

Navigating the Complexity Trap: Deep Learning Optimization for Mapping Historic Surface Mines in Appalachian Terrain

Mohammad Nahyan

`nahyan-arc@sust.edu`

Shahjalal University of Science and Technology

Aaron Maxwell

West Virginia University

Michael Strager

West Virginia University

Vincenzo Cribari

University of Georgia

Stefania Staniscia

West Virginia University

Research Article

Keywords: Landform identification, post-mining landscapes, artificial neural networks, image segmentation, anthropo-genic geomorphic features

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10535406/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Navigating the Complexity Trap: Deep Learning Optimization for Mapping Historic Surface Mines in Appalachian Terrain

Mohammad Nahyan^{1*}, Aaron Maxwell², Michael Strager³, Vincenzo Cribari⁴, Stefania Staniscia³

¹Department of Architecture, Shahjalal University of Science & Technology, Sylhet, Bangladesh.

²Department of Geology and Geography, Eberly College of Arts and Sciences, West Virginia University, Morgantown, WV, USA.

³Davis College of Agriculture and Natural Resources, West Virginia University, Morgantown, WV, USA.

⁴College of Environment + Design (CED), University of Georgia, Athens, GA, USA.

*Corresponding Author: Mohammad Nahyan, Email: nahyan-arc@sust.edu

Abstract

Anthropogenic landform classification is critical for hazard mitigation, heritage conservation, and ecological restoration in complex geological terrains like the Appalachian region. Within these areas, extensive surface mining has drastically altered local topography, creating a complex mosaic of natural and 'synthetic' landscapes. Specifically, operations predating the Surface Mining Control and Reclamation Act (SMCRA) of 1977 lacked strict backfilling requirements, leaving enduring anthropogenic footprints of exposed highwalls and mine benches. Today, these geomorphological remains are recognized as an integral part of the region's 'mining heritage', serving as permanent, tangible records of human interaction with the environment. However, accurate mapping of these post-mining landscapes is challenging due to dense forest canopy regrowth and their morphological similarities to the rugged natural terrain. Consequently, existing historic maps and databases remain incomplete for these surface mines. To address this, an automated deep learning (DL) framework was developed in this study to detect and delineate relict surface mining features. High spatial resolution light detection and ranging (LiDAR)-derived land surface parameters (LSPs) were utilized to train fully convolutional networks across southern West Virginia. A comparative analysis of four U-Net architectures (Base U-Net, ResU-Net, DiU-Net, and DiResU-Net) coupled with systematic hyperparameter optimization was conducted to determine the optimal configuration. Results indicate that the optimized Base U-Net bypassed the 'complexity trap' of deeper networks, achieving the highest performance with an F1-score of 0.452 and an overall accuracy of 81.4%. Conversely, more complex models struggled with over-segmentation and noise. While the generated inference maps can isolate anthropogenic landforms, future improvements- such as hybrid data fusion with optical imagery, transfer learning across diverse geologic settings, and topology-aware training- may further refine segmentation accuracy, thereby advancing post-mining landscape inventories.

Keywords Landform identification, post-mining landscapes, artificial neural networks, image segmentation, anthropogenic geomorphic features

1 Introduction

As a primary driver of anthropogenic geomorphology, mining represents one of the most intensive forms of land transformation, leaving an enduring topographic signature on existing landscape. In the eastern United States, the Appalachian region is geologically characterized by rich coal reserves embedded within complex topography (Ali et al., 2025; Milici, 2005; Tewalt, et. al., 2014). Specifically, the Central Appalachian region, mostly covering West Virginia and Kentucky historically relied on the coal industry as its primary economic engine (Lutz et al.,

2013; Pericak et al., 2018). As a result of the extensive coal extraction from underground and surface mining, the geomorphology of this region has been greatly altered, and it is often described as a 'synthetic landscapes' (Jaeger & Ross, 2021). The evolution from underground mining to extensive surface mining marked a critical turning point in the region's history. Beginning in the mid-20th century, a distinct strip mining culture emerged where small-scale operators utilized powered machinery to aggressively mine steep slopes (Munn, 1973). This practice involved excavating overburden to create a level bench for coal removal and, mechanically severing the natural slope to leave a precipitous highwall on the uphill side (Brown, 1962). This transformation is particularly pronounced for operations predating the Surface Mining Control and Reclamation Act (SMCRA) of 1977, when reclamation regulations, such as backfilling and recontouring requirements were not enforced by law (Skousen & Zipper, 2014). During this pre-SMCRA era, minimal reclamation requirements led to a proliferation of abandoned mine lands where mine benches and highwalls remained exposed.

Subsequently, the industry transitioned to a more intensive large-scale surface mining operation: Mountaintop Removal (MTR) (Bower et al., 2024; Burns, 2005; Ross et al., 2016). MTR process removes the complete mountaintop including vegetation, soils, and bedrock to access coal seams beneath massive overburden, with the resulting overburden either regraded on-site or disposed of in adjacent valleys called 'valley fills' (Palmer et al., 2010). Surface mining, including both strip mining and MTR, alter the landforms and slope characteristics of the terrain surface (Palmer et al., 2010; Ross et al., 2016). Consequently, these features persist as enduring post-mining legacies that continue to dictate this region's geomorphology and cultural landscape.

These surface mining activities have adverse environmental effects within the Appalachian region (Bernhardt & Palmer, 2011). The regional water quality is severely degraded by the cumulative extent of surface mining and widespread biological impairments has been documented throughout Central Appalachia (Bernhardt et al., 2012). Furthermore, despite the engineering consideration mainly being focused on mining safety, it is demonstrated that specific mine spoil characteristics in reclaimed areas are prone to slope instability (Ali et al., 2025; Reed & Kite, 2020). The structural complexity of valley fills also creates novel subsurface flow paths that complicate hydrological processes in these areas (Jaeger & Ross, 2021). Therefore, to effectively support ecological restoration and hazard mitigation, the accurate mapping of post-mining landscapes, including features resulting from pre-SMCRA resource extraction, is critical.

The significance of these anthropogenic features extends beyond their geomorphic footprint; they are the physical manifestation of the 'coal monoculture' of West Virginia (Burns, 2005). For over a century, the Appalachian identity has been inextricably linked to the coal industry, creating a society where the landscape itself serves as a record of economic survival and industrial dominance. The geomorphological remains of the surface mines are part of mining heritage that encompasses these tangible remains as permanent records of human interaction with the landscape (Velikić et al., 2022). This is an important characterizing feature which can inform spatial planning and heritage conservation (Cribari, et al., 2025). Consequently, these mine lands are not merely environmental hazards to be remediated, but integral components of a cultural landscape that document the profound human alteration of the Appalachia.

Several topographic maps and databases documenting surface mines in the United States are maintained by federal and state agencies, yet significant gaps persist regarding historic features. The National Mine Map Repository (NMMR), maintained by the Office of Surface Mining Reclamation and Enforcement (OSMRE), contains over 275,000 digital archives (OSMRE, 2024). However, because these records are mainly donated by mining companies and surveyors, the OSMRE explicitly notes that they cannot be verified for accuracy and may omit prior or subsequent mining activities. As a result, for pre-SMCRA operations, these historic maps typically depict theoretical permit boundaries rather than the true final geomorphic footprint of the mines. Similarly, while regional datasets have been updated by the West Virginia Geological and Economic Survey (WVGES) and the West Virginia Department of Environmental Protection (WVDEP), these agencies explicitly caution that their interactive inventories are not a complete documentation of all historic mining and contain known data gaps (WVGES, 2024).

To map surface mines, remote sensing efforts have been explored, yet optical methods remain inherently limited to historic mining features. For instance, the full archive of Landsat imagery (1985–2015) was processed by Pericak et al. (2018) using Google Earth Engine to map mining extent based on vegetation removal. Similarly, Cribari et al. (2021) utilized high spatial resolution historical aerial imagery to successfully track land-cover transitions associated with surface mining. While this optical approach is highly effective for mapping two-dimensional land classes (e.g., barren land versus forest), delineating the actual physical landforms requires the addition of topographic data. While successful for detecting active disturbance, these approaches cannot resolve the geomorphic signature of historic, pre-SMCRA mines that have since been obscured by dense canopy regrowth (Cribari et al., 2021; Maxwell & Strager, 2013; Skousen et al., 2006) or historic reforestation (Brown, 1962). Object-based image analysis (OBIA) has been documented to be effective for local-scale geomorphometry and mapping, however, it is less suitable for large regional inventories due to limitations in transferability and automation (Drăguț & Blaschke, 2006; Saha et al., 2002). OBIA is heavily reliant on the manual construction of semantic rule sets and the heuristic selection of segmentation scale parameters (Blaschke, 2010; Hay et al., 2003), which require extensive tuning when applied to heterogeneous landscapes. Given the vast and topographically complex nature of the Appalachian regions, this requirement for expert-guided parameterization renders OBIA less scalable with rigid, pre-defined geometric rules. Therefore, exploring OBIA methods applied to optical data is not the focus of the present study.

During the past several years, convolutional neural networks (CNNs) and deep learning (DL) have achieved significant advances in natural image processing (He et al., 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014). However, because remote sensing data presents more complex textures and background noise than typical natural images, applying CNNs often requires specific adaptations or careful hyperparameter tuning to maximize performance (Cheng et al., 2017). Despite these domain-specific complexities, CNNs are widely applicable to remotely sensed data, driving notable progress in building extraction (Vakalopoulou et al., 2015; Yuan, 2016), road detection (Zhang et al., 2018), and water-body segmentation (Isikdogan et al., 2017).

For landform mapping, the standard architecture often requires modification to address the complexity of geomorphic features. For instance, Yi and Zhang (2020) applied LandsNet, a specialized network designed to extract geomorphic features directly from standard satellite imagery, thereby eliminating the need for DEMs. Conversely, for detailed landform classification in transitional areas, Li et al. (2020) utilized the U-Net architecture, a form of DL, originally proposed by Ronneberger et al. (2015) as the base network structure due to its effectiveness and simplicity. However, it is necessary to modify certain layers to adjust for geomorphological research, demonstrating that DL models can be successfully trained to learn landform features when provided with a dataset combining images, DEMs, and terrain derivatives, often referred to as land surface parameters (LSPs) (Maxwell & Strager, 2013; Maxwell et al., 2023). Addressing the need for specialized architectures in anthropogenic mapping, Farhadpour and Maxwell (2025) evaluated various U-Net configurations for mapping agricultural terraces, mine benches and valley fills when using a feature space consisting of LSPs derived from high spatial resolution LiDAR-derived DEMs. Their findings indicate that while Base U-Net models perform adequately with large training datasets, the incorporation of advanced modules significantly enhances segmentation performance in scenarios involving limited training data or complex geomorphic landscapes, thereby offering a robust solution for mapping anthropogenic features where extensive ground truth data may be scarce.

The application of DL to digital terrain data is still limited (Maxwell et al., 2020). Several studies have focused primarily on extracting ground returns for digital terrain model (DTM) generation (Dai et al., 2023; Rizaldy et al., 2018) or identifying discrete objects such as historic structures (Küçükdemirci et al., 2021; Trier et al., 2019). However, there is a distinct need to investigate mapping anthropogenic and natural terrain features specifically from digital terrain data using DL-based methods that have the potential to outperform traditional shallow learners (e.g. Random Forest) for complex terrain analysis (Behrens et al., 2018; Maxwell et al., 2020). Provided that, researchers augment workflows to handle the computational demands and unique parameterization required for processing large datasets as image samples (Ma et al., 2019; Maxwell et al., 2018). However, the application of DL to regional-scale geological mapping imposes substantial computational demands due to the high

dimensionality of geospatial data (Reichstein et al., 2019). To mitigate this, fundamental optimization strategies are required, such as the 'overlap-tile' processing method originally proposed by Ronneberger et al. (2015) to ensure seamless segmentation of large raster, or the use of transfer learning with pre-trained encoders to accelerate model convergence in data-sparse environments (Iglovikov & Shvets, 2018). Furthermore, while transfer learning has been identified as a promising mechanism for overcoming training data scarcity in remote sensing (Ma et al., 2019), its effectiveness in 'domain adaptation' - transferring learned geomorphic features across disparate geologic settings without extensive retraining, remains a complex and underexplored in quantitative geomorphology (Maxwell et al., 2020).

The primary goal of this study is to develop a robust, automated DL framework for the detection and delineation of relict surface mining features, specifically focusing on strip mine benches, highwalls, and regraded areas from historic (pre-SMCRA) periods in southern West Virginia. While mountaintop removal (MTR) mining has created prominent landscape alterations- such as active mine complexes, flattened reclaimed mountaintops, and valley fills- these represent distinct geomorphic features typically associated with more recent, large-scale operations. Furthermore, previous studies have already successfully addressed the automated mapping of specific MTR features; for example, Maxwell et al. (2020) achieved high accuracy in delineating valley fill faces using a Mask Region-based CNN (R-CNN) approach. Therefore, to avoid redundancy and maintain a precise geomorphological focus, all remnants of MTR mining are explicitly excluded from the scope of this research. The present study is solely dedicated to the detection and delineation of unmapped, relict pre-SMCRA surface mines.

While previous deep learning efforts successfully extracted generalized 'surface mine disturbance' footprints from the cartographic symbology of historic topographic maps (Maxwell et al., 2020), this study advances the field by directly segmenting the physical geomorphic remnants (i.e., highwalls and benches) of these pre-SMCRA operations using high spatial resolution topography. By addressing the limitations of optical remote sensing and the incompleteness of existing inventories, this research utilizes high spatial resolution LiDAR-derived DTMs and LSPs to identify the narrow benches and associated steep highwalls that are often obscured by dense forest canopy and absent from current geospatial databases. To achieve this goal, the specific objectives of this research are:

1. Develop a specialized DL framework capable of detecting complex, relict surface mines- specifically highwalls and benches- using high spatial resolution LiDAR-derived LSPs.
2. Evaluate the effectiveness of different neural network architectures by comparing a Base U-Net baseline against advanced modifications (including residual connections and dilated convolutions) to determine which structure best captures the complex geometry of anthropogenic landforms.
3. Optimize the model training process by systematically tuning key hyperparameters and scaling the training dataset, ensuring the final model achieves the best possible balance between computational efficiency and segmentation accuracy.
4. Validate the reliability of the generated maps through rigorous statistical testing and qualitative visual assessment, confirming that the model's predictions are stable and accurately reflect the physical characteristics of the mining features.

2 Material and Methods

To briefly describe the overall methodology, in this study, historic surface mine maps were utilized as reference data to train a fully convolutional network (FCN) for the binary semantic segmentation of relict surface mines,

specifically strip mine bench and highwall features, within the post-mining landscape of southern West Virginia. To identify the optimal feature extraction strategy, four distinct U-Net configurations were systematically evaluated using LSPs derived from high spatial resolution DTMs. This evaluation benchmarked a Base U-Net architecture (Ronneberger et al., 2015) against advanced variants, including a Residual U-Net (ResU-Net) (He et al., 2016; Zhang et al., 2018), a Dilated U-Net (DiU-Net) utilizing atrous convolutions (Yu & Koltun, 2015), and a hybrid Dilated-Residual U-Net (DiResU-Net) (Diakogiannis et al., 2020). Furthermore, model performance was assessed across varying training sample sizes and batch allocations to determine the computational trade-offs of processing 512×512 -pixel image chips. Ultimately, the optimal model configuration was applied to generate a comprehensive inference map of the historic surface mining features remaining across the study area. The methodological diagram of this research is shown in Figure 1. Methodological details including the different U-Net architecture tested in this research are explained in the following sections.

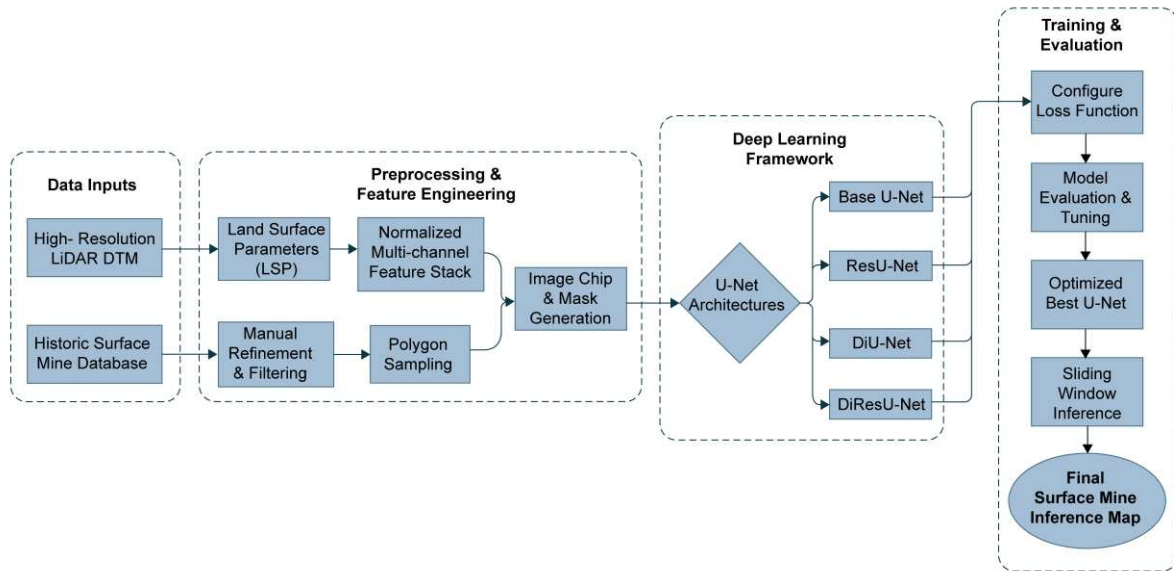


Figure 1. Overall methodological diagram of this research.

2.1 Study area

The study area is situated within the National Coal Heritage Area (NCHA) of southern West Virginia (Figure 2), a region characterized by the rugged topography of the Appalachian Plateau and a century-long legacy of resource extraction. While the region possesses significant environmental and cultural assets- including the New River Gorge National Park and Preserve, the Bluestone National Scenic River, eleven state parks, and the Coal Heritage Trail National Scenic Byway, the landscape has been substantially altered by anthropogenic disturbance. The prevalence of surface mining has impacted on the visual integrity of these scenic corridors and contributed to the fragmentation of the central Appalachian forests, which are recognized as one of the most biodiverse non-tropical ecosystems globally (Drummond & Loveland, 2010; Zipper et al., 2020).

Within this domain, the study specifically targets post-mining anthropogenic landforms generated during the pre-SMCRA era. These features are morphologically distinct from the natural ridges and valleys of the region, consisting primarily of excavated highwalls- steep, vertical rock faces left after coal extraction- and their associated mine benches, which represent the flat, excavated terraces created to access the coal seams (Figure 3).

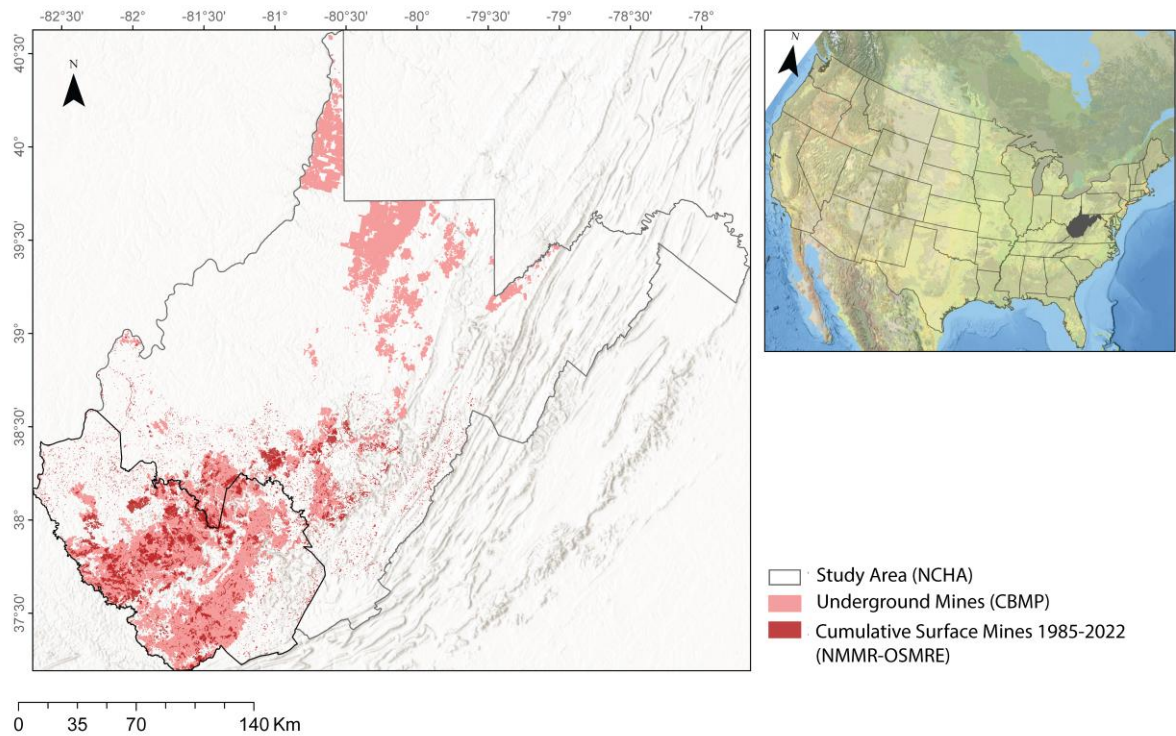


Figure 2. The study area of this research and extensive underground and surface mining activities in West Virginia.

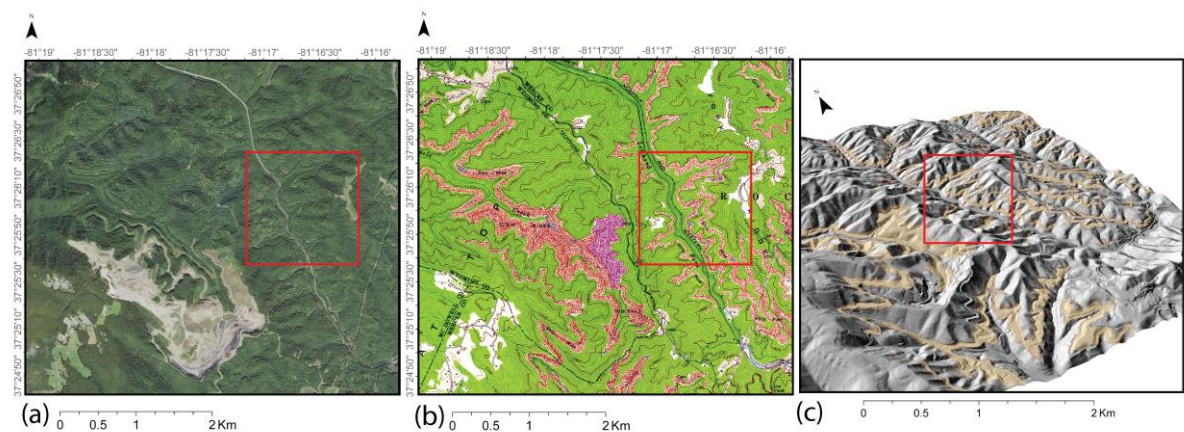


Figure 3. Example of surface mine features shown on a) areal 1m Orthophoto imagery (West Virginia GIS Technical Center, 2020) and b) historic topographic map (Crumpler Quadrangle, West Virginia, 1967, 1:24,000 scale) sourced from the USGS National Map Historical Topographic Map Collection. c) The same surface mine features highlighted on a 3D hillshade model created using ArcGIS Pro, with corresponding areas indicated by red boxes.

2.2 Data Sets and Sampling

The primary topographic data utilized in this study consisted of a high spatial resolution DTM acquired from the West Virginia GIS Technical Center (WVGISTC). This DTM was generated from LiDAR at a 1 m spatial resolution (WVGISTC, n.d.). Surface mining reference data was obtained from the Coal Bed Mapping Project (CBMP) of the WVGES. This dataset was digitized from a compilation of underground and surface mine maps provided by the West Virginia Office of Miners' Health, Safety, and Training and the OSMRE (WVGES, 2024).

For this study, specific surface features such as mine benches, high walls were extracted from the CBMP dataset. To generate a reliable 'ground truth' for DL, the original dataset was manually refined to correct for instances where the records depicted regulatory permit boundaries rather than the precise geo-morphic extent of the mining features. This correction was performed through visual inspection of high spatial resolution LiDAR-derived hill-shade imagery, resulting in a final curated dataset of 3,301 reference polygons representing visually verified, distinct relict mine features.

To ensure rigorous model evaluation, a sampling lattice consisting of 6×6 km square grids was superimposed over the study area (Figure 4), creating 127 distinct plots. To prevent spatial autocorrelation and ensure an unbiased evaluation of model performance, these plots were partitioned using a spatially stratified random sampling design. This approach decoupled test data from training clusters, ensuring that performance metrics (e.g., Precision, Recall, F1-scores) reflected the model's generalization capability across the entire study area rather than overfitting to specific local clusters.

From these plots, image chips with a resolution of 512×512 pixels were generated. Given that surface mining features in this region typically appear as narrow, contour-parallel structures amidst substantial natural background terrain, a filtering criterion was applied: only image chips containing at least one pixel of positive mining features were retained. This constraint was implemented to minimize class imbalance and focus the training process on relevant geomorphic signatures. The specific distribution of sampling plots and the resulting image chips is de-tailed in Table 1.

Table 1. Distribution of image chips across training, validation, and testing partitions.

	<i>Tarin</i>	<i>Validation</i>	<i>Test</i>	<i>Total</i>
Grid Plots	86	19	22	127
Samples	4873	1037	1099	7009

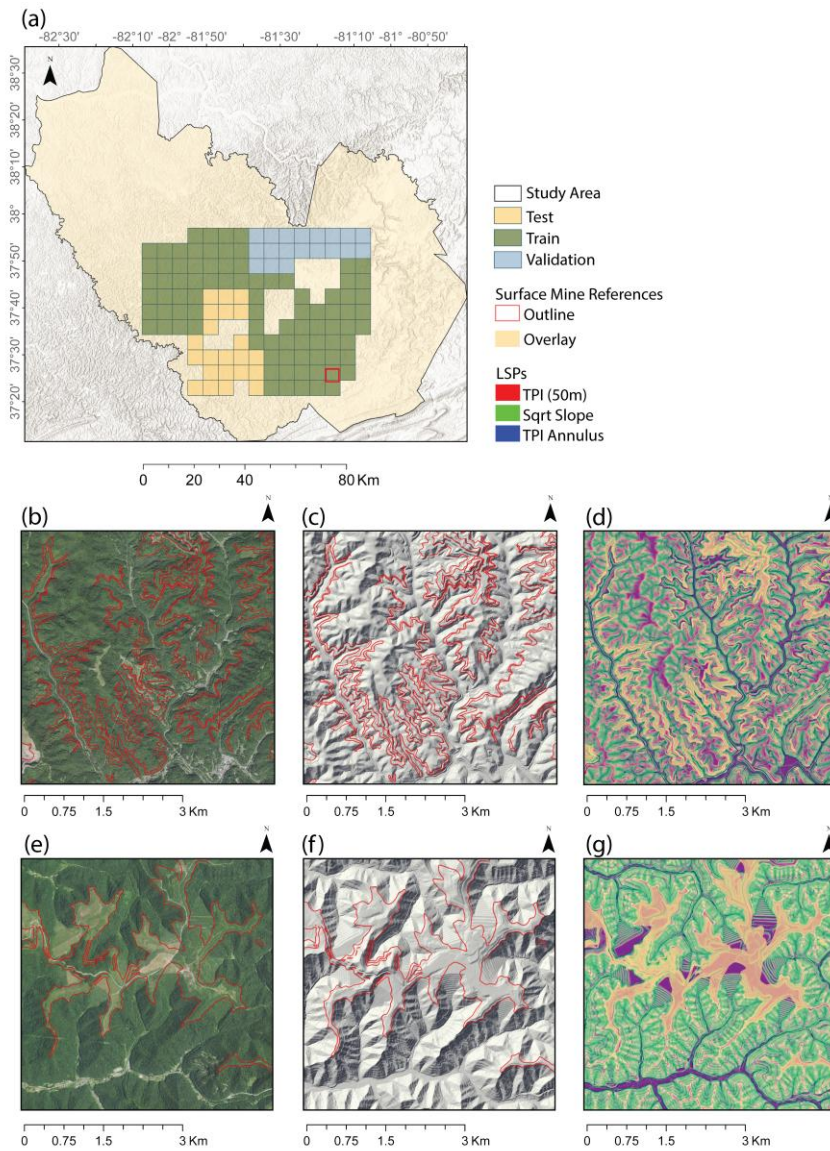


Figure 4. a) Grid plots for sampling. b-g) Surface mine reference data derived from CBMP dataset in an example extent, areal 1-m Orthophoto imagery (West Virginia GIS Technical Center, 2020) (b,e), high spatial resolution LiDAR-derived hillshade (c,f) and LSPs (d,g).

2.3 Data Analysis

2.3.1 Data Preprocessing and Feature Engineering

Data preprocessing and geomorphic feature engineering were conducted in the R environment (R Core Team, 2024) using the terra package (Hijmans, 2023). To enhance the visibility of anthropogenic landforms against the natural topography, a multi-channel input stack was derived from the DTM. Following the suggestions of Maxwell

et al. (2023), a three-layer composite of LSPs was adopted, as this specific feature space has been demonstrated to yield higher F1-scores and improved generalization compared to traditional visualization inputs such as hill-shades.

The composite consisted of three normalized layers:

- Topographic Position Index (TPI): Calculated using a 50-meter circular neighborhood to characterize hillslope-scale conditions. Values were clamped between -10 and 10 to minimize outliers.
- Square Root of Slope: Computed from slope degrees to suppress extreme values and accentuate the subtle gradient changes characteristic of strip-mine benches. These values were clamped to a ceiling of 10.
- Local TPI: Calculated using an annulus kernel (2-meter inner and 10-meter outer radius) to capture micro-topographic variability and terrain rugosity.

All derivative layers were linearly normalized to a range of 0 to 1 prior to model ingestion to ensure consistent feature scaling. From this composite, image chips of 512×512 pixels were generated using a stride equal to the chip size to create non-overlapping tiles. This field of view corresponds to a spatial footprint of approximately 0.26 km², which was determined to be appropriate for encompassing the morphological variability of the target landforms. To mitigate the effects of extreme class imbalance, chips containing zero mine pixels were filtered out during the generation process, ensuring that the models were trained exclusively on samples containing positive instances of the target class.

2.3.2 Deep Learning Architectures

To address the complexity of identifying anthropogenic terrain features, this study compares the Base U-Net architecture with three modified configurations designed to enhance feature propagation and contextual understanding. The Base U-Net, originally proposed by Ronneberger et al. (2015), serves as the baseline for this study. It consists of a contracting encoder path to capture context and a symmetric expanding decoder path that enables precise localization. Spatial downsampling in the encoder is achieved using (2×2 max pooling / strided convolutions), which halves the spatial dimensions of the feature maps. Conversely, the decoder employs (2×2 transposed convolutions / bilinear interpolation) to upsample the feature maps back to the original image resolution.

To mitigate the vanishing gradient problem in deeper networks, the ResU-Net replaces standard convolutional blocks with Residual Blocks. These blocks introduce a ‘shortcut’ connection that adds the input directly to the output, facilitating the flow of information through the network (He et al., 2016). This structure has proven particularly effective in remote sensing for extracting linear and complex features, such as road networks (Zhang et al., 2018). Additionally, Batch Normalization (Ioffe & Szegedy, 2015) was applied after each convolution to stabilize the learning process and accelerate convergence.

To capture multi-scale geomorphic context- such as distinguishing wide mine benches from narrow roads- without reducing spatial resolution, the DiU-Net incorporates Atrous (Dilated) Convolutions. By inserting spaces (holes) between kernel elements, the receptive field is expanded without increasing the parameter count (Yu & Koltun, 2015). This technique is widely used in semantic segmentation to aggregate multi-scale context (Chen et al., 2017), allowing the model to distinguish large-scale mine benches from smaller, locally similar terrain features.

Finally, the hybrid architecture (DiResU-Net) was developed to combine the deep feature extraction capabilities of residual learning with the expanded context of dilated convolutions (Diakogiannis et al., 2020). This model

utilizes a residual backbone paired with a dilated residual bottleneck. Instead of standard convolutions, this bottleneck employs sequential dilated convolutions with a fixed dilation rate (rate=2). This modification expands the receptive field of the network without increasing the number of parameters or degrading spatial resolution, allowing the model to capture broader contextual information critical for identifying large-scale post-mining terrain features. The following Figure 5 shows a schematic representation of the different U-Net architectures tested in this study.

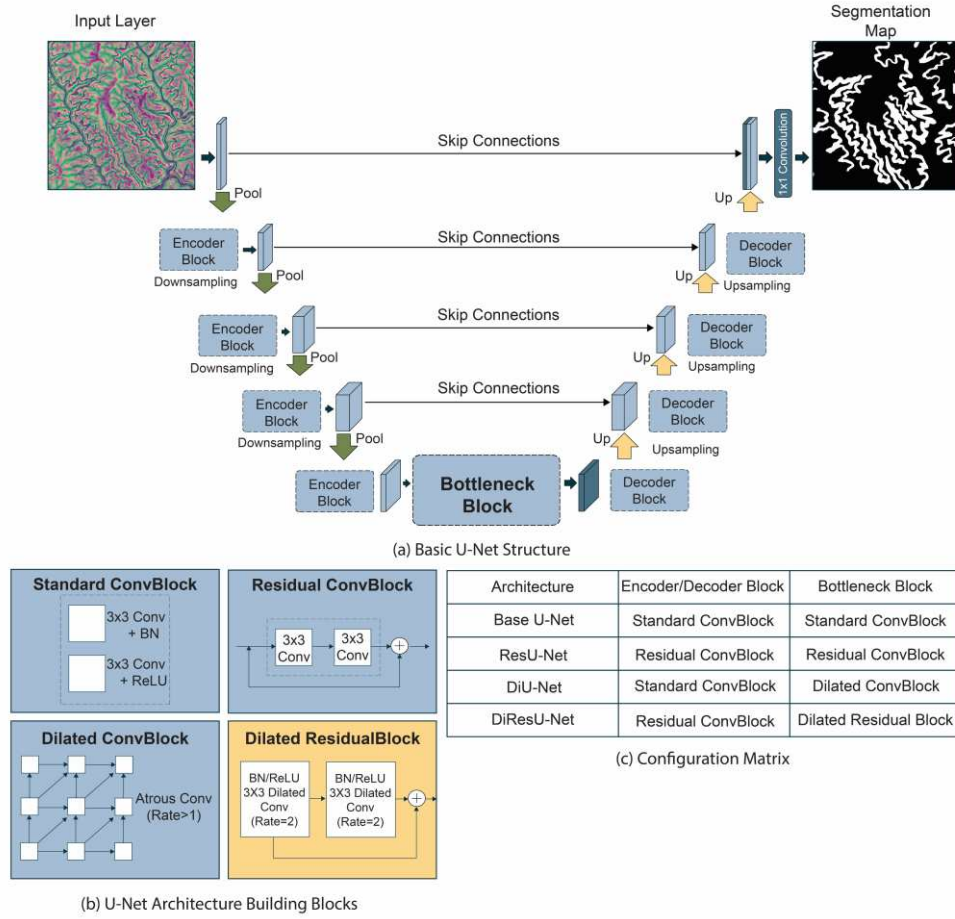


Figure 5. Schematic representation of the tested U-Net architectures. (a) The basic U-Net architecture relies on a standard encoder-decoder structure with skip connections. (b) U-Net architectures varied based on the model: the Base U-Net utilizes standard ConvBlocks; the ResU-Net replaces these with Residual ConvBlocks to mitigate vanishing gradients; the DiU-Net inserts Dilated ConvBlocks at the bottleneck to expand the receptive field; and the hybrid DiResU-Net combines a residual backbone with a Dilated Residual bottleneck, utilizing sequential dilated convolutions to capture broader contextual information while maintaining deep feature extraction. (c) Configuration matrix of the tested U-Net architectures.

2.3.3 Experimental Design

The DL workflow was implemented in the R computing environment utilizing the tensorflow (Allaire & Tang, 2022) and keras package (Allaire & Chollet, 2025). A three-stage experimental design was employed to isolate architectural performance from hyperparameter tuning.

- Stage 1 (Sensitivity Analysis/Architecture Selection): To determine the optimal architecture, all four U-Net configurations (Base U-Net, ResU-Net, DiU-Net, DiResU-Net) were first evaluated using a controlled, fixed configuration: a batch size of 2 and a limited random sample of 500 image chips. This constraint was applied to avoid computational bottlenecks while isolating the influence of architectural changes.
- Stage 2 (Hyperparameter Optimization): The best-performing model identified in Stage 1 was subsequently subjected to targeted hyperparameter testing. While DL models contain numerous tunable parameters (e.g., learning rate, data augmentation degree, and loss function parameterization), these were held constant based on established best practices and preliminary empirical trials to prevent an exponentially large grid search. Instead, this study specifically focused on optimizing the mini-batch size (testing sizes of 2, 6, and 10). This parameter was prioritized because it represents the primary computational bottleneck when processing high spatial resolution LiDAR tensors, directly dictating the balance between gradient stability and GPU memory efficiency. Following this evaluation, the optimized model configuration was trained on the full dataset to maximize generalization capability.
- Stage 3 (Bootstrap Validation): To rigorously assess model stability and quantify the variance in performance metrics, a bootstrap analysis was conducted on the final optimized configuration. The model was iteratively trained and evaluated on randomized subsets of the data (resampling with replacement). This process ensured that the reported accuracy metrics represented the model's true generalization capability rather than artifacts of a specific training-validation split.

2.3.4 Training and Evaluation

To optimize the models, the Adam optimizer (Kingma & Ba, 2014) was utilized with a learning rate of $1e-4$. Given the significant imbalance between mining features and the background terrain, a Hybrid Focal Tversky Loss was implemented. This loss function combines the precision-recall balance of the Tversky index (Salehi et al., 2017) with the hard-example mining capability of Focal Loss (Lin et al., 2017; Abraham & Khan, 2019). To ensure an unbiased tradeoff between false positives and false negatives, the Tversky index was parameterized with $\alpha=0.5$ and $\beta=0.5$, effectively aligning the metric with the Dice coefficient. A focusing parameter of $\gamma=0.85$ was applied to the focal component to down-weight easy background examples while maintaining stable gradient flow for the minority class.

$$Loss = (1 - Tversky\ Index)^\gamma + Binary\ Cross\ Entropy \quad (1)$$

To improve model generalization, data augmentation was applied on-the-fly during training, including random horizontal/vertical flips, rotations (factor 0.25), and contrast adjustments. Training was conducted for 25 epochs with a dynamic learning rate reduction (ReduceLROnPlateau) triggered if validation loss stagnated for 5 epochs. Model checkpoints were utilized to save the weights yielding the lowest validation loss.

2.3.4.1 Evaluation Metrics

To carefully assess the performance of the DL models, a comprehensive suite of pixel-based metrics derived from the confusion matrix was employed. These metrics quantify the correspondence between the model predictions and the ground truth reference polygons, where True Positives (TP) represent correctly identified mine pixels, True Negatives (TN) represent correctly identified background pixels, False Positives (FP) represent natural terrain incorrectly classified as mining features, and False Negatives (FN) represent mining features missed by the model.

First, Precision and Recall (Sensitivity) were calculated to evaluate the model's ability to minimize commission and omission errors, respectively. Overall Accuracy (OA) was also computed to provide a general measure of correct classification across both classes.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$Overall Accuracy (OA) = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Given the inherent trade-off between Precision and Recall, the F1-Score (harmonic mean of precision and recall) was utilized as the primary metric. To maximize this score, probability maps were generated for the validation set using an iterative thresholding approach (0.01 to 0.99) to identify the optimal cutoff value. Furthermore, the Matthews Correlation Coefficient (MCC) was adopted to reliably assess model accuracy given the inherent class imbalance of the terrain data (Chicco et al., 2021; Meena et al., 2022).

$$F1 Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (6)$$

All evaluation metrics were subsequently calculated on the independent test set (unseen samples withheld entirely from the training process) using the optimized threshold. This ensures that the reported metrics represent an unbiased assessment of the model's generalization performance on new data.

2.3.4.2 Inference and Visualization

Upon completion of the quantitative evaluation, the model configuration yielding the highest performance metrics was deployed to generate a comprehensive inference map of relict surface mines. To ensure a rigorous assessment of model generalization, this inference process was applied across the full study area, specifically targeting the unseen test plots that were strictly withheld during the training and validation phases.

The inference workflow was executed in the R computing environment using the terra package to generate the raster maps. Inference was conducted using a sliding window approach with a stride equal to the chip size to ensure seamless coverage of the region. Final probability maps were exported to ArcGIS Pro software (Esri Inc., 2018) for visualization and qualitative assessment against reference areal 1-meter Orthophoto imagery (West Virginia GIS Technical Center, 2020) and 1-meter LiDAR-derived hillshade imagery.

3 Results

3.1 Model Performance Assessment

The comparative analysis of DL architectures revealed that the Base U-Net was the most effective model for identifying relict surface mining features in this specific terrain (Table 2).

Table 2. Comparative performance metrics of DL models on the independent test set. The Base U-Net (Batch 6) trained on the full dataset achieved the highest overall performance.

Model Selection							
	Sample Size	Batch Size	Precision	Recall	F1_Score	OA	MCC
Base U-Net	500	2	0.358	0.554	0.435	0.806	0.336
ResU-Net	500	2	0.342	0.529	0.415	0.800	0.312
DiU-Net	500	2	0.286	0.625	0.392	0.740	0.286
DiResU-Net	500	2	0.262	0.474	0.338	0.751	0.213
Hyperparameter Optimization							
Base U-Net	500	6	0.372	0.562	0.448	0.814	0.352
Base U-Net	500	10	0.297	0.571	0.391	0.761	0.281
Final Model							
Base U-Net	4873 (All samples)	6	0.373	0.572	0.452	0.814	0.357

Note: Bold values indicate the highest performing configuration.

3.1.1 Sensitivity Analysis

In the initial sensitivity phase (fixed batch size of 2), the Base U-Net achieved the highest F1-score of 0.435, surpassing the ResU-Net (0.415) and significantly outperforming the Hybrid DiResU-Net (0.338). While the Dilated U-Net (DiU-Net) exhibited the highest sensitivity (Recall = 0.625), successfully identifying a large proportion of mine features, this came at the cost of substantial over-segmentation. Consequently, the DiU-Net suffered from the lowest precision (0.286) among the top performers. Conversely, Base U-Net achieved the most balanced performance, effectively minimizing false positives while maintaining a competitive recall of 0.554.

3.1.2 Hyperparameter Optimization

Subsequent optimization of the Base U-Net demonstrated that batch size played a critical role in model convergence (Figure 6). As illustrated in the training profiles, the standard batch size of 2 resulted in noisy validation curves. Increasing the batch size to 6 yielded a notable improvement, raising the F1-score to 0.448 and the Matthews Correlation Coefficient (MCC) to 0.352. This suggests that the larger batch size provided a more stable gradient estimation. However, further increasing the batch size to 10 resulted in performance degradation (F1 = 0.391), likely due to the reduced frequency of weight updates limiting the model's ability to escape local minima in the complex loss landscape.

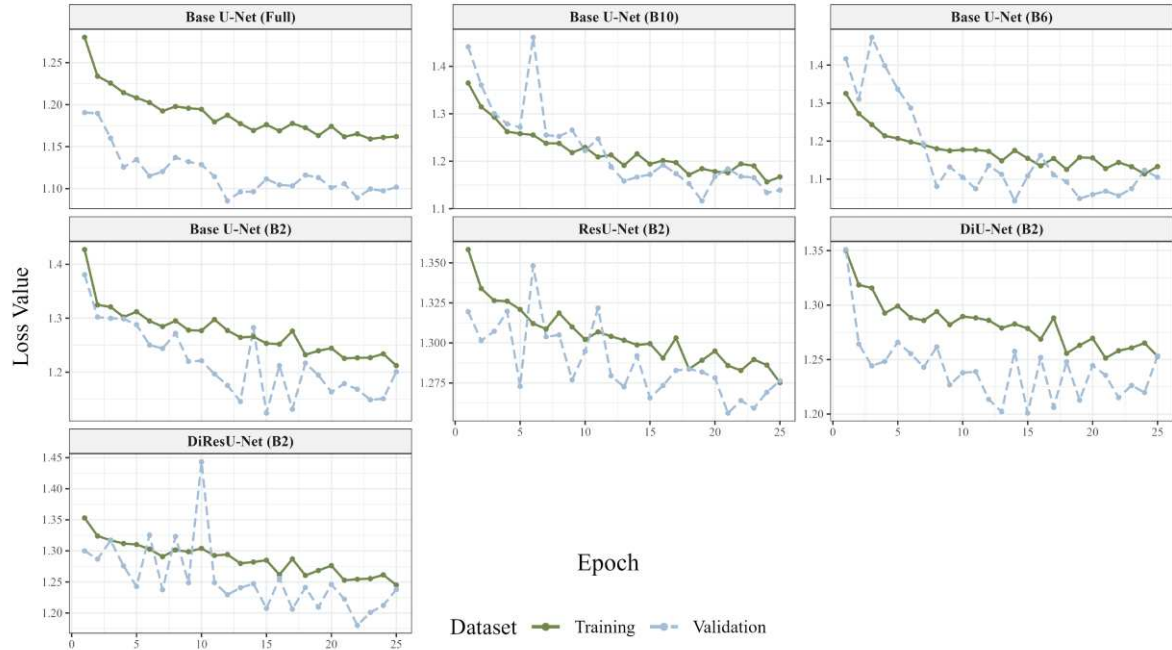


Figure 6. Comparative learning curves for the U-Net configurations over 25 epochs.

3.1.3 Final Model Validation

The final model, trained on the full dataset using the optimized Base U-Net (Batch 6) configuration, confirmed these findings. It achieved the highest overall performance with a top F1-score of 0.452 and an MCC of 0.357. This slight improvement over the sensitivity test confirms that the inclusion of the full dataset enhanced the model's generalization capability without introducing overfitting.

3.2 Bootstrap Analysis

To validate statistical reliability, bootstrap analysis (n=1,000 iterations of resampled test image chips) quantified performance variance across randomized test subsets (Figure 7). The results reveal that the Base U-Net (Full Dataset) and the optimized Base U-Net (Batch 6) configurations exhibit the highest median F1-scores and narrowest interquartile ranges, indicating superior stability and generalization. In contrast, complex architectures like the DiU-Net and ResU-Net demonstrated significant volatility, characterized by wide variances in precision and recall that reflect inconsistent performance and over-segmentation. These findings empirically validate the selection of Base U-Net, confirming that its performance advantage is robust and not an artifact of specific test splits.

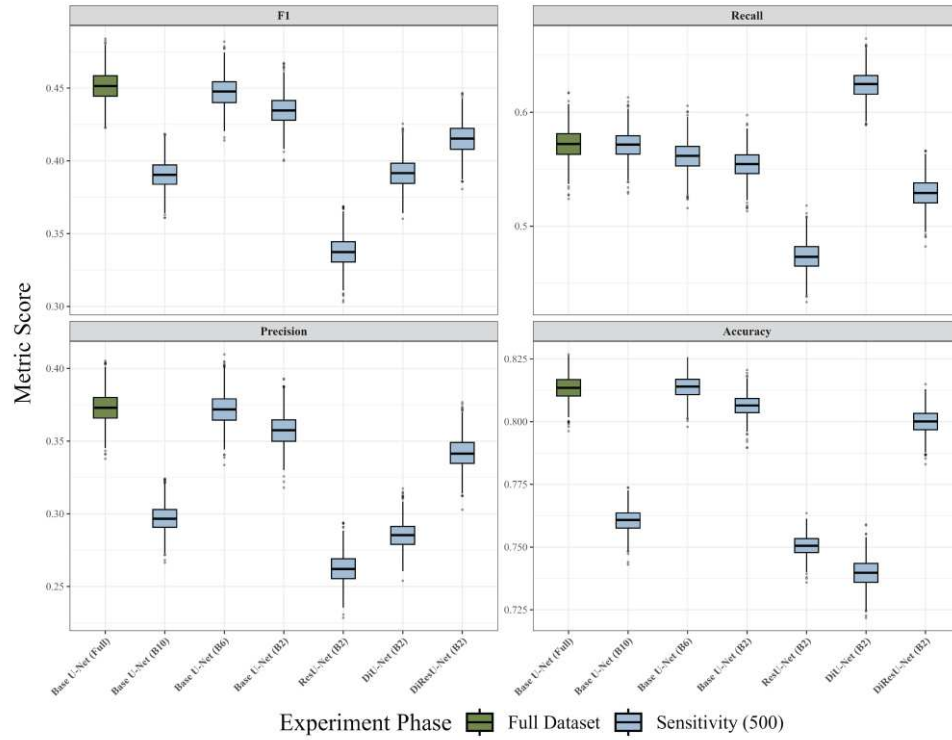


Figure 7. Bootstrap performance distributions (n=1,000 iterations) for the tested U-Net configurations. The Base U-Net (Full Dataset) demonstrates the highest stability (narrowest box) and highest median F1-score. Note the high recall but low precision variance in the DiU-Net, contrasting with the balanced performance of the Base U-Net.

3.3 Visual Assessment

While quantitative metrics provided a baseline for model selection, a qualitative visual assessment was further conducted to evaluate the spatial coherence of the predictions (Figure 8). Confirming the quantitative results, the Base U-Net (Batch 6) trained on the full dataset exhibited the most precise identification of surface mining features (Figure 8a). Conversely, the Base U-Net models trained with limited samples (Batch 6 and 10) produced similar results but introduced notable false positives, incorrectly classifying valley bottoms and foothills as surface mines (Figure 8b-c). Similarly, the Base U-Net (Batch 2) and DiU-Net suffered from substantial over-segmentation, erroneously labeling extensive valley and foothill regions as mining features (Figure 8d, f). The ResU-Net model struggled significantly with feature extraction, resulting in scattered false negatives that indicate difficulty in learning contiguous geometries (Figure 8e). Meanwhile, the hybrid DiResU-Net adopted a conservative prediction strategy; while it generated fewer false positives, it failed to detect a substantial portion of the mining features (Figure 8g).

Across all models, the complex curvilinear morphology of the strip mines was not fully captured. The continuity of these linear features was frequently compromised at sharp bends (resulting in false negatives), while ‘spillover’ effects occurred at ambiguous boundaries (resulting in false positives). Notably, however, all models successfully identified an existing surface mine in the northeast that was absent from the reference dataset, confirming the models' ability to generalize beyond the training labels.

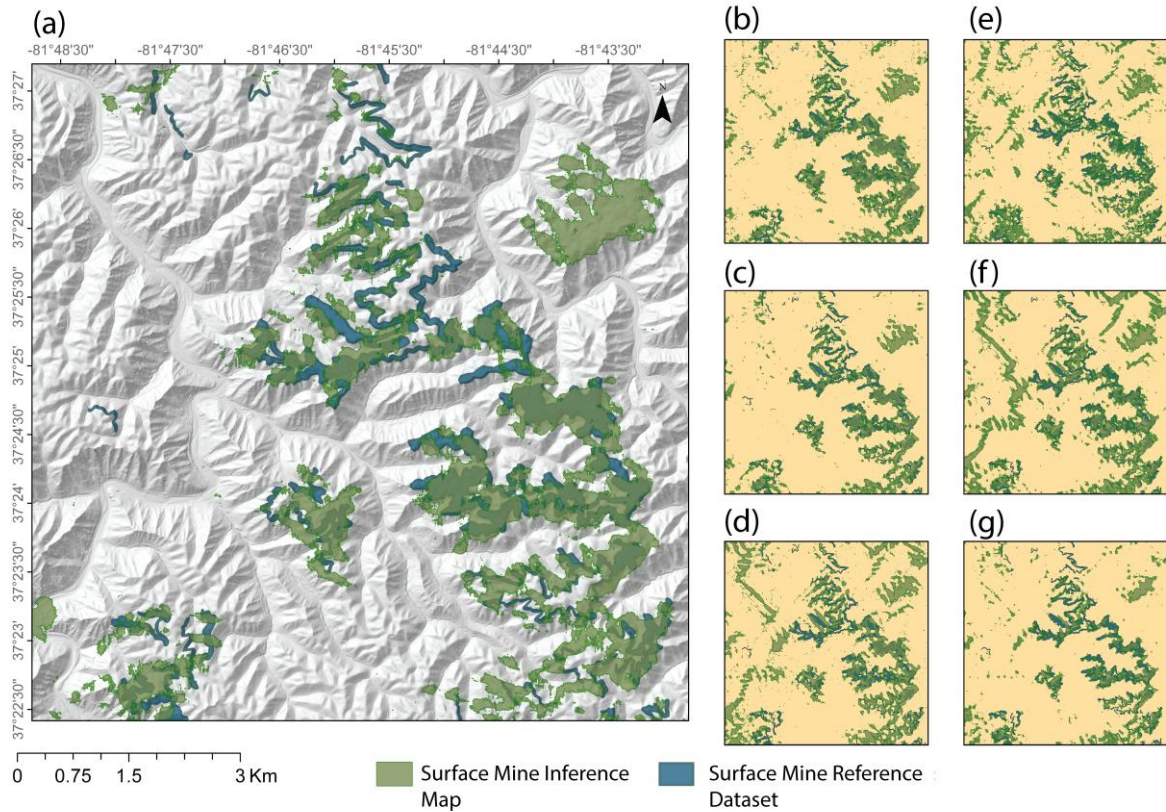


Figure 8. Comparative inference maps of segmentation results. (a) The optimized Base U-Net (Batch 6) trained on the full dataset. (b-g) Models trained on a limited subset (500 samples) for sensitivity analysis: (b) Base U-Net (Batch 10), (c) Base U-Net (Batch 6), (d) Base U-Net (Batch 2), (e) ResU-Net (Batch 2), (f) DiU-Net (Batch 2), and (g) DiResU-Net (Batch 2).

4 Discussion

4.1 Efficacy of Deep Learning in Complex Anthropogenic Terrain

This study demonstrates that DL frameworks driven by high spatial resolution LiDAR-derived terrain data can effectively segment relict surface mining features in the complex topography of the central Appalachians. The final model achieved an F1-score of 0.452 and an Overall Accuracy of 81.4%, a significant achievement given the subtle, degraded nature of pre-SMCRA mining features obscured by over 40 years of reforestation. While active surface mining operations are frequently mapped using optical imagery (Maxwell et al., 2020; Pericak et al., 2018)- though spectral confusion with other barren land covers and developed areas remains a persistent challenge- relict benches and highwalls beneath dense canopy are spectrally indistinguishable from the surrounding forest. By relying exclusively on geometric signatures encoded in LSPs, the model successfully learned the structural morphology of these legacy anthropogenic disturbances.

These results are comparable to recent benchmarks in similarly complex environments. For instance, Li et al. (2020, p. 11) achieved an overall accuracy of 87% when classifying broad landform categories (e.g., ridges,

valleys) in transitional loess areas. While their accuracy is numerically higher, their target features were distinct, large-scale landforms, whereas this study targets narrow, discontinuous anthropogenic cuts that are morphologically ambiguous and mimic natural breaks-in-slope.

Furthermore, Farhadpour and Maxwell (2025, p. 18) reported F1-scores ranging from 0.47 to 0.61 when extracting historic mine benches using various U-Net configurations. While their results are slightly higher, this performance gap is likely attributable to distinct differences in geography, landscape history, and dataset construction. First, Farhadpour and Maxwell (2025) focused on northern West Virginia, a distinctly different landscape that lacks the extensive, recent mountaintop removal (MTR) mining that heavily overprints the older pre-SMCRA disturbances found in our southern West Virginia study area. Differentiating legacy features from a background heavily scarred by modern MTR naturally increases segmentation difficulty. Second, they utilized a 'purpose-built' dataset generated entirely through manual on-screen digitization of high spatial resolution LiDAR, ensuring pristine boundary precision specifically for DL training. In contrast, this study relied on refining federal inventories (CBMP), which inherently retain some spatial noise and boundary generalizations from their paper-map origins.

4.2 The ‘Complexity Trap’: Why Simpler Architectures Prevailed

A critical finding of this research is the counter-intuitive performance of the tested architectures. Contrary to the prevailing trend in computer vision where deeper, more complex networks typically yield superior results (Zhang et al., 2018; He et al., 2016), this study found that the Base U-Net outperformed its more sophisticated variants, including the ResU-Net, DiU-Net, and the hybrid DiResU-Net.

The DiU-Net, designed to expand the receptive field and capture larger contextual features, exhibited the highest recall (0.625) but the lowest precision (0.286). This suggests that while the dilated convolutions successfully aggregated multi-scale context, they aggressively over-smoothed the feature boundaries. In the rugged terrain of West Virginia, where natural ridgelines often mimic the linear geometry of highwalls, this excessive context integration led the model to conflate natural steep slopes with anthropogenic cuts, resulting in substantial commission errors.

Similarly, the Hybrid DiResU-Net significantly underperformed ($F1 = 0.338$). This drop in accuracy is not due to a bloated model size, as residual blocks add very few trainable parameters compared to the Base U-Net. Instead, the issue lies in how the architecture processes data. Residual connections are designed to pass early, unfiltered information deeper into the network. While this helps ultra-deep models analyze complex photographs, it works against this specific terrain data. By bypassing the network's filters, these connections likely dragged forward ‘noise’ from the rugged natural mountains. As a result, the model struggled to suppress the natural landscape to cleanly isolate the human-made mining features, leading to poor generalization on unseen data.

The Base U-Net, with its balanced encoder-decoder structure, proved to be the ‘Goldilocks’ solution. It possessed sufficient depth to learn the hierarchical features of mining landscapes (e.g., the sharp break-in-slope of a highwall) without introducing architectural redundancies (like unnecessary residual pathways) that carry forward noise, or aggressive receptive field expansion that blurs fine-scale boundaries. This finding aligns with Li et al. (2020), who also found that Base U-Net architectures often suffice for landform classification when the input data (LSPs) already explicitly highlights the features of interest.

4.3 The Critical Role of Hyperparameter Optimization

Beyond architecture, this study highlights the often-overlooked importance of hyperparameter tuning, specifically batch size, in geospatial deep learning. The results demonstrated a non-monotonic relationship between batch size and model performance.

This finding contributes to an ongoing discussion in deep learning literature regarding the impact of mini-batch size on model convergence and memory utilization. In semantic segmentation tasks, particularly those utilizing high spatial resolution remote sensing imagery, batch sizes are inherently constrained by GPU memory limitations (Ghosh et al., 2018). Consequently, foundational architectures like the original U-Net were initially designed and trained using micro-batch sizes of a single image (Ronneberger et al., 2015). More recent remote sensing applications continue to demonstrate successful anthropogenic feature extraction using batch sizes as small as two (Doghmane & Boukari, 2024), supported by the theoretical understanding that continuous convolutions on large spatial grids naturally necessitate small batches (Wu & He, 2018). However, as our results indicate, simply defaulting to the smallest possible batch size does not always yield optimal segmentation.

Increasing the batch size from 2 to 6 resulted in a statistically significant performance gain (+1.3% F1). This improvement is attributed to Gradient Stability. In a dataset dominated by ‘background’ pixels (forest), a small batch size (e.g., 2) often contains unrepresentative samples- some batches might contain only forest, while others contain complex mine features. This variance causes the gradient descent algorithm to oscillate wildly (‘noisy updates’), making it difficult for the model to settle into a global minimum. Increasing the batch size to 6 stabilized these updates, providing a more accurate estimate of the true error gradient.

However, further increasing the batch size to 10 degraded performances. This phenomenon, often described as the ‘generalization gap’ in DL literature (Keskar et al., 2017), occurs because large batch sizes tend to converge to ‘sharp’ minima in the loss landscape- solutions that work well for training data but fail when the data shifts slightly (as seen in the test set). Smaller to moderate batch sizes (like 6) introduce a beneficial amount of noise that pushes the model into ‘flat’ minima, which is more robust and generalize better to new data. Ultimately, this study identified an optimal configuration that effectively balances segmentation performance with computational demand. However, this optimization is specific to the fixed input scale used (512×512); future investigations utilizing varying chip sizes are necessary to characterize the potential ‘scale effect’ in U-Net architectures, determining if larger spatial contexts yield diminishing returns or further stability improvements.

Furthermore, the adoption of a hybrid Focal Tversky Loss function proved essential for mitigating the severe class imbalance inherent in the dataset. As noted by Farhadpour and Maxwell (2025), standard loss functions like Cross-Entropy often bias the model toward the majority class (background forest), leading to high overall accuracy but poor sensitivity for the minority class (mining features). By integrating the focal parameter (γ) to down-weight easy examples and the Tversky index to penalize false negatives, this hybrid loss forced the model to prioritize the difficult, sparse boundaries of highwalls and benches. This alignment with recent successful applications in geomorphometry confirms that specialized loss functions are as critical as architectural depth for segmenting subtle anthropogenic features.

4.4 Limitations and Future Directions

Despite the model's success, qualitative inspection of the inference maps reveals persistent challenges, primarily driven by the morphological ambiguity between natural cliffs and anthropogenic highwalls. In the steep terrain of the Appalachian Plateau, natural sandstone outcrops often form vertical scarps geometrically identical to mine

highwalls. Lacking spectral context or stratigraphic data, the model occasionally misclassifies these natural outcrops as mining features. Additionally, the model struggled with feature discontinuity, where relict benches segmented by erosion gullies or landslides were detected as fragmented segments rather than continuous linear features. While morphological post-processing could mitigate this, it risks merging distinct adjacent features. Additionally, the "ground truth" itself presents a limitation; reference polygons derived from historic topographic maps often lack precise alignment with the recent geomorphic conditions, potentially penalizing predictions. Furthermore, mine bench widths and configurations vary significantly depending on the local topography, leading to heterogeneous representations within small input chips. Although extensive manual corrections were applied to the training dataset, treating disparate sub-regional landforms uniformly likely hindered the model's ability to learn topology-specific feature patterns.

Despite these limitations, the automated framework developed in this study represents a significant advancement for regional reclamation planning, offering a solution to the incomplete inventories that currently focus on post-1977 permits. By delineating pre-SMCRA mining features, this model provides the foundational data needed to calculate cut-and-fill volumes, assess landslide risks, and prioritize sites for outcome-based funding, such as the Abandoned Mine Land Economic Revitalization (AMLER) program.

Future research should focus on four key areas to overcome current limitations:

- Hybrid Data Fusion: Integrating optical imagery (e.g., leaf-off winter orthophotos) as a secondary input channel could help distinguish between bare-rock natural cliffs and vegetated mine highwalls, potentially reducing commission errors.
- Transfer Learning: Testing the model's transferability to different geologic settings- such as the contour mines of Kentucky or the area mines of the Illinois Basin- would validate its utility as a universal tool for Appalachian relict mine mapping.
- Optimizing Input Dimensions: Systematically experimenting with larger input chip sizes (e.g., 1024×1024) could enhance the model's ability to capture the broader landscape context required to distinguish between extensive anthropogenic highwalls and localized natural scarps. Larger input dimensions would also minimize 'edge effects,' potentially reducing the fragmentation of linear features observed along chip boundaries.
- Topology-Aware Training: Implementing a zone-wise training approach stratified by sub-regional landform types (e.g., low hills, high mountains, and plateaus) would account for topographic variability. This targeted training strategy would enable the model to more effectively associate specific mining feature configurations with their corresponding topographic patterns, improving overall segmentation accuracy across diverse terrains.

5 Conclusions

This study presents a robust deep learning framework for the automated mapping of historic, pre-SMCRA surface mining features in the complex terrain of the central Appalachians. Contrary to the prevailing trend of utilizing increasingly deep neural networks, the Base U-Net outperformed more complex configurations, such as the ResU-Net and DiU-Net, in this geomorphic context. By avoiding the 'complexity trap,' the Base U-Net successfully learned the hierarchical geometries of legacy mine benches and highwalls without over-segmenting morphologically similar natural terrain. Furthermore, this research highlights that targeted hyperparameter tuning and specialized loss functions are just as critical as architectural choice; utilizing a moderate batch size stabilized

gradient update, while the implementation of a Hybrid Focal Tversky Loss effectively mitigated severe class imbalances by forcing the model to prioritize sparse anthropogenic boundaries. Ultimately, by relying exclusively on geometric signatures encoded in high spatial resolution LiDAR-derived land surface parameters (LSPs), this framework overcomes the limitations of optical remote sensing in densely forested areas. This approach provides a scalable solution to update incomplete historical databases, supplying essential geomorphic data for regional reclamation, heritage conservation, and spatial planning.

Funding:

This work was supported by the USDA National Institute of Food and Agriculture, Hatch project accession number 7004979, and the West Virginia Agricultural and Forestry Experiment Station.

Clinical trial number: not applicable.

Consent to Publish declaration: not applicable

Ethics and Consent to Participate declarations: not applicable

References:

Abraham, N., & Khan, N. M. (2019, April). A novel focal tversky loss function with improved attention u-net for lesion segmentation. In 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019) (pp. 683-687). IEEE. <https://doi.org/10.1109/ISBI.2019.8759329>

Ali, M., Maxwell, A. E., Shobe, C. M., & Farhadpour, S. (2025). Slope Instability Below Historic Mine Benches in the Northern Coalfields of West Virginia. *Environmental & Engineering Geoscience*, 31(4), 267-283.

Allaire, J. J., & Tang, Y. (2022). tensorflow: R interface to TensorFlow (R package version 2.10.1) [Computer software]. <https://cran.r-project.org/package=tensorflow>

Allaire, J. J., & Chollet, F. (2025). keras: R interface to Keras (R package version 2.16.0) [Computer software]. <https://cran.r-project.org/package=keras>

Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., & MacMillan, R. A. (2018). Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, 69(5), 757–770.

Bernhardt, E. S., & Palmer, M. A. (2011). The environmental costs of mountaintop mining valley fill operations for aquatic ecosystems of the Central Appalachians. *Annals of the New York Academy of Sciences*, 1223(1), 39-57.

Bernhardt, E. S., Lutz, B. D., King, R. S., Fay, J. P., Carter, C. E., Helton, A. M., ... & Amos, J. (2012). How many mountains can we mine? Assessing the regional degradation of central Appalachian rivers by surface coal mining. *Environmental science & technology*, 46(15), 8115-8122.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16.

- Bower, S. J., Shobe, C. M., Maxwell, A. E., & Campforts, B. (2024). The uncertain future of mountaintop-removal-mined landscapes 2: Modeling the influence of topography and vegetation. *Geomorphology*, 446, 108985.
- Brown, J. H. (1962). Success of tree planting on strip-mined areas in West Virginia (Bulletin 473). West Virginia University Agricultural Experiment Station.
- Burns, S. L. S. (2005). Bringing down the mountains: the impact of mountaintop removal surface coal mining on southern West Virginia communities, 1970–2004. West Virginia University.
- Cheng, G., Han, J., & Lu, X. (2017). Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10), 1865–1883.
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining*, 14(1), 13.
- Cribari, V., Staniscia, S., & Nahyan, M. (2025). Mapping landscape characters through spatial data and multivariate analysis: the case of the Coal River Watershed in West Virginia, USA. *Landscape Research*, 1-22.
- Cribari, V., Strager, M. P., Maxwell, A. E., & Yuill, C. (2021). Landscape changes in the southern coalfields of West Virginia: Multi-level intensity analysis and surface mining transitions in the Headwaters of the Coal River from 1976 to 2016. *Land*, 10(7), 748. <https://doi.org/10.3390/land10070748>
- Dai, H., Hu, X., Shu, Z., Qin, N., & Zhang, J. (2023). Deep ground filtering of large-scale ALS point clouds via iterative sequential ground prediction. *Remote Sensing*, 15(4), 961.
- Doghmane, R., & Boukari, K. (2024). Enhanced U-Net model for accurate aerial road segmentation. *Machine Graphics and Vision*, 33(1), 1–15.
- Diakogiannis, F. I., Waldner, F., Caccetta, P., & Wu, C. (2020). ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162, 94-114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>
- Drăguț, L., & Blaschke, T. (2006). Automated classification of landform elements using object-based image analysis. *Geomorphology*, 81(3-4), 330-344.
- Drummond, M. A., & Loveland, T. R. (2010). Land-use pressure and a transition to forest-cover loss in the eastern United States. *BioScience*, 60(4), 286-298.
- Esri Inc. (2018). ArcGIS Pro (Version 2.3) [Computer software]. Westlands, CA: Environmental Systems Research Institute. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>
- Farhadpour, S., & Maxwell, A. E. (2025). Comparing UNet configurations for anthropogenic geomorphic feature extraction from land surface parameters. *PLOS One*, 20(6), e0325904. <https://doi.org/10.1371/journal.pone.0325904>
- Ghosh, A., Ehrlich, M., Shah, S., Davis, L. S., & Chellappa, R. (2018). Stacked U-Nets for ground material segmentation in remote sensing imagery. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 257-261).

- Hay, G. J., Blaschke, T., Marceau, D. J., & Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, 57(5-6), 327–345.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Hijmans, R. (2024). terra: Spatial data analysis (R package version 1.7-83) [Computer software]. <https://CRAN.R-project.org/package=terra>
- Iglovikov, V., & Shvets, A. (2018). Ternaunet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv preprint arXiv:1801.05746*.
- Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). PMLR. <http://proceedings.mlr.press/v37/ioffe15.html>
- Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11), 4909–4918.
- Jaeger, K. L., & Ross, M. R. V. (2021). Identifying Geomorphic Process Domains in the Synthetic Landscapes of West Virginia, USA. *Journal of Geophysical Research: Earth Surface*, 126(3), e2020JF005851. <https://doi.org/10.1029/2020JF005851>
- Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2017). On large-batch training for deep learning: Generalization gap and sharp minima. In *Proceedings of the 5th International Conference on Learning Representations (ICLR)*. <https://openreview.net/forum?id=H1oyRIYgg>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://doi.org/10.48550/arXiv.1412.6980>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Küçükdemirci, M., Landeschi, G., Dell’Unto, N., & Ohlsson, M. (2021). Mapping archeological signs from airborne Lidar data using deep neural networks: Primary results. *ArcheoSciences*, 451(1), 291–293.
- Li, S., Xiong, L., Tang, G., & Strobl, J. (2020). Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. *Geomorphology*, 354, 107045. <https://doi.org/10.1016/j.geomorph.2020.107045>
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980-2988). <https://doi.org/10.1109/ICCV.2017.324>
- Lutz, B. D., Bernhardt, E. S., & Schlesinger, W. H. (2013). The environmental price tag on a ton of mountaintop removal coal. *PloS one*, 8(9), e73203.
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177.

- Maxwell, A. E., Bester, M. S., Guillen, L. A., Ramezan, C. A., Carpinello, D. J., Fan, Y., ... & Pylon, J. L. (2020). Semantic segmentation deep learning for extracting surface mine extents from historic topographic maps. *Remote Sensing*, 12(24), 4145. <https://doi.org/10.3390/rs12244145>
- Maxwell, A. E., Odom, W. E., Shobe, C. M., Doctor, D. H., Bester, M. S., & Ore, T. (2023). Exploring the influence of input feature space on CNN-based geomorphic feature extraction from digital terrain data. *Earth and Space Science*, 10, e2023EA002845. <https://doi.org/10.1029/2023EA002845>
- Maxwell, A. E., Shobe, C. M., & Bohnenstiehl, C. F. (2020). Assessing the generalization capability of a deep learning model for mapping valley fills in the central Appalachians. *Remote Sensing*, 12(14), 2269.
- Maxwell, A. E., & Strager, M. P. (2013). Assessing landform alterations induced by mountaintop mining. *Natural Science*, 5(2A), 229–237.
- Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International journal of remote sensing*, 39(9), 2784–2817.
- Meena, S. R., Soares, L. P., Grohmann, C. H., van Westen, C., Bhuyan, K., Singh, R. P., Floris, M., & Catani, F. (2022). Landslide detection in the Himalayas using machine learning algorithms and U-Net. *Landslides*, 19, 1209–1229.
- Milici, R.C. (2005). Appalachian coal assessment: Defining the coal systems of the Appalachian basin, in Warwick, P.D., ed., *Coal systems analysis: Geological Society of America Special Paper 387*, p. 9–30.
- Munn, R. (1973). The First Fifty Years of Strip Mining in West Virginia, 1916-1965. *West Virginia History*, 35, 66-74.
- Office of Surface Mining Reclamation and Enforcement. (2024). National Mine Map Repository. U.S. Department of the Interior. <https://www.osmre.gov/programs/national-mine-map-repository>
- Palmer, M. A., et al. (2010). Mountaintop mining consequences. *Science*, 327(5962), 148–149.
- Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R. V., Clinton, N. E., Campagna, D. J., Franklin, Y., Bernhardt, E. S., & Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. *PLOS ONE*, 13(7), e0197758.
- R Core Team. (2024). R: A language and environment for statistical computing (Version 4.4.1) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reed, M., & Kite, S. (2020). Peripheral gully and landslide erosion on an extreme anthropogenic landscape produced by mountaintop removal coal mining. *Earth Surface Processes and Landforms*, 45(9), 2078–2090.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195–204.
- Rizaldy, A., Persello, C., Gevaert, C., Oude Elberink, S., & Vosselman, G. (2018). Ground and multi-class classification of airborne laser scanner point clouds using fully convolutional networks. *Remote sensing*, 10(11), 1723.

- Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Cham: Springer international publishing.
- Ross, M. R., McGlynn, B. L., & Bernhardt, E. S. (2016). Deep impact: Effects of mountaintop mining on surface topography, bedrock structure, and downstream waters. *Environmental science & technology*, 50(4), 2064-2074.
- Saha, A. K., Gupta, R. P., & Arora, M. K. (2002). GIS-based landslide hazard zonation in the Bhagirathi (Ganga) valley, Himalayas. *International journal of remote sensing*, 23(2), 357-369.
- Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In Q. Wang, Y. Shi, H.-I. Suk, & K. Suzuki (Eds.), *Machine Learning in Medical Imaging (MLMI 2017, Lecture Notes in Computer Science, Vol. 10541, pp. 379-387)*. Springer. https://doi.org/10.1007/978-3-319-67389-9_44
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Skousen, J., & Zipper, C. (2014). Post-mining lands use options in the Appalachian Coal Mining Region of the USA. In *Proceedings of Life of Mine Conference* (pp. 153-164).
- Skousen, J., renton, K., & Brown, T. (2006). Tree recruitment and growth on 20-year-old, unreclaimed surface mined lands in West Virginia. *International Journal of Mining, Reclamation and Environment*, 20(2), 142-154.
- Tewalt, S.J., and Ruppert, L.F. (2014). Coal assessments and coal research in the Appalachian basin, chap. D.4 of Ruppert, L.F., and Ryder, R.T., eds., *Coal and petroleum resources in the Appalachian basin; Distribution, geologic framework, and geochemical character: U.S. Geological Survey Professional Paper 1708*, 15 p., <http://dx.doi.org/10.3133/pp1708D.4>.
- Trier, Ø. D., Cowley, D. C., & Waldeland, A. U. (2019). Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. *Archaeological Prospection*, 26(2), 165-175.
- Vakalopoulou, M., Karantzas, K., Komodakis, N., & Paragios, N. (2015). Building detection in very high resolution multispectral data with deep learning features. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1873-1876.
- Velikić, I. M., Rabrenović, D., Srećković-Batočanin, D., & Spahić, D. (2022). Geoheritage and mining heritage in the promotion of theme parks: an example of the National Park Đerdap (Carpathian-Balkan Thrust-And-Fold Belt, Eastern Serbia). *Geoheritage*, 14(3), 102.
- West Virginia GIS Technical Center. (n.d.). Digital Elevation Model 1-Meter Statewide Mosaic. Retrieved September 5, 2025, from <https://wvgis.wvu.edu/data/dataset.php?ID=477>
- West Virginia GIS Technical Center. (2020). West Virginia NAIP 2020 60cm Imagery [Map service]. West Virginia University. https://services.wvgis.wvu.edu/arcgis/rest/services/Imagery_BaseMaps_Earth-Cover/wv_imagery_NAIP_2020_60cm/MapServer

- West Virginia Geological and Economic Survey. (2024). Coal Bed Mapping Project (CBMP). Retrieved July 6, 2025, from <http://www.wvgs.wvnet.edu/www/coal/cbmp/coalims.html>
- Wu, Y., & He, K. (2018). Group normalization. In Proceedings of the European conference on computer vision (ECCV) (pp. 3-19).
- Yi, Y., & Zhang, W. (2020). A New Deep-Learning-Based Approach for Earthquake-Triggered Landslide Detection From Single-Temporal RapidEye Satellite Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6166–6176. <https://doi.org/10.1109/JSTARS.2020.3028855>
- Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.
- Yuan, J. (2016). Automatic building extraction in aerial scenes using convolutional networks. arXiv preprint arXiv:1602.06564.
- Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5), 749-753.
- Zipper, C. E., Adams, M. B., & Skousen, J. (2020). The Appalachian coalfield in historical context. In *Appalachia's coal-mined landscapes: Resources and communities in a new energy era* (pp. 1–26). Springer International Publishing. https://doi.org/10.1007/978-3-030-57780-3_1