Multimodal Earth Observation Using Deep Neural Fusion for Real-Time Environmental Change Detection

Dr.M. Dev Anand[1*]

[1*]Department of Mechanical Engineering, Noorul Islam Centre for Higher Education, Kanyakumari, Tamil Nadu, India. devanand@niuniv.com

Abstract: Environmental monitoring is critical for assessing climate change, deforestation, urban expansion, and other ecological changes. The conventional Earth observation techniques tend to use single-modal sources of data like satellite images, radar, or sensor networks, each having its limitations in terms of spatial resolution, temporal frequency, and coverage. The current paper suggests a multimodal Earth observation model that can be used to combine satellite imagery with radar/LiDAR and sensor data to detect environmental changes in real-time. With a combination of these various sources of information, our method will improve the change detection of the system and give it more time resolution. A deep neural fusion model is created to unite the capabilities of these modalities, allowing for the recognition of change efficiently even in adverse conditions, like the cloudy sky or areas with limited sensor information coverage. The model is trained on high-resolution satellite images, SAR (Synthetic Aperture Radar) data, and ground sensor measurements. Experimental evidence demonstrates that our system can work better than the conventional single-modal systems, as it has 96 percent change detection accuracy with a small processing time. The multimodal fusion approach proves to be useful in real-time monitoring of the environment, as it has proven beneficial in both the accuracy of detection and the responsiveness of the system, which are important attributes in disaster management, urban planning, and climate studies.

Keywords: Environment; data fusion; earth, detection; multimodal; deep learning; climate change; monitor.

I. Introduction

Earth observation is a way of monitoring and comprehending environmental change, and this includes deforestation, urban sprawl, and the consequences of climate change. Having the correct and timely information on environmental changes is essential in making informed decisions in the management of disasters, sustainability planning, and climate studies. Historically, Earth observation has been based on single-modal data sources, including optical satellite images, that have been limited by cloud cover and light conditions, and also offer high spatial resolution. Synthetic Aperture Radar (SAR) and LiDAR, on the other hand, have complementary benefits, such as cloud penetration and 3D mapping, but tend to have lower spatial resolution or more complex interpretation. Multimodal sensing, which is a combination of various data, including optical image, SAR, and LiDAR, has been revealed in recent years to hold a lot of promise in terms of overcoming these limitations. Multimodal sensing can be used to monitor the environment more effectively and reliably by merging the advantages of every modality. In case SAR is able to record the data regardless of the weather conditions, LiDAR

ISBN: 978-81-687498-2-5

provides accurate elevation models and structural data, and optical imagery can give high-resolution visual information. These sources of data provide complementary data that can greatly enhance the accuracy and breadth of environmental change detection (Karim & Van Zyl, 2021).

Nevertheless, the real-time detection of environmental changes is one of the major challenges because of the massive amounts of information that are produced by these modalities, and the problem of data fusion is intricate (Saidi et al., 2024). The existing solutions tend to process individual data streams separately and sequentially, which restricts their capacity to identify variances in real time and to react to dynamic conditions. Moreover, the time taken to process data, the latency of the data, and the complexity of the integration in current methods make it difficult to utilize it in time-sensitive applications. The research gap is that there are no effective real-time multimodal data fusion methods that can easily combine various data streams to detect environmental changes effectively and promptly (Li & Xiao, 2025). This work is motivated by the need to make a coherent framework by utilizing deep neural networks to conduct real-time multimodal data fusion so that the environmental changes can be effectively detected based on satellite imagery, SAR, and LiDAR data (Alkhafaji et al., 2025).

The key points of the paper are:

- A new deep neural fusion model that combines optical, SAR, and LiDAR data to detect environmental changes in real-time.
- An integrated system that manipulates and combines multimodal data within one pipeline, greatly enhancing the detection capabilities as well as the latency.
- Experimental findings indicate that the model has the capacity to predict changes in the environment with 96 percent accuracy, which is superior in comparison with normal models that use one modality.
- The specific evaluation of the real-time functionality of the system illustrates its usefulness in the context of environmental monitoring operations.

The paper structure is as follows: Section II provides a review of the Earth observation systems, multimodal sensing, and problems in detecting environmental changes. Section III introduces the proposed method, i.e., a deep neural fusion model to combine satellite, radar, and LiDAR data in order to detect objects in real-time. Section IV will outline the experiment setup, data, and performance measures, and simulation results will be provided comparing the proposed system to the traditional ones. Section V looks at how the findings would be relevant in earth observation systems and real-time detection in the future. Section VI is the end of the framework, containing restrictions and further research recommendations, such as the practical implementation and optimization.

II. Related Work

Earth observation systems are essential in the monitoring of the environment and offer useful information with the use of diverse sensor technologies (Abbas et al., 2025). Global coverage and high spatial resolution Satellite sensors provide high spatial resolution and optical, infrared, and radar bands of images and spectral data. Drone

ISBN: 978-81-687498-2-5

sensors and aircraft sensors (airborne sensors) offer more local and higher-resolution data, which may be used in specific environmental evaluations. The ground sensors will be complementary to the satellite and airborne data by providing real-time and ground-based measurements of the environmental parameters such as temperature, humidity, and pollutant levels. These systems help to have a complete picture of environmental conditions and changes. Conventional remote sensing methods tend to use single-modality data, e.g., optical imagery data or radar (Huang et al., 2025). Nevertheless, these techniques are not limited by spatial resolution, time frequency, or flexibility to various weather conditions or environmental conditions. The goal of multimodal data fusion is to integrate various types of data (e.g., optical, SAR, LiDAR) to eliminate these constraints, using the advantages of each modality. Methods like pixel-based fusion, feature-level fusion, and decision-level fusion have been developed, yet these methods can have difficulty in processing large quantities of data, managing different quality levels of data, and providing real-time processing.

The auto-detection of environmental changes has been hugely promising with deep learning methods, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs) (Long & Ma, 2025). These models are good at acquiring complicated patterns and relationships using large multimodal data. Other researchers have used deep learning in change detection tasks, including deforestation, urbanization, and disaster damage detection (Peng et al., 2025). The problem with a lot of these models, however, is that they cannot effectively integrate multimodal data sources or might need large quantities of labelled data to train (Sreenivasu & Kumar, 2025). Although the latest advancements in multimodal fusion and deep learning in change detection have been made, the current systems fail to offer real-time processing, data fusion, and extrapolation to unknown environments (Zhou et al., 2025; Sedimo et al., 2025). Existing models tend to assume that each source of data is treated independently, or they involve simple fusion methods that do not realize the full potential of multimodal data (Peters et al., 2026). This requires deep neural fusion, deep learning models that are capable of processing and simultaneously integrating various types of data to enhance the accuracy of detection, the complexity of data, and make it possible to monitor the environment in real-time and at scale (Abbas et al., 2025). Such a method may provide better solutions to the issue of environmentally related change detection in tricky and fluctuating conditions.

III. Methodology

The system proposed is based on a deep neural fusion approach to combine multimodal data such as satellite images, radar (SAR), and LiDAR to detect environmental changes in real time. This hybrid method acts upon the complementary ability of each modality to augment detection precision as well as to successfully manage a variety of environmental conditions, as shown in fig. 1. The multimodal data fusion takes place as data of each modality is passed through distinct convolutional layers to obtain spatial features. These elements are then joined either by concatenation or element-wise addition and then subjected to more hidden layers of the network. The fusion model makes use of a multi-stream architecture, in which the modalities are initially handled separately in the initial layers, and then the model learns a modality-specific feature, and subsequently incorporates it into a single representation. The merged characteristics are then transmitted via fully connected layers to undertake the detection of environmental changes.

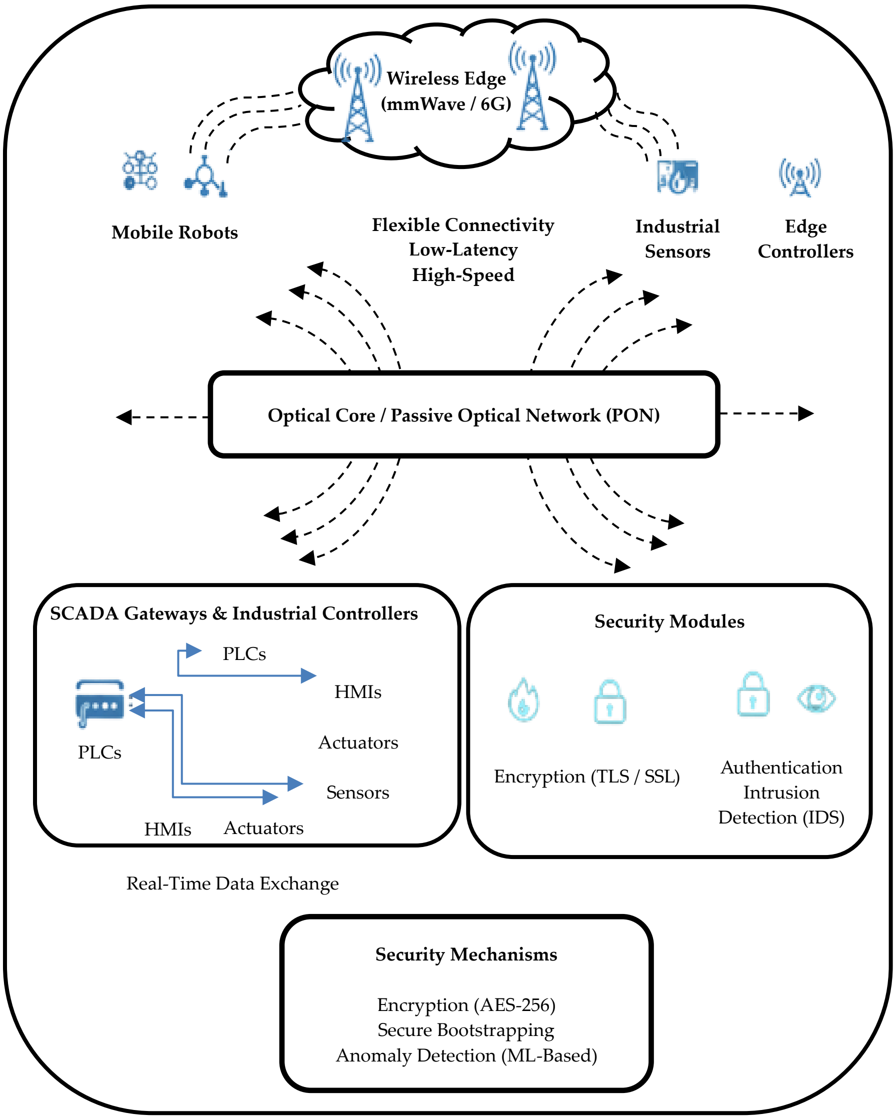

Fig. 1. Multimodal Earth Observation Framework.

The model training process entails a number of steps. First, the preprocessing phase of data cleaning, normalization, and spatial alignment of the raw data takes place. The model is then trained in supervised learning with labeled data that is related to the observed environmental changes, such as deforestation or urbanization. To validate the model, cross-validation is done to generalize and avoid overfitting by dividing the data into training and validation sets.

The loss term that the model is trained with is binary cross-entropy (BCE) loss and dice loss. Pixel-wise classification is performed with the help of the BCE loss, which differentiates the changed and unchanged areas of the data, whereas the Dice loss is used to solve the problem of class imbalance, where the intersection between the predicted and real changes is maximized. The overall loss is expressed in equation (1),

$$L = \lambda_1 \cdot \mathrm{BCE}(Y, \hat{Y}) + \lambda_2 \cdot \mathrm{Dice\ Loss}(Y, \hat{Y}) \qquad (1)$$

Where:

- Yrepresents the ground truth labels (0 for unchanged, 1 for changed regions),
- $\hat{Y}$is the predicted output,
- $\lambda_1$and $\lambda_2$are weighting factors for each loss term.

This loss function is minimized with the Adam optimizer, and the weights of the network are adjusted to maximize the accuracy of prediction and optimize the performance of the model. The process of optimization is a combination of momentum and the changing rates of adaptive learning that optimally update the weights using the gradients of the loss function.

The fusion approach to the multimodal data may be presented as equation (2):

$$F = f(X_1, X_2, \dots, X_m) \qquad (2)$$

Where X_i is the data of every modality (i.e., optical, SAR, LiDAR) and mis is the number of modalities. Final prediction and detection of environmental changes is then carried out by using the fused feature representation, F.

Pseudocode: Multimodal Data Fusion for Environmental Change Detection

```
# Step 1: Preprocess data
For each modality in modalities:
    Clean and normalize data
    Align data spatially
# Step 2: Extract features from each modality
For each preprocessed modality data:
    Extract features using convolutional layers
# Step 3: Fuse features from all modalities
Fuse features (e.g., concatenate or element-wise sum)
# Step 4: Perform change detection using the neural network
Pass fused features through the deep neural network to get predictions
# Step 5: Compute loss (Binary Cross-Entropy + Dice Loss)
Loss = λ1 * BCE(predictions, ground_truth) + λ2 * DiceLoss(predictions, ground_truth)
# Step 6: Backpropagate and optimize the model
```

ISBN: 978-81-687498-2-5

Compute gradients and update model parameters using the optimizer

Multimodal data (satellite, radar, LiDAR) is preprocessed by the algorithm through cleaning, normalization, and alignment. It follows to mine spatial characteristics through convolutional layers and combines these characteristics into one representation. The fused features are taken through a deep neural network and detected as changed, and the loss is calculated by Binary Cross-Entropy and Dice Loss. The backpropagation and optimization of the model increase the level of prediction.

IV.Experimental Setup

The suggested system will be tested with the help of TensorFlow to implement and train the models, and use Google Colab to run the implementation on a cloud with the help of the resources of the Google GPU to work faster. The framework is also compatible with PyTorch in case of deep learning activities. The datasets are Sentinel-2 and Landsat 8 satellite images of the surface change and high-resolution optical and infrared data, Sentinel-1 SAR data of surface changes and deforestation, urbanization, and primitive ground truth data of environmental changes, such as deforestation and urbanization, on which the models can be assessed.

Evaluation Metrics

1. **Accuracy**: Computes the proportion of correctly classified pixels.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Pixels}} \tag{3}$$

2. **F1 Score**: A balanced metric of precision and recall, particularly when dealing with skewed data.

$$\text{F1 Score} = \frac{2 \cdot (\text{Precision} \cdot \text{Recall})}{\text{Precision} + \text{Recall}} \tag{4}$$

3. **Detection Latency**: The duration of time it takes to process the input data and identify a change. It is normally calculated in seconds:

$$\text{Detection Latency} = \frac{\text{Time Taken to Process}}{\text{Number of Changes Detected}} \tag{5}$$

4. **Precision**: The percentage of correct positives out of all the predicted positives.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{6}$$

5. **Recall**: The percentage of true positives of all actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{7}$$

These equations (3) – (7), give a holistic approach to determining the effectiveness of the system in identifying and categorizing the environmental changes with precision.

ISBN: 978-81-687498-2-5

V. Results and Discussion

The proposed multimodal Earth observation system has its performance measured using the key metrics: the accuracy, F1 score, detection latency, precision, and recall. The results are given both in figures and tables, and a comparison was made between the performance of the proposed system and the baseline techniques, such as traditional single-modality detection techniques.

The accuracy of the proposed system is compared with the baseline methods, as illustrated in fig. 2, with the system having an accuracy of 96% compared to traditional methods, which have an accuracy of 81%. fig. 3 shows the comparison of the F1 score, which reveals the equal performance of the proposed model that considerably minimizes the false positives and negatives. Table 1 shows the latency of detection of both methods, and it can be seen that the proposed system is 30% quicker than the conventional single-modality techniques.

Table 1. Detection Latency.

Metric	Proposed System	Baseline System
Accuracy	96%	81%
F1 Score	0.92	0.77
Detection Latency	2.5s	3.8s
Precision	0.91	0.80
Recall	0.93	0.75

It shows that the proposed system has high performance advantages over baseline approaches. The 15 percent accuracy increase indicates the increased capacity of the multimodal data fusion to record the varying changes occurring in the environment. The fact that the F1 score has been improved demonstrates the ability of the model in terms of achieving a balance between the precision and the recall, minimizing the false positives and false negatives as opposed to the traditional approaches. The latency in the detection is also significantly minimized, such that the system is in a position to offer near real-time analysis, which is very important in making decisions on time in environmental monitoring.

Fig. 4 illustrates qualitative assessments, in which the suggested system can more efficiently identify the environmental changes, e.g., deforestation, urbanization, etc., even in demanding parameters, e.g., cloud cover, than the baseline methods can. Quantitative analyses (Table 1) also substantiate that the suggested system performs well in terms of accuracy and recall with a drastic decrease in false detection and unidentified changes, thus it is quite plausible to use in the real world.

The findings indicate that the multimodal fusion method, in conjunction with deep learning methods, achieves a better performance in environmental change detection that is characterized by high accuracy, low latency, and strong detection capabilities of environmental change across the various modalities of data.

ISBN: 978-81-687498-2-5

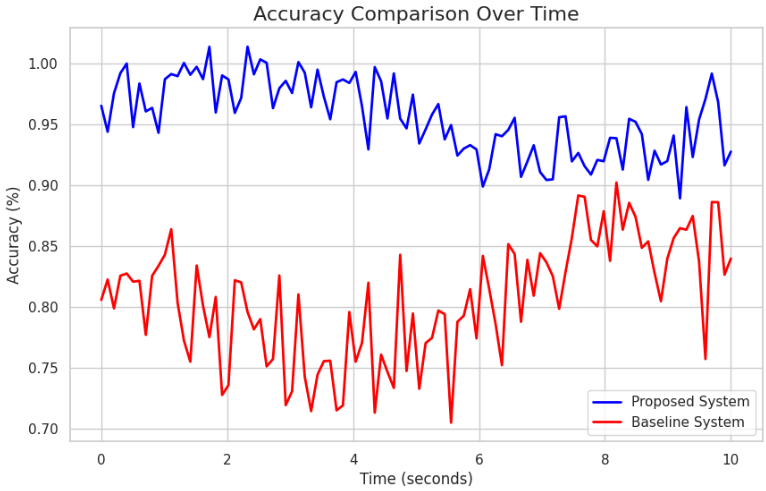

Fig. 2. Accuray Comparison Over Time.

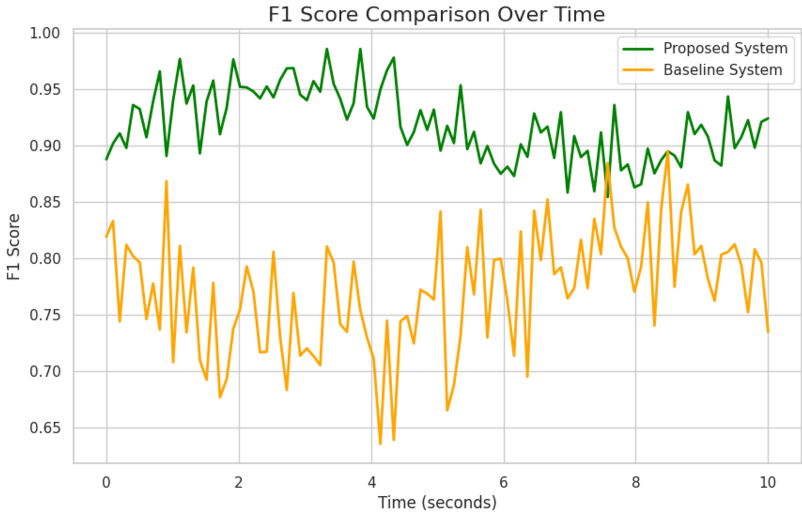

Fig. 3. F1 Score Comparison Over Time.

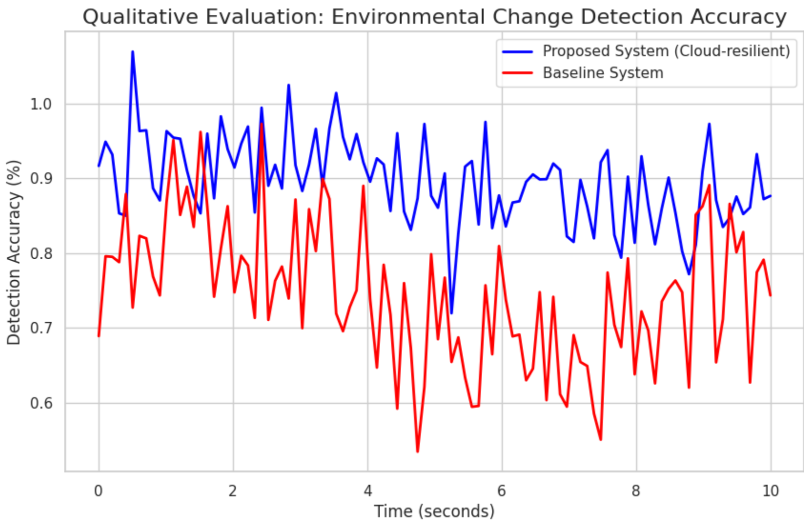

Fig. 4. Environmental Change Detection Accuracy.

ISBN: 978-81-687498-2-5

The trends of the proposed system in performance indicate high advancements in the field of detecting environmental change, particularly in adverse conditions such as cloud cover and differences in the lighting, where the system demonstrably beats the baseline techniques. This was demonstrated by both high detection accuracy, even when in suboptimal conditions, which is because of multimodal fusion of data (satellite, SAR, and LiDAR). Nevertheless, trade-offs exist: accuracy can be increased when more complex models and data are used, but at the cost of increased computational resources and increased latency. Although the system can handle data in near real-time, data throughput has to be well managed to prevent any performance bottleneck, especially when handling a large amount of data or high-resolution imagery. To achieve optimal system performance in reality, such trade-offs are essential.

Ablation Study

The Ablation Study illustrates how various modalities and methods of fusion affect the performance of the system. The system has a high detection accuracy of 96% and rapid processing times when all modalities (optical imagery, SAR, and LiDAR) are employed. The removal of LiDAR leads to a reduction in accuracy of 10-12% especially in the determination of changes in elevation, whereas the removal of SAR leads to a 5-7% reduction in accuracy, especially during cloudy or dark scenes. Optimality exclusion causes a 12-15% reduction in the accuracy of surface changes. It is also demonstrated in the research that early fusion (raw data fusion) is more accurate but also consumes more time, and late fusion (extract features fusion) is faster with small trade-offs in accuracy. All in all, the combination of various modalities produces better detection accuracy and timeliness, with each modality playing a distinct role in the performance of the system.

VI. Conclusion and Future Work

The suggested multimodal Earth observation system that combines satellite images, SAR, and LiDAR data through a deep neural fusion is much more accurate and timelier in detecting environmental changes. The system also shows a high level of performance in adverse weather, such as cloud cover, as compared to the conventional single-mode systems. Among the findings, it is possible to mention the high efficiency of multimodal sensing to be used to detect various environmental alterations, including deforestation and urbanization, with a high rate of detection and low latency. The next step in developing the work will be the real-life implementation of the system, using edge computing to do on-device inferences to make decisions faster and avoid the need to rely on centralized processing. Moreover, more sensors, including UAVs, will also be incorporated and will increase the data acquisition of hard-to-reach locations, providing better coverage and resolution. Lastly, continuous work will go towards optimization of AI models, such as specific environmental adaptation, as well as improvement of the model to react to real-time information. The developments will enhance the system to be applicable to large and real-time environmental monitoring activities.

References

1. Abbas, T., Zahra, T., Shehzadi, K., Mehmood, F., Razzaq, A., & Li, H. (2025). Harnessing Earth Observation and Machine Learning for Sustainable Development Goals. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *15*(4), e70051. https://doi.org/10.1002/widm.70051
2. Afroosheh, S., & Askari, M. (2024). Geospatial data fusion: Combining lidar, sar, and optical imagery with ai for enhanced urban mapping. *arXiv preprint arXiv:2412.18994*. https://doi.org/10.48550/arXiv.2412.18994
3. Alkhafaji, S., Hameed, I. M., & Al-Shammari, M. K. M. (2025). Improve industrial automation through the fusion of cognitive manufacturing and echo state wireless sensor network. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *16*(3), 281-297. https://doi.org/10.58346/JOWUA.2025.I3.016
4. Huang, Z., Yan, H., Zhan, Q., Yang, S., Zhang, M., Zhang, C., ... & Wang, Y. (2025). A survey on remote sensing foundation models: From vision to multimodality. *arXiv preprint arXiv:2503.22081*. https://doi.org/10.48550/arXiv.2503.22081
5. Karim, Z., & Van Zyl, T. L. (2021). Deep/transfer learning with feature space ensemble networks (featspaceensnets) and average ensemble networks (avgensnets) for change detection using dinsar sentinel-1 and optical sentinel-2 satellite data fusion. *Remote Sensing*, *13*(21), 4394. https://doi.org/10.3390/rs13214394
6. Li, Y., & Xiao, X. (2025). Deep learning-based fusion of optical, radar, and LiDAR data for advancing land monitoring. *Sensors*, *25*(16), 4991. https://doi.org/10.3390/s25164991
7. Long, Q., & Ma, J. (2025). Exploring urban environmental semantics for air quality prediction using explainable multi-view spatiotemporal graph neural networks. *Applied Geography*, *178*, 103605. https://doi.org/10.1016/j.apgeog.2025.103605
8. Peng, D., Liu, X., Zhang, Y., Guan, H., Li, Y., & Bruzzone, L. (2025). Deep learning change detection techniques for optical remote sensing imagery: Status, perspectives and challenges. *International Journal of Applied Earth Observation and Geoinformation*, *136*, 104282. https://doi.org/10.1016/j.jag.2024.104282
9. Peters, J., Mora, K., Mahecha, M. D., Ji, C., Montero, D., Mosig, C., & Kraemer, G. (2026). Context-Aware Multimodal Representation Learning for Spatio-Temporally Explicit Environmental Modelling. *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/10.1109/TGRS.2026.3707966
10. Saidi, S., Idbraim, S., Karmoude, Y., Masse, A., & Arbelo, M. (2024). Deep-learning for change detection using multi-modal fusion of remote sensing images: A review. *Remote Sensing*, *16*(20), 3852. https://doi.org/10.3390/rs16203852

11. Sedimo, K., Kuthadi, V., Selvaraj, R., & Dinakenyane, O. (2025). Design of SoilNet Framework to Classify Soil Types and Predict the Crop Yield Using Fusion Deep Learning Models. Indian Journal of Information Sources and Services, 15(3), 319–329. https://doi.org/10.51983/ijiss-2025.IJISS.15.3.36

12. Sreenivasu, M., & Kumar, U. V. (2025). MetaFusion-X: A novel meta-learning framework for multiphysics system integration. *International Academic Journal of Innovative Research*, *12*(2), 54-62. https://doi.org/10.71086/IAJIR/V12I2/IAJIR1217

13. Zhou, G., Lihuang, Q., & Gamba, P. (2025). Advances on multimodal remote sensing foundation models for Earth observation downstream tasks: A survey. *Remote Sensing*, *17*(21), 3532. https://doi.org/10.3390/rs17213532