

---

# EarthScape: A Multimodal Dataset for Surficial Geologic Mapping and Earth Surface Analysis (Supplementary Material)

---

**Matthew A. Massey**  
Kentucky Geological Survey  
University of Kentucky  
Lexington, KY 40506-0053  
matthew.massey@uky.edu

**Nusrat Munia**  
Department of Computer Science  
University of Kentucky  
Lexington, KY 40506-0633  
nusrat.munia@uky.edu

**Abdullah-Al-Zubaer Imran**  
Department of Computer Science  
University of Kentucky  
Lexington, KY 40506-0633  
aimran@uky.edu

## 1 Code Availability and Reproducibility

All code used for data preprocessing, patch extraction, model training, and evaluation is publicly available at <https://github.com/masseygeo/earthscape>. The repository includes clear documentation and instructions for reproducing all experiments presented in the main paper and supplemental material. The codebase provides tools for downloading and aligning multimodal data (including GeoTIFF imagery and vector layers), generating spatially independent patch splits, and computing terrain derivatives. It also includes baseline model implementations of SGMap-Net using both ResNeXt-50 and ViT-B/16 backbones, along with scripts for training, evaluation, and visualization. Additional utilities support focal loss configuration, per-class performance metrics, and spatial overlays of predictions. The full EarthScape dataset is publicly available at [https://uknowledge.uky.edu/kgs\\_data/16/](https://uknowledge.uky.edu/kgs_data/16/). The dataset archive includes geospatially registered input images, multilabel target masks, class proportion tables, a README, and a detailed data dictionary describing all included modalities.

## 2 Exploring the EarthScape Dataset

### 2.1 Geographic Extent

Fig. 1 illustrates the current and planned geographic extent of the EarthScape dataset. At present, the dataset includes two spatially independent regions in central Kentucky: Warren County, which contains the largest number of image patches, and Hardin County, which serves as an independent test area with similar geologic and geomorphic conditions. This separation enables evaluation of cross-region generalizability. The EarthScape dataset is designed as a “living” resource and 14 additional 7.5-minute quadrangles will soon be added, nearly tripling the number of patches (Fig. 1).

### 2.2 Geologic Relevance

While EarthScape is highly relevant to the Interior Low Plateaus and, potentially, the adjacent Appalachian Plateaus to the east, geologic differences limit its transferability into other major regions (Fig. 1). Specifically, the glaciated Central Lowlands to the north and the Coastal Plain to the west

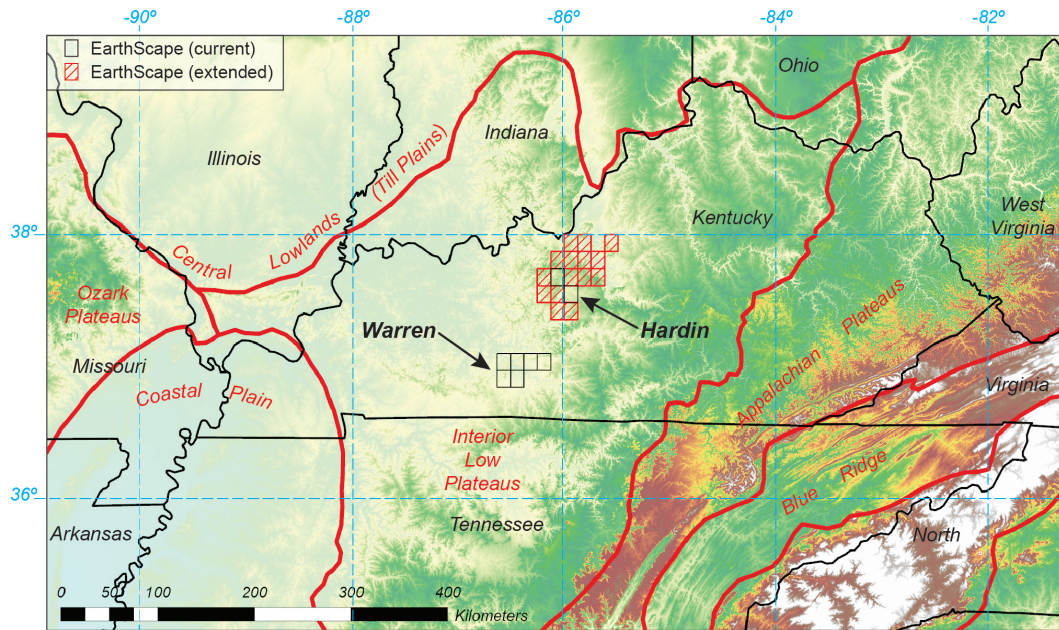


Figure 1: Map of the central United States showing the publicly available 1:24,000-scale surficial geologic maps. Red lines show boundaries of major geologic provinces, which provide geological constraints for generalizability. EarthScape-trained models are expected to generalize effectively throughout the Interior Low Plateaus and adjacent Appalachian Plateaus, based on shared terrain, bedrock, and geomorphic processes. In contrast, the glaciated Central Lowlands and Coastal Plain are characterized by fundamentally different surficial processes and materials.

are characterized by fundamentally different surficial materials and geomorphic processes. These boundaries are highlighted to clarify both the applicability and the current limitations of the dataset for broader generalization studies.

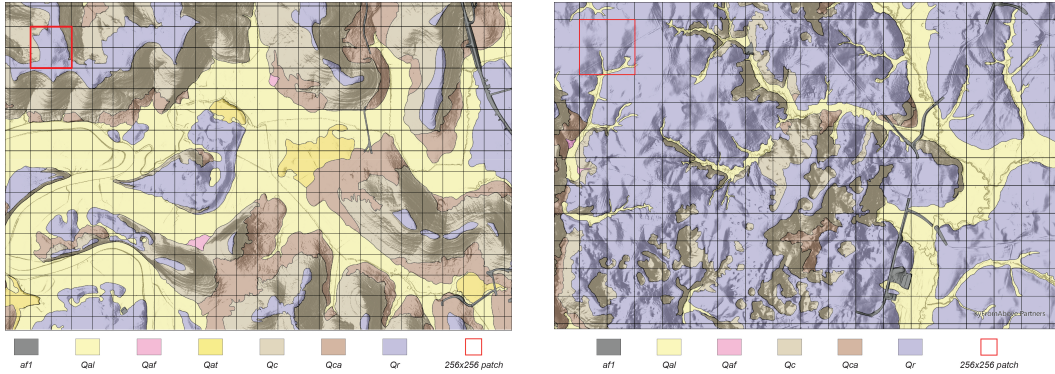
Importantly, the bedrock-dominated, unglaciated terrain represented in EarthScape is not unique to Kentucky and the immediate surrounding area. Similar landscapes characterized by carbonate bedrock, dissected plains, and mixed fluvial-colluvial systems are found in regions such as the Ozark Plateau (USA), parts of the Carpathians (Eastern Europe), the Dinaric Alps (Balkans), and select areas of central China and southeastern Australia. As such, EarthScape may serve as a valuable pretraining or adaptation resource for similar geologic regions worldwide.

### 2.3 Surficial Geology

Fig. 2 presents two examples of surficial geologic maps from the EarthScape dataset, shown as semi-transparent overlays atop multi-directional hillshade images. This visualization emphasizes the strong relationship between surficial geology and topography. Distinct landforms, such as river valleys, plains, and steep hillslopes, are spatially correlated with specific surficial geologic units. EarthScape leverages this relationship to frame surficial geologic mapping as a vision task, where computer vision models can learn to associate surface patterns with underlying geological processes.

The EarthScape dataset currently includes seven surficial geologic map units, each representing distinct surface processes. Although the maps are from Kentucky, the units reflect fluvial deposition, gravitational transport, and in-situ weathering processes that are active in many landscapes worldwide.

1. *Artificial fill (af1)*: Manmade deposits consisting of transported or excavated material placed or removed for engineering, mining, or other anthropogenic structures. Includes road embankments, building pads, quarries, and areas of significant topographic modification. Often exhibits sharp, angular boundaries. The spatial extent of af1 can be below the mapping resolution and inconsistently captured on expert-curated surficial geologic maps.



(a) Surficial geologic map of part of Warren County. (b) Surficial geologic map of part of Hardin County.

Figure 2: Example surficial geologic maps showing the distribution of unconsolidated materials overlaid on hillshade images to emphasize topographic context. The spatial correspondence between SG map units and landscape features, such as valleys and slopes, is visually apparent. The black grid indicates the layout of EarthScape patches, each measuring  $1280 \times 1280$  feet ( $256 \times 256$  pixels) with 50% overlap. Red squares in the upper left corners highlight a single patch

2. *Alluvium (Qal)*: Unconsolidated sediments, typically consisting of clay-, silt-, sand-, and gravel-sized particles, deposited by modern rivers and streams. Qal is commonly found in active floodplains and valley bottoms and reflects recent sedimentation from overbank flooding and channel migration. These areas are generally flat, vegetated, and hydrologically dynamic.
3. *Alluvial fans (Qaf)*: Fan-shaped deposits formed at the base of tributaries or drainages, where sediment-laden water rapidly spreads and loses energy. These deposits are typically coarse-grained, poorly sorted, and associated with debris flows or flash floods. Although geologically significant, Qaf are often small, making them inconsistently represented on typical 1:24,000-scale maps.
4. *Terrace deposits (Qat)*: Relict alluvial sediments preserved on elevated flat surfaces above modern stream channels. These deposits reflect former floodplain levels and subsequent stream incision. Compositionally similar to Qal, but usually expressed as distinct landforms above modern flood plains.
5. *Colluvium (Qc)*: Hillslope-derived sediments that accumulate at the base of slopes due to gravity-driven processes such as soil creep, slopewash, and shallow landslides. Qc deposits are unsorted and variable in thickness, typically found on slopes  $> 12^\circ$ . Qc is considered an active geomorphic unit.
6. *Colluvial aprons (Qca)*: Slope-derived material deposited across lower hillslopes. Qca typically occurs downslope from Qc and is more stable, having accumulated over longer time periods. These deposits may be partially weathered, with poorly defined lower boundaries that grade into Qr due to extended weathering and lower erosion rates.
7. *Residuum (Qr)*: Weathered material formed in place from the physical, chemical, and biological breakdown of underlying bedrock or older unconsolidated deposits. Qr lacks significant sediment transportation and is commonly found in upland areas with minimal active erosion. Qr is commonly gradational and poorly defined where it grades into Qc or Qca, leading to interpretive ambiguity during mapping.

## 2.4 Available Modalities

Figs. 3 and 4 showcase the diverse, multimodal data available for each of the 31,018 EarthScape patches. Each patch includes 38 co-registered channels, comprising expert-labeled geologic masks, high-resolution aerial RGB and NIR imagery, a DEM, terrain features derived from the DEM at multiple spatial scales, and rasterized vector data representing hydrologic and infrastructure features. Among these modalities, the DEM and its derived terrain features provide critical context for

Table 1: Descriptions of surficial geologic units represented in EarthScape.

Class	Name	Dominant Process	Visual Cues
af1	Artificial fill	Anthropogenic	Sharp, angular edges; linear or rectilinear shapes; DEM anomalies inconsistent with natural terrain.
Qal	Alluvium	Water-dominated	Relatively wide, flat-bottomed valleys; active stream channels; low relative elevations.
Qaf	Alluvial fans	Water-dominated (acute)	Small, isolated, lobate landforms; located at slope-base transitions.
Qat	Terrace deposits	Water-dominated (relict)	Flat benches above floodplains; stepped margins; often dissected.
Qc	Colluvium	Gravity-dominated (active)	Steep slopes ( $> 12^\circ$ ); may include landslides or erosional hazards.
Qca	Colluvial aprons	Gravity-dominated (stable)	Wedge-shaped landforms along slope bases with concave profiles; transitional between slope and plain.
Qr	Residuum	In-situ weathering	Broad, low-relief uplands; little drainage or erosion; variable surface texture.

understanding surface processes and interpreting surficial geologic units. Five terrain variables were computed at six spatial scales to capture localized and regional landform variability.

1. Slope ( $S$ ) is the first derivative of elevation, measuring the rate of change of elevation over a horizontal distance. It quantifies the steepness of the terrain, providing insight into processes like erosion and material movement.

$$S = \tan^{-1} \left( \sqrt{\left( \frac{\partial z}{\partial x} \right)^2 + \left( \frac{\partial z}{\partial y} \right)^2} \right) \quad (1)$$

Where  $\frac{\partial z}{\partial x}$  and  $\frac{\partial z}{\partial y}$  are the partial derivatives of elevation in the x and y directions, respectively.

2. Profile curvature ( $PrC$ ) is a directional second derivative of elevation, measured along the direction of the steepest slope. It quantifies how slope changes in that direction, reflecting the acceleration or deceleration of flow, and influencing erosion and deposition patterns.

$$PrC = \frac{p^2 r + 2pqs + q^2 t}{(p^2 + q^2)^{3/2}} \quad (2)$$

Where  $p = \frac{\partial z}{\partial x}$  and  $q = \frac{\partial z}{\partial y}$  are the first-order partial derivatives of elevation in the x and y directions, and  $r = \frac{\partial^2 z}{\partial x^2}$ ,  $s = \frac{\partial^2 z}{\partial x \partial y}$ , and  $t = \frac{\partial^2 z}{\partial y^2}$  are the corresponding second-order partial derivatives.

3. Planform curvature ( $PlC$ ) is another directional second derivative of elevation, measured perpendicular to the direction of the steepest slope. It describes the curvature of contour lines (lines of equal elevation) and reflects how flow paths converge or diverge across the landscape.

$$PlC = \frac{q^2 r - 2pqs + p^2 t}{(p^2 + q^2)^{3/2}} \quad (3)$$

Where  $p = \frac{\partial z}{\partial x}$  and  $q = \frac{\partial z}{\partial y}$  are the first-order partial derivatives of elevation in the x and y directions, and  $r = \frac{\partial^2 z}{\partial x^2}$ ,  $s = \frac{\partial^2 z}{\partial x \partial y}$ , and  $t = \frac{\partial^2 z}{\partial y^2}$  are the corresponding second-order partial derivatives.

4. Elevation percentile ( $EP$ ) measures the relative elevation of a point within a defined neighborhood, expressed as a percentile rank (0–100%) of the elevation among neighboring values.  $EP$  helps distinguish between landforms defined by relative topography, such as ridges, valleys, or sinkholes.

$$EP = 100 \cdot \frac{|\{z_i \in Z \mid z_i < z\}|}{N} \quad (4)$$



Where  $z$  is the elevation at the center cell,  $Z$  is the set of elevations in the neighborhood,  $z_i$  are the individual neighboring elevations, and  $N$  is the total number of neighbors. The numerator counts the number of neighbors with elevation less than  $z$ .

5. *Standard deviation of slope (SDS)* is a measure of roughness and quantifies the variability in slope angle within a local window. *SDS* represents how rugged or uneven the surface is, highlighting areas with complex topography that may correlate with diverse geologic materials or processes.

$$SDS = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \bar{S})^2} \quad (5)$$

Where  $S_i$  is the slope angle (in degrees or radians) of the  $i^{th}$  cell in the neighborhood,  $\bar{S}$  is the mean slope within that neighborhood, and  $N$  is the total number of cells used in the calculation window.

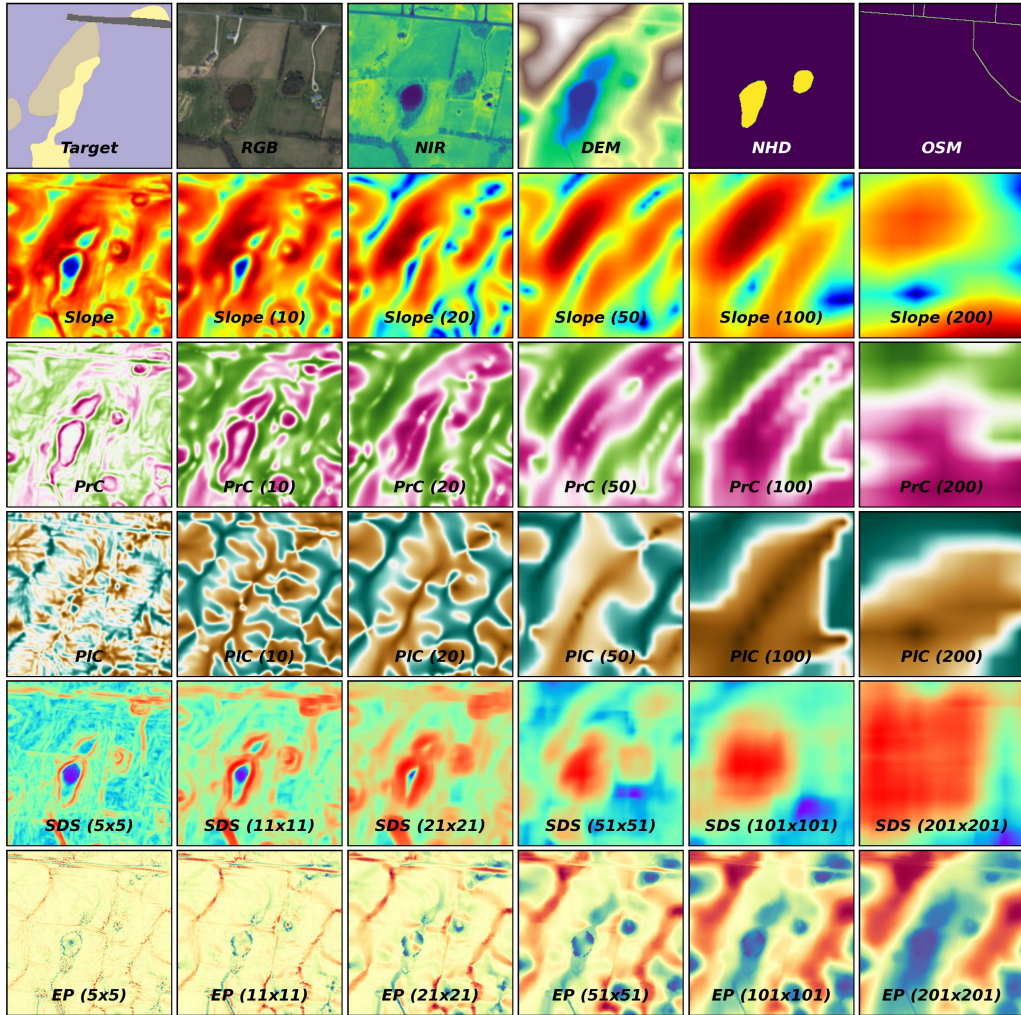


Figure 3: Example patch from the Warren County area showcasing the 38 channels available in EarthScape. Channels are displayed from top left to bottom right: target mask, RGB aerial imagery, NIR aerial imagery, DEM, hydrologic features (NHD), infrastructure (OSM), multiple scales of  $S$ ,  $PrC$ , and  $PIC$  derived from downsampled DEMs, and multiple scales of  $SDS$  and  $EP$  calculated using multiple window sizes with the original DEM.

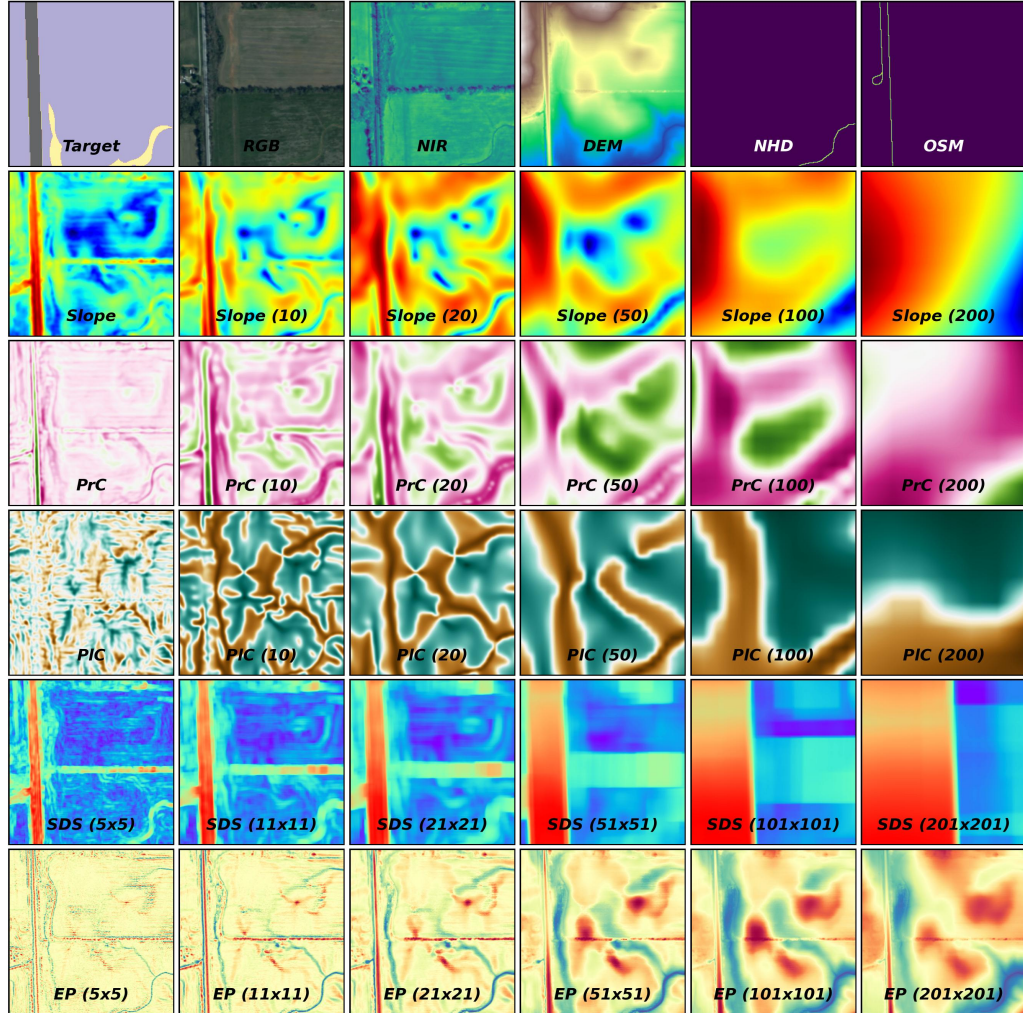


Figure 4: Example patch from the Hardin County area showcasing the 38 channels available in EarthScape. Channels are displayed from top left to bottom right: target mask, RGB aerial imagery, NIR aerial imagery, DEM, hydrologic features (NHD), infrastructure (OSM), multiple scales of  $S$ ,  $PrC$ , and  $PIC$  derived from downsampled DEMs, and multiple scales of  $SDS$  and  $EP$  calculated using multiple window sizes with the original DEM.



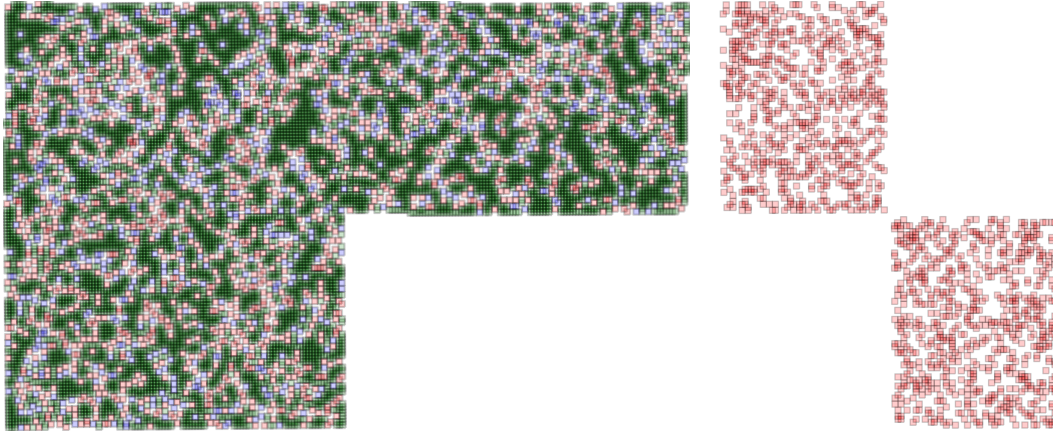
### 3 Additional Experimental Details

#### 3.1 Geospatial Patch Selection and Experimental Design

To ensure robust and geographically fair model evaluation, EarthScape patches were split into spatially independent training, validation, and test sets. The Warren County region was used for in-domain training and evaluation due to its broader spatial coverage and diversity of surficial geologic units. We first randomly selected 1,536 test patches, followed by 768 validation patches that did not spatially intersect with the test set, and then assigned the remaining 8,416 non-overlapping patches to the training set (Fig. 5). These split sizes were chosen through iterative selection to satisfy several practical constraints: (1) all splits had to be spatially non-overlapping; (2) patch counts needed to be divisible by common batch sizes (e.g., 16 or 32) to support efficient model training; (3) the resulting proportions had to be reasonably balanced and typical for supervised learning workflows (Table 2).

To assess geographic generalization, we created a cross-domain test set consisting of 1,536 randomly selected patches from the Hardin County region (Fig. 5). Although geologically similar, Hardin County is located approximately 85 km from Warren County and is spatially independent. This separate region enables testing model performance under domain shift, simulating real-world conditions in which models are applied beyond the area used for training.

Figure 6 shows the class distributions for each data split. All subsets reflect the inherent class imbalance typical of surficial geologic mapping, driven by the localized nature of surface processes. Importantly, the class distributions are consistent across the training, validation, and both test sets, ensuring that evaluation performance is not biased by differences in class representation.



(a) Training, validation, and in-domain test patches from the Warren County region. (b) Cross-domain test patches from the Hardin County region.

Figure 5: Spatial distribution of selected patches for EarthScape experiments. All splits are spatially independent: no patch overlaps between splits, though patches within the same split may partially overlap due to the 50% patch stride. See Figure 1 for geographic locations.

Table 2: Patch counts and split proportions for training, validation, and testing based on the total number of patches used for in-domain training and evaluation. An additional test set from the spatially independent Hardin County region was used to assess cross-domain generalization.

Split	Region	Patch Count ( $n$ )	In-domain Proportion (%)
Training	Warren	8,416	78.5
Validation	Warren	768	7.2
In-domain Testing	Warren	1,536	14.3
Cross-domain Testing	Hardin	1,536	-

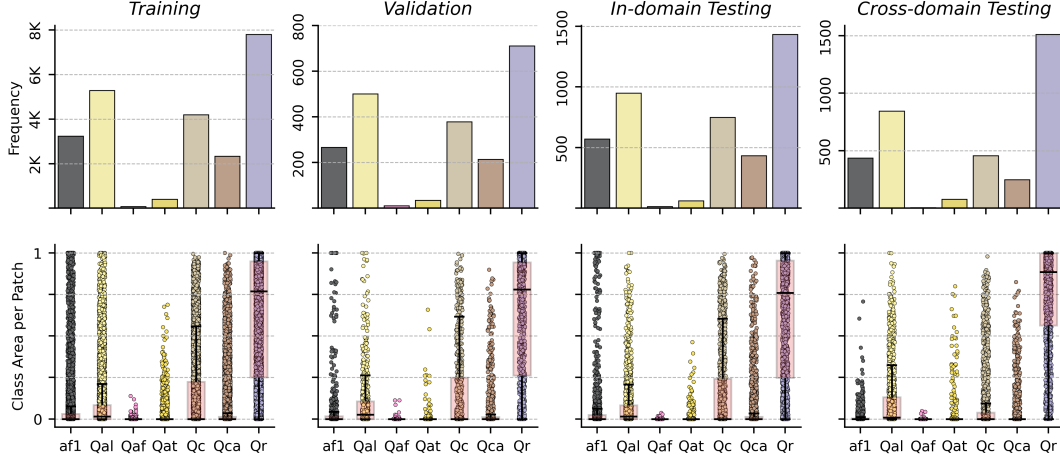
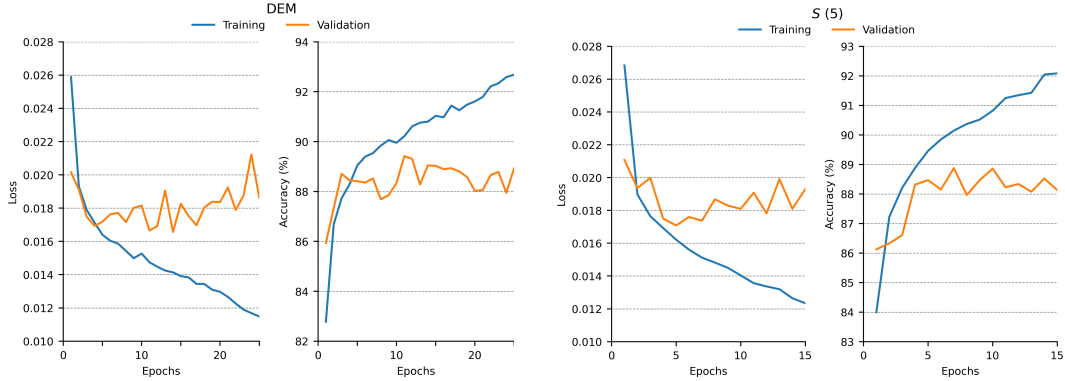


Figure 6: Class distribution and intra-patch composition across EarthScape data splits. Top row: Bar plots showing the frequency of each surficial geologic unit in the training, validation, in-domain test, and cross-domain test sets. Bottom row: Swarm plots overlaid with box plots showing the proportion of each patch occupied by each class. All splits display consistent patterns in both overall frequency and within-patch composition, supporting fair evaluation across subsets.

### 3.2 Hardware and Training Configuration

All experiments were implemented in Python using the PyTorch framework. Models were trained and evaluated on a machine equipped with an Intel Xeon processor, 128 GB of RAM, and two NVIDIA RTX A4000 GPUs. Initial training experiments were run for 25 epochs to observe convergence behavior (Fig. 7). Across all configurations, we found that model performance generally stabilized within the first 10 epochs (Fig. 7). Based on these observations, we standardized all subsequent experiments to 15 epochs, which provided a balance between sufficient training and computational efficiency.



(a) DEM model trained for 25 epochs. Early convergence is evident by epoch 10, with decreased performance thereafter.

(b)  $S(5)$  model trained for 15 epochs, demonstrating stable convergence and alignment between training and validation performance.

Figure 7: Training and validation loss and accuracy curves across epochs. Each subplot shows model loss (left panel) and accuracy (right panel) behavior for a different input modality, with training curves shown in blue and validation curves in orange.

### 3.3 Focal Loss

To address the significant class imbalance in EarthScape, we adopted focal loss. Initial tuning was conducted using the validation set and DEM modality only, a ResNeXt-50 backbone, the Adam optimizer, and a fixed learning rate of 0.001 to explore the effects of focal loss parameters. We

evaluated values of  $\gamma \in 1.0, 1.5, 2.0, 2.5, 3.0$  and tested several strategies for the class-balancing factor ( $\alpha$ ), including a fixed scalar ( $\alpha = 0.25$ ), inverse class frequency (ICF), square root of ICF ( $\sqrt{\text{ICF}}$ ), and class-balanced focal loss with  $\beta = 0.999$  (CBFL) (Table 3). The combination of  $\alpha = \sqrt{\text{ICF}}$  and  $\gamma = 2.0$  yielded the best performance for the DEM-only configuration. However, when this setting was applied to other modalities, training became unstable, and convergence was inconsistent. To ensure comparability across all experiments and isolate the effects of modality and fusion design, we adopted the original focal loss settings ( $\alpha = 0.25, \gamma = 2.0$ ) for all remaining runs.

Table 3: Per-class and macro-averaged validation set F1 scores for different focal loss configurations using the DEM modality and a ResNeXt-50 backbone. These results were used to guide focal loss tuning, although the best-performing configuration did not generalize well across modalities. As a result, we adopted the original focal loss settings ( $\alpha = 0.25, \gamma = 2.0$ ) for all subsequent experiments.

$\alpha$	$\gamma$	F1								AUC							
		af1	Qal	Qaf	Qat	Qc	Qca	Qr	AVG.	af1	Qal	Qaf	Qat	Qc	Qca	Qr	AVG.
0.25	1	0.743	0.848	0.267	0.436	0.899	0.778	0.968	0.706	0.861	0.862	0.907	0.923	0.967	0.923	0.937	0.911
0.25	1.5	0.726	0.855	0.250	0.354	0.914	0.751	0.968	0.688	0.866	0.874	0.915	0.884	0.964	0.909	0.932	0.906
0.25	2	0.749	0.841	0.229	0.400	0.914	0.778	0.965	0.697	0.868	0.859	0.929	0.919	0.970	0.929	0.912	0.912
0.25	2.5	0.690	0.866	0.275	0.387	0.895	0.767	0.971	0.693	0.844	0.887	0.944	0.895	0.965	0.920	0.945	0.914
0.25	3	0.709	0.851	0.267	0.323	0.890	0.772	0.970	0.683	0.853	0.863	0.895	0.890	0.962	0.925	0.924	0.902
ICF	1	0.524	0.804	0.204	0.390	0.831	0.640	0.961	0.622	0.639	0.730	0.921	0.851	0.912	0.828	0.851	0.819
ICF	2	0.596	0.805	0.286	0.314	0.839	0.687	0.961	0.641	0.731	0.737	0.934	0.828	0.916	0.854	0.869	0.838
ICF	2.5	0.589	0.799	0.267	0.326	0.843	0.671	0.962	0.637	0.711	0.716	0.923	0.838	0.919	0.842	0.848	0.828
$\sqrt{\text{ICF}}$	1	0.696	0.845	0.286	0.348	0.879	0.763	0.965	0.683	0.843	0.867	0.912	0.905	0.955	0.925	0.922	0.904
$\sqrt{\text{ICF}}$	1.5	0.688	0.838	0.333	0.409	0.877	0.766	0.974	0.698	0.834	0.844	0.961	0.909	0.951	0.914	0.924	0.905
$\sqrt{\text{ICF}}$	2	0.726	0.841	0.444	0.460	0.905	0.749	0.962	0.727	0.850	0.853	0.945	0.931	0.961	0.921	0.913	0.911
$\sqrt{\text{ICF}}$	2.5	0.709	0.835	0.293	0.487	0.901	0.760	0.963	0.707	0.849	0.844	0.956	0.940	0.962	0.926	0.893	0.910
CBFL	1	0.720	0.831	0.412	0.427	0.893	0.733	0.973	0.713	0.864	0.839	0.965	0.903	0.962	0.902	0.924	0.908
CBFL	1.5	0.715	0.841	0.286	0.412	0.908	0.764	0.971	0.700	0.844	0.854	0.940	0.906	0.971	0.920	0.947	0.912
CBFL	2	0.727	0.866	0.357	0.455	0.914	0.792	0.965	0.725	0.867	0.890	0.918	0.923	0.971	0.921	0.914	0.915
CBFL	2.5	0.711	0.844	0.455	0.372	0.911	0.753	0.968	0.716	0.846	0.857	0.970	0.908	0.967	0.928	0.930	0.915

### 3.4 Comprehensive Results

**Single Modality Models:** Tables 4, 5, and 6 present single-modality results across F1, AUC, precision, recall, mAP, and accuracy for both in-domain (Warren County region) and cross-domain (Hardin County region) test sets, using ResNeXt-50 and ViT-B/16 backbones. Results highlight significant performance differences between modalities and backbones, especially under domain shift (Fig. 8). Imagery-based models (RGB and NIR) show substantial degradation when transferred across regions. For example, RGB drops from 0.599 to 0.394 in macro-averaged F1 ( $\Delta_{F1} = -0.205$ ), and its AUC declines by 0.258 (Table 4). NIR shows a smaller, but still notable, performance drop. In contrast, models trained on DEM retain more performance across domains, with only a 0.105 decline in F1 and a 0.153 decline in AUC (Table 4). Terrain features derived from the DEM, including  $S$ ,  $EP$ ,  $PIC$ ,  $PrC$ , and  $SDS$ , outperform raw elevation and spectral imagery in both accuracy and generalization. Among these,  $S$  and  $EP$  stand out as the most stable and informative inputs.  $S$  (5) achieves an in-domain F1 of 0.645 and a cross-domain F1 of 0.575 with ResNeXt-50, indicating strong generalization.  $EP$   $51 \times 51$  provides high in-domain scores, but it suffers a larger drop under domain shift. In terms of backbones, ResNeXt-50 consistently achieves higher in-domain performance. However, ViT-B/16 narrows the generalization gap in most cases (Fig. 8). This suggests that ResNeXt-50 is more effective at capturing localized patterns, whereas ViT-B/16 offers slightly more robustness in unfamiliar geologic settings.

**Multi-scale Fusion Models:** Tables 7, 8, and 9 show F1, AUC, precision, recall, mAP, and accuracy results for multi-scale fusion experiments for both in-domain (Warren County region) and cross-domain (Hardin County region) test sets. Each experiment was conducted using either a ResNeXt-50 or ViT-B/16 backbone, and one of two fusion strategies: early channel stacking or attention-based fusion. Across nearly all configurations, early channel stacking consistently outperforms attention-based fusion in terms of in-domain performance. For example, with ResNeXt-50 on  $EP$ , stacking improves the in-domain F1 score from 0.494 (attention-based) to 0.640, while cross-domain performance remains stable (0.426 vs. 0.425) (Table 7). Terrain features  $S$  and  $SDS$  achieve high in-domain F1 scores (0.636–0.637) and exhibit comparatively small performance drops across domains ( $\Delta_{F1} = 0.043$ –0.048), underscoring their robustness (Fig. 8). In contrast,  $PIC$  and  $PrC$  remain weak performers, even with multi-scale fusion, reinforcing earlier findings that these features are less discriminative in isolation.



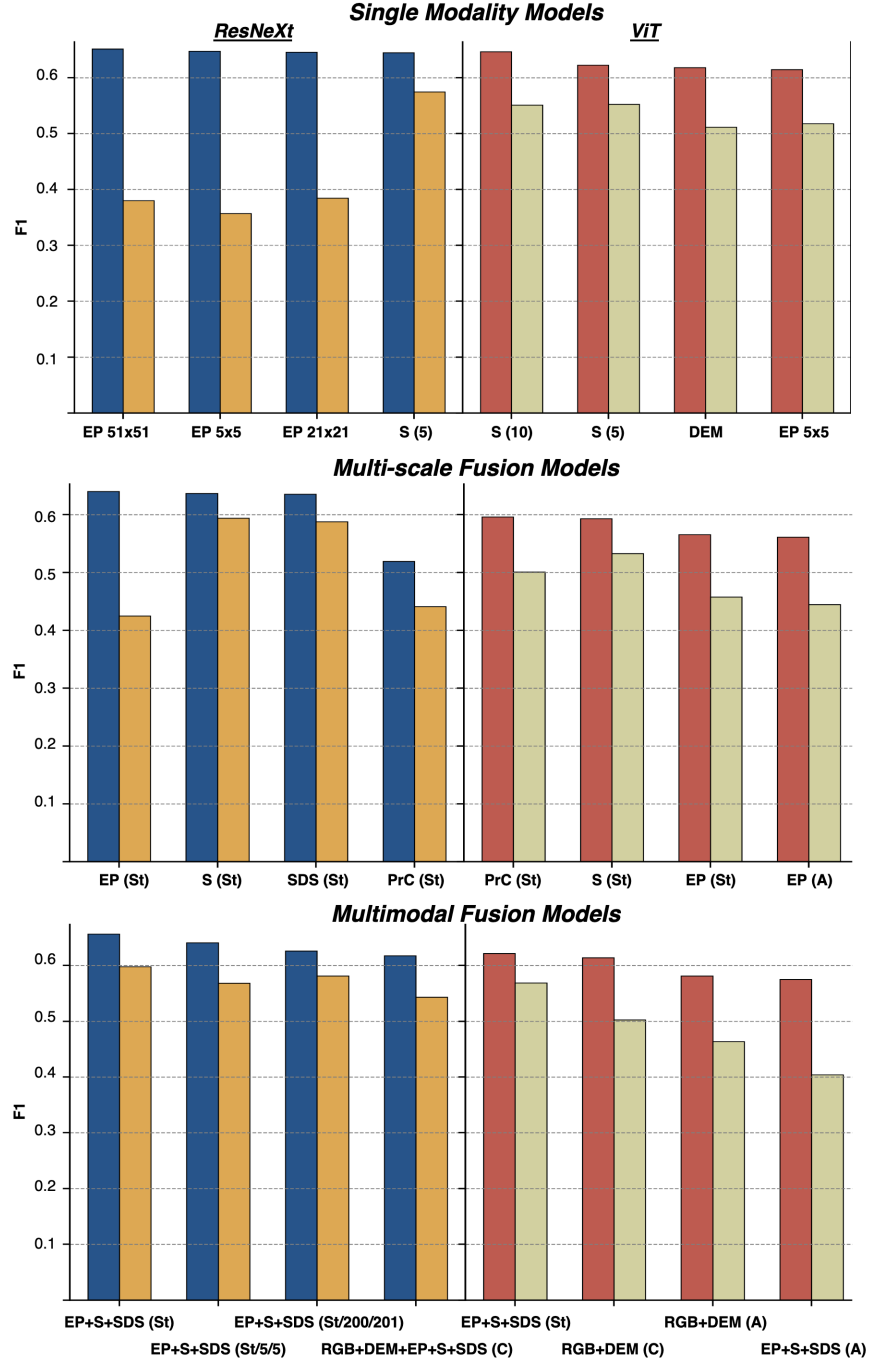


Figure 8: In-domain and cross-domain F1 scores for the top four models for single-modality (top), multi-scale fusion (middle), and multimodal fusion (bottom) experiments. Columns show comparisons of ResNeXt-50 (left) vs. ViT-B/16 (right) backbones. Each subplot shows the four best-performing models based on in-domain (Warren County) F1 scores (ResNeXt: blue; ViT: red). Cross-domain bars reflect the same models selected based on in-domain performance to illustrate domain shift (ResNeXt: orange; ViT: beige). Model names are shown beneath each group and indicate the input modality or modality combination. Spatial scale is denoted by the DEM resolution (for *S*, *PrC*, and *PrC*) or kernel size (for *EP* and *SDS*). For multi-scale and multimodal models, labels include the fusion strategy (St: early channel stacking, C: mid-level concatenation, A/A2: mid-level attention), followed by the DEM resolution and kernel size used for each modality.

Results using the ViT-B/16 backbone largely mirror those of ResNeXt-50, with early stacking again yielding modest gains in in-domain performance. However, profile curvature ( $PrC$ ) stands out as the best-performing multi-scale input with ViT, despite performing poorly with ResNeXt (Fig. 8). This contrast suggests that performance on certain terrain features may be more sensitive to architectural differences than fusion strategy alone. In all cases, attention-based fusion strategies fail to match the simplicity and effectiveness of early fusion via channel stacking.

**Multimodal Models:** Tables 10, 11, and 12 show the full results for multimodal fusion experiments across F1, AUC, precision, recall, mAP, and accuracy. Each experiment uses one of four fusion strategies: early channel stacking, mid-level concatenation, mid-level attention with a shared encoder, and mid-level attention with separate encoders. Fusion strategy plays a critical role in shaping both peak performance and generalization. Early channel stacking consistently delivers the highest in-domain results (e.g.,  $F1=0.657$  with ResNeXt-50), but also incurs a moderate domain gap ( $\Delta_{F1}=0.059$ ). In contrast, mid-level attention with a shared encoder yields lower in-domain performance (e.g.,  $F1=0.561$ ), but narrows the generalization gap ( $\Delta_{F1}=0.029$ ). Mid-level concatenation offers a balanced compromise, achieving moderate in-domain performance (e.g.,  $F1=0.596$ ), while maintaining the smallest domain shift observed ( $\Delta_{F1}=0.028$ ).

Backbone choice also influences outcomes. ResNeXt-50 typically provides a slight in-domain advantage over ViT-B/16 (e.g., 0.657 vs. 0.621) (Fig. 8). Modality ablation further highlights the importance of feature selection. Pairing RGB and DEM results in poor generalization, with domain gaps as large as  $\Delta_{F1}=0.211$ . In contrast, fusing engineered terrain features  $EP$ ,  $S$ , and  $SDS$  yields the best overall performance. This shape-centric modality combination achieves an in-domain F1 of 0.657 and reduces the cross-domain drop to just  $\Delta_{F1}=0.059$  (Fig. 8). Strong results persist for this configuration even when using single-scale versions of these features, underscoring their standalone value compared to raw elevation or overhead imagery (Fig. 8). While several single-modality models achieve strong in-domain performance, multimodal models slightly outperform them overall and, more importantly, generalize substantially better to unseen regions in cross-domain tests (Fig. 8).

**Per Class Performance Analysis:** Class-wise AUC analysis across backbones and fusion strategies reveals broadly consistent patterns across both ResNeXt-50 and ViT-B/16 backbones (Fig. 9; Tables 13 and 14). Units such as Qc, Qca, and Qr consistently achieve the highest discriminability, while Qal, Qat, and Qaf are more challenging to classify. Single-modality models typically yield the highest in-domain AUC scores, but exhibit greater sensitivity to domain shift. Both multi-scale and multimodal fusion models help reduce the generalization gap across most classes, though this improvement can occasionally come at the expense of peak in-domain performance. Interestingly, several units show better AUC on the cross-domain test set than in the in-domain evaluation, suggesting that the training region (Warren County) may be more geomorphologically complex than the test region (Hardin County), or that the models learn representations that transfer well despite this complexity.

Preferred input modalities also vary by class. For ResNeXt models, most classes favor  $S$ , though the spatial scale varies. For example, af1 and Qal perform best with  $S$  at the original 5ft/pixel resolution. Qat prefers  $EP$  at the smallest scale, consistent with identification from its relative height above active floodplains. Qc and Qca favor  $S$  as expected, though Qca benefits from a coarser 100ft/pixel resolution, which possibly reflects its broader landform morphology. Qr prefers  $S$ , aligning with its expression as a low-relief surficial map deposit. Surprisingly, Qaf performs best with  $PlC$  at the largest scale, despite being a small and spatially confined unit. This may indicate that large-scale curvature helps the model recognize depositional fans in a broader geomorphic context. With ViT-B/16, af1, Qal, and Qat perform best with  $EP$ , while Qaf prefers  $S$  at a moderate 50ft/pixel resolution. Qc and Qca again favor  $S$ , but at much different scales, similar to ResNeXt. Qr shows a unique preference for  $PrC$ , which is intuitive given that Qr typically forms in relatively flat areas.

In multimodal ResNeXt models, all classes perform best with the shape-centric input combination of  $EP + S + SDS$ , consistent with earlier findings. One exception is Qca, which prefers a larger input set that includes RGB and DEM. This is somewhat unexpected, as Qca is typically a slope-derived landform not visually distinct in aerial imagery. The benefit of RGB for Qca may be incidental or related to correlated infrastructure or vegetation patterns in the training region. For ViT-based multimodal models, fewer combinations were tested, but several meaningful trends emerge. af1 and Qat both favor RGB+DEM inputs, which is consistent with their distinctive visual patterns in aerial imagery and association with human-modified or floodplain-related terrain. The other five classes

all perform best with the  $EP + S + SDS$  combination, reinforcing the importance of terrain-based modalities even within transformer architectures.

Overall, these results highlight the complexity of class-specific modality preferences and suggest that optimal model configuration varies not only by architecture and fusion strategy, but also by the geomorphic expression and internal variability of each surficial geologic unit. While no single input combination or fusion method performs best across all units, the shape-derived terrain features  $EP$ ,  $S$ , and  $SDS$  remain consistently strong predictors of unit separability.

**Exploratory Comparisons with Existing Models:** We experimented with SatMAE and SatMAE++ models by fine-tuning their pre-trained models on our dataset. Similar to their models, we grouped the modalities into three groups. For each group, we selected the modalities that performed better in single-modality experiments. We formed group 1 with RGB and DEM, group 2 with  $EP\ 5 \times 5$ ,  $EP\ 51 \times 51$ ,  $EP\ 101 \times 101$ , and  $EP\ 201 \times 201$ , and group 3 with  $S\ (5)$ , and  $SDS\ 5 \times 5$ . We evaluated SatMAE and SatMAE++ on the same training, validation, and testing splits using these ten modalities. These models are designed for grouped multispectral satellite imagery and rely on representations that differ significantly from our LiDAR-based terrain features. As such, the SatMAE input configuration was not optimized for geological interpretation and differs in both sensor type and feature derivation from our modality sets. SatMAE achieved macro F1 scores of 0.614 (in-domain) and 0.427 (cross-domain), while SatMAE++ reached 0.656 and 0.454, respectively (Table 15). While these in-domain scores are comparable to our best ResNeXt-based models, cross-domain performance is substantially lower, highlighting the challenge of generalization in geologically diverse regions.

For context, we tested a similar input configuration with SGMap-Net, including RGB, DEM, and all six spatial scales of  $EP$ ,  $S$ , and  $SDS$ , using a ResNeXt backbone with both mid-level concatenation and attention-based fusion. These configurations achieved up to 0.618 F1 in-domain and 0.543 cross-domain. While SatMAE++ marginally outperformed this setup in-domain, its generalization gap was notably larger. These results suggest that masked autoencoder (MAE) architectures may offer strong in-domain performance but could benefit from pairing with geologically informed modality design and fusion strategies to improve robustness. We consider these comparisons exploratory and encourage future work on adapting foundation models for Earth surface analysis tasks.

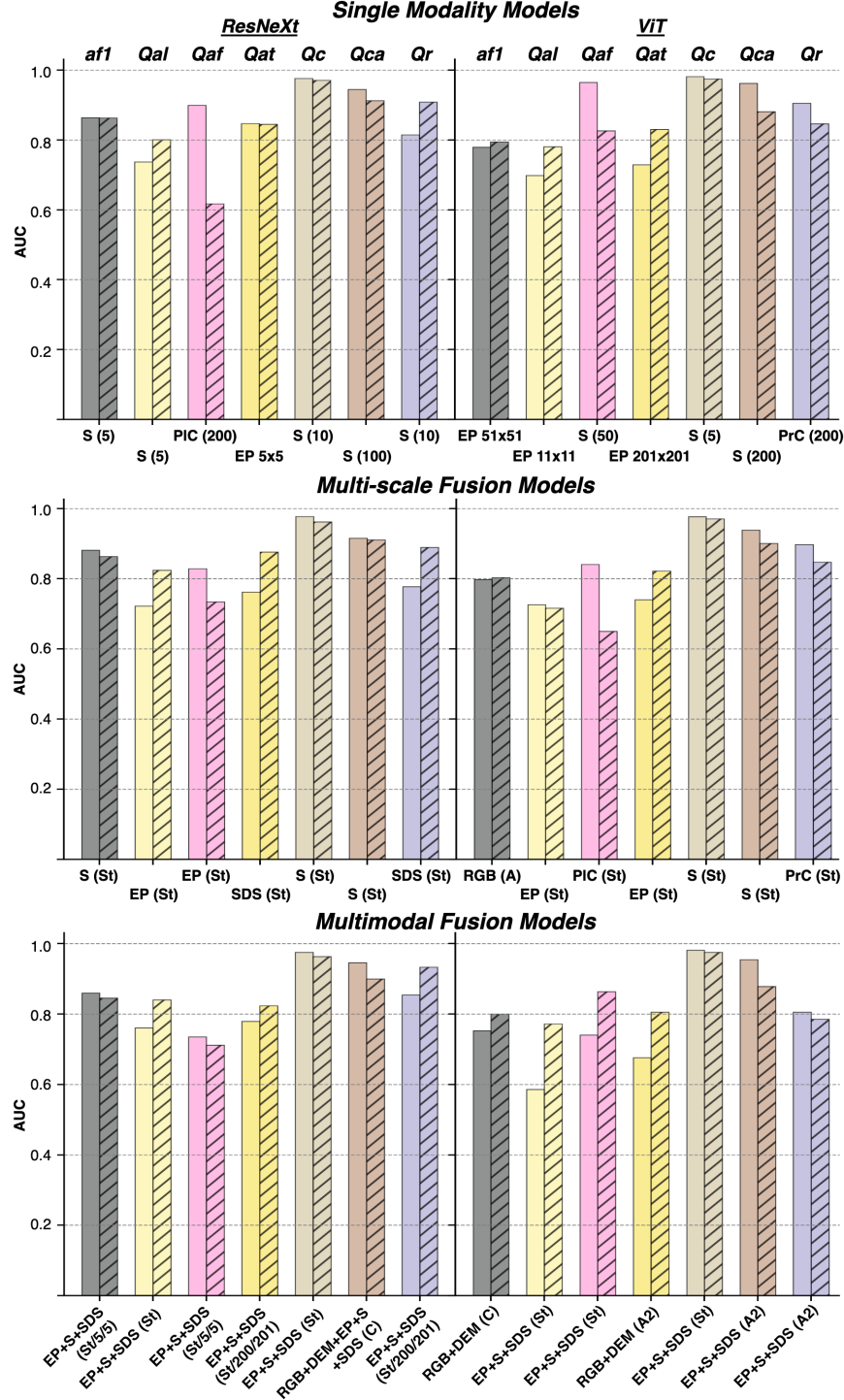


Figure 9: Class-wise AUC scores for the best-performing models across different experiment types and backbone architectures. Rows correspond to single-modality (top), multi-scale fusion (middle), and multimodal fusion (bottom) experiments. Columns represent ResNeXt-50 (left) and ViT-B/16 (right) backbones. Each bar pair shows the in-domain (solid) and cross-domain (hatched) AUC scores for the best model selected for each class based on in-domain performance. Model names, including modality and spatial scale, are shown below each bar group. Fusion strategies used in multi-scale and multimodal models are abbreviated in parentheses: early channel stacking (St), mid-level concatenation (C), and attention-based fusion (A, A2).

Table 4: Macro F1 and AUC for *single modality* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	F1 (ResNeXt)			F1 (ViT)			AUC (ResNeXt)			AUC (ViT)		
	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>RGB</i>	0.599	0.394	0.205	0.579	0.332	0.267	0.815	0.557	0.258	0.793	0.526	0.267
<i>NIR</i>	0.613	0.468	0.145	0.579	0.275	0.274	0.815	0.650	0.166	0.784	0.509	0.274
<i>DEM</i>	0.632	0.527	0.105	0.618	0.512	0.237	0.883	0.730	0.153	0.857	0.620	0.237
<i>EP</i> (101)	0.619	0.476	0.143	0.589	0.477	0.075	0.857	0.739	0.118	0.819	0.744	0.075
<i>EP</i> (11)	0.639	0.425	0.214	0.603	0.519	0.082	0.879	0.675	0.203	0.850	0.768	0.082
<i>EP</i> (201)	0.610	0.391	0.219	0.584	0.472	0.062	0.869	0.724	0.145	0.799	0.737	0.062
<i>EP</i> (21)	0.645	0.384	0.261	0.608	0.503	0.079	0.877	0.695	0.183	0.838	0.759	0.079
<i>EP</i> (51)	0.651	0.380	0.271	0.604	0.489	0.078	0.876	0.663	0.213	0.835	0.757	0.078
<i>EP</i> (5)	0.648	0.357	0.291	0.614	0.518	0.117	0.872	0.582	0.290	0.854	0.738	0.117
<i>PlC</i> (10)	0.494	0.426	0.068	0.524	0.457	0.007	0.501	0.500	0.001	0.621	0.614	0.007
<i>PlC</i> (100)	0.488	0.420	0.068	0.484	0.422	-0.008	0.511	0.470	0.041	0.532	0.540	-0.008
<i>PlC</i> (20)	0.495	0.425	0.070	0.513	0.453	0.005	0.488	0.485	0.002	0.632	0.627	0.005
<i>PlC</i> (200)	0.488	0.433	0.055	0.495	0.427	-0.039	0.474	0.528	-0.054	0.500	0.539	-0.039
<i>PlC</i> (50)	0.488	0.425	0.063	0.495	0.426	0.016	0.472	0.459	0.013	0.560	0.544	0.016
<i>PlC</i> (5)	0.491	0.425	0.066	0.517	0.452	0.013	0.514	0.513	0.001	0.603	0.590	0.013
<i>PrC</i> (10)	0.492	0.421	0.071	0.497	0.425	0.023	0.486	0.520	-0.034	0.517	0.493	0.023
<i>PrC</i> (100)	0.510	0.418	0.092	0.540	0.431	0.035	0.553	0.491	0.062	0.613	0.578	0.035
<i>PrC</i> (20)	0.496	0.415	0.081	0.495	0.426	-0.055	0.508	0.463	0.046	0.389	0.444	-0.055
<i>PrC</i> (200)	0.495	0.425	0.071	0.549	0.431	0.028	0.417	0.428	-0.011	0.626	0.599	0.028
<i>PrC</i> (50)	0.492	0.417	0.074	0.494	0.426	-0.022	0.440	0.398	0.042	0.466	0.487	-0.022
<i>PrC</i> (5)	0.493	0.433	0.060	0.494	0.426	-0.039	0.554	0.516	0.038	0.407	0.446	-0.039
<i>S</i> (10)	0.619	0.570	0.049	0.647	0.551	0.127	0.875	0.779	0.096	0.841	0.713	0.127
<i>S</i> (100)	0.594	0.536	0.058	0.578	0.528	0.061	0.811	0.710	0.102	0.765	0.705	0.061
<i>S</i> (20)	0.617	0.555	0.061	0.614	0.555	0.102	0.861	0.804	0.057	0.833	0.731	0.102
<i>S</i> (200)	0.543	0.485	0.058	0.578	0.514	0.093	0.601	0.578	0.023	0.770	0.676	0.093
<i>S</i> (50)	0.612	0.537	0.075	0.600	0.554	0.081	0.841	0.744	0.096	0.812	0.731	0.081
<i>S</i> (5)	0.645	0.575	0.070	0.623	0.552	0.093	0.876	0.808	0.068	0.855	0.762	0.093
<i>SDS</i> (101)	0.611	0.571	0.040	0.535	0.502	0.037	0.848	0.756	0.092	0.718	0.681	0.037
<i>SDS</i> (11)	0.631	0.575	0.056	0.599	0.543	0.080	0.846	0.786	0.061	0.803	0.723	0.080
<i>SDS</i> (201)	0.613	0.527	0.086	0.548	0.508	0.064	0.837	0.713	0.124	0.735	0.671	0.064
<i>SDS</i> (21)	0.633	0.573	0.060	0.591	0.552	0.074	0.854	0.786	0.067	0.809	0.735	0.074
<i>SDS</i> (51)	0.603	0.533	0.069	0.554	0.536	0.038	0.841	0.746	0.095	0.727	0.689	0.038
<i>SDS</i> (5)	0.613	0.567	0.045	0.569	0.513	0.072	0.850	0.804	0.046	0.786	0.713	0.072



Table 5: Macro precision and recall for *single modality* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Precision (ResNeXt)			Precision (ViT)			Recall (ResNeXt)			Recall (ViT)		
	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>RGB</i>	0.553	0.405	0.148	0.522	0.296	0.235	0.672	0.418	0.254	0.664	0.429	0.235
<i>NIR</i>	0.564	0.486	0.078	0.521	0.273	0.384	0.698	0.514	0.184	0.668	0.284	0.384
<i>DEM</i>	0.621	0.460	0.161	0.551	0.432	0.125	0.661	0.653	0.008	0.800	0.674	0.125
<i>EP</i> (101)	0.570	0.480	0.090	0.539	0.421	0.102	0.727	0.551	0.176	0.674	0.572	0.102
<i>EP</i> (11)	0.602	0.474	0.128	0.552	0.449	0.060	0.748	0.428	0.320	0.690	0.631	0.060
<i>EP</i> (201)	0.593	0.465	0.127	0.520	0.425	0.092	0.634	0.364	0.270	0.707	0.615	0.092
<i>EP</i> (21)	0.629	0.455	0.173	0.548	0.435	0.089	0.737	0.416	0.321	0.706	0.617	0.089
<i>EP</i> (51)	0.612	0.382	0.230	0.565	0.440	0.087	0.705	0.389	0.316	0.664	0.577	0.087
<i>EP</i> (5)	0.617	0.450	0.167	0.556	0.452	0.112	0.706	0.333	0.373	0.733	0.621	0.112
<i>PlC</i> (10)	0.391	0.333	0.059	0.432	0.370	0.119	1.000	1.000	0.000	0.871	0.752	0.119
<i>PlC</i> (100)	0.390	0.332	0.058	0.392	0.334	-0.045	0.856	0.809	0.047	0.795	0.840	-0.045
<i>PlC</i> (20)	0.393	0.333	0.060	0.429	0.365	0.052	0.892	0.889	0.003	0.853	0.801	0.052
<i>PlC</i> (200)	0.389	0.337	0.052	0.393	0.335	-0.022	0.842	0.921	-0.079	0.973	0.995	-0.022
<i>PlC</i> (50)	0.390	0.334	0.057	0.403	0.338	-0.029	0.823	0.834	-0.010	0.765	0.794	-0.029
<i>PlC</i> (5)	0.390	0.333	0.057	0.419	0.359	0.078	0.837	0.829	0.007	0.806	0.728	0.078
<i>PrC</i> (10)	0.394	0.335	0.059	0.406	0.336	0.000	0.819	0.853	-0.034	0.919	0.919	0.000
<i>PrC</i> (100)	0.430	0.337	0.092	0.456	0.348	0.074	0.679	0.639	0.040	0.731	0.657	0.074
<i>PrC</i> (20)	0.396	0.328	0.068	0.392	0.333	-0.001	0.739	0.719	0.020	0.997	0.998	-0.001
<i>PrC</i> (200)	0.392	0.332	0.060	0.464	0.350	0.100	0.896	0.854	0.042	0.748	0.648	0.100
<i>PrC</i> (50)	0.392	0.331	0.061	0.391	0.333	0.000	0.759	0.718	0.041	1.000	1.000	0.000
<i>PrC</i> (5)	0.392	0.341	0.052	0.391	0.333	0.000	0.967	0.946	0.021	1.000	1.000	0.000
<i>S</i> (10)	0.590	0.507	0.084	0.614	0.490	0.041	0.654	0.662	-0.009	0.693	0.653	0.041
<i>S</i> (100)	0.523	0.464	0.059	0.508	0.464	0.054	0.744	0.679	0.065	0.717	0.663	0.054
<i>S</i> (20)	0.592	0.497	0.095	0.553	0.491	0.072	0.670	0.671	0.000	0.791	0.720	0.072
<i>S</i> (200)	0.469	0.409	0.060	0.500	0.436	0.064	0.697	0.651	0.047	0.736	0.672	0.064
<i>S</i> (50)	0.550	0.478	0.072	0.537	0.484	-0.027	0.749	0.664	0.085	0.774	0.801	-0.027
<i>S</i> (5)	0.616	0.506	0.110	0.578	0.489	0.051	0.681	0.687	-0.006	0.726	0.674	0.051
<i>SDS</i> (101)	0.566	0.490	0.075	0.459	0.409	-0.009	0.775	0.716	0.058	0.710	0.719	-0.009
<i>SDS</i> (11)	0.596	0.499	0.097	0.545	0.460	0.084	0.689	0.698	-0.008	0.769	0.685	0.084
<i>SDS</i> (201)	0.558	0.452	0.107	0.459	0.411	0.044	0.709	0.660	0.048	0.796	0.752	0.044
<i>SDS</i> (21)	0.578	0.486	0.092	0.529	0.469	-0.006	0.768	0.740	0.027	0.690	0.696	-0.006
<i>SDS</i> (51)	0.578	0.471	0.108	0.482	0.443	0.022	0.638	0.646	-0.008	0.740	0.718	0.022
<i>SDS</i> (5)	0.580	0.487	0.093	0.518	0.435	-0.025	0.661	0.707	-0.047	0.641	0.666	-0.025

Table 6: Mean average precision (mAP) and accuracy for *single modality* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	mAP (ResNeXt)			mAP (ViT)			Accuracy (ResNeXt)			Accuracy (ViT)		
	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>RGB</i>	0.509	0.367	0.143	0.489	0.336	0.109	0.832	0.781	0.051	0.815	0.706	0.109
<i>NIR</i>	0.513	0.387	0.125	0.485	0.337	0.020	0.833	0.809	0.025	0.812	0.792	0.020
<i>DEM</i>	0.554	0.442	0.111	0.516	0.431	0.022	0.873	0.827	0.046	0.808	0.785	0.022
<i>EP</i> (101)	0.528	0.401	0.128	0.500	0.385	0.034	0.835	0.812	0.024	0.818	0.784	0.034
<i>EP</i> (11)	0.551	0.397	0.154	0.510	0.409	0.024	0.854	0.832	0.022	0.829	0.805	0.024
<i>EP</i> (201)	0.535	0.381	0.154	0.476	0.367	0.041	0.858	0.838	0.019	0.791	0.750	0.041
<i>EP</i> (21)	0.565	0.386	0.179	0.504	0.398	0.029	0.860	0.828	0.031	0.827	0.798	0.029
<i>EP</i> (51)	0.546	0.377	0.169	0.507	0.395	0.034	0.862	0.818	0.044	0.837	0.803	0.034
<i>EP</i> (5)	0.549	0.385	0.164	0.516	0.417	0.019	0.858	0.831	0.026	0.829	0.810	0.019
<i>PlC</i> (10)	0.391	0.333	0.059	0.418	0.353	0.005	0.392	0.333	0.059	0.631	0.626	0.005
<i>PlC</i> (100)	0.392	0.333	0.059	0.392	0.334	0.064	0.524	0.467	0.057	0.586	0.521	0.064
<i>PlC</i> (20)	0.393	0.333	0.060	0.416	0.353	-0.001	0.494	0.452	0.043	0.617	0.619	-0.001
<i>PlC</i> (200)	0.390	0.335	0.055	0.393	0.335	0.062	0.525	0.471	0.054	0.456	0.395	0.062
<i>PlC</i> (50)	0.391	0.334	0.057	0.397	0.335	0.053	0.533	0.482	0.051	0.644	0.591	0.053
<i>PlC</i> (5)	0.391	0.333	0.058	0.411	0.354	0.015	0.551	0.502	0.049	0.643	0.628	0.015
<i>PrC</i> (10)	0.393	0.332	0.060	0.400	0.334	0.051	0.527	0.466	0.061	0.452	0.401	0.051
<i>PrC</i> (100)	0.406	0.339	0.067	0.431	0.345	0.055	0.714	0.674	0.040	0.726	0.671	0.055
<i>PrC</i> (20)	0.392	0.333	0.059	0.392	0.333	0.062	0.645	0.581	0.064	0.395	0.334	0.062
<i>PrC</i> (200)	0.392	0.333	0.059	0.433	0.345	0.045	0.510	0.463	0.047	0.723	0.677	0.045
<i>PrC</i> (50)	0.393	0.334	0.059	0.391	0.333	0.059	0.644	0.591	0.054	0.392	0.333	0.059
<i>PrC</i> (5)	0.392	0.340	0.052	0.391	0.333	0.059	0.411	0.402	0.009	0.392	0.333	0.059
<i>S</i> (10)	0.543	0.472	0.071	0.542	0.465	0.025	0.867	0.852	0.015	0.850	0.825	0.025
<i>S</i> (100)	0.501	0.447	0.053	0.485	0.452	-0.001	0.793	0.784	0.009	0.792	0.793	-0.001
<i>S</i> (20)	0.539	0.463	0.077	0.523	0.466	0.019	0.857	0.844	0.013	0.812	0.793	0.019
<i>S</i> (200)	0.450	0.398	0.052	0.481	0.435	0.003	0.742	0.752	-0.010	0.784	0.780	0.003
<i>S</i> (50)	0.517	0.455	0.062	0.506	0.463	0.012	0.807	0.799	0.008	0.794	0.781	0.012
<i>S</i> (5)	0.552	0.468	0.084	0.525	0.456	0.021	0.871	0.848	0.023	0.840	0.819	0.021
<i>SDS</i> (101)	0.525	0.461	0.064	0.448	0.400	-0.017	0.820	0.808	0.012	0.734	0.751	-0.017
<i>SDS</i> (11)	0.533	0.466	0.068	0.504	0.434	0.011	0.850	0.839	0.011	0.806	0.795	0.011
<i>SDS</i> (201)	0.520	0.427	0.093	0.446	0.402	-0.019	0.834	0.805	0.030	0.710	0.729	-0.019
<i>SDS</i> (21)	0.531	0.454	0.078	0.491	0.435	0.007	0.836	0.819	0.017	0.816	0.809	0.007
<i>SDS</i> (51)	0.529	0.436	0.093	0.459	0.418	0.002	0.855	0.824	0.031	0.754	0.752	0.002
<i>SDS</i> (5)	0.527	0.459	0.068	0.484	0.420	0.011	0.853	0.833	0.020	0.820	0.809	0.011

Table 7: Macro F1 and AUC for *multi-scale fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	F1 (ResNeXt)			F1 (ViT)			AUC (ResNeXt)			AUC (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP</i>	A	0.494	0.426	0.068	0.561	0.445	0.117	0.500	0.500	0.000	0.759	0.664	0.095
<i>EP</i>	St	0.640	0.425	0.215	0.566	0.458	0.108	0.862	0.717	0.145	0.756	0.693	0.063
<i>PlC</i>	A	0.494	0.426	0.068	0.505	0.435	0.070	0.500	0.500	0.000	0.578	0.581	-0.003
<i>PlC</i>	St	0.490	0.426	0.063	0.493	0.429	0.063	0.525	0.521	0.004	0.511	0.536	-0.026
<i>PrC</i>	A	0.494	0.426	0.068	0.531	0.410	0.121	0.500	0.500	0.000	0.594	0.562	0.032
<i>PrC</i>	St	0.519	0.441	0.078	0.596	0.501	0.095	0.579	0.497	0.082	0.816	0.727	0.089
<i>S</i>	A	0.494	0.426	0.068	0.557	0.519	0.038	0.500	0.500	0.000	0.615	0.594	0.021
<i>S</i>	St	0.637	0.594	0.043	0.593	0.533	0.061	0.864	0.804	0.061	0.798	0.705	0.093
<i>SDS</i>	A	0.493	0.451	0.042	0.494	0.426	0.068	0.618	0.618	0.001	0.500	0.500	0.000
<i>SDS</i>	St	0.636	0.588	0.048	0.619	0.571	0.048	0.878	0.792	0.086	0.672	0.644	0.028

Table 8: Macro precision and recall for *multi-scale fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	Precision (ResNeXt)			Precision			Recall (ResNeXt)			Recall (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP</i>	A	0.391	0.333	0.059	0.483	0.375	0.108	1.000	1.000	0.000	0.700	0.612	0.088
<i>EP</i>	St	0.606	0.556	0.051	0.493	0.380	0.112	0.703	0.426	0.277	0.712	0.636	0.076
<i>PlC</i>	A	0.391	0.333	0.059	0.405	0.341	0.064	1.000	1.000	0.000	0.874	0.868	0.006
<i>PlC</i>	St	0.391	0.335	0.056	0.391	0.335	0.056	0.738	0.738	0.000	0.872	0.940	-0.067
<i>PrC</i>	A	0.391	0.333	0.059	0.431	0.325	0.106	1.000	1.000	0.000	0.738	0.678	0.060
<i>PrC</i>	St	0.429	0.353	0.076	0.530	0.435	0.095	0.697	0.694	0.003	0.743	0.642	0.101
<i>S</i>	A	0.391	0.333	0.058	0.489	0.440	0.049	1.000	1.000	0.000	0.745	0.688	0.057
<i>S</i>	St	0.607	0.535	0.072	0.525	0.455	0.070	0.730	0.682	0.047	0.714	0.681	0.033
<i>SDS</i>	A	0.432	0.380	0.052	0.391	0.332	0.057	0.801	0.748	0.053	1.000	1.000	0.000
<i>SDS</i>	St	0.588	0.509	0.079	0.575	0.472	0.103	0.742	0.729	0.013	0.675	0.674	0.001

Table 9: Mean average precision (mAP) and accuracy for *multi-scale fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	mAP (ResNeXt)			mAP (ViT)			Accuracy (ResNeXt)			Accuracy (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP</i>	A	0.391	0.333	0.059	0.450	0.362	0.088	0.391	0.333	0.059	0.766	0.727	0.039
<i>EP</i>	St	0.555	0.403	0.152	0.460	0.360	0.099	0.865	0.828	0.037	0.774	0.724	0.050
<i>PlC</i>	A	0.391	0.333	0.059	0.401	0.338	0.062	0.391	0.333	0.059	0.598	0.541	0.057
<i>PlC</i>	St	0.392	0.335	0.057	0.392	0.335	0.057	0.634	0.588	0.046	0.534	0.465	0.069
<i>PrC</i>	A	0.391	0.333	0.059	0.407	0.333	0.074	0.391	0.333	0.059	0.691	0.625	0.065
<i>PrC</i>	St	0.416	0.348	0.069	0.504	0.423	0.081	0.717	0.666	0.051	0.794	0.768	0.027
<i>S</i>	A	0.391	0.333	0.058	0.472	0.434	0.038	0.391	0.333	0.059	0.742	0.747	-0.005
<i>S</i>	St	0.557	0.491	0.066	0.498	0.453	0.045	0.856	0.860	-0.004	0.810	0.803	0.006
<i>SDS</i>	A	0.416	0.357	0.059	0.391	0.333	0.058	0.630	0.666	-0.036	0.391	0.333	0.058
<i>SDS</i>	St	0.540	0.470	0.070	0.522	0.447	0.075	0.846	0.839	0.007	0.851	0.826	0.025

Table 10: Macro F1 and AUC for *multimodal fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	F1 (ResNeXt)			F1 (ViT)			AUC (ResNeXt)			AUC (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP+S+SDS</i>	A	0.561	0.532	0.029	0.567	0.538	0.029	0.677	0.707	-0.030	0.776	0.678	0.098
<i>EP+S+SDS</i>	A2	0.561	0.532	0.029	-	-	-	0.677	0.707	-0.030	-	-	-
<i>EP+S+SDS</i>	C	0.596	0.569	0.028	-	-	-	0.829	0.750	0.079	-	-	-
<i>EP+S+SDS</i>	St	0.657	0.598	0.059	0.621	0.569	0.053	0.882	0.806	0.076	0.860	0.774	0.086
<i>EP(5)+S(5)+SDS(5)</i>	St	0.641	0.568	0.073	-	-	-	0.848	0.812	0.036	-	-	-
<i>EP(201)+S(200)+SDS(201)</i>	St	0.626	0.582	0.045	-	-	-	0.885	0.812	0.073	-	-	-
<i>RGB+DEM</i>	A	0.551	0.457	0.094	0.575	0.404	0.171	0.714	0.552	0.163	0.787	0.622	0.165
<i>RGB+DEM</i>	A2	0.559	0.474	0.085	0.581	0.464	0.118	0.763	0.641	0.122	0.810	0.724	0.085
<i>RGB+DEM</i>	C	0.600	0.389	0.211	0.614	0.503	0.111	0.808	0.535	0.273	0.870	0.721	0.149
<i>RGB+DEM+EP+S+SDS</i>	A2	0.494	0.426	0.068	-	-	-	0.500	0.500	0.000	-	-	-
<i>RGB+DEM+EP+S+SDS</i>	C	0.618	0.543	0.074	0.621	0.528	0.093	0.858	0.739	0.118	0.735	0.615	0.120

Table 11: Macro precision and recall for *multimodal fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	Precision (ResNeXt)			Precision (ViT)			Recall (ResNeXt)			Recall (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP+S+SDS</i>	A	0.487	0.451	0.036	0.507	0.466	0.041	0.734	0.723	0.011	0.752	0.693	0.059
<i>EP+S+SDS</i>	A2	0.487	0.451	0.036	-	-	-	0.734	0.723	0.011	-	-	-
<i>EP+S+SDS</i>	C	0.542	0.529	0.013	-	-	-	0.694	0.640	0.054	-	-	-
<i>EP+S+SDS</i>	St	0.626	0.546	0.080	0.568	0.491	0.077	0.735	0.666	0.068	0.761	0.711	0.050
<i>EP(5)+S(5)+SDS(5)</i>	St	0.606	0.531	0.074	-	-	-	0.697	0.623	0.074	-	-	-
<i>EP(201)+S(200)+SDS(201)</i>	St	0.588	0.529	0.059	-	-	-	0.721	0.674	0.048	-	-	-
<i>RGB+DEM</i>	A	0.495	0.445	0.050	0.515	0.387	0.129	0.647	0.555	0.092	0.686	0.582	0.105
<i>RGB+DEM</i>	A2	0.498	0.411	0.087	0.513	0.434	0.079	0.656	0.595	0.061	0.720	0.607	0.113
<i>RGB+DEM</i>	C	0.537	0.373	0.163	0.558	0.420	0.137	0.715	0.437	0.278	0.706	0.661	0.045
<i>RGB+DEM+EP+S+SDS</i>	A2	0.391	0.333	0.059	-	-	-	1.000	1.000	0.000	-	-	-
<i>RGB+DEM+EP+S+SDS</i>	C	0.563	0.496	0.067	0.574	0.485	0.090	0.740	0.644	0.096	0.621	0.622	-0.001

Table 12: Mean average precision (mAP) and accuracy for *multimodal fusion* models on in-domain (WC) and cross-domain (HC) test sets. Results are shown for ResNeXt-50 and ViT-B/16 backbones using four fusion strategies: channel stacking (St), concatenation (C), and attention with shared (A) or separate encoders (A2). WC–HC differences ( $\Delta$ ) are also reported. Spatial scale is noted in parentheses as resolution (feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	Fusion	mAP (ResNeXt)			mAP (ViT)			Accuracy (ResNeXt)			Accuracy (ViT)		
		WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$	WC	HC	$\Delta$
<i>EP+S+SDS</i>	A	0.474	0.442	0.033	0.488	0.456	0.032	0.747	0.758	-0.011	0.750	0.752	-0.002
<i>EP+S+SDS</i>	A2	0.474	0.442	0.033	-	-	-	0.747	0.758	-0.011	-	-	-
<i>EP+S+SDS</i>	C	0.505	0.451	0.053	-	-	-	0.822	0.836	-0.015	-	-	-
<i>EP+S+SDS</i>	St	0.571	0.495	0.076	0.534	0.463	0.070	0.875	0.867	0.008	0.834	0.823	0.011
<i>EP(5)+S(5)+SDS(5)</i>	St	0.551	0.471	0.080	-	-	-	0.865	0.856	0.009	-	-	-
<i>EP(201)+S(200)+SDS(201)</i>	St	0.552	0.480	0.072	-	-	-	0.858	0.852	0.006	-	-	-
<i>RGB+DEM</i>	A	0.459	0.389	0.070	0.478	0.360	0.118	0.784	0.776	0.008	0.799	0.745	0.054
<i>RGB+DEM</i>	A2	0.464	0.389	0.075	0.486	0.388	0.098	0.795	0.793	0.002	0.795	0.775	0.020
<i>RGB+DEM</i>	C	0.495	0.360	0.135	0.524	0.415	0.109	0.815	0.809	0.007	0.838	0.796	0.042
<i>RGB+DEM+EP+S+SDS</i>	A2	0.391	0.333	0.059	-	-	-	0.391	0.333	0.059	-	-	-
<i>RGB+DEM+EP+S+SDS</i>	C	0.525	0.458	0.067	0.537	0.449	0.088	0.833	0.805	0.028	0.827	0.824	0.003

Table 13: Class-wise AUC scores for in-domain (Warren County region) performance across single-modality, multi-scale fusion, and multimodal fusion models, using both ResNeXt-50 and ViT-B/16 backbones. Four fusion strategies are shown: early channel stacking (St), mid-level concatenation (C), mid-level attention with a shared encoder (A), and mid-level attention with separate encoders (A2). Spatial scale is indicated in parentheses, either as DEM resolution (in feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	ResNeXt							ViT						
	af1	Qal	Qaf	Qat	Qc	Qca	Qr	af1	Qal	Qaf	Qat	Qc	Qca	Qr
<i>RGB</i>	0.834	0.713	0.684	0.815	0.912	0.857	0.886	0.816	0.679	0.744	0.780	0.891	0.834	0.805
<i>NIR</i>	0.816	0.698	0.782	0.793	0.907	0.866	0.842	0.760	0.664	0.797	0.799	0.886	0.816	0.763
<i>DEM</i>	0.845	0.832	0.820	0.887	0.964	0.922	0.910	0.663	0.771	0.926	0.871	0.956	0.923	0.888
<i>EP</i> (101)	0.853	0.806	0.759	0.886	0.904	0.904	0.890	0.757	0.751	0.758	0.860	0.850	0.884	0.874
<i>EP</i> (11)	0.868	0.816	0.833	0.888	0.936	0.905	0.902	0.778	0.781	0.834	0.882	0.891	0.889	0.898
<i>EP</i> (201)	0.846	0.812	0.844	0.879	0.901	0.894	0.904	0.734	0.750	0.756	0.830	0.789	0.872	0.864
<i>EP</i> (21)	0.856	0.807	0.842	0.883	0.945	0.908	0.900	0.783	0.776	0.799	0.858	0.888	0.885	0.880
<i>EP</i> (51)	0.860	0.825	0.827	0.870	0.921	0.906	0.924	0.794	0.766	0.791	0.858	0.877	0.888	0.870
<i>EP</i> (5)	0.837	0.805	0.845	0.845	0.947	0.905	0.920	0.791	0.783	0.838	0.865	0.914	0.885	0.903
<i>PlC</i> (100)	0.517	0.490	0.604	0.473	0.501	0.524	0.465	0.469	0.567	0.650	0.515	0.531	0.529	0.465
<i>PlC</i> (10)	0.501	0.501	0.500	0.500	0.501	0.501	0.500	0.445	0.494	0.769	0.675	0.499	0.773	0.689
<i>PlC</i> (200)	0.462	0.413	0.617	0.414	0.479	0.494	0.439	0.461	0.627	0.620	0.382	0.524	0.482	0.402
<i>PlC</i> (20)	0.459	0.516	0.491	0.497	0.455	0.505	0.490	0.451	0.478	0.746	0.712	0.668	0.719	0.649
<i>PlC</i> (50)	0.526	0.505	0.362	0.387	0.547	0.476	0.500	0.466	0.523	0.655	0.620	0.578	0.575	0.505
<i>PlC</i> (5)	0.440	0.491	0.610	0.515	0.513	0.514	0.516	0.438	0.509	0.719	0.610	0.575	0.725	0.645
<i>PrC</i> (100)	0.515	0.432	0.465	0.608	0.530	0.681	0.640	0.501	0.341	0.523	0.845	0.512	0.738	0.833
<i>PrC</i> (10)	0.549	0.555	0.324	0.537	0.341	0.554	0.539	0.545	0.501	0.630	0.400	0.420	0.613	0.508
<i>PrC</i> (200)	0.482	0.499	0.494	0.244	0.473	0.474	0.253	0.511	0.326	0.558	0.859	0.601	0.682	0.846
<i>PrC</i> (20)	0.526	0.494	0.445	0.503	0.472	0.539	0.579	0.493	0.602	0.487	0.190	0.541	0.224	0.186
<i>PrC</i> (50)	0.443	0.423	0.602	0.522	0.145	0.377	0.567	0.501	0.429	0.477	0.378	0.501	0.499	0.476
<i>PrC</i> (5)	0.465	0.566	0.569	0.473	0.564	0.516	0.724	0.444	0.545	0.546	0.236	0.501	0.347	0.233
<i>S</i> (100)	0.619	0.750	0.803	0.831	0.957	0.912	0.807	0.623	0.707	0.791	0.725	0.947	0.869	0.696
<i>S</i> (10)	0.816	0.805	0.840	0.870	0.971	0.915	0.908	0.770	0.759	0.772	0.829	0.975	0.910	0.868
<i>S</i> (200)	0.416	0.535	0.681	0.595	0.838	0.815	0.324	0.626	0.666	0.818	0.750	0.909	0.880	0.738
<i>S</i> (20)	0.778	0.809	0.764	0.877	0.974	0.921	0.905	0.718	0.765	0.809	0.853	0.975	0.910	0.803
<i>S</i> (50)	0.648	0.788	0.842	0.873	0.966	0.926	0.842	0.641	0.750	0.826	0.796	0.974	0.908	0.789
<i>S</i> (5)	0.863	0.800	0.813	0.870	0.968	0.905	0.910	0.794	0.748	0.853	0.854	0.974	0.900	0.864
<i>SDS</i> (101)	0.814	0.732	0.860	0.813	0.964	0.882	0.874	0.659	0.608	0.804	0.659	0.891	0.751	0.655
<i>SDS</i> (11)	0.839	0.751	0.774	0.866	0.946	0.877	0.871	0.792	0.671	0.817	0.757	0.933	0.853	0.800
<i>SDS</i> (201)	0.802	0.679	0.812	0.833	0.967	0.897	0.870	0.633	0.605	0.855	0.666	0.913	0.741	0.729
<i>SDS</i> (21)	0.842	0.750	0.842	0.841	0.953	0.889	0.860	0.769	0.685	0.853	0.767	0.934	0.837	0.816
<i>SDS</i> (51)	0.832	0.719	0.851	0.800	0.951	0.883	0.852	0.675	0.620	0.777	0.684	0.889	0.759	0.689
<i>SDS</i> (5)	0.855	0.733	0.789	0.860	0.944	0.890	0.883	0.772	0.665	0.800	0.757	0.921	0.833	0.751
<i>EP</i> (St)	0.823	0.824	0.734	0.878	0.945	0.911	0.917	0.823	0.824	0.734	0.878	0.945	0.911	0.917
<i>EP</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PlC</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PlC</i> (St)	0.504	0.500	0.641	0.501	0.514	0.500	0.514	0.504	0.500	0.641	0.501	0.514	0.500	0.514
<i>PrC</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PrC</i> (St)	0.494	0.653	0.567	0.721	0.628	0.791	0.201	0.494	0.653	0.567	0.721	0.628	0.791	0.201
<i>S</i> (A)	0.499	0.501	0.500	0.500	0.501	0.499	0.500	0.499	0.501	0.500	0.500	0.501	0.499	0.500
<i>S</i> (St)	0.863	0.787	0.760	0.870	0.962	0.911	0.900	0.863	0.787	0.760	0.870	0.962	0.911	0.900
<i>SDS</i> (A)	0.552	0.576	0.801	0.602	0.679	0.540	0.580	0.552	0.576	0.801	0.602	0.679	0.540	0.580
<i>SDS</i> (St)	0.839	0.766	0.917	0.876	0.964	0.898	0.889	0.839	0.766	0.917	0.876	0.964	0.898	0.889
<i>EP+S+SDS</i> (A)	0.486	0.575	0.726	0.641	0.930	0.784	0.599	0.698	0.623	0.831	0.660	0.961	0.879	0.785
<i>EP+S+SDS</i> (A2)	0.486	0.575	0.726	0.641	0.930	0.784	0.599	-	-	-	-	-	-	-
<i>EP+S+SDS</i> (C)	0.723	0.802	0.746	0.809	0.959	0.879	0.885	-	-	-	-	-	-	-
<i>EP+S+SDS</i> (St)	0.866	0.840	0.790	0.858	0.975	0.913	0.933	0.780	0.772	0.864	0.847	0.976	0.890	0.890
<i>EP</i> (201)+ <i>S</i> (200)+ <i>SDS</i> (201) (St)	0.846	0.802	0.840	0.903	0.961	0.911	0.933	-	-	-	-	-	-	-
<i>EP</i> (5)+ <i>S</i> (5)+ <i>SDS</i> (5) (St)	0.845	0.797	0.712	0.829	0.964	0.904	0.886	-	-	-	-	-	-	-
<i>RGB+DEM</i> (A)	0.687	0.476	0.747	0.762	0.837	0.801	0.692	0.711	0.629	0.813	0.815	0.886	0.842	0.811
<i>RGB+DEM</i> (A2)	0.752	0.617	0.780	0.816	0.825	0.786	0.764	0.704	0.692	0.856	0.806	0.930	0.841	0.839
<i>RGB+DEM</i> (C)	0.821	0.708	0.804	0.803	0.871	0.845	0.803	0.800	0.756	0.874	0.899	0.949	0.901	0.911
<i>RGB+DEM+EP+S+SDS</i> (A2)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	-	-	-	-	-	-	-
<i>RGB+DEM+EP+S+SDS</i> (C)	0.837	0.774	0.842	0.827	0.963	0.899	0.860	0.746	0.755	0.875	0.878	0.975	0.910	0.921



Table 14: Class-wise AUC scores for cross-domain (Hardin County region) performance across single-modality, multi-scale fusion, and multimodal fusion models, using both ResNeXt-50 and ViT-B/16 backbones. Four fusion strategies are shown: early channel stacking (St), mid-level concatenation (C), mid-level attention with a shared encoder (A), and mid-level attention with separate encoders (A2). Spatial scale is indicated in parentheses, either as DEM resolution (in feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*.

Model	ResNeXt							ViT						
	af1	Qal	Qaf	Qat	Qc	Qca	Qr	af1	Qal	Qaf	Qat	Qc	Qca	Qr
<i>RGB</i>	0.757	0.576	0.403	0.486	0.654	0.515	0.507	0.575	0.527	0.782	0.650	0.270	0.381	0.494
<i>NIR</i>	0.733	0.519	0.490	0.550	0.703	0.824	0.727	0.502	0.578	0.474	0.641	0.466	0.348	0.554
<i>DEM</i>	0.804	0.613	0.612	0.472	0.969	0.907	0.733	0.587	0.549	0.379	0.210	0.958	0.947	0.710
<i>EP</i> (101)	0.851	0.716	0.726	0.769	0.621	0.748	0.742	0.745	0.633	0.815	0.629	0.759	0.842	0.789
<i>EP</i> (11)	0.790	0.687	0.801	0.763	0.463	0.563	0.662	0.763	0.698	0.734	0.667	0.807	0.870	0.840
<i>EP</i> (201)	0.786	0.737	0.805	0.752	0.573	0.697	0.717	0.698	0.676	0.821	0.729	0.718	0.818	0.701
<i>EP</i> (21)	0.818	0.700	0.846	0.746	0.392	0.668	0.694	0.778	0.696	0.725	0.662	0.817	0.842	0.796
<i>EP</i> (51)	0.821	0.676	0.769	0.778	0.409	0.519	0.672	0.779	0.633	0.798	0.684	0.771	0.851	0.786
<i>EP</i> (5)	0.769	0.635	0.782	0.847	0.291	0.352	0.399	0.764	0.651	0.626	0.622	0.860	0.882	0.757
<i>PlC</i> (100)	0.513	0.514	0.324	0.516	0.497	0.472	0.454	0.517	0.527	0.809	0.512	0.536	0.527	0.349
<i>PlC</i> (10)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.511	0.305	0.733	0.638	0.500	0.791	0.819
<i>PlC</i> (200)	0.510	0.472	0.899	0.537	0.465	0.503	0.311	0.501	0.562	0.831	0.515	0.554	0.523	0.285
<i>PlC</i> (20)	0.517	0.480	0.529	0.478	0.474	0.492	0.426	0.530	0.304	0.758	0.701	0.627	0.703	0.766
<i>PlC</i> (50)	0.511	0.464	0.275	0.397	0.557	0.497	0.511	0.517	0.470	0.711	0.600	0.537	0.532	0.442
<i>PlC</i> (5)	0.492	0.501	0.599	0.548	0.487	0.509	0.453	0.514	0.340	0.650	0.518	0.561	0.792	0.752
<i>PrC</i> (100)	0.506	0.448	0.150	0.505	0.552	0.631	0.646	0.532	0.428	0.566	0.528	0.463	0.664	0.867
<i>PrC</i> (10)	0.597	0.379	0.797	0.612	0.277	0.363	0.614	0.574	0.508	0.507	0.448	0.372	0.539	0.505
<i>PrC</i> (200)	0.467	0.543	0.464	0.435	0.431	0.429	0.225	0.534	0.424	0.569	0.574	0.573	0.612	0.905
<i>PrC</i> (20)	0.498	0.490	0.408	0.414	0.478	0.491	0.459	0.417	0.644	0.348	0.468	0.584	0.393	0.256
<i>PrC</i> (50)	0.493	0.493	0.426	0.551	0.136	0.248	0.438	0.500	0.458	0.476	0.496	0.500	0.500	0.482
<i>PrC</i> (5)	0.426	0.559	0.263	0.418	0.679	0.710	0.559	0.412	0.633	0.219	0.362	0.500	0.592	0.404
<i>S</i> (100)	0.550	0.549	0.746	0.537	0.965	0.945	0.675	0.533	0.559	0.704	0.372	0.945	0.959	0.859
<i>S</i> (10)	0.781	0.731	0.531	0.696	0.976	0.922	0.815	0.683	0.563	0.528	0.530	0.981	0.937	0.772
<i>S</i> (200)	0.467	0.545	0.541	0.365	0.802	0.890	0.435	0.524	0.533	0.607	0.348	0.919	0.962	0.842
<i>S</i> (20)	0.713	0.704	0.889	0.706	0.976	0.924	0.717	0.621	0.569	0.786	0.708	0.981	0.941	0.509
<i>S</i> (50)	0.625	0.619	0.674	0.665	0.974	0.936	0.718	0.529	0.551	0.964	0.477	0.971	0.952	0.673
<i>S</i> (5)	0.863	0.737	0.611	0.754	0.975	0.915	0.801	0.759	0.579	0.646	0.667	0.981	0.923	0.778
<i>SDS</i> (101)	0.809	0.611	0.443	0.788	0.960	0.886	0.795	0.656	0.474	0.491	0.644	0.954	0.871	0.677
<i>SDS</i> (11)	0.861	0.671	0.587	0.701	0.971	0.905	0.804	0.762	0.538	0.556	0.631	0.957	0.863	0.753
<i>SDS</i> (201)	0.752	0.579	0.503	0.645	0.964	0.804	0.744	0.641	0.479	0.477	0.640	0.942	0.870	0.647
<i>SDS</i> (21)	0.838	0.673	0.749	0.794	0.969	0.869	0.613	0.741	0.543	0.658	0.694	0.952	0.853	0.704
<i>SDS</i> (51)	0.822	0.649	0.608	0.605	0.959	0.834	0.749	0.670	0.515	0.511	0.673	0.943	0.824	0.686
<i>SDS</i> (5)	0.858	0.637	0.805	0.737	0.963	0.886	0.744	0.776	0.561	0.503	0.593	0.958	0.864	0.739
<i>EP</i> (St)	0.769	0.722	0.828	0.722	0.603	0.701	0.671	0.769	0.722	0.828	0.722	0.603	0.701	0.671
<i>EP</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PlC</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PlC</i> (St)	0.479	0.524	0.603	0.489	0.553	0.567	0.432	0.479	0.524	0.603	0.489	0.553	0.567	0.432
<i>PrC</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>PrC</i> (St)	0.496	0.567	0.301	0.440	0.687	0.788	0.202	0.496	0.567	0.301	0.440	0.687	0.788	0.202
<i>S</i> (A)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
<i>S</i> (St)	0.881	0.711	0.643	0.741	0.977	0.915	0.759	0.881	0.711	0.643	0.741	0.977	0.915	0.759
<i>SDS</i> (A)	0.558	0.592	0.699	0.679	0.626	0.602	0.568	0.558	0.592	0.699	0.679	0.626	0.602	0.568
<i>SDS</i> (St)	0.843	0.679	0.629	0.762	0.966	0.889	0.777	0.843	0.679	0.629	0.762	0.966	0.889	0.777
<i>EP+S+SDS</i> (A)	0.555	0.525	0.674	0.552	0.921	0.907	0.816	0.653	0.483	0.500	0.377	0.973	0.955	0.805
<i>EP+S+SDS</i> (A2)	0.555	0.525	0.674	0.552	0.921	0.907	0.816	-	-	-	-	-	-	-
<i>EP+S+SDS</i> (C)	0.701	0.693	0.498	0.689	0.962	0.902	0.804	-	-	-	-	-	-	-
<i>EP+S+SDS</i> (St)	0.857	0.760	0.612	0.736	0.972	0.914	0.792	0.734	0.586	0.740	0.650	0.982	0.922	0.805
<i>EP</i> (201)+ <i>S</i> (200)+ <i>SDS</i> (201) (St)	0.859	0.717	0.699	0.685	0.962	0.911	0.855	-	-	-	-	-	-	-
<i>EP</i> (5)+ <i>S</i> (5)+ <i>SDS</i> (5) (St)	0.860	0.638	0.735	0.760	0.960	0.899	0.833	-	-	-	-	-	-	-
<i>RGB+DEM</i> (A)	0.708	0.527	0.274	0.130	0.836	0.740	0.647	0.671	0.513	0.271	0.548	0.916	0.901	0.531
<i>RGB+DEM</i> (A2)	0.743	0.482	0.325	0.499	0.905	0.835	0.695	0.688	0.498	0.695	0.676	0.941	0.894	0.677
<i>RGB+DEM</i> (C)	0.788	0.460	0.173	0.406	0.661	0.621	0.635	0.752	0.554	0.545	0.611	0.930	0.923	0.732
<i>RGB+DEM+EP+S+SDS</i> (A2)	0.500	0.500	0.500	0.500	0.500	0.500	0.500	-	-	-	-	-	-	-
<i>RGB+DEM+EP+S+SDS</i> (C)	0.841	0.644	0.452	0.493	0.964	0.946	0.833	0.660	0.540	0.687	0.594	0.965	0.933	0.825

Table 15: Macro F1 and AUC scores for SatMAE, SatMAE++, and SGMMap-Net architectures. SGMMap-Net models include one with a similar set of input modalities and another representing the best overall configuration. Spatial scale is indicated in parentheses, either as DEM resolution (in feet) for *PlC*, *PrC*, and *S*, or kernel size for *EP* and *SDS*. SGMMap-Net models use concatenation (C) or channel stacking (St) for fusion. Metrics are reported for in-domain (Warren County, WC) and cross-domain (Hardin County, HC) test sets, along with the difference ( $\Delta$ ) between WC and HC scores. **Bolded** values indicate the highest performance per metric; underlined values represent the second best.

Model	Modalities	F1			AUC		
		WC	HC	$\Delta$	WC	HC	$\Delta$
SatMAE	<i>RGB+DEM+EP(5)+EP(51)+EP(101)+EP(201)+S(5)+SDS(5)</i>	0.614	0.427	0.187	0.864	0.735	0.129
SatMAE++	<i>RGB+DEM+EP(5)+EP(51)+EP(101)+EP(201)+S(5)+SDS(5)</i>	<u>0.656</u>	0.454	0.202	<b>0.904</b>	<u>0.762</u>	0.142
SGMMap-Net	<i>RGB+DEM+EP+ S+SDS (C)</i>	0.618	<u>0.543</u>	<u>0.074</u>	0.858	0.739	<u>0.118</u>
SGMMap-Net	<i>EP+S+SDS (St)</i>	<b>0.657</b>	<b>0.598</b>	<b>0.059</b>	<u>0.882</u>	<b>0.806</b>	<b>0.076</b>