seasonal variation from persistent land-cover change. Recurrent neural networks, including LSTM-based models, are designed to learn dependencies across ordered sequences. Research comparing CNN, RNN, and vision-transformer architectures for Sentinel time series highlights the importance of temporal representation when multi-temporal satellite observations are used [7]. Russo et al. further studied spatio-temporal Sentinel-1 features for land-cover mapping using seasonal feature sequences [8].

Based on these research directions, this paper presents an integrated framework combining dilated CNN feature extraction, channel and spatial attention, and LSTM-based temporal modelling. The proposed workflow is based on the project presentation supplied for this study. It includes satellite-image input, preprocessing, multi-scale feature extraction, attention, temporal modelling, pixel-wise segmentation, post-processing, boundary detection, area calculation, and output-map generation. The main objective is to provide a structured architecture for accurate LULC mapping and optional temporal change analysis.

II. RELATED WORK AND LITERATURE REVIEW

Research in remote-sensing LULC mapping has moved from conventional pixel classifiers toward deep architectures that learn spatial, spectral, and temporal representations. The following studies are particularly relevant to the proposed combination of dilated convolution, attention, and LSTM.

Liu et al. introduced the Dense Dilated Convolutions' Merging Network (DDCM-Net), which combines dilated convolution branches to capture multi-scale context for land-cover classification. Their work addresses the difficulty of distinguishing objects with similar colour and texture and provides a direct foundation for the multi-scale feature extractor considered here [2].

Russo et al. investigated land-cover mapping using spatio-temporal Sentinel-1 features and a Swin-Unet-based architecture. Their study emphasizes the value of temporal observations and seasonal feature sequences, which is relevant to the temporal modelling stage of this project [8].

Liu et al. proposed a lightweight remote-sensing image segmentation network using dilated convolution and a data-dependent attention decoder. This work is relevant to the proposed integration of multi-scale convolutional features and attention-based feature selection [3].

Attention has also been applied to land-cover change detection. A multiscale attention network guided by a change-gradient image combines multi-scale information and position-channel attention to improve change maps [5]. The GLSANet approach uses global and local self-attention to represent long-range relationships and fine spatial details in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

remote-sensing semantic segmentation [4]. Boundary-supervision-aided multiscale channel-wise cross-attention has been studied to reduce boundary blurring and improve segmentation of complex aerial objects [6].

Temporal modelling is supported by research on recurrent networks and satellite time series. Ienco et al. studied land-cover classification using multi-temporal spatial data and deep recurrent neural networks [9]. A comparative evaluation of CNN, RNN, and ViT models for Sentinel time-series representation further demonstrates the relevance of architectures that preserve temporal information [7]. These studies support the motivation for including an LSTM stage when multiple aligned dates are available.

Other related research includes self-supervised feature learning for multimodal land-cover classification [10], collaborative attention-based heterogeneous gated fusion for land-cover classification [11], and cross-resolution deep learning for high-resolution land-cover mapping [12]. EuroSAT provides a widely used Sentinel-2 benchmark for land-use and land-cover classification [13]. Recent IEEE work also investigates real-time semantic segmentation, temporal-domain adaptation, cross-city generalization, and efficient architecture search for remote-sensing images [14]–[18]. These studies highlight the importance of accuracy, generalization, temporal consistency, and computational efficiency.

III. RESEARCH GAP AND MOTIVATION

The reviewed studies address important parts of the LULC mapping problem, including multi-scale convolution, attention-based feature fusion, temporal sequence learning, boundary recovery, and cross-region generalization. However, these components are often investigated as separate architectural choices or evaluated under different datasets and protocols. The project therefore explores an integrated pipeline in which dilated convolution extracts multi-scale spatial information, attention emphasizes informative channel and spatial responses, and LSTM models temporal dependencies when multi-date imagery is available.

A further challenge is that land-cover categories may have similar appearance while their boundaries can be narrow, irregular, or partially obscured by noise. The proposed workflow includes preprocessing, segmentation, post-processing, and boundary extraction rather than relying only on a final pixel classifier. It also includes class-wise area estimation and optional change analysis to make the resulting map more useful for geospatial interpretation.

The supplied presentation does not specify the satellite sensor, study-area coordinates, image acquisition dates, dataset size, train-validation-test split, hardware, hyperparameters, or reproducibility settings. These details must be recorded during implementation before the reported performance can be independently validated. The research gap is therefore framed as an architectural and workflow integration objective, not as a claim that the proposed model has already surpassed all published methods.

IV. PROBLEM STATEMENT

Identifying land-cover categories from satellite imagery is challenging because different classes may have similar spectral characteristics, image boundaries may be unclear, and observations may contain noise, shadows, cloud effects, or seasonal variation. Conventional methods may fail to capture both fine spatial details and broad contextual relationships. When images from multiple dates are available, temporal changes must also be modelled without confusing seasonal variation with genuine land-cover conversion. An automated deep-learning framework is therefore required to perform multi-scale feature extraction, focus on informative regions, learn temporal patterns, generate pixel-wise LULC maps, and provide boundary and area-based summaries.

V. OBJECTIVES

- To classify different land-cover types from satellite imagery at pixel level.
- To extract multi-scale spatial features using dilated convolution blocks.
- To apply channel and spatial attention to emphasize informative land-cover features.
- To use LSTM-based temporal modelling for ordered multi-date imagery, where such data are available.
- To improve the delineation of land-cover boundaries through post-processing and boundary extraction.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

- To automate LULC map generation and reduce manual image interpretation effort.
- To calculate class-wise area, percentage distribution, and optional year-to-year land-cover change statistics.

VI. PROPOSED METHODOLOGY

The methodology follows a sequential processing pipeline, beginning with satellite-image acquisition and ending with a classified LULC map and statistical reports. The project presentation identifies data collection, preprocessing, feature/model design, training, evaluation, and final output as the major stages. The detailed workflow below expands those stages while retaining the architecture described in the supplied material.

A. Data Acquisition

Satellite images are collected for the selected study area. If temporal analysis is required, imagery from multiple acquisition dates is organized into an ordered sequence. Corresponding labelled masks are required for supervised segmentation. The source, sensor, acquisition date, spatial resolution, spectral bands, projection, and study-area extent should be recorded for every image. The supplied presentation does not identify a specific satellite sensor or dataset; these must be specified in the final experimental setup.

B. Image Preprocessing

Preprocessing prepares the imagery for consistent model input. The presentation specifies image resizing, radiometric correction and normalization, noise removal, data augmentation using rotation, flipping and cropping, and patch extraction. For multi-temporal input, images from different dates must be co-registered and aligned to a common grid so that corresponding pixels represent the same geographic locations. Cloud and no-data regions should be identified and handled consistently. Normalization parameters should be calculated using training data only to avoid data leakage.

C. Dilated Convolutional Feature Extraction

The feature extractor begins with a standard convolution block and then applies dilated convolution blocks with increasing dilation rates. As shown in the supplied architecture, the dilation rates are 2, 4, and 8. Dilated convolution inserts spaces between kernel elements, increasing the effective receptive field while retaining the feature-map resolution. This enables the network to represent both local texture and broader spatial context. Multi-scale feature information is combined before the attention stage.

D. Attention Mechanism

The architecture contains channel attention implemented using a Squeeze-and-Excitation (SE) block and a spatial-attention component. Channel attention estimates the relative importance of feature channels and reweights them so that useful spectral-spatial responses are emphasized. Spatial attention produces a spatial weighting map that highlights relevant regions in the feature representation. The combination is intended to reduce the influence of irrelevant background responses and improve discrimination among visually similar classes.

E. LSTM Temporal Modelling

The LSTM module receives an ordered feature sequence and learns temporal dependencies between observations. In the proposed workflow, the feature vectors or feature maps extracted from multiple dates are arranged in chronological order. LSTM gates regulate the retention and updating of temporal information, allowing the model to represent changes across observations. If only a single-date image is available, the temporal stage cannot provide genuine year-to-year change information; the system should then operate in a spatial segmentation mode or use a clearly defined sequence strategy.

F. Classification and Segmentation

The classification stage uses a fully connected prediction layer followed by dropout and softmax, as indicated in the supplied architecture. The softmax output represents the class probabilities for each pixel. The predicted class is selected from the LULC categories used by the project. The presentation identifies forest, agriculture, urban, water bodies, barren land, and roads as output classes. The exact class definitions and label encoding should be documented with the dataset.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

G. Post-Processing and Boundary Detection

The predicted segmentation map is refined using conditional random field (CRF) processing or majority filtering, followed by smoothing and refinement. Boundary extraction is performed using Canny edge detection or a morphological-gradient operation, as specified in the architecture diagram. Post-processing should be applied consistently and evaluated separately so that its contribution to the final map can be measured.

H. Area Calculation and Reporting

The final classified map is used to calculate class-wise area in hectares or square kilometres. The percentage occupied by each class is computed relative to the total valid mapped area. For multi-temporal outputs, a change-analysis stage can compare class maps from different dates and summarize transitions. Reports may include maps, tables, bar charts, and pie charts. Area calculation must account for the map projection and pixel ground area to avoid inaccurate totals.

VII. DATA ORGANIZATION AND INPUT PREPARATION

A. Spatial and Temporal Organization

The input data should be organized so that every image tile represents a known geographic footprint and every label corresponds to the same footprint. For multi-date analysis, the images should be arranged chronologically. Spatial alignment is important because a pixel-wise comparison between dates is meaningful only when corresponding ground locations occupy corresponding image positions. The supplied project presentation identifies multi-temporal satellite imagery as an input option, but it does not specify a particular satellite mission, sensor, acquisition interval, spatial resolution, or geographic study boundary. These items therefore remain project implementation decisions and should be documented when the dataset is finalized.

Each image should be associated with its acquisition date, source, band information, coordinate reference system, spatial resolution, and preprocessing history. Label masks should use a consistent class-code convention throughout training, validation, and testing. The class names shown in the project architecture are Forest, Agriculture, Urban, Water Bodies, Barren Land, and Roads. The same class order should be used by the annotation process, model output layer, confusion matrix, and area-reporting module. If a class is absent from a particular tile, it should still retain its defined class identifier so that class indices do not change between samples.

B. Quality Control of Satellite Images and Labels

Before model training, image tiles and their corresponding masks should be checked for missing files, invalid pixel values, inconsistent dimensions, and incorrect label identifiers. Cloud, shadow, haze, and no-data regions require particular attention in optical satellite imagery because they can obscure the ground surface or introduce misleading spectral patterns. The presentation includes noise removal and radiometric correction in the preprocessing stage. The exact correction procedure should be selected according to the sensor and data product rather than assumed to be identical for every source.

Image-mask pairs must undergo identical geometric operations. For example, when a tile is rotated, flipped, cropped, or divided into patches, the label mask must receive the same transformation. Image interpolation should preserve continuous-valued pixel data, whereas categorical masks should use nearest-neighbour interpolation to avoid creating invalid intermediate class codes. These checks help maintain pixel-level correspondence between the input image and the reference segmentation.

VIII. MODEL DESIGN AND PROCESSING FLOW

A. Multi-Scale Feature Representation

The proposed feature extractor uses dilated convolution to increase the receptive field without requiring the same increase in kernel size. A standard convolution samples neighbouring positions in a feature map. A dilated convolution inserts gaps between sampled positions, allowing the filter to gather information from a wider spatial neighbourhood. The architecture shown in the supplied presentation uses dilation rates of 2, 4, and 8 after a standard 3 × 3 convolution

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

block. This arrangement is intended to represent both local texture and broader spatial context that may help distinguish land-cover regions.

Multi-scale context is relevant to LULC mapping because classes can appear at different spatial scales. A road may be represented by a narrow linear structure, while an agricultural field or water body may cover a much larger region. The use of multiple dilation rates can provide feature responses over different neighbourhood sizes. However, the effective benefit depends on the image resolution, patch dimensions, and object sizes in the selected study area. These parameters should be evaluated on the project dataset rather than treated as universally optimal.

B. Channel and Spatial Attention

The attention stage follows feature extraction and is represented by channel attention and spatial attention. A Squeeze-and-Excitation block summarizes channel responses and learns channel-wise weights. These weights can increase the contribution of channels that contain useful information and reduce the contribution of less informative channels. Spatial attention performs a complementary operation by assigning weights to locations within a feature map. In a land-cover image, this may help the network emphasize regions containing relevant boundaries, textures, or spatial patterns.

The two attention operations address different dimensions of the feature representation. Channel attention asks which feature channels are useful, while spatial attention asks where useful information is located. Their combined use is intended to refine the multi-scale features before temporal modelling and pixel-wise prediction. The supplied presentation identifies these components but does not provide layer dimensions, reduction ratios, activation functions, or a complete mathematical specification. Such implementation details should be included in the final code documentation and experimental configuration.

C. Temporal Feature Learning with LSTM

When images are available for multiple dates, the model can use their chronological order to learn temporal dependencies. An LSTM maintains an internal cell state and uses input, forget, and output gates to regulate information flow. This mechanism is designed to retain useful information across a sequence and reduce the difficulty of learning long-range dependencies compared with a basic recurrent unit. In the proposed workflow, spatial features from successive observations are supplied to the LSTM module so that the prediction can use both the appearance of a location and its temporal context.

Temporal modelling requires that observations are sufficiently comparable. Images from different dates should be co-registered, and major differences caused by sensor characteristics, atmospheric conditions, seasonal timing, or resolution should be considered. A temporal model should not interpret misregistration or acquisition artefacts as actual land-cover change. If only one date is available, the LSTM component cannot provide meaningful multi-date information; the implementation should then clearly identify whether the temporal module is bypassed or whether a multi-date dataset is required.

D. Pixel-Wise Prediction and Map Refinement

The segmentation stage converts learned features into class scores for each pixel. The supplied architecture identifies a fully connected prediction layer, dropout, and softmax. Softmax converts the scores into class probabilities, and the class with the highest probability can be selected as the predicted label for each valid pixel. The resulting raster should preserve the geographic reference of the input tile so that it can be viewed and measured in a GIS environment.

The proposed post-processing stage includes conditional random field processing or majority filtering, followed by smoothing and refinement. These operations can reduce isolated predictions and improve local consistency, but excessive smoothing may remove narrow features such as roads or small water channels. Boundary extraction is shown as a subsequent step using Canny edge detection or a morphological gradient. The extracted boundaries should be interpreted as derived visual products; their accuracy depends on the quality of the segmentation map and the selected edge parameters.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

IX. TRAINING, VALIDATION, AND REPRODUCIBILITY

A reproducible implementation should separate the labelled dataset into training, validation, and test partitions. The test partition should remain isolated from model fitting and hyperparameter selection. Where images are spatially adjacent or originate from repeated observations of the same area, partitioning should be designed to reduce spatial or temporal leakage. Otherwise, near-duplicate regions may appear in both training and testing data and produce an overly optimistic estimate of generalization performance.

During training, the model parameters are updated using the training samples and a defined optimization objective. For multi-class segmentation, a pixel-wise classification loss such as categorical cross-entropy may be considered; class imbalance may motivate a weighted loss or an overlap-based loss. The supplied presentation does not identify the optimizer, learning rate, batch size, loss function, number of epochs, patch size, or stopping criterion. These values must be selected experimentally, recorded, and reported in the final implementation instead of being inferred from the architecture diagram.

Validation data can be used to monitor convergence, compare model configurations, and select a checkpoint. The final test evaluation should then be performed once using the selected configuration. All compared models should use the same dataset partitions, label definitions, and evaluation procedure if the comparison is intended to establish a fair performance difference. If published values from separate studies are included for context, they should be clearly distinguished from results obtained under a common experimental protocol.

For reproducibility, the project should retain the dataset source and version, acquisition dates, preprocessing steps, class mapping, split strategy, model configuration, software framework and version, hardware environment, random seed, training duration, and saved model checkpoint. Reporting these details enables another researcher to repeat the experiment and helps identify whether differences in reported metrics arise from the architecture, data, or evaluation protocol.

X. OUTPUT PRODUCTS AND INTERPRETATION

The expected primary output is a thematic LULC raster in which each valid pixel is assigned to one of the defined land-cover classes. A colour legend should map each class identifier to a consistent display colour. The map should include the study-area boundary, coordinate reference information, scale, north arrow, acquisition date or date range, and data-source note when it is prepared for reporting. The supplied architecture also indicates a boundary layer and class-wise area statistics as supporting outputs.

Area calculation should be performed using the appropriate map projection or geodesic area calculation. For a projected raster with known pixel dimensions, the area represented by a pixel can be derived from its ground sampling dimensions, subject to the projection and raster validity. The area for a class is then obtained by summing the valid pixels assigned to that class. The percentage share is calculated by dividing the class area by the total valid mapped area and multiplying by 100. No class-area totals can be stated in this paper because the supplied project material does not include a completed study-area map or pixel counts.

For multi-date outputs, change analysis can compare aligned classification maps from two or more dates. A transition matrix can summarize the number or area of pixels moving from one class at an earlier date to another class at a later date. Such comparisons require consistent class definitions and careful handling of invalid pixels. Classification errors from either date can propagate into the change map, so change statistics should be interpreted together with class-wise accuracy and, where possible, independent reference information.

The reporting module can present the classified map, extracted boundaries, class-wise area table, percentage distribution, and optional change summaries. Charts can help communicate the relative contribution of each class, while a confusion matrix and class-wise metrics are more suitable for evaluating classification quality. These products serve different purposes: visual maps support spatial interpretation, area tables quantify land-cover composition, and accuracy metrics assess agreement with reference labels.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

XI. SYSTEM ARCHITECTURE

The system architecture is organized into ten connected stages: (1) satellite-image input, (2) preprocessing, (3) dilated convolutional feature extraction, (4) channel and spatial attention, (5) LSTM temporal modelling, (6) classification and segmentation, (7) post-processing, (8) boundary detection, (9) area calculation and statistics, and (10) final output generation. The diagram supplied with the project presentation is reproduced below as the proposed architecture.

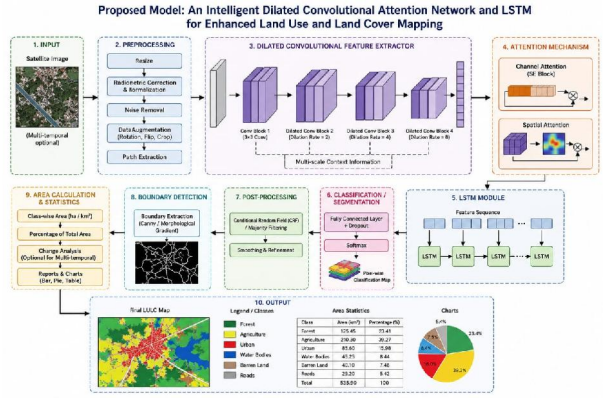


Fig. 1. Proposed dilated convolutional attention network and LSTM architecture for LULC mapping (reproduced from the supplied project presentation).

XII. SYSTEM MODULES

Module 1: Data Acquisition. Collect satellite images from selected dates and their corresponding labelled masks. Record source, sensor, acquisition date, resolution, projection, and study area.

Module 2: Image Preprocessing. Resize image tiles, perform radiometric correction and normalization, handle noise and cloud/no-data regions, apply augmentation, and extract patches.

Module 3: Dilated CNN Feature Extraction. Extract spatial features at multiple scales using convolution blocks and dilated convolution rates of 2, 4, and 8.

Module 4: Attention Mechanism. Apply channel attention through an SE block and spatial attention to emphasize informative feature channels and image regions.

Module 5: LSTM and Segmentation. Learn temporal dependencies from ordered aligned image features and generate pixel-wise class probabilities.

Module 6: Evaluation and Reporting. Calculate accuracy, precision, recall, F1-score, and IoU; generate segmented maps, boundaries, class-wise area, and optional change summaries.

XIV. COMPARATIVE PERFORMANCE ANALYSIS OF DIFFERENT MODELS

The supplied comparative-performance figure reports Overall Accuracy (OA), mean Intersection over Union (mIoU), and Macro F1 for five methods: Random Forest, U-Net, SegNet, DeepLabV3+, and the proposed model. The numerical values below are transcribed from the user-provided figure. The source figure itself notes that the values are reproduced from the supplied report and do not constitute a controlled cross-paper comparison.

Model	OA (%)	mIoU (%)	Macro F1 (%)
Random Forest	71.50	36.80	51.20
U-Net	78.40	44.20	59.80
SegNet	76.80	42.10	57.50
DeepLabV3+	82.30	49.50	65.20
Proposed Model	99.18	97.88	98.92

TABLE I. PERFORMANCE VALUES REPRODUCED FROM THE SUPPLIED PROJECT FIGURE.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

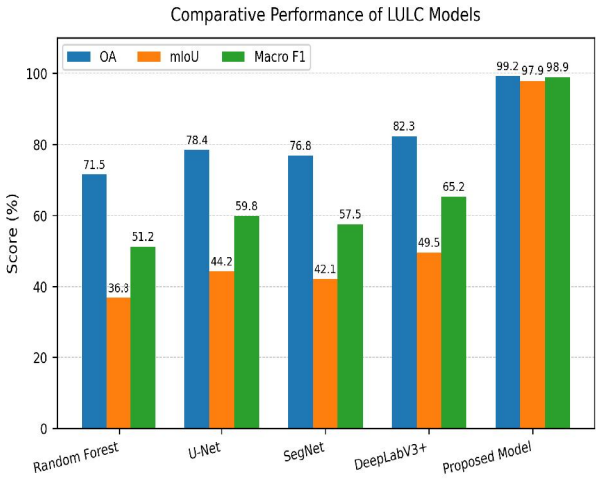


Fig. 2. Comparative performance chart supplied with the project material.

According to the supplied values, the proposed model has the highest reported OA, mIoU, and Macro F1 among the five listed methods. Its reported OA is 99.18%, mIoU is 97.88%, and Macro F1 is 98.92%. DeepLabV3+ is the strongest of the four listed comparison baselines in the supplied table, with OA of 82.30%, mIoU of 49.50%, and Macro F1 of 65.20%. Random Forest, U-Net, and SegNet have lower reported values in the same supplied comparison.

These figures must be interpreted carefully. The supplied material does not provide the dataset name, number of images, class distribution, train-validation-test split, image resolution, model configurations, training epochs, random seeds, or whether all models were evaluated on exactly the same test set. Therefore, the table is reported as project-supplied comparative values only. A fair experimental comparison requires training or evaluating every model on the same dataset split with identical preprocessing and documented evaluation code. The unusually high proposed-model values should be independently reproduced before being presented as validated experimental findings.

XV. EVALUATION METRICS

Overall Accuracy (OA) is the proportion of correctly classified pixels among all valid pixels. It provides a global measure but can be dominated by large classes when the dataset is imbalanced.

Intersection over Union (IoU) measures the overlap between predicted and ground-truth pixels for a class. Mean IoU is calculated by averaging IoU across the evaluated classes and is useful for segmentation tasks with multiple land-cover categories.

Precision measures the proportion of predicted pixels for a class that are correct, while recall measures the proportion of ground-truth pixels of that class that are detected. The F1-score is the harmonic mean of precision and recall. Macro F1 calculates the F1-score independently for each class and averages the class scores, giving each class equal weight.

For reproducible evaluation, the project should report per-class precision, recall, F1-score, and IoU in addition to OA, mIoU, and Macro F1. A confusion matrix can identify class pairs that are frequently confused. Boundary quality can be assessed using a boundary-specific metric if ground-truth boundary annotations are available.

XIII. EXPERIMENTAL CONSIDERATIONS AND ERROR ANALYSIS

The three reported measures in the supplied comparison describe different aspects of segmentation performance. Overall Accuracy summarizes the proportion of correctly labelled pixels across the map. Mean Intersection over Union gives each evaluated class a contribution to the average overlap score, while Macro F1 averages class-level F1 values and therefore provides a class-balanced view when calculated consistently. Reporting all three measures is useful because a model can achieve high overall accuracy while still performing poorly on a small or visually similar class.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

A class-wise confusion matrix should be examined to identify systematic confusion between categories. For example, errors may occur between barren land and built-up surfaces, or between vegetation and agricultural land, depending on season, spatial resolution, and the class definitions used by the annotation scheme. These are potential error categories to inspect, not confirmed errors in the current project. The supplied presentation does not provide a confusion matrix or per-class results, so the actual error distribution cannot yet be established.

Boundary quality should also be assessed separately from pixel-level aggregate metrics. A segmentation map may achieve a strong global score while producing jagged edges, small gaps, or inaccurate outlines around narrow objects. Visual inspection of representative tiles, class-wise IoU, and a boundary-sensitive measure can provide complementary evidence. Any such evaluation should use reference boundaries or masks created independently of the model prediction.

The comparative values reproduced in this paper show the proposed method at 99.18% OA, 97.88% mIoU, and 98.92% Macro F1. These are values transcribed from the supplied project figure. Since the figure does not state the dataset, split, sample count, class distribution, or common evaluation protocol for all listed methods, the numbers should be treated as reported comparison values rather than proof of a controlled benchmark advantage. A controlled experiment using identical data partitions is required before making a definitive performance claim.

Recommended Ablation and Robustness Checks

An ablation experiment can measure the contribution of each major component. The baseline may use a conventional CNN segmentation backbone; subsequent variants can add dilated convolution, channel and spatial attention, and LSTM temporal modelling one component at a time. The complete architecture can then be compared with these variants using the same training and test data. This design helps determine whether observed gains are associated with multi-scale context, attention, temporal information, or interactions between the components.

Robustness checks can evaluate the model across different acquisition dates, geographic subregions, and class distributions. If the dataset permits, a geographically separate test region can help assess whether the learned representation transfers beyond the training locations. Performance should also be examined for small classes and narrow structures, which may be underrepresented in the training samples. These checks are recommendations for the project evaluation; they are not experiments reported as completed in the supplied presentation.

XVI. RESULTS AND DISCUSSION

The project presentation describes the intended output as a pixel-wise LULC map with clear boundaries, class-wise area estimates, percentage distribution, and optional year-to-year change analysis. The architecture also provides charts and tabular summaries to support interpretation of the final map. These are the expected outputs of the proposed workflow; actual maps and quantitative results should be generated from the implemented model and the documented test set.

The reported comparative chart suggests that the proposed model performs strongly on the three supplied metrics. A plausible architectural explanation is that dilated convolutions provide multi-scale spatial context, attention highlights relevant feature channels and spatial regions, and LSTM captures temporal dependencies when multi-date imagery is supplied. However, the current project material does not include ablation experiments. Consequently, the separate contribution of dilated convolution, channel attention, spatial attention, LSTM, and post-processing cannot yet be quantified.

A suitable ablation study should compare a baseline CNN, CNN with dilated blocks, CNN with attention, CNN with LSTM, and the complete proposed architecture. This would establish whether each component contributes measurable improvement. The study should also report computational cost, parameter count, inference time, and performance by class, especially for small or under-represented classes.

XVII. APPLICATIONS

Urban and regional planning through identification of built-up and expanding urban regions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Agricultural monitoring through mapping of cultivated areas and changes in agricultural land.
Forest and vegetation monitoring for land-resource management and environmental assessment.
Water-body and barren-land mapping for watershed and land-management studies.
Temporal land-cover change assessment when properly aligned multi-date satellite imagery is available.

XVIII. LIMITATIONS AND FUTURE SCOPE

The proposed framework depends on the quality, resolution, and consistency of satellite imagery and labelled training masks. Cloud cover, seasonal variation, class imbalance, mixed pixels, and geographic domain shifts can affect segmentation quality. LSTM-based temporal analysis requires multiple aligned observations; it should not be described as year-to-year change detection when only single-date imagery is used. The current supplied material also does not document the dataset, implementation details, or the experimental protocol behind the comparison chart.

Future work can evaluate the model on publicly available datasets such as EuroSAT, ISPRS Potsdam, ISPRS Vaihingen, LoveDA, or other suitable land-cover benchmarks, subject to matching class definitions and input requirements. Cross-region evaluation can test generalization to unseen areas. Additional work may investigate class-balanced losses, boundary-aware training, uncertainty estimation, semi-supervised learning for limited labels, and lightweight model variants for faster inference. Temporal evaluation can be expanded using more acquisition dates and consistent change labels.

The proposed architecture may also be extended to fuse optical and SAR data, include transformer-based feature extractors, or provide interactive GIS visualizations. Any extension should be evaluated with clearly documented datasets, training procedures, and baseline comparisons.

XIX. CONCLUSION

This paper presents a proposed LULC mapping framework that integrates dilated convolution, channel and spatial attention, and LSTM-based temporal modelling. The workflow includes satellite-image acquisition, radiometric correction and normalization, noise handling, augmentation, patch extraction, multi-scale feature extraction, attention, temporal learning, pixel-wise classification, post-processing, boundary detection, and class-wise area reporting. The supplied architecture and comparison figures have been incorporated into the paper. The supplied comparison reports OA of 99.18%, mIoU of 97.88%, and Macro F1 of 98.92% for the proposed model; these values require independent validation under a documented and controlled experimental protocol. The integrated framework is intended to support detailed land-cover mapping and, where multi-temporal data are available, land-cover change analysis.

REFERENCES

[1] N. Zang, Y. Cao, Y. Wang, B. Huang, L. Zhang, and P. T. Mathiopoulos, “Land-Use Mapping for High-Spatial Resolution Remote Sensing Image Via Deep Learning: A Review,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 5372–5391, 2021. https://doi.org/10.1109/JSTARS.2021.3078631
[2] Q. Liu, M. Kampffmeyer, R. Jenssen, and A.-B. Salberg, “Dense Dilated Convolutions’ Merging Network for Land Cover Classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 58, pp. 6309–6320, 2020. https://doi.org/10.1109/TGRS.2020.2976658
[3] L. Liu, S. Saypadith, I. Taniguchi, J. Zhou, H. Nishikawa, and T. Onoye, “Lightweight Remote Sensing Image Segmentation Network With Dilated Convolution and Data-Dependent Attention Decoder,” IEEE Access, 2025. https://doi.org/10.1109/ACCESS.2025.3631504
[4] “GLSANet: Global-Local Self-Attention Network for Remote Sensing Image Semantic Segmentation,” IEEE Geoscience and Remote Sensing Letters, vol. 20, 2023. https://doi.org/10.1109/LGRS.2023.3235117
[5] “Multiscale Attention Network Guided With Change Gradient Image for Land Cover Change Detection Using Remote Sensing Images,” IEEE Geoscience and Remote Sensing Letters, vol. 20, 2023. https://doi.org/10.1109/LGRS.2023.3267879

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, November 2026 Impact Factor: 8.2

[6] "Remote Sensing Semantic Segmentation via Boundary Supervision-Aided Multiscale Channelwise Cross Attention Network," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, 2023. https://doi.org/10.1109/TGRS.2023.3292112
[7] "CNN, RNN, or ViT? An Evaluation of Different Deep Learning Architectures for Spatio-Temporal Representation of Sentinel Time Series," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 44–56, 2023. https://doi.org/10.1109/JSTARS.2022.3219816
[8] L. Russo, A. Sorriso, S. L. Ullo, and P. Gamba, "A Deep Learning Architecture for Land Cover Mapping Using Spatio-Temporal Sentinel-1 Features," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 10562–10581, 2025. https://doi.org/10.1109/JSTARS.2025.3557687
[9] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land Cover Classification via Multitemporal Spatial Data by Deep Recurrent Neural Networks," IEEE Geoscience and Remote Sensing Letters, 2017. https://doi.org/10.1109/LGRS.2017.2728698
[10] "Self-Supervised Feature Learning for Multimodal Remote Sensing Image Land Cover Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, 2022. https://doi.org/10.1109/TGRS.2022.3190466
[11] X. Li et al., "Collaborative Attention-Based Heterogeneous Gated Fusion Network for Land Cover Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 59, pp. 3829–3845, 2021. https://doi.org/10.1109/TGRS.2020.3015389
[12] X. Qi et al., "High-Resolution Land-Cover Mapping Based on a Cross-Resolution Deep Learning Framework and Available Low-Resolution Labels," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024. https://doi.org/10.1109/JSTARS.2023.3342994
[13] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 7, pp. 2217–2226, 2019. https://doi.org/10.1109/JSTARS.2019.2918242
[14] "Real-Time Semantic Segmentation: A Brief Survey and Comparative Study in Remote Sensing," IEEE Geoscience and Remote Sensing Magazine, vol. 11, no. 4, pp. 94–124, 2023. https://doi.org/10.1109/MGRS.2023.3321258
[15] "Temporal-Domain Adaptation for Satellite Image Time-Series Land-Cover Mapping With Adversarial Learning and Spatially Aware Self-Training," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 3645–3675, 2023. https://doi.org/10.1109/JSTARS.2023.3263755
[16] "Cross-City Semantic Segmentation (C2Seg) in Multimodal Remote Sensing: Outcome of the 2023 IEEE WHISPERS C2Seg Challenge," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 8851–8862, 2024. https://doi.org/10.1109/JSTARS.2024.3388464
[17] "Hierarchical Shared Architecture Search for Real-Time Semantic Segmentation of Remote Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, 2024. https://doi.org/10.1109/TGRS.2024.3373493
[18] "Land Cover Change Detection Based on Vector Polygons and Deep Learning With High-Resolution Remote Sensing Images," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, 2024. https://doi.org/10.1109/TGRS.2023.3346968
[19] "RRSIS: Referring Remote Sensing Image Segmentation," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, 2024. https://doi.org/10.1109/TGRS.2024.3369720
[20] "Deep Spatiotemporal Subpixel Mapping Network by Integrating a Prior Fine Land Cover Map With Change Detection," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, 2025. https://doi.org/10.1109/TGRS.2025.3578351
[21] "Multilevel Multimodal CNN–Transformer Fusion Network for Land Use and Land Cover Classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 63, 2025. https://doi.org/10.1109/TGRS.2025.3609638

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

[22] "CCD-TBLCA: A Land Cover Continuous Change Detection Method Fusing Transformer and Bi-Directional LSTM via Cross-Attention," IEEE Transactions on Geoscience and Remote Sensing, vol. 64, 2026. https://doi.org/10.1109/TGRS.2026.3665051