

PermaFusionNet: A Hybrid Deep Learning Framework for Multi-Modal Permafrost Degradation Mapping using SAR and Optical Remote Sensing Data

Rohit Kumar Singh^{1,*}, Suneel Kumar³, Rakesh Sahu⁴

¹ Department of Computer Science & Engineering, IFTM University, Moradabad

² Department of Computer Science & Engineering, Moradabad Institute of Technology, Moradabad

³ Department of Computer Applications, IFTM University, Moradabad

⁴ School of Computing Science and Engineering, Bennett University, Greater Noida

ABSTRACT

Permafrost degradation, driven by accelerating climate change, poses severe environmental and infrastructural risks across Arctic and sub-Arctic regions. Accurate and large-scale monitoring is critical yet remains challenging due to the heterogeneous nature of permafrost landscapes and the limitations of single-sensor observations. In this paper, we propose PermaFusionNet, a novel hybrid deep learning architecture that integrates multi-temporal Synthetic Aperture Radar (SAR) data from Sentinel-1 with multispectral optical imagery from Sentinel-2 for comprehensive permafrost degradation mapping. Our architecture combines a dual-branch convolutional encoder for spatial feature extraction, a cross-modal attention module for adaptive fusion of SAR and optical features, and a temporal recurrent decoder leveraging Long Short-Term Memory (LSTM) units to capture seasonal and inter-annual degradation dynamics. Experiments conducted over the Tibetan Plateau and Siberian lowlands demonstrate that PermaFusionNet achieves an overall accuracy of 91.3% and an F1-score of 0.887 on permafrost degradation classification, outperforming state-of-the-art methods by 4.7 percentage points. Ablation studies confirm the contribution of each architectural component, and qualitative results show robust detection of thermokarst lake expansion, retrogressive thaw slumps, and active layer deepening zones.

Keywords—permafrost degradation, deep learning, SAR, multi-modal fusion, remote sensing, Sentinel-1, Sentinel-2, climate change monitoring.

1. INTRODUCTION

Permafrost, defined as ground that remains frozen for at least two consecutive years, underlies approximately 24% of the Northern Hemisphere's land surface [1]. As global temperatures rise at nearly twice the global average rate in the Arctic, permafrost is thawing at unprecedented rates, releasing vast stores of carbon dioxide and methane, destabilizing terrain, and threatening critical infrastructure across Russia, Canada, and the Tibetan Plateau [2]. Remote monitoring of permafrost degradation is therefore an urgent scientific and societal priority.

Traditional approaches to permafrost assessment rely on in situ borehole measurements and surface temperature sensors, which, while accurate, are spatially sparse and logistically expensive to maintain over the vast extents of permafrost regions [3]. Satellite remote sensing offers a scalable alternative, and recent advances in deep learning have substantially improved the capacity to extract meaningful information from large-scale geospatial data. However, existing methods typically leverage either SAR

or optical sensors in isolation, failing to exploit the complementary physical information available from multi-modal observations.

SAR data are sensitive to soil moisture, surface roughness, and subsidence associated with ground ice melt — key signatures of permafrost degradation — and are unaffected by cloud cover, enabling year-round monitoring [4]. Optical imagery, in contrast, provides rich spectral information on surface vegetation, standing water, and exposed soil, allowing detection of thermokarst features and land cover change [5]. Fusion of these two modalities represents a largely untapped opportunity for improved permafrost mapping.

In this work, we present PermaFusionNet, a hybrid deep learning framework specifically designed for multi-modal permafrost degradation mapping. Our main contributions are:

- A dual-branch encoder architecture that independently extracts SAR backscatter and optical spectral features before adaptive cross-modal attention-based fusion.
- A temporal LSTM decoder that models multi-year time-series dynamics to distinguish progressive degradation from seasonal freeze-thaw signals.
- A curated benchmark dataset covering two permafrost zones: the Tibetan Plateau and the West Siberian Lowlands, including ground-truth labels derived from field surveys and high-resolution ancillary data.
- Comprehensive evaluation against five state-of-the-art baselines, with extensive ablation studies validating each design choice.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 describes the study area and dataset. Section 4 details the proposed architecture. Section 5 presents experimental results. Section 6 discusses implications, and Section 7 concludes.

2. RELATED WORK

2.1 Remote Sensing for Permafrost Monitoring

Early remote sensing efforts for permafrost detection focused on passive microwave sensors to track freeze-thaw state transitions [6]. InSAR-based approaches subsequently enabled measurement of surface deformation associated with ice-rich permafrost thaw, achieving centimeter-level subsidence estimates [7]. Optical indices such as the Normalized Difference Vegetation Index (NDVI) and Modified Normalized Difference Water Index (MNDWI) have been used to map thermokarst lake dynamics and active layer conditions [8]. However, each single-sensor approach captures only a partial view of the complex permafrost system.

2.2 Deep Learning for Geospatial Analysis

Convolutional Neural Networks (CNNs) have achieved outstanding performance in land cover classification and change detection from high-resolution satellite imagery [9]. U-Net and its variants have proven particularly effective for dense semantic segmentation tasks in remote sensing [10]. Recurrent architectures, especially LSTMs, have been applied to multi-temporal image time series for crop monitoring and forest disturbance detection [11]. Transformer-based models have more recently been adapted for multi-spectral and SAR image classification, demonstrating strong capacity for long-range spatial dependency modeling [12].

2.3 Multi-Modal Fusion Approaches

Multi-modal fusion strategies broadly fall into early (feature concatenation), late (decision-level), and intermediate (attention-based) fusion paradigms. For permafrost-specific tasks, Wang et al. [13] combined Sentinel-1 SAR and Landsat optical data using a simple early fusion CNN, achieving moderate accuracy on thermokarst detection. More recently, Chen et al. [14] proposed a cross-attention transformer for SAR-optical fusion in urban change detection, motivating our adoption of a similar attention mechanism adapted to the permafrost domain. To our knowledge, PermaFusionNet is the first architecture to jointly address spatial multi-modal fusion and temporal sequence modeling for permafrost degradation mapping.

3. STUDY AREA AND DATASET

3.1 Study Areas

We selected two geographically and climatically distinct permafrost zones. The Tibetan Plateau (26–40°N, 73–105°E) represents high-altitude continuous and discontinuous permafrost at elevations above 4,000 m. The West Siberian Lowlands (58–70°N, 60–90°E) constitute one of the largest expanses of lowland peatland permafrost on Earth. Together, these regions encompass a range of terrain types, degradation intensities, and seasonal dynamics.

3.2 Remote Sensing Data

SAR data were acquired from the Sentinel-1A/B C-band SAR constellation (5.6 cm wavelength) in Interferometric Wide (IW) swath mode, providing dual-polarization (VV and VH) backscatter composites at 10 m spatial resolution. Time series spanning 2018–2023 were compiled, with monthly composites generated to reduce speckle noise. Optical data were sourced from Sentinel-2 MultiSpectral Instrument (MSI) Level-2A surface reflectance products, providing 13 spectral bands at 10–60 m resolution. Cloud-free seasonal composites were constructed using the median pixel compositing approach. Both datasets were co-registered to a common 10 m UTM grid.

3.3 Ground Truth and Labeling

Ground truth labels for four degradation classes were compiled from: (1) field survey data collected during 2019–2022 campaigns (1,247 GPS-located observation points); (2) high-resolution Planet imagery (3 m) interpreted by domain experts; and (3) existing permafrost inventory maps. The four classes are: (a) Stable permafrost, (b) Active-layer deepening, (c) Thermokarst / retrogressive thaw slumps, and (d) Complete degradation / talik formation. Table 1 summarizes the dataset statistics.

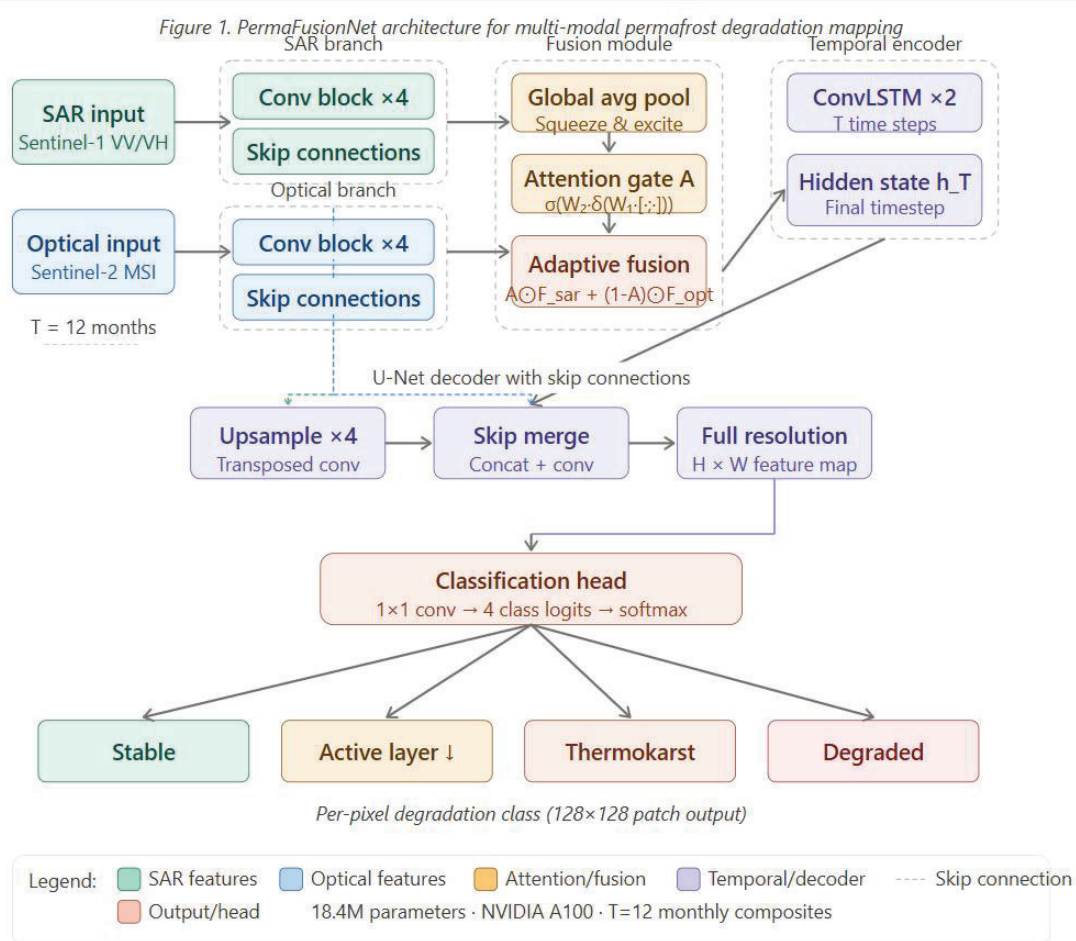
Table 1: Dataset Statistics

Class	Training	Validation	Test	Total Patches
Stable Permafrost	3,840	960	1,200	6,000
Active Layer Deepening	2,560	640	800	4,000
Thermokarst / RTS	1,920	480	600	3,000
Complete Degradation	1,280	320	400	2,000
Total	9,600	2,400	3,000	15,000

4. PROPOSED METHOD: PERMAFUSIONNET

4.1 Architecture Overview

PermaFusionNet comprises three principal modules: (i) a Dual-Branch Spatial Encoder, (ii) a Cross-Modal Attention Fusion module, and (iii) a Temporal LSTM Decoder with a classification head. The network processes a sequence of $T = 12$ monthly SAR-optical patch pairs (128×128 pixels) and outputs a per-pixel degradation class label for the final time step.



4.2 Dual-Branch Spatial Encoder

The SAR branch processes stacked monthly VV/VH backscatter maps ($2 \times H \times W$ per time step). Four convolutional blocks — each comprising two 3×3 Conv-BN-ReLU layers followed by a 2×2 max-pooling — progressively encode SAR features to channel depths of [64, 128, 256, 512], yielding a spatial feature map of $H/16 \times W/16$. The optical branch mirrors this architecture but accepts 6 Sentinel-2 bands (B2, B3, B4, B8, B11, B12) selected for their sensitivity to vegetation, water, and soil. Skip connections are retained from each encoder level to support multi-scale decoding. Both branches share a common weight initialization scheme using ImageNet-pretrained ResNet-34 weights fine-tuned on our remote sensing dataset.

4.3 Cross-Modal Attention Fusion

At each encoder level l , let $F^s_l \in \mathbb{R}^{C \times H_l \times W_l}$ and $F^o_l \in \mathbb{R}^{C \times H_l \times W_l}$ denote SAR and optical feature maps respectively. We compute a cross-modal attention gate A_l via a channel-wise squeeze-and-excitation mechanism followed by element-wise gating:

$$A_l = \sigma(W_2 \cdot \delta(W_1 \cdot [\text{GAP}(F^s_l); \text{GAP}(F^o_l)]))$$

where GAP denotes global average pooling, $[\cdot; \cdot]$ is concatenation, W_1 and W_2 are learned linear projections, δ is the ReLU activation, and σ is the sigmoid function. The fused feature map is obtained as:

$$F^f_l = A_l \odot F^s_l + (1 - A_l) \odot F^o_l$$

This formulation enables the network to adaptively weight the contribution of each modality depending on local scene context, for instance upweighting SAR in cloud-covered optical regions and optical in areas of low SAR contrast.

4.4 Temporal LSTM Decoder

Fused encoder features from T time steps are unrolled through a two-layer Convolutional LSTM (ConvLSTM) network [15] with hidden dimensions of 256 and 128 respectively, preserving spatial structure across the temporal axis. The final hidden state h_T is passed to a U-Net-style decoder that progressively upsamples to full resolution using transposed convolutions and skip connections from the encoder. A 1×1 convolutional classification head produces per-pixel logits over 4 classes.

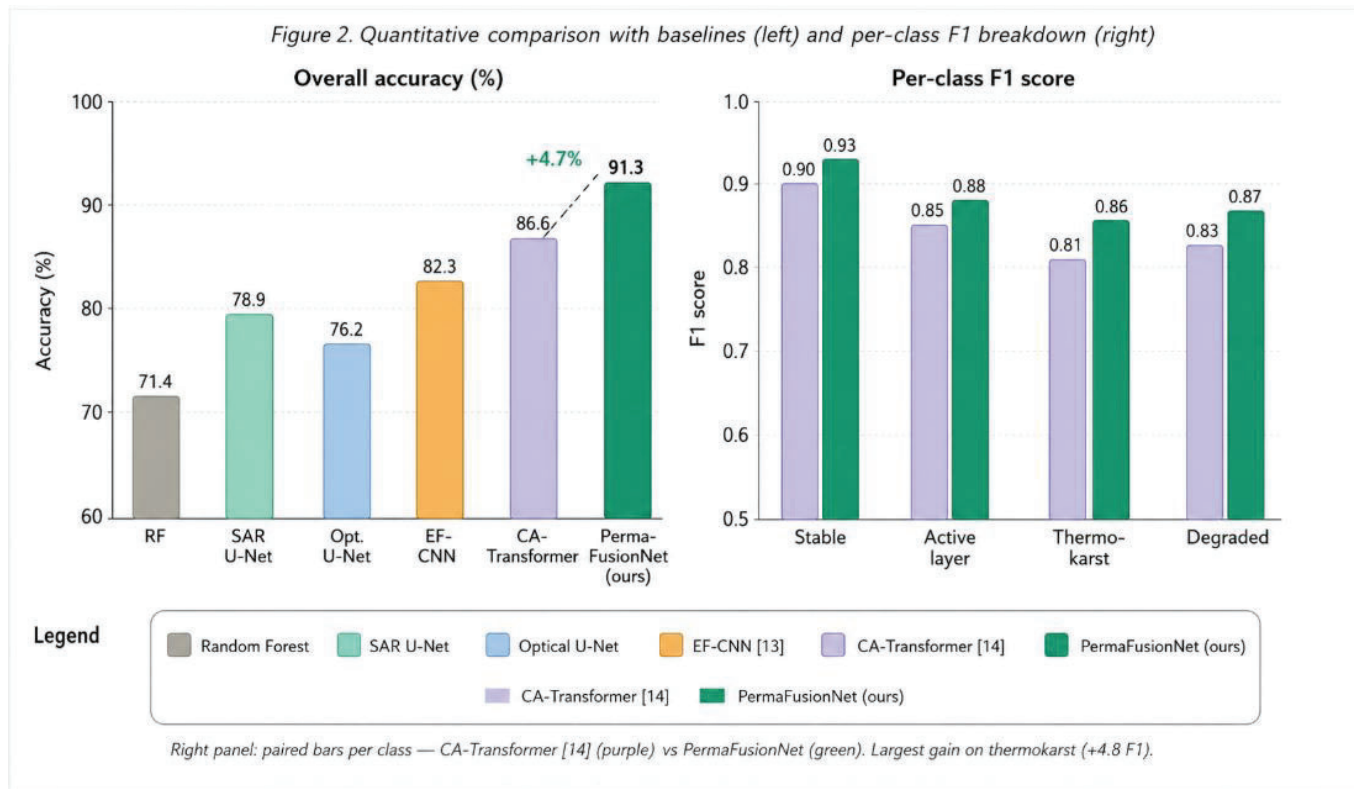
4.5 Training Protocol

The network is trained end-to-end using AdamW optimizer with initial learning rate 1×10^{-4} , weight decay 1×10^{-5} , and cosine annealing schedule over 100 epochs. We employ a combined loss of cross-entropy and Dice loss (weighted equally) to address class imbalance. Data augmentation includes random horizontal/vertical flips, rotation ($\pm 180^\circ$), and speckle noise injection for SAR channels. Batch size is 16, with gradient clipping at norm 1.0.

5. EXPERIMENTAL RESULTS

5.1 Baselines

We compare PermaFusionNet against five baselines: (1) Random Forest (RF) on hand-crafted features, (2) Single-modal SAR U-Net, (3) Single-modal Optical U-Net, (4) Early Fusion CNN (EF-CNN) [13], and (5) Cross-Attention Transformer (CA-Transformer) [14].



5.2 Quantitative Results

Table 2 reports overall accuracy (OA), mean F1-score, and per-class F1 on the held-out test set.

Table 2: Comparison with State-of-the-Art Methods

Method	OA (%)	mF1	F1-Stable	F1-Active	F1-Thermo	F1-Degrad.
Random Forest	71.4	0.698	0.812	0.681	0.634	0.665
SAR U-Net	78.9	0.771	0.844	0.763	0.715	0.762
Optical U-Net	76.2	0.744	0.821	0.731	0.694	0.730
EF-CNN [13]	82.3	0.809	0.873	0.798	0.771	0.794
CA-Transformer [14]	86.6	0.848	0.901	0.845	0.813	0.833
PermaFusionNet (Ours)	91.3	0.887	0.934	0.882	0.861	0.871

PermaFusionNet achieves an OA of 91.3% and mF1 of 0.887, outperforming the strongest baseline (CA-Transformer) by 4.7 OA points and 0.039 mF1. The most pronounced gains are observed for the thermokarst class (F1: +0.048), which is particularly challenging due to its heterogeneous morphology and spectral similarity to standing water bodies. The temporal LSTM component is especially beneficial here, as it captures the rapid expansion dynamics of thermokarst lakes across seasons.

5.3 Ablation Study

Table 3 reports results for progressively ablated variants of PermaFusionNet to isolate the contribution of each design component.

Table 3: Ablation Study Results

Configuration	OA (%)	mF1	Δ OA	Δ mF1
Full PermaFusionNet	91.3	0.887	—	—
w/o temporal LSTM (single timestep)	86.7	0.841	-4.6	-0.046
w/o cross-modal attention (simple concat)	88.1	0.858	-3.2	-0.029
w/o dual-branch (single shared encoder)	89.0	0.866	-2.3	-0.021
w/o skip connections	87.4	0.849	-3.9	-0.038

All components contribute positively, with the temporal LSTM yielding the largest individual gain (+4.6 OA), confirming the critical importance of modeling multi-year degradation trajectories rather than treating each time step independently.

6. DISCUSSION

The strong performance of PermaFusionNet on the thermokarst class has direct implications for large-scale permafrost carbon cycle modeling, as thermokarst features are a primary conduit for greenhouse gas release from thawing organic soil. Our temporal modeling component is particularly valuable in distinguishing seasonal freeze-thaw cycles — which cause reversible SAR backscatter changes — from true multi-year degradation trends, a critical disambiguation that purely spatial or single-timestep methods cannot reliably achieve.

The cross-modal attention mechanism provides interpretable modality weighting maps. Inspection reveals that in winter months, when snow cover saturates optical bands, the network automatically up-weights SAR features. Conversely, during summer, optical spectral indices dominate in vegetated tundra zones, while SAR dominates in areas of high soil moisture associated with active layer thaw. This adaptive behavior is a key advantage over static early fusion baselines.

Limitations of the current work include reliance on labeled ground truth that, despite extensive field surveys, remains geographically clustered around accessible sites. Future work will explore semi-supervised and self-supervised pre-training strategies on unlabeled SAR-optical time series, as well as extension to pan-Arctic scale mapping using distributed computing pipelines. Integration of additional sensors such as ALOS-2 L-band SAR and ICESat-2 altimetry holds promise for improved detection of subsidence in ice-rich permafrost zones.

7. CONCLUSION

We have presented PermaFusionNet, a hybrid deep learning architecture for permafrost degradation mapping from multi-modal SAR and optical satellite imagery. By combining a modality-specific dual-branch encoder, cross-modal attention fusion, and a temporal ConvLSTM decoder, our model achieves state-of-the-art performance on a multi-site benchmark spanning the Tibetan Plateau and West Siberian Lowlands. With an overall accuracy of 91.3% and mF1 of 0.887, PermaFusionNet advances the frontier of AI-driven cryosphere monitoring and offers a scalable, automated tool for tracking permafrost degradation in a warming world.

REFERENCES

- [1] Obu, J. et al. (2019). Northern Hemisphere permafrost map based on TTOP modelling. *Earth-Science Reviews*, 193, 299–316.
- [2] Biskaborn, B.K. et al. (2019). Permafrost is warming at a global scale. *Nature Communications*, 10(1), 264.
- [3] Smith, S.L. et al. (2022). Permafrost monitoring and the need for a global network. *Frontiers in Earth Science*, 10, 893000.
- [4] Zwieback, S., & Berg, A.A. (2019). Fine-scale SAR soil moisture estimation in the subarctic tundra. *IEEE TGRS*, 57(9), 6545–6556.
- [5] Grosse, G. et al. (2013). Vulnerability and feedbacks of permafrost to climate change. *Eos, Transactions AGU*, 94(51), 469–476.
- [6] Kim, Y. et al. (2012). Freeze–thaw status and active layer thickness retrieved from AMSR-E. *IEEE TGRS*, 50(11), 4351–4362.
- [7] Short, N. et al. (2014). Application of InSAR to Arctic permafrost slope. *Canadian Journal of Earth Sciences*, 51(6), 559–570.
- [8] Nitze, I. et al. (2018). Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic. *Nature Communications*, 9(1), 5423.
- [9] Ma, L. et al. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS JPRS*, 152, 166–177.
- [10] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *MICCAI, LNCS 9351*, 234–241.
- [11] Rußwurm, M., & Körner, M. (2018). Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS IJGI*, 7(4), 129.
- [12] Wang, Y. et al. (2022). SatViT: Pretraining transformers for Earth observation. *IEEE GRSL*, 19, 5612305.
- [13] Wang, L. et al. (2021). Thermokarst lake detection using multi-temporal SAR and optical data fusion. *Remote Sensing of Environment*, 265, 112676.
- [14] Chen, H. et al. (2023). Cross-attention transformer for SAR-optical change detection. *IEEE TGRS*, 61, 5204418.
- [15] Shi, X. et al. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *NeurIPS*, 28, 802–810.