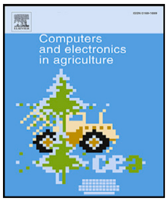
Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Original papers

Deep learning solutions for mapping contour levee rice production systems from very high resolution imagery

Dakota S. Dale[a], Lu Liang[b], Liheng Zhong[c,1], Michele L. Reba[d], Benjamin R.K. Runkle[e,*]

[a] *Department of Computer Science and Computer Engineering, University of Arkansas, Fayetteville, AR, USA*
[b] *Department of Geography and the Environment, University of North Texas, Denton, TX, USA*
[c] *Department of Water Resources, State of California, Sacramento, CA, USA*
[d] *USDA Agricultural Research Service, Delta Water Management Research Unit, Jonesboro, AR, USA*
[e] *Department of Biological and Agricultural Engineering, University of Arkansas, Fayetteville, AR, USA*

ARTICLE INFO

Keywords:
Remote sensing
Agriculture
Irrigation
ResNet

ABSTRACT

The construction of contour levees for rice irrigation represents a major landscape management activity with impacts on irrigation water use efficiency, crop management decisions, and food production. However, levee distribution information traditionally relies on local field surveys because remote sensing approaches are complicated by irregular spacing, shape, and landscape variability within the field. In this paper the authors develop a deep learning approach capable of identifying rice fields with contour style levee irrigation practices from open-source aerial imagery. To generate a levee-identification scheme, a hybrid ResNet/Unet model is built from the commonly known Residual Network (ResNet) architecture for multi-layer deep learning strategies. The model takes a 320 × 320 RGB aerial landscape image from the US National Agricultural Imagery Program as input along with label data to then generate a probability map of the distribution of farm fields that use contour levees within the image. In performing this task, the model generates a 0.991 receiver operating characteristic curve score. The model continues to perform well under the introduction of clouds, data augmentation, or minor reductions in spatial resolution. Throughout these tests, the model performed within 0.2 of its original score, except for when the image quality was reduced to 60 m wherein the model score dropped to 0.691. Via these tests the model demonstrates potential to function well given different spatial extents or potential satellite remote sensing with moderate (10 m) resolutions. This model provides a proof-of-concept for the use of aerial imagery and a deep learning strategy for irrigation-type mapping practices.

1. Introduction

Agriculture accounts for 70% of freshwater withdrawal globally; therefore it is paramount to minimize the amount of freshwater used while maximizing crop yield (Campbell et al., 2017). However, each crop requires different growing conditions in terms of soil composition, climate, and water availability. For example, as one of the major staple foods that supply 20% of the calories consumed globally (Kubo and Purevdorj, 2004), rice requires up to 2–3 times as much water as other cereal crops like wheat and maize (Bouman and Tuong, 2001). This high water use is associated with widespread anaerobic cultivation practices as a weed control mechanism and driver of greenhouse gas emissions. Between 24% and 30% of global freshwater resources are used to irrigate rice (Bouman et al., 2007). Thus, many precision irrigation and conservation efforts aim to reduce the consumptive water requirements of rice production while limiting the potential for yield loss.

As there are only rare instances where the field topography will allow for standing water for rice production, farmers must use other means to prevent water from draining off the field. In the United States with larger field sizes and a higher degree of mechanization, farmers typically implement various forms of levee systems which essentially act as small dams that hold water in each section of the field. The most common levee system consists of generally curved contour levees that follow the topographical elevation lines of the field to ensure equal elevation within each section. This system accounts for 47.9% of rice acreage in Arkansas—the largest rice growing state in the U.S. (Norman and Moldenhauer, 2019). Second, precision land grading on some fields allows the straight levee system with parallel levees perpendicular to

[*] Corresponding author.
E-mail address: brrunkle@uark.edu (B.R.K. Runkle).
[1] Present Address: Ant Group, World Financial Center, Beijing 100000, China.

https://doi.org/10.1016/j.compag.2023.107954
Received 11 January 2023; Received in revised form 19 May 2023; Accepted 24 May 2023
Available online 17 June 2023

the consistent slope (0.1–0.5%). The third system is known as "Zero-grade", whereby precise leveling eliminates the need for levees entirely. The Zero-grade system enables faster water distribution and a constant flood depth across the field and can lead to water savings of 40% compared to contour or straight levees (Henry et al., 2016). Fourth, and less common in rice production, some farmers forgo levee systems and use a pivot sprinkler to deliver the precise amount of needed water (Vories et al., 2013). Finally, furrow irrigation uses surface irrigation through small furrow ditches without levee use, and has increased to over 15% of irrigation practices in Arkansas (Chlapecka et al., 2021; Stevens et al., 2018). Each of these systems requires different levels of labor, power, and water, and knowledge about shifts among them over time can inform water resources planning and sustainability assessments (Moreno-García et al., 2021). However, there is not currently a viable system for efficient, widespread categorization of these fields.

Examining the distribution and prevalence of these systems provides a baseline dataset to guide improved efficiency in water use, e.g., by the implementation of multiple inlet rice irrigation (Shew et al., 2021), furrow irrigation (Chlapecka et al., 2021), alternate wetting and drying (Atwill et al., 2020; Carrijo et al., 2017), or modeling the effects of these approaches on water use (Carroll et al., 2020). Historically, the primary method of identification was the use of surveys by county or extension offices relying on owner-reporting of land use, or via ground-truthing with windshield surveys (Smith et al., 2007). For any large spatial extent where laborious hand-labeling is inefficient, remote sensing offers a viable solution to detect large-scale land cover and land use, as well as their changes, in an efficient manner (Weiss et al., 2020). However, traditional pixel-based methods have difficulty addressing this classification, because most irrigation infrastructure or equipment can only be holistically seen from a landscape view. Additional difficulties lie in the complex visual characteristics of various irrigation types regarding their geometry (e.g., field size, levee width and length, and levee curvedness), photometry, and texture. A typical contour-levee rice field usually only has 3 to 10% of the land in levees that are 36 to 40 in wide (Massey, 2023). Applying deep learning approaches to high-resolution images has provided an opportunity to classify rice irrigation strategies through image segmentation and feature extraction (Liang et al., 2021; Meyarian et al., 2022). Thus, the objective of this study is to develop a deep learning-based method to classify irrigation practices using high-resolution aerial imagery.

We next move toward model portability, where the developed method can be adapted to data acquired from other platforms, with variations in greeness, spatial resolution, and image contamination. Hence, our second objective is to test the model's effectiveness in different image quality scenarios and scalability to be applied at different geographical locations. These scenarios include coarsened imagery resolution or the introduction of various forms of noise to the image.

2. Method

The research workflow includes phases of image annotation, preprocessing, deep learning classification, hyperparameter tuning, and scalability testing (Fig. 1). First, our annotators labeled the imagery according to the presence of fields using contour levee irrigation. These images are then converted to grayscale and divided with their respective labels according to a 5×5 grid composed of $320{\times}320\ images$. After dividing the images into training and testing sets, we began training the model, adjusting the hyperparameters as needed. Once the final model was developed, we stress-tested it according to three criteria: resolution reduction, cloud noise addition, and gaussian noise addition.

2.1. Study area

We chose Lonoke County in Arkansas, where rice is the dominant agricultural crop type and the region has diverse levee systems, to train the model and assess its scalability by testing the model in distinct physio-geographical areas (Fig. 2). Lonoke County is located in central Arkansas, on the west side of the Lower Mississippi Alluvial Plain's Grand Prairie, 32.5% of its harvested agricultural land is planted in rice (USDA-NASS, 2021). The region has intense water demands for crop production, declining groundwater table levels, and the increasing stress of climate change which collectively has amplified needs for irrigation strategies and conservation management approaches. With inter-annual and inter-region variability, rice is usually planted in late March to mid-May and harvested in mid-August to mid-October. At the state level in 2015, only 21% of rice acreage will continue growing rice in the following year and the majority will be rotated into soybeans (72%). The remaining small percentage (7%) follows other crops such as corn, grain sorghum, cotton, wheat, oats, and fallow (Norman and Moldenhauer, 2016). The high rotation rate of rice presents a challenge in moderate resolution remote sensing mapping as the locations of ground truth points vary year to year.

2.2. Image selection and annotation

2.2.1. Input imagery

This study utilized the freely accessible United States Department of Agriculture (USDA) National Agricultural Imagery Program (NAIP) aerial photographic imagery that was acquired between 31 July and 25 August 2015 at 1-m spatial resolution (USDA, 2017). We consider NAIP to a be very high resolution imagery dataset due to its 1-m scale as defined in Fu et al. (2017). The NAIP program began in 2003 with a 5-year cycle and transitioned to a three-year cycle in 2009 with one year publication latency to acquire aerial imagery during the agricultural growing season in a wide range of areas in the U.S. Although the selected 2015 NAIP images come with four bands (red, green, blue, and near-infrared (NIR)), we tested the method only with the three natural color bands to reduce data volume. Moreover, some studies have shown only minor performance gains when deep learning algorithms trained on multispectral images were applied over trichromatic images. For instance, one landscape segmentation study found less than a 1% increase in the accuracy of grass identification from an image and a 3.9% increase for soil identification in multispectral models versus RGB models (Salamati et al., 2012). Conversely, a study to identify rice lodging from UAV imagery found that an RGB model achieved a measure of similarity that was 2% higher than its multispectral counterpart (Zhao et al., 2019). Several other studies found comparable results between multispectral and trichromatic imagery (Elihos et al., 2018; Liu et al., 2020). Lastly, constraining the analysis to the RGB domain increases the transferability of the proposed method to NAIP images acquired in years without the NIR band and to other RGB data sources.

2.2.2. Image annotation and sample selection

Our images were visually annotated using multiple people who were trained to provide consistent annotation and deliver high accuracy of field type labeling. Here we focused on the identification of contour-levee fields due to their strenuous water requirements (annual irrigation for rice: 892 mm) that are 8% greater than the next highest irrigation technique, straight levees (Reba and Massey, 2020). Contour-levee irrigation is a conventional irrigation method where levees are typically 30–45 cm in height to maintain flood between levees. From an aerial view, these fields are characterized by distinguishable lines that seem to resemble a topographical pattern used by the annotators to classify irrigation type. We also used the CDL layer as supporting information to provide the user with the estimated distribution of crops such as rice and soybeans, with very high classification accuracy on the major crops (Lark et al., 2021). This information aids the user in the labeling process because, depending on the area, levee systems are nearly exclusive to rice and soybean production fields. Additionally, for field patches with vague or unidentifiable patterns, the analysts also

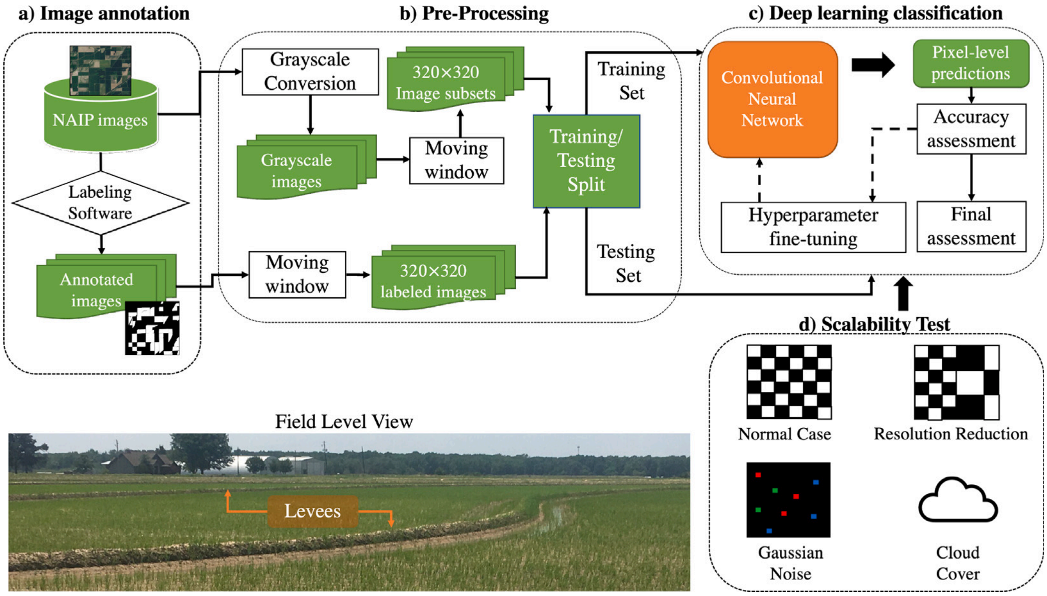


Fig. 1. Workflow diagram; (a) The dataset is fed into the labeling program for image annotation. (b) In pre-processing, the images are converted to grayscale, then both the images and the labels are split into 320 × 320 *subsets*. (c) Model tuning and accuracy assessment. (d) Different scalability analysis scenarios are created for testing in the classification phase and final assessment. Field-scale image of a contour levee rice field from B. Moreno-García in Lonoke County, Arkansas, 12 May 2019.

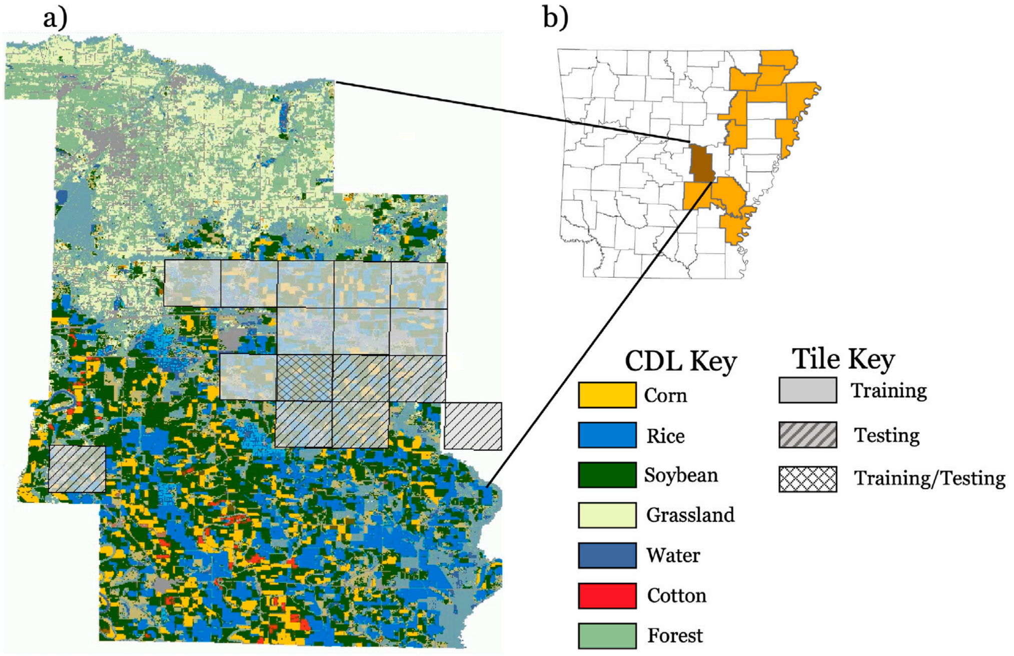
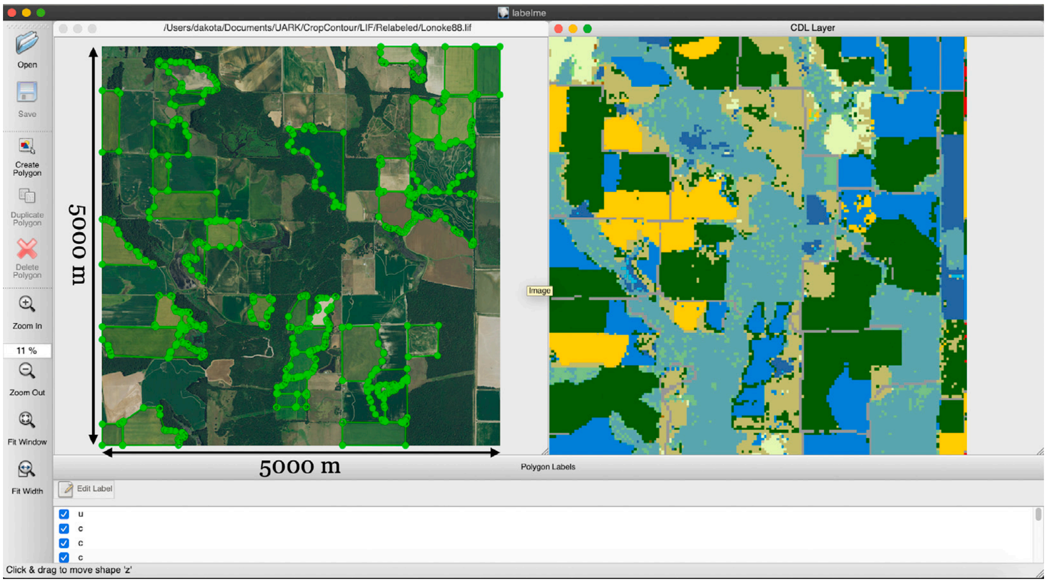


Fig. 2. We chose 11 Arkansas counties (in orange) for testing the algorithm and Lonoke County (indicated in dark orange) for training. The zoom-in view of Lonoke County has the 2015 Cropland Data Layer (CDL) to show the major crops, along with the location of analyzed data tiles in gray. Note: One tile in neighboring Prairie county was also used for training and one tile was partitioned into the training and testing sets (denoted as Training/Testing).

used the CDL to assist in image interpretation - i.e., by ensuring that those fields were planted in either rice or soybean.

To support the selection and interpretation of training and validation samples, we designed a labeling program (Fig. 3), where the user can upload the image for display alongside the CDL (Boryan et al., 2011) for the corresponding spatial extent. The trained analyst first draws polygons following the edges of each field over the NAIP image. Next, the analyst is prompted to label the field's irrigation and/or levee system. Fields that are difficult to identify were labeled as unknown and discarded in the subsequent process. To ensure the integrity of the dataset and mitigate human error during labeling, the interpretation is repeated multiple times using a wall-to-wall method, where a new analyst inspects each field across the image and labels any crop fields that the previous analyst may have missed. To guarantee the best interpretation, at least two analysts were trained together for sample selection by the project leader. An initial training sample selection was performed by one analyst and the interpretation results were passed to another analyst for cross-checking. Cross-checked results were submitted to the final analyst, who is the quality controller for final checking. To gauge the degree of agreement between our annotators, we implemented Cohen's Kappa Coefficient. This statistical measure serves to measure the agreement between annotators for qualitative



☑ u
☑ c
☑ c
☑ c

Click & drag to move shape 'z'

Fig. 3. A screen capture of the LabelMe program, with annotated crop polygons superimposed on a subset of 5000*5000 NAIP image (left) located in Lonoke County centered around (34.7654, −91.6952) and Cropland Data Layer (right) with the same spatial extent as the NAIP area. Colors represent different crop types following the legend in Fig. 2. The label corresponds as follows: c for contour levees, s for straight levees, z for zero grade, and u for unknown.

items, such as the field typings. Cohen's Kappa is expressed as

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \tag{1}$$

Where Pr(a) is the probability of overall agreement of the annotators and Pr(e) is the probability of overall agreement if by chance (Vieira et al., 2010).

2.2.3. Image pre-processing

Due to the variation in the greenness of crop fields caused by the differences in acquisition date, cameras, flight height, and weather conditions, we used grayscale images in classification. Grayscale images can decrease the algorithm's sensitivity to color changes between images, as well as any color-associated errors that occurred during capture. Additionally, the grayscale image emphasizes the shapes and patterns present within the fields more than colors, which is beneficial in distinguishing the specific levee systems. One study found nearly a 2% increase in accuracy with grayscale images over RGB images in identifying colorectal polyps with a CNN (Hsu et al., 2021). Similarly, a facial recognition study found that RGB and grayscale images had equivalent rates of recognition (Torres et al., 1999). To convert images from RGB to grayscale, we used a formula proposed in Pascale (2003),

$$I = 0.299 \times R + 0.587 \times G + 0.114 \times B \tag{2}$$

Furthermore, we subset the raw 5000 × 5000 images into a 5 × 5 grid creating 25 1000 × 1000 subimages for computational efficiency. These images were then downsampled to 320 × 320 using the skip count method with a factor of 3, which extracted every 3rd pixel value and discarded the rest.

2.2.4. Training/test sample split

The quality of training and test samples is paramount to the model's success. In this study, training samples were selected to comprise as much diversity of land use types in as few images as possible to reduce model training time and time spent on manual sample annotation. Test samples are also diverse when in the season the images were taken, contributing to a robust classification evaluation. The Lonoke county images have the widest image capture interval with dates ranging from 13 July to 26 August, which in this region generally is after canopy closure and prior to harvest. When training deep learning models, providing the appropriate balance between the size of the training and testing sets is crucial to avoiding over- and underfitting. Here, overfitting with an overly complex model could occur by detecting fields during only a particular growth phase or season and can be detected through very high accuracy for the training and/or validation sets with low accuracies for the testing set (Hawkins, 2004). Conversely, underfitting could manifest as the model predicting every field as having contour levees simply because of the similarity in color. Despite its importance, there is little consensus as to the proper split for the data (Joseph, 2022). Studies have suggested that anywhere from 30% (Nguyen et al., 2021) to nearly 100% of the data should be for training (Dubbs, 2021). Thus, to optimize the trade off between model performance, training time, and avoiding overfitting, we trained the model 10 times for each 10% increment of the training fraction ranging from 10% to 80% (Supplemental Fig. 3). Through this process, we identified the 60/40 training/testing split as the most optimal. As shown in Fig. 2, whole tile representations were assigned to either training or testing, with the exception of one tile in which 15 of its subtiles were assigned to training and 10 were assigned to testing.

2.3. Network architecture

2.3.1. Model selection

To select an appropriate deep learning architecture design, we performed an initial comparison between the VGG and ResNet network designs, both of which come with the Tensorflow-Keras installation (Abadi et al., 2016), a common deep learning framework. The main difference between these models is the number of weighted neural layers, which are 16 and 50 for VGG and ResNet, respectively. The appropriate number of layers depends on the complexity of the problem, which is not known a priori. Neither of these models classify on a per-pixel basis, but rather assign a label to the entire input subimage. To effectively identify the label of each image object, we created UNet hybrid architectures that utilize native encoders from ResNet and VGG coupled with custom decoders similar to those in UNet. UNet was first introduced in Ronneberger et al. (2015) and is named for its architecture's shape. The model down-samples the input (the downslope of the "u") to extract the feature maps then up-samples the input from previous layers and applies the feature maps (the upslope of the "u"). The upsampling phase of the architecture allows it to classify

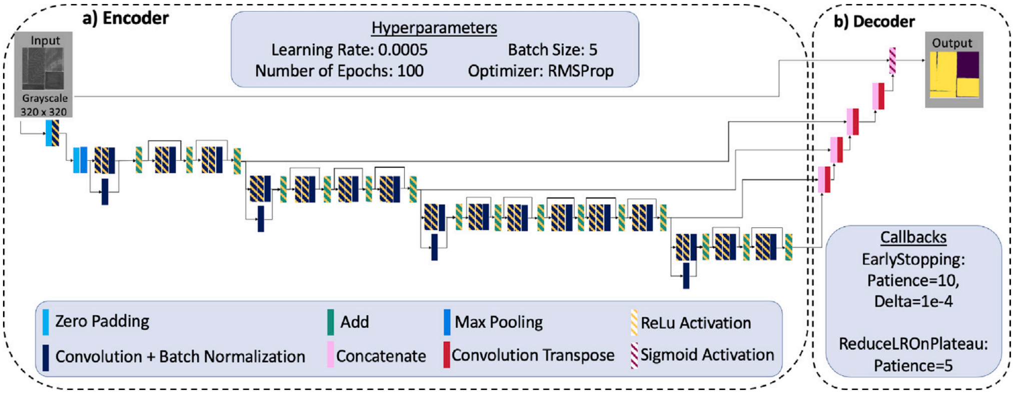


Fig. 4. Block Diagram of the hybrid ResNet/UNet model's architecture; (a) the encoder segment of the architecture passes data through blocks of convolution, ReLu activation, and pooling to extract the patterns of interest before being upsampled in (b) the decoder to generate the output, a probability map classifying different segments of the grayscale input image, yellow for contour field and purple for background.

on a per-pixel basis. Thus, these hybrids preserve the portion of the VGG and ResNet models responsible for their high performance while also extending them with UNet decoders that allow for pixel-based classification.

To compare the VGG/Unet and ResNet/Unet hybrids, we trained them using a custom loss function which sums (1) the intersection-over-union (IoU) loss, that essentially measures the similarity between the labeled and predicted bounding boxes, and (2) the binary cross-entropy functions that are described in more detail in the next section. Preliminary results showed that the ResNet/Unet model consistently surpassed the VGG/Unet model throughout the training epochs, or iterations, by converging at 0.06 in 21 epochs versus 0.33 in 55 epochs (Supplemental Fig. 2).

The ResNet architecture was first introduced in an effort to address the accuracy degradation that many deep models face. Essentially, as the depth (or number of layers) increases, the accuracy will become stagnant and then quickly decrease. To combat this problem, ResNet uses identity mapping, wherein each block or grouping of layers uses both the raw input and output of the previous block. This redundancy provides the retention of useful information while also preventing the over-abstraction of the data. Overall, ResNet has outperformed many models including Fitnet and Highway with more than a 1% reduction in error (He et al., 2016). Our hybrid ResNet/Unet (Fig. 4) uses 7 specific types of layers from the tensorflow library (Abadi et al., 2016): Zero-padding, 2-dimensional convolution, batch normalization, addition, concatenate, max pooling, and convolutional transpose. Zero padding appends zeros to the top, bottom, left, and right of the image, effectively centering it in a larger array. Next, the 2-dimensional convolution layer provides the functionality for the model to extract feature maps. Third, the batch normalization layer standardizes the input to maintain a mean close to zero and a standard deviation close to one (Ioffe and Szegedy, 2015). Fourth, the addition layer adds two arrays of the same shape to return an array of the same shape as the inputs. Next, the concatenate layer also takes two arrays as inputs and appends one input to the other along a specified axis. Next, the max pooling layer effectively downsamples its input by extracting the maximum value of a moving window. Lastly, the convolutional transpose, also known as deconvolution, works in the opposite direction as the regular convolutional layer. Thus, the layer essentially applies the feature maps that the convolutional layers had previously extracted.

2.3.2. Model tuning

The ResNet/Unet model hyperparameters were set as follows. First, the learning rate that controls how much the weights of the network are updated within each training iteration is set to 0.0005. Second, the number of these training iterations, known as epochs, acts as the maximum number of times that the weights can be updated, and is set to 100. Third, the batch size, referring to how many samples of the training set will be used within that epoch, was set to 5 images. Fourth, we established gradient accumulation into an exponentially weighted average to guide model learning using RMSProp (Tieleman et al., 2012; Zaheer and Shaziya, 2019), which is an optimizer protocol that was found to achieve the highest validation accuracy with the VGG-16 model on a common image classification dataset (Li et al., 2021).

Additionally, several self-tuning callbacks were used to dynamically augment the training process. First, we implemented early stopping, which cuts the training process short if the model stops improving for more than two epochs. Second, through the "reduce" callback we were able to lower the learning rate during a learning plateau, which prevents the model from overshooting the minimum loss.

2.3.3. Determining the optimal threshold for binary mapping

To determine whether a field contains a contour levee patch from the probability map, we establish a threshold value to convert the output into a binary map. The probability value of each pixel will be adjusted to either 1, should its probability be higher than the threshold, or 0 in the case that its probability is lower. For example, with a threshold value of 0.6, all pixels with values above 0.6 were adjusted to 1 and all others were adjusted to 0. By iterating over the "binary metrics" (Accuracy, F1, IoU, and BER, discussed in a later section) for varying thresholds, we were able to identify the value that yields the greatest performance.

2.4. Model evaluation and accuracy assessment

2.4.1. Probabilistic evaluation

We evaluated the ResNet/Unet model's performance based on our loss function consisting of the IoU, also known as the Jaccard index (van Beers and Marco A. Wiering, 2019) and the Binary Cross Entropy function. The IoU is a widely used metric that shows the model's object detection by returning the similarities between the ground truth bounding box and the bounding box of the prediction. Binary Cross Entropy, also known as the Log loss, takes the negative log of the prediction probabilities. If the probability is 100%, the loss will be zero. As the probability decreases, the loss will increase thus penalizing uncertainty.

Furthermore, we examined the model's raw output, which is the probability map that predicts the likelihood of each pixel belonging to a field that contains contour levees, against the corresponding ground truth labels. The metric employed is the set of Receiver Operating Characteristic (ROC) curves (Bradley, 1997; Flach, 2016) that compare the true positive rate (the proportion of pixels predicted and labeled as contour fields) and the false positive rate (the proportion of pixels predicted as fields but labeled as a non-contour field). By thresholding the model's output as described in Section 2.3.3, we are able to convert the probability map into a binary map. Next, we can compare the binary predictions to their respective labels to calculate the true and false positive rates. This process is repeated for each possible value for the decision threshold (within the range of (0,1)) leaving a set of true/false positive rate pairings that we can plot.

Smaller false positive rates indicate more true negatives while higher true positive rates indicate more true positives. Thus, as the area under the receiver operating characteristic (AUC-ROC) curve approaches 100%, the ResNet/Unet model's effectiveness increases as shown in supplemental fig. 3. A 1:1 line is also drawn to signify a random predictor (Mas et al., 2013). By visualizing and quantifying these metrics, a holistic assessment of model skill can be estimated.

2.4.2. Binary accuracy assessment on the classified map

The binary map after thresholding was assessed using four metrics, separately overall accuracy, F1, IoU, and balanced error rate.

Overall accuracy (OA) is a relatively standard accuracy metric for binary classification that is calculated by dividing the number of true classifications by the total number of classifications. The F1 score is the harmonic mean of Precision and Recall, which are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

Where TP indicates true positive, FP indicates false positive, and FN indicates false negatives (Sasaki, 2007). Precision shows the ratio of samples that are correctly classified as positive out of all of the samples labeled as positive. Recall shows the number of actual positives compared to the total number of labeled positives in the set. The values are further used to calculate the F1 score which allows us to examine the combined effects of misclassification. However, this metric is independent of the value of true negatives, so in our case, the pixels that are correctly classified as within non-contour fields will not be reflected in this metric. This metric is less prone to issues related to class imbalances than the standard overall accuracy metric, but it does not distinguish between false positive or false negative model errors (Japkowicz, 2006). Thus, in cases where the drawbacks of false positive and false negative classifications are not equal, this metric can be misleading. In our project, however, a false positive and a false negative are comparable.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5}$$

Additionally, we calculated the Balanced Error Rate (BER) metric. This metric shows the average proportion of error between the classes, which is given by:

$$BER = (1 - \frac{1}{2}(\frac{TP}{NP} \times \frac{TN}{NN})) \times 100 \tag{6}$$

Where Np and Nn are the number of Field pixels and the number of non-field pixels. Because of the disproportionate number of field pixels per non-field pixels, the BER metric gives a less biased result than the standard mean accuracy. Since this metric displays the proportion of error, lower values indicate better model performance.

2.4.3. Comparison to other products

We also evaluated performance by comparing our algorithm's accuracy with previous studies. We first conducted a literature review in the field of agricultural landscape identification and summarized the accuracy values for the highest performed models. We then compared the mapping outcomes with two relevant studies. The first study took a deep network-based method IrrNet-Bi-Seg that used a bi-stream encoder–decoder architecture to quantify the contour-levee fields. By applying the IrrNet_Bi_Seg method on the same NAIP dataset, it achieved an average accuracy of 86.23% and 15%–17% improvement over benchmark methods (Liang et al., 2021). The second study is Landsat-based Irrigation Dataset (LANID-US) - 30 m resolution annual maps of irrigation distribution for the US centered around detecting irrigated fields (Xie and Lark, 2021). It used a semi-automatic training approach for training sample generation and ecoregion-stratified random forest classification using Google Earth Engine. The annual maps there achieved a mean Kappa value of 0.88, overall accuracy of 94%, and producer's and user's accuracy of the irrigation class of 97.3% and 90.5%, respectively, at the sub-national (large aquifer) level.

2.5. Scalability test

The quality of remotely sensed images could be affected by various atmospheric conditions or the mechanical-optical systems, resulting in different levels of noise and obstructions being added to the image. Using the clear, undistorted image as our "normal" benchmark scenario, we conducted tests on a series of image artifacts, including:

(1) Sensitivity to various spatial resolutions

Our first sensitivity experiments consisted of reducing the image resolution to account for changes in sensing equipment. To simulate a reduction in quality of sensing equipment, we tested the ResNet/Unet model on reductions of resolution from 1-m to 10-m, 20-m, 30-m, and 60-m. To achieve this reduction, we created a duplicate image where the pixel values were averaged across varying window sizes following the step in Rosa et al. (2021). For example, to reduce the image to 10-m, we averaged consecutive windows of size 10px by 10px. This same method was applied for tests on 20-m, 30-m, and 60-m resolutions.

(2) Sensitivity to various image contamination scenarios

To test different degrees of noise via random noise augmentation to the images, we created a separate array which contains random samples from a normal gaussian distribution with a standard deviation (which can also be interpreted as the spread) of 0.3. Then, once combined with the image it creates a static effect. This occurs when the signal to noise ratio is low, which may not be common in modern equipment, but is still possible in older instruments (Curran and Hay, 1986).

We test cloudiness simulations given that approximately 45% of the Earth's land surface is covered by clouds at any moment (Stubenrauch et al., 2013). This effect may also be more relevant to continuous satellite observations (e.g., the Planet constellation, Houborg and McCabe, 2016) than to aerial imagery taken deliberately on clear days. To simulate this contamination, a python library imgaug was used (Jung et al., 2020). We set the opacity of the layer to only 40% to achieve more representative conditions that can be encountered by a low flying plane, based on our visual assessment of likely opacity conditions in the resultant images.

(3) Geographic scalability

Lastly, we tested the ResNet/Unet model in samples from a different location, expanding from the singular location used in the training and testing sets. We seek to develop a model that applies globally, requiring that the model's spatial transferability be scrutinized. Thus, we selected five images with a high degree of rice production from 11 different counties across Arkansas: Arkansas, Clay, Craighead, Crittenden, Desha, Greene, Jackson, Jefferson, Lawrence, Mississippi, and Woodruff. These images, referred to as the 55 testing tiles, were annotated and evaluated using the same annotation scheme and metrics described previously. In an effort to explain any variation in performance when applying the ResNet/Unet model to imagery outside of Lonoke we implemented three greeness metrics: excess greenness (EG), green chromatic coordinate (GCC) (Reid et al., 2016), and the Greenness Index (GI) (Louhaichi et al., 2001). These metrics are defined as:

$$EG = (2 \times G) - (R + B) \tag{7}$$

$$GCC = \frac{G}{R + G + B} \tag{8}$$

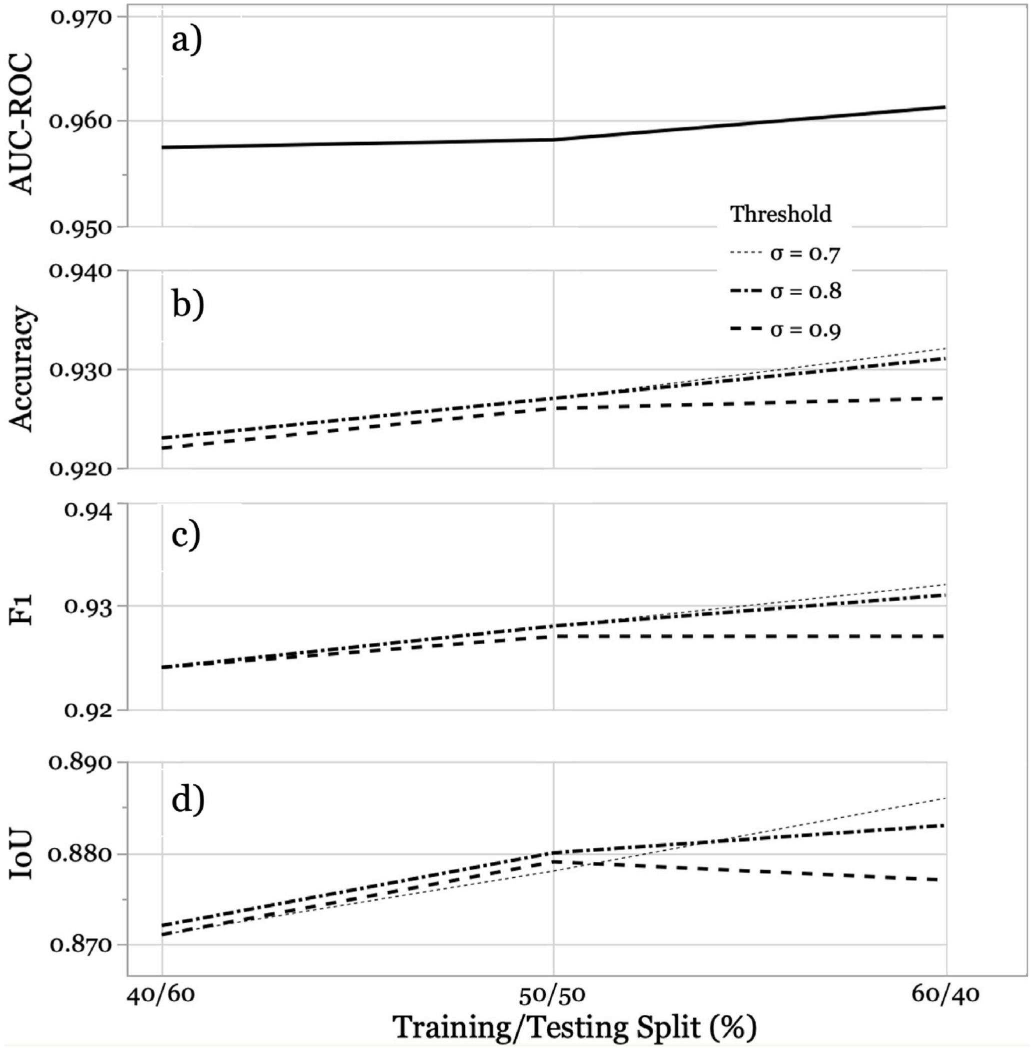


Fig. 5. The ResNet/Unet model performance for each training split and confidence threshold (when applicable, denoted by σ). (a) Area Under the Receiver Operating characteristic curve (AUC-ROC) is used to compare performance across Training/Testing splits. The binary performance metrics are: (b) Accuracy, (c) F1, and (d) Intersection over Union (IoU).

$$GI = \frac{2 \times G - R - B}{2 \times R + G + B} \tag{9}$$

where R, G, and B represent the red, green, and blue bands of the input imagery respectively.

3. Results

3.1. Dataset and model optimization

Our final training and test dataset contains 723 annotated polygons with 461 labeled as contour fields and the remaining 262 as other irrigation methods (228 as zero grade, 25 as straight levee, 0 as center pivot, and 9 as unknown) that will be grouped with background. Each polygon encompasses a crop field and the polygons averaged 20.35 ha in size with a standard deviation of 14.28 ha; their sizes ranged from 0.36 ha to 126.75 ha. These annotations were inspected by two trained annotators and received an average 0.841 Cohen Kappa inter-rater agreement score, which falls in the 0.81–0.99 interval of almost perfect agreement (Landis and Koch, 1977). After the images were subsetted, there were a total of four hundred 320 × 320 px images.

To balance the training and testing sets, we calculated the averaged AUC score over 10 training sessions for varying training/testing splits. Because model performance reached a plateau (AUC score: 0.957) after 40% of the data was designated for training (supplemental Fig. 1), we elected to use this ratio for later tests. Due to the probabilistic nature of the raw model output, we also defined the confidence threshold (σ) to convert the model's predictions from probabilities to binary outputs. Fig. 5 shows the effects of three different σ values across three different metrics where a binary map is required. The difference in performance between these thresholds was within 0.01 for all cases. Considering the 0.7 threshold yielded the highest accuracy and F1 regardless of training/testing split, we used this threshold for our remaining tests. With the 40/60 training/testing split on the Lonoke County tiles and a threshold value of 0.7, the ResNet/Unet model produced a 0.991 AUC, 0.548 BER, 0.970 Accuracy, 0.924 F1 (composed of Precision of 0.945 and Recall of 0.903), and 0.858 IoU.

3.2. Results of scalability tests

In our scalability experiments, we assessed the ResNet/Unet model's feasibility given different environmental and technological scenarios. Overall, the model presented high sensitivity to resolution reduction (Fig. 6 a), but continued to perform well given some degree of cloud cover or a small spread of gaussian noise (Fig. 6 b). Spatial scalability could present a challenge for the model considering that the worst performing of the 55 tiles tested has received a 83.1% AUC-ROC (Min Diff Site in Fig. 6 c), but on average the model continues to perform well with approximately an 8% reduction in the average case (Avg Diff Site in Fig. 6 c).

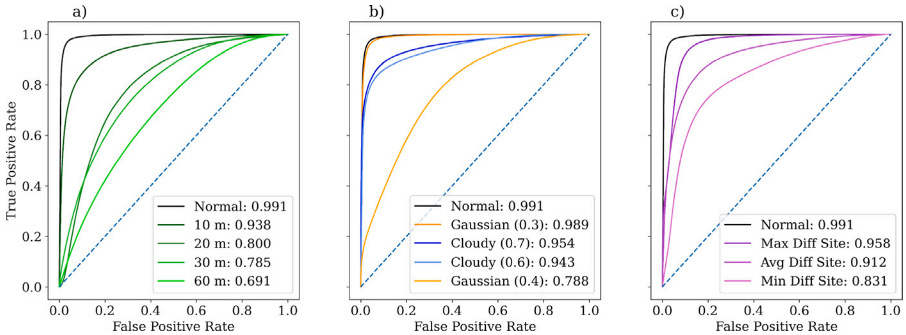


Fig. 6. The receiver operating characteristic (ROC) curves of the ResNet/Unet model under different scenarios: (a) the effects of coarsening the resolution; (b) the effects of noise addition; (c) varying spatial extents (where Max or Min Diff Site denotes the best or worst performing tiles and Avg Diff Site is the mean performance across all tiles). The legends display the values of the area under the ROC curve (AUC-ROC). The dashed diagonal line shows the ROC for a random classifier.

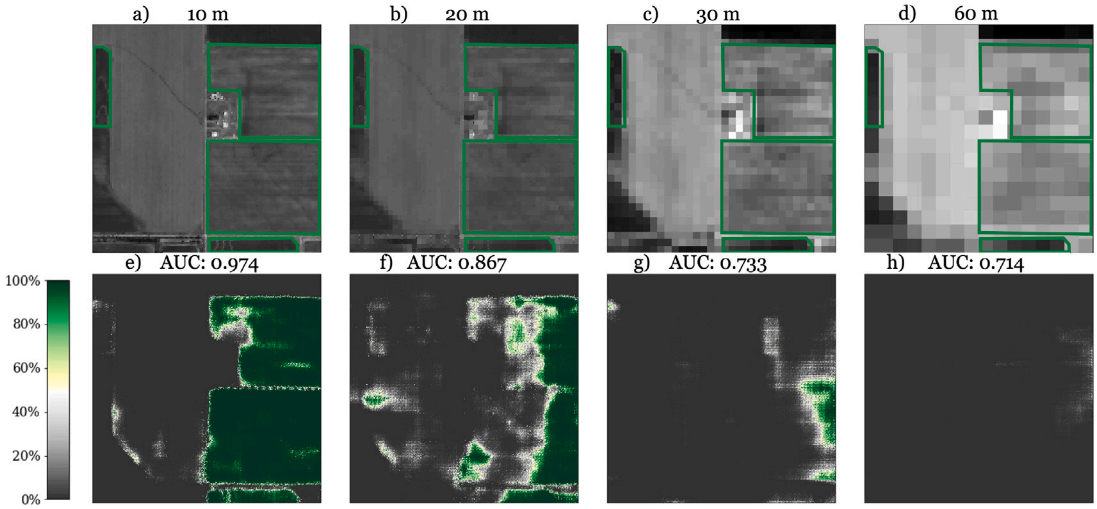


Fig. 7. Demonstration of the resolution reduction tests. (a)–(d) The augmented input image coarsened to the specified resolution, with associated visual annotation (green outline), compared to (e)–(h) the ResNet/Unet model's prediction at the respective level of coarsening. The color bar shows the probability the pixel represents a field with contour levees.

3.2.1. Sensitivity to different spatial resolution

In testing the impact of image spatial resolution, we observed a quick degradation in performance after resampling to less than a tenth of the ResNet/Unet model's native resolution (i.e., from 1 m to 10 m or coarser). Specifically, AUC changes from 99.1% in the original 1-m resolution to 93.8%, 80.0%, 78.5%, 69.1% in the 10-m, 20-m, 30-m, and 60-m resolutions, respectively (Fig. 6). Fig. 7 provides a qualitative demonstration of the resampled images and the corresponding model predictions.

3.2.2. Sensitivity to different atmospheric conditions

Next, we tested the ResNet/Unet model's sensitivity to different simulated atmospheric conditions. For both cloud contamination and gaussian static noise, we tested across a range of parameter values to find the threshold at which model performance decays. In the case of cloud cover, the maximum opacity of the cloud layer (denoted by α) was between 60% and 70%, receiving a 94.3% and 95.8% AUC-ROC respectively (Fig. 6a, 8a,b,e,f). For gaussian noise, spread values (σ_G) between 0.3 and 0.4 degraded model performance to below 80% AUC-ROC (Fig. 6b, 8c–d,8g–h).

3.2.3. Sensitivity to geographic scalability

When testing the model outside of the training region, the performance slightly decreased from Lonoke County (AUC of 99.1%) to the 55-tile set (AUC of 91.2%). The values for OA, F1, IoU and BER are given in Table 1 and represented in the brown sets in Fig. 9. We investigated performance within the 55 tiles in terms of imagery capture date, excess greenness (EG) and green chromatic coordinate (GCC), the Greenness Index, as well as the presence of straight levees, but none provided explanatory power over the model residuals beyond 0.9–7.7%. Examples of these images are provided in (Supplemental Fig. 4).

Table 1
Performance metric values, ranges, and standard deviations for the 55 testing tiles.

Metric	Average +–Std	Min	Max
OA	90.4% +–13.1%	22.6%	100%
F1	30.0% +–38.3%	0.0%	99.2%
IoU	25.3% +–34.5%	0.0%	98.3%
BER	70.2% +–16.8%	50.0%	100.0%

3.3. Comparison with other irrigation products

From our literature search, we selected eight papers that are most relevant to irrigation or agricultural landscape classification, and we summarize their subject area and reported accuracies in Table 2. By comparing our Lonoke County mapping results against IrrNet_Bi_Seg results using the same testing data. IrrNet_Bi_Seg yielded an AUC, overall accuracy, F1, IoU, and BER of 93.5%, 94.4%, 88.7% (a precision of 85.7% and recall of 91.9%), 79.7%, and 54.1% respectively. The ResNet/UNet and IrrNet_Bi_Seg models performed comparably with ResNet/UNet having a slight increase in four of the five metrics. The comparison with 30 m resolution LANID data demonstrates that LANID

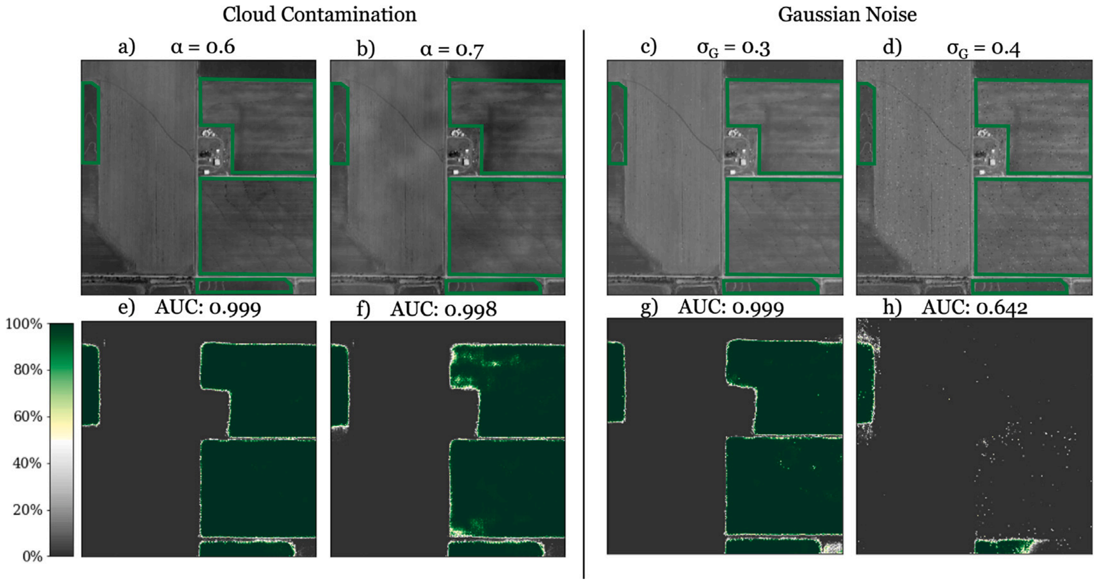


Fig. 8. Demonstration of the atmospheric noise addition tests, given cloud contamination parameter α values of (a, e) 0.6 and (b, f) 0.7 and gaussian noise parameter σ_G values of (c, g) 0.3 and (d, h) 0.4. (a)–(d) The augmented input image with associated labels and (e)–(h) the model's prediction across noise addition scenarios. The color bar shows the probability the pixel represents a field with contour levees.

Table 2
Reported accuracy of different model architecture types in land cover, plant morphology, and image analysis.

Model	Subject area	Accuracy
CNN - 5 Layers Setup 2 (Grinblat et al., 2016)	Plant Morphology	96.9
2-D CNNs (Kussul et al., 2017)	Agricultural Landscape Identification	94.6
CNN256 (Martins et al., 2020)	General Land Cover Mapping	93.2
CNN - 5 Layers Setup 1 (Grinblat et al., 2016)	Plant Morphology	93.0
CNN- University of Pavia (Li et al., 2017)	Agricultural Landscape Identification	92.27
CNN-Salinas (Li et al., 2017)	Agricultural Landscape Identification	89.28
CNN-Indian Pines (Li et al., 2017)	Agricultural Landscape Identification	86.44
SegNet-Basic-Encoder Addition (Badrinarayanan et al., 2017)	General Image Segmentation	84.2

detected 85.2% of the fields our ResNet/UNet model detected (Xie and Lark, 2021). In other words, 14.8% of contour field pixels detected by our model did not show up on the LANID map.

4. Discussion

4.1. How does our model compare with other studies?

This work is part of a package of approaches to understand rice production systems in the US mid-south, from rice field mapping in google earth engine with multi-year training datasets (Liang et al., 2019) to using machine learning techniques to identify contour levee fields (Liang et al., 2021), and now testing different levee-identification approaches. Overall our system shows preliminary success in segmenting contour levee irrigation fields, as evidenced through the ROC curve, overall accuracy, F1, IoU and BER scores, receiving 99.1%, 97.0%, 92.4%, 85.8%, and 54.8% respectively. The model performed on par with many of the aforementioned models in the field of agriculture landscape segmentation and surpassed the highest model we found with a similar goal of striation detection within a segmented image (Table 2).

The comparison with LANID data helps illustrate the role of spatial resolution in classification. Theoretically, all of the contour-levee fields our model detects should be a subset of the fields LANID detected. Further investigation revealed that a large number of the pixels belong to a single field with very apparent contour levees (Supplemental Fig. 5b.) that were thus likely misclassified by LANID. Any remaining pixels fall on the edges of other fields and are likely mismatched due to the difference in resolution. Coarser resolution imagery (e.g., 30 m for LANID) will have blurrier edges between fields, and the field boundaries are likely to blend with the crop fields, which causes overestimation in the size of the actual field (Supplemental Fig. 5a). Thus, a field boundary detection step using higher resolution imagery is likely important, and could follow recent model development in this area (e.g., Jong et al., 2022).

Several challenges remain for model validation (Supplemental Figs. 4–6). Upon calculating the true positive and true negative scores of the Lonoke testing set, we can see that the model accurately discerns what is labeled as background, correctly classifying background 98.6% of the time, but it is slightly weaker in classifying contour fields (90.3%). There are a few scenarios that could cause this weaker detection rate, some of which are demonstrated in Supplemental Fig. 6. First, structures such as tractor paths or roads that may cut through the field could present as levees. Additionally, the model easily discerns formations such as buildings or streams as not being a field at all, let alone one irrigated with contour levees. However, fields using one of the other irrigation systems or even a field that was recently harvested and still has tractor tracks present would be more difficult to classify. Though our imagery was largely captured during times with a fully closed canopy, these field management effects are sometimes still visible, as discovered in the tests detailed in Section 3.2.3. Additionally, the

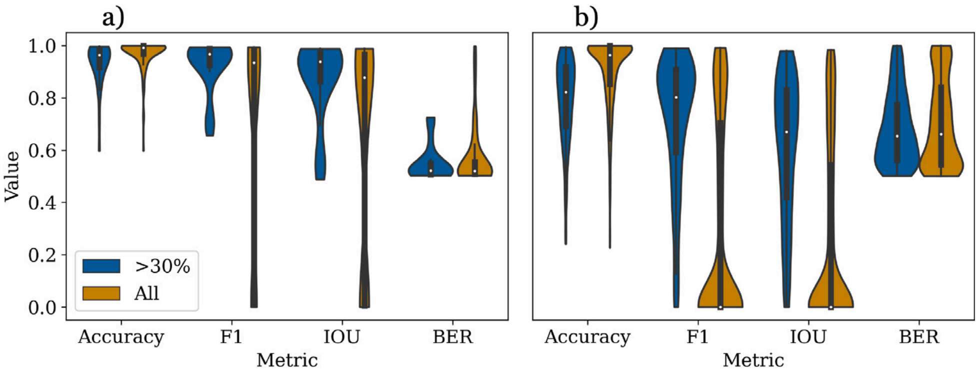


Fig. 9. Shows the accuracy, F1, IoU, and BER performance metrics for (a) the 240 sub-tile model development testing set from Lonoke County and (b) the 1375 sub-tile scalability test set drawn from the 55 testing tiles across the state. Performance metrics are shown to compare a case with a subset of contour-dense tiles (i.e., at least 30% of the image label contains a contour field) vs. the full set of testing tiles. The contour-dense set (i.e., ">30%") contains 33 images in plot a and 258 in plot b, and its metrics typically have a smaller range and smoother distribution compared to the inclusive set ("All"), thus insinuating a higher consistency amongst its subtiles. In the violin plots, the white dot represents the median of the samples, the thick portion of the line represents the interquartile range, and the thin portion represents the 1.5× interquartile range. The kernel density in blue or orange is represented by the shape around the box plot.

use of the higher resolution commercial satellites such as Planet could give us year-round imagery which may alleviate these issues, when they are available. It is also important to note that the model's confusion may be a byproduct of filtering out non-contour levee fields in the training process (Supplemental Fig. 6). Had we preserved the other irrigation types, the more ambiguous areas of the image would likely have had labels, thus providing additional classes, such as straight levees, for the model to learn rather than relying solely on contour levees.

Another potential cause of the weaker true positive rate of some of these images is the lack of positive pixels within a sub-tile. The labeling in 1b–5b of Supplemental Fig. 6 shows that these images contain only a small portion of a much larger contour field. Thus, due to the label corresponding to such a small area of the field, it is possible that no levees were even present in the image, but rather only part of the contour-leveed field. This issue can be attributed to the way the images were broken into a grid, and could be improved in future studies with a different type of labeling and image segmentation process (e.g., with fraction thresholds).

Upon further inspection, we noticed that the worst-performing images from the 55 testing tiles had less than 30% of the label occupied by contour labels (some example images demonstrating this impact are in Supplemental Fig. 6). It is likely that the image subsetting function used to subset the large 5000 × 5000 tiles into smaller sub-images skews the metric values, as the only positive case present in the label is a small portion of a field on the edge of the image in these cases (e.g., row b in Supplemental Fig. 6). We now compare the accuracy, F1, IoU, and BER for sets including and excluding samples with less than 30% of the label containing contour fields for both the Lonoke testing tiles and the 55 tiles from across eastern Arkansas (Fig. 9). The exclusive set, meaning tiles with greater than 30% contour levee pixels, received 77.2%, 68.2%, 51.7%, and 72.6% for Overall Accuracy, F1, IoU, and BER respectively (blue sets in Fig. 9b). Contrary to its inclusive counterpart discussed in Section 4.1, the Lonoke testing set with at least 30% of the label containing a contour field received an Accuracy, F1, IoU, and BER of 93.2%, 93.0%, 86.8%, and 54.8% respectively. Thus, while not uniformly better (i.e., across all four metrics), we recommend setting a threshold value of 30% labeled for each tile to minimize edge effects in future studies. Alternatively, instead of breaking images into a grid, a form of sliding window could be implemented. This approach could alleviate edge cases by ensuring that a more complete image of the field will also be seen by the model. Additionally, this approach would eliminate the risk that a small but complete contour levee field that occupies less than 30% of the pixels in a subtile could be discarded.

A potential candidate for comparison with our study is the shifted windows Transformer (Swin) that brings a hierarchical to image processing (Liu et al., 2021). This architecture's hierarchical approach builds on existing transformer models, which rely on fixed-sized patches of a given image, by dividing the image into non-overlapping patches of varying sizes. The model then refines its understanding of the input image at varying scales: smaller patches at the lower hierarchical levels and larger ones at the higher levels. This process is quite different than how the ResNet portion of our model convolves overlapping patches of an image. Although it is not clear whether CNN or transformer-based neural networks are more suited for a given task, there have been several studies leveraging Swin/UNet architectures for remote sensing applications in recent years. Similar to our use of ResNet for the extraction of feature maps, these studies use the Swin transformer for the same purpose. For instance, two recent studies proposed Swin/Unet hybrid architectures (He et al., 2022; Gao et al., 2021) to evaluate different datasets: the Vaihingen dataset and the Potsdam dataset (Chen et al., 2014b). The first contains 33 aerial images of varying sizes with near-infrared, red, and green bands. The images are labeled according to many different object classes such as buildings, roads, and trees. The second dataset contains 38 images of uniform size with red, green, and blue color bands. This dataset was also labeled according to many urban development-related classes such as buildings and roads. For the Vaihingen dataset, He et al. (2022) found mean IoU and mean F1 scores of 67.32% and 79.94% while Gao et al. (2021) achieved 66.66% and 78.67% for the same metrics. With respect to the Potsdam dataset, He et al. (2022) achieved 75.97% mean IoU and 86.13% mean F1 scores whereas Gao et al. (2021) yielded 71.46% and 82.08%. Although these values are slightly lower than what we found in this study, given the change in subject area from urban development to crop field mapping, the Swin transformer could provide valuable insight into the most applicable architectures for mapping irrigation patterns.

4.2. *What drives model sensitivity in high-resolution mapping?*

The model's capability is retained even with variations and noise in the data. Reductions in resolution and other visual disturbances created minor changes in the performance metrics; the greatest change was 0.203 in the AUC when the gaussian noise with a 0.4 spread was added to the input imagery. The model seems to maintain its ability up to a resolution reduction to 10 m at which the performance begins to degrade. This larger resolution threshold (i.e., much greater than levee thickness) calls into question what characteristic of the labeled fields the model uses to make its prediction. Perhaps it weighs the curvature and overall pattern of the field more heavily than the thickness of the individual levees. In practice, this finding implies that the ResNet/Unet model is suitable for use with Sentinel-like images or finer (de Moura et al., 2022), but not Landsat Pouliot et al. (2019) or MODIS (Zhang et al., 2021). These outcomes are consistent with a recent machine learning approach to rice field identification and mapping in Pakistan that demonstrates at least an 8% F1-score improvement when using Sentinel-2 compared to Landsat and MODIS (Waleed et al., 2022). The model is robust against different levels of simulated clouds but handles differences in imaging systems better than environmental differences. The introduction of cloud cover, a likely scenario in practice, yielded results comparable to that of a 25% reduction in image resolution. Static noise, however, proved to minimally affect the results, likely because the static is spread over the entire image rather than in one concentrated area. Thus, the model seems to mostly disregard the affected pixel in favor of the classification of the neighboring pixels. Insensitivity to cloud cover further bolsters the model's suitability for high-resolution satellite imagery, as cloud cover typically presents a challenge with this data source (Zhang et al., 2020). However, the stronger sensitivity to gaussian static could present a challenge due to random noise being an often overlooked limitation of aerial and satellite imagery (Anikeeva and Chibunichev, 2021).

4.3. Theoretical implications and practical applications

Precision farming requires a mix of historical and contemporary information, including harvest estimates, land use history, farmland sales, and landscape modification (Finger et al., 2019). Thus, our model could provide value to non-profit, public sector, or commercial actors who aid farmers in making informed decisions when dealing with critical natural resources. This identification could help map out and prioritize water and carbon conservation programs, due to the tight interplay between irrigation and the water and carbon cycles in these landscapes (Runkle et al., 2019; Moreno-García et al., 2021; Henry et al., 2016). Ongoing identification work can help monitor the success of programs that seek to induce practice changes or to aid in new program development by identifying regions slower to implement conservation land grading methods. Indeed, changing contour levee irrigation to other, more efficient delivery approaches such as multiple-inlet rice irrigation can improve irrigation efficiency and reduce water use by up to 24% (Massey et al., 2018). Improved levee detection could also improve hydrological models that require a clear understanding of water flow pathways through these agricultural environments, including estimates of seepage and percolation (LaHue and Linquist, 2021), flood mitigation potential (Chen et al., 2014a), and nitrogen runoff associated with irrigation (Ouyang et al., 2020; Kim et al., 2021). Better landform classification can also be useful in a geomorphological context, improving digital elevation models with implications in hydrology and erosion studies (Li et al., 2020).

4.4. Future research directions

Since this research has primarily focused on methods development, there is still work to be done to further assess the scalability of the model. First, we need to examine the temporal scalability. All of the data we used in this study was acquired in a 26 day span in 2015, thus, the model should be tested on data from years other than 2015. This inter-annual test would provide necessary insight into how the model will perform in practice. Second, though we tested the model with images outside of the training county within Arkansas with the 55 testing tiles, a complete assessment of the model's performance outside of Arkansas would demonstrate the extent of the geographic scalability. A state such as California, which ranks second in terms of rice production behind Arkansas (Illsley, 2020), would provide ample data for us to test while also being geographically distinct from Araknsas. Third, the model's transferability into non-native resolutions should also be validated. We speculate that the model would perform well on Sentinel-like imagery, but further testing could prove this hypothesis. Lastly, this product has the potential to enable whole-region classification of irrigation systems. Thus, we have made the project open-source to make it both accessible to the people who need it and to allow people to make the necessary changes to suit their specific means. The code is open source and available at https://github.com/dsdale/CropContourLeveeMapping on publication.

5. Conclusion

In this study we investigated the use of a neural network to identify rice contour levee systems from aerial imagery. The results of our approach are promising, as they show that our model maintains a high level of performance despite added noise and reductions in image resolution. This open-source model lays the groundwork for a future region-wide landscape classification system. Our results demonstrate that we are on track to support irrigation and water resource management in the Midsouth USA region. Additionally, the model architecture is robust enough that it is moderately scalable to different environments provided the resolution of the imagery used is greater than 10 m. The noise addition analyses show that the model is moderately accepting of environmental variance such as cloud cover. The work enables analysis of landscape use patterns to drive models of agricultural productivity and sustainability in rice growing regions by providing a county-wide proof of concept and improving approximately 5.6% over existing methods. It also offers a guide for approaching any type of deep learning pattern identification problem, no matter the discipline.

CRediT authorship contribution statement

Dakota S. Dale: Methodology, Investigation, Data curation, Writing – original draft. **Lu Liang:** Writing – review & editing, Conceptualization, Supervision, Visualization. **Liheng Zhong:** Software, Methodology, Visualization. **Michele L. Reba:** Writing – review & editing, Resources. **Benjamin R.K. Runkle:** Conceptualization, Resources, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the United States National Science Foundation under CBET Award 1752083, NASA Water Resources Award 80NSSC22K0923, and the U.S. Geological Survey under Cooperative Agreement G20AC00448 and G21AC10729. We thank Drew McFall at the University of Arkansas and Abolfazl Meyarian at the University of North Texas for work on image classification, and many student volunteers for their annotation work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compag.2023.107954.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467.

Anikeeva, I., Chibunichev, A., 2021. Random noise assessment in aerial and satellite images. ISPRS - Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 43B2, 771–775.

Atwill, R.L., Krutz, L.J., Bond, J.A., Golden, B.R., Spencer, G.D., Bryant, C.J., Mills, B.E., Gore, J., 2020. Alternate wetting and drying reduces aquifer withdrawal in Mississippi rice production systems. Agron. J. 112 (6), 5115–5124.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A deep convolutional Encoder-Decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26 (5), 341–358. http://dx.doi.org/10.1080/10106049.2011.562309.

Bouman, B.A.M., Humphreys, E., Tuong, T.P., Barker, R., 2007. Rice and water. In: Sparks, D.L. (Ed.), Advances in Agronomy, Vol. 92. Academic Press, pp. 187–237.

Bouman, B.A.M., Tuong, T.P., 2001. Field water management to save water and increase its productivity in irrigated lowland rice. Agricult. Water Manag. 49 (1), 11–30.

Bradley, A.P., 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 30 (7), 1145–1159. http://dx.doi.org/10.1016/S0031-3203(96)00142-2, URL: https://www.sciencedirect.com/science/article/pii/S0031320396001422.

Campbell, B.M., Beare, D.J., Bennett, E.M., Hall-Spencer, J.M., Ingram, J.S.I., Jaramillo, F., Ortiz, R., Ramankutty, N., Sayer, J.A., Shindell, D., 2017. Agriculture production as a major driver of the Earth system exceeding planetary boundaries. Ecol. Soc. 22 (4).

Carrijo, D.R., Lundy, M.E., Linquist, B.A., 2017. Rice yields and water use under alternate wetting and drying irrigation: A meta-analysis. Field Crops Res. 203, 173–180.

Carroll, S.R., Le, K.N., Moreno-García, B., Runkle, B.R.K., 2020. Simulating Soybean–Rice rotation and irrigation strategies in arkansas, USA using APEX. Sustain. Sci. Pract. Policy 12 (17), 6822.

Chen, S.K., Chen, R.S., Yang, T.Y., 2014a. Application of a tank model to assess the flood-control function of a terraced paddy field. Hydrol. Sci. J. 59 (5), 1020–1031.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014b. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv preprint arXiv:1412.7062.

Chlapecka, J.L., Hardke, J.T., Roberts, T.L., Mann, M.G., Ablao, A., 2021. Scheduling rice irrigation using soil moisture thresholds for furrow irrigation and intermittent flooding. Agron. J. 113 (2), 1258–1270. http://dx.doi.org/10.1002/agj2.20600, URL: https://acsess.onlinelibrary.wiley.com/doi/abs/10.1002/agj2.20600.

Curran, Hay, 1986. The importance of measurement error for certain procedures in remote sensing at optical wavelengths. Photogramm. Eng. Remote Sens. 52 (2), 229–241.

de Moura, N.V.A., de Carvalho, O.L.F., Gomes, R.A.T., Guimarães, R.F., de Carvalho Júnior, O.A., 2022. Deep-water oil-spill monitoring and recurrence analysis in the Brazilian territory using Sentinel-1 time series and deep learning. Int. J. Appl. Earth Obs. Geoinf. 107, 102695.

Dubbs, A., 2021. Test set sizing via random matrix theory. http://dx.doi.org/10.48550/ARXIV.2112.05977, URL: https://arxiv.org/abs/2112.05977.

Elihos, A., Alkan, B., Balci, B., Artan, Y., 2018. Comparison of image classification and object detection for passenger seat belt violation detection using NIR & RGB surveillance camera images. In: 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance. AVSS, pp. 1–6.

Finger, R., Swinton, S.M., El Benni, N., Walter, A., 2019. Precision farming at the nexus of agricultural production and the environment. Annu. Rev. Resour. Econ. 11 (1), 313–335.

Flach, P.A., 2016. ROC analysis. In: Sammut, C., Webb, G.I. (Eds.), Encyclopedia of Machine Learning and Data Mining. Springer US, Boston, MA, pp. 1–8. http://dx.doi.org/10.1007/978-1-4899-7502-7_739-1.

Fu, G., Liu, C., Zhou, R., Sun, T., Zhang, Q., 2017. Classification for high resolution remote sensing imagery using a fully convolutional network. Remote Sens. 9 (5), http://dx.doi.org/10.3390/rs9050498, URL: https://www.mdpi.com/2072-4292/9/5/498.

Gao, L., Liu, H., Yang, M., Chen, L., Wan, Y., Xiao, Z., Qian, Y., 2021. STransFuse: Fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 10990–11003.

Grinblat, G.L., Uzal, L.C., Larese, M.G., Granitto, P.M., 2016. Deep learning for plant identification using vein morphological patterns. Comput. Electron. Agric. 127, 418–424. http://dx.doi.org/10.1016/j.compag.2016.07.003.

Hawkins, D.M., 2004. The problem of overfitting. J. Chem. Inf. Comput. Sci. 44 (1), 1–12. http://dx.doi.org/10.1021/ci0342472.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 770–778. http://dx.doi.org/10.1109/CVPR.2016.90.

He, X., Zhou, Y., Zhao, J., Zhang, D., Yao, R., Xue, Y., 2022. Swin transformer embedding UNet for remote sensing image semantic segmentation. IEEE Trans. Geosci. Remote Sens. 60, 1–15. http://dx.doi.org/10.1109/TGRS.2022.3144165.

Henry, C.G., Hirsh, S.L., Anders, M.M., Vories, E.D., Reba, M.L., Watkins, K.B., Hardke, J.T., 2016. Annual irrigation water use for Arkansas rice production. J. Irrig. Drain. Eng. 142 (11), 05016006.

Houborg, R., McCabe, M.F., 2016. High-resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. Remote Sens. 8 (9), 768.

Hsu, C.M., Hsu, C.C., Hsu, Z.M., Shih, F.Y., Chang, M.L., Chen, T.H., 2021. Colorectal polyp image detection and classification through grayscale images and deep learning. Sensors 21 (18).

Illsley, C.L., 2020. The leading rice growing states in the United States. WorldAtlas, https://www.worldatlas.com/articles/the-leading-rice-growing-states-in-the-united-states.html. (Accessed 29 November 2022).

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (Eds.), Proceedings of the 32nd International Conference on Machine Learning. In: Proceedings of Machine Learning Research, vol. 37, PMLR, Lille, France, pp. 448–456.

Japkowicz, N., 2006. Why question machine learning evaluation methods. In: AAAI Workshop on Evaluation Methods for Machine. aaai.org.

Jong, M., Guan, K., Wang, S., Huang, Y., Peng, B., 2022. Improving field boundary delineation in ResUNets via adversarial deep learning. Int. J. Appl. Earth Obs. Geoinf. 112, 102877. http://dx.doi.org/10.1016/j.jag.2022.102877, URL: https://www.sciencedirect.com/science/article/pii/S1569843222000796.

Joseph, V.R., 2022. Optimal ratio for data splitting. Statistical Analysis and Data Mining: The ASA Data Science Journal 15 (4), 531–538. http://dx.doi.org/10.1002/sam.11583, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sam.11583, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/sam.11583.

Jung, A.B., Wada, K., Crall, J., Tanaka, S., Graving, J., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Borovec, J., et al., 2020. Imgaug. https://github.com/aleju/imgaug.

Kim, D.H., Jang, T., Hwang, S., Jeong, H., Choi, S.-K., 2021. APEX-Paddy model simulation of hydrology, total nitrogen, and rice yield for different agricultural activities in paddy fields. Paddy Water Environ. 19 (4), 609–622.

Kubo, M., Purevdorj, M., 2004. The Future of Rice Production and Consumption, Vol. 35. Technical Report 856-2016-57064, pp. 128–142.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782. http://dx.doi.org/10.1109/LGRS.2017.2681128.

LaHue, G.T., Linquist, B.A., 2021. The contribution of percolation to water balances in water-seeded rice systems. Agricult. Water Manag. 243, 106445.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.

Lark, T.J., Schelly, I.H., Gibbs, H.K., 2021. Accuracy, bias, and improvements in mapping crops and cropland across the United States using the USDA cropland data layer. Remote Sens. 13 (5), 968.

Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., 2021. A survey of convolutional neural networks: Analysis, applications, and prospects. IEEE Trans. Neural Netw. Learn. Syst. PP.

Li, W., Wu, G., Zhang, F., Du, Q., 2017. Hyperspectral image classification using deep pixel-pair features. IEEE Trans. Geosci. Remote Sens. 55 (2), 844–853. http://dx.doi.org/10.1109/TGRS.2016.2616355.

Li, S., Xiong, L., Tang, G., Strobl, J., 2020. Deep learning-based approach for landform classification from integrated data sources of digital elevation model and imagery. Geomorphology 354, 107045.

Liang, L., Meyarian, A., Yuan, X., Runkle, B.R., Mihaila, G., Qin, Y., Daniels, J., Reba, M.L., Rigby, J.R., 2021. The first fine-resolution mapping of contour-levee irrigation using deep bi-stream convolutional neural networks. Int. J. Appl. Earth Obs. Geoinf. 105, 102631. http://dx.doi.org/10.1016/j.jag.2021.102631, URL: https://www.sciencedirect.com/science/article/pii/S030324342100338X.

Liang, L., Runkle, B.R.K., Sapkota, B.B., Reba, M.L., 2019. Automated mapping of rice fields using multi-year training sample normalization. Int. J. Remote Sens. 40 (18), 7252–7271. http://dx.doi.org/10.1080/01431161.2019.1601286.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

Liu, Z., Wu, J., Fu, L., Majeed, Y., Feng, Y., Li, R., Cui, Y., 2020. Improved kiwifruit detection using pre-trained VGG16 with RGB and NIR information fusion. IEEE Access 8, 2327–2336.

Louhaichi, M., Borman, M.M., Johnson, D.E., 2001. Spatially located platform and aerial photography for documentation of grazing impacts on wheat. Geocarto Int. 16 (1), 65–70.

Martins, V.S., Kaleita, A.L., Gelder, B.K., da Silveira, H.L.F., Abe, C.A., 2020. Exploring multiscale object-based convolutional neural network (multi-OCNN) for remote sensing image classification at high spatial resolution. ISPRS J. Photogramm. Remote Sens. 168, 56–73.

Mas, J.F., Soares Filho, B., Pontius, R.G., Farfán Gutiérrez, M., Rodrigues, H., 2013. A suite of tools for ROC analysis of spatial models. ISPRS Int. J. Geo-Inf. 2 (3), 869–887. http://dx.doi.org/10.3390/ijgi2030869, URL: https://www.mdpi.com/2220-9964/2/3/869.

Massey, J., 2023. Private communication.

Massey, J.H., Smith, M.C., Vieira, D.A.N., Adviento-Borbe, M.A., Reba, M.L., Vories, E.D., 2018. Expected irrigation reductions using Multiple-Inlet rice irrigation under rainfall conditions of the lower Mississippi river valley. J. Irrig. Drain. Eng. 144 (7), 04018016.

Meyarian, A., Yuan, X., Liang, L., Wang, W., Gu, L., 2022. Gradient convolutional neural network for classification of agricultural fields with contour levee. Int. J. Remote Sens. 43 (1), 75–94. http://dx.doi.org/10.1080/01431161.2021.2003467.

Moreno-García, B., Coronel, E., Reavis, C.W., Suvočarev, K., Runkle, B.R., 2021. Environmental sustainability assessment of rice management practices using decision support tools. J. Clean. Prod. 315, 128135. http://dx.doi.org/10.1016/j.jclepro.2021.128135, URL: https://www.sciencedirect.com/science/article/pii/S0959652621023532.

Nguyen, Q.H., Ly, H.B., Ho, L.S., Al-Ansari, N., Van Le, H., Tran, V.Q., Prakash, I., Pham, B.T., 2021. Influence of data splitting on performance of machine learning models in prediction of shear strength of soil. Math. Probl. Eng. 2021.

Norman, R.J., Moldenhauer, K.A.K., 2016. B.R. Wells Arkansas Rice Research Studies 2015. In: Arkansas Agricultural Experiment Station Research Series, https://scholarworks.uark.edu/aaesser/20.

Norman, Moldenhauer, 2019. B.R. Wells Arkansas Rice Research Studies. https://scholarworks.uark.edu/aaesser/154.

Ouyang, W., Wei, P., Gao, X., Srinivasan, R., Yen, H., Xie, X., Liu, L., Liu, H., 2020. Optimization of SWAT-Paddy for modeling hydrology and diffuse pollution of large rice paddy fields. Environ. Model. Softw. 130, 104736.

Pascale, D., 2003. A review of rgb color spaces... from xyy to r'g'b'. Babel Color 18, 136–152.

Pouliot, D., Latifovic, R., Pasher, J., Duffe, J., 2019. Assessment of convolution neural networks for wetland mapping with landsat in the Central Canadian Boreal Forest Region. Remote Sens. 11 (7), 772.

Reba, M.L., Massey, J.H., 2020. Surface irrigation in the lower Mississippi river basin: Trends and innovations. Trans. ASABE.

Reid, A.M., Chapman, W.K., Prescott, C.E., Nijland, W., 2016. Using excess greenness and green chromatic coordinate colour indices from aerial images to assess lodgepole pine vigour, mortality and disease occurrence. Forest Ecol. Manag. 374, 146–153.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation, CoRR abs/1505.04597. URL: http://arxiv.org/abs/1505.04597.

Rosa, L.G., Zia, J.S., Inan, O.T., Sawicki, G.S., 2021. Machine learning to extract muscle fascicle length changes from dynamic ultrasound images in real-time. PLoS One 16 (5), e0246611.

Runkle, B.R.K., Suvočarev, K., Reba, M.L., Reavis, C.W., Smith, S.F., Chiu, Y.L., Fong, B., 2019. Methane emission reductions from the alternate wetting and drying of rice fields detected using the eddy covariance method. Environ. Sci. Technol. 53 (2), 671–681. http://dx.doi.org/10.1021/acs.est.8b05535.

Salamati, N., Larlus, D., Csurka, G., Süsstrunk, S., 2012. Semantic image segmentation using visible and Near-Infrared channels. In: Computer Vision – ECCV 2012. Workshops and Demonstrations. Springer Berlin Heidelberg, pp. 461–471.

Sasaki, Y., 2007. The truth of the F-measure. Teach. Tutor. Mater.

Shew, A.M., Nalley, L.L., Durand-Morat, A., Meredith, K., Parajuli, R., Thoma, G., Henry, C.G., 2021. Holistically valuing public investments in agricultural water conservation. Agricult. Water Manag. 252, 106900.

Smith, M.C., Massey, J.H., Branson, J., Epting, J.W., Pennington, D., Tacker, P.L., Thomas, J., Vories, E.D., Wilson, C., 2007. Water use estimates for various rice production systems in Mississippi and Arkansas. Irrig. Sci. 25 (2), 141–147.

Stevens, G., Rhine, M., Heiser, J., 2018. Rice production with furrow irrigation in the Mississippi river delta region of the USA. In: Shah, F., Khan, Z.H., Iqbal, A. (Eds.), Rice Crop. IntechOpen, Rijeka, http://dx.doi.org/10.5772/intechopen.74820 (Chapter 5).

Stubenrauch, C.J., Rossow, W.B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B.C., Menzel, W.P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., Zhao, G., 2013. Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX radiation panel. Bull. Am. Meteorol. Soc. 94 (7), 1031–1049.

Tieleman, T., Hinton, G., et al., 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Netw. Mach. Learn. 4 (2), 26–31.

Torres, L., Reutter, J.Y., Lorente, L., 1999. The importance of the color information in face recognition. In: Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), Vol. 3. pp. 627–631.

USDA, 2017. NAIP imagery. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-programs/naip-imagery/. (Accessed 20 January 2022).

USDA-NASS, 2021. Census of agriculture. Quick stats. 2020 survey. National Agriculture Statistics Service. Available at https://www.nass.usda.gov/AgCensus/index.php.

van Beers, E.O., Marco A. Wiering, F., 2019. Deep neural networks with intersection over union loss for binary image segmentation. https://www.scitepress.org/Papers/2019/73475/73475.pdf. (Accessed 22 July 2020).

Vieira, S.M., Kaymak, U., Sousa, J.M.C., 2010. Cohen's kappa coefficient as a performance measure for feature selection. In: International Conference on Fuzzy Systems. pp. 1–8.

Vories, E.D., Stevens, W.E., Tacker, P.L., Griffin, T.W., Counce, P.A., 2013. Rice production with center pivot irrigation. Appl. Eng. Agric. 29 (1), 51–60.

Waleed, M., Mubeen, M., Ahmad, A., Habib-Ur-Rahman, M., Amin, A., Farid, H.U., Hussain, S., Ali, M., Qaisrani, S.A., Nasim, W., Javeed, H.M.R., Masood, N., Aziz, T., Mansour, F., El Sabagh, A., 2022. Evaluating the efficiency of coarser to finer resolution multispectral satellites in mapping paddy rice fields using GEE implementation. Sci. Rep. 12 (1), 13210.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sens. Environ. 236, 111402.

Xie, Y., Lark, T.J., 2021. Mapping annual irrigation from landsat imagery and environmental variables across the conterminous United States. Remote Sens. Environ. 260, 112445.

Zaheer, R., Shaziya, H., 2019. A study of the optimization algorithms in deep learning. In: 2019 Third International Conference on Inventive Systems and Control. ICISC, pp. 536–539.

Zhang, C., Marzougui, A., Sankaran, S., 2020. High-resolution satellite imagery applications in crop phenotyping: An overview. Comput. Electron. Agric. 175, 105584.

Zhang, L., Ren, Z., Dong, R., Xu, B., Fu, H., 2021. Monitoring daily nighttime light based on modis and deep learning: A Belgium case study. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. pp. 5032–5035.

Zhao, X., Yuan, Y., Song, M., Ding, Y., Lin, F., Liang, D., Zhang, D., 2019. Use of unmanned aerial vehicle imagery and deep learning UNet to extract rice lodging. Sensors 19 (18).