GIScience & Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgrs20

Classification of drainage crossings on high-resolution digital elevation models: A deep learning approach

Di Wu, Ruopu Li, Banafsheh Rekabdar, Claire Talbert, Michael Edidem & Guangxing Wang

To cite this article: Di Wu, Ruopu Li, Banafsheh Rekabdar, Claire Talbert, Michael Edidem & Guangxing Wang (2023) Classification of drainage crossings on high-resolution digital elevation models: A deep learning approach, GIScience & Remote Sensing, 60:1, 2230706, DOI: 10.1080/15481603.2023.2230706

To link to this article: https://doi.org/10.1080/15481603.2023.2230706

Published online: 03 Jul 2023.

GISCIENCE & REMOTE SENSING
2023, VOL. 60, NO. 1, 2230706
https://doi.org/10.1080/15481603.2023.2230706

Taylor & Francis
Taylor & Francis Group

Classification of drainage crossings on high-resolution digital elevation models: A deep learning approach

Di Wu[a], Ruopu Li[a], Banafsheh Rekabdar[b], Claire Talbert[a], Michael Edidem[a] and Guangxing Wang[a]

[a]School of Earth Systems and Sustainability, Southern Illinois University, Carbondale, IL, USA; [b]Department of Computer Science, Portland State University, Portland, OR, USA

ABSTRACT

High-Resolution Digital Elevation Models (HRDEMs) have been used to delineate fine-scale hydrographic features in landscapes with relatively level topography. However, artificial flow barriers associated with roads are known to cause incorrect modeled flowlines, because these barriers substantially increase the terrain elevation and often terminate flowlines. A common practice is to breach the elevation of roads near drainage crossing locations, which, however, are often unavailable. Thus, developing a reliable drainage crossing dataset is essential to improve the HRDEMs for hydrographic delineation. The purpose of this research is to develop deep learning models for classifying the images that contain the locations of flow barriers. Based on HRDEMs and aerial orthophotos, different Convolutional Neural Network (CNN) models were trained and compared to assess their effectiveness in image classification in four different watersheds across the U.S. Midwest. Our results show that most deep learning models can consistently achieve over 90% accuracies. The CNN model with HRDEMs as the sole input feature was found to be the best-fit one. The addition of aerial orthophotos and their derived spectral indices is insignificant to or even worsens the model's accuracy. The selected best-fit model exhibits excellent transferability over different geographic contexts. This work can be applied to improve elevation-derived hydrography mapping at fine spatial scales.

ARTICLE HISTORY
Received 3 October 2022
Accepted 22 June 2023

KEYWORDS
Deep learning; Convolutional neural network; LiDAR; HRDEM; Hydrography

1. Introduction

Hydrologic connectivity is critical for environmental management issues such as overland nutrient transport and ecological conservation (Good, Noone, and Bowen 2015; Pringle 2001; Stieglitz et al. 2003). For this, Digital Elevation Models (DEMs) have been widely utilized to represent topographic structure and simulate hydrologic connectivity (Callow, Van Niel, and Boggs 2007; Habtezion, Tahmasebi Nasab, and Chu 2016). Conventional medium-resolution DEMs are unsuitable for resolving subtle terrain variations in a flat topography and mapping hydrographic features in fine scales (Liu, Peterson, and Zhang 2005; Regnauld and Mackaness 2006). In contrast, High-resolution DEMs (HRDEMs) products, mostly generated from Light Detection and Ranging (LiDAR), allow highly accurate representation of terrain owing to their fine spatial resolution and high vertical accuracy. However, accurate delineation of hydrographic features using HRDEMs exhibits a challenge arising from unrepresented features in HRDEMs. Numerous studies have shown that hydrological delineation using HRDEMs is susceptible to flow barriers such as roads, which function as "digital dams" (Duke et al. 2003; Li et al. 2013; Sofia, Fontana, and Tarolli 2014). Because only the elevation of the road surface rather than its underneath channel at drainage structures such as culverts and bridges are typically represented on the HRDEMs, the simulated drainage flowlines often terminate before the roads or cross the roads at incorrect locations (Poppenga and Worstell 2016).

Research has shown that incorporating drainage crossing locations can effectively breach flow barriers and improve the spatial accuracy of HRDEM-derived watershed boundaries and flowlines (Aristizabal et al. 2018; Bhadra et al. 2021; Li et al. 2013; Lindsay and Dhun 2015). Breaching roads at those locations results in a preprocessed, "hydrologic" version of HRDEMs. However, the drainage crossing location dataset is either unavailable or only available with variable quality (Poppenga and Worstell 2016) in most areas. To identify the locations of flow barriers, current efforts

CONTACT Ruopu Li ruopu.li@siu.edu

have been directed toward manual on-screen digitization (Shore et al. 2013) and field survey (Wang et al. 2011), which are labor-intensive and costly especially over large geographic areas. Thus, it is imperative to develop a cost-effective method to efficiently identify flow barriers for improving HRDEM preprocessing for hydrographic mapping.

As an emerging technique in geospatial information science, deep learning has shown its superiority to solve complex image classification and object detection problems (Hu et al. 2020; Krizhevsky, Sutskever, and Hinton 2012). It uses multiple layers to progressively extract higher-level features from raw inputs, which could be adaptively used as "knowledge" to create outputs (Sinha, Pandey, and Pattnaik 2018). For example, Li et al. (2017) used deep learning to support automatic craters detection and classification, providing important information for estimating relative regional ages. Ye et al. (2019) proposed a deep learning framework with constraints to detect landslides on hyperspectral images, resulting in a higher overall accuracy than traditional classification methods. Li et al. (2020) employed a deep learning strategy to detect landforms in the landscape of Loess Plateau using integrated data sources of digital elevation model and imagery. Deep learning has been found capable of producing results comparable to human performance from large samples in a supervised, semi-supervised, or unsupervised manner. In particular, Convolutional Neural Networks (CNN), as a supervised deep learning architecture, has its eminence in automatic features selection, extraction, and generalization to avoid overfitting (Tang et al. 2018; Widyantoko et al. 2021). The rapidly developing computing infrastructure and tools (e.g. TensorFlow and Keras) also make CNN a viable option to address complex classification tasks based on big training datasets (Li et al. 2017). In hydrography mapping, Xu et al. (2021) adopted an attention U-net and high-accuracy LiDAR data for developing detailed streamline detection, resulting in an improved performance over traditional method learning methods such as ANN and SVM. To generate more accurate drainage networks, Stanislawski, Brockmeyer, and Shavers (2018) utilized a set of existing roads and stream valleys data to train a model with advanced CNN architecture, then using the trained model to extract all road and stream valley features from HRDEMs. The flow barriers from road embankments were recognized as an important issue to be addressed for future research. CNN has also been applied to assess the debris-related blockage conditions of drainage culverts (Iqbal et al. 2022), which developed CNN algorithms to automate the process of classifying visual blockage images captured by HD cameras. Talafha et al. (2021) used several advanced CNN techniques to distinguish images with drainage crossings (i.e. bridges and culverts) in a small watershed in the northeast Nebraska based on both DEMs and National Agriculture Imagery Program (NAIP) aerial orthophotos. However, it is unclear if those developed CNN models are transferrable to broader geographic areas.

All these studies support that deep learning techniques can be leveraged to facilitate the identification of drainage crossing locations for improving the modeled hydrography. However, the modeled channel continuity depends on the availability of high-accuracy supplemental training data (e.g. road and flowline data), which are often unavailable at fine scales. The purpose of this study is to develop a deep learning model that can classify the images with under-road drainage crossings and evaluate the performance and transferability of the model in different geographic areas. Drainage crossings were focused on those locations where natural streams or artificial canals pass roads through underpass structures such as bridges or culverts. We expect that the knowledge gained in this study can be used to guide the application of deep learning models for improving hydrologic connection in broader geographic areas.

2. Study area

The experiments were conducted in four study areas, including West Fork Big Blue Watershed, Nebraska, Vermilion River Watershed, Illinois, Maple River Watershed, North Dakota, and Sacramento-Stone Corral Watershed, California (Figure 1). The landscapes of these areas are dominated by intensive agriculture and relatively level topography. Dense road networks over these areas complicate the flow patterns and segment hydrologic features represented on HRDEMs, resulting in a critical need for breaching drainage barriers at culverts and bridges. West Fork Big Blue, Nebraska lies on gently undulating loess plain, descending from its west of around

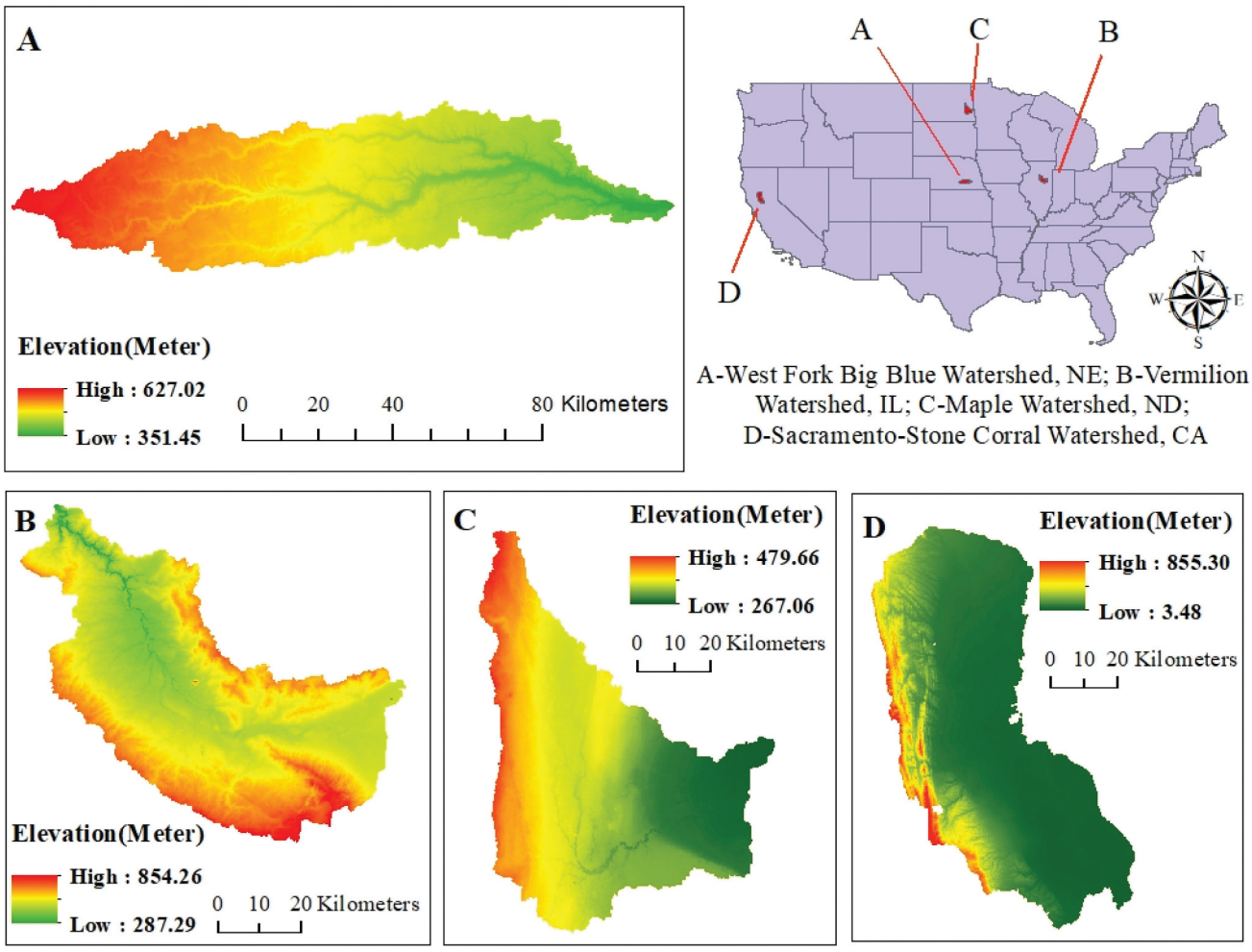

Figure 1. Topography and locations of four study areas.

626 m to its east of around 410 m with the total area of 3,471 km^2. Land use within the watershed is primarily agriculture with over 80% of the land devoted to row crops. The hydrology of the region is dominated by a poorly developed drainage system with many depressional wetlands (Stutheit et al. 2004). The Vermilion River Watershed is located in North-Central Illinois, encompassing more than 3,453 km^2 that drain northwesterly into the Illinois River. Its elevation ranges from 134 to 260 m. About 95% of the watershed is dominated by cultivated crops (88%) and pasture (7%). Understanding the routing and eroding effect of surface runoff is a top priority for local natural resources conservation. The Maple River Watershed is mostly within the ecoregion of Northern Glaciated Plains of North Dakota, which features flat to undulating landforms and shallow river valleys. The watershed has a total of 4,089 km^2, and generally descends toward the east with elevations ranging from about 480 to 267 m. Approximately 93% of the area is dominated by row crops and pasture owing to its deep and fertile soils. The Sacramento-Stone Corral watershed is located in northern Central Valley, California, one of the most productive U.S. agricultural regions. Within an area of 4,880 km^2, the terrain elevation ranges from 857 to 5 m, descending from its mountainous west to the east plain. Most of the watershed is dominated by low-lying plains, which support many agricultural commodities ranging from rice to row crops.

3. Methods

3.1 Datasets

The sources and specification of HRDEMs and 4-band Digital Orthophotos used in this study are listed in Table 1. Compared with conventional DEMs, LiDAR-

Table 1. Data sources of LIDAR-derived HRDEMs and aerial orthophotos.

Data	Locations	Sources	Spatial resolution	Vertical accuracy RMSE
DEM	West Fork Big Blue Watershed, Nebraska	Nebraska Department of Natural Resource	1.0m	0.185m
	Vermilion River Watershed, Illinois	Illinois Geospatial Data Clearinghouse	0.30m	1.19ft (~0.36m)
	Maple River Watershed, North Dakota	North Dakota GIS Hub Data Portal	2.0ft (~0.61m)	0.15m
	Sacramento-Stone Corral Watershed, California	USGS	1.0m	0.196m
Aerial Orthophotos	Four Watersheds	USGS National Agriculture Imagery Program (NAIP)	1.0m	-

derived HRDEMs show remarkable improvement in representing subtle topographic details of land surface. The National Agriculture Imagery Program (NAIP) under the USDA Farm Service Agency (FSA) acquired color infrared aerial orthophotos at a resolution of 1-meter ground sample distance.

3.2 Input features

Unique topographic and spectral signatures at locations of flow barriers such as microtopographic patterns and soil erosion near drainage crossing locations are commonly used as indicators to guide manual digitization of these structures on HRDEMs and aerial orthophotos (Gelder, Zhou, and Yu 2015; Poppenga et al. 2010). In this paper, we tested four different combinations of features as listed in Table 2. In addition to different band combinations, we also derived new features to enhance the contrasts of land covers near flow barriers, including Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI). NDVI has been well known to effectively distinguish vegetation and bare soil that are often found near drainage crossings, while NDWI can indicate differential moisture contents caused by vegetation and topographic variations. To prepare the datasets, we resampled all HRDEMs from four study areas to 1-meter resolution and manually digitized drainage crossing locations. Using a square bounding box of 100 m near each drainage crossing location, five-band samples with the identical size (100 by 100-pixel) were clipped as True samples (Figure 2a). We also randomly generated equal amounts of False samples without a drainage crossing along the roads (Figure 2b). The sample size for each watershed is shown in Table 3.

3.3 Model development

We selected the CNN deep learning architecture due to its excellent flexibility and effectiveness in complex data intensive tasks. The combination that includes convolution, activation, and pooling layers can be regarded as a convolution unit of CNN (Li et al. 2017). In our study, depending on the complexities in the images, this unit could be repeated, resulting in a "deep stacking" to capture low-level details, albeit at the cost of increasing computational requirement.

Figure 3 demonstrates the architecture of our CNN models, which contains four convolution layers, one flatten layer, and two fully connected layers. Convolution layers automatically extract the feature map that is the sum of dot product of all elements in sub-regions and the kernels (Hussain, Bird, and Faria 2018; Li et al. 2017) by applying a group of filters (a.k.a., convolutional kernels) with the same predefined size to image sub-regions. For feature inputs, each convolutional layer contains a different set of image features. In our model, 3×3 and 5×5 filters were used, respectively, in different convolution layers. Rectified Linear Unit (ReLU) was also utilized in our model and applied to the output feature map from a convolution layer to change all negative values

Table 2. Input features for flow barriers identification.

Input features	Applied methods	References
HRDEM (1 band) HRDEM, NAIP Orthophotos (R, G, B, NIR) (5 bands)	To remove image dependency on lighting geometry and illumination color, the data were normalized to a scale of (0,1). $X_{scaler} = \frac{X - X_{min}}{X_{max} - X_{min}}$ $X_{scaled} = X_{scaler} * (max - min) + min$ $max = 1, min = 0$	(Finlayson, Schiele, and Crowley 1998; Patro and Sahu 2015; Saranya and Manikandan 2013).
HRDEM, NDVI (2 bands)	$NDVI = \frac{NIR - Red}{NIR + Red}$	(Brown 2015).
HRDEM, NDWI (2 bands)	$NDWI = \frac{GREEN - NIR}{GREEN + NIR}$	(McFeeters 1996).

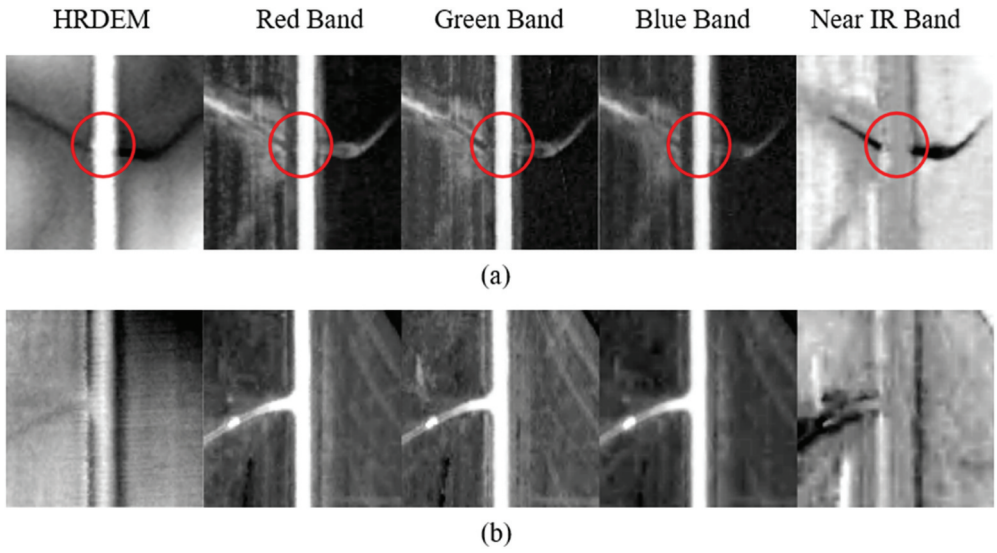

Figure 2. Examples of 5-band samples, where (a) is the true sample and (b) is false sample. The red circles point to the locations of drainage crossings.

Table 3. Training sample sizes for four watersheds located in California, Illinois, Nebraska, and North Dakota.

Locations	*True sample (Label 1, with culvert)*	*False sample (Label 0, without culvert)*	*Total*
Nebraska	2022	2022	4044
Illinois	1011	1011	2022
North Dakota	613	613	1226
California	2388	2388	4776

within the feature map to zero. As a non-linear activation function, it can improve the non-linear properties of the model (Agarap 2018).

Batch normalization was implemented between convolution operations and average pooling layers in the model. Batch normalization is the process of applying normalization among layers in a neural

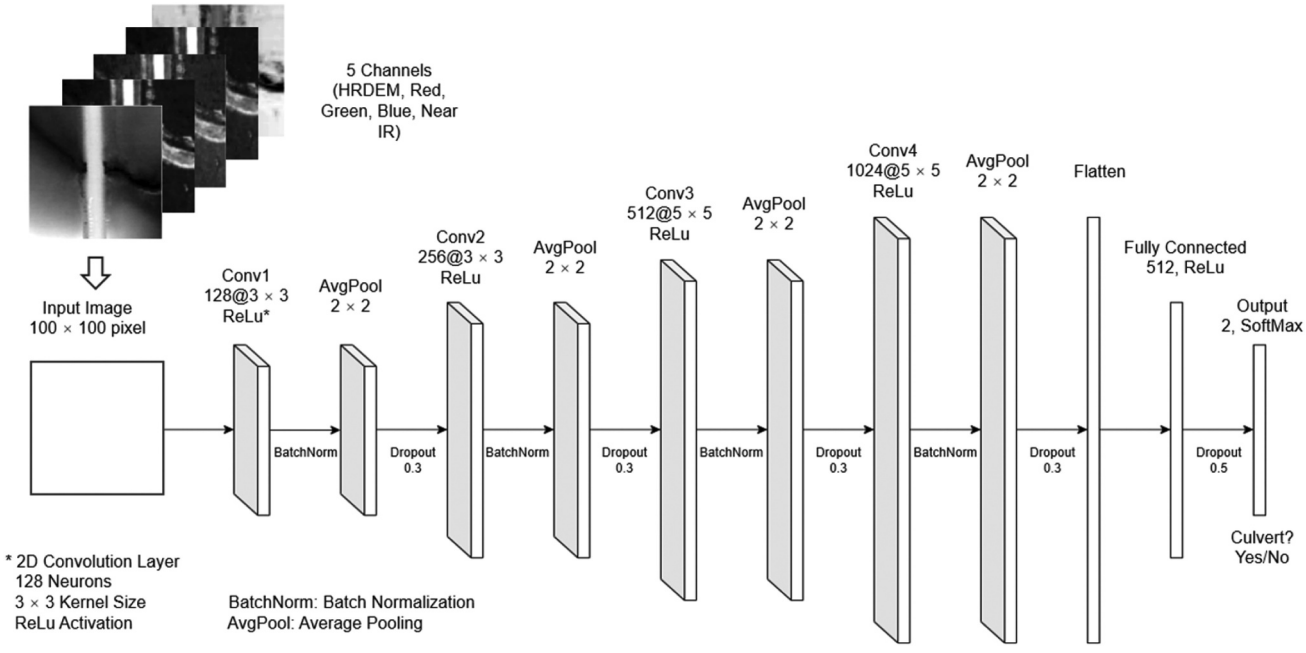

Figure 3. Architecture of the CNN-based deep learning model for this study.

network, which helps stabilize the learning process, reduce overfitting through regularization, and speeds up training substantially. Then, the feature map is further processed by an average pooling layer for reducing the spatial size of the convolved features. In order to "pool" the feature map, the feature map has been divided into non-overlapping smaller regions by a 2×2 kernel and the average of values in the regions are returned to represent each sub-region. The application of pooling layer can decrease the computational power requirement for data processing through dimensionality reduction. It also helps maintain the training process effectively by retaining the most important information such as the dominant features that are rotational and positional invariant. A fully connected layer (a.k.a., the classifier) operates on a flattened input where each input is connected to all neurons. Based on the errors in the prediction of SoftMax function, backpropagation is applied to every iteration of training to adjust the bias values and weights.

To train a CNN model, data should be divided into training, validation, and test sets. The validation set is used for the purposes of tuning hyperparameters of the model (Barry-Straume et al. 2018). To avoid the randomness from estimates produced by imbalanced data and overfitting, K-fold cross-validation was used (Berrar 2019; Koul, Becchio, and Cavallo 2018). In a sparse dataset, splitting the data only once could yield unstable estimates because of unsuitable train-test split selected and inadequate amount of observational data (Ziggah et al. 2019). The results from a single train-and-test experiment may not provide enough evidence of the generalization ability of model. K-fold cross-validation is one way to combat these defects. In this technique, the whole dataset is divided into k parts of equal size and each partition is called a fold. One of these (k) folds is chosen as validation dataset while the others (k-1) are used for the training model. The process is repeated until each fold has been used as the validation set and we will get the mean of k number of performance estimates. Different hyperparameters of the model will undergo a cross-validation test and hyperparameters with best performance will be selected for different input combination. The workflow is shown in Figure 4.

For each watershed, we randomly selected 80% among the dataset as training dataset, and 20% as the test dataset. In the training dataset we apply 4-fold cross-validation, which ensures that there are 20% of whole data as the validation dataset (training data: validated data: test data = 60%: 20%: 20%) (Kumar 2020). Adam optimizer, an adaptive learning rate optimization algorithm was utilized through a learning rate of 0.001 (Talafha et al. 2021). Different batch sizes (e.g. 8, 16, and 32) are examined

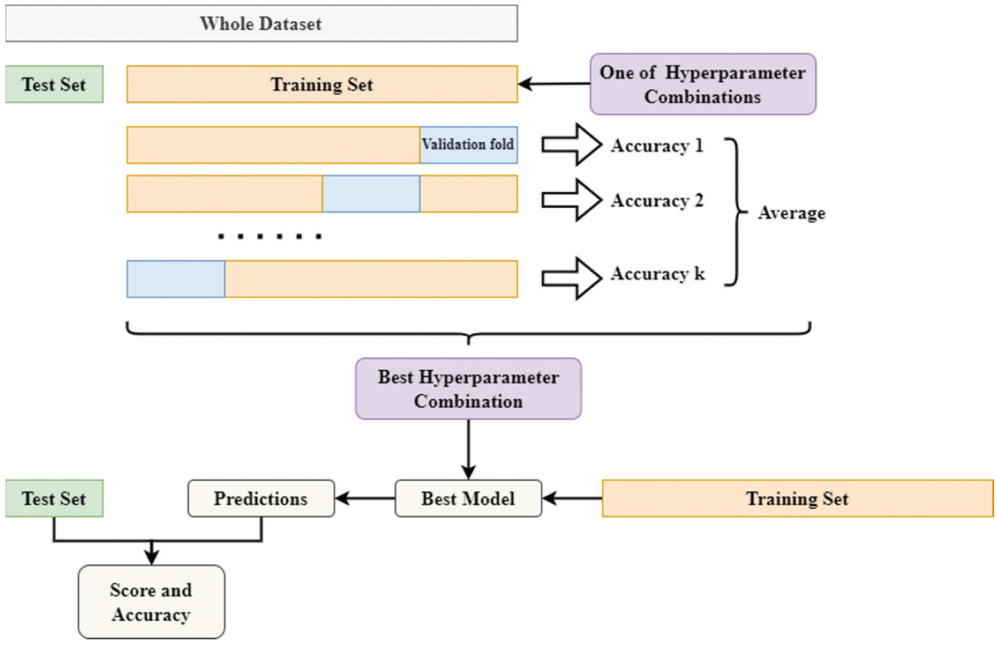

Figure 4. The workflow of the CNN model development that involves model training, testing, and validation.

and the results will be compared to find the optimum. Using the data samples from West Fork Big Blue Watershed, Nebraska, we trained the best-fit CNN model, which was then used to test its transferability of classifying drainage crossing locations in the other three watersheds.

3.4 Model evaluation

To evaluate the performance of CNN models, we separate the entire image samples into training, validation, and test datasets. The validation process is performed on the validation dataset after every epoch, which gives information that helps tune the model's hyperparameters and prevents model from overfitting. The test dataset is used to test the model after completing the training, providing an unbiased final model performance metric in terms of accuracy. Since truth labels are encoded with integers, sparse categorical cross entropy method was used to calculate the loss of model (Eq.1). Entropy of a random variable presents the level of uncertainty inherent in the possible outcome of variables. Cross entropy loss, also called logarithmic loss, measures the performance of a classification model whose output is a probability value between 0 and 1. It compares each predicted class probability with the actual class. The loss would increase as the predicted probability diverges from the actual label. A binary confusion matrix was applied to represent accuracy where positive samples represent those samples containing flow barriers while negative samples are those without flow barriers. A perfect model is expected to only produce "true positives" (TP) and "true negatives" (TN), and not include any "false positives" (FP) or "false negatives" (FN). The accuracy was calculated by Eq.2.

$$Loss = -\sum_{i=1}^{n} y_i \log(p_i) \tag{1}$$

Where n is number of categories, y_i is the truth label and $p_i$is the Softmax probability for the *ith* class.

$$Acc = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

To estimate the contribution of each input features to the classification outcome and the relationships among them, a multicollinearity analysis was conducted. The multicollinearity is the linear relationship between two or more variables, where variables are all influencing each other and not independent (Chen et al. 2019). Since a high multicollinearity increases the difficulties for models to establish the relationships between inputs and outputs, those input features that have strong correlation with each other should be reduced. Tolerance (TOL) (Eq.3) and the variance inflation factor (VIF) (Eq.4) are commonly used indexes for multicollinearity diagnoses. Theoretically, VIF > 10 or TOL < 0.1 is regarded as a threshold for a multicollinearity problem (Tang et al. 2019; Yu, Jiang, and Land 2015).

$$TOL_i = 1 - R_i^2 \tag{3}$$

where R_i^2 is the coefficient of determination from the regression of factor i on all the other factors.

$$VIF_i = \frac{1}{TOL_i} \tag{4}$$

To assume the performance of the CNN model, we compare it with Support Vector Machine (SVM). SVM is a supervised learning model for classification problems. SVM exhibits balanced predictive performance and complexity, even if the sample sizes are limited (Pisner and Schnyer 2020). For classification, if the data were linearly separable, the SVM model would be used as a non-probabilistic binary linear classifier. For nonlinear data, the SVM uses nonlinear kernel functions for transforming input data to a high-dimensional feature space, in which the input data become more separable compared to the original input space (Suthaharan 2016). A SVM with linear kernel was applied to the West Fork Big Blue Watershed in Nebraska. The train-test dataset split was 0.8 to 0.2.

4. Results

4.1 Model performance

Different combinations of model parameters and input features were used to construct CNN models for West Fork Big Blue Watershed, Nebraska. It was found that the model consistently achieved high accuracies (>90%) when the number of epochs was set as 100. CNN models reached over 99% training accuracies regardless of the combination of feature inputs and batch sizes (Table 4). According to the validation accuracies, the optimal batch size for the model with inputs HRDEM, HRDEM and NAIP aerial orthophotos, HRDEM and NDVI, HRDEM and NDWI are

Table 4. Results of k-fold cross validation (k = 4). The total sample size is 4044 (50% False, 50% True), where train samples are 2426 (60%), validate sample are 809 (20%), and test samples are 809 (20%). The model is with optimizer adam, learning rate 0.001, and Epoch 100.

		Training		Validation	
Feature inputs	Batch size	Loss	Accuracy	Loss	Accuracy
HRDEM	8	0.0159	0.9968	0.3745	0.9607
	16	0.0117	0.9970	0.3826	0.9638
	32	0.0094	0.9975	0.2998	0.9601
HRDEM+NAIP Aerial Orthophoto	8	0.0193	0.9965	0.2806	0.9713
	16	0.0208	0.9965	0.2545	0.9607
	32	0.0053	0.9988	0.3312	0.9592
HRDEM+NDVI	8	0.0108	0.9981	0.3372	0.9710
	16	0.0121	0.9965	0.2997	0.9682
	32	0.0084	0.9974	0.3141	0.9608
HRDEM+NDWI	8	0.0087	0.9972	0.3884	0.9737
	16	0.0219	0.9955	0.1546	0.9743
	32	0.0098	0.9973	0.2024	0.9734

16, 8, 8, and 16, respectively. The test accuracies of best-fit models with different combinations of input features are in Table 5, which shows that the model with input HRDEM and batch size 16 has the best testing accuracy (93.33%). Among all combinations of input features, HRDEM appears to be a dominant feature to accurately classify images with under-road drainage crossings being present. The results suggest that most CNN models can properly perform the classification tasks. The CNN model with HRDEMs as the sole feature allows the best model-fit while maintaining relatively a lower computation cost. The addition of aerial imagery and derived spectral indices show little improvement on the model accuracy. To assess the contributions of aerial orthophoto bands to model accuracy, we trained CNN models with four aerial orthophoto bands, NDVI, and NDWI (with 16 as the batch size). The outcomes shown in Table 6 and Figure 5 suggest that NAIP aerial orthophotos and their derived spectral indices can be interpedently used as features to classify drainage crossing locations albeit with relatively lower accuracies. Both HRDEMs and aerial orthophotos appear to contribute similarly to the classification of images with and without drainage crossings.

4.2 Model transferability

Based on the results in Tables 4 and 5, we applied the best-fit CNN model (Input: HRDEM; Learning rate: 0.001; Epoch: 100; Batch size: 16) to the other three watersheds in Illinois, North Dakota, and California, respectively, to test the selected model's transferability in broader geographic contexts. As the results

Table 5. Classification results of CNN model training, validation, and testing with the combination of HRDEMs and NAIP-related spectral features. The total sample size is 4044 (50% False, 50% True), where train samples are 2,426 (60%), validate sample are 809 (20%), and test samples are 809 (20%).

		Training		Testing	
Feature inputs	Batch size	Loss	Accuracy	Loss	Accuracy
HRDEM	16	0.0065	0.9969	0.7398	0.9333
HRDEM +NAIP Aerial Orthophotos	8	0.0062	0.9981	1.0977	0.9246
HRDEM +NDVI	8	0.0269	0.9947	0.8423	0.9197
HRDEM +NDWI	16	0.0187	0.9966	0.7979	0.9271

Table 6. Classification results of CNN model training, validation, and testing with spectral features from NAIP aerial orthophotos. The total sample size is 4044 (50% False, 50% True), where train samples are 2,426 (60%), validate sample are 809 (20%), and test samples are 809 (20%). The model was set up with batch size 16, optimizer adam, learning rate 0.001, and epoch 100.

	Training		Validation		Testing	
Feature inputs	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
NAIP Aerial Orthophotos	0.0242	0.9922	0.9638	0.8591	1.0914	0.8492
NDVI	0.0353	0.9885	1.1865	0.8022	1.1080	0.8232
NDWI	0.0548	0.9802	0.6269	0.8665	0.8215	0.8381

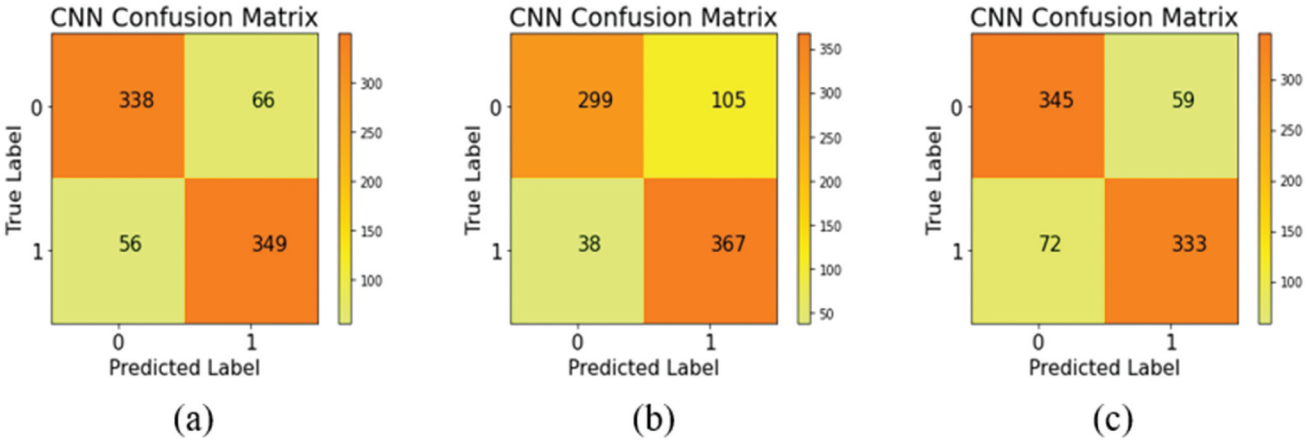

Figure 5. Accuracy matrices for the CNN models with sole input feature from aerial orthophotos and their related spectral indexes respectively. The inputs feature for model (a) is NAIP orthophotos, (b) is NDVI, (c) is NDWI.

Table 7. Classification results in other watersheds. The model has been trained by the HRDEM from Nebraska, with optimizer adam, learning rate 0.001, epoch 100, and batch size 16. The HRDEM from Illinois, North Dakota, and California are applied as test samples, respectively.

Location	Testing loss	Testing accuracy
Vermilion River Watershed, Illinois	0.7995	0.8734
Maple River Watershed, North Dakota	1.1490	0.8320
Sacramento-Stone Corral Watershed, California	2.3472	0.7067

shown in Table 7, the best-fit CNN model developed in Nebraska can be transferable to classify flow barriers in watersheds from the other three. The test resulted in 87.34% accuracy in the Illinois watershed, 83.20% in the North Dakota watershed, and 70.67% in the California watershed. Among three watersheds, California watershed shows the lowest test accuracy. Around 30% of image samples with flow barriers were classified into the wrong category according to results, which are explained in the Discussion.

4.3 Features affecting model accuracies

Table 5 shows that HRDEMs itself allow more accurate classification of images containing flow barriers than the combinations of HRDEMs with NAIP aerial photos and their derived spectral indices. Adding features derived from aerial photos barely improved or even decreased the model accuracy. Upon a close examination of those false positives and true negatives, we found that dense vegetation near roads make drainage-crossing patterns less visible in aerial photos at some locations. To better understand whether HRDEMs and digital numbers (DNs) of aerial photos exhibit interdependencies in terms of image classification, we conducted a multicollinearity test. The multicollinearity analysis is shown in Table 8. As shown in Table 8, TOL values are all larger than 0.1 and VIF values for three models are far less than 10. Therefore, there is no serious multicollinearity being detected between HRDEMs and aerial photo-derived features.

Table 8. Multicollinearity statistics of different input features.

		Collinearity statistics	
Model input	Features	TOL	VIF
HRDEM+NAIP	HRDEM	0.994	1.006
	NAIP-R	0.802	1.248
	NAIR-G	0.787	1.271
	NAIP-B	0.816	1.226
	NAIP-NIR	1.000	1.000
HRDEM+NDVI	HRDEM	1.000	1.000
	NDVI	1.000	1.000
HRDEM+NDWI	HRDEM	1.000	1.000
	NDWI	1.000	1.000

4.4 Baseline model comparison

The classification accuracy of using SVM for the Nebraska watershed is shown in Table 9. Based on the results, SVM cannot properly distinguish the samples without flow barriers. The CNN model turned out to be superior to SVM in terms of classification accuracy.

Table 9. Classification statistics of using SVM as the baseline model. The total sample size is 4044 (50% False, 50% True), where train samples are 3235 (80%), and test samples are 809 (20%).

Class	Precision	Recall	F1-score
False Sample	0.00	0.00	0.00
True Sample	0.50	1.00	0.67
Overall Accuracy		0.5006	

5. Discussion

The best-performing CNN model developed in the Nebraska watershed has HRDEM as the sole input feature with an accuracy of approximately 93%. This level of accuracy is consistent with our earlier tests on a smaller watershed (Talafha et al. 2021). The application of the model to other watersheds confirmed the model's transferability in other geographic regions. It is also interesting to note that testing accuracies for Illinois and North Dakota watersheds are higher than that for the California watershed. This may be explained by the fact that the hydrography in the California watershed exhibits more complex drainage crossing patterns than that in the selected watersheds in Illinois, Nebraska, and North Dakota. This performance variability is also an example of out-of-distribution problems common in deep learning. As exemplified in Figure 6(a-b), natural streams dominate the hydrography in those watersheds spanning North Dakota, Nebraska, and Illinois in North-Central U.S. The common pattern of drainage crossing is a low-lying natural stream channel intersecting with an elevated rural road, which is relatively visible. However, in the California watershed, artificial canals often intersect with roads with more complex crossing patterns, in which artificial structures including canals, canal banks and roads form clustered drainage crossings as shown in Figure 6(c-d). This finding on the model's transferability also suggests that an image training dataset (with flow barriers) collected from broader geographic regions can lead to a robust deep learning model. An image dataset containing image training samples with variable appearances, positions, viewpoints, as well as background clutter and occlusions, is imperative for image analysis algorithms (Deng et al. 2009). It has been found that models being developed based on a dataset with high variability (e.g. ImageNet) usually perform better on computer vision tasks (Kornblith, Shlens, and Le 2019). Thus, collecting flow barriers training data from broader geographic contexts can increase data variability, and thus improve the transferability of the model.

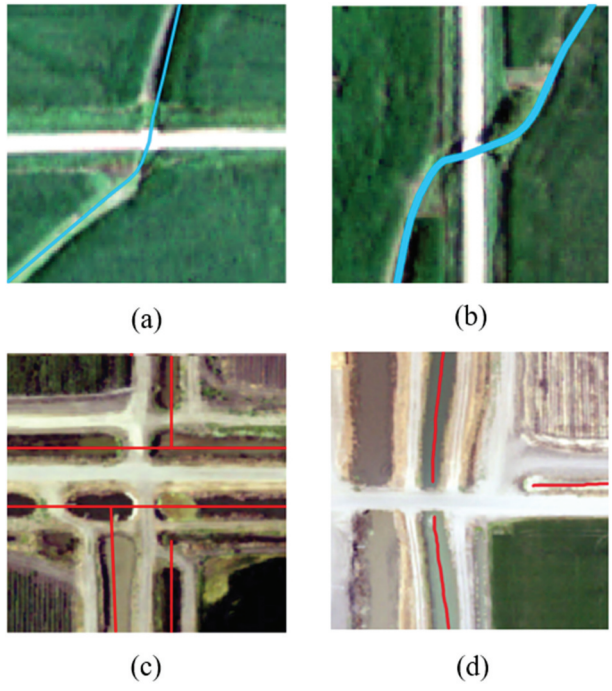

Figure 6. Drainage crossing examples from watersheds in Nebraska and California. The blue lines in (a) and (b) are natural streams in the Nebraska watershed, whereas red lines in (c) and (d) are canal ditches in the California watershed.

Regarding the model's feature selection, the CNN model with HRDEMs as the sole feature turned out the best-fit (~93% accuracy). In particular, the addition of NAIP aerial orthophoto bands and their derived spectral indices was found to be insignificant to or even worsen the model's performance, although they can be used as sole features to achieve 82–85% accuracies (Table 6). As the interdependence between HRDEMs and aerial photo-derived features is low based on the multicollinearity analysis, we argue that two issues may explain the comparability of HRDEMs and NAIP aerial orthophotos in terms of model performance. First, at the locations of flow barriers, the drastic change in elevation between a road and a stream is a typical characteristic of a "digital dam." In NAIP aerial orthophotos, although visible surface characteristics (e.g. erosion, ponding, and dense grass near culverts) can help reveal culverts locations, the unique stream-road crossing pattern appears to be not as explicit as that represented in HRDEMs because streams are often covered by grass, shrubs, and trees near crossing locations. Rafique, Zhu, and Jacobs (2022) found that only HRDEMs-derived topographic features, rather than aerial orthophotos, achieved the success in detecting the features of interest in CNN-based deep learning models. In addition, NAIP aerial orthophotos exhibit data quality issues. The acquisition of NAIP images usually involves multiple flights with different sensors that may last weeks or potentially even months (Davies et al. 2010), resulting in the DN values inconsistent among image tiles associated with atmospheric interference, viewing geometry, illumination, shadows, and even plant phenology. Some images turn out to be blurrier and have lower color contrast than others (Hogland et al. 2018; Maxwell et al. 2017; Yang et al. 2018). Although a conversion from DN values to surface reflectance could help mitigate the issue (Wulder et al. 2019), there is a general absence of sensor information for such a purpose (Zhang, Zimba, and Nzewi 2019).

For future work, we plan to focus on two areas. First, we will further refine the feature selection for the deep learning model. For example, as our tests suggested potential data quality issue associated with NAIP aerial orthophotos, we plan to incorporate the high-resolution satellite images via Google Maps Static API (Google 2023) as a new input feature. Second, we will test different advanced CNN models (e.g. Faster-RCNN and U-net) to locate drainage crossing locations with bounding boxes, which can be used for improving the elevation-derived hydrography research (Stanislawski, Brockmeyer, and Shavers 2018). Feature selection, model calibration, and transferability outcomes of the CNN models in this study are expected to guide the object detection of drainage crossing locations in broad geographic areas.

6. Conclusions

Drainage crossings are hydrological features widely observed in our agricultural landscapes. Identifying their locations is critical for mapping hydrography with high quality and managing environmental issues. In our study, we explored image classification for identifying drainage crossing locations automatically based on high-resolution DEMs. We trained CNN models and labeled each sample images as "true" (with a drainage crossing) or "false" (without a drainage crossing). We also evaluated the performance of the CNN models with different input features and hyperparameters. Furthermore, we applied our best-fit model in a Nebraska watershed to different watersheds in California, Illinois, and North Dakota to test our model's broad geographic transferability. The results suggest that:

(1) Compared with baseline machine learning techniques like SVM, CNN-based deep learning model is a promising tool to accurately classify drainage crossings from our image samples.
(2) Among all models for the study area in Nebraska, the one with a batch size of 16, a learning rate of 0.01, an epoch of 100, and the HRDEM as the sole input feature exhibits the best performance (~93% accuracy).
(3) The best-fit CNN model shows its transferability in different geographic regions, although the classification accuracy appears to link to their hydrography similarity.
(4) The addition of aerial orthophotos and their derived spectral indices was found to be insignificant to or even worsen the best-fit model's performance, which may be explained by the interference of vegetation and data quality issues.

Acknowledgments

This work is supported by the National Science Foundation under Grant #1951741.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

The data that support the findings of this study are available from the corresponding author, RL, upon reasonable request.

References

Agarap, A. F. 2018. "Deep Learning Using Rectified Linear Units (Relu)." arXiv preprint arXiv:1803.08375.

Aristizabal, F., L. E. Grimley, J. Bales, D. Tijerina, T. Flowers, and E. P. Clark. 2018. "National Water Centers Innovators Program Summer Institute Report." *Consortium of Universities for the Advancement of Hydrologic Science*, Inc. Technical Report No 15.

Barry-Straume, J., A. Tschannen, D. W. Engels, and E. Fine. 2018. "An Evaluation of Training Size Impact on Validation Accuracy for Optimized Convolutional Neural Networks." *SMU Data Science Review* 1 (4): 12.

Berrar, D. 2019. "Cross-Validation." *Encyclopedia of Bioinformatics and Computational Biology* 1:542–545.

Bhadra, S., R. Li, D. Wu, G. Wang, and B. Rekabdar. 2021. "Assessing the Roles of Anthropogenic Drainage Structures on Hydrologic Connectivity Using High-Resolution Digital Elevation Models." *Transactions in GIS* 25 (5): 2596–2611. https://doi.org/10.1111/tgis.12832.

Brown, J. 2015. "NDVI, the Foundation for Remote Sensing Phenology." *USGS*. Accessed November 27, 2018. https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology.

Callow, J. N., K. P. Van Niel, and G. S. Boggs. 2007. "How Does Modifying a DEM to Reflect Known Hydrology Affect Subsequent Terrain Analysis?" *Journal of Hydrology* 332 (1–2): 30–39. https://doi.org/10.1016/j.jhydrol.2006.06.020.

Chen, W., X. Zhao, H. Shahabi, A. Shirzadi, K. Khosravi, H. Chai, S. Zhang, et al. 2019. "Spatial Prediction of Landslide Susceptibility by Combining Evidential Belief Function, Logistic Regression and Logistic Model Tree." *Geocarto International* 34 (11): 1177–1201. https://doi.org/10.1080/10106049.2019.1588393 .

Davies, K. W., S. L. Petersen, D. D. Johnson, D. B. Davis, M. D. Madsen, D. L. Zvirzdin, and J. D. Bates. 2010. "Estimating Juniper Cover from National Agriculture Imagery Program (NAIP) Imagery and Evaluating Relationships Between Potential Cover and Environmental Variables." *Rangeland Ecology & Management* 63 (6): 630–637. https://doi.org/10.2111/REM-D-09-00129.1.

Deng, J., W. Dong, R. Socher, L. J. Li, K. Li, and F. Li. 2009, June. "Imagenet: A Large-Scale Hierarchical Image Database." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, https://doi.org/10.1109/CVPR.2009.5206848.

Duke, G. D., S. W. Kienzle, D. L. Johnson, and J. M. Byrne. 2003. "Improving Overland Flow Routing by Incorporating Ancillary Road Data into Digital Elevation Models." *Journal of Spatial Hydrology* 3 (2). Article 2. Available at. https://scholarsarchive.byu.edu/josh/vol3/iss2/2.

Finlayson, G. D., B. Schiele, and J. L. Crowley. 1998, June. *Comprehensive Colour Image Normalization*. *In European conference on computer vision* (pp. 475–490). Springer, Berlin, Heidelberg.

Gelder, B. K., Z. Zhou, and A. Yu. 2015. "Automation of DEM Cutting for Hydrologic/Hydraulic Modeling." *Tech Transfer Summaries* 62 (1): 62–77. https://doi.org/10.1002/aic.15050.

Good, S. P., D. Noone, and G. Bowen. 2015. "Hydrologic Connectivity Constrains Partitioning of Global Terrestrial Water Fluxes." *Science: Advanced Materials and Devices* 349 (6244): 175–177. https://doi.org/10.1126/science.aaa5931.

Google. 2023. Maps Static API in *Google for Developers*. Online access on May 19, 2023 at https://developers.google.com/maps/documentation/maps-static/start.

Habtezion, N., M. Tahmasebi Nasab, and X. Chu. 2016. "How Does DEM Resolution Affect Microtopographic Characteristics, Hydrologic Connectivity, and Modelling of Hydrologic Processes?" *Hydrological Processes* 30 (25): 4870–4892. https://doi.org/10.1002/hyp.10967.

Hogland, J., N. Anderson, J. St Peter, J. Drake, and P. Medley. 2018. "Mapping Forest Characteristics at Fine Resolution Across Large Landscapes of the Southeastern United States Using NAIP Imagery and FIA Field Plot Data." *ISPRS International Journal of Geo-Information* 7 (4): 140. https://doi.org/10.3390/ijgi7040140.

Hu, J., H. Niu, J. Carrasco, B. Lennox, and F. Arvin. 2020. "Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning." *IEEE Transactions on Vehicular Technology* 69 (12): 14413–14423. https://doi.org/10.1109/TVT.2020.3034800.

Hussain, M., J. J. Bird, and D. R. Faria. 2018. September. "A Study on CNN Transfer Learning for Image Classification". In *UK Workshop on Computational Intelligence* 191–202 Cham: Springer. https://doi.org/10.1007/978-3-319-97982-3_16.

Iqbal, U., J. Barthelemy, and P. Perez. 2022. "Prediction of Hydraulic Blockage at Culverts from a Single Image Using Deep Learning." *Neural Computing and Applications* 34 (23): 1–17. https://doi.org/10.1007/s00521-022-07593-8.

Kornblith, S., J. Shlens, and Q. V. Le. 2019. "Do Better Imagenet Models Transfer Better?" In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR)* Long Beach, CA (pp. 2656–2666).

Koul, A., C. Becchio, and A. Cavallo. 2018. "Cross-Validation Approaches for Replicability in Psychology." *Frontiers in Psychology* 9:1117. https://doi.org/10.3389/fpsyg.2018.01117.

Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In *Proceedings of the Advances in Neural Information Processing Systems*, Lake Tahoe, NV (pp. 1097–1105).

Kumar, S. 2020. "Data Splitting Technique to Fit Any Machine Learning Model." *Towards Data Science*, Aprl 30. Accessed September 11. https://towardsdatascience.com/data-splitting-technique-to-fit-any-machine-learning-model-c0d7f3f1c790.

Lindsay, J. B., and K. Dhun. 2015. "Modelling Surface Drainage Patterns in Altered Landscapes Using LiDar." *International Journal of Geographical Information Science* 29 (3): 397–411. https://doi.org/10.1080/13658816.2014.975715.

Li, R., Z. Tang, X. Li, and J. Winter. 2013. "Drainage Structure Datasets and Effects on LiDAR-Derived Surface Flow Modeling." *ISPRS International Journal of Geo-Information* 2 (4): 1136–1152. https://doi.org/10.3390/ijgi2041136.

Liu, X., J. Peterson, and Z. Zhang. 2005, December. "High-Resolution DEM Generated from LiDAR Data for Water Resource Management." In *Proceedings of the International Congress on Modelling and Simulation (MODSIM05)* (pp. 1402–1408). Melbourne, Australia: Modelling and Simulation Society of Australia and New Zealand Inc.

Li, S., L. Xiong, G. Tang, and J. Strobl. 2020. "Deep Learning-Based Approach for Landform Classification from Integrated Data Sources of Digital Elevation Model and Imagery." *Geomorphology* 354:107045. https://doi.org/10.1016/j.geomorph.2020.107045 .

Li, W., B. Zhou, C. Y. Hsu, Y. Li, and F. Ren. 2017, November. "Recognizing Terrain Features on Terrestrial Surface Using a Deep Learning Model: An Example with Crater Detection." In *Proceedings of the 1st Workshop on Artificial Intelligence and Deep Learning for Geographic Knowledge Discovery*, Los Angeles, California (pp. 33–36).

Maxwell, A. E., T. A. Warner, B. C. Vanderbilt, and C. A. Ramezan. 2017. "Land Cover Classification and Feature Extraction from National Agriculture Imagery Program (NAIP) Orthoimagery: A Review." *Photogrammetric Engineering & Remote Sensing* 83 (11): 737–747. https://doi.org/10.14358/PERS.83.10.737.

McFeeters, S. K. 1996. "The Use of the Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features." *International Journal of Remote Sensing* 17 (7): 1425–1432. https://doi.org/10.1080/01431169608948714.

Patro, S., and K. K. Sahu. 2015. "Normalization: A Preprocessing Stage." *IARJSET* 20–22. arXiv preprint arXiv:1503.06462. https://doi.org/10.17148/IARJSET.2015.2305 .

Pisner, D. A., and D. M. Schnyer. 2020. "Support Vector Machine." In *Machine Learning*, 101–121. Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

Poppenga, S. K., and B. B. Worstell. 2016. "Hydrologic Connectivity: Quantitative Assessments of Hydrologic-Enforced Drainage Structures in an Elevation Model." *Journal of Coastal Research* 76:90–106. https://doi.org/10.2112/SI76-009.

Poppenga, S. K., B. B. Worstell, J. M. Stoker, and S. K. Greenlee. 2010. "Using Selective Drainage Methods to Extract Continuous Surface Flow from 1-Meter Lidar-Derived Digital Elevation Data." *US Geological Survey Scientific Investigations Report 2010–5059* 12 p.

Pringle, C. M. 2001. "Hydrologic Connectivity and the Management of Biological Reserves: A Global Perspective." *Ecological Applications* 11 (4): 981–998. 10.1890/1051-0761(2001)011[0981:HCATMO]2.0.CO;2.

Rafique, M. U., J. Zhu, and N. Jacobs. 2022. "Automatic Segmentation of Sinkholes Using a Convolutional Neural Network." *Earth & Space Science* 9 (2): e2021EA002195. https://doi.org/10.1029/2021EA002195.

Regnauld, N., and W. A. Mackaness. 2006. "Creating a Hydrographic Network from Its Cartographic Representation: A Case Study Using Ordnance Survey MasterMap Data." *International Journal of Geographical Information Science* 20 (6): 611–631. https://doi.org/10.1080/13658810600607402.

Saranya, C., and G. Manikandan. 2013. "A Study on Normalization Techniques for Privacy Preserving Data Mining." *International Journal of Engineering and Technology (IJET)* 5 (3): 2701–2704.

Shore, M., P. N. C. Murphy, P. Jordan, P. E. Mellander, M. Kelly-Quinn, M. Cushen, S. Mechan, O. Shine, and A. R. Melland. 2013. "Evaluation of a Surface Hydrological Connectivity Index in Agricultural Catchments." *Environmental Modelling & Software* 47:7–15. https://doi.org/10.1016/j.envsoft.2013.04.003 .

Sinha, R. K., R. Pandey, and R. Pattnaik. 2018. "Deep Learning for Computer Vision Tasks: A Review." *arXiv preprint arXiv: 1804.03928*.

Sofia, G., G. D. Fontana, and P. Tarolli. 2014. "High-Resolution Topography and Anthropogenic Feature Extraction: Testing Geomorphometric Parameters in Floodplains." *Hydrological Processes* 28 (4): 2046–2061. https://doi.org/10.1002/hyp.9727.

Stanislawski, L., T. Brockmeyer, and E. Shavers. 2018. "Automated Road Breaching to Enhance Extraction of Natural Drainage Networks from Elevation Models Through Deep Learning." *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 42 (4): 597–601. https://doi.org/10.5194/isprs-archives-XLII-4-597-2018.

Stieglitz, M., J. Shaman, J. McNamara, V. Engel, J. Shanley, and G. W. Kling. 2003. "An Approach to Understanding Hydrologic Connectivity on the Hillslope and the Implications for Nutrient Transport." *Global Biogeochemical Cycles* 17 (4). https://doi.org/10.1029/2003GB002041 .

Stutheit, R. G., M. C. Gilbert, K. L. Lawrence, and P. M. Whited. 2004. "A Regional Guidebook for Applying the Hydrogeomorphic Approach to Assessing Wetland Functions of Rainwater Basin Depressional Wetlands in Nebraska." *US Army Corps of Engineers* 137. https://digitalcommons.unl.edu/usarmyceomaha/137.

Suthaharan, S. 2016. "Support Vector Machine." In *Machine Learning Models and Algorithms for Big Data Classification*, 207–235. Boston: Springer US. https://doi.org/10.1007/978-1-4899-7641-3_9.

Talafha, S., D. Wu, B. Rekabdar, R. Li, and G. Wang. 2021, September. "Classification and Feature Extraction for Hydraulic Structures Data Using Advanced CNN Architectures." In *2021 Third International Conference on Transdisciplinary AI (TransAI)*, Laguna Hills, CA (pp. 137–146). IEEE.

Tang, X., H. Hong, Y. Shu, H. Tang, J. Li, and W. Liu. 2019. "Urban Waterlogging Susceptibility Assessment Based on a PSO-SVM Method Using a Novel Repeatedly Random Sampling Idea to Select Negative Samples." *Journal of Hydrology* 576:583–595. https://doi.org/10.1016/j.jhydrol.2019.06.058 .

Tang, M., F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov 2018. "On Regularized Losses for Weakly-Supervised Cnn Segmentation." In *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany (pp. 507–522).

Wang, J., L. Li, Z. Hao, and J. J. Gourley. 2011. "Stream Guiding Algorithm for Deriving Flow Direction from DEM and Location of Main Streams." *IAHS-AISH Publication* 346:198–206.

Widyantoko, Z., T. P. Widowati, I. Isnaini, and P. Trapsiladi. 2021. "Expert Role in Image Classification Using CNN for Hard to Identify Object: Distinguishing Batik and Its Imitation." *IAES International Journal of Artificial Intelligence (IJ-AI)* 10 (1): 93. https://doi.org/10.11591/ijai.v10.i1.pp93-100.

Wulder, M. A., T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Alleng, et al. 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.

Xu, Z., S. Wang, L. V. Stanislawski, Z. Jiang, N. Jaroenchai, A. M. Sainju, E. Shavers, et al. 2021. "An Attention U-Net Model for Detection of Fine-Scale Hydrologic Streamlines." *Environmental Modelling & Software* 140:104992. https://doi.org/10.1016/j.envsoft.2021.104992.

Yang, H. L., J. Yuan, D. Lunga, M. Laverdiere, A. Rose, and B. Bhaduri. 2018. "Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (8): 2600–2614. https://doi.org/10.1109/JSTARS.2018.2835377.

Ye, C., Y. Li, P. Cui, L. Liang, S. Pirasteh, J. Marcato, W. N. Gonçalves, et al. 2019. "Landslide Detection of Hyperspectral Remote Sensing Data Based on Deep Learning with Constrains." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (12): 5047–5060. https://doi.org/10.1109/JSTARS.2019.2951725.

Yu, H., S. Jiang, and K. C. Land. 2015. "Multicollinearity in Hierarchical Linear Models." *Social Science Research* 53:118–136. https://doi.org/10.1016/j.ssresearch.2015.04.008.

Zhang, H., P. V. Zimba, and E. U. Nzewi. 2019. "A New Pseudoinvariant Near-Infrared Threshold Method for Relative Radiometric Correction of Aerial Imagery." *Remote Sensing* 11 (16).

Ziggah, Y. Y., H. Youjian, A. R. Tierra, and P. B. Laari. 2019. "Coordinate Transformation Between Global and Local Data Based on Artificial Neural Network with K-Fold Cross-Validation in Ghana." *Earth Sciences Research Journal* 23 (1): 67–77. https://doi.org/10.15446/esrj.v23n1.63860.