© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

An Integrated Deep Learning Framework Using U-Net++ And Hybrid Loss for Class-Imbalanced Semantic Segmentation in Satellite Imagery

Akshat Yadav[1], Varsha Dang[2], Mihir Mohite[3]
[1,2,3]*Department of CSE/AI/ML, Vishwakarma Institute of Technology Pune, India*

***Abstract*—Accurate semantic segmentation of satellite imagery is essential for environmental monitoring, agricultural planning, water-resource assessment, and climate-driven land-use analysis. However, a major challenge persists in Deep Learning–based Earth Observation: class imbalance, where small but environmentally critical classes (e.g., water bodies, wetlands, deforested patches) appear far less frequently than dominant classes such as agricultural or barren land. Traditional U-Net models trained with Cross-Entropy loss tend to ignore such minority classes, resulting in low recall, imprecise boundaries, and unreliable predictions for rare features.**

This research proposes a comprehensive deep-learning framework integrating U-Net++ architecture, class-aware sampling, targeted augmentation, and a three-part hybrid loss function combining Area-Weighted Binary Cross Entropy (AWBCE), Tversky Loss, and Boundary Loss. Training stability is further enhanced using the Ranger optimizer, which combines RAdam and Lookahead to better navigate complex loss landscapes. Experiments conducted on LISS-IV satellite imagery demonstrate substantial improvement in overall accuracy, mean IoU, and—most importantly—minority-class IoU. Water-body IoU improves by more than 23%, and the proposed method outperforms standard U-Net by 7.3% mean IoU.

The results validate that addressing class imbalance requires multi-level redesigns across data, architecture, optimization, and loss formulation.

***Index Terms*—Earth Observation, semantic segmentation, U-Net++, hybrid loss, minority-class detection, remote sensing, deep learning.**

I. INTRODUCTION

The increasing reliance on satellite imagery for environmental analysis, agricultural forecasting, natural resource management, and urban planning has led to a heightened demand for accurate semantic segmentation models capable of extracting pixel-level land-cover information. Deep learning architectures, especially convolutional neural networks (CNNs), have proven highly effective for vision-based segmentation tasks. U-Net and its derivatives have achieved remarkable success in biomedical and geospatial applications due to their encoder–decoder structures and skip connections, which preserve spatial continuity and enable multi-scale feature representation. However, remote-sensing datasets pose a unique challenge that greatly diminishes the performance of conventional models: severe class imbalance. Majority classes such as agriculture, forest cover, and barren land dominate satellite images, whereas minority classes—often the most environmentally significant—appear in small, fragmented, or thin structures. These minority regions include narrow rivers, ponds, wetlands, and early-stage deforested sites that are crucial indicators for policy making and environmental health.

Traditional deep learning systems struggle under such imbalance because the abundance of dominant classes biases the optimization process, causing gradients derived from rare classes to become negligible. As a result, the model tends to maximize overall accuracy by predicting majority classes at the expense of minority-class recall. Although the model may report seemingly high overall accuracy, it performs poorly in detecting classes that contain essential environmental information. This failure is problematic for applications in flood prediction, water-resource mapping, and biodiversity conservation, where accurate delineation of small features directly influences decision-making. To overcome these limitations, there is a need for segmentation frameworks explicitly designed to handle imbalance through improvements at the architectural, loss-function, and data-processing levels.

The present study introduces such a framework, constructed around the U-Net++ architecture, which uses nested skip pathways to reduce the semantic gap between encoder and decoder representations. This enhanced feature fusion significantly improves the recovery of small objects that are often lost through successive down sampling operations in standard CNNs. To further reinforce minority-class learning, the model is trained with a hybrid loss combining AWBCE, Tversky, and Boundary losses, each compensating for specific limitations of traditional Cross-Entropy loss. Additionally, we introduce class-aware sampling to ensure that the training batches provide balanced exposure to minority classes. The Ranger optimizer, combining the stability of RAdam with the foresight of Lookahead, is employed to handle the complex gradient landscape created by the hybrid loss. Together, these innovations form a unified pipeline designed to effectively address class imbalance in semantic segmentation of remote-sensing data.

II. LITERATURE REVIEW

Remote-sensing image classification has undergone a dramatic evolution over the past two decades. Earlier methodologies relied heavily on handcrafted features derived from spectral indices such as NDVI, NDBI, and NDWI, combined with traditional machine-learning algorithms including Support Vector Machines, Random Forests, and K-means clustering. Although these approaches were effective in simple contexts, they lacked robustness when dealing with heterogeneous landscapes, complex spatial structures, and high intra-class variability. The advent of deep learning marked a paradigm shift, with Fully Convolutional Networks (FCNs) enabling end-to-end learning of dense pixel-wise predictions without manual feature engineering. U-Net became a widely adopted architecture for segmentation tasks due to its dual-path structure, which effectively captures both high-level semantics and detailed spatial features through skip connections.

Subsequent U-Net variants introduced several improvements. UNet++ refined the skip-connection design by inserting a series of shallow convolutional nodes between the encoder and decoder, enabling more gradual semantic alignment and improved multi-scale feature integration. This design proved advantageous in detecting small structures, making it a strong candidate for imbalanced remote-sensing tasks. Attention U-Net introduced selective focus mechanisms that automatically highlight relevant spatial regions while suppressing noise, improving segmentation in scenes with cluttered backgrounds. DeepLab series networks incorporated atrous convolutions and CRFs to enhance contextual understanding and boundary precision. These architectural advancements collectively contributed to more accurate segmentation models, yet they did not inherently resolve the issue of extreme class imbalance.

Parallel to architectural advancements, significant research has focused on loss functions tailored for imbalanced data. Dice loss and its variants prioritize region overlap, providing better performance than Cross-Entropy for small objects. Tversky loss introduces tunable parameters to adjust penalties for false positives and false negatives, making it particularly effective for recall-focused applications. Focal loss shifts attention toward hard or rare examples by down-weighting easy predictions. Boundary loss represents another category of geometric losses that emphasize proximity to ground-truth edges, resulting in sharper segmentation boundaries. Despite their individual strengths, each loss function addresses only one aspect of the imbalance problem, leaving other weaknesses unresolved.

Efforts to mitigate imbalance also include data-centric strategies such as patch-based sampling, minority-focused augmentation, and curriculum learning. While these contribute to improved training dynamics, they do not eliminate the need for architectural and loss-level interventions. The literature reveals that no singular approach sufficiently addresses the full complexity of class imbalance in satellite imagery. This gap forms the motivation for the integrated framework proposed in this research.

In addition to architectural innovations, several studies have explored hybrid optimization strategies to improve segmentation performance in complex remote-sensing environments. Researchers have combined region-based and pixel-based losses to create adaptive formulations capable of handling noisy labels and heterogeneous landscapes. For instance, recent work on hybrid Dice–Focal losses has shown improved robustness in detecting small objects across modalities such as medical imaging and aerial surveillance. These approaches reinforce the necessity of multi-objective optimization, where different loss components compensate for one another's weaknesses and collectively strengthen model learning, especially in imbalanced scenarios. However, most existing hybrid-loss studies remain tailored to domain-specific datasets and lack generalizability to satellite imagery, where spatial variability, atmospheric disturbances, and class imbalance introduce significantly greater complexity.

Another important research theme involves multi-scale context modeling. While classical CNNs offer strong local feature extraction, they struggle to capture long-range dependencies essential for understanding large geographic formations. Techniques such as Pyramid Pooling Modules, Atrous Spatial Pyramid Pooling (ASPP), and attention mechanisms have been shown to improve global context representation. These innovations help distinguish between spectrally similar but spatially distinct regions, a common challenge in land-cover segmentation. Nevertheless, these solutions often increase the computational burden, making them less suitable for large-scale or real-time applications. U-Net++, with its nested skip connections, offers a favorable balance between multi-scale representation and computational efficiency, making it well aligned with the needs of remote-sensing segmentation.

Recent research has also explored transformer-based architectures, which excel at capturing long-range spatial relationships. Vision Transformers (ViT), Swin Transformers, and hybrid ConvFormer models have demonstrated impressive performance in segmentation tasks. Although transformers outperform CNNs on large datasets, their effectiveness declines when data is limited or heavily imbalanced, as is often the case in Earth Observation datasets where labeling is expensive. This reinforces the need for frameworks like the one proposed in this work, where architectural modifications, loss balancing, and sampling strategies collectively compensate for training limitations. The literature clearly highlights that no single technique—architectural, data-driven, or loss-based—is sufficient on its own, providing strong justification for the integrated approach presented in this research.

III. METHODOLOGY

A. Dataset and Preprocessing

The dataset used in this study consists of high-resolution LISS-IV images with six land-cover categories: urban, agriculture, rangeland, forest, water bodies, and barren land. A detailed exploratory analysis indicated severe imbalance, with water bodies comprising less than five percent of the total pixels. Preprocessing involved resizing images to 256×256 pixels, normalizing with ImageNet statistics, and encoding label masks using integer class indices. To enhance generalization, extensive data augmentation was applied using the Albumentations library, including geometric and photometric transformations. Special augmentation strategies such as copy-paste were selectively applied to water-body regions to artificially increase their occurrence without compromising spatial realism.

B. Class-Aware Sampling

To ensure equitable learning across classes, a class-aware sampling mechanism was designed. Each tile in the dataset was examined to compute its class distribution, after which minority-rich tiles were grouped separately from majority-rich tiles. During batch formation, samples were drawn from both groups in controlled proportions, ensuring that minority-class examples appeared more frequently than their natural occurrence in the dataset. This method provided a balanced gradient flow without artificially duplicating images.

C. U-Net++ Architecture

The segmentation network is based on U-Net++ with a ResNet-34 encoder. The defining feature of U-Net++ is its nested skip pathways, which introduce intermediate convolutional nodes that gradually align encoder features with decoder features. This design reduces the semantic discrepancy between downsampled and upsampled representations, enabling more accurate reconstruction of small structures. The decoder uses densely connected convolutional blocks to aggregate multi-scale contextual information, and a 1×1 convolutional layer produces class-wise logits for segmentation.

D. Hybrid Loss Function

A hybrid loss was constructed to reconcile multiple imbalance factors simultaneously. AWBCE assigns inverse-area weights to minority-class regions, ensuring that their gradients are proportionally amplified. Tversky loss contributes to improved recall by penalizing false negatives more strongly, aligning well with the requirements of detecting small or thin structures. Boundary loss enhances contour alignment by minimizing discrepancies between predicted and actual object boundaries using distance maps. The combined loss provides a balanced optimization signal that improves regional accuracy, recall, and edge definition.

E. Training Configuration

Training was conducted using the Ranger optimizer, which incorporates the rectified adaptive moment estimation of RAdam and the weight-interpolation mechanism of Lookahead. This combination stabilizes training in complex loss landscapes and accelerates convergence. The model was trained for 50–100 epochs with a learning rate of 1×10^{-4} and a batch size of four, optionally using cosine annealing scheduling

IV. RESULTS AND DISCUSSION

Extensive experiments were conducted using a GPU-accelerated environment. The dataset was split into training, validation, and test subsets, ensuring no spatial overlap to prevent data leakage. Evaluation metrics included overall accuracy, macro F1-score, macro recall, mean IoU, and class-wise IoU. These metrics provided comprehensive insight into both global performance and class-specific behavior.

The proposed system demonstrated significant improvement over the baseline U-Net trained with Cross-Entropy loss. The mean IoU increased from 0.7989 to 0.8721, while macro recall improved substantially, indicating enhanced sensitivity to minority classes. The IoU for water bodies rose from 0.6234 to 0.8534, a remarkable improvement reflecting the effectiveness of the hybrid loss and architectural enhancements. Urban class IoU similarly increased from 0.7845 to 0.8912, demonstrating that the method not only benefits minority classes but also improves performance on structurally complex classes. These quantitative improvements confirm that the proposed multi-level framework effectively mitigates the bias intrinsic to imbalanced datasets.

Qualitative results further validate these findings. Visual inspection of predicted masks reveals sharper boundaries, more continuous water streams, distinct urban edges, and better separation of adjacent land-cover types. The proposed model exhibits fewer misclassifications along class boundaries, especially between forest and rangeland, which often share similar spectral characteristics. The hybrid loss plays a crucial role in enforcing precise boundary alignment, resolving ambiguities that frequently arise in remote-sensing segmentation.

Ablation studies affirm the contribution of each component in the framework. Removing AWBCE leads to diminished detection of small water bodies. Excluding Tversky loss results in lower recall across all minority classes. The absence of boundary loss yields less distinct contours and lower IoU, emphasizing its role in refining segmentation quality. Replacing Ranger with Adam degrades performance, confirming that the optimizer contributes meaningfully to stable convergence in the proposed multi-component loss scenario. These observations demonstrate that the full integration of data-level, model-level, and loss-level strategies is critical for achieving the highest performance.

Component Removed	Effect on Performance
Remove AWBCE	Small-object detection drops sharply
Remove Tversky	Recall decreases for minority classes
Remove Boundary Loss	Edges become fuzzy; IoU drops
Replace Ranger with Adam	Slower convergence, lower IoU

A. Quantitative Evaluation

The proposed pipeline is compared with Standard U-Net baselines.

Table 1 — Performance Comparison

Metric	Standard U-Net	Proposed U-Net++ + Hybrid Loss
Overall Accuracy	0.8234	0.8834
Macro F1-Score	0.8456	0.9156
Macro Recall	0.8123	0.8943
Mean IoU	0.7989	0.8721
Water IoU	0.6234	0.8534
Urban IoU	0.7845	0.8912

The proposed system improves mean IoU by 7.3% and water-body IoU by 23%, demonstrating substantial gains in minority-class recognition.

B. Class-wise IoU Scores

Land Cover	IoU
Urban	0.8912
Agriculture	0.8645
Rangeland	0.8523
Forest	0.8789
Water	0.8534
Barren	0.8923

The model produces consistently high IoU across all classes.

The results of this study highlight not only the performance improvements achieved through the proposed pipeline but also the importance of integrating multiple strategies when dealing with complex remote-sensing problems. The substantial gains observed in minority-class IoU affirm that architectural refinement alone is insufficient; rather, effective segmentation in imbalanced datasets emerges from a combination of architectural strength, loss-function design, and intelligent sampling of training data. The use of U-Net++ contributed significantly by enabling deeper multi-scale feature aggregation and reducing the semantic gap between encoder and decoder outputs, which is often responsible for the loss of fine structures. The hybrid loss served as an essential complement by aligning optimization objectives with the real-world priorities of geospatial analysis—namely, recovering small, environmentally crucial features with high precision. The Boundary loss component proved particularly effective in sharpening river edges, wetland boundaries, and urban outlines, which are often blurred in standard segmentation outputs. Meanwhile, the class-aware sampling mechanism ensured that minority classes were sufficiently represented throughout the training process, preventing gradient starvation and allowing the model to form balanced discriminative representations. These combined components produced a segmentation system that outperformed baseline approaches across nearly every metric. The findings illustrate the multifaceted nature of class imbalance and demonstrate that solutions must address its effects at multiple stages of the pipeline rather than relying on isolated enhancements.

V. PROBLEM STATEMENT

The core problem addressed in this research is the persistent failure of deep-learning-based semantic segmentation models to accurately detect and classify minority land-cover classes in high-resolution satellite imagery due to severe class imbalance. In typical Earth Observation datasets, dominant classes such as agriculture, barren land, and forest occupy the majority of pixels, while environmentally critical but spatially limited classes such as water bodies, wetlands, narrow rivers, and fragmented urban structures appear in extremely small proportions. Conventional architectures, particularly the standard U-Net trained with Cross-Entropy loss and optimizers like Adam, become strongly biased toward majority classes, leading the network to minimize overall error by overlooking minority regions. This results in poor recall, weak boundary localization, and a significant degradation in the Intersection over Union (IoU) for rare classes, despite achieving deceptively high overall accuracy. The challenge is further amplified by the thin, irregular, and fragmented geometry of minority classes, which is difficult to capture through standard encoder–decoder pathways that suffer from semantic gaps during upsampling. Therefore, the main problem is to design a segmentation framework that can reliably detect small and rare land-cover structures without being overshadowed by the overwhelming presence of dominant classes, and to achieve this without compromising majority-class accuracy. This requires a holistic solution that addresses imbalance at the data level, improves multi-scale feature fusion at the architectural level, and incorporates loss functions that explicitly emphasize minority-class representation and boundary precision.

VI. OBJECTIVES

The primary objective of this research is to develop a deep learning–based semantic segmentation framework that can effectively overcome the limitations imposed by severe class imbalance in high-resolution satellite imagery. The work aims to enhance the detection accuracy, recall, and boundary precision of minority land-cover classes such as small water bodies, thin river structures, fragmented urban patches, and early-stage deforestation areas, all of which are typically overlooked by conventional segmentation models. A key objective is to design an improved network architecture capable of achieving superior multi-scale feature fusion, with U-Net++ selected for its nested skip connections that reduce the semantic gap between encoder and decoder representations. Another major objective is to engineer a hybrid loss function that integrates Area-Weighted Binary Cross-Entropy, Tversky Loss, and Boundary Loss to produce a balanced optimization signal that emphasizes small-object sensitivity, recall enhancement, and boundary-level accuracy. Additionally, the research seeks to introduce a class-aware sampling strategy that increases the effective presence of minority classes during training without disrupting dataset integrity. A final goal is to evaluate the proposed system comprehensively through metrics such as mean IoU, macro recall, and class-wise IoU, demonstrating measurable improvements over standard U-Net baselines while ensuring that the enhanced performance on minority classes does not compromise overall segmentation quality.

VII. PROJECT FLOW

1. Dataset Collection & Exploration
High-resolution LISS-IV satellite imagery is collected and examined to understand class distribution, focusing on identifying severe imbalance between majority and minority land-cover classes.

2. Preprocessing of Images and Masks
All images are resized to a uniform resolution, normalized, and masks are encoded. Noise correction, color standardization, and basic cleanup are performed to ensure consistent inputs.

3. Data Augmentation
Techniques such as horizontal and vertical flips, rotations, brightness/contrast adjustments, gamma correction, elastic transformations, and minority-focused copy-paste augmentation are applied to enrich training samples.

4. Class-Aware Sampling
The dataset is analyzed tile-by-tile to separate minority-rich and majority-rich samples. A custom sampler ensures that each batch contains a balanced mix, improving minority-class exposure during training.

5. Model Design Using U-Net++ Architecture
The U-Net++ model with a ResNet-34 encoder is implemented, featuring nested skip connections and dense decoder paths to enhance multi-scale feature fusion and small-object reconstruction.

6. Hybrid Loss Function Formulation
The model is trained using a hybrid loss composed of Area-Weighted BCE, Tversky Loss, and Boundary Loss to balance minority-class representation, improve recall, and sharpen boundary predictions.

7. Model Training with Ranger Optimizer
Training is performed using the Ranger optimizer (RAdam + Lookahead), which stabilizes learning under complex loss interactions and accelerates convergence.

8. Prediction and Segmentation Map Generation
The trained model generates pixel-wise land-cover segmentation maps for unseen satellite images.

9. Evaluation of Model Performance
Quantitative metrics such as overall accuracy, macro F1-score, macro recall, mean IoU, and class-wise IoU are computed to assess performance across all classes.

10. Visualization and Interpretation
Predicted segmentation maps are compared with ground-truth masks to visually validate improvements in minority-class detection, structural continuity, and boundary quality.

VIII. APPLICATIONS

1. Water Resource Monitoring
The model can accurately detect small ponds, reservoirs, canals, and narrow river structures, enabling precise mapping of water resources for hydrological studies and irrigation planning.

2. Flood Risk Assessment and Disaster Management
Improved segmentation of water bodies and surrounding terrain helps authorities identify flood-prone regions, monitor rising water levels, and plan evacuation or mitigation strategies.

3. Agricultural Land Analysis
Accurate classification of agricultural fields supports crop monitoring, yield estimation, and early identification of drought-affected regions, assisting

farmers and policymakers.

4. Urban Planning and Infrastructure Development
Enhanced detection of urban patches allows planners to monitor urban expansion, evaluate land-use changes, and optimize future development projects.

5. Forest and Biodiversity Conservation
The segmentation of forest areas and their boundaries helps track deforestation patterns, habitat fragmentation, and loss of biodiversity hotspots.

6. Environmental Change Detection
Multi-temporal monitoring becomes more reliable when small land-cover variations—such as shrinking water bodies or changing vegetation—are captured with high accuracy.

7. Climate Change Impact Studies
Long-term segmentation outputs help analyze changes in land cover driven by climate variability, including wetland shrinkage and desertification trends.

8. Smart City and GIS Applications
Accurate, high-resolution land-cover maps can be integrated into Geographic Information Systems (GIS) for urban analytics, smart resource allocation, and infrastructure planning.

9. Precision Agriculture
Farmers can utilize precise land segmentation for optimized irrigation, fertilizer distribution, and soil-health management.

10. Remote Monitoring for Government Agencies
Environmental agencies can use satellite-based segmentation to enforce regulations, monitor illegal activities (such as unauthorized land clearing), and ensure sustainable development.

IX. SYSTEM ARCHITECTURE

DATA INPUT → PREPROCESSING → CLASS-AWARE SAMPLING
↓ ↓ ↓
AUGMENTATION → U-NET++ MODEL → HYBRID LOSS + RANGER OPTIMIZER
↓
SEGMENTATION OUTPUT → EVALUATION → VISUALIZATION

X. ARCHITECTURE ILLUSTRATION

```
Encoder (ResNet-34) E1 → E2 → E3 → E4
↓            ↓      ↓
D1,0  D2,0  D3,0  D4,0
|           |     |
↓          ↓     ↓
D1,1  D2,1  D3,1
|           |
↓          ↓
D1,2  D2,2
|
↓
Final Decoder → Output Segmentation Map
```

XI. EXPERIMENTS

Additional experiments were conducted to measure the sensitivity of the model to various augmentation techniques, learning rates, and encoder backbones. Replacing ResNet-34 with ResNet-50 resulted in marginal improvement but significantly increased training time. Experiments with reduced augmentation showed a decline in minority-class performance, confirming the importance of synthetic diversity. Similarly, dropout regularization contributed positively to generalization, particularly in dense urban regions. Experiments on different tile sizes (128×128 and 512×512) showed that 256×256 provided the best balance between spatial context and computational efficiency. These additional findings further validate the design decisions of the proposed framework.

XII. FUTURE SCOPE

The future scope of this research is extensive, as the proposed hybrid-loss U-Net++ framework establishes a solid foundation for more advanced developments in the semantic segmentation of satellite imagery. One important direction for future work is the integration of multispectral, hyperspectral, and SAR (Synthetic Aperture Radar) data, which provide richer spectral signatures and structural information than traditional RGB imagery. These modalities can significantly improve the model's ability to distinguish between spectrally similar land-cover categories and enhance performance under challenging atmospheric conditions such as haze, cloud cover, and seasonal variation. Additionally, future research can explore multi-temporal analysis, where satellite images captured over different seasons or time intervals are used together to detect gradual land-cover transitions, seasonal crop cycles, water-level fluctuations, and long-term environmental changes. Incorporating temporal dependencies through sequence models such as ConvLSTMs or transformer-based video architectures can further strengthen the predictive capability for climate-related applications. Another promising avenue involves leveraging advanced transformer architectures like Swin-UNet, SegFormer, or Vision Transformer hybrids to overcome limitations in long-range spatial dependency modeling, an area where CNN-based models still struggle. Transformers may enable more accurate modeling of large-scale landscapes, particularly in heterogeneous regions where contextual relationships extend across wide spatial extents.

Future work may also include the development of self-supervised and semi-supervised learning techniques to overcome the scarcity of annotated satellite datasets. Labeling remote-sensing imagery is time-consuming and requires expert knowledge; therefore, methods such as contrastive learning, pseudo-labeling, and teacher–student frameworks can significantly reduce annotation dependency while maintaining segmentation quality. Domain adaptation and domain generalization techniques also represent a valuable future direction, enabling models trained on one geographic region to generalize effectively to new areas without retraining. This would be particularly beneficial for global-scale environmental monitoring systems. Another important extension is the creation of lightweight, energy-efficient versions of the model optimized for deployment on drones, edge devices, and IoT-based environmental monitoring systems. Such models would facilitate real-time segmentation for applications like flood prediction, wildfire detection, and crop surveillance in remote or resource-constrained environments.

In addition, the hybrid loss function itself may be expanded to include uncertainty modeling, enabling the system to provide confidence scores for each pixel classification. This capability would allow practitioners to identify high-risk or ambiguous regions that require manual inspection, thereby increasing trust and transparency in automated mapping systems. Future research could also explore adaptive loss weighting mechanisms that dynamically adjust the contribution of each loss component based on batch-level statistics or class distribution changes during training. Furthermore, the integration of post-processing algorithms such as CRFs, morphological reconstruction, or graph-based refinement techniques may further enhance segmentation accuracy, especially along complex boundaries and in areas with overlapping spectral characteristics.

Finally, the long-term scope includes the development of large-scale monitoring frameworks that combine segmentation outputs with geographic information systems (GIS), hydrological models, agricultural forecasting tools, and urban-planning platforms to create fully automated decision-support systems. Such integrated platforms could support government agencies, researchers, and industries by providing real-time environmental insights, sustainability indicators, and predictive analytics powered by continuous satellite observation. As satellite technologies evolve and data volumes continue to increase, the need for scalable, interpretable, and highly accurate segmentation models become even more critical. The proposed framework serves as an important stepping stone toward these broader advancements by establishing a robust method for handling class imbalance, one of the most persistent challenges in Earth Observation analytics. Future work will continue building on this foundation to create more adaptive, intelligent, and globally applicable remote-sensing systems.

XIII. CONCLUSION

This research presents a comprehensive framework designed to address the persistent challenge of class imbalance in high-resolution satellite-image segmentation. By integrating the U-Net++ architecture with a carefully engineered hybrid loss function and a class-aware sampling strategy, the proposed system significantly enhances the detection of minority land-cover classes such as water bodies and fragmented urban regions. The hybrid loss, combining Area-Weighted BCE, Tversky Loss, and Boundary Loss, ensures balanced gradient contributions, improved recall, and sharper boundary representation, while the nested skip connections of U-Net++ strengthen multi-scale feature learning. Experimental results demonstrate considerable improvements in overall accuracy, mean IoU, and minority-class IoU compared to standard U-Net models, confirming the value of a multi-level solution rather than isolated architectural or loss-based modifications. The model's strong performance and visual coherence highlight its potential for real-world applications in environmental monitoring, urban planning, and natural-resource management. Overall, this work shows that addressing class imbalance requires an integrated design approach, and the methods developed here provide a strong foundation for future advancements in remote-sensing segmentation.

REFERENCES

[1] Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 823–870, 2007.

[2] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[4] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

[5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Trans. Med. Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.