

Appendix C:

Machine Learning for Remote Sensing: Foundations and Applications for Battlefield Damage Assessment of Urban Structures

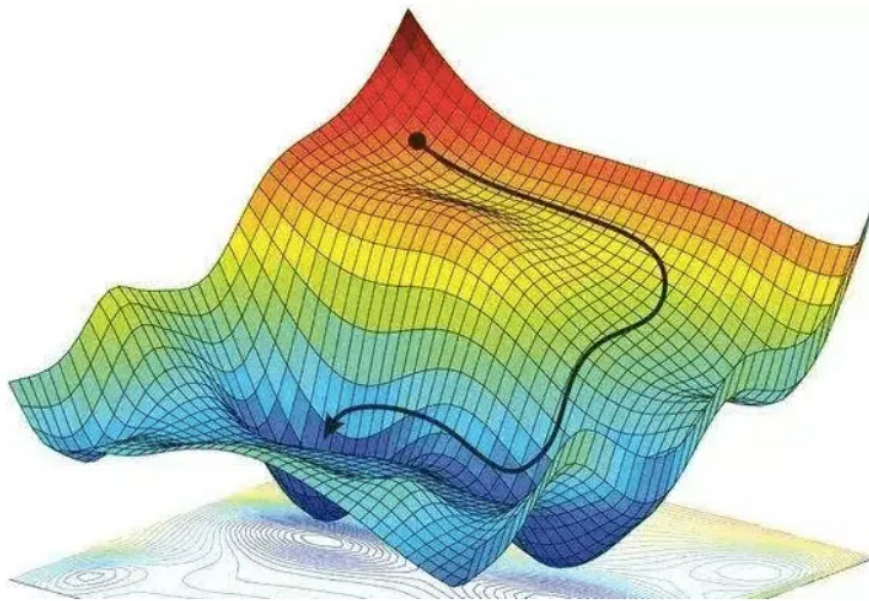


Figure 1: Gradient descent in 3 Dimensions. Source:(Alexander Amini & Daniela Rus, 2018)

Author : Marco Heinzen

Contents

Appendix C:	1
1. Foundational Methods	4
1.1. Dimensionality Reduction Beyond PCA: t-SNE and Autoencoders.....	4
1.2. Gradient Descent: Optimization in Machine Learning	4
1.3. Laplacian and Gaussian operators	6
1.4. Kernels in Digital Image Processing and Remote Sensing	8
1.5. Principal Component Analysis (PCA) for Dimensionality Reduction.....	9
2. Classical and Traditional Machine Learning for Remote Sensing and Land Use Classification.....	12
2.1. Land Use and Land Cover (LULC) Classification Workflows.....	12
2.2. Support Vector Machines (SVM) for Remote Sensing Classification.....	13
2.3. K-Nearest Neighbors (KNN) and Instance-Based Learning.....	14
2.4. Decision Trees in Remote Sensing	15
2.5. Linear and Polynomial Regression in Remote Sensing.....	16
2.6. Logistic Regression in Remote Sensing	17
2.7. Naïve Bayes in Remote Sensing	17
2.8. Linear Discriminant Analysis (LDA) in Remote Sensing	18
2.9. Pixel-Based Image Analysis (PBIA) with Classical Classifiers.....	18
2.10. Object-Based Image Analysis (OBIA) with Classical Classifiers	19
3. Foundations of Neural Networks.....	20
3.1. Artificial Neural Networks.....	20
3.2. Perceptron.....	20
3.3. Activation Functions	21
3.4. Loss Functions.....	22
3.5. Backpropagation.....	24
3.6. ANN Architectures, Input Layers, Hidden Layers, Output Layers	24
3.7. From Gradients to Convolution	25
4. Core Architectures.....	26
4.1. Convolutional Neural Networks (CNNs): Foundations and Applications.....	26
4.2. U-Net and Variants: Semantic Segmentation for Remote Sensing	28
4.3. SegNet, UNet++, and V-Net: Encoder–Decoder Architectures for Image Segmentation	30
4.4. ResNet: Deep Residual Learning for Feature Extraction	31
4.5. DenseNet and EfficientNet: Modern Backbone Networks.....	32
5. Advanced and Emerging Approaches	34
5.1. Transfer Learning.....	34

5.2. Siamese Networks and Attention Mechanisms for Change Detection.....	35
5.3. Generative Adversarial Networks (GANs) for Data Augmentation and Damage Inference	36
5.4. Graph Neural Networks (GNNs) for Spatial Relations in Urban Damage Assessment	38
5.5. Contrastive and Self-Supervised Learning in Remote Sensing	39
5.6. Ensemble and Hybrid Methods: Random Forests, Gradient Boosting, Stacking, Bagging and Model Fusion.....	40
5.7. BDANet and other Specialized Architectures for Building Damage Assessment.....	42
6. Validation and Testing	45
6.1. Root Mean Square Error (RMSE) for Quantitative Accuracy	45
6.2. Precision, Recall, and F1 Score in Classification Tasks.....	45
6.3. Intersection over Union (IoU) and Segmentation Metrics	46
6.4. Overall Accuracy and Kappa Coefficient	46
6.5. Cross-Validation and Train/Test Splits	46
6.6. Confusion Matrices.....	47
6.7. Ground Truth and Benchmark Dataset Validation	48
6.8. AUC, SSIM, and Task-Specific Metrics.....	49
7. Glossary and Definitions	50
8. Acronyms and Abbreviations.....	56
9. List of Figures.....	59
10. List of Tables	59
11. References.....	60
12. Disclaimer	82

1. Foundational Methods

Foundational methods provide the mathematical and algorithmic basis upon which more complex machine learning models are built. Optimization techniques such as gradient descent enable algorithms to iteratively learn from vast amounts of data by minimizing error functions, a crucial process when dealing with the complexity of remote sensing imagery (I. Goodfellow et al., 2016). Equally important are dimensionality reduction strategies, which confront the challenge of high-dimensional data typical of multispectral and hyperspectral sensors. Linear approaches like Principal Component Analysis (PCA) compress redundant information into fewer dimensions (Bioucas-Dias et al., 2013), while nonlinear techniques such as t-SNE (van der Maaten & Hinton, 2008) or autoencoders uncover hidden structures and nonlinear relationships in heterogeneous datasets (Hinton & Salakhutdinov, 2006)

1.1. Dimensionality Reduction Beyond PCA: t-SNE and Autoencoders

While Principal Component Analysis (PCA) remains a standard linear method, many remote sensing problems involve nonlinear relationships that PCA cannot fully capture. Techniques such as t-distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008, Visualizing Data using t-SNE) provide a nonlinear approach for mapping high-dimensional data into low-dimensional spaces by preserving local neighborhood structure. In hyperspectral analysis, t-SNE has been used to reveal class separability in feature space, offering intuitive visualizations of spectral clusters that highlight land-cover types or damaged structures (Zhang et al., 2017, Hyperspectral Dimensionality Reduction Using Locality Preserving Embedding and t-SNE). Unlike PCA, which emphasizes variance, t-SNE emphasizes similarity between data points, making it useful for exploratory analysis of complex multimodal or multi-temporal imagery.

Another widely adopted nonlinear approach is the use of autoencoders, neural networks designed to compress input data into a low-dimensional latent space and reconstruct it back with minimal loss (Hinton & Salakhutdinov, 2006, Reducing the Dimensionality of Data with Neural Networks). Autoencoders have been applied to multispectral and hyperspectral imagery to learn compact, discriminative representations that retain essential spectral-spatial features while suppressing noise (Paoletti et al., 2019, Deep Learning Classifiers for Hyperspectral Imaging: A Review). Variants such as denoising autoencoders and convolutional autoencoders are particularly well suited for remote sensing tasks, where they enable effective dimensionality reduction prior to classification or segmentation. Together, t-SNE and autoencoders extend beyond PCA by uncovering nonlinear structures in the data, making them powerful tools for modern machine learning pipelines in Earth observation.

1.2. Gradient Descent: Optimization in Machine Learning

Gradient descent is a fundamental optimization method in machine learning that relies on concepts from calculus, specifically partial differentiation. In a one-dimensional setting, optimization is performed by moving iteratively in the opposite direction of the derivative of a function, which indicates the steepest descent toward a local minimum. Extending this to higher dimensions requires partial derivatives. For a function of two variables, $f(x, y)$ the

gradient is defined as the vector of its partial derivatives (Nocedal Jorge & Wright Stephen J., 2006):

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)$$

and for a function of three variables, $f(x, y, z)$ the gradient extends to:

$$\nabla f(x, y, z) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right)$$

These gradient vectors point in the direction of the steepest increase of the function. Gradient descent uses their negative direction to iteratively update parameters, thereby approaching a minimum (I. Goodfellow et al., 2016; Nocedal Jorge & Wright Stephen J., 2006).

The main principle of gradient descent is to optimize an objective function $J(\theta)$ typically representing error or loss in machine learning, by adjusting parameters θ step by step. The update rule is:

$$\theta_{t+1} = \theta_t - \eta \nabla J(\theta_t)$$

Where η is the learning rate controlling the step size. A small η results in slow convergence, while a large one may cause divergence. This iterative optimization enables algorithms to minimize complex loss functions arising in classification, regression, or neural network training (Rumelhart et al., 1986).

In remote sensing and damage assessment, gradient descent plays a pivotal role by enabling deep learning models to iteratively adjust their internal parameters in response to high-dimensional inputs such as multispectral, hyperspectral, or synthetic aperture radar (SAR) imagery (Ball et al., 2017; I. Goodfellow et al., 2016). These data sources often contain thousands of correlated spectral features, and without efficient optimization, models would struggle to converge toward meaningful representations (Bioucas-Dias et al., 2013).

By following the negative gradient of a loss function, gradient descent ensures that convolutional neural networks and related architectures progressively refine their filters to detect relevant (spatial) patterns (Lecun et al., 1998) such as building edges, road networks, or debris fields (Xu et al., 2019), as well as spectral cues associated with material properties (X. X. Zhu et al., 2017).

Variants of gradient descent address the computational challenges posed by large-scale Earth observation datasets. Batch gradient descent is conceptually straightforward but requires processing entire datasets before weight updates, which is impractical with terabyte-scale imagery (Nocedal Jorge & Wright Stephen J., 2006). Stochastic Gradient Descent (SGD) alleviates the computational burden of full-batch updates by adjusting parameters after each sample rather than waiting for an entire dataset to be processed. Its inherent stochasticity injects noise into the optimization trajectory, which in practice can help models escape shallow local minima and saddle points. This behavior is not a guarantee of reaching a global minimum but reflects empirical properties of SGD in high-dimensional deep learning landscapes. (I. Goodfellow et al., 2016; Jin et al., 2017; Keskar et al., 2017; Robbins, Herbert

& Monro, Sutton, 1951). Foundational work introduced SGD as a stochastic approximation method (Robbins, Herbert & Monro, Sutton, 1951), while subsequent analyses demonstrated its ability to avoid sharp minima and improve generalization (I. Goodfellow et al., 2016; Keskar et al., 2017). Theoretical results further show that SGD-type methods can escape saddle points efficiently under mild conditions (Jin et al., 2017). Mini-batch gradient descent strikes a balance, making optimization scalable on GPUs while retaining stability (I. Goodfellow et al., 2016). Adaptive optimizers such as Adam (Kingma & Ba, 2017) and RMSProp (Ruder, 2017) further accelerate convergence in deep architectures. In the context of damage assessment from satellite imagery, these optimization techniques allow networks such as U-Net, ResNet, and BDANet to efficiently learn robust features across diverse urban environments. This enables reliable detection of collapsed or damaged structures, classification of damage severity, and robust change detection in multi-temporal imagery (Ahmadi et al., 2023; Kerle et al., 2020; Shen et al., 2022).

In image processing, the gradient operator is mostly applied for edge detection, since edges represent abrupt spatial transitions in brightness or reflectance values (Canny, 1986). In a one-dimensional signal, these correspond to peaks or valleys in the first derivative (Marr & Hildreth, 1980), while in two-dimensional imagery the gradient is defined by the partial derivatives with respect to x and y , capturing both the magnitude and direction of the steepest intensity change (Gonzalez & Woods, 2018). This makes the gradient vector perpendicular to the orientation of an edge (Jain et al., 1995). By thresholding the gradient magnitude, it is possible to delineate edges corresponding to object boundaries, such as building outlines or road networks in aerial and satellite imagery (Richards, 2022). In remote sensing, gradient-based edge detection serves as a critical preprocessing step for segmentation, object extraction, and change detection, particularly when mapping infrastructure damage in urban environments (Blaschke, 2010).

1.3. Laplacian and Gaussian operators

While the gradient operator relies on first-order derivatives, an alternative approach for edge detection uses the second derivative (Canny, 1986; Marr & Hildreth, 1980). In one dimension, the second derivative identifies changes in the slope of the first derivative, with zero-crossings marking the location of edges (Gonzalez & Woods, 2018). Extending this to two dimensions leads to the Laplacian operator, defined as the sum of the second derivatives in the x and y directions (Jain et al., 1995; Richards, 2022):

1.3.1. Laplacian Operator:

$$\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$$

The Laplacian highlights regions of rapid intensity change regardless of direction, producing edge maps that are less sensitive to orientation than gradient-based methods. However, because it amplifies high-frequency signals, it is also more sensitive to noise (Marr & Hildreth, 1980).

1.3.2. Gaussian Kernel (2D):

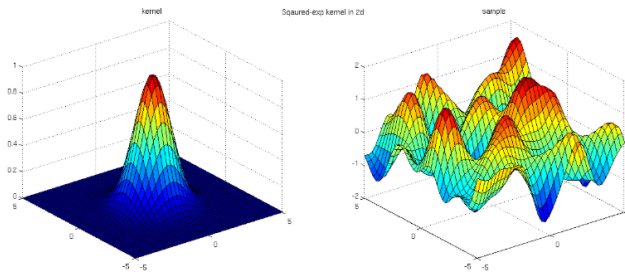


Figure 2: Gaussian Kernel in 3d. Source:(Kernel Cookbook, n.d.)

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

To address noise sensitivity, the Gaussian operator is often employed as a smoothing filter (Gonzalez & Woods, 2018). The Gaussian function, shaped like a bell curve, weights nearby pixels more strongly than distant ones, effectively reducing high-frequency fluctuations while preserving larger-scale structures (Lindeberg, 1994). When an image is convolved with a Gaussian kernel, the result is a smoothed version in which edges remain detectable, but small, spurious variations are suppressed (Marr & Hildreth, 1980). This is particularly valuable in remote sensing imagery, where atmospheric effects or sensor noise can obscure subtle features such as building cracks or burn scars (Y. Chen et al., 2011; Richards, 2022).

1.3.3. Laplacian of Gaussian (LoG):

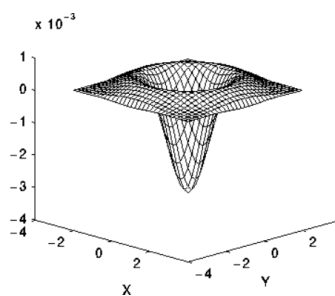


Figure 3: "Mexican Sombrero" - LoG. Source: (Spatial Filters - Laplacian/Laplacian of Gaussian, n.d.)

$$\nabla^2 G(x, y) = \frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} * \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

The power of linearity allows these operations to be combined. The Laplacian-of-Gaussian (LoG) operator first applies Gaussian smoothing to reduce noise and then the Laplacian to

detect edges in a single convolution. Conceptually, the LoG resembles a “Mexican hat” function in two dimensions, capturing edges as zero-crossings of the smoothed second derivative. This approach is widely used in computer vision and remote sensing tasks where robust edge detection is essential, such as delineating building footprints from very high-resolution aerial imagery or identifying rubble boundaries in post-disaster scenes (Marr & Hildreth, 1980; Richards, 2022).

Finally, both gradient-based and Laplacian-based approaches are often applied together in modern image analysis (Canny, 1986; Marr & Hildreth, 1980). Gradient methods provide edge orientation and strength, while Laplacian and Gaussian operators improve robustness to noise and highlight edges in a scale-invariant manner (Gonzalez & Woods, 2018). This combination underpins many advanced edge detectors and segmentation frameworks used in Earth observation and damage assessment workflows, serving as a bridge between low-level image processing and higher-level machine learning models (Blaschke, 2010; Richards, 2022).

1.4. Kernels in Digital Image Processing and Remote Sensing

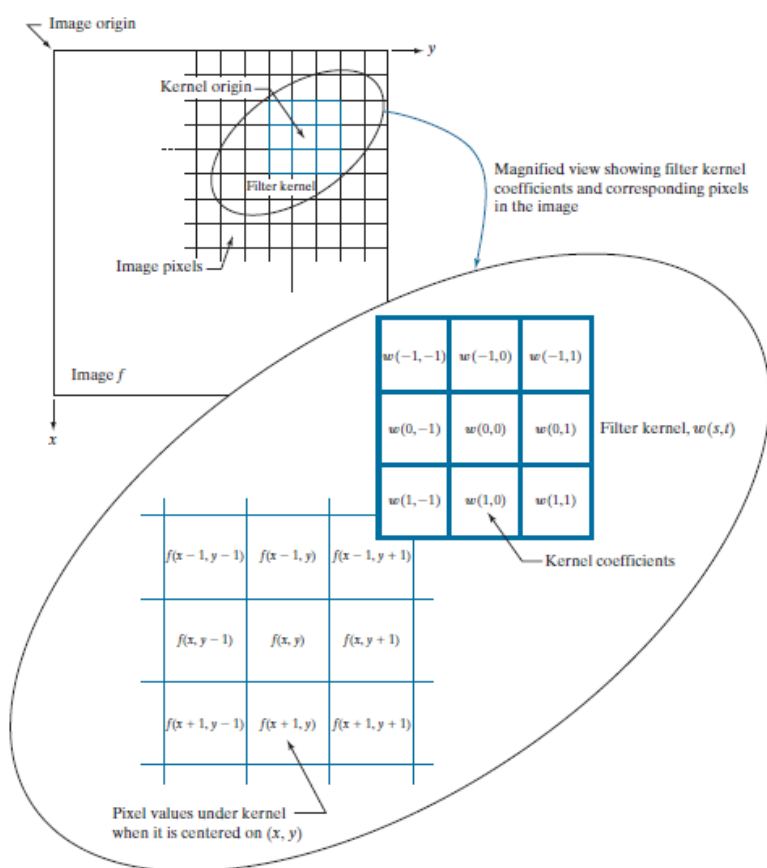


Figure 4: The mechanics of linear spatial filtering using a 3×3 kernel. The pixels are shown as squares to simplify the graphics. Note that the origin of the image is at the top left, but the origin of the kernel is at its center. Placing the origin at the center of spatially symmetric kernels simplifies writing expressions for linear filtering. (Gonzalez & Woods, 2018)

In digital image processing, continuous mathematical functions must be translated into operations on discrete pixel grids. Convolution, which in continuous mathematics integrates a function with a filter, becomes in practice the application of a kernel, a small sliding window

that moves across the image. At each position, the pixel values of the image are multiplied by the corresponding entries of the kernel, and the products are summed to produce a new pixel value. In this way, kernels provide a practical implementation of convolution for discrete images, enabling tasks such as smoothing, sharpening, and edge detection (Gonzalez & Woods, 2018).

Kernels are flexible because they approximate continuous operators like the gradient, Laplacian, or Gaussian in discrete space. Simple kernels such as Roberts or Prewitt filters compute first derivatives over a 2×2 or 3×3 neighborhood, while the Sobel operator extends this idea by incorporating smoothing into the derivative estimation, thereby combining edge detection with noise suppression. More complex kernels, like the Laplacian-of-Gaussian (“Mexican hat”), encode both smoothing and second derivative calculations into a single operator. These kernel-based filters form the basis of many common image processing operations in remote sensing, where they are used to enhance roads, delineate field boundaries, or prepare imagery for segmentation and classification (Richards, 2022).

The power of kernels lies not only in their mathematical basis but also in their adaptability. Operators can be combined - Gaussian smoothing before a Laplacian, or Sobel filters after Gaussian noise reduction - and because convolution is linear, these combinations can be precomputed as single kernels. Furthermore, the size of the kernel (e.g. 3×3 , 5×5 , 7×7) determines the scale of features emphasized: small kernels highlight fine details such as cars or narrow roads, while larger kernels capture broader structures like fields or forest patches. This scale-dependence makes kernel design a crucial step in preprocessing workflows for Earth observation, ensuring that subsequent algorithms for segmentation, object detection, or change analysis receive cleaner and more interpretable inputs (Blaschke, 2010).

1.5. Principal Component Analysis (PCA) for Dimensionality Reduction

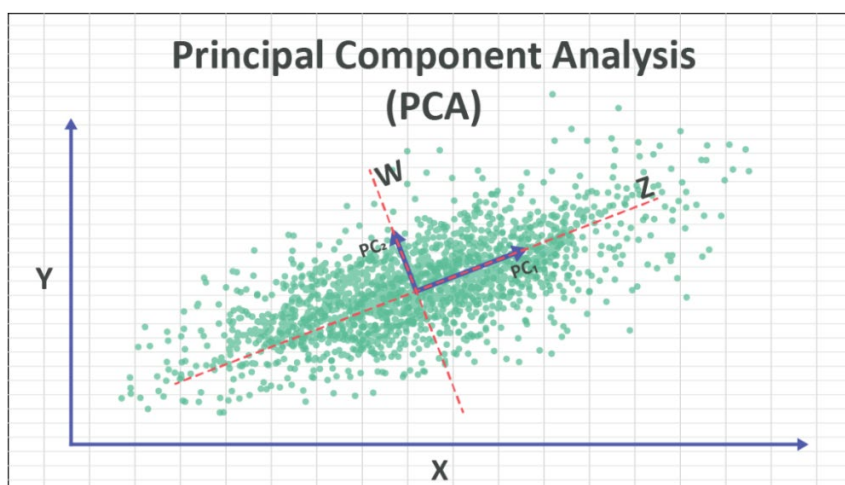


Figure 5: PCA (Marridi, 2016)

Principal Component Analysis (PCA) is one of the most widely used dimensionality reduction methods in remote sensing, particularly for hyperspectral imagery where hundreds of spectral

bands are available. Richards & Jia (originally in 2006) describe PCA as a fundamental statistical method for reducing data dimensionality while preserving variance, and note its long-standing adoption in remote sensing (Richards, 2022). Hyperspectral sensors such as Hyperion (Pearlman et al., 2003), AVIRIS (R. O. Green et al., 1998), or EnMAP (Guanter et al., 2015) routinely capture hundreds of narrow and often correlated spectral channels. While this spectral richness is valuable for discriminating materials, vegetation types, or subtle damage signatures, it also introduces redundancy and noise, making direct analysis computationally expensive and prone to overfitting (Bioucas-Dias et al., 2013).

PCA addresses this challenge by transforming correlated input bands into a new set of linearly uncorrelated variables called principal components (PCs). These components are ranked by the variance they explain, with the first few capturing the majority of spectral variability and subsequent components increasingly dominated by noise (Bioucas-Dias et al., 2013; Richards, 2022). In hyperspectral applications, the first principal component often reflects overall scene brightness, while subsequent components highlight differences between land-cover types or material properties (A. A. Green et al., 1988; Richards, 2022).

1.5.1. Mathematical Foundations

The foundation of PCA lies in the Eigen-decomposition of the covariance matrix of the data (I. T. Jolliffe, 2002). Suppose we have a dataset X with n observations and p correlated variables (e.g., spectral bands). The first step is to compute the covariance matrix:

$$\Sigma = \frac{1}{n-1} X^{\{T\}} X$$

This matrix describes how spectral bands vary together. PCA then finds the eigenvalues λ_i and eigenvectors e_i of Σ :

$$\Sigma e_i = \lambda_i * e_i$$

The eigenvectors define the new orthogonal axes (principal components), and the eigenvalues indicate the amount of variance explained by each component. The data are projected into this new basis by:

$$Z = XE$$

where E is the matrix of eigenvectors, and Z contains the principal component scores. In practice, the first few components (those with the largest λ_i are retained, while later ones are often discarded as noise (I. T. Jolliffe, 2002).

1.5.2. Applications in Remote Sensing

In hyperspectral remote sensing, the first principal component typically reflects overall scene brightness, while subsequent components highlight differences between land-cover types or damaged and intact structures (Richards, 2022). Analysts can visualize these components in RGB composites to emphasize patterns such as urban destruction, vegetation stress, or

burned areas (A. A. Green et al., 1988). PCA also serves as a powerful preprocessing step: compressing hyperspectral data to a smaller set of uncorrelated features makes downstream classification with SVMs, Random Forests, or CNNs more efficient and less prone to overfitting (Plaza et al., 2009).

Beyond hyperspectral analysis, PCA is widely applied to multispectral imagery and radar data. For example, PCA reduces the dimensionality of Sentinel-2 or Landsat images for vegetation and urban mapping (Jolliffe & Cadima, 2016) and in polarimetric SAR, it helps identify dominant scattering mechanisms by decorrelating polarization channels (Lee & Pottier, 2017). Variants such as the Minimum Noise Fraction (MNF) transform, implemented in ENVI, extend PCA by separating signal-dominated and noise-dominated components, allowing reconstruction of denoised spectra for more accurate classification (A. A. Green et al., 1988). Despite its linearity assumption and the risk of discarding low-variance but meaningful information, PCA remains a cornerstone of dimensionality reduction in Earth observation, balancing efficiency with interpretability for tasks such as land-cover mapping, urban monitoring, and building damage assessment (I. T. Jolliffe, 2002; Richards, 2022).

2. Classical and Traditional Machine Learning for Remote Sensing and Land Use Classification

Before the rise of deep learning architectures, a broad range of classical machine learning methods formed the backbone of remote sensing analysis, particularly in land use and land cover classification. These approaches offer robust and interpretable tools for handling the spectral, spatial, and textural variability present in satellite and aerial imagery (Meyer, 2019). Techniques such as support vector machines (Platt, 2000) and k-nearest neighbors (KNN) (Cover & Hart, 1967), along with decision trees (Friedl & Brodley, 1997; Thakur & Panse, 2022) and logistic regression (CART) (Breiman et al., 2017), provided flexible classifiers adaptable to different data types and resolutions. Object-based image analysis (OBIA) workflows extended these methods by shifting from pixel-level to object-level interpretation, stabilizing change analyses in dense built environments (Al-Khudhairy et al., 2005). Methods like Random Forests remain widely used, either as standalone classifiers or as components of ensemble frameworks, due to their efficiency and resilience to noise (Dietrich et al., 2025).

2.1. Land Use and Land Cover (LULC) Classification Workflows

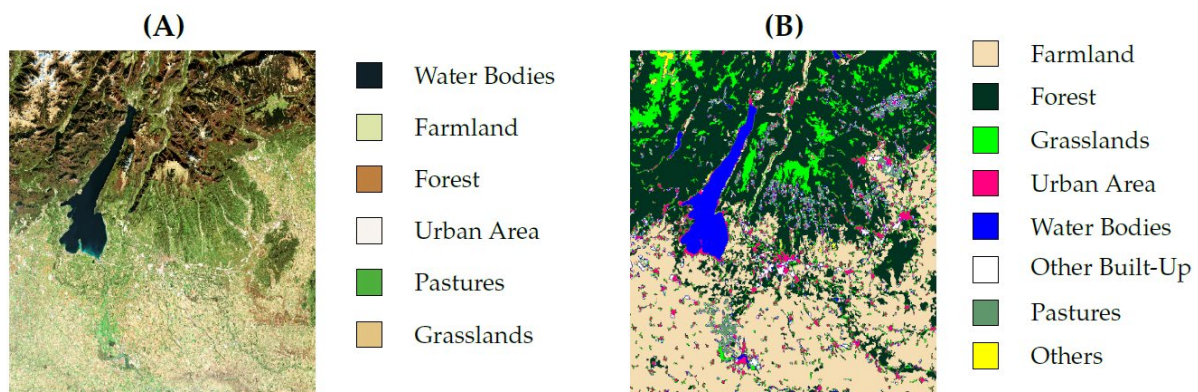


Figure 6: True colour (A) and land classification (B) images of the studied area.

Land Use and Land Cover (LULC) classification forms one of the most established applications of remote sensing and provides the methodological foundation for many modern approaches to damage assessment. Traditional workflows typically begin with data preprocessing, including radiometric and geometric corrections, atmospheric adjustment, and often noise filtering to ensure comparability across space and time (Richards, 2022). Following preprocessing, features are extracted from satellite or aerial imagery that represent the spectral, spatial, or textural characteristics of the landscape. These features may include vegetation indices such as NDVI, texture measures, or dimensionality-reduced components obtained through PCA or Minimum Noise Fraction (MNF) transforms (Bioucas-Dias et al., 2013).

Classification itself can proceed along two main paradigms: pixel-based classification, which assigns each pixel to a class using its spectral signature, and object-based classification (OBIA), which groups neighboring pixels into meaningful image objects prior to classification (Blaschke, 2010). Pixel-based approaches such as maximum likelihood classification or support

vector machines (SVMs) remain popular for multispectral and hyperspectral data (Mountrakis et al., 2011; Richards, 2022), while OBIA has gained traction with very high-resolution imagery because it integrates spectral and spatial information (Blaschke, 2010). Both paradigms can employ a range of classifiers, from traditional statistical methods to ensemble learning (Gislason et al., 2006) and modern deep learning networks, depending on the scale and complexity of the application (X. X. Zhu et al., 2017).

LULC classification workflows are not only central to mapping vegetation, urban structures, and land-cover changes but also serve as a baseline for post-disaster analysis. For example, classifying built-up versus non-built-up areas enables subsequent damage detection algorithms to focus on relevant structures, while distinguishing cropland from forest or water bodies helps quantify secondary impacts of conflict or natural hazards (L. Liu et al., 2021).

2.2. Support Vector Machines (SVM) for Remote Sensing Classification

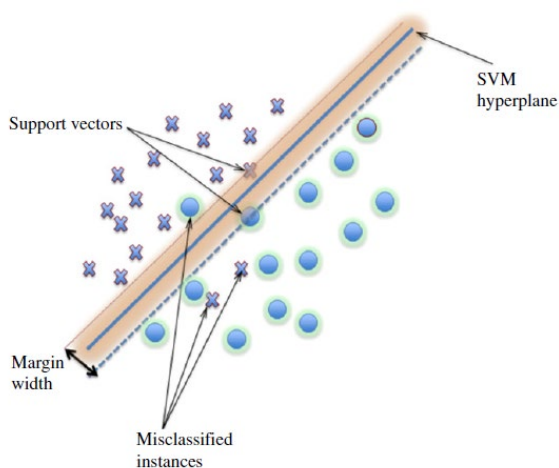


Figure 7: Linear support vector machine example. (Mountrakis et al., 2011)

Support Vector Machines (SVMs) are a widely used supervised learning method in remote sensing, particularly valued for their ability to handle high-dimensional and nonlinear data. SVMs operate by finding an optimal hyperplane that maximizes the margin between different classes in the feature space. For linearly inseparable cases, kernel functions such as radial basis functions (RBF), polynomial, or sigmoid kernels are applied to transform the input data into higher-dimensional spaces where separation becomes feasible (Cortes & Vapnik, 1995). This ability to adapt to complex decision boundaries makes SVMs highly suitable for spectral datasets with many correlated variables, such as hyperspectral imagery (Pal & Mather, 2005).

In the context of remote sensing, SVMs have been applied successfully to both multispectral and hyperspectral image classification tasks. Unlike parametric classifiers such as maximum likelihood, SVMs are non-parametric and do not require assumptions about the underlying data distribution. This property is particularly advantageous for hyperspectral data, where the number of spectral bands far exceeds the number of training samples (the “Hughes phenomenon”) (Mountrakis et al., 2011). Numerous studies demonstrate that SVMs

outperform traditional classifiers in mapping land-cover types, vegetation classes, and urban areas, owing to their robustness in high-dimensional feature spaces (Pal & Mather, 2005).

SVMs have also played an important role in advancing damage assessment and change detection workflows (Mountrakis et al., 2011). By exploiting kernel methods to capture subtle spectral and spatial differences, SVMs have been successfully applied to discriminate between damaged and intact structures (Li et al., 2009; Taskin Kaya et al., 2011), to detect burned areas (Petropoulos et al., 2011) and to identify land-cover transitions in post-disaster contexts (Taskin Kaya et al., 2011). While deep learning approaches increasingly dominate recent literature, SVMs remain relevant due to their low computational demands (Pal & Mather, 2005) strong generalization with limited training samples (Foody & Mathur, 2004) and proven track record across diverse sensors and applications (Mantero et al., 2005). As such, SVMs continue to bridge traditional LULC workflows and advanced machine learning pipelines in Earth observation (Mountrakis et al., 2011).

2.3. K-Nearest Neighbors (KNN) and Instance-Based Learning

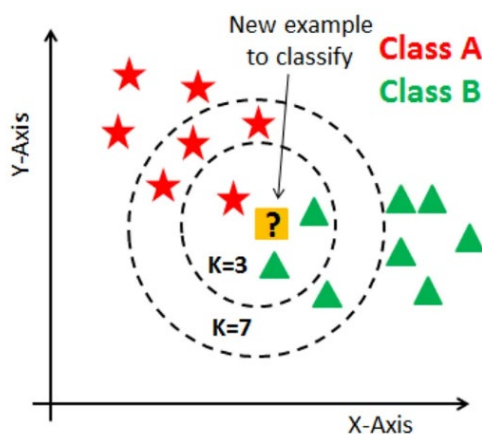


Figure 8: KNN Illustration. Source: (Shah, 2021)

The K-Nearest Neighbors (KNN) algorithm is one of the simplest and most intuitive machine learning classifiers, yet it remains widely used in remote sensing for land-cover classification and image interpretation. KNN is an instance-based learning method, meaning it does not learn explicit decision boundaries during training but instead classifies new observations by comparing them to the most similar examples in the training dataset. Classification is achieved by identifying the k nearest samples in the feature space—typically based on Euclidean or Mahalanobis distance—and assigning the class label most common among these neighbors (Cover & Hart, 1967).

In remote sensing, KNN is attractive for its non-parametric nature and its ability to handle complex and non-linear distributions of spectral signatures without assuming a statistical model. This makes it suitable for both multispectral and hyperspectral datasets, where different land-cover types may form irregular clusters in spectral space. Applications include vegetation mapping, urban land-use classification, and environmental monitoring, where KNN has been shown to perform competitively with more complex classifiers under certain conditions (Fix & Hodges, 1989; Pal & Mather, 2005).

Despite its simplicity, KNN has also been used in more advanced workflows, such as change detection and damage mapping, by leveraging spectral–spatial features. Because KNN directly compares observations to training data, its performance is sensitive to the quality and representativeness of the training set, as well as the choice of distance metric and the value of k . It can struggle with high-dimensional data due to the “curse of dimensionality,” though techniques such as feature selection, PCA, or metric learning can mitigate these effects (Hughes, 1968).

2.4. Decision Trees in Remote Sensing

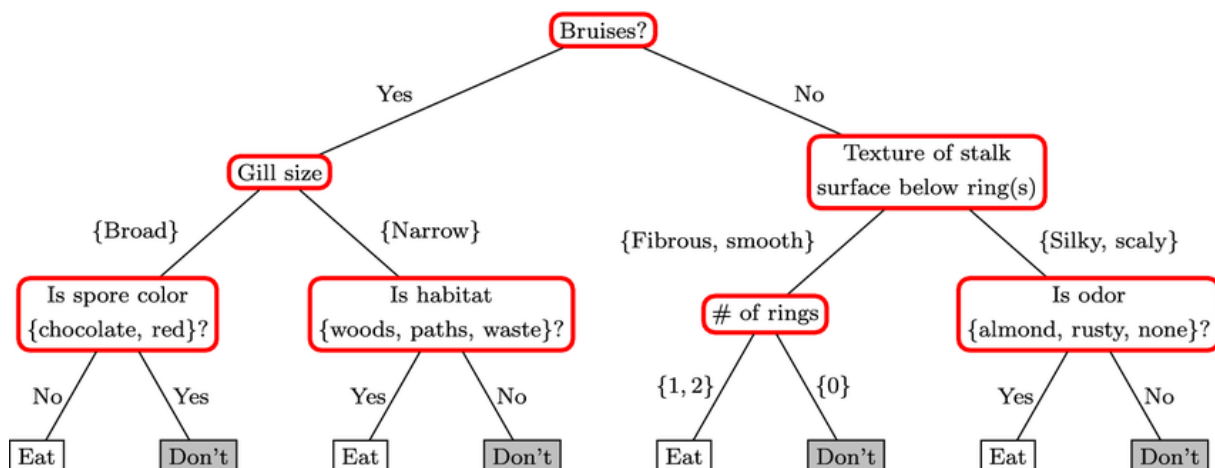


Figure 9: Decision Tree Example for Mushrooms. Source: (Gunluk et al., 2021)

Decision trees are supervised, non-parametric classifiers that partition feature space into hierarchical regions using a sequence of threshold rules. Each internal node represents a decision on one of the input variables, while each terminal node assigns a class label. In remote sensing, decision trees have been widely adopted for land-cover and land-use classification because they can accommodate nonlinear relationships, handle both continuous and categorical variables, and integrate spectral, spatial, and contextual features within a unified framework. Their hierarchical, rule-based output also makes them more interpretable than many statistical classifiers, allowing analysts to identify which spectral bands or indices contribute most to class separation. Their rule-based structure is also relatively robust to missing data, which is advantageous when combining heterogeneous datasets across regions and seasons (Pal & Mather, 2003; Quinlan, 1986).

One of the strengths of decision trees in Earth observation is their ability to integrate multi-source and multi-resolution data. For example, tree-based models can simultaneously use spectral reflectance from Landsat or Sentinel imagery, textural information from very-high-resolution optical data, and ancillary variables such as elevation or slope. This flexibility has enabled applications ranging from vegetation type mapping and forest monitoring to urban growth analysis and agricultural crop classification (Hansen et al., 2000).

Despite these strengths, single decision trees are prone to overfitting and can yield unstable predictions when small changes in the training data occur. These limitations motivated the development of ensemble tree methods such as Random Forests (Breiman, 2001) and Gradient Boosting (Friedman, 2001), which aggregate multiple trees to reduce variance and

improve generalization. In practice, decision trees are still employed as standalone classifiers in remote sensing studies, particularly when interpretability is prioritized or when serving as a baseline for comparison with more advanced methods. Even where ensemble approaches dominate performance metrics, decision trees remain central to remote sensing workflows because they represent the conceptual foundation of many of today's most powerful classification algorithms (Belgiu & Drăguț, 2016; C. Liu et al., 2020).

2.5. Linear and Polynomial Regression in Remote Sensing

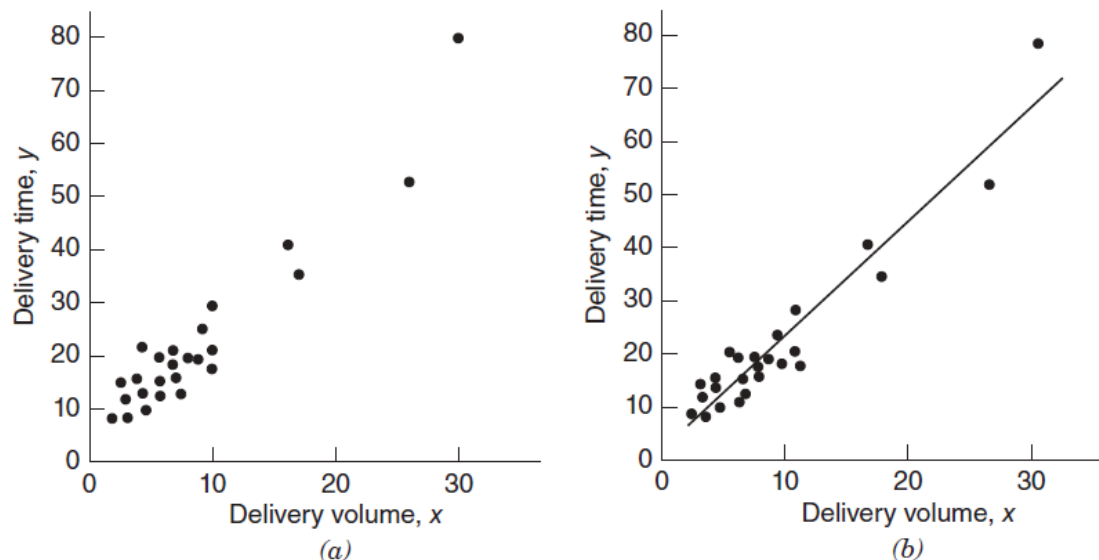


Figure 10: Example of Linear Regression: (a) Scatter diagram for delivery volume. (b) Straight - line relationship between delivery time and delivery volume. (Khuri, 2013)

Linear regression is one of the most fundamental predictive models in remote sensing, applied primarily to estimate continuous biophysical variables from spectral data. By fitting a linear relationship between predictor variables such as reflectance values or vegetation indices and a response variable such as above-ground biomass, soil moisture, or land surface temperature, linear regression provides a transparent and computationally efficient framework (Khuri, 2013). In multispectral and hyperspectral applications, it has been widely used for vegetation studies—for example, relating NDVI to biomass (Tucker, 1979), retrieving foliar chemistry parameters (Curran, 1989), or estimating crop yields from spectral bands (Schlerf et al., 2005). Its simplicity and statistical rigor make it an essential baseline for quantitative remote sensing tasks (Foody, 2002).

Polynomial regression extends this approach by introducing nonlinear terms, enabling the model to capture more complex relationships between spectral features and environmental variables. Quadratic and cubic terms have been used to describe vegetation growth dynamics, nonlinear soil reflectance–moisture relationships, and spectral mixing effects in water quality or canopy structure (Curran, 1989; Verrelst et al., 2015). However, polynomial models are prone to overfitting, particularly when applied to high-dimensional imagery, and must be applied with caution to ensure generalizability to unseen data (Khuri, 2013).

In Earth observation practice, linear and polynomial regression models are often employed as benchmark approaches before advancing to more sophisticated machine learning methods. Their interpretability, low computational cost, and clear statistical assumptions ensure they

remain relevant in operational workflows, especially in environmental monitoring or agricultural policy contexts where transparency is critical (Foody, 2002).

2.6. Logistic Regression in Remote Sensing

Logistic regression is a classical statistical classifier that models the probability of a pixel or object belonging to a given class as a logistic function of the input features. Unlike linear regression, which predicts continuous outcomes, logistic regression outputs probabilities constrained between 0 and 1, making it well suited for binary or multi-class land-cover classification (Hosmer et al., 2013). In remote sensing, logistic regression has been applied to problems such as vegetation–non-vegetation separation, urban sprawl detection, and land-use change monitoring (Pontius & Schneider, 2001; Richards, 2022).

One of its main advantages is interpretability: coefficients provide direct insight into the relationship between spectral variables and class probabilities. This has made logistic regression especially useful in cases where the goal is not only to classify but also to understand the drivers of land-cover transitions. For example, logistic regression has been used to quantify the likelihood of deforestation or urban expansion based on spectral indices and ancillary socioeconomic or topographic variables (Pontius & Schneider, 2001).

Although logistic regression is less powerful than ensemble learning or deep learning models for high-dimensional data, it remains an important tool in remote sensing. It is computationally efficient, robust with limited training data, and provides a statistical foundation for more advanced probabilistic classifiers. In many workflows, logistic regression continues to serve as a baseline method, ensuring that classification outputs can be compared against a transparent and statistically rigorous model (McRoberts & Walters, 2012).

2.7. Naïve Bayes in Remote Sensing

Naïve Bayes is a simple probabilistic classifier based on Bayes' theorem, which assigns class labels by computing posterior probabilities given the observed features. The “naïve” assumption refers to conditional independence among features, meaning that each spectral band or input variable is assumed to contribute independently to the classification decision (Manning et al., 2008). Naïve Bayes often performs surprisingly well in practice, especially in high-dimensional datasets, due to its efficiency and robustness.

In remote sensing, Naïve Bayes has been applied to multispectral and hyperspectral image classification, particularly for land-cover mapping where computational efficiency is important. Because it estimates class-conditional probability distributions directly from training data, Naïve Bayes is effective with relatively small training sets, making it a useful choice in data-scarce environments. Applications include vegetation type classification, crop monitoring, and preliminary urban mapping, where the method provides a lightweight alternative to more complex classifiers (Storvik et al., 2005).

While Naïve Bayes may be outperformed by ensemble methods or deep learning models in complex tasks, it remains relevant as a baseline classifier in comparative studies. Its transparency, low computational cost, and ease of implementation allow researchers to establish benchmarks for classification accuracy or to test the separability of spectral classes before applying more sophisticated techniques. In workflows for Earth observation, Naïve Bayes thus serves as a link between classical statistical methods and modern machine learning, offering an interpretable and efficient option for initial classification experiments.

(Richards, 2022)(Richards & Jia, 2006, Remote Sensing Digital Image Analysis: An Introduction).

2.8. Linear Discriminant Analysis (LDA) in Remote Sensing

Linear Discriminant Analysis (LDA) is a supervised classification technique that seeks linear combinations of features which maximize class separability. Mathematically, LDA finds a projection that maximizes the between-class scatter relative to the within-class scatter, producing linear decision boundaries in the projected space (Fisher, 1936). Under its classical assumptions - approximately Gaussian class distributions with similar covariance structures - this projection yields low-dimensional, discriminative representations that are efficient to estimate (Fukunaga, 1990).

In remote sensing, LDA has been widely used for multispectral and hyperspectral image classification both as a classifier and as a discriminative dimensionality-reduction step. By reducing dimensionality while preserving class separation, LDA helps mitigate the “Hughes phenomenon,” where the number of spectral bands can far exceed the number of labeled samples (Fukunaga, 1990; Richards, 2022). Applications include vegetation discrimination, crop mapping, and urban classification, where LDA provides an interpretable and computationally efficient baseline. Although LDA is limited by its linearity assumption and may underperform when spectral classes are not linearly separable, it remains an important method in remote sensing workflows. LDA often serves as a benchmark classifier against which more flexible methods such as Support Vector Machines (SVMs), Random Forests, or neural networks are compared. Moreover, LDA-derived features are sometimes used as inputs to other machine learning models, combining the statistical rigor of discriminant analysis with the predictive power of modern classifiers (Richards, 2022).

2.9. Pixel-Based Image Analysis (PBIA) with Classical Classifiers

Pixel-Based Image Analysis (PBIA) is the traditional approach to remote sensing classification, where each pixel is treated as an independent observation characterized by its spectral values. In this framework, classifiers assign class labels directly to pixels based on their position in spectral feature space, without considering spatial context or neighborhood relationships. PBIA has been foundational for land-cover and land-use mapping, particularly with multispectral sensors such as Landsat, MODIS, and Sentinel-2, where pixel sizes correspond to broad landscape elements (Richards, 2022).

Classical classifiers commonly used in PBIA include parametric methods such as Maximum Likelihood Classification (MLC), which assumes Gaussian distributions of classes, and non-parametric methods such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Decision Trees, and Random Forests. These approaches have been extensively applied for vegetation mapping, urban monitoring, agricultural crop classification, and change detection (Foody, 2002; Pal & Mather, 2005). The main strength of PBIA is its simplicity and efficiency, especially for medium-resolution imagery where spectral information alone often suffices to distinguish broad land-cover classes.

However, PBIA faces limitations when applied to very high-resolution (VHR) data, where single pixels may capture only fractions of real-world objects such as buildings, roads, or trees. This often results in high intra-class variability and “salt-and-pepper” effects in classification outputs. To mitigate these issues, pixel-based workflows are often complemented with filtering, post-classification smoothing, or integration of textural indices. Despite these

challenges, PBIAs remain a crucial baseline in remote sensing research, and its transparent integration with classical classifiers continues to make it an essential component in comparative studies and operational land-cover mapping pipelines (Gong et al., 2013; L. Liu et al., 2021).

2.10. Object-Based Image Analysis (OBIA) with Classical Classifiers

Object-Based Image Analysis (OBIA) is an approach to remote sensing classification that shifts the analytical focus from individual pixels to meaningful image objects. These objects are created through segmentation algorithms that group neighboring pixels based on spectral, spatial, and contextual similarity. This methodology has become especially important with the advent of very high-resolution (VHR) satellite and aerial imagery, where single pixels often represent only fractions of real-world objects, reducing the effectiveness of traditional pixel-based methods. By working at the object level, OBIA enables classifiers to incorporate not only spectral signatures but also shape, texture, and neighborhood information, providing a richer feature set for land-cover and land-use mapping (Blaschke, 2010).

In many workflows, OBIA is combined with classical classifiers such as decision trees, support vector machines, random forests, or k-nearest neighbors to assign labels to segmented objects. The advantage of this integration is that segmentation reduces intra-class spectral variability while enhancing inter-class differences, thereby improving the separability of classes in the feature space. OBIA pipelines have been successfully applied to extract urban features such as buildings and roads, delineate agricultural fields, and monitor forest degradation by combining object-level segmentation with supervised classifiers (Burnett & Blaschke, 2003; Whiteside et al., 2011).

Despite the growing dominance of deep learning approaches, OBIA continues to play a critical role in Earth observation, particularly in applications where interpretability and expert knowledge are essential. Analysts can fine-tune segmentation parameters and directly relate classification rules to landscape features, providing transparency that complements the predictive accuracy of machine learning. Moreover, OBIA remains an effective way to preprocess and structure high-resolution data for integration into advanced frameworks, including hybrid methods that combine object-based features with convolutional neural networks (Blaschke et al., 2014).

3. Foundations of Neural Networks

Artificial neural networks (ANNs) provide the mathematical and conceptual basis for modern deep learning in remote sensing. Starting with the perceptron (Rosenblatt, 1958) which could only solve linearly separable problems (Minsky & Papert, 1969), the field advanced with multilayer perceptrons and the backpropagation algorithm that enabled learning of non-linear functions (Rumelhart et al., 1986). Core elements include activation functions, which introduce nonlinearity (Nair & Hinton, 2010), and loss functions, which guide optimization, from cross-entropy for classification to Dice loss for segmentation (Milletari et al., 2016; Ronneberger et al., 2015). The universal approximation theorem established the theoretical power of ANNs to model complex relationships (Hornik, 1991), while their architectural organization into input, hidden, and output layers enabled structured learning (Haykin, 1994). The transition to convolutional operations introduced spatial weight sharing, allowing convolutional neural networks (CNNs) to process images efficiently and learn hierarchical features from edges to semantic patterns (Lecun et al., 1998; Maggiori et al., 2017).

3.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational systems inspired by biological neural networks, designed to approximate complex mappings between inputs and outputs. An ANN consists of interconnected layers of artificial neurons, where each neuron computes a weighted sum of its inputs and applies a nonlinear activation function (I. Goodfellow et al., 2016). The architecture typically includes an input layer, one or more hidden layers, and an output layer, with weights optimized during training to minimize a predefined loss function. In remote sensing, ANNs have been applied since the 1990s to problems such as land-cover mapping, vegetation classification, and hyperspectral image analysis. Early studies demonstrated that ANNs outperform traditional classifiers in handling non-linear and noisy spectral data (Atkinson & Tatnall, 1997). Their flexibility is supported by the universal approximation theorem, which guarantees that multilayer feedforward networks with nonlinear activations can approximate any continuous function to arbitrary accuracy (Hornik, 1991).

3.2. Perceptron

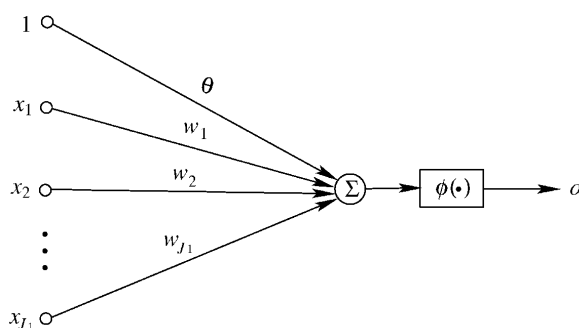


Figure 11: Perceptron - The McCulloch–Pitts neuron model (Du et al., 2022)

The perceptron, proposed by Rosenblatt (1958), is the earliest model of an artificial neuron. It computes a weighted linear combination of inputs and applies a step activation function to decide class membership, functioning as a binary classifier. This model could learn linear decision boundaries and adapt its weights through iterative updates (Rosenblatt, 1958).

However, the perceptron's restriction to linearly separable problems was rigorously analyzed by Minsky and Papert (1969), who demonstrated that it could not solve tasks such as the XOR problem. Their critique significantly slowed neural network research for over a decade (Minsky & Papert, 1969). The field was revitalized in the 1980s with the advent of multilayer perceptrons (MLPs) trained using the backpropagation algorithm, which enabled neural networks to learn non-linear functions and overcome the shortcomings of the single-layer perceptron (Rumelhart et al., 1986). These advances laid the foundation for the development of modern deep learning architectures.

3.3. Activation Functions


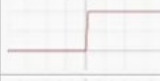







Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
Tanh		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parametric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

Figure 12: Different Activation Functions (Everton Gomedé, 2024)

Activation functions introduce nonlinearity into artificial neural networks, enabling them to approximate complex relationships beyond linear mappings. Early network designs commonly relied on sigmoid and hyperbolic tangent (tanh) activations, which map outputs into bounded ranges and thereby supported probability-like interpretations (Goodfellow, Bengio, & Courville, 2016, *Deep Learning*). However, both functions are prone to the vanishing-gradient problem, limiting their effectiveness as network depth increases.

The introduction of the Rectified Linear Unit (ReLU) significantly improved optimization in deep networks by propagating positive inputs unchanged while mapping negative values to zero, thereby accelerating convergence and mitigating gradient decay (Nair & Hinton, 2010,

Rectified Linear Units Improve Restricted Boltzmann Machines). To address the “dying ReLU” problem, variants such as Leaky ReLU (Maas, Hannun, & Ng, 2013, *Rectifier Nonlinearities Improve Neural Network Acoustic Models*) and Exponential Linear Units (ELU) were proposed, allowing a small gradient flow or smooth negative values, respectively.

In remote sensing applications, the choice of activation function is central to enabling neural networks to extract discriminative spectral and spatial features from high-dimensional data. Nonlinear activations facilitate the representation of both fine-grained spectral variations and spatial textures, which are crucial for classification, detection, and damage assessment tasks in Earth observation (Zhu et al., 2017, *Deep Learning in Remote Sensing: A Comprehensive Review*).

3.4. Loss Functions

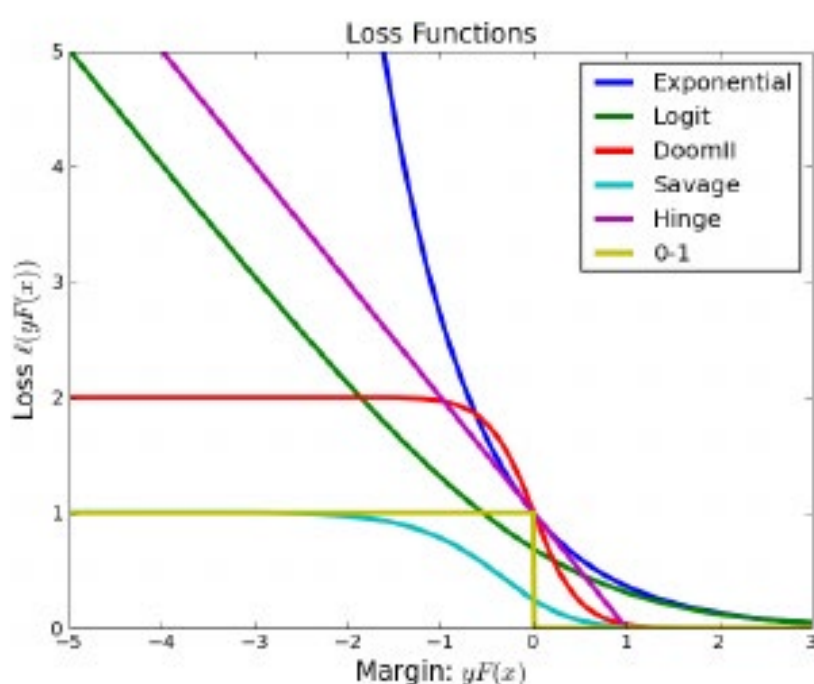


Figure 13: Different loss functions used in boosting and supervised machine learning methods. (Leistner et al., 2009)

Loss functions quantify the discrepancy between predictions and reference labels and thus guide the optimization of neural networks. For regression tasks, Mean Squared Error (MSE) and Mean Absolute Error (MAE) are the most common objectives; for classification, cross-entropy loss is widely used because it penalizes divergence between predicted class probabilities and true classes and is compatible with probabilistic outputs such as softmax (I. Goodfellow et al., 2016).

In remote sensing, loss selection is closely tied to task characteristics and data constraints. For semantic segmentation problems (e.g., building footprint extraction or damage mapping), pixel-wise cross-entropy remains the default, often with class weighting to address foreground–background imbalance (Ronneberger et al., 2015). When labels are severely imbalanced, as in building damage assessment where the no-damage class dominates the label space, surrogates that correlate more directly with overlap metrics - such as Dice-family or IoU-aligned losses - are frequently adopted alongside or in place of cross-entropy to improve minority-class recall while maintaining precision (Ahmadi et al., 2023; Gupta et al., 2019; Su et al., 2020). Because Intersection over Union (IoU) is the standard evaluation metric for segmentation quality (Everingham et al., 2010), optimizing with overlap-aware losses

better aligns training with deployment-time objectives in Earth-observation pipelines. In practice, BDA systems combine class-balanced cross-entropy with an overlap-aware term and report IoU/F1 on benchmarks such as xBD to ensure that optimization and validation are consistent with operational mapping needs (Ahmadi et al., 2023; Gupta et al., 2019; Su et al., 2020).

Task / Scenario	Typical Loss Function(s)	Use Case	Key References
Regression (e.g., vegetation index, LST, building height)	Mean Squared Error (MSE), Mean Absolute Error (MAE)	MSE penalizes large deviations strongly; MAE more robust to outliers	Goodfellow et al., 2016, <i>Deep Learning</i>
Multi-class classification (pixels or objects)	Cross-Entropy	Aligns with probabilistic softmax outputs; widely adopted in CNNs	Goodfellow et al., 2016, <i>Deep Learning</i>
Semantic segmentation (balanced datasets)	Pixel-wise Cross-Entropy	Standard for per-pixel supervision; implemented in U-Net	Ronneberger et al., 2015, <i>U-Net</i>
Semantic segmentation (imbalanced classes)	Weighted Cross-Entropy, Dice Loss, Focal Loss (not in library, optional)	Handles foreground/background imbalance; Dice directly optimizes overlap	Ronneberger et al., 2015, <i>U-Net</i> ; Gupta et al., 2019, <i>xBD</i>
Segmentation aligned with evaluation metrics	IoU Loss, Dice-IoU hybrids	Optimizes overlap measures directly; reduces false negatives in rare classes	Everingham et al., 2010, <i>PASCAL VOC</i> ; Ahmadi et al., 2023, <i>BD-SKUNet</i>
Building Damage Assessment (severe imbalance: no-damage vs. destroyed)	Combined Weighted Cross-Entropy + IoU/Dice	Mitigates imbalance, improves recall on destroyed/major classes	Gupta et al., 2019, <i>xBD</i> ; Ahmadi et al., 2023, <i>BD-SKUNet</i>

Table 1: Loss Functions and Use Cases in BDA

3.5. Backpropagation

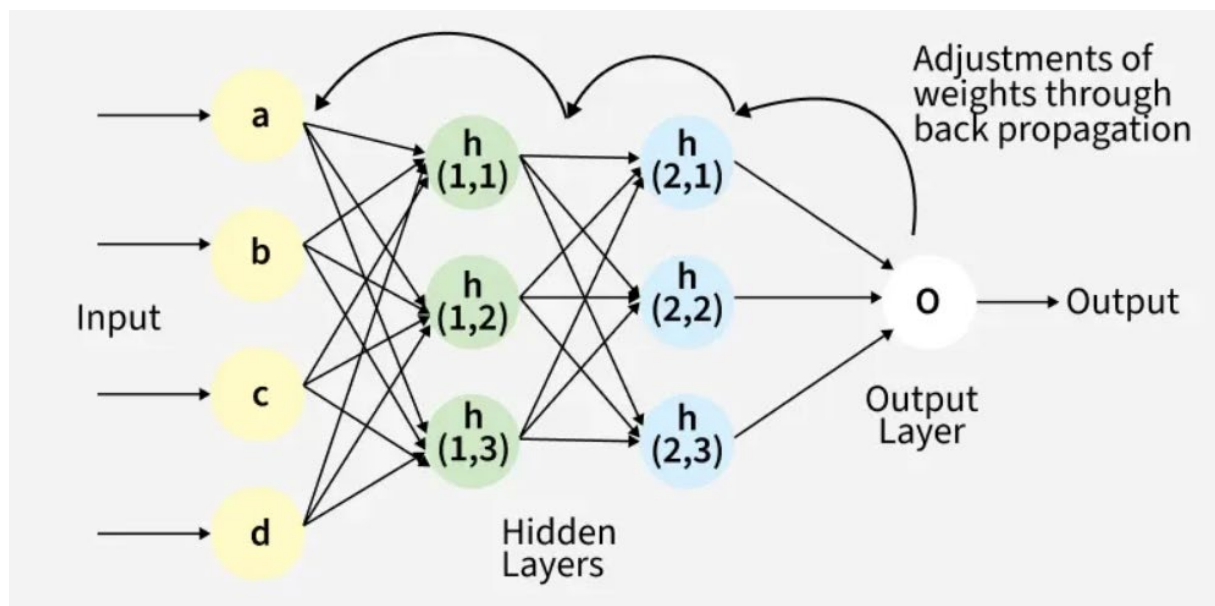


Figure 14: A simple illustration of how the backpropagation works by adjustments of weight. Source: (Backpropagation in Neural Network, 16:07:27+00:00)

Backpropagation is the algorithm that enables neural networks to be trained at scale. It applies the chain rule of calculus to efficiently compute gradients of the loss function with respect to each parameter in the network (Rumelhart et al., 1986). By propagating errors backward from the output layer through hidden layers, backpropagation provides the information required to update weights using gradient descent or its many variants.

Historically, backpropagation was the breakthrough that revitalized ANN research in the 1980s, allowing multilayer perceptrons to solve problems beyond the capability of single-layer perceptrons. In remote sensing, backpropagation-trained networks were among the earliest ANN approaches for multispectral and hyperspectral image classification, and they demonstrated improved accuracy compared with traditional statistical classifiers (PAOLA & SCHOWENGERDT, 1995). Today, backpropagation remains fundamental to all modern deep learning frameworks, underpinning the optimization of convolutional, recurrent, and transformer-based architectures.

3.6. ANN Architectures, Input Layers, Hidden Layers, Output Layers

Artificial Neural Network (ANN) architectures are organized into three principal layer types. The input layer receives raw or preprocessed features such as spectral bands, vegetation indices, or spatial textures. Hidden layers apply weighted transformations followed by nonlinear activation functions to learn intermediate feature representations of increasing abstraction, while the output layer produces the final predictions, typically through a softmax activation for classification or a linear activation for regression tasks (Haykin, 1999).

The capacity of an ANN is largely determined by the depth and width of its hidden layers. Shallow networks may underfit when learning complex feature relationships, while deeper architectures can capture richer, hierarchical representations but risk overfitting when training data are scarce. In remote sensing, architecture design must balance this trade-off between model complexity and generalizability, particularly for high-dimensional data with limited ground truth availability. Specialized designs have emerged in this context, including spectral-only ANNs optimized for hyperspectral image classification and hybrid spectral–

spatial architectures for tasks such as object detection and land-cover mapping (X. X. Zhu et al., 2017).

3.7. From Gradients to Convolution

The optimization of ANNs is grounded in gradient descent, where weights are updated iteratively based on computed gradients of the loss function. While fully connected ANNs can handle tabular or spectral data, they are inefficient for spatially structured inputs like images, as they fail to exploit spatial locality. Convolutional Neural Networks (CNNs) address this by introducing convolutional layers, where small kernels slide across the input image, learning local features such as edges, corners, and textures (Lecun et al., 1998).

Stacking convolutional layers enables CNNs to build hierarchical feature representations, progressing from low-level spatial structures to high-level semantic patterns. In remote sensing, CNNs have revolutionized tasks such as building footprint extraction, land-cover classification, and damage detection, where both spectral and spatial cues are critical (Maggiori et al., 2017). This transition from fully connected networks to convolutional models marked a pivotal shift, bridging the gap between classical ANNs and modern deep learning architectures tailored to Earth observation imagery.

4. Core Architectures

Convolutional neural networks (CNNs) and their architectural variants form the central toolkit for extracting spatial and semantic information from remote sensing imagery. By exploiting local connectivity and hierarchical feature extraction, CNNs are particularly well suited for identifying structural patterns such as buildings, roads, and damage signatures in complex urban environments (Lecun et al., 1998; Maggiori et al., 2017). Encoder–decoder designs like U-Net, SegNet, and V-Net extend this capability by enabling dense pixel-wise predictions, which are essential for segmentation tasks such as mapping building footprints or delineating urban structures (Ronneberger et al., 2015). Residual learning frameworks, exemplified by ResNet, tackle the challenges of training very deep networks, allowing models to capture increasingly abstract representations without degradation (He et al., 2016). More recent backbone architectures such as DenseNet (Huang et al., 2017) and EfficientNet (Tan & Le, 2019) emphasize parameter efficiency and scalable performance, making it feasible to train high-performing models even with limited or heterogeneous datasets.

The evolution of neural network architectures reflects a trajectory from foundational CNNs to specialized encoder–decoders and modern backbones. CNNs introduced hierarchical feature extraction, unlocking scalable image analysis for Earth observation. U-Net and its numerous derivatives became the dominant segmentation framework, refined by attention, Siamese, and selective-kernel enhancements tailored to disaster contexts. Encoder–decoder alternatives such as SegNet, UNet++, and V-Net demonstrate architectural adaptability for efficient and 3D volumetric analysis. ResNet’s residual learning enabled very deep feature extractors, powering hybrid segmentation networks, while DenseNet and EfficientNet emphasized efficiency and scalability. Collectively, these architectures form a versatile toolkit for remote sensing, enabling tasks from land-cover mapping to fine-grained building damage assessment in post-disaster and battlefield scenarios.

4.1. Convolutional Neural Networks (CNNs): Foundations and Applications

Convolutional Neural Networks (CNNs) are the core of modern image understanding because they exploit spatial stationarity via weight-sharing convolutions that detect localized patterns across an image; stacked convolution and pooling layers build hierarchical representations that progress from edges and simple textures to higher-level semantic structures (e.g., rooftops, road networks, rubble), as established in classic work on gradient-based learning for vision (Lecun et al., 1998) and confirmed across remote-sensing surveys emphasizing multi-scale features and orientation variability in Earth observation scenes (Al Shafian & Hu, 2024).

Unlike fully connected neural networks, CNNs exploit the spatial structure of imagery by applying convolutions, where small filters (kernels) slide across the input to detect localized patterns. Each convolutional layer extracts increasingly abstract representations, beginning with edges, corners, or simple textures, and building toward higher-level semantic concepts such as rooftops, road networks, or rubble fields (Lecun et al., 1998). Pooling layers reduce

spatial resolution while retaining salient information, thereby providing a degree of translation invariance and allowing deeper layers to capture larger-scale relationships. By stacking convolution and pooling operations, CNNs construct hierarchical feature representations that are particularly well suited to remote sensing, where features manifest across multiple scales and orientations (Al Shafian & Hu, 2024; I. Goodfellow et al., 2016; Lecun et al., 1998; X. X. Zhu et al., 2017).

CNNs revolutionized Earth observation by outperforming traditional classifiers such as Maximum Likelihood or SVMs, especially on very high-resolution (VHR) imagery where context is critical. Maggiori et al. (2017), demonstrated CNN scalability for semantic segmentation across urban and rural scenes (Maggiori et al., 2017). Ji et al. (2018), showed CNNs' effectiveness for collapsed building detection after the Haiti earthquake, outperforming pixel-based methods (Ji et al., 2018). Mueller et al. (2021), applied CNNs to systematically quantify war-related destruction in Syria, highlighting CNNs as not only general-purpose image classifiers but also essential tools for battlefield monitoring and humanitarian response (Mueller et al., 2021).

Formally, a 2-D convolution computes a feature map:

$$F(x, y) = \sum_{i=1}^m \sum_{j=1}^n K(i, j) I(x + i, y + j)$$

with a kernel $K \in \mathbb{R}^{m \times n}$ sliding over an image I ; weight sharing drastically reduces parameters and underpins approximate translation equivariance, while nonlinear activation functions (e.g., ReLU) and pooling/downsampling enable deeper layers to aggregate broader spatial context—properties widely used to move from raw spectral values to semantically meaningful objects in remote sensing. The same set of weights is shared across spatial locations, reducing the number of parameters compared to fully connected layers and ensuring translation invariance (Al Shafian & Hu, 2024; Lecun et al., 1998). Early layers capture low-level features such as edges and lines, while deeper layers capture higher-level semantic patterns, including complex objects like collapsed structures (Lecun et al., 1998). In remote sensing, this hierarchical representation is essential to transform raw spectral values into semantically meaningful objects such as building footprints or road networks (Ji et al., 2018; Maggiori et al., 2017).

CNNs also enable multi-modal fusion by combining optical, SAR, and ancillary geospatial data, which is operationally crucial in crisis situations where optical imagery may be obscured by clouds, smoke, or debris. Adriano et al. (2021), demonstrated that CNNs can jointly learn from SAR and optical imagery to improve building footprint extraction (Adriano et al., 2021). More recent work has extended this principle to disaster contexts: cross-modal scenarios that pair pre-disaster optical imagery with post-disaster SAR achieve performance comparable to single-mode optical pipelines, underlining the operational value of learned fusion for building-damage mapping (Adriano et al., 2021). Practical implementations often integrate GIS data such as OpenStreetMap or settlement masks to restrict analysis to built-up areas, thereby improving signal-to-noise in conflict environments (Aimaiti et al., 2022).

At the same time, CNNs remain sensitive to annotation quality: noisy or inconsistent ground-truth labels degrade performance in damage detection, particularly when distinguishing between adjacent severity classes (Alisjahbana et al., 2024). Imbalanced disaster datasets, where intact buildings dominate, further exacerbate overfitting risks, motivating strategies such

as augmentation, transfer learning, and regularization. Reviews of the last decade highlight additional challenges in achieving real-time processing, model generalization across geographies and hazards, and robust data fusion (Al Shafian & Hu, 2024). Finally the inherently local receptive fields of CNNs limit their ability to capture long-range dependencies, prompting increasing adoption of attention mechanisms and transformer-based backbones in remote sensing. For example, multi-modal vision transformers have been applied successfully in high-noise, dynamic contexts such as spacecraft and debris recognition, illustrating their potential to extend beyond CNN constraints (AlDahoul et al., 2022).

Despite their success, CNNs face challenges. Overfitting is a major risk when training on small or imbalanced disaster datasets, where the majority of buildings may be intact (Alisjahbana et al., 2024). Data augmentation, transfer learning, or dropout regularization are commonly applied to mitigate this (X. X. Zhu et al., 2017). Computational cost is another limitation: training large CNNs on very-high-resolution (VHR) imagery is resource-intensive, particularly for multi-temporal damage assessment pipelines (Adriano et al., 2021). Finally, CNNs have limited ability to capture long-range dependencies such as spatial context across entire urban blocks, which has motivated the development of attention-based networks and transformer architectures for overhead imagery (Al Shafian & Hu, 2024; AlDahoul et al., 2022).

4.2. U-Net and Variants: Semantic Segmentation for Remote Sensing

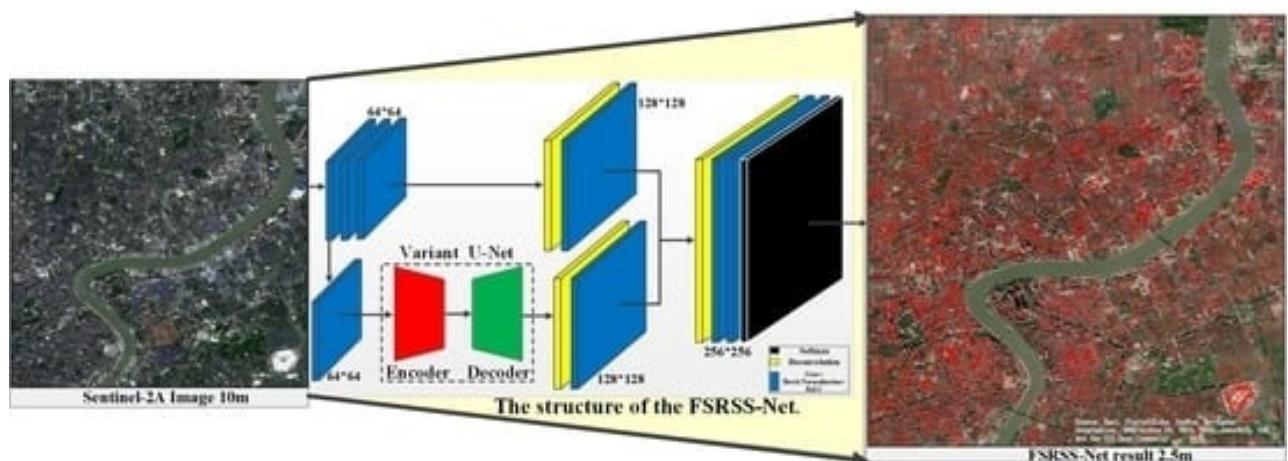


Figure 15: Mapping Buildings with U-Net. (Zhang et al., 2021)

The U-Net architecture, introduced by Ronneberger et al. (2015) has become one of the most widely adopted encoder–decoder models for semantic segmentation in computer vision. Its design was originally tailored for biomedical image segmentation, where precise delineation of structures such as cells or organs was needed from relatively few annotated samples (Ronneberger et al., 2015). The design of U-Net follows an encoder–decoder structure, where the contracting path (encoder) captures progressively abstract semantic features through repeated convolution, activation, and pooling operations, while the expanding path (decoder) reconstructs spatial resolution through upsampling and convolution layers. The hallmark innovation of U-Net lies in the use of skip connections, which concatenate encoder feature

maps with corresponding decoder layers to preserve fine spatial details that would otherwise be lost during pooling. A central contribution of U-Net was its ability to achieve pixel-level classification with limited training data by combining encoder–decoder convolutional pathways with skip connections that preserve spatial resolution (Çiçek et al., 2016; Ronneberger et al., 2015). The efficient use of annotated data, robustness in settings with small sample sizes, and high precision in boundary localization made U-Net particularly attractive for remote sensing applications, where annotated training data are expensive to acquire and high spatial precision is critical (T. Chen et al., 2022; Ronneberger et al., 2015; Xie et al., 2022). Early surveys of deep learning in Earth observation identified U-Net as particularly suited for pixel-wise land-cover classification, building footprint extraction, and damage mapping, given its ability to integrate multi-scale spatial features (Maggiori et al., 2017; X. X. Zhu et al., 2017). The adaptation of the U-Net architecture for building damage assessment has been driven by the need to segment damaged structures from VHR imagery with fine boundary detail (Gupta et al., 2019).

Mathematically, the core building block of U-Net can be expressed as a series of convolutions followed by nonlinear activation. At layer l , the feature extraction step is defined as:

$$h^{(l)} = \sigma(W^{(l)} * h^{(l-1)} + b^{(l)})$$

where $h^{(l)}$ denotes the feature map at layer l , $W^{(l)}$ the convolutional kernel, $b^{(l)}$ the bias, and $*$ the convolution operation. In the decoder stage, upsampling (via transposed convolution) produces a higher-resolution feature map, which is then concatenated with the encoder feature of the same spatial resolution:

$$h_s^{(\text{dec})} = \text{concat}\left(h_s^{(\text{enc})}, \text{Up}\left(h_{s+1}^{(\text{dec})}\right)\right)$$

This formulation emphasizes the encoder–decoder symmetry and the role of skip connections in preserving spatial detail during reconstruction (Çiçek et al., 2016; Ronneberger et al., 2015). This operation combines semantic information from deeper layers with fine-grained spatial details, enabling accurate segmentation of object boundaries such as building edges or rubble contours. This property stems directly from the skip connections introduced in the original U-Net (Ronneberger et al., 2015). The effectiveness of the skip-connected encoder–decoder has been generalized by nnU-Net, which automatically configures pre-/post-processing, architecture, and training, achieving SOTA on diverse segmentation tasks and consistently achieves state-of-the-art performance in segmentation benchmarks (Isensee et al., 2018).

Several variants of U-Net have been developed to address specific limitations. Attention U-Net introduces spatial and channel attention gates that emphasize relevant image regions, such as damaged areas, while suppressing background noise — improving segmentation quality in cluttered urban environments (Oktay et al., 2018; C. Wu et al., 2021). Siamese U-Nets process pre- and post-disaster imagery in parallel encoders, learning difference features in the decoder, which makes them particularly effective for change detection tasks (Daudt et al., 2018; Gholami et al., 2022). Another extension, UNet++, redesigns skip connections as nested, dense pathways, which better align semantic and spatial features and yield improved delineation of irregular structures such as partially collapsed buildings (Z. Zhou et al., 2018). For 3D data, 3D U-Net extends convolutions into the volumetric domain, enabling segmentation of UAV-derived point clouds or LiDAR volumes, which are increasingly important in post-disaster urban modeling (Çiçek et al., 2016).

(Gupta et al., 2019) The success of U-Net in remote sensing arises from its ability to integrate contextual and spatial information at multiple scales. Unlike traditional pixel-based classifiers, which treat each pixel independently, U-Net performs dense prediction across entire images, allowing contextual cues such as neighboring structures and textures to influence classification (X. X. Zhu et al., 2017). Empirical evidence from large-scale studies has confirmed this advantage: Maggiori et al. (2017) showed that encoder–decoder CNNs outperform shallow classifiers for land-cover mapping (Maggiori et al., 2017). In the specific domain of building damage assessment (BDA), U-Nets have become a standard baseline. Gupta et al. (2019) demonstrated that U-Net-based architectures achieved strong performance across earthquakes, hurricanes, and floods, confirming their generalizability across diverse disaster contexts (Gupta et al., 2019). Similarly, Jones & Saniie (2019), validated U-Net’s effectiveness in disaster response by generating building damage maps from pre- and post-event imagery (Jones & Saniie, 2019). Pan et al. (2020), further demonstrated the ability of U-Nets to integrate high-resolution WorldView imagery with auxiliary GIS data for accurate urban segmentation (Pan et al., 2020).

Numerous adaptations of U-Net address damage-mapping limitations. Attention U-Nets introduce channel and/or spatial attention (e.g., CBAM), steering the model toward damaged regions and suppressing background clutter. Remote-sensing instantiations show consistent gains: attention-enhanced U-Nets improve robustness to shadows, occlusions and debris in cluttered urban scenes (Shen et al., 2022; Woo et al., 2018; C. Wu et al., 2021). Siamese U-Nets process pre- and post-disaster imagery in parallel, enabling explicit modeling of change features (Gholami et al., 2022). Domain-specific innovations such as BDANet incorporate selective kernel units and ordinal loss functions to rank building damage severity (Shen et al., 2022). Related architectures like BD-SKUNet enhance feature fusion across scales to capture both fine-grained cracks and large rubble regions (Ahmadi et al., 2023). Collectively, U-Net and its derivatives dominate post-disaster building segmentation, bridging high-resolution remote sensing with actionable intelligence.

4.3. SegNet, UNet++, and V-Net: Encoder–Decoder Architectures for Image Segmentation

Beyond the classical U-Net, several encoder–decoder architectures have been developed to extend semantic segmentation capabilities across domains such as biomedical imaging, computer vision, and remote sensing. Among the most prominent are SegNet, UNet++, and V-Net, which represent three different directions in addressing the limitations of U-Net: computational efficiency, semantic feature alignment, and volumetric data representation.

SegNet, introduced by Badrinarayanan et al. (2017), maintains the overall encoder–decoder symmetry but introduces a key innovation in its decoder design. Instead of transferring full feature maps from the encoder to the decoder, SegNet stores the pooling indices from the max-pooling operations of the encoder. During decoding, these indices are reused for upsampling, after which trainable convolutional filters densify the sparse feature maps. This design reduces memory consumption and parameter count while retaining spatial detail, making SegNet significantly more efficient than architectures with heavy skip connections (Badrinarayanan et al., 2017). SegNet was originally motivated by road scene understanding, where fast inference and limited hardware resources are crucial, but has since been applied

to large-scale Earth observation, such as continental urban mapping and cropland monitoring (Badrinarayanan et al., 2017; Xie et al., 2022).

UNet++ addresses a fundamental challenge in the original U-Net design: the semantic gap between encoder and decoder features (Z. Zhou et al., 2018). Shallow encoder features contain fine spatial information but limited semantics, while deeper decoder features represent high-level semantics but lack spatial precision. UNet++ reduces this gap through a system of nested and dense skip connections. Instead of directly concatenating encoder outputs with decoder inputs, intermediate convolutional blocks progressively refine the transferred features before fusion. The network can further employ deep supervision, where segmentation outputs are generated at multiple decoder stages, improving gradient flow and training stability. UNet++ has proven particularly effective in scenarios requiring precise delineation of irregular boundaries, such as collapsed or partially destroyed roofs. While Zhou et al. (2020) demonstrated its architectural improvements in medical segmentation, subsequent evaluations on disaster damage datasets have shown that UNet++ outperforms classical U-Net. On the xBD benchmark, later studies reported that UNet++ with deep supervision achieved better discrimination between moderate and severe damage classes (Dietrich et al., 2025; Gupta et al., 2019; Kaur et al., 2023; Khvedchenya & Gabruseva, 2021). V-Net, introduced by Milletari et al. (2016), extends the encoder–decoder paradigm into the three-dimensional domain. It processes volumetric inputs using 3D convolutional layers, enabling voxel-wise segmentation. V-Net’s principal innovations include the integration of residual learning, which improves convergence and network stability, and the adoption of the Dice loss function as an optimization objective, which effectively mitigates class imbalance—especially important when small structures are embedded within large background volumes first demonstrated in medical imaging contexts (Milletari et al., 2016).

Although originally targeted at medical imaging, V-Net’s volumetric design has inspired geospatial applications. Liu et al. (2020), utilized UAV-derived photogrammetric point clouds to assess building damage post-disaster, demonstrating the critical need for voxel-wise volumetric analysis to estimate rubble volumes and detect collapsed structures—an outcome conceptually aligned with V-Net’s architecture, even though the study did not apply V-Net directly (C. Liu et al., 2020).

4.4. ResNet: Deep Residual Learning for Feature Extraction

ResNet introduced residual connections, where identity shortcuts allow gradients to propagate across layers without vanishing. This innovation enabled the training of networks with hundreds of layers, significantly increasing representational capacity. The basic residual block reformulates the learning objective as a residual mapping,

$$y = F(x) + x,$$

where $F(x)$ represents the learned residual function and x is the shortcut identity connection. This simple addition allows the network to learn residual functions (differences) rather than complete mappings, stabilizing training of very deep networks. Without residual connections, adding more layers often degrades performance due to vanishing gradients. With ResNet, gradients can flow directly through identity connections, enabling networks with 50, 101, or even more layers. In practical terms, this means that ResNet can capture both low-level patterns and high-level semantic abstractions in the same architecture. The design stabilizes convergence and prevents the degradation of accuracy that typically occurs in very deep

networks (He et al., 2016). In remote sensing, this enables robust extraction of both texture-level details (e.g., rubble) and scene-level patterns (e.g., neighborhood-wide destruction). For this reason, ResNet backbones are widely used in encoder–decoder pipelines for disaster damage detection. Qing et al. (2022), integrated ResNet-50 into a change-detection pipeline for earthquake damage mapping (Qing et al., 2022). Pretraining ResNet on ImageNet accelerates convergence and provides robust feature initialization, which is crucial when annotated disaster datasets are small (Shao et al., 2020). Bhardwaj et al. (2025), combined ResNet encoders with U-Net decoders, demonstrating superior accuracy in delineating damage severity classes (Bhardwaj et al., 2024).

ResNet’s success lies in its generalization capacity: its architecture is not domain-specific but has been seamlessly adapted across a variety of remote sensing modalities. Originally introduced for large-scale natural image classification (He et al., 2016). ResNet backbones have since been employed in optical satellite image classification, segmentation, and building damage assessment tasks (Bhardwaj et al., 2024; Gupta et al., 2019; Gupta & Shah, 2021). In the radar domain, ResNet architectures have been incorporated into SAR change detection frameworks and target recognition pipelines, demonstrating their adaptability beyond purely optical inputs (Parikh et al., 2020; Passah et al., 2022). For multispectral optical imagery, ResNet encoders are used for segmentation and classification tasks—from ship detection and maritime mapping (Ciocarlan & Stoian, 2021) to scene-level classification on multispectral datasets (Zhao et al., 2022) and land-cover mapping with ResNet-50 backbones (Gharahbagh et al., 2025). For battlefield damage assessment applications, residual learning enables models to capture subtle textural cues of rubble or collapsed roofs against cluttered urban backdrops. This robustness has cemented ResNet as a default encoder in hybrid pipelines for disaster and conflict monitoring (Bhardwaj et al., 2024; Gupta et al., 2019; Gupta & Shah, 2021).

4.5. DenseNet and EfficientNet: Modern Backbone Networks

DenseNet introduced dense connectivity, where each layer reuses the outputs of all preceding layers, encouraging feature reuse and improving gradient propagation. This design yields more compact networks with fewer parameters, reducing overfitting and improving computational efficiency (Huang et al., 2017). In remote sensing, DenseNet backbones have proven useful for hyperspectral and high-dimensional imagery, where redundancy and instability often challenge conventional CNNs. For example, Gao et al. (2021) and Cai et al. (2021), demonstrated that DenseNet encoders deliver robust spectral–spatial fusion and improved segmentation reliability in noisy environments (Cai et al., 2021; Gao et al., 2021).

The Dense Convolutional Network (DenseNet) proposed by Huang et al. (2017), rethinks feature propagation by introducing dense connectivity (Huang et al., 2017). Unlike conventional feed-forward CNNs, where each layer receives input only from its immediate predecessor, DenseNet establishes direct connections from each layer to all subsequent layers. Concretely, the feature maps produced by the l -th layer are concatenated with all previous outputs $[x_0, x_1, x_2 \dots, x_{l-1}]$ and passed forward, ensuring that later layers have access to both low-level and high-level representations. This architecture addresses vanishing gradients, promotes extensive feature reuse, and reduces redundancy by limiting the need for repeated relearning of similar patterns. Importantly, this also yields more parameter-efficient models compared to equally deep residual networks, since DenseNet requires fewer filters per layer while still achieving competitive representational capacity. In remote sensing, such connectivity has proven particularly useful for hyperspectral and high-dimensional imagery, where DenseNet backbones enhance spectral–spatial fusion and stabilize training under noisy conditions (Cai et al., 2021; Gao et al., 2021; Huang et al., 2017).

EfficientNet, introduced by Tan & Le (2019), addresses the problem of scaling CNN architectures in a principled manner (Tan & Le, 2019). EfficientNet proposed a compound scaling strategy, balancing network depth, width, and input resolution to achieve an optimal accuracy–efficiency trade-off and the EfficientNet-B0 to B7 family ranges from lightweight to high-capacity models, designed to generalize across tasks under varying computational budgets. Prior approaches typically scaled models by increasing either depth, width or input resolution in isolation, often leading to suboptimal trade-offs between accuracy and efficiency. EfficientNet proposes compound scaling, where a single scaling coefficient uniformly balances all three dimensions based on a simple, but effective heuristic derived from empirical optimization. The base model, EfficientNet-B0, is constructed via neural architecture search, after which larger variants (B1–B7) are generated by applying the compound scaling rule. This approach achieves state-of-the-art accuracy on ImageNet while using an order of magnitude fewer parameters and FLOPs than traditional backbones (Tan & Le, 2019). For remote sensing, EfficientNet’s design is operationally attractive: its lightweight variants can be deployed for rapid inference on limited hardware, while larger versions support high-capacity tasks like disaster damage assessment. Chen et al. (2022), highlight how encoder choices affect cross-region transferability in change detection and demonstrated that EfficientNet backbones provide robust cross-region transferability in change detection pipelines. Adriano et al. (2021), highlighted how encoder efficiency directly impacts the scalability of large-area multimodal fusion for disaster monitoring (Adriano et al., 2021).

5. Advanced and Emerging Approaches

Building on the foundations of CNNs and encoder–decoder architectures, recent advances in machine learning have introduced methods that expand both the capability and flexibility of remote sensing analysis. Attention mechanisms and Siamese networks allow models to explicitly compare multi-temporal imagery and focus on the most relevant features, which is particularly valuable for change detection after disasters (Khvedchenya & Gabruseva, 2021). Generative adversarial networks (GANs) contribute by synthesizing realistic samples or enhancing scarce training data, thereby improving robustness in scenarios with limited ground truth (I. Goodfellow et al., 2014). Wu et al (2021) extend the analysis beyond grid-based imagery by modeling spatial relations between objects, such as building-to-building interactions within urban environments (Z. Wu et al., 2021). At the same time, contrastive and self-supervised learning techniques reduce the reliance on large annotated datasets by leveraging inherent structures in unlabeled data (Cui et al., 2023). Ensemble methods, including Random Forests, gradient boosting, and stacking, continue to play a complementary role by combining predictions from multiple models for improved stability (Ahmad et al., 2024). Finally, domain-specific innovations such as BDANet (Shen et al., 2022), BD-SKUNet (Ahmadi et al., 2023) (Ahmadi et al., 2023), BDD-Net (Shao et al., 2020), and U-Net derivatives for urban mapping (Pan et al., 2020) integrate these principles into specialized architectures tailored for building damage assessment.

5.1. Transfer Learning

Transfer learning has become a cornerstone of deep learning in remote sensing, particularly for disaster damage assessment, because it addresses the challenge of limited annotated data. Training deep convolutional neural networks from scratch is impractical in most Earth observation scenarios due to the scarcity of large, labeled datasets and the high computational cost. Instead, transfer learning leverages weights pre-trained on large-scale natural image collections such as ImageNet, which contains over 14 million labeled images across thousands of categories (Deng et al., 2009). These pre-trained weights encode general visual representations - edges, corners, textures, and shapes - that transfer effectively to satellite and aerial imagery domains. Numerous surveys confirm that fine-tuning pre-trained networks on remote sensing datasets yields competitive performance with far fewer labeled samples compared to training from scratch (X. X. Zhu et al., 2017).

Methodologically, transfer learning operates in two main modes. In feature transfer, convolutional layers trained on generic data are reused as frozen feature extractors, and only the final task-specific layers (e.g. fully connected classifiers) are retrained. In fine-tuning, some or all of the pre-trained network is unfrozen, and gradients adapt its parameters to the new dataset. This initialization provides a strong inductive bias: mathematically, task-specific weights W_{target} are initialized close to the pre-trained weights W_{source} learned on dataset D_S , reducing the risk of overfitting when training on smaller remote sensing datasets D_T . This approach is especially valuable for disaster imagery, where annotated data are scarce and often noisy (X. X. Zhu et al., 2017).

In battlefield and post-disaster contexts, transfer learning enables rapid deployment of robust models without requiring extensive ground truth. Gupta et al. (2019), demonstrated that ImageNet-pretrained backbones such as ResNet significantly improve building damage classification and segmentation across multiple disaster types (Gupta et al., 2019). Adriano et al. (2021), similarly employed pre-trained CNN encoders to improve generalization in multimodal fusion of optical and SAR data, showing that initialization with external weights stabilized convergence on their global disaster dataset (Adriano et al., 2021). Recent work further confirmed that transfer learning supports scalability across diverse hazards: for example, Alisjahbana et al. (2024), designed a modular two-step pipeline in which pre-trained encoders provided stable performance across earthquakes, floods, and typhoons with minimal retraining (Alisjahbana et al., 2024). This strategy allows operational systems to generalize across geographies, sensor modalities, and disaster scenarios, reducing retraining time and computational burden. Transfer learning thereby accelerates the transition from research prototypes to battlefield and humanitarian applications, where timely, accurate assessment is critical for decision-making and response (Adriano et al., 2021; Gupta et al., 2019).

5.2. Siamese Networks and Attention Mechanisms for Change Detection

Siamese networks are a specialized neural architecture designed to perform pairwise comparisons between two inputs by learning feature embeddings in a shared latent space (Bromley et al., 1993). In the context of remote sensing change detection, a Siamese CNN processes pre- and post-event images of the same region through identical subnetworks with shared weights (Daudt et al., 2018). This design ensures that both images are transformed into comparable embeddings, enabling the network to emphasize differences caused by structural or environmental changes while suppressing irrelevant variations (T. Chen et al., 2022). The approach is particularly well suited to damage assessment, where the objective is to distinguish collapsed or destroyed buildings from intact ones within dense urban settings (Aimaiti et al., 2022; Gupta et al., 2019).

Technically, a Siamese network operates by mapping each input image to a feature space and then quantifying their similarity. This paradigm was first introduced by Bromley et al. (1993), who showed that two subnetworks with shared weights can produce comparable embeddings for pairwise similarity tasks (Bromley et al., 1993). In the context of remote sensing change detection, let $f_{\theta}(x)$ denote the feature embedding of an input image x , produced by a CNN encoder with parameters θ . Given a pre-disaster image x_{pre} and a post-disaster image x_{post} , the embeddings are:

$$h_{pre} = f_{\theta}(x_{pre}) \text{ and } h_{post} = f_{\theta}(x_{post})$$

A distance metric such as the L_1 norm or cosine similarity computes a change score:

$$d = ||h_{pre} - h_{post}||$$

This metric-based formulation has been widely adopted in remote sensing, where fully convolutional Siamese networks produce dense pixel-wise change maps (Bromley et al., 1993; Daudt et al., 2018).

Attention mechanisms can be integrated into this pipeline by applying learnable masks A that reweight embeddings spatially or spectrally, yielding: $h' = A \odot h$. This refinement directs the model to emphasize regions most relevant for damage (e.g., collapsed rooftops, rubble fields) while down-weighting irrelevant differences such as illumination or vegetation changes (T. Chen et al., 2022; Woo et al., 2018). Chen et al. (2022) explicitly combined Siamese encoders with attention modules, showing that attention improves the detection of weak changes and reduces noise in high-resolution disaster imagery (T. Chen et al., 2022).

The foundational contribution of this architecture to remote sensing came with the introduction of fully convolutional Siamese networks for change detection, which established the feasibility of end-to-end learning for pixel-wise change maps (Daudt et al., 2018). Since then, the integration of attention modules has further improved robustness by guiding the model toward informative spatial regions or spectral bands. Transformer-based encoders have been incorporated into Siamese architectures to capture long-range dependencies across an entire urban block, extending the receptive field beyond localized kernels (T. Chen et al., 2022). The Vision Transformer laid the foundation for this paradigm, and subsequent remote sensing work has adapted it to multi-temporal damage mapping tasks (Dosovitskiy et al., 2021).

For battlefield and post-disaster applications, these developments are highly significant. Siamese networks provide an interpretable and scalable means of monitoring destruction by directly contrasting pre- and post-event imagery. This is critical in operational environments where timely intelligence is required to assess infrastructure losses. For instance, Adriano et al. (2021) demonstrated that pre-trained CNN encoders in a Siamese framework could generalize across disaster types and sensor modalities (Adriano et al., 2021), while Aimaiti et al. (2022), showed that combining optical and SAR imagery within a change detection pipeline enables robust mapping under conflict conditions, where clouds, smoke, or intentional camouflage may obscure optical observations (Aimaiti et al., 2022).

5.3. Generative Adversarial Networks (GANs) for Data Augmentation and Damage Inference

Generative Adversarial Networks (GANs) are a major advance in generative modeling. Introduced by Goodfellow et al. (2014), GANs comprise a generator, which synthesizes samples, and a discriminator, which distinguishes real from synthetic, trained jointly in a minimax game; under ideal conditions, the generator recovers the data distribution, and the discriminator is uniform at equilibrium (I. J. Goodfellow et al., 2014). This adversarial setup enables the synthesis of highly realistic images and is particularly valuable when annotated data are scarce or imbalanced, because generative augmentation can expand minority classes without changing task semantics (Hao et al., 2023). In remote sensing, GANs and related deep generative methods are used to generate realistic synthetic imagery for training augmentation; studies show that such augmentation improves downstream detection/recognition and directly addresses data imbalance (Hao et al., 2023; Yates et al., 2022). For damage assessment specifically, widely used datasets such as xBD exhibit pronounced class imbalance across damage levels, which degrades performance on rare, severe classes; in such settings, oversampling severe-damage categories with GAN-generated samples is a principled remedy (Gupta et al., 2019; Su et al., 2020).

In remote sensing GANs have been recognized as a powerful tool to mitigate the lack of labeled training data, which remains a critical bottleneck for the application of deep learning in damage assessment (X. X. Zhu et al., 2017).

The training of GANs is formulated as a minimax optimization problem between generator G and discriminator D :

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]$$

Here, x denotes real data samples drawn from the distribution p_{data} , and z represents latent noise vectors sampled from a prior distribution (e.g., Gaussian). The generator $G(z)$ transforms noise into synthetic samples, while the discriminator $D(x)$ outputs the probability that a given sample is real. Through iterative training, the generator improves until its outputs are nearly indistinguishable from real data (I. J. Goodfellow et al., 2014; Mirza & Osindero, 2014).

Extensions of this framework include Conditional GANs (cGANs), where conditioning variables y (such as building footprints or pre-disaster imagery) guide the generator to produce semantically meaningful outputs $G(z, y)$ (Mirza & Osindero, 2014). Other architectures relevant to Earth observation include Variational Autoencoder-GAN hybrids for spectral feature synthesis (X. Wang et al., 2020).

In Earth observation, GANs are used to mitigate two persistent hurdles - limited labeled data and strong class imbalance - by synthetically expanding minority classes and improving downstream classifiers' robustness (Antoniou et al., 2018; Douzas & Bacao, 2018). Post-disaster datasets such as xBD are indeed valuable but skewed: the dataset's damage taxonomy (no/minor/major/destroyed) is dominated by "no-damage," with minority classes much smaller, which complicates learning (Gupta et al., 2019) and pixel-level proportions in comparative analyses show extreme sparsity of damaged categories (e.g., ~0.1% each for minor/major/destroyed in xBD), underscoring the imbalance problem. In practice, generative augmentation is an established strategy to oversample underrepresented categories, alleviating bias toward frequent classes (Antoniou et al., 2018; Douzas & Bacao, 2018). Beyond simple resampling, image-to-image translation with adversarial learning (e.g., CycleGAN) can synthesize target-domain-like imagery to support model training under distribution shifts that are common in damage mapping (different disasters, sensors, and contexts) (J.-Y. Zhu et al., 2017).

In conflict- and disaster-mapping, damage labels are scarce and strongly imbalanced: in xBD the distribution of damage categories is skewed toward "no damage" (313,033 polygons) versus "minor" (36,860), "major" (29,904) and "destroyed" (31,560). At the pixel level, "minor/major/destroyed" each account for only ~0.1% while background and no-damage dominate. This imbalance is also reflected in downstream evaluations: F1 scores typically rank no-damage > destroyed > major > minor, with mid-level classes underperforming due to intra-class ambiguity and sample imbalance; collapsing "minor/major" into a single "damage" class improves performance (three-class setting) (Benson & Ecker, 2020; Gupta et al., 2019; Su et al., 2020).

A further remedy is multi-modal fusion. SAR and optical imagery are complementary: SAR provides all-weather/day-night sensing and captures structural backscatter governed by wavelength-dependent penetration, whereas optical collects texture and color. Empirically, combining SAR with very-high-resolution optical data improves building-damage detection compared to single-mode inputs (e.g., Izmit/Bam, L'Aquila, Wenchuan case studies) (Balz & Liao, 2010). Consistent with this complementarity, cross-modal and fusion scenarios in a

global multimodal dataset (optical + SAR, pre/post) yield competitive results for damage mapping (Adriano et al., 2021).

Adversarial image-to-image translation further helps when distributions differ across disasters/sensors: CycleGAN-based style transfer has been used to transform post-disaster satellite images to the target domain to promote generalization on xBD, with experiments reporting gains relative to no-transfer baselines (while noting occasional negative transfer).

This is particularly valuable in conflict zones where cloud cover, night-time acquisitions, or denied-access conditions often limit optical availability (Flores et al., 2019).

5.4. Graph Neural Networks (GNNs) for Spatial Relations in Urban Damage Assessment

While CNNs capture local grid-based patterns, they struggle with relational reasoning. Graph-based learning addresses relational reasoning between two objects by representing scenes as nodes (e.g., buildings) and edges encoding spatial or functional relationships (adjacency, shared utilities, road connectivity) (Selvakumaran et al., 2025). Graph Neural Networks (GNNs) extend deep learning to non-Euclidean domains by representing data as graphs rather than regular grids. Unlike convolutional neural networks, which operate on fixed pixel neighborhoods, GNNs exploit node–edge structures to capture relational and topological information. (Adriano et al., 2021; Selvakumaran et al., 2025). For example, Zheng et al. (2021) applied graph-based semantic change detection to urban damage mapping, while Chen & Parent (2020) demonstrated graph CNNs for modeling road networks. Such studies illustrate the practical potential of GNNs for capturing building-to-building and infrastructure relations in remote sensing workflows .

In a GNN, node features $h_i^{(l)}$ are updated by aggregating neighbor information:

$$h_i^{(l+1)} = \sigma \left(W \cdot \text{AGG} \left(h_i^{(l)} : j \in N(i) \cup h_i^{(l)} \right) \right)$$

where AGG is mean/sum/attention, W are trainable weights, and σ a nonlinearity (Kipf & Welling, 2017).

For building damage assessment, initial node features are typically derived from convolutional neural network (CNN) embeddings of per-building image chips from pre- and post-disaster pairs, while edges encode spatial adjacency or infrastructure links. Message passing in the graph propagates contextual cues across the urban topology, enabling the inference of plausible damage patterns even when direct evidence is incomplete (Selvakumaran et al., 2025; Zheng et al., 2021). Graph models are particularly powerful for urban and battlefield contexts, where cascading effects of destruction and network dependencies matter. If a cluster of buildings shows heavy damage, adjacent structures are statistically more likely to be compromised even when pixel-level evidence is weak. This improves robustness against occlusion, cloud cover, or sensor noise (Selvakumaran et al., 2025). Object-centric pipelines that operate at the building level, rather than on pixel grids, also improve semantic consistency across scenes and disaster types (Zheng et al., 2021).

In urban environments, entities such as buildings, roads, and infrastructure interact and influence each other’s vulnerability during disasters. In battlefield or post-disaster contexts, the damage state of one structure often correlates with its neighbors through blast waves,

fire spread, or seismic cascades, making graph-based reasoning especially valuable (Selvakumaran et al., 2025). Recent developments in Earth observation highlight the integration of GNNs with multimodal data sources - including optical, SAR, multispectral, and hyperspectral imagery - allowing contextual information to propagate across object-level graphs for improved classification under sparse or noisy observations (Adriano et al., 2021). Extending to multispectral, hyperspectral and SAR, graph and object-centric approaches complement CNNs in two important ways for conflict and post-disaster mapping. First, multisensor integration benefits from explicit graph structure: multimodal and multitemporal datasets demonstrate that cross-modal and fusion scenarios (optical + SAR, pre/post) are competitive for damage mapping, and graph reasoning encodes cross-sensor consistency at the object level (Adriano et al., 2021). Second, when spectral richness is available, per-building feature vectors derived from multispectral or hyperspectral data capture material and debris signatures that can be propagated through the graph. Recent work shows that deep fusion of hyperspectral and SAR features improves classification by suppressing heterogeneity and speckle while retaining complementary information (Bioucas-Dias et al., 2013). For battlefield damage assessment of urban structures with denied access or persistent cloud cover, SAR is central, and object-level analysis using SAR - augmented by very-high-resolution optical data when available - enhances situational awareness for urban damage assessment (Aimaiti et al., 2022; Balz & Liao, 2010). Because disaster imagery can vary widely across hazards, geographies, and sensors, out-of-domain generalization remains a critical limitation. Evaluations on cross-event splits show that in-distribution accuracy significantly overestimates real-world robustness, underscoring the value of structural priors such as graph constraints to support transfer (Benson & Ecker, 2020).

5.5. Contrastive and Self-Supervised Learning in Remote Sensing

Supervised learning in remote sensing is constrained by the limited availability of annotated datasets, especially for rare events such as armed conflict or earthquake-induced building damage. Labeling satellite imagery at building or pixel level is costly, time-consuming, and in conflict zones often impossible due to access restrictions. Self-supervised learning (SSL) has emerged as a promising paradigm to mitigate these limitations by exploiting large volumes of unlabeled data to pretrain feature extractors on auxiliary tasks, before fine-tuning them on small, labeled datasets. Early SSL approaches in computer vision formulated surrogate objectives such as predicting the relative position of image patches (Doersch et al., 2015) or reconstructing shuffled image tiles (Noroozi & Favaro, 2017).

A major breakthrough was contrastive learning, where models are trained to maximize the similarity between embeddings of augmented views of the same image, while minimizing similarity to other samples. The SimCLR framework introduced the InfoNCE loss as a tractable objective and demonstrated that SSL can rival fully supervised pretraining on ImageNet (T. Chen et al., 2020). Building on this, subsequent methods such as MoCo (He et al., 2020) and BYOL (Grill et al., 2020) removed the need for large batch sizes or even negative samples, further widening the applicability of SSL to remote sensing archives.

In remote sensing, SSL has been adapted to geospatial imagery with encouraging results. Ayush et al. (2021) introduced GeoSSL, demonstrating that geolocation-aware contrastive pretraining improves performance across multiple satellite benchmarks (Ayush et al., 2021). For disaster contexts, Cui et al. (2023), showed that pretraining with contrastive objectives on large pools of unlabeled remote sensing data significantly boosts Swin Transformer accuracy for damage classification on the xBD dataset (Cui et al., 2023). For battlefield damage

assessment and post-disaster applications, SSL provides three major advantages. First, it enables robust feature learning from vast repositories of unlabeled multispectral, SAR, or hyperspectral imagery, where annotation is infeasible. Second, SSL pretraining has been shown to improve out-of-domain generalization, a key challenge in building damage detection where models trained on one event often fail on others (Benson & Ecker, 2020). Third, contrastive frameworks can integrate multimodal data. For instance, recent multimodal fusion datasets demonstrate that combining optical and SAR imagery with SSL objectives improves damage classification consistency across hazards and regions (Adriano et al., 2021). Expanding beyond optical data, SSL has also been applied in hyperspectral and multispectral domains, where spectral redundancy can be exploited through masked autoencoding and contrastive spectral reconstruction (Bioucas-Dias et al., 2013). In SAR, SSL approaches are particularly valuable since speckle noise and coherence artifacts reduce supervised transferability; pretraining on large unlabeled Sentinel-1 archives enhances downstream tasks such as coherence-based change detection (Plank, 2014).

5.6. Ensemble and Hybrid Methods: Random Forests, Gradient Boosting, Stacking, Bagging and Model Fusion

Ensemble learning refers to a family of machine learning strategies that combine multiple models to improve predictive performance, reduce variance, and enhance generalization (Breiman, 1996, 2001; Freund & Schapire, 1997; Friedman, 2001; Z.-H. Zhou, 2012). In remote sensing, ensemble and hybrid models are particularly valuable because they can handle high-dimensional spectral inputs, a feature well-demonstrated by Random Forests (Belgiu & Drăguț, 2016) and integrate heterogeneous sensor data, as shown in SAR-optical fusion studies (Ahmad et al., 2024). Ensemble Methods have the ability to cope with extreme class imbalances, such as in post-earthquake building-damage mapping, has been demonstrated using random forests in InSAR-based damage classification (Rao et al., 2023) and through bagging-based ensembles in aerial-image change detection with scarce damaged samples (Seo et al., 2019). By integrating complementary models, ensembles leverage diversity in representation and decision boundaries, often achieving robustness that single models cannot attain (Adriano et al., 2021; Aimaiti et al., 2022; Al Shafian & Hu, 2024; Benedetti et al., 2018; Benson & Ecker, 2020; C. Liu et al., 2020; Selvakumaran et al., 2025).

5.6.1. Bagging and Random Forests

Bootstrap aggregating (bagging) was introduced by Breiman (1996) as a method to reduce variance by training multiple classifiers on bootstrapped subsets of the data and averaging their predictions (Breiman, 1996). Random Forests (RF) extend this by incorporating random feature selection at each tree split, decorrelating trees and improving performance on high-dimensional data (Breiman, 2001). In remote sensing, Random Forests are widely used for classification tasks because they can handle non-linear feature interactions, high-dimensional spectral bands, and small training sets without overfitting (Belgiu & Drăguț, 2016; Gislason et al., 2006). Belgiu and Drăguț (2016) provided a comprehensive review of RF applications in remote sensing, highlighting its adoption for land cover mapping, urban feature extraction, and disaster assessment (Belgiu & Drăguț, 2016).

Hybrid methods combine CNN-derived features with Random Forest classifiers when labeled data are scarce. Liu et al. (2020) demonstrated RF-based classification of UAV photogrammetric point clouds for earthquake damage, integrating 3D geometric features and

supervoxel segmentation (C. Liu et al., 2020). In the context of battlefield damage assessment, RF classifiers have been employed on SAR-optical fused features to detect urban impervious surfaces or rubble proxies with robust performance under noisy conditions (Ahmad et al., 2024).

5.6.2. Boosting and Gradient Boosting Methods

Boosting refers to sequentially training weak learners, typically shallow trees, to correct errors from previous iterations. AdaBoost introduced reweighting of misclassified samples (Freund & Schapire, 1997), while Gradient Boosting Machines (GBM) reframed boosting as stagewise additive modeling with gradient descent in function space (Friedman, 2001).

Modern gradient boosting frameworks such as XGBoost (T. Chen & Guestrin, 2016), LightGBM (Ke et al., 2017), and CatBoost (Prokhorenkova et al., 2019) have been optimized for speed, regularization, and handling of categorical features. In Earth observation (EO), gradient boosting has proven effective in small-sample contexts or where tabular features (e.g., building footprint area, structural attributes) are combined with image-based embeddings (Adriano et al., 2021; Ahmad et al., 2024). Ahmad et al. (2024) demonstrated that ensembles of RF, AdaBoost, and XGBoost achieved high accuracy for impervious surface extraction using Sentinel-1 and Sentinel-2 fusion (Ahmad et al., 2024).

5.6.3. Stacking and Meta-Learning

Stacking, or stacked generalization, integrates heterogeneous base learners through a higher-level meta-learner (Ting & Witten, 1999; Wolpert, 1992). The method requires careful cross-validation to prevent information leakage but can exploit complementary error structures across models. In damage assessment, stacking has been used to fuse CNN-based image features with RF or GBM classifiers operating on contextual attributes (e.g., adjacency, structural type). Multimodal and multitemporal stacking was applied to optical and SAR features for global damage mapping, showing performance gains over single-sensor pipelines (Adriano et al., 2021).

5.6.4. Model Fusion in Remote Sensing

Fusion approaches can occur at multiple levels:

- **Early fusion** combines features (e.g., SAR backscatter + multispectral indices) (Ahmad et al., 2024).
- **Intermediate fusion** integrates CNN embeddings from different modalities (Benedetti et al., 2018).
- **Late fusion** aggregates decisions from separate classifiers (Adriano et al., 2021).

Benedetti et al. (2018) proposed M³Fusion, a deep architecture for multiscale multimodal fusion, demonstrating improved land-cover mapping by combining Sentinel-2 time series with VHR imagery (Benedetti et al., 2018). In damage mapping, similar principles apply: Aimaiti et al. (2022) fused Sentinel-1 radar with Sentinel-2 optical imagery to detect war-related building damage in Kyiv, reporting improved precision under cloud cover (Aimaiti et al., 2022).

5.6.5. Hybrid Pipelines for Damage and Conflict Mapping

Hybrid models often combine deep feature extraction with ensemble classifiers. CNN encoders can extract semantic features from pre- and post-disaster imagery, which are then classified by RF or GBM to exploit their ability to handle tabular and sparse data (Bhardwaj et al., 2024; C. Liu et al., 2020). Bhardwaj et al. (2024) integrated ResNet and U-Net for damage segmentation, a design that could be extended with ensemble classifiers for robust classification (Bhardwaj et al., 2024). UAV-based 3D features have also been fused with RF for detailed building-level assessments, bridging 2D optical and 3D structural perspectives (C. Liu et al., 2020).

Graph-based pipelines add another layer of hybridization: Selvakumaran et al. (2025) integrated graph convolutional networks with multimodal features, but decision-level fusion with ensemble classifiers remains a practical way to incorporate contextual graph-derived attributes (Selvakumaran et al., 2025).

In battlefield and disaster damage assessment contexts, ensembles mitigate dataset imbalance, domain shift, and noise. RF and GBM allow weighting of minority classes, while stacking permits heterogeneous input features (e.g., SAR backscatter statistics + CNN embeddings). However, computational complexity increases, and careful cross-validation is required for generalization. In-distribution accuracy can overestimate robustness; ensemble fusion improved out-of-domain generalization, particularly across disasters unseen during training (Benson & Ecker, 2020).

5.7. BDANet and other Specialized Architectures for Building Damage Assessment

Specialized architectures for building damage assessment refine generic encoder–decoder designs to exploit the particular signal characteristics of pre/post-event satellite imagery and the structure of the output label space. In post-disaster and battlefield damage assessment contexts, these models must cope with multi-scale urban morphology, subtle appearance changes between damage grades, severe class imbalance, and heterogeneous sensors such as very-high-resolution optical (VHR), Sentinel-2 multispectral, Sentinel-1 SAR, and UAV-derived 3D products. Recent work converges on a handful of design patterns: attention-augmented CNNs tuned to building morphology and debris cues (Shen et al., 2022), Siamese change-pair networks that learn invariances across pre/post conditions (T. Chen et al., 2022) multibranch or two-stage pipelines that separate footprint localization from per-building damage inference (Alisjahbana et al., 2024), U-Net variants with selective or spatial attention for sharper boundaries and mid-level semantics (Ahmadi et al., 2023), and transformer-based backbones that improve long-range context aggregation (Cui et al., 2023; Dietrich et al., 2025).

BDANet represents a class of attention-augmented encoder–decoders designed specifically for building damage mapping. In contrast to plain U-Nets, BDANet variants insert cross-directional or spatial-channel attention modules into the skip connections and decoder in order to emphasize salient building and debris cues while suppressing background clutter that confounds damage grades. The rationale is that damage evidence is often local and anisotropic (e.g., façade spalling along edges, roof-hole patterns, rubble fields adjacent to load-bearing walls), so attention gates are tasked with reweighting features conditioned on spatial context and inter-channel dependencies before fusion in the decoder (Shen et al., 2022). Empirical studies on xBD-like setups report that attention reduces confusion between mid-level classes by sharpening boundary evidence and enhancing mid-level texture responses; the practical burden shifts from deeper backbones to better skip-path feature

selection. In operational terms, BDANet is trained end-to-end on paired pre/post tiles; an image-level or pixel-level cross-entropy is combined with boundary or attention regularizers, and class-balanced sampling is used to counteract the dominance of “no-damage” in the label distribution (Shen et al., 2022).

Joint segmentation-and-damage pipelines recognize that errors in footprint extraction propagate into damage classification and therefore optimize both sub-tasks in a single graph. RescueNet is emblematic: a shared encoder branches into a footprint head and a damage head, allowing backbone features to be supervised by both geometry and semantics. This joint supervision tends to (i) stabilize building localization under clutter and occlusion and (ii) constrain damage predictions to building interiors without requiring costly post-processing. Architecturally, RescueNet retains a U-Net-like topology but augments it with multi-task losses (footprint IoU loss + per-pixel damage cross-entropy) and feature sharing across heads; training uses paired pre/post inputs or post-only inputs depending on availability, with the pre/post case typically outperforming due to explicit change cues (RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery). In disaster deployments, the ability to export both masks in one pass reduces latency for emergency mapping workflows where footprint and grade must be delivered together (Gupta et al., 2019; Gupta & Shah, 2021; Weber & Kané, 2020).

Subclass-aware designs address a persistent failure mode: conflation of visually similar damage levels. BSSNet reframes building damage as a hierarchical problem by decoupling localization from subclass inference and injecting additional supervision to align boundaries and interior evidence. The network divides the task into a localization subnetwork that predicts building extent and a classification subnetwork that assigns fine-grained building subclasses; a spatial-gradient fusion module regularizes the localization stream by supervising the predicted gradient map of the mask, which in practice produces crisper boundaries around roof-holes and collapsed regions. Crucially, BSSNet integrates contrastive learning within the classification stream to increase feature consistency for the same subclass under geographic and sensor shifts. This combination improves subclass separability on both a bespoke Hainan dataset and xBD, particularly for mid-level categories where texture differences are minor relative to background variation (Xie et al., 2022).

Siamese U-Net and its variants exploit the paired nature of pre- and post-event imagery to learn change-selective representations. Two weight-sharing encoders process pre and post tiles, and their mid-level feature maps are fused—by concatenation, differencing, or attention—before the decoder predicts either a binary change mask or a multiclass damage map. The Siamese formulation suppresses static background appearance and reduces reliance on absolute texture cues that are brittle under seasonal or sensor shifts. In remote sensing change detection, Siamese-U-Net-style models repeatedly outperform single-image baselines on high-resolution data, and they provide a strong foundation for damage mapping when combined with class-balanced losses and change-aware augmentations. In battlefield damage assessment contexts, where cloud cover and imaging angles vary, the shared encoder and explicit pre/post fusion often deliver superior cross-event generalization compared with post-only segmenters (T. Chen et al., 2022).

Two-stage and multibranch pipelines pursue robustness by decomposing the task. DeepDamageNet exemplifies a pragmatic two-step strategy for the xView2 problem: first detect buildings with a semantic segmentation CNN; then classify per-building damage with a dedicated CNN head using pre/post crops. The separation permits targeted hard-negative mining and class-aware sampling at the object-level and simplifies domain adaptation by allowing the first stage to be fine-tuned to a new area while keeping the second stage largely

intact. When class imbalance is severe, class-balanced focal losses or cost-sensitive logits at the second stage help reduce minority-class collapse. Extensive participation in the xView2 challenge (xView2, n.d.) shows that such modular designs can outperform monolithic end-to-end models in low-label settings by leveraging object-level priors and tailored augmentations at each step (Alisjahbana et al., 2024). A related pattern in operational deployments uses a deep encoder–decoder for footprints and damage, exports the penultimate feature maps per building, and trains a lightweight tree ensemble (RF/GBM) meta-classifier on those features together with contextual tabular attributes (height proxies, adjacency indicators); this hybridization has been reported to increase cross-event robustness when labeled target data are scarce (Adriano et al., 2021).

Attention-enhanced U-Nets remain the default template for specialized variants. BD-SKUNet inserts selective-kernel modules into the encoder–decoder to adaptively adjust receptive fields and channel emphasis across scales. This is particularly relevant for VHR imagery where small roof perforations and large-area collapses must be detected simultaneously. By dynamically weighting multiple kernel sizes, the network suppresses aliasing in upsampling paths and improves delineation of damage patterns at disparate scales. Evaluations on xBD and an external earthquake dataset show that selective-kernel attention narrows the performance gap between easy extremes (no-damage vs. destroyed) and mid-level classes by focusing features on geometrically consistent patterns rather than spurious textures (Ahmadi et al., 2023). Closely related are models that pair spatial/channel attention with boundary-aware losses or refine skip-connection fusion via attention gates; these produce crisper masks in dense urban cores, where boundary confusion is common, and reduce mislocalization of damaged pixels onto shadows or adjacent roofs (Bhardwaj et al., 2024).

Transformer-based backbones, notably Swin Transformer, bring global modeling and hierarchical windowed self-attention to damage mapping. Their ability to capture long-range context helps resolve ambiguous mid-level classes where local texture is insufficient. In practice, Swin backbones are either used in a U-shaped segmentation head (Swin-U-Net) or within two-branch change-pair pipelines; self-supervised pretraining on large unlabeled satellite corpora further improves feature invariance across regions and sensors, yielding measurable gains when fine-tuned on limited labeled damage data. In conflict scenarios, where scene layout and building typologies vary across theaters, the inductive bias of self-attention to relate distant, semantically similar regions can reduce false positives from repetitive urban textures and improve alignment of pre/post features under geometric noise (Cui et al., 2023).

Across these families, workflows converge on a few practical themes. Pre/post alignment, even if approximate, consistently benefits Siamese and attention-based models; coarse co-registration suffices when the fusion module is designed to tolerate residual parallax (T. Chen et al., 2022). Class imbalance requires explicit handling; focal losses, reweighting, and curriculum sampling are common in BDANet-style and BD-SKUNet pipelines (Ahmadi et al., 2023; Shen et al., 2022). Multimodal fusion remains beneficial: optical+SAR combinations support robustness under cloud or night acquisitions, either via early feature fusion in intermediate layers or via late decision-level stacking with a meta-learner (Adriano et al., 2021; Aimaiti et al., 2022). Operational advantage of modular architectures: two-stage designs and hybrid CNN-plus-ensemble classifiers absorb domain shifts better than monolithic models and can be retuned quickly to new geographical regions with minimal labels (Alisjahbana et al., 2024; Benson & Ecker, 2020).

6. Validation and Testing

The credibility of machine learning methods in remote sensing depends not only on the sophistication of architectures but also on rigorous validation against reference data. In building damage assessment, where results inform time-critical decisions, evaluation metrics must capture both numerical accuracy and operational reliability. Classical measures such as RMSE, precision, recall, F1, IoU, overall accuracy, and Kappa remain central, while recent work highlights the value of benchmark datasets (xBD, xView2, UNOSAT) and tailored loss functions for imbalanced, ordinal, or perceptual tasks (Congalton, 1991; Foody & Mathur, 2004; Gupta et al., 2019).

6.1. Root Mean Square Error (RMSE) for Quantitative Accuracy

The Root Mean Square Error (RMSE) is one of the most widely used statistical measures for evaluating model performance in both machine learning and remote sensing. It quantifies the average magnitude of errors between predictions and reference observations, providing an interpretable estimate of how far, on average, predictions deviate from the ground truth. RMSE is especially important in regression tasks (e.g., predicting vegetation indices, land surface temperature, or building height) but has also been adapted to measure accuracy in image reconstruction, segmentation, and classification contexts. Its broad applicability has made it a standard metric in Earth observation studies (Richards, 2022; Willmott & Matsuura, 2005). RMSE evaluates the mean squared deviation between predictions and reference values, penalizing large errors more strongly than small ones. This sensitivity to outliers is advantageous in remote sensing tasks where large deviations are unacceptable. In practice, RMSE has long been applied in digital elevation model validation, where it quantifies the difference between satellite-derived and surveyed heights (Toutin, 2006), and in hyperspectral analysis, where it measures reconstruction accuracy following dimensionality reduction (Bioucas-Dias et al., 2013).

For building damage assessment (BDA), RMSE is particularly valuable when models produce continuous or ordinal outputs rather than binary labels. For example, when estimating a damage severity score ranging from minor cracks to complete collapse, RMSE captures how closely predicted values align with field-based or annotated references (Ci et al., 2019). In SAR-based applications, regression models frequently estimate intensity differences or coherence loss between pre- and post-event images, and RMSE provides a transparent measure of how well these predictions fit observed ground-truth metrics (Plank, 2014).

6.2. Precision, Recall, and F1 Score in Classification Tasks

For classification tasks, particularly under class imbalance, precision, recall, and F1 are indispensable. Precision quantifies reliability of positive predictions, recall measures sensitivity to actual damaged structures, and the F1 score balances the two. These measures are grounded in the confusion matrix (Sokolova & Lapalme, 2009). In building damage assessment, precision is critical to avoid false alarms in destroyed categories, while recall ensures damaged buildings are not missed, both of which affect disaster response priorities (Gupta et al., 2019). F1 scores therefore serve as the most balanced indicator, with recent models such as BD-SKUNet explicitly reporting improved trade-offs across classes (Ahmadi et al., 2023).

6.3. Intersection over Union (IoU) and Segmentation Metrics

IoU, also known as the Jaccard Index, has become the standard for segmentation in remote sensing, measuring spatial overlap between predictions and ground truth. Unlike overall pixel accuracy, IoU penalizes both under- and over-segmentation, making it more robust for imbalanced classes (Everingham et al., 2010). In damage mapping, IoU ensures predicted masks for collapsed structures align closely with annotated polygons in datasets such as xBD, where failure to capture precise boundaries directly affects operational mapping (Gupta et al., 2019). Extensions such as Generalized IoU provide stricter evaluations for small and irregular structures, which are common in urban battlefield environments (Rezatofghi et al., 2019).

6.4. Overall Accuracy and Kappa Coefficient

Overall accuracy (OA) reports the proportion of correct classifications across all classes, while Cohen's Kappa adjusts for chance agreement (Cohen, 1960). Both have long been used in remote sensing (Congalton, 1991). In highly imbalanced datasets like xBD, OA can be misleading since predicting all buildings as intact yields high scores. Kappa partially mitigates this by penalizing chance agreement, but it too has limitations under skewed distributions (Pontius & Millones, 2011). Nonetheless, both remain baseline metrics in disaster mapping studies and are frequently reported alongside IoU and F1 to ensure comparability (Ahmadi et al., 2023).

6.5. Cross-Validation and Train/Test Splits

Validation design is as important as the metrics themselves. The simplest approach is the hold-out or train/test split, where data are partitioned into disjoint subsets for training and evaluation. This strategy, widely formalized in early ML validation studies (Kohavi, 1995), provides an efficient estimate of model generalization while ensuring that test data remain unseen during training. In remote sensing, standardized splits such as those in the xBD benchmark enable comparability across studies (Gupta et al., 2019). Single splits may introduce bias through spatial autocorrelation or uneven representation of disasters. To address this, cross-validation rotates training and test folds across the dataset, yielding more stable error estimates (Hastie et al., 2009).

6.6. Confusion Matrices

		reference data classes			
		A	B	C	sum
thematic map classes	A	35	2	2	39
	B	10	37	3	50
	C	5	1	41	47
	sum	50	40	46	136

$$\text{overall accuracy} = (35+37+41)/136 \approx 83.1\%$$

producer's accuracies

$$A \quad 35/50 \approx 70.0\%$$

$$B \quad 37/40 \approx 92.5\%$$

$$C \quad 41/46 \approx 89.1\%$$

user's accuracies

$$A \quad 35/39 \approx 89.7\%$$

$$B \quad 37/50 \approx 74.0\%$$

$$C \quad 41/47 \approx 87.2\%$$

Figure 16: Using an error matrix to summarise classifier performance and map accuracy (Richards, 2022)

The confusion matrix remains the foundation of classification evaluation, offering detailed class-level diagnostics. From it, precision, recall, F1, IoU, and OA are derived (Richards, 2022). In building damage assessment, confusion matrices reveal systematic weaknesses: intact classes are often over-predicted while severe categories suffer from under-detection. Such patterns were observed in xBD baselines (Gupta et al., 2019) and mitigated in later architectures such as BDANet and BD-SKUNet, which reported reduced confusion between “major” and “destroyed” (Ahmadi et al., 2023). This transparency is vital in operational contexts, as misclassifying destroyed as minor damage has far more serious consequences than the reverse.

For a binary classification problem (e.g., damaged vs. undamaged), the confusion matrix is a 2×2 table:

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 2: Confusion Matrix, derived from (Richards, 2022)

For multi-class problems (e.g., no damage, minor, major, destroyed), the confusion matrix expands to a $k \times k$ table where k is the number of classes. The diagonal entries n_i represent correct classifications for class i , while off-diagonal entries capture misclassifications (Richards, 2022). From this table, all major metrics - including precision, recall, F1, overall accuracy, and Kappa - can be derived, making it the most versatile evaluation tool.

In remote sensing applications, confusion matrices have long been used to validate land-use and land-cover (LULC) classifications, vegetation mapping, and urban extraction. They remain

a requirement in most classification accuracy assessment. For example, in multispectral classification of agricultural crops, confusion matrices highlight confusion between spectrally similar classes (e.g., wheat vs. barley). In hyperspectral mapping, they help diagnose whether dimensionality reduction (e.g., PCA, MNF) improves separability. In damage assessment, confusion matrices are essential for revealing systematic biases, such as models consistently underestimating “major damage” while overestimating “minor damage” (Foody & Mathur, 2004).

6.7. Ground Truth and Benchmark Dataset Validation

The credibility of any machine learning model in remote sensing and battlefield damage assessment (BDA) of urban structures depends not only on its architecture and metrics, but also on the quality and representativeness of the ground truth data used for validation. Ground truth refers to reference data against which model predictions are compared, typically derived from field surveys, manual annotation of satellite or UAV imagery, or authoritative datasets produced by agencies such as UNOSAT (UNOSAT, n.d.) or the Copernicus Emergency Management Service (EMS) (Chawanji et al., 2022). Without accurate ground truth, even sophisticated models risk being validated against flawed labels, undermining their utility in operational settings (Foody & Atkinson, 2002).

In remote sensing, ground truth collection has traditionally relied on field campaigns, where GPS points, in-situ surveys, or air photographs are used to annotate training and validation data. In urban contexts, cadastral maps, building footprints, and land-use records provide additional reference layers (Congalton, 1991). With the expansion of high-resolution satellite imagery, manual annotation by experts or trained volunteers has become the dominant approach, enabling the creation of large benchmark datasets. For example, the SpaceNet challenge provided building footprints derived from commercial very-high-resolution imagery, establishing a standard for urban object extraction (Etten et al., 2018).

For building damage assessment, several benchmark datasets now underpin the field:

- **xBD dataset:** Gupta et al. (2019) introduced *xBD: A Dataset for Assessing Building Damage from Satellite Imagery*, which contains pre- and post-disaster imagery across 19 global events with building-level damage labels (no damage, minor, major, destroyed). It is the most widely used benchmark for training and validating CNNs, U-Nets, and BDANet architectures (Gupta et al., 2019; Su et al., 2020).
- **xView2 challenge dataset:** IBM developed the *xView2 Building Damage Assessment Challenge*, extending xBD by providing standardized splits and evaluation protocols, explicitly focusing on post-disaster operational performance (xView2, n.d.).
- **UNOSAT datasets:** UNOSAT regularly provides authoritative post-disaster damage assessments for earthquakes, floods, and conflicts, often used as ground truth in humanitarian validation pipelines (UNOSAT, n.d.; Voigt et al., 2007, 2011, 2016).
- **Copernicus EMS datasets:** The Copernicus Emergency Management Service produces rapid mapping products for European disasters, serving as additional validation references in flood, fire, and earthquake mapping (Chawanji et al., 2022).

Ground truth validation in BDA operates at multiple levels of granularity. At the pixel level, predicted damage masks are compared directly with annotated damage maps, though this approach is sensitive to boundary inconsistencies. At the object level, building footprints serve

as units of analysis, aligning predictions with annotated damage grades — the strategy adopted in xBD and xView2. At the scene or regional level, aggregated statistics (e.g., percentage of destroyed buildings) are compared to field reports, supporting situational awareness but sacrificing per-object fidelity (Voigt et al., 2011).

Ground truth data, however, introduces uncertainties. Manual annotations can be subjective, particularly in distinguishing adjacent categories such as “major” versus “destroyed.” Inter-annotator agreement varies, introducing label noise into benchmark datasets (Gupta et al., 2019; Su et al., 2020). Temporal mismatches between imagery acquisition and field surveys further complicate validation, especially in contexts where reconstruction begins quickly. These issues highlight the need for probabilistic evaluation or uncertainty quantification approaches (Foody & Atkinson, 2002).

Despite these limitations, benchmark datasets remain essential for standardizing evaluation and comparing methods. They enable reproducibility, foster competition through challenges such as SpaceNet (Etten et al., 2018) and xView2 (xView2, n.d.) and accelerate methodological progress by providing consistent baselines. Yet, for operational adoption, validation must move beyond benchmarks, incorporating independent data from new disasters, cross-sensor imagery that integrates optical and SAR (Plank, 2014), and multi-temporal sequences. Ultimately, reliable models must demonstrate not only strong performance on curated datasets but also robustness under diverse and unforeseen battlefield and disaster conditions.

6.8. AUC, SSIM, and Task-Specific Metrics

While classical metrics such as overall accuracy, F1 score, and IoU remain the backbone of model evaluation in remote sensing and building damage assessment (BDA), advanced and task-specific metrics provide additional perspectives that are critical in specialized contexts. Among the most widely adopted are the Area Under the Curve (AUC) from Receiver Operating Characteristic (ROC) analysis, the Structural Similarity Index (SSIM) from image quality assessment, and custom task-specific metrics tailored to disaster mapping. These measures capture nuances that traditional metrics often miss, particularly in imbalanced datasets, image reconstruction tasks, and operational decision-making scenarios.

Advanced metrics complement classical ones by addressing operational and perceptual gaps. AUC evaluates discrimination across thresholds, avoiding biases from class imbalance that inflate accuracy (Bradley, 1997, *The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms*). In imbalanced disaster datasets such as xBD, AUC provides a measure of ranking reliability rather than threshold-dependent accuracy. Structural similarity (SSIM) focuses on perceptual quality and has been applied to assess SAR-to-optical translation and GAN-based synthetic imagery, ensuring generated maps retain critical building features (Wang et al., 2004, *Image Quality Assessment: From Error Visibility to Structural Similarity*). Finally, task-specific approaches such as ordinal regression–based loss functions penalize errors more severely across ordered classes, better reflecting operational needs where misclassifying “destroyed” as “no damage” is far more critical than confusing “minor” with “major” (Feng et al., 2020, *Building Damage Assessment with Ordinal Regression*).

7. Glossary and Definitions

Activation Function

A nonlinear function applied to a neuron's output to allow networks to approximate complex relationships. Examples include sigmoid, tanh, and ReLU (Nair & Hinton, 2010).

Artificial Neural Network (ANN)

A layered computational model of interconnected neurons designed to learn mappings between inputs and outputs. Forms the basis of deep learning (I. Goodfellow et al., 2016).

Area Under the Curve (AUC)

A metric summarizing the performance of a classifier across thresholds by plotting the true positive rate against the false positive rate (Bradley, 1997).

Attention Mechanism

A neural module that reweights features spatially or spectrally, focusing computation on the most relevant information, such as damaged regions in imagery (Vaswani et al., 2023).

Autoencoder

A neural network trained to compress data into a low-dimensional latent representation and reconstruct it, uncovering hidden structures in multispectral and hyperspectral datasets (Hinton & Salakhutdinov, 2006).

Backpropagation

An algorithm using the chain rule to compute gradients of the loss function with respect to weights, enabling efficient optimization of multilayer networks (Rumelhart et al., 1986).

Battlefield Damage Assessment (BDA)

The systematic evaluation of damage inflicted by armed conflict – in the context of this publication especially to urban structures using remote sensing and AI. BDA integrates multi-sensor imagery and machine learning to quantify destruction, classify severity, and support military or humanitarian response.

Building Segmentation

A semantic segmentation task in remote sensing where algorithms delineate individual building footprints from imagery. It is the prerequisite for per-building damage classification in disaster and conflict monitoring (Ahmad et al., 2024; Pan et al., 2020).

Confusion Matrix

A matrix summarizing classification performance at the class level, used to derive precision, recall, F1, and accuracy metrics (Richards, 2022).

Convolution

A mathematical operation applying small kernels to an input grid (e.g., images) to extract local features such as edges or textures. The basis of CNNs (Lecun et al., 1998).

Convolutional Neural Network (CNN)

A neural network using convolutional layers for hierarchical feature extraction. Core to land-cover classification, object detection, and damage mapping (Maggiori et al., 2017).

Cross-Validation

A resampling method that partitions data into folds to estimate model generalization, reducing bias in accuracy assessment (Kohavi, 1995).

Damage Assessment

The process of evaluating the extent and severity of destruction to buildings or infrastructure after disasters or conflicts. Remote sensing-based damage assessment uses optical and SAR imagery combined with ML for rapid and scalable mapping (Voigt et al., 2007, 2011, 2016).

Decision Tree

A hierarchical classifier that splits data according to feature thresholds. Applied in urban land-use and crop monitoring (Quinlan, 1986).

DenseNet

A CNN design where each layer connects to all previous layers, encouraging feature reuse and efficient gradient flow (Huang et al., 2017).

EfficientNet

A CNN family that scales depth, width, and resolution uniformly, achieving high accuracy with fewer parameters. Effective for disaster damage assessment pipelines (Tan & Le, 2019).

Ensemble Learning

Combining predictions from multiple models (e.g., Random Forests, Gradient Boosting) to improve robustness and generalization (Z.-H. Zhou, 2012).

F1 Score

The harmonic mean of precision and recall, balancing the two in imbalanced classification settings (Rijsbergen, 1979).

False Colour Composite (FCC)

A remote sensing visualization technique where non-visible spectral bands (e.g., near-infrared) are assigned to visible colour channels, enhancing features like vegetation health or urban structures (James B. Campbell & Randolph H. Wynne, 2011).

Gaussian Operator

A smoothing filter based on the Gaussian distribution, used to reduce noise and preserve larger-scale structures in imagery. Essential for preprocessing remote sensing data before edge detection or segmentation (Lindeberg, 1994).

Generative Adversarial Network (GAN)

A generative model with adversarial training between a generator and discriminator, producing realistic synthetic imagery for data augmentation (I. J. Goodfellow et al., 2014).

Gradient Descent

An iterative optimization method that adjusts parameters in the negative direction of the gradient of a loss function. Essential for training deep models on large, high-dimensional remote sensing data (Nocedal Jorge & Wright Stephen J., 2006).

Graph Neural Network (GNN)

A deep learning framework that models data as graphs, propagating information across nodes and edges. Applied to urban networks for building-to-building relational damage inference (Kipf & Welling, 2017).

Ground Truth

Reference data collected via field surveys, expert annotation, or authoritative agencies, used to validate predictions (Congalton, 1991).

Hyperspectral Imagery

Remote sensing data acquired in hundreds of narrow, contiguous spectral bands. Provides detailed spectral signatures for materials and vegetation, useful for damage detection and land-use classification (Goetz et al., 1985).

Intersection over Union (IoU)

A segmentation metric calculating the ratio of overlap to union between predicted and reference regions. Standard for building segmentation in disaster mapping (Everingham et al., 2010).

K-Nearest Neighbors (KNN)

An instance-based algorithm that assigns class labels to new samples based on the majority label among their nearest neighbors in feature space (Cover & Hart, 1967).

Kappa Coefficient

A measure of classification agreement adjusted for chance, providing a more robust metric than overall accuracy alone (Cohen, 1960).

Kernel (Image Processing)

A small matrix or filter applied across an image to perform convolution, enabling smoothing, sharpening, or feature extraction such as edge detection (Gonzalez & Woods, 2018).

Land Use and Land Cover (LULC) Classification

A workflow that assigns pixels or objects to categories such as vegetation, urban, or water, based on spectral, spatial, and contextual features. Serves as a foundation for post-disaster mapping (Richards, 2022).

Land Use Classification

The categorization of land according to human activity or economic function (e.g., residential, industrial, agricultural). Often derived from satellite-based LULC maps and applied in damage assessment and reconstruction planning (Foody, 2002; Richards, 2022).

Linear Discriminant Analysis (LDA)

A dimensionality reduction and classification method projecting features to maximize class separation. Helps mitigate the Hughes phenomenon in hyperspectral classification (Fisher, 1936).

Linear Regression

A predictive model that estimates continuous variables, such as vegetation biomass, from spectral features. Assumes a linear relationship between predictors and response (Foody, 2002; Pearson, 1896).

Logistic Regression

A probabilistic classifier that models the likelihood of class membership using a logistic function. Commonly used for binary land-cover classification (Pontius & Schneider, 2001).

Loss Function

A mathematical function quantifying the error between predicted outputs and ground truth labels, guiding optimization. Common examples: cross-entropy, mean squared error, Dice loss (I. Goodfellow et al., 2016).

Multispectral Imagery

Remote sensing data captured in a limited number of broad spectral bands (typically 4–13). Forms the backbone of Earth observation programs such as Landsat and Sentinel-2 (Wulder et al., 2019).

Naïve Bayes

A probabilistic classifier based on Bayes' theorem, assuming feature independence. Useful for rapid, lightweight land-cover classification (Manning et al., 2008).

Object-Based Image Analysis (OBIA)

An approach grouping pixels into meaningful objects based on spectral and spatial similarity, improving classification of very-high-resolution imagery (Blaschke, 2010).

Overall Accuracy (OA)

The proportion of correctly classified samples across all classes. Commonly reported but sensitive to imbalance (Congalton, 1991).

Perceptron

The first artificial neuron model, introduced by Rosenblatt (1958). It computes weighted sums of inputs and applies a step activation function, solving only linearly separable problems (Minsky & Papert, 1969).

Pixel-Based Image Analysis (PBIA)

A classification paradigm treating each pixel as an independent unit based solely on spectral information, widely applied in medium-resolution imagery (Congalton, 1991).

Polynomial Regression

An extension of linear regression that includes higher-order terms to capture non-linear relationships between predictors and response variables. Applied in vegetation dynamics, soil reflectance modeling, and crop yield estimation (Curran, 1989).

Precision

The proportion of predicted positives that are true positives, reflecting reliability of model predictions (Sokolova & Lapalme, 2009).

Principal Component Analysis (PCA)

A statistical technique that transforms correlated features into a smaller set of uncorrelated components, retaining the majority of variance. Commonly used to reduce hyperspectral imagery dimensionality (I. T. Jolliffe, 2002).

Recall (Sensitivity)

The proportion of actual positives correctly identified by the model, reflecting its ability to detect damaged structures (Foody & Mathur, 2004).

Remote Sensing

The science of acquiring information about the Earth's surface without direct contact, typically through satellite, airborne, or UAV sensors, by measuring reflected or emitted electromagnetic radiation (James B. Campbell & Randolph H. Wynne, 2011).

ResNet

A CNN architecture introducing residual (identity) connections that allow training of very deep networks by stabilizing gradient propagation (He et al., 2016).

Root Mean Square Error (RMSE)

A statistical metric measuring the average squared deviation between predicted and reference values, sensitive to large errors. Common in regression-based remote sensing applications (Willmott & Matsuura, 2005).

SAR (Synthetic Aperture Radar)

An active remote sensing technology that synthesizes a long antenna by exploiting platform motion to generate high-resolution microwave imagery. SAR penetrates clouds and operates day-and-night, making it crucial for battlefield and disaster monitoring (Flores et al., 2019; Moreira et al., 2013).

Segmentation (Remote Sensing)

The partitioning of imagery into homogeneous regions or objects, forming the basis for classification and object extraction. In remote sensing, segmentation enables object-based analysis of urban features or damaged structures (Blaschke, 2010; James B. Campbell & Randolph H. Wynne, 2011).

SegNet

An encoder–decoder segmentation architecture that uses pooling indices for upsampling, making it memory-efficient while retaining spatial detail (Badrinarayanan et al., 2017).

Self-Supervised Learning (SSL)

A training paradigm leveraging unlabeled data by designing pretext tasks (e.g., contrastive learning) to learn representations, later fine-tuned on limited labeled samples (T. Chen et al., 2020).

Siamese Network

An architecture using two identical subnetworks with shared weights to learn differences between paired inputs, enabling change detection from pre- and post-disaster imagery (Bromley et al., 1993; T. Chen et al., 2022; Daudt et al., 2018; Khvedchenya & Gabruseva, 2021).

Structural Similarity Index (SSIM)

A perceptual metric comparing image quality based on luminance, contrast, and structure, used to evaluate image translation methods such as SAR-to-optical GANs (Z. Wang et al., 2004).

Support Vector Machine (SVM)

A supervised learning model that classifies data by finding the optimal separating hyperplane in feature space. Robust in high-dimensional imagery such as hyperspectral datasets (Cortes & Vapnik, 1995; Mountrakis et al., 2011; Pal & Mather, 2005; Platt, 2000).

Transfer Learning

A method where models pretrained on large datasets (e.g., ImageNet) are fine-tuned on smaller domain-specific datasets. Widely used in remote sensing to overcome data scarcity (Bouchard et al., 2022; Bukhsh et al., 2021; Deng et al., 2009; Enayatullah Meskinyaar, n.d.; Khankeshizadeh et al., 2024; Seo et al., 2019, 2019; Tilon et al., 2020).

True Colour Composite (TCC)

An image visualization method in which the red, green, and blue channels of an optical sensor are displayed as they appear to the human eye. Useful for intuitive interpretation of satellite imagery (James B. Campbell & Randolph H. Wynne, 2011; Richards, 2022).

U-Net

An encoder–decoder CNN architecture with skip connections that preserve spatial detail, enabling precise segmentation of objects such as buildings and rubble (Oktay et al., 2018; Ronneberger et al., 2015)

V-Net

A 3D CNN architecture designed for volumetric segmentation, using 3D convolutions and Dice loss to segment voxel-based inputs such as LiDAR volumes (Milletari et al., 2016).

Very High Resolution (VHR)

Satellite or aerial imagery with spatial resolution finer than ~1 m. Enables object-level tasks such as individual building segmentation for disaster and battlefield analysis (Wulder et al., 2019).

8. Acronyms and Abbreviations

Acronym	Full Term	Explanation / Context
ANN	Artificial Neural Network	Computational model inspired by biological neurons, used for supervised and unsupervised learning.
AUC	Area Under the Curve	Evaluation metric summarizing classifier performance across thresholds (ROC curve).
BDA	Battlefield Damage Assessment	Systematic evaluation of structural damage in conflict zones using remote sensing and AI.
BDANet	Building Damage Assessment Network	Specialized deep learning architecture with attention mechanisms for post-disaster building damage mapping.
BYOL	Bootstrap Your Own Latent	Self-supervised learning method without negative pairs, used in representation learning.
CNN	Convolutional Neural Network	Core architecture for spatial feature extraction from imagery.
DL	Deep Learning	Subfield of machine learning using deep neural network architectures.
EO	Earth Observation	Acquisition of data about Earth's surface via satellites, aircraft, or UAV sensors.
ELU	Exponential Linear Unit	Activation function variant addressing vanishing gradients.
FCC	False Colour Composite	Visualization of remote sensing data using non-visible bands in RGB channels (e.g., NIR-Red-Green).
F1	F1 Score	Harmonic mean of precision and recall, used for classification performance.
GAN	Generative Adversarial Network	Generative model framework used for data augmentation and image synthesis.
GIS	Geographic Information System	Framework for managing, analyzing, and visualizing spatial data.
GNN	Graph Neural Network	Deep learning model that learns from graph-structured data such as building adjacency.
IoU	Intersection over Union	Segmentation metric measuring spatial overlap between prediction and ground truth.
KNN	K-Nearest Neighbors	Instance-based machine learning classifier.

LDA	Linear Discriminant Analysis	Method for classification and dimensionality reduction.
LULC	Land Use and Land Cover	Classification of surface types based on satellite imagery.
MLP	Multilayer Perceptron	A class of feedforward artificial neural networks.
MNF	Minimum Noise Fraction	Dimensionality reduction method for separating signal from noise in hyperspectral data.
MSE	Mean Squared Error	Loss function used in regression tasks.
OA	Overall Accuracy	Percentage of correctly classified samples across all classes.
OBIA	Object-Based Image Analysis	Classification approach using segmented image objects rather than individual pixels.
PCA	Principal Component Analysis	Statistical method for dimensionality reduction in spectral data.
PBIA	Pixel-Based Image Analysis	Traditional classification method assigning labels per pixel.
ReLU	Rectified Linear Unit	Widely used activation function in deep learning.
RF	Random Forest	Ensemble learning method based on decision trees.
RMSE	Root Mean Square Error	Statistical metric measuring average magnitude of prediction errors.
ROC	Receiver Operating Characteristic	Curve plotting TPR vs. FPR across classification thresholds.
RS	Remote Sensing	Acquisition of information about Earth's surface via satellite or airborne sensors.
SAR	Synthetic Aperture Radar	Active microwave imaging system for all-weather, day-night observation.
SGD	Stochastic Gradient Descent	Optimization method used in training deep learning models.
SNR	Signal-to-Noise Ratio	Ratio used to quantify the level of desired signal relative to background noise.
SSL	Self-Supervised Learning	Training paradigm using unlabeled data to learn representations.
SVM	Support Vector Machine	Supervised machine learning algorithm for classification.
TCC	True Colour Composite	Visualization of imagery in natural colors (RGB).

TP/TN/FP/FN	True Positive / True Negative / False Positive / False Negative	Elements of a confusion matrix for classification.
UAV	Unmanned Aerial Vehicle	Drone platforms used for aerial imaging.
VHR	Very High Resolution	Satellite/aerial imagery with spatial resolution finer than ~1 m.
XAI	Explainable Artificial Intelligence	Methods to make AI model predictions interpretable.
xBD	xView Building Damage Dataset	Large-scale dataset for training and evaluating damage assessment models.
XGB	Extreme Gradient Boosting (XGBoost)	Scalable gradient boosting framework.

Table 3: Acronymns and Abbreviations

9. List of Figures

Figure 1: Gradient descent in 3 Dimensions. Source:(Alexander Amini & Daniela Rus, 2018) ..	1
Figure 2: Gaussian Kernel in 3d. Source:(Kernel Cookbook, n.d.)	7
Figure 3: "Mexican Sombrero" - LoG. Source: (Spatial Filters - Laplacian/Laplacian of Gaussian, n.d.)	7
Figure 4: The mechanics of linear spatial filtering using a 3×3 kernel. The pixels are shown as squares to simplify the graphics. Note that the origin of the image is at the top left, but the origin of the kernel is at its center. Placing the origin at the center of spatially symmetric kernels simplifies writing expressions for linear filtering. (Gonzalez & Woods, 2018)	8
Figure 5: PCA (Marridi, 2016)	9
Figure 6: True colour (A) and land classification (B) images of the studied area.	12
Figure 7: Linear support vector machine example. (Mountrakis et al., 2011)	13
Figure 8: KNN Illustration. Source: (Shah, 2021).....	14
Figure 9: Decision Tree Example for Mushrooms. Source: (Gunluk et al., 2021)	15
Figure 10: Example of Linear Regression: a) Scatter diagram for delivery volume. (b) Straight - line relationship between delivery time and delivery volume. (Khuri, 2013)	16
Figure 11: Perceptron - The McCulloch–Pitts neuron model (Du et al., 2022)	20
Figure 12: Different Activation Functions (Everton Gomedé, 2024)	21
Figure 13: Different loss functions used in boosting and supervised machine learning methods. (Leistner et al., 2009).....	22
Figure 14: A simple illustration of how the backpropagation works by adjustments of weight. Source: (Backpropagation in Neural Network, 16:07:27+00:00)	24
Figure 15: Mapping Buildings with U-Net. (Zhang et al., 2021)	28
Figure 16: Using an error matrix to summarise classifier performance and map accuracy (Richards, 2022).....	47

10. List of Tables

Table 1: Loss Functions and Use Cases in BDA.....	23
Table 2: Confusion Matrix, derived from (Richards, 2022)	47
Table 3: Acronymns and Abbreviations	58

11. References

1. Adriano, B., Yokoya, N., Xia, J., Miura, H., Liu, W., Matsuoka, M., & Koshimura, S. (2021). Learning from multimodal and multitemporal earth observation data for building damage mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175, 132–143. <https://doi.org/10.1016/j.isprsjprs.2021.02.016>
2. Ahmad, M. N., Shao, Z., Xiao, X., Fu, P., Javed, A., & Ara, I. (2024). A novel ensemble learning approach to extract urban impervious surface based on machine learning algorithms using SAR and optical data. *International Journal of Applied Earth Observation and Geoinformation*, 132, 104013. <https://doi.org/10.1016/j.jag.2024.104013>
3. Ahmadi, S. A., Mohammadzadeh, A., Yokoya, N., & Ghorbanian, A. (2023). BD-SKUNet: Selective-Kernel UNets for Building Damage Assessment in High-Resolution Satellite Images. *Remote Sensing*, 16(1), Article 1. <https://doi.org/10.3390/rs16010182>
4. Aimaiti, Y., Sanon, C., Koch, M., Baise, L. G., & Moaveni, B. (2022). War Related Building Damage Assessment in Kyiv, Ukraine, Using Sentinel-1 Radar and Sentinel-2 Optical Images. *Remote Sensing*, 14(24), Article 24. <https://doi.org/10.3390/rs14246239>
5. Al Shafian, S., & Hu, D. (2024). Integrating Machine Learning and Remote Sensing in Disaster Management: A Decadal Review of Post-Disaster Building Damage Assessment. *Buildings*, 14(8), Article 8. <https://doi.org/10.3390/buildings14082344>
6. AlDahoul, N., Karim, H. A., & Momo, M. A. (2022). RGB-D based multi-modal deep learning for spacecraft and debris recognition. *Scientific Reports*, 12(1), 3924. <https://doi.org/10.1038/s41598-022-07846-5>
7. Alexander Amini & Daniela Rus. (2018). *AI researchers allege that machine learning is alchemy* [Graphic]. <https://www.science.org/content/article/ai-researchers-allege-machine-learning-alchemy>
8. Alisjahbana, I., Li, J., Ben, Strong, & Zhang, Y. (2024). *DeepDamageNet: A two-step deep-learning model for multi-disaster building damage segmentation and classification using satellite imagery*. arXiv. <https://doi.org/10.48550/ARXIV.2405.04800>
9. Al-Khudhairi, D. H. A., Caravaggi, I., & Giada, S. (2005). Structural Damage Assessments from Ikonos Data Using Change Detection, Object-Oriented Segmentation, and Classification

- Techniques. *Photogramm Eng Remote Sensing*, 71(7), Article 7.
<https://doi.org/10.14358/PERS.71.7.825>
10. Antoniou, A., Storkey, A., & Edwards, H. (2018). *Data Augmentation Generative Adversarial Networks* (No. arXiv:1711.04340). arXiv.
<https://doi.org/10.48550/arXiv.1711.04340>
11. Atkinson, P. M., & Tatnall, A. R. L. (1997). Introduction Neural networks in remote sensing. *International Journal of Remote Sensing*, 18(4), 699–709.
<https://doi.org/10.1080/014311697218700>
12. Ayush, K., UzKent, B., Meng, C., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021). Geography-Aware Self-Supervised Learning. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 10161–10170. <https://doi.org/10.1109/ICCV48922.2021.01002>
13. *Backpropagation in Neural Network*. (16:07:27+00:00). GeeksforGeeks.
<https://www.geeksforgeeks.org/machine-learning/backpropagation-in-neural-network/>
14. Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
15. Ball, J. E., Anderson, D. T., & Chan, C. S. (2017). A Comprehensive Survey of Deep Learning in Remote Sensing: Theories, Tools and Challenges for the Community. *Journal of Applied Remote Sensing*, 11(04), 1. <https://doi.org/10.1117/1.JRS.11.042609>
16. Balz, T., & Liao, M. (2010). Building-damage detection using post-seismic high-resolution SAR satellite data. *International Journal of Remote Sensing*, 31(13), 3369–3391.
<https://doi.org/10.1080/01431161003727671>
17. Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.
<https://doi.org/10.1016/j.isprsjprs.2016.01.011>
18. Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G., & Dupuy, S. (2018). Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12), 4939–4949. <https://doi.org/10.1109/JSTARS.2018.2876357>

19. Benson, V., & Ecker, A. (2020). *Assessing out-of-domain generalization for robust building damage detection* (No. arXiv:2011.10328). arXiv. <https://doi.org/10.48550/arXiv.2011.10328>
20. Bhardwaj, D., Nagabhooshanam, N., Singh, A., Selvalakshmi, B., Angadi, S., Shargunam, S., Guha, T., Singh, G., & Rajaram, A. (2024). Enhanced satellite imagery analysis for post-disaster building damage assessment using integrated ResNet-U-Net model. *Multimed Tools Appl*, 84(5), Article 5. <https://doi.org/10.1007/s11042-024-20300-0>
21. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). Hyperspectral Remote Sensing Data Analysis and Future Challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2), 6–36. <https://doi.org/10.1109/MGRS.2013.2244672>
22. Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1), 2–16. <https://doi.org/10.1016/j.isprsjprs.2009.06.004>
23. Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., Van Der Meer, F., Van Der Werff, H., Van Coillie, F., & Tiede, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. <https://doi.org/10.1016/j.isprsjprs.2013.09.014>
24. Bouchard, I., Rancourt, M.-È., Aloise, D., & Kalaitzis, F. (2022). On Transfer Learning for Building Damage Assessment from Satellite Imagery in Emergency Contexts. *Remote Sensing*, 14(11), Article 11. <https://doi.org/10.3390/rs14112532>
25. Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7), 1145–1159. [https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)
26. Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
27. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
28. Breiman, L., Friedman, J., Olshen, R. A., & Stone, C. J. (2017). *Classification and Regression Trees*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>

29. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature verification using a “Siamese” time delay neural network. *Proceedings of the 7th International Conference on Neural Information Processing Systems*, 737–744.
30. Bukhsh, Z. A., Jansen, N., & Saeed, A. (2021). Damage detection using in-domain and cross-domain transfer learning. *Neural Computing and Applications*, 33(24), 16921–16936. <https://doi.org/10.1007/s00521-021-06279-x>
31. Burnett, C., & Blaschke, T. (2003). A multi-scale segmentation/object relationship modelling methodology for landscape analysis. *Ecological Modelling*, 168(3), 233–249. [https://doi.org/10.1016/S0304-3800\(03\)00139-X](https://doi.org/10.1016/S0304-3800(03)00139-X)
32. Cai, H., Chen, T., Niu, R., & Plaza, A. (2021). Landslide Detection Using Densely Connected Convolutional Networks and Environmental Conditions. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 5235–5247. <https://doi.org/10.1109/JSTARS.2021.3079196>
33. Canny, J. (1986). A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679–698. <https://doi.org/10.1109/TPAMI.1986.4767851>
34. Chawanji, S., Fleischer, K., & Ullrich, J. (2022). Copernicus Emergency Management Service (CEMS) – Risk and Recovery Mapping. *Abstracts of the ICA*, 5, 1–2. <https://doi.org/10.5194/ica-abs-5-118-2022>
35. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
36. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*, 119, 1597–1607.
37. Chen, T., Lu, Z., Yang, Y., Zhang, Y., Du, B., & Plaza, A. (2022). A siamese network based u-net for change detection in high resolution remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2357–2369. <https://doi.org/10.1109/JSTARS.2022.3157648>

38. Chen, Y., Nasrabadi, N. M., & Tran, T. D. (2011). Hyperspectral Image Classification Using Dictionary-Based Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10), 3973–3985. <https://doi.org/10.1109/TGRS.2011.2129595>
39. Ci, T., Liu, Z., & Wang, Y. (2019). Assessment of the Degree of Building Damage Caused by Disaster Using Convolutional Neural Networks in Combination with Ordinal Regression. *Remote Sensing*, 11(23), 2858. <https://doi.org/10.3390/rs11232858>
40. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, & W. Wells (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (pp. 424–432). Springer International Publishing. https://doi.org/10.1007/978-3-319-46723-8_49
41. Ciocarlan, A., & Stoian, A. (2021). Ship Detection in Sentinel 2 Multi-Spectral Images with Self-Supervised Learning. *Remote Sensing*, 13(21), 4255. <https://doi.org/10.3390/rs13214255>
42. Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
43. Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, 37(1), 35–46. [https://doi.org/10.1016/0034-4257\(91\)90048-B](https://doi.org/10.1016/0034-4257(91)90048-B)
44. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
45. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
46. Cui, L., Jing, X., Wang, Y., Huan, Y., Xu, Y., & Zhang, Q. (2023). Improved Swin Transformer-Based Semantic Segmentation of Postearthquake Dense Buildings in Urban Areas Using Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, 16, 369–385. <https://doi.org/10.1109/JSTARS.2022.3225150>
47. Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, 30(3), 271–278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2)

48. Daudt, R. C., Saux, B. L., & Boulch, A. (2018). *Fully Convolutional Siamese Networks for Change Detection* (No. arXiv:1810.08462). arXiv. <https://doi.org/10.48550/arXiv.1810.08462>
49. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
50. Dietrich, O., Peters, T., Sainte Fare Garnot, V., Sticher, V., Ton-That Whelan, T., Schindler, K., & Wegner, J. D. (2025). An open-source tool for mapping war destruction at scale in Ukraine using Sentinel-1 time series. *Commun Earth Environ*, 6(1), Article 1. <https://doi.org/10.1038/s43247-025-02183-7>
51. Doersch, C., Gupta, A., & Efros, A. A. (2015). Unsupervised Visual Representation Learning by Context Prediction. *2015 IEEE International Conference on Computer Vision (ICCV)*, 1422–1430. <https://doi.org/10.1109/ICCV.2015.167>
52. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. <https://doi.org/10.48550/arXiv.2010.11929>
53. Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Applications*, 91, 464–471. <https://doi.org/10.1016/j.eswa.2017.09.030>
54. Du, K.-L., Leung, C.-S., Mow, W. H., & Swamy, M. N. S. (2022). Perceptron: Learning, Generalization, Model Selection, Fault Tolerance, and Role in the Deep Learning Era. *Mathematics*, 10(24), 4730. <https://doi.org/10.3390/math10244730>
55. Enayatullah Meskinyaar. (n.d.). *FEW-SHOTS LEARNING FOR POST-EARTHQUAKE BUILDING DAMAGE ASSESSMENT USING METRICBASED AND TRANSFER LEARNING METHODS*.
56. Etten, A., Lindenbaum, D., & Bacastow, T. (2018). *SpaceNet: A Remote Sensing Dataset and Challenge Series*. <https://doi.org/10.48550/arXiv.1807.01232>
57. Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338. <https://doi.org/10.1007/s11263-009-0275-4>

58. Everton Gomedé. (2024, March 6). An Overview of Activation Functions in Deep Learning. *The Modern Scientist*. <https://medium.com/the-modern-scientist/an-overview-of-activation-functions-in-deep-learning-97a85ac00460>
59. Fisher, R. A. (1936). THE USE OF MULTIPLE MEASUREMENTS IN TAXONOMIC PROBLEMS. *Annals of Eugenics*, 7(2), 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
60. Fix, E., & Hodges, J. L. (1989). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties. *International Statistical Review / Revue Internationale de Statistique*, 57(3), 238–247. <https://doi.org/10.2307/1403797>
61. Flores, A., Herndon, K., Thapa, R., & Cherrington, E. (2019). *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. <https://doi.org/10.25966/NR2C-S697>
62. Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1), 185–201. [https://doi.org/10.1016/S0034-4257\(01\)00295-4](https://doi.org/10.1016/S0034-4257(01)00295-4)
63. Foody, G. M., & Atkinson, P. M. (Eds.). (2002). *Uncertainty in Remote Sensing and GIS* (1st ed.). Wiley. <https://doi.org/10.1002/0470035269>
64. Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1), 107–117. <https://doi.org/10.1016/j.rse.2004.06.017>
65. Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
66. Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61(3), 399–409. [https://doi.org/10.1016/S0034-4257\(97\)00049-7](https://doi.org/10.1016/S0034-4257(97)00049-7)
67. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
68. Fukunaga, K. (1990). *Introduction to statistical pattern recognition* (2. ed). Acad. Press.
69. Gao, X., Chen, T., Niu, R., & Plaza, A. (2021). Recognition and Mapping of Landslide Using a Fully Convolutional DenseNet and Influencing Factors. *IEEE Journal of Selected Topics in*

- Applied Earth Observations and Remote Sensing*, 14, 7881–7894.
<https://doi.org/10.1109/JSTARS.2021.3101203>
70. Gharahbagh, A. A., Hajhashemi, V., Machado, J. J. M., & Tavares, J. M. R. S. (2025). Land Cover Classification Model Using Multispectral Satellite Images Based on a Deep Learning Synergistic Semantic Segmentation Network. *Sensors*, 25(7), 1988.
<https://doi.org/10.3390/s25071988>
71. Gholami, S., Robinson, C., Ortiz, A., Yang, S., Margutti, J., Birge, C., Dodhia, R., & Ferres, J. L. (2022). On the Deployment of Post-Disaster Building Damage Assessment Tools using Satellite Imagery: A Deep Learning Approach. *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1029–1036. <https://doi.org/10.1109/ICDMW58026.2022.00134>
72. Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, 27(4), 294–300.
<https://doi.org/10.1016/j.patrec.2005.08.011>
73. Goetz, A. F., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging spectrometry for Earth remote sensing. *Science (New York, N.Y.)*, 228(4704), 1147–1153.
<https://doi.org/10.1126/science.228.4704.1147>
74. Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., ... Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7), 2607–2654.
<https://doi.org/10.1080/01431161.2012.748992>
75. Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing*. Pearson.
76. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
77. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 27.
https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

78. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. *Advances in Neural Information Processing Systems*, 3. <https://doi.org/10.1145/3422622>
79. Green, A. A., Berman, M., Switzer, P., & Craig, M. D. (1988). A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65–74. <https://doi.org/10.1109/36.3001>
80. Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., Faust, J. A., Pavri, B. E., Chovit, C. J., Solis, M., Olah, M. R., & Williams, O. (1998). Imaging Spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, 65(3), 227–248. [https://doi.org/10.1016/S0034-4257\(98\)00064-9](https://doi.org/10.1016/S0034-4257(98)00064-9)
81. Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap your own latent a new approach to self-supervised learning. *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 21271–21284.
82. Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., ... Sang, B. (2015). The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sensing*, 7(7), 8830–8857. <https://doi.org/10.3390/rs70708830>
83. Gunluk, O., Kalagnanam, J., Li, M., Menickelly, M., & Scheinberg, K. (2021). Optimal decision trees for categorical data via integer programming. *Journal of Global Optimization*, 81, 1–28. <https://doi.org/10.1007/s10898-021-01009-y>
84. Gupta, R., Hosfelt, R., Sajeev, S., Patel, N., Goodman, B., Doshi, J., Heim, E., Choset, H., & Gaston, M. (2019). *xBD: A Dataset for Assessing Building Damage from Satellite Imagery*. arXiv. <https://doi.org/10.48550/ARXIV.1911.09296>
85. Gupta, R., & Shah, M. (2021). RescueNet: Joint Building Segmentation and Damage Assessment from Satellite Imagery. *2020 25th International Conference on Pattern Recognition (ICPR)*, 4405–4411. <https://doi.org/10.1109/ICPR48806.2021.9412295>

86. Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21(6–7), 1331–1364. <https://doi.org/10.1080/014311600210209>
87. Hao, X., Liu, L., Yang, R., Yin, L., Zhang, L., & Li, X. (2023). A Review of Data Augmentation Methods of Remote Sensing Image Target Recognition. *Remote Sensing*, 15(3), 827. <https://doi.org/10.3390/rs15030827>
88. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
89. Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice hall PTR.
90. Haykin, S. (1999). *Neural networks: a comprehensive foundation* (2nd ed). Pearson Education.
91. He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum Contrast for Unsupervised Visual Representation Learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9726–9735. <https://doi.org/10.1109/CVPR42600.2020.00975>
92. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
93. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786), 504–507. <https://doi.org/10.1126/science.1127647>
94. Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2), 251–257. [https://doi.org/10.1016/0893-6080\(91\)90009-T](https://doi.org/10.1016/0893-6080(91)90009-T)
95. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Third edition). Wiley.
96. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
97. Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1), 55–63. <https://doi.org/10.1109/TIT.1968.1054102>

98. I. T. Jolliffe. (2002). *Principal Component Analysis*. Springer-Verlag.
<https://doi.org/10.1007/b98835>
99. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., & Maier-Hein, K. H. (2018). *nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation* (No. arXiv:1809.10486). arXiv.
<https://doi.org/10.48550/arXiv.1809.10486>
100. Jain, R., Kasturi, R., & Schunck, B. (1995). *Machine vision*.
101. James B. Campbell & Randolph H. Wynne. (2011). *Introduction to Remote Sensing* (5th ed., Vol. 5). Guilford Press, New York. <https://www.mdpi.com/2072-4292/5/1/282>
102. Ji, M., Liu, L., & Buchroithner, M. (2018). Identifying Collapsed Buildings Using Post-Earthquake Satellite Imagery and Convolutional Neural Networks: A Case Study of the 2010 Haiti Earthquake. *Remote Sensing*, 10(11), Article 11. <https://doi.org/10.3390/rs10111689>
103. Jin, C., Ge, R., Netrapalli, P., Kakade, S. M., & Jordan, M. I. (2017). How to Escape Saddle Points Efficiently. *Proceedings of the 34th International Conference on Machine Learning*, 1724–1732. <https://proceedings.mlr.press/v70/jin17a.html>
104. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
105. Jones, S., & Saniie, J. (2019). Using Deep Learning and Satellite Imagery to Assess the Damage to Civil Structures After Natural Disasters. *2019 IEEE International Conference on Electro Information Technology (EIT)*, 189–193. <https://doi.org/10.1109/EIT.2019.8833724>
106. Kaur, N., Lee, C., Mostafavi, A., & Mahdavi-Amiri, A. (2023). Large-scale building damage assessment using a novel hierarchical transformer architecture on satellite images. *Computer Aided Civil Eng*, 38(15), Article 15. <https://doi.org/10.1111/mice.12981>
107. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 30.
https://papers.nips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

108. Kerle, N., Nex, F., Gerke, M., Duarte, D., & Vetrivel, A. (2020). UAV-Based Structural Damage Mapping: A Review. *ISPRS International Journal of Geo-Information*, 9(1), 14. <https://doi.org/10.3390/ijgi9010014>
109. *Kernel Cookbook*. (n.d.). <https://www.cs.toronto.edu/~duvenaud/cookbook/>
110. 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* (No. arXiv:1609.04836). arXiv. <https://doi.org/10.48550/arXiv.1609.04836>
111. Khankeshizadeh, E., Mohammadzadeh, A., Arefi, H., Mohsenifar, A., Pirasteh, S., Fan, E., Li, H., & Li, J. (2024). A Novel Weighted Ensemble Transferred U-Net Based Model (WETUM) for Postearthquake Building Damage Assessment From UAV Data: A Comparison of Deep Learning- and Machine Learning-Based Approaches. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–17. <https://doi.org/10.1109/TGRS.2024.3354737>
112. Khuri, A. I. (2013). Introduction to Linear Regression Analysis, Fifth Edition by Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining. *International Statistical Review*, 81(2), 318–319. https://doi.org/10.1111/insr.12020_10
113. Khvedchenya, E., & Gabruseva, T. (2021). *Fully convolutional Siamese neural networks for buildings damage assessment from satellite images*. arXiv. <https://doi.org/10.48550/ARXIV.2111.00508>
114. Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (No. arXiv:1412.6980). arXiv. <https://doi.org/10.48550/arXiv.1412.6980>
115. Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks* (No. arXiv:1609.02907). arXiv. <https://doi.org/10.48550/arXiv.1609.02907>
116. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, 1137–1143.
117. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>

118. Lee, J.-S., & Pottier, E. (2017). *Polarimetric Radar Imaging: From Basics to Applications*. CRC Press. <https://doi.org/10.1201/9781420054989>
119. Leistner, C., Saffari, A., Roth, P. M., & Bischof, H. (2009). On robustness of on-line boosting - a competitive study. *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 1362–1369. <https://doi.org/10.1109/ICCVW.2009.5457451>
120. Li, P., Xu, H., Liu, S., & Guo, J. (2009). Urban building damage detection from very high resolution imagery using one-class SVM and spatial relations. *2009 IEEE International Geoscience and Remote Sensing Symposium*, 5, V-112–V-114. <https://doi.org/10.1109/IGARSS.2009.5417719>
121. Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Springer US. <https://doi.org/10.1007/978-1-4757-6465-9>
122. Liu, C., Sui, H., & Huang, L. (2020). Identification of Building Damage from UAV-Based Photogrammetric Point Clouds Using Supervoxel Segmentation and Latent Dirichlet Allocation Model. *Sensors*, 20(22), Article 22. <https://doi.org/10.3390/s20226499>
123. Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., & Mi, J. (2021). Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects. *Journal of Remote Sensing*, 2021, 2021/5289697. <https://doi.org/10.34133/2021/5289697>
124. Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2), 645–657. <https://doi.org/10.1109/TGRS.2016.2612821>
125. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge university press.
126. Mantero, P., Moser, G., & Serpico, S. B. (2005). Partially Supervised classification of remote sensing images through SVM-based probability density estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 559–570. <https://doi.org/10.1109/TGRS.2004.842022>

127. Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society (Great Britain)*, 207, 187–217. <https://doi.org/10.1098/rspb.1980.0020>
128. Marridi, Z. (2016, November 7). Principal Component Analysis (PCA) 101. *NumXL*. <https://numxl.com/blogs/principal-component-analysis-pca-101/>
129. McRoberts, R. E., & Walters, B. F. (2012). Statistical inference for remote sensing-based estimates of net deforestation. *Remote Sensing of Environment*, 124, 394–401. <https://doi.org/10.1016/j.rse.2012.05.011>
130. Meyer, F. (2019). Spaceborne Synthetic Aperture Radar: Principles, data access, and basic processing techniques. *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*, 21–64.
131. Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation (No. arXiv:1606.04797). arXiv. <https://doi.org/10.48550/arXiv.1606.04797>
132. Minsky, M. & Papert. (1969). *Perceptrons*. MIT press.
133. Mirza, M., & Osindero, S. (2014). *Conditional Generative Adversarial Nets* (No. arXiv:1411.1784). arXiv. <https://doi.org/10.48550/arXiv.1411.1784>
134. Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., & Papathanassiou, K. P. (2013). A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, 1(1), 6–43. <https://doi.org/10.1109/MGRS.2013.2248301>
135. Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>
136. Mueller, H., Groeger, A., Hersh, J., Matranga, A., & Serrat, J. (2021). Monitoring war destruction from space using machine learning. *Proc. Natl. Acad. Sci. U.S.A.*, 118(23), Article 23. <https://doi.org/10.1073/pnas.2025400118>
137. Nair, V., & Hinton, G. E. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*.

138. Nocedal Jorge & Wright Stephen J. (2006). *Numerical Optimization*. Springer New York.
<https://doi.org/10.1007/978-0-387-40065-5>
139. Noroozi, M., & Favaro, P. (2017). *Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles* (No. arXiv:1603.09246). arXiv.
<https://doi.org/10.48550/arXiv.1603.09246>
140. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). *Attention U-Net: Learning Where to Look for the Pancreas* (No. arXiv:1804.03999). arXiv.
<https://doi.org/10.48550/arXiv.1804.03999>
141. Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86(4), 554–565.
[https://doi.org/10.1016/S0034-4257\(03\)00132-9](https://doi.org/10.1016/S0034-4257(03)00132-9)
142. Pal, M., & Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5), 1007–1011.
<https://doi.org/10.1080/01431160512331314083>
143. Pan, Z., Xu, J., Guo, Y., Hu, Y., & Wang, G. (2020). Deep Learning Segmentation and Classification for Urban Village Using a Worldview Satellite Image Based on U-Net. *Remote Sensing*, 12(10), Article 10. <https://doi.org/10.3390/rs12101574>
144. PAOLA, J. D., & SCHOWENGERDT, R. A. (1995). A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing*, 16(16), 3033–3058. <https://doi.org/10.1080/01431169508954607>
145. Parikh, H., Patel, S., & Patel, V. (2020). Classification of SAR and PolSAR images using deep learning: a review. *International Journal of Image and Data Fusion*, 11(1), 1–32.
<https://doi.org/10.1080/19479832.2019.1655489>
146. Passah, A., Sur, S. N., Paul, B., & Kandar, D. (2022). SAR Image Classification: A Comprehensive Study and Analysis. *IEEE Access*, 10, 20385–20399.
<https://doi.org/10.1109/ACCESS.2022.3151089>

147. Pearlman, J. S., Barry, P. S., Segal, C. C., Shepanski, J., Beiso, D., & Carman, S. L. (2003). Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), 1160–1173. <https://doi.org/10.1109/TGRS.2003.815018>
148. Pearson, K. (1896). VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 187, 253–318. <https://doi.org/10.1098/rsta.1896.0007>
149. Petropoulos, G. P., Kontoes, C., & Keramitsoglou, I. (2011). Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. *International Journal of Applied Earth Observation and Geoinformation*, 13(1), 70–80. <https://doi.org/10.1016/j.jag.2010.06.008>
150. Plank, S. (2014). Rapid Damage Assessment by Means of Multi-Temporal SAR — A Comprehensive Review and Outlook to Sentinel-1. *Remote Sensing*, 6(6), Article 6. <https://doi.org/10.3390/rs6064870>
151. Platt, J. (2000). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv. Large Margin Classif.*, 10.
152. Plaza, A., Benediktsson, J. A., Boardman, J. W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J. C., & Trianni, G. (2009). Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113, S110–S122. <https://doi.org/10.1016/j.rse.2007.07.028>
153. Pontius, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429. <https://doi.org/10.1080/01431161.2011.552923>
154. Pontius, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, 85(1), 239–248. [https://doi.org/10.1016/S0167-8809\(01\)00187-6](https://doi.org/10.1016/S0167-8809(01)00187-6)
155. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019). *CatBoost: unbiased boosting with categorical features* (No. arXiv:1706.09516). arXiv. <https://doi.org/10.48550/arXiv.1706.09516>

156. Qing, Y., Ming, D., Wen, Q., Weng, Q., Xu, L., Chen, Y., Zhang, Y., & Zeng, B. (2022). Operational earthquake-induced building damage assessment using CNN-based direct remote sensing change detection on superpixel level. *International Journal of Applied Earth Observation and Geoinformation*, 112, 102899. <https://doi.org/10.1016/j.jag.2022.102899>
157. Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. <https://doi.org/10.1007/BF00116251>
158. Rao, A., Jung, J., Silva, V., Molinario, G., & Yun, S.-H. (2023). Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning. *Natural Hazards and Earth System Sciences*, 23(2), 789–807. <https://doi.org/10.5194/nhess-23-789-2023>
159. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 658–666. <https://doi.org/10.1109/CVPR.2019.00075>
160. Richards, J. A. (2022). *Remote Sensing Digital Image Analysis*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-82327-6>
161. Rijsbergen, C. J. V. (1979). *Information Retrieval*. Butterworths.
162. Robbins, Herbert & Monro, Sutton. (1951). *A Stochastic Approximation Method*.
163. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
164. Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), 386–408. <https://doi.org/10.1037/h0042519>
165. Ruder, S. (2017). *An overview of gradient descent optimization algorithms* (No. arXiv:1609.04747). arXiv. <https://doi.org/10.48550/arXiv.1609.04747>
166. Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). *Learning representations by back-propagating errors*.

167. Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 95(2), 177–194. <https://doi.org/10.1016/j.rse.2004.12.016>
168. Selvakumaran, S., Rolland, I., Cullen, L., Davis, R., Macabuag, J., Chakra, C. A., Karageozian, N., Gilani, A., Geiß, C., Bravo-Haro, M., & Marinoni, A. (2025). Improving operational use of post-disaster damage assessment for Urban Search and Rescue by integrated graph-based multimodal remote sensing data analysis. *Progress in Disaster Science*, 25, 100404. <https://doi.org/10.1016/j.pdisas.2025.100404>
169. Seo, J., Lee, S., Kim, B., & Jeon, T. (2019). *Revisiting Classical Bagging with Modern Transfer Learning for On-the-fly Disaster Damage Detector* (No. arXiv:1910.01911). arXiv. <https://doi.org/10.48550/arXiv.1910.01911>
170. Shah, R. (2021, March 5). *Introduction to k-Nearest Neighbors (kNN) Algorithm*. Medium. <https://ai.plainenglish.io/introduction-to-k-nearest-neighbors-knn-algorithm-e8617a448fa8>
171. Shao, J., Tang, L., Liu, M., Shao, G., Sun, L., & Qiu, Q. (2020). BDD-Net: A General Protocol for Mapping Buildings Damaged by a Wide Range of Disasters Based on Satellite Imagery. *Remote Sensing*, 12(10), Article 10. <https://doi.org/10.3390/rs12101670>
172. Shen, Y., Zhu, S., Yang, T., Chen, C., Pan, D., Chen, J., Xiao, L., & Du, Q. (2022). BDANet: Multiscale Convolutional Neural Network With Cross-Directional Attention for Building Damage Assessment From Satellite Images. *IEEE Trans. Geosci. Remote Sensing*, 60, 1–14. <https://doi.org/10.1109/TGRS.2021.3080580>
173. Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
174. *Spatial Filters - Laplacian/Laplacian of Gaussian*. (n.d.). <https://homepages.inf.ed.ac.uk/rbf/HIPR2/log.htm>
175. Storvik, G., Fjortoft, R., & Solberg, A. H. S. (2005). A bayesian approach to classification of multiresolution remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), 539–547. <https://doi.org/10.1109/TGRS.2004.841395>

176. Su, J., Bai, Y., Wang, X., Lu, D., Zhao, B., Yang, H., Mas, E., & Koshimura, S. (2020). Technical Solution Discussion for Key Challenges of Operational Convolutional Neural Network-Based Building-Damage Assessment from Satellite Imagery: Perspective from Benchmark xBD Dataset. *Remote Sensing*, 12(22), Article 22. <https://doi.org/10.3390/rs12223808>
177. Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proceedings of the 36th International Conference on Machine Learning*, 6105–6114. <https://proceedings.mlr.press/v97/tan19a.html>
178. Taskin Kaya, G., Musaoglu, N., & Ersoy, O. K. (2011). Damage assessment of 2010 Haiti earthquake with post-earthquake satellite image by support vector selection and adaptation. *Photogrammetric Engineering & Remote Sensing*, 77(10), 1025–1035.
179. Thakur, R., & Panse, P. (2022). Classification Performance of Land Use from Multispectral Remote Sensing Images using Decision Tree, K-Nearest Neighbor, Random Forest and Support Vector Machine Using EuroSAT Data. *International Journal of Intelligent Systems and Applications in Engineering*, 10(1s), 67–77. <https://ijisae.org/index.php/IJISAE/article/view/2238>
180. Tilon, S., Nex, F., Kerle, N., & Vosselman, G. (2020). Post-Disaster Building Damage Detection from Earth Observation Imagery Using Unsupervised and Transferable Anomaly Detecting Generative Adversarial Networks. *Remote Sensing*, 12(24), Article 24. <https://doi.org/10.3390/rs12244193>
181. Ting, K. M., & Witten, I. H. (1999). Issues in Stacked Generalization. *Journal of Artificial Intelligence Research*, 10, 271–289. <https://doi.org/10.1613/jair.594>
182. Toutin, T. (2006). Elevation Modeling from Satellite Data. In *Encyclopedia of Analytical Chemistry*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9780470027318.a2305>
183. Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150.
184. UNOSAT. (n.d.). <https://unosat.org/products/3371>
185. van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86), 2579–2605. <http://jmlr.org/papers/v9/vandermaaten08a.html>

186. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (No. arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>
187. Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J. P., Veroustraete, F., Clevers, J. G. P. W., & Moreno, J. (2015). Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 108, 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>
188. Voigt, S., Giulio-Tonolo, F., Lyons, J., Kučera, J., Jones, B., Schneiderhan, T., Platzeck, G., Kaku, K., Hazarika, M. K., Czarán, L., Li, S., Pedersen, W., James, G. K., Proy, C., Muthike, D. M., Bequignon, J., & Guha-Sapir, D. (2016). Global trends in satellite-based emergency mapping. *Science*, 353(6296), 247–252. <https://doi.org/10.1126/science.aad8728>
189. Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite Image Analysis for Disaster and Crisis-Management Support. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6), 1520–1528. <https://doi.org/10.1109/TGRS.2007.895830>
190. Voigt, S., Schneiderhan, T., Twele, A., Gähler, M., Stein, E., & Mehl, H. (2011). Rapid Damage Assessment and Situation Mapping: Learning from the 2010 Haiti Earthquake. *Photogrammetric Engineering & Remote Sensing*, 77(9), 923–931. <https://doi.org/10.14358/PERS.77.9.923>
191. Wang, X., Tan, K., Du, Q., Chen, Y., & Du, P. (2020). CVA 2 E: a conditional variational autoencoder with an adversarial training process for hyperspectral imagery classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(8), 5676–5692.
192. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. <https://doi.org/10.1109/TIP.2003.819861>
193. Weber, E., & Kané, H. (2020). *Building Disaster Damage Assessment in Satellite Imagery with Multi-Temporal Fusion*. arXiv. <https://doi.org/10.48550/ARXIV.2004.05525>
194. Whiteside, T. G., Boggs, G. S., & Maier, S. W. (2011). Comparing object-based and pixel-based classifications for mapping savannas. *International Journal of Applied Earth Observation and Geoinformation*, 13(6), 884–893. <https://doi.org/10.1016/j.jag.2011.06.008>

195. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79–82. <https://www.jstor.org/stable/24869236>
196. Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5(2), 241–259. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1)
197. Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). *CBAM: Convolutional Block Attention Module*. 3–19. https://openaccess.thecvf.com/content_ECCV_2018/html/Sanghyun_Woo_Convolutional_Block_Attention_ECCV_2018_paper.html
198. Wu, C., Zhang, F., Xia, J., Xu, Y., Li, G., Xie, J., Du, Z., & Liu, R. (2021). Building Damage Detection Using U-Net with Attention Mechanism from Pre- and Post-Disaster Remote Sensing Datasets. *Remote Sensing*, 13(5), Article 5. <https://doi.org/10.3390/rs13050905>
199. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>
200. Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., ... Zhu, Z. (2019). Current status of Landsat program, science, and applications. *Remote Sensing of Environment*, 225, 127–147. <https://doi.org/10.1016/j.rse.2019.02.015>
201. Xie, H., Hu, X., Jiang, H., & Zhang, J. (2022). BSSNet: Building subclass segmentation from satellite images using boundary guidance and contrastive learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 7700–7711. <https://doi.org/10.1109/JSTARS.2022.3202524>
202. Xu, J. Z., Lu, W., Li, Z., Khaitan, P., & Zaytseva, V. (2019). *Building Damage Detection in Satellite Imagery Using Convolutional Neural Networks*. arXiv. <https://doi.org/10.48550/ARXIV.1910.06444>
203. xView2. (n.d.). <https://xview2.org/>

204. Yates, M., Hart, G., Houghton, R., Torres Torres, M., & Pound, M. (2022). Evaluation of synthetic aerial imagery using unconditional generative adversarial networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, 231–251. <https://doi.org/10.1016/j.isprsjprs.2022.06.010>
205. Zhang, T., Tang, H., Ding, Y., Li, P., Ji, C., & Xu, P. (2021). FSRSS-Net: High-Resolution Mapping of Buildings from Middle-Resolution Satellite Images Using a Super-Resolution Semantic Segmentation Network. *Remote Sensing*, 13(12), 2290. <https://doi.org/10.3390/rs13122290>
206. Zhao, Y., Zhang, X., Feng, W., & Xu, J. (2022). Deep Learning Classification by ResNet-18 Based on the Real Spectral Dataset from Multispectral Remote Sensing Images. *Remote Sensing*, 14(19), 4883. <https://doi.org/10.3390/rs14194883>
207. Zheng, Z., Zhong, Y., Wang, J., Ma, A., & Zhang, L. (2021). Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters. *Remote Sensing of Environment*, 265, 112636. <https://doi.org/10.1016/j.rse.2021.112636>
208. Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). *UNet++: A Nested U-Net Architecture for Medical Image Segmentation* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1807.10165>
209. Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms* (1st ed.). Chapman & Hall/CRC.
210. Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2242–2251. <https://doi.org/10.1109/ICCV.2017.244>
211. Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>

12. Disclaimer

The preparation of this thesis involved a structured workflow combining extensive manual reading, systematic literature management, and selective use of AI-based tools to support efficiency and consistency. The primary foundation of the work rests on the author's direct engagement with the scientific literature, critical assessment of sources, and manual drafting of summaries and arguments. Artificial intelligence was employed as an auxiliary tool, never as a replacement for scholarly judgment.

Approximately 165 scientific papers relevant to the topic of battlefield damage assessment of urban structures were initially identified via Google Scholar searches. These publications were imported into Zotero, where each entry was supplemented with a full-text PDF and an associated structured note. Notes followed a standardized template including metadata, extracted data fields (keywords, models, datasets, satellites, methods), and a layered summary (context, methodology, results, conclusions). Each referenced paper was read and evaluated by the author; summaries were drafted manually, supported where appropriate by Adobe AI-generated abstracts, which were always verified and corrected against the original texts.

Markdown notes were processed through a custom Python pipeline that normalized metadata, extracted datasets and satellites, created keywords, and ensured interoperability with Obsidian for relational graph visualization. AI support was used here primarily Regex in the Python pipeline. For technical tasks such as metadata normalization, DOI resolution, and error checking of bibliographic entries Python scripts were used. For the identification of highly cited primary works, ChatGPT 4.0/5 (Version of July / August 2025) was employed in conjunction with relational graph analyses; candidate references were cross-verified manually via Google Scholar citation counts before inclusion.

Draft chapters of the appendices and literature overview were written manually, based on the structured notes and critical reading. ChatGPT was subsequently used in an editorial capacity to propose rephrasings, ensure consistency in academic style, and generate alternative formulations. The AI-generated text was not adopted verbatim; instead, it was carefully fact-checked, compared with the cited sources, and revised by the author. Similarly, scite.ai was used as a supplementary tool to verify the context of citations within the scientific discourse. All final references were inserted manually via Zotero to ensure accuracy and compliance with citation standards.

This integrated workflow allowed for efficient processing of a large body of literature while ensuring academic integrity. The critical reading, interpretation, selection of relevant findings, structuring of arguments, and final phrasing of all content were conducted by the author. AI tools contributed to metadata processing, drafting of preliminary summaries, language refinement, and bibliometric checks, but the responsibility for accuracy, interpretation, and argumentation rests fully with the author.