ESTIMATION OF CROP RESIDUE COVER IN HIGH-RESOLUTION RGB IMAGES USING FEATURES FROM A PRE-TRAINED CONVOLUTIONAL NEURAL NETWORK

A Thesis

Presented to

The Faculty of the Graduate School

at the University of Missouri - Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

TIMOTIUS ANDREAN PATRICK LAGAUNNE

Dr. John Lory, Thesis Supervisor

MAY 2022

The undersigned, appointed by the dean of the Graduate School,

have examined the Thesis entitled

Estimation of crop residue cover in high-resolution RGB Images using features from a pre-trained Convolutional Neural Network

Presented by Timotius Andrean Patrick Lagaunne

A candidate of the degree of

Master of Science

And hereby certify that, in their opinion, it is worthy of acceptance

_____

Dr. John Lory

_____

Dr. Harley Naumann

_____

Dr. Guilherme DeSouza

Acknowledgements

Firstly, I would like to thank Jesus for loving me, supporting me, being with me and always giving me hope during my process of finishing my degree. I am grateful for his grace and love that manifest through everyone surrounding me.

Secondly, I would like to deeply thank Dr. John Lory for the past four and half years that give me chance to recover from my mental health and give me opportunity to finish my thesis and my degree. Without his support, patience, knowledge and guidance, I would not be able to finish the degree. Also, I want to thank Dr. Harley Naumann and Dr. Guilherme DeSouza for their guidance throughout my process as committee members.

Thirdly, I would like to thank Dr. David Mendoza Cozatl for helping me whenever I had difficulties in regards to my mental health and struggled with administrative things related to my degree and semester extensions. Thank you for giving advice and being supportive. I also thank Parth Upadhyay, my fellow graduate student, for friendship and assistance. I wish you the best as fellow graduate student. I also would thank David Kleinsorge, Krystal Burkett-Tysdal, and Theresa Musket for helping me a lot during data collection. I learned a lot about how to work as a team from all of you. I also want to thank Dewi Kharismawati, Isaac Shee, Ahmed Krgo, and Robert Philips who gave me the chance to learn how to guide fellow workers and friends as we worked together.

Finally, I would like to thank my family that supported me from far away and my friends and family at University of Missouri in ICF (International Christian Fellowship) and MISA (Missouri Indonesian Student Association). Thank you for supporting me and being a friend especially during times I struggled with mental health.

Table of Contents

List of Tables

Table **Page**

List of Figures

Figure **Page**

Abstract

Plant residue on the soil surface increases the sustainability food and fiber production in agricultural systems. Automated assessments of residue cover based on imagery has the potential to reduce labor and human bias associated with in-field measurements. Our objective was to evaluate the capacity of a transfer learning strategy to improve estimates of residue cover derived from high-resolution RGB images. The imagery for the project was collected from 88 locations in 40 row crop fields in five Missouri counties between mid-April and early July in 2018 and 2019. At each field location, 50 contiguous 0.3 m x 0.2 m region of interest (ROI) images (ground sampling distance of 0.014 cm pixel^{-1}) were extracted from imagery resulting in a dataset of 4,400 ROI images; 3,000 used for cross validation and training (data collected in 2018) and 1,400 used for testing (data collected in 2019). The percent residue for each ROI image (ground truth) was determined by a bullseye grid method (n = 100). Features were extracted from ROI images using the VGGNet16 model, a convolutional neural network model. To reduce feature numbers, we averaged the features based on each kernel resulting 1,472 feature dataset per ROI. After the extraction, we compared three feature selection strategies: recursive feature elimination support vector machine classification (RFE-SVM), sequential forward feature selection classification (SFFS-SVM) and forward regression feature selection (FRFS). Best locations outcomes were obtained with RFE-SVM (r^2 = 0.93, MAE = 4.9, with three outliers) and FRFS (r^2 = 0.94, MAE = 5.2, with two outliers). The three models had no apparent pattern of correlation among selected features and limited overlap in outliers suggesting unique characteristics among the three selected feature sets. These results were superior to previous research based the same data set

using 70 manually extracted known features. This suggested that transfer learning through features extracted from VGGNet16 pre-trained on ImageNet was a successful strategy for estimating residue cover. This research also confirmed the utility of high-resolution RGB imagery to quantify residue cover in agricultural systems.

Introduction

Retention of plant residue on the soil surface provides significant eco-service benefits including reduced evaporation, protection of soil from water and wind erosion, and improved soil structure and infiltration through increases in soil organic matter (Bronick & Lal, 2005; Cherubin et al., 2018; Ranaivoson et al., 2017; Searle & Bitnere, 2017; Singh & Rengel, 2007). The importance of residue in agricultural systems is represented by its role as a key input in soil erosion models such as the Watershed Erosion Prediction Project (WEPP), Revised Universal Soil Loss Equation (RUSLE) and RUSLE2 (Dabney, Yoder, Vieira, & Bingner, 2011; Flanagan, Gilley, & Franti, 2007; Weltz et al., 2020). The Food Security Act of 1985 (Glaser, 1985) established a requirement to maintain "sustainable erosion rates on cropland, hay land and pasture" defined as "highly erodible land" (HEL) with the objective of protecting the Nation's long-term capability to produce food and fiber. Residue assessments are part of the documentation of farmer HEL compliance; when farmers fail to comply with requirements from the Food Security Act of 1985 they may lose access to assistance payments, conservation program benefits, and other Federal subsidies.

The Natural Resource Conservation Service (NRCS) is the primary agency documenting farmer HEL compliance. They assess the percentage of plant residue cover as a critical compliance variable in Conservation and Compliance assessments (USDA-NRCS, 2011) To determine percent residue cover, trained technical staff visit fields and make three to five determinations of residue cover in representative areas using the line-transect method. (USDA-NRCS, 2011) The line-transact method requires NRCS personnel to read 100 points evenly distributed along a 15.2- or 30.4-m tape laid at a 45-

degree angle to the direction of farming (USDA-NRCS, 2011). The NRCS manual emphasizes that readers should view transect points from directly above and only count residue with a surface coverage diameter greater than or equal to 2.4 mm and that touches the pin-head sized transect points on the tape. In practice, the line-transect method requires careful attention to correctly read the tape and is prone to reader bias to overestimate residue (Laamrani, Joosse, & Feisthauer, 2017; Laflen, Amemiya, & Hintz, 1981; Lory et al., 2021; Richards, Walter, & Muck, 1984). Alternatives to the line-transect method that also require human judgement include visual estimate, point intercept, meter stick, spiked wheel, photograph comparison, and photographic-grid methods(Dickey, Shelton, Meyer, & Fairbanks, 1989; Laamrani et al., 2017; Laflen et al., 1981; Morrison Jnr, Huang, Lightle, & Daughtry, 1993).

One advantage of automated systems is that they can limit human judgement and bias. Most automated systems deploy optical sensors on aircraft, satellite platforms, or unmanned aerial vehicles (UAV's) to collect RGB, multispectral, and/or hyperspectral imagery to estimate residue cover (Zheng, Campbell, Serbin, & Galbraith, 2014). Most research has focused on the development of indices based on the dissimilarity of the reflection and absorption of multiple wavelengths from the images to detect residue (Chi & Crawford, 2014; Gausman et al., 1975; Hively et al., 2019; Pacheco & McNairn, 2010; Quemada, Hively, Daughtry, Lamb, & Shermeyer, 2018; Quemada & Daughtry, 2016; Zheng et al., 2014). Alternatively, traditional machine learning methods have been used to build a prediction model from RGB images based on known color, shape and texture features extracted from the image (Kavoosi et al., 2020; Najafi, Feizizadeh, & Navid, 2021; Upadhyay, Lory, DeSouza, Lagaunne, & Spinka, 2022). The recent interest in

RGB imagery is likely because of the easy access to low cost RGB images from tools like UAV's and smart phones.

Automated systems using RGB images have applied a diverse set of strategies using limited data sets with mixed results. Baeur and Strauss (2014) used Object Based Image Analysis (OBIA) that applied segmentation and a rule-based approach to estimate residue cover. Using a dataset of 61 images that included four different soil colors, and soil cover ranging from 0 – 50%, they obtained an r^2 of 0.75 compared to a ground truth based on a photographic grid method. Laamrani, Lara, Berg, Branson & Joosse (2018) tested the mobile phone application Crop Residue Estimator (FieldTraks Solution, Ottawa, CA) that used a pixel-by-pixel classification of residue versus soil based on color thresholding to estimate percent residue cover in an image. A test of the system using 54 images obtained from 18 fields had an r^2 of 0.86 when compared to a ground truth based on a photographic grid method. Riegler-Nurscher, J. Prankl, Baeur, Strauss & H. Prankl (2018) used 200 training images to develop a pixel-by-pixel classifier using entangled random forest, a modification of random forest. The resulting model had an r^2 of 0.84 for a 99-image testing data set. The ground truth for both training and testing was obtained using a photographic grid method. Kavoosi et al. (2020) used vegetation indices derived from RGB images to estimate residue and obtained an $r^2 = 0.84$ on images obtained 5- to 10-m above the ground compared with ground truth generated from the line transect method. In contrast to others working with RGB imagery, Upadhyay et al. (2022) divided the data into three classes and used support vector machine (SVM) to classify images based on the selected features. The model was trained on 3,000 region of interest (ROI) images and tested on 1,400 ROI images. Classification accuracy for testing ROI

images was 0.81. Upadhyay et al. (2022) also performed location-wise estimation by combining 50 images from the same location using a Bayesian model and obtained an r^2 = 0.90 for 28 ((1,400 ROI/50=28) testing locations.

Early research using low resolution (80-m ground sampling distance (GSD)) imagery from the Landstat-1 multi spectral scanner was unable to estimate residue cover on agricultural soils (Gausman et al., 1975). Daughtry et al. (2006) used a Cellulose Absorption Index (CAI) applied to imagery from the EO-1 Hyperion satellite (GSD=30 m) and obtained residue prediction of r^2 = 0.85 for May data and of r^2 = 0.77 for June data compared with ground truth acquired from line transect method. They also converted the data into three classes (residue cover < 15%, 15% < residue cover < 30%, and residue cover > 30%) and obtained an accuracy score of 66% - 68%; accuracy increased to 80% - 82% when class 1 and 2 were combined. Zheng, Campbell, & de Beurs (2012) used a minimum Normalized Difference Tillage Index (minNDTI) and acquired r^2 = 0.89 when compared to ground truth from the line transect method. Kavoosi et al. (2020) used multi-spectral imagery from Landsat-8 (30-m GSD) and concluded a Dead Fuel Index (DFI) had the highest performance score (r^2 = 0.96) compared with ground truth generated with the line transect method.

In summary, RGB strategies tested to date, typically relied on limited data sets with sub-meter resolution and reported r^2 values between 0.75 - 0.90. Hyperspectral and multispectral methods have used much lower resolution imagery ($\geq$ 30 m) strategies and have typically been tested in both lab and field conditions with more extensive datasets and achieved r^2 between 0.77 - 0.96. While both approaches have shown promise, an RGB-based system has the benefit that practitioners can obtain relatively low-cost high

resolution images during field visits using a hand-held camera, smart phone, or UAV (Laamrani et al., 2018; Upadhyay et al., 2022). Research to date has documented the potential for high-resolution RGB images to provide accurate estimates of residue cover.

Machine learning is a subfield of artificial intelligence focused on building algorithms that allow the computer to learn (Segaran, 2007). Machine learning has typically focused on solving problems through the collection of data (ground truth) and then deriving a statistical algorithm based on the data that solves the problem. (Burkov, 2019). Traditionally machine learning, especially supervised machine learning, requires labeling the ground truth and the construction of features or patterns that are extracted from the data to support building the model. These features are then filtered and selected using feature selection methods and the optimum model is built using the selected features with the machine learning method. An example of a common machine learning strategy using regression is linear regression; and some common methods used for classification include SVM, multi-layer perceptron, and random forest (Kotsiantis, Zaharakis, & Pintelas, 2006). There is a fundamental difference between machine learning using regression versus classification. Regression requires the ground truth and the prediction value to be a continuous variable while classification requires the ground truth and the prediction value to be categorical variable.

The benefits of traditional machine learning approaches are that the resulting automated system can reduce human bias and labor. However, challenges of traditional machine learning strategies include the need for a large amount of data, the need to identify and develop manual features for extraction, and labelling the ground truth, especially for supervised machine learning algorithms.

However, it is also possible to use machine learning methods without manual feature extraction. An example of a model that has that capability is a Neural Network or Deep Learning (Shaheen, Verma, & Asafuddoula, 2016). In simple terms, a neural network is a model that mimics the human brain. A desirable feature of neural networks is that they can self-construct their own features (Pouyanfar et al., 2019). There are different neural network architectures available including Convolutional Neural Network (CNN), Generative Adversarial Network, and Recurrent Neural Network (Pouyanfar et al., 2019; Simonyan & Zisserman, 2015). The selection of a particular architecture is usually based on the type of data being processed and the desired output. For image prediction, CNN is the most common among architecture. A Convolutional Neural Network is a network that does parameter sharing using convolution as the operator. Parameter sharing results in a lower number of parameters than a fully connected network. Some examples of CNNs are VGGNet, AlexNet and ResNet (Alzubaidi et al., 2021).

It is common in the field of neural networks to use a method called transfer learning. Transfer learning uses knowledge from a previously trained model to solve a new problem faster or better (Kaya et al., 2019; Weiss, Khoshgoftaar, & Wang, 2016; Zhuang et al., 2021). One common way to do transfer learning with neural network is to use a pre-trained model as feature extractor (Kaya et al., 2019; Zhuang et al., 2021). A pre-trained model has been previously trained on a different dataset. With transfer learning, the information that is acquired by training the initial model is then leveraged to solve another problem. This is why the method is called transfer learning. Benefits of transfer learning include reducing the amount of data needed to train the new model and eliminating the need for manual feature extraction. Concerns with transfer learning

include that features tend to be hard to be interpret, especially when neural networks (usually considered as black box) are used for transfer learning (Q. shi Zhang & Zhu, 2018; Y. Zhang, Tino, Leonardis, & Tang, 2021).

In this project we used features derived from the pre-trained VGGNet16 model (Simonyan & Zisserman, 2015) and then evaluated models derived from those features using two classification strategies and one regression strategy. The objective was to evaluate the capacity of a transfer learning strategy to improve estimates of residue cover derived from high-resolution RGB images. Additionally, we hypothesized that regression methods may be better adapted to capturing the continuous nature of residue cover compared to a standard classification method used in a previous project. Results were also compared to previous reported results using a classification strategy to derive a model from 70 known features using the same image data set (Upadhyay et al., 2022). We also tested the uniqueness and superiority of the initial features selected from the CNN by comparing results derived from a sequential iteration strategy.

Materials and Methods

DATA ACQUISITION

From late April through early July in 2018 and 2019 the project team obtained residue imagery of corn (*Zea mays* L.) and soybean (*Glycine max* [L.] Merr.) from fields in five central Missouri counties (Table 1). Before we obtained data, NRCS personnel identified and obtained permission for the project team to access the farmer fields. Our objective was to visit fields after planting but prior to corn or soybean reaching growth stage V3. Additonal location details are reported in Table 1.

In each field, the team placed up to five 15.2-m tapes at 45 degrees to the crop planted row direction with the goal of capturing the range of variation in residue cover within a field. For the purposes of this research, each tape was considered a location. Imagery was obtained at an elevation of one meter above the surface of the soil using a tripod-mounted Canon EOS Rebel T6i Digital SLR camera (Canon USA, Melville, NY) with a 24 mm lens and 24.2 MP resolution, producing a 6,000- x 4,000-pixel image. The images had a calculated ground sampling distance (GSD) of 0.014 cm pixel^{-1}. We obtained 50 or 51 images from each tape by centering the camera over the 0 point on the tape horizontal to the soil surface, taking an image, and then repeating the process every 30 cm along the length of the tape. We acquired imagery from 60 field locations in 2018 and 28 locations in 2019. Dominant residue (Table 1; Fig. 1) in 2018 was corn (28 locations), soybean (23 locations), or winter small grains (9 locations); and in 2019 was corn (4 locations), soybean (19 locations), or winter small grains (5 locations).

Using Adobe Photoshop (Adobe inc., San Jose, CA) 50 sequential 0.305 m X 0.20 m images (approximately 2,400 x 1,600 pixels) were cropped from the part of the image

Table 1. Approximate location, sampling date, crop, and residue details for 2018 and 2019 study sites. Exact locations were not provided to protect farmer privacy.

ID	County	Approximate Site Location	Sampling Date	Locations at Site	Location Number	Crop (Stage)[1]	Dominant Residue[1]	Other Issues	Location Residue Ground Truth (%)
2018-01	Audrain	39.260, -92.272	8-May-18	3	39, 40, 41	C (V2)	SB	-	14, 9, 5
2018-02	Audrain	39.257, -92.267	20-May-18	3	42, 43, 44	NE	C, TL	-	48, 57, 54
2018-03	Audrain	39.264, -92.270	8-May-18	4	45, 46, 47, 48	C (V2)	SB	-	10, 9, 21, 1
2018-04	Audrain	39.269, -92.265	8-May-18	3	49, 50, 51	NE	C, SB	-	41, 44, 49
2018-06	Audrain	39.267, -92.266	10-May-18	4	52, 53, 54, 55	NE	C	OE (4)	57, 58, 54, 27
2018-07	Audrain	39.270, -92.212	14-May-18	3	56, 57, 58	C (V2)	SB	S (1,2), UE (3)	13, 13, 5
2018-09	Audrain	39.257, -92.150	29-May-18	3	59, 60, 61	SB (V2.5)	C	S (1,2,3)	23, 23, 23
2018-11	Callaway	39.029, -92.078	30-May-18	3	62, 63, 64	SB (V1)	WSG, C	WL (1,2)	97, 91, 89
2018-15	Boone	39.215, -92.204	23-May-18	3	65, 66, 67	SB (V1)	C	S (1)	57, 66, 51
2018-16	Boone	39.212, -92.203	30-May-18	3	68, 69, 70	SB (V1)	C	S (1,2,3)	57, 60, 57
2018-20	Cooper	38.778, -92.657	17-May-18	3	71, 72, 73	C (V3)	SB	-	21, 27, 26
2018-21	Cooper	38.782, -92.654	17-May-18	3	74, 75, 76	C (V3)	WSG	-	25, 27, 24
2018-22	Cooper	38.817, -92.623	18-May-18	3	77, 78, 79	SB (VE)	C, W, SB	WL (1)	85, 88, 94
2018-25	Cooper	38.815, -92.623	18-May-18	3	80, 81, 82	NE	C, W	S (1,2), WL (1)	92, 85, 89
2018-26	Callaway	39.032, -92.079	4-Jun-18	3	83, 84, 85	NE	WSG	-	94, 91, 96
2018-27	Cooper	38.822, -92.620	18-May-18	3	86, 87, 88	C (V3.5)	SB, W	-	18, 20, 15
2018-40	Boone	38.901, -92.214	14-Jun-18	3	29, 31, 32	NE	SB, W	S (2,3)	3, 4, 2
2018-41	Boone	38.901, -92.210	14-Jun-18	2	34, 35	NE	C, W	S (2)	44, 41
2018-42	Boone	38.900, -92.206	14-Jun-18	4	33, 36, 37, 38	NE	SB, W	S (4), WL (3,4)	19, 12, 7, 15
2018-43	Boone	38.901, -92.210	14-Jun-18	1	30	SB (VE)	C, W	-	25
2019-01	Howard	39.037, -92.694	14-May-19	2	1, 2	C (V1.5)	SB, C	-	9, 24
2019-02	Howard	39.039, -92.694	14-May-19	1	3	C (V1.5)	SB, C	S	15
2019-03	Cooper	38.914, -92.736	15-May-19	1	4	C (V3)	SB	-	7
2019-05	Howard	38.864, -92.843	17-May-19	1	5	C (V2)	SB	S	9
2019-06	Cooper	38.864, -92.843	20-May-19	1	6	C (V2)	SB	S, WL	19
2019-07	Cooper	38.864, -92.843	20-May-19	1	7	C (V2)	SB, C	-	11
2019-08	Howard	39.020, -92.573	29-May-19	2	8, 9	SB (VC)	SB	S (1), WL (1,2)	68, 60
2019-09	Cooper	38.774, -92.940	30-May-19	1	10	SB (VE)	C, W	S, WL	7
2019-10	Cooper	38.791, -92.931	30-May-19	1	11	SB (VE)	C, W	S	6
2019-11	Audrain	39.283, -91.880	31-May-19	1	12	C (V2)	SB	WL	6
2019-12	Callaway	38.969, -92.103	3-Jun-19	2	13, 14	C (V2)	SB, W, C	S (1), WL (1)	73, 71
2019-13	Callaway	39.040, -91.838	3-Jun-19	1	15	NE	SB	-	3
2019-14	Audrain	39.299, -91.970	4-Jun-19	2	16, 17	NE	SB, C	S (1,2)	13, 42
2019-15	Audrain	39.241, -91.752	5-Jun-19	1	18	NE	SB, W	S	4
2019-16	Audrain	39.223, -91.830	5-Jun-19	2	19, 20	NE	WSG	S (1,2)	27, 40
2019-17	Audrain	39.334, -91.740	11-Jun-19	1	21	SB (VE)	SB, W	S	41
2019-18	Audrain	39.332, -91.756	11-Jun-19	2	22, 23	SB (VE)	SB, W	S (1,2)	51, 77
2019-19	Cooper	38.915, -92.739	18-Jun-19	3	24, 25, 26	SB (V1)	WSG, C	WL (2)	66, 80, 82
2019-20	Audrain	39.178, -91.742	18-Jun-19	1	27	NE	C, W	-	56
2019-21	Audrain	32.299, -91.972	28-Jun-19	1	28	SB (V1)	C, W	S	51

[1] Crops and residue types: C=corn; SB=soybean; TL=tree leaves; WSG=winter small grain; W=weeds. Note for residue, the first entry is residue from the previous year's crop. Growth Stages: NE=not emerged; VE=vegetative stage emerged; V#= vegetative stage. Other issues: WL=weeds live; S=significant shadow; OE/UE=over/under exposure in 0.014 GSD images; (numbers in parenthesis are location numbers affected).

contiguous to, but not including the tape, starting at the 0 point on the tape. The cropped image was obtained from the side of the tape with an oblique angle to the sun to eliminate

tape shadow and the cropped portion was from the image where the target area was most central to the image to minimize paralax effect. The resulting data set had 4,400 images (88 location x 50 images).

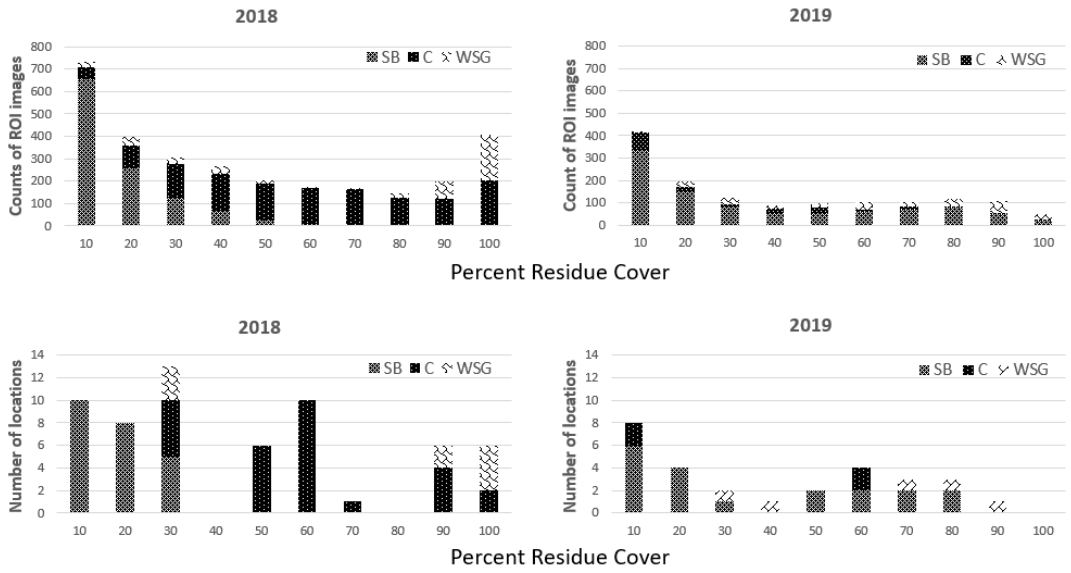

Figure 1. Distribution of dominant residue type based on three categories: SB=Soybean, C=Corn, WSG=Winter Small Grain. Data collected in 2018 (left) and 2019 (right). Report number of region-of-interest (ROI) images (top) and locations (bottom).

A bullseye grid method with n = 100 grid points (Lory et al., 2021) was used to obtain image-wise estimates of residue cover used for ground truth for the 4,400 cropped images. Residue estimates were based on two readers assessing unique 50-point grids and averaging their results. In this project, we estimated residue cover only, ignoring live plant contributions (either grain crop or weeds) to ground cover. The field-location residue estimate was then calculated as the mean of the 50 cropped images along the tape. Fig. 1 summarizes the distribution of ROI images and location images for the 2018 and 2019 data sets. More details on location residue cover and the accuracy and precision of image-wise residue estimates are available in Lory et al. (2021).

POST DATA COLLECTION

Post data collection project workflow is summarized in Figure 2.

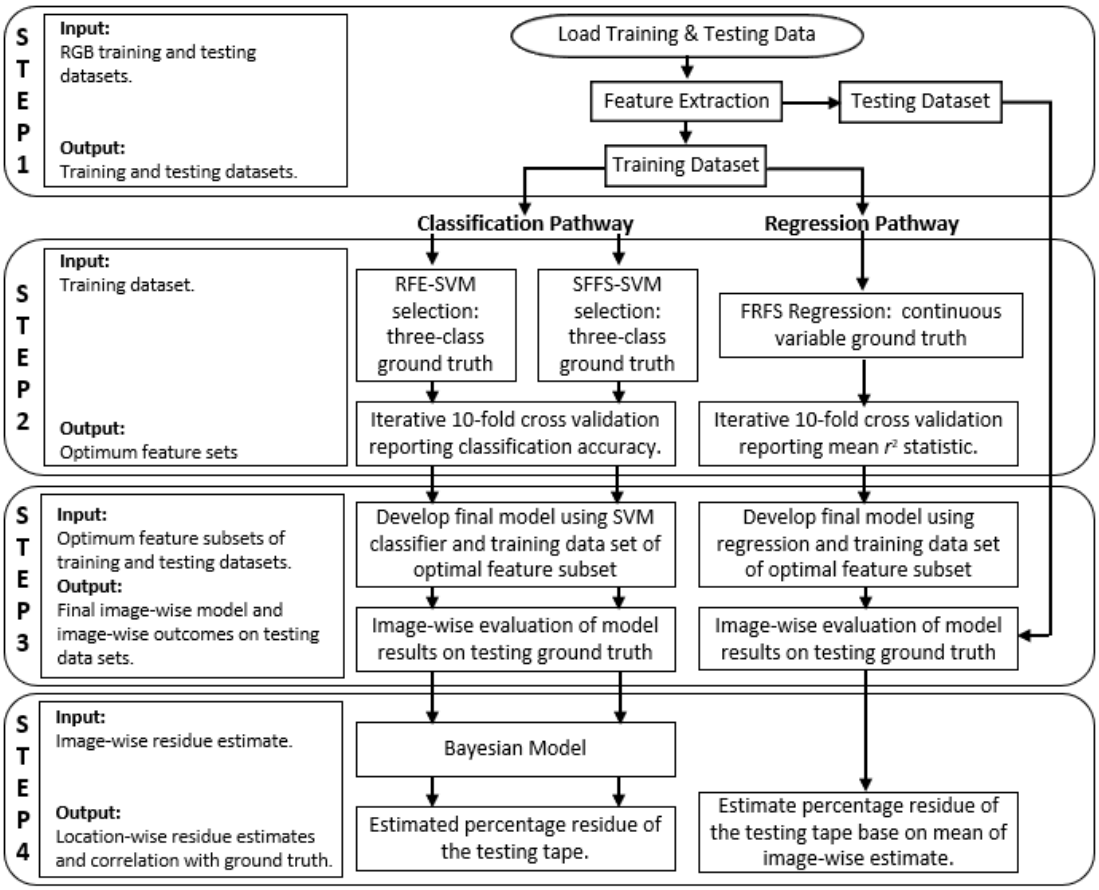

Figure 2. Post data collection project flowchart. Result was three strategies for model building and testing, two based on support vector machine classification (Recursive Feature Elimination support vector machine (RFE-SVM) and Sequential Forward Feature Selection support vector machine (SFFS-SVM)) and one based on forward regression feature selection (FRFS).

Feature Extraction (Step 1)

In step 1, features were extracted from images using a transfer learning methodology. Features were extracted from the 4,400 images using pre-trained CNN, VGGNet16 (Zimonyan and Zisserman, 2015). The model required input images be sized to 224 x 224 pixels. We resized the original images (~2400 x ~1600 pixels) to 2,240 x 1,568 pixels to facilitate analysis of the image as 70 (7 x 10) 224- x 224-pixel sub images. To resize the images, a nearest neighbor interpolation method (Keras ver 2.3.1; Chollet, 2015)was used in which pixel points of the new grid size are based on the value of their nearest neighbor. The resulting image was divided into 70 components of 224 x 224 sub-images in preparation for processing with VGGNet16 (Fig. 3).

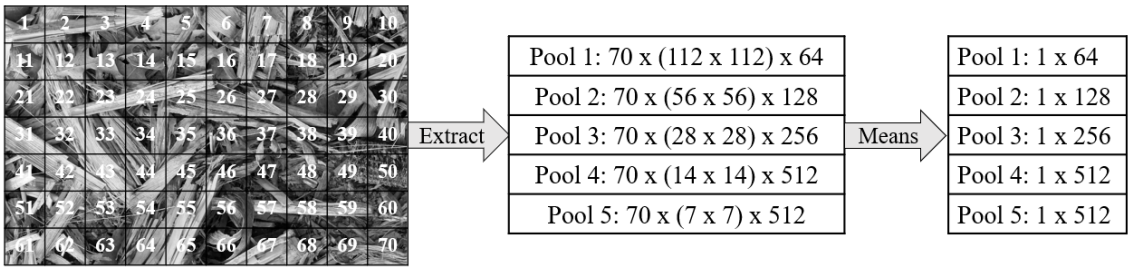

Figure 3. Flow chart showing how the five max pooling layers were extracted into 1,472 features. Features were initially extracted from 70 sub-images using the five max pooling layers of VGGNet16 which were then averaged across the 70 sub-images resulting in 1,472 features (64 + 128 + 256 + 512+ 512 = 1,472).

The VGGNet16 contains 13 convolution layers, five max pooling layers and three fully connected layers (Simonyan & Zisserman, 2015).We used the five max-pooling layers for this project because max pooling is spatially invariant (Nagi et al., 2011). In VGGNet16, the first pooling layer has a dimension of 112 x 112 cells with 64 kernels. After extraction, the first pooling layer of the 70 sub-images was recombined producing a matrix with a dimension of 1,120 x 784 with 64 kernels (Fig. 3). For each kernel, we calculated the average value of the 878,080 cells in the 1,120 x 784 matrix.

Consequently, the first max pooling layer of an image was represented by 64 values, where each value was the average of the 878,080 cells derived from the 70 sub-images. This was repeated for the other four max pooling layers with kernel numbers of 128, 256, 512, and 512, respectively. Each of the 4,440 images was then represented by a 1 x 1,472 feature matrix (64 + 128 + 256 + 512 + 512=1,472; see Fig. 3).

The data set was then divided into the training data set (2018 data of 3,000 images) and the testing data set (2019 data, 1,400 images). In preparation for feature selection analysis, using the 2018 training data, the minimum and maximum value were determined for each of the 1,472 features and all feature values were scaled to a range of 0 to 1 based on the min-max range. The same min-max range was then used to scale the 2019 data set. All subsequent analysis used scaled data. Additionally, for the classification pathway, image-wise ground truth data was divided into three classes: images with residue < 33.3% defined as class 1; images with residue between 33.3% and 66.6%, inclusive, defined as class 2; and images with residue > 66.6% defined as class 3. The number of training images was 1,549, 626, and 825 and the number of testing images was 772, 302, and 362 for classes 1, 2 and 3, respectively. All processes were implemented using Jupyter Notebooks (ver. 6.2.0; Kluyver et al., 2016).

Feature Selection (Step 2)

There were two pathways in the workflow from this step: the classification-based workflow and the regression-based workflow.

Feature selection for the classification pathway was done two ways. In the first classification approach, recursive feature elimination and the SVM classifier (linear

kernel) was used to sequentially order from 1 to 1472 (RFE-SVM; Isabelle, Jason, Stephen, & Vladimir, 2002) . In the second classification approach, sequential forward feature selection and the SVM classifier was used to put the first 40 selected features in rank order from 1 to 40 (SFFS-SVM; Ferri, Pudil, Hatef, & Kittler, 1994). In both classification pathways, ten-fold cross validation was then used to determine the cross-validation score for the addition of each ranked feature, starting from the first ranked feature and evaluating up to the 40^{th} ranked feature. In this step the SVM classifier was used with the radial basis function kernel, the penalty parameter C selected as '1'and the kernel parameter γ selected as 'auto'. The optimum feature set was determined from the 10-fold cross-validation scores using an accuracy threshold of 0.005; when the increase in the accuracy score dropped to less than 0.005 for an added feature, no additional features were added to the optimum features set. Classification operations using the SVM model used the 'SVC' package in the "sklearn" library (Pedregosa et al., 2011) and were implemented using the Jupyter Notebook (ver. 6.2.0, (Kluyver et al., 2016).

In the regression pathway, a natural log transformation was applied to the ground truth data;

$$\text{LogGT} = \ln\left(\frac{(GT+0.5)}{(100.5-GT)}\right) \qquad \textit{Eq. 1}$$

where GT is the ground truth value. Forward regression feature selection (FRFS) was used on the training data using the LogGT value as the ground truth and the 1,472 scaled features as the potential explanatory variables. Similar to the classification pathway, ten-fold cross validation was used and an accuracy score was calculated based on the average of the ten r^2 values generated in the 10-fold cross validation; the r^2 value was based on the difference between the predicted value and LogGT. We evaluated the first 40

selected features. Similar to the classification method, when the increase in the accuracy score for adding a feature dropped below 0.005, no additional features were added to the optimum feature set. All regression analysis was programmed using python in Jupyter Notebooks (ver. 6.2.0; Kluyver et al., 2016).

This process resulted in three optimum feature sets, two from the classification pathway (RFE-SVM and SFFS-SVM) and one from the regression pathway (FRFS).

Final model and image-wise testing (Step 3)

The optimum feature sets for the two classification approaches and the regression approach from Step 2 were then used to develop the final model. To train the model, the 3,000 training images were used with respective model building strategies. The model was then tested on the 1,400 testing images (2019 dataset) to obtain the image-wise classification score (RFE-SVM and SFFS-SVM models) or the image-wise estimate of residue (FRFS model). Imagewise classification accuracy (classification methods) or linear regression fit statistics to the ground truth data (regression method) were reported.

Location-wise assessment (Step 4)

For the classification approach, a Bayesian multinomial Gaussian model, based on the Bummer model, was developed to estimate location residue (Upadhyay et al., 2022; Vasko, Toivonen, & Korhola, 2000). The model was developed using the 2018 training data set and estimated location-wise residue values based on probability distributions image-wise classification scores, as determined by the final classification models developed in step 3. The scaling factor (alpha) was set to one, the optimum factors (betas) were assumed to be normally distributed, the tolerance factors (gammas) were assumed

to have a gamma distribution, and non-informative priors were used, consistent with Vasko et al. (2000). This Bayesian model was then used to obtain location-wise residue estimates for the 2019 testing dataset with the image-wise classification estimates from step 3 as the input. The model was developed using proc MCMC (SAS ver. 9.4, SAS Institute Inc., Cary NC). For the regression approach, location-wise residue estimates were the mean of the 50 image-wise estimates of each tape predicted from the final regression model. Location-wise model performance for all three methods were then evaluated using regression fit statistics to the ground truth data.

Model Iteration

For all three methods, we repeated the analysis (steps 1 to 4) three times to analyze the uniqueness and the superiority of the selected feature sets. In the first iteration, all 1,472 features were used. In the second iteration, the optimum selected features from the first iteration were removed from the data set and the process was repeated on the remaining features. In the third iteration, the optimum features selected in the second iteration were also removed from the data set and the process was repeated.

Result and Discussion

Feature Selection and ROI Residue Estimates

The first objective was to compare the models derived from three feature selection strategies selecting from the 1,472 features generated using the pre-trained VGGNet16 CNN. Fig 4 summarizes mean accuracy scores from the two classification strategies (RFE-SVM and SFFS-SVM), and the mean r^2 for the regression strategy (FRFS), all

based on 10-fold cross validation. Seven features were selected using RFE-SVM, 10 using SFFS-SVM, and seven using FRFS (Fig. 4; Table 2) based on the selection criterion (>0.005 unit improvement in accuracy/model fit). Features selected by classification did not overlap with features selected by regression. There were two common features selected using RFE-SVM compared to SFFS-SVM (Table 2).

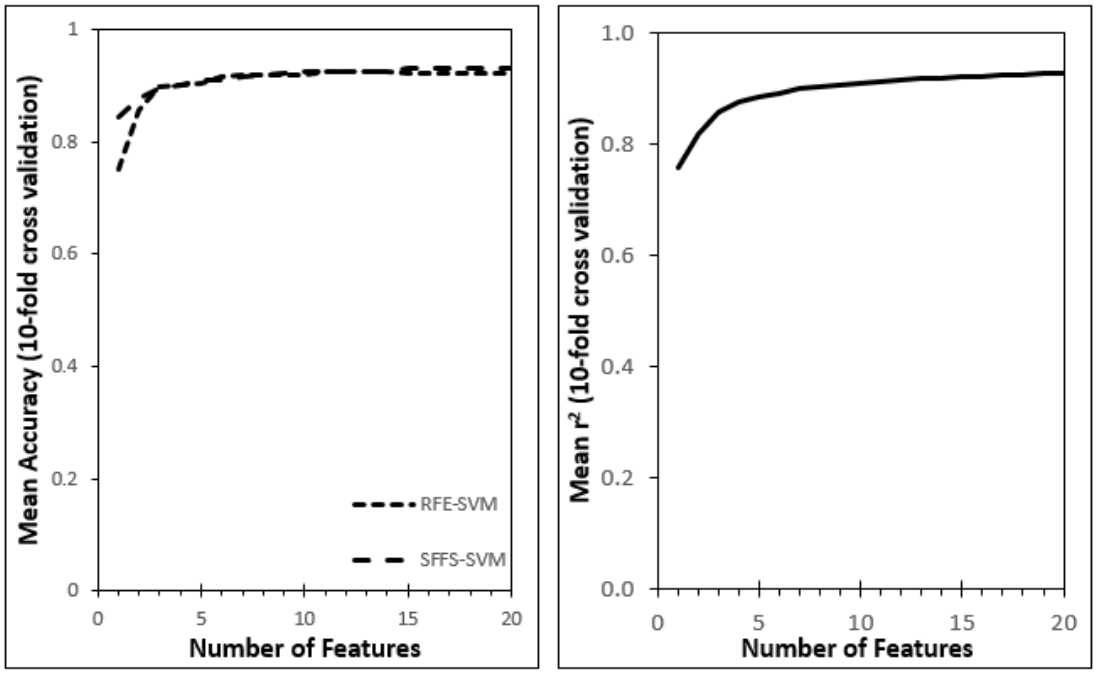

Figure 4. Ten-fold cross validation scores for classification pathways (RFE-SVM and SFFS-SVM) and regression pathway (FRFS) calculated for first 20 selected features. The reported metric is the mean Accuracy (classification) and mean r^2 (regression) of the ten folds of the cross-validation.

To direcly compare the performance of the three models, we determined the classification accuracy of the regression model based on the three-classes used for classification strategies (Table 2). This required converting the regression results into the three-classes to calculate the three-class accuracy score. Although validation and classification scores were higher for the two classification models, the regression model had the highest testing score and the least change between training and testing score. The higher testing score suggested that the regression model was the superior model for the

ROI images. The significant drop in testing score from the training score in the classification methods suggests some over-fitting of those models. The RFE-SVM model had a higher testing score than the SFFS-SVM model suggesting it was the superior classification strategy (Table 2).

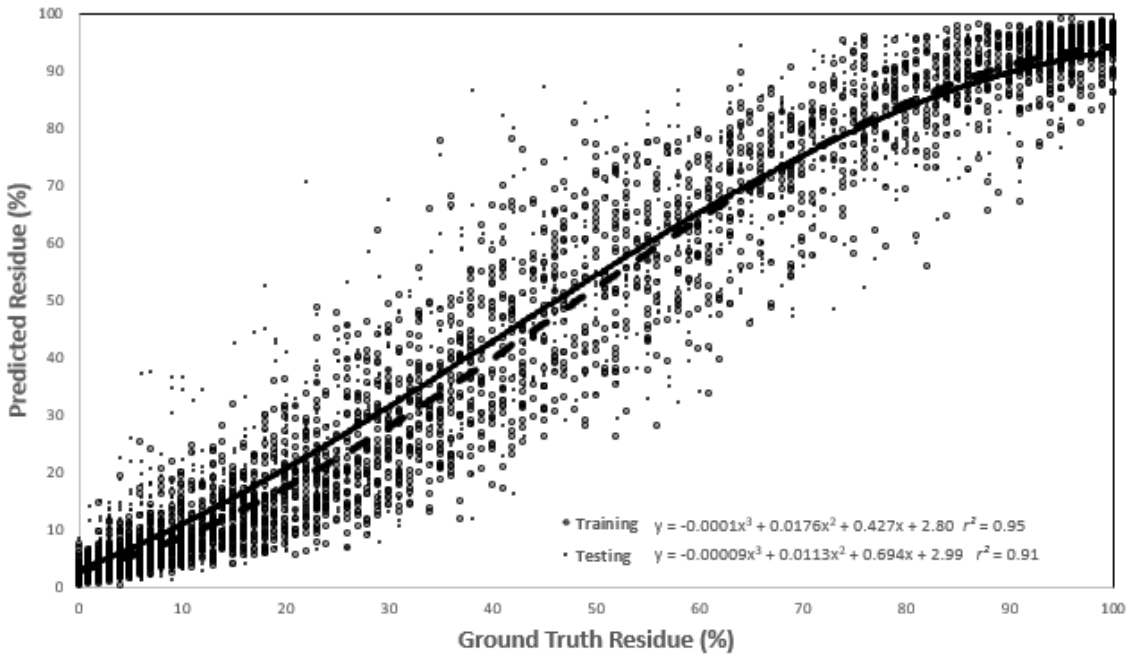

Figure 5. Ground truth percent residue versus predicted percent residue of region of interest images based on forward regression feature selection for training and testing datasets. Data has been converted back to 0-100% basis.

To better understand the performance of regression analysis, we then compared both the training and the testing ROI datasets with the ground truth (Fig. 5). The resulting model fit documented a lack of sensitivity to differences in residue when residue was < 15% or > 85%. When comparing ROI model fit between the 2018 training data and 2019 testing data, there was more variability and some evidence that the model overestimated residue when residue was between 15 to 45 % (Fig.5).

Table 2. Selected information on model results and performance for two classification-based model selection methods and one regression-based model selection method using the region of interest (ROI) image dataset.

Feature Selection				Classification Scores		
Method[1]	**Strategy**	**Number**	**Selected Feature ID numbers[2]**	**Validation**	**Training**	**Testing**
RFE-SVM	Classification	7	785(4), 457(4), 764(4), 713(4), 705(4), 868(4), 198(3)	0.92	0.92	0.85
SFFS-SVM	Classification	10	646(4), 764(4), 482(4), 814(4), 735(4), 386(3), 1357(5), 636(4), 1060(5), 868(4)	0.92	0.92	0.81
FRFS	Regression	7	481(4), 462(4), 263(3), 784(4), 18(1), 1254(5), 669(4)	0.88[3]	0.88[3]	0.87[3]

[1] SFFS-SVM=sequential forward feature selection support vector machine; RFE-SVM=recursive feature elimination support vector machine; FRFS=forward regression feature selection. [2] Number in parentheses is the max pooling layer number. [3]Model performance scores for FRFS calculated by converting region of interest residue estimates into the appropriate three-class level used for the classification methods and comparing with the ground truth class.

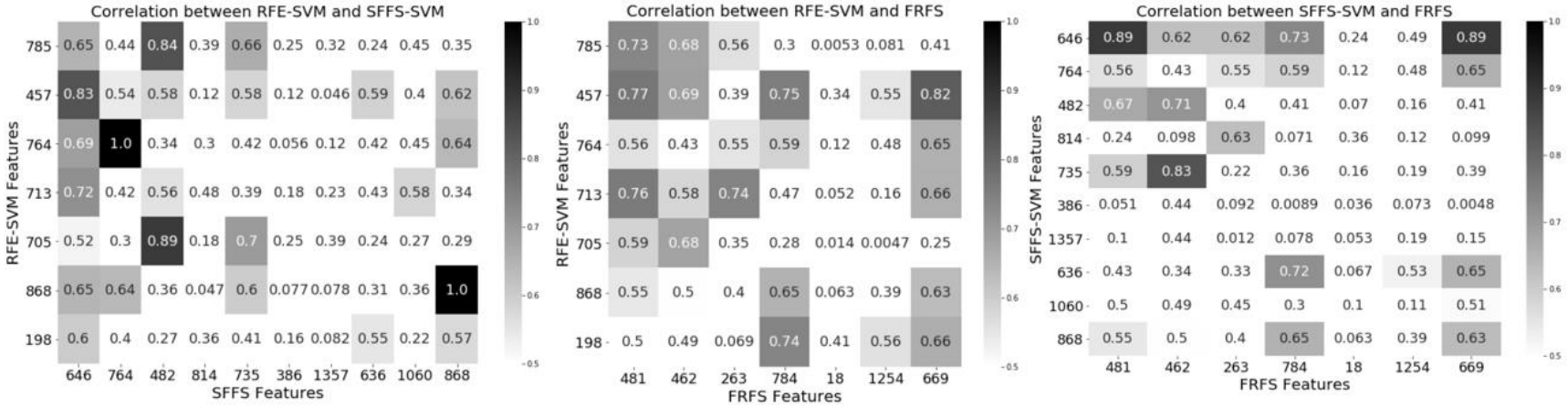

Figure 6. Correlation among features selected by three feature selection methods: recursive feature elimination using support vector machine (RFE-SVM), sequential forward feature selection using SVM (SFFS-SVM) and forward regression feature selection (FRFS). The first selected feature is at the left of the *x*-axis and top of the *y*-axis.

We then investigated correlations among the three feature sets (Fig. 6). High correlation among all features selected by the three models would imply that the models used similar strategies to quantify residue. Instead, we observed few clear patterns in correlation among the three features sets (Fig. 6). The most apparent similarity was that both forward selection methods selected highly correlated first features (r=0.89). In forward selection, the first selected feature will always be highly correlated with ground truth. Beyond that, there was no apparent correlation pattern to any of the other selected features. We concluded that there is little commonality among the features selected by the three models.

Almost all the features selected by classification methods were from the max-pooling layer four (Table 2). The four not from max-pooling layer four were from max pooling layer three (two features) or five (two features). For the regression approach, four of the seven features were from max-pooling layer four with the other three features from layers one, three and five. Generally, features from higher numbered layers are more complex (Zeiler & Fergus, 2014).

In summary, we concluded that the feature selection strategies selected features that were not highly correlated among the models. Most of the selected features had similar complexity coming from maxpooling layer 4. Based on the ROI image-wise analysis, the regression result was best followed by RFE-SVM.

LOCATION-WISE RESIDUE ESTIMATES

Location-wise results are reported in Fig. 7 and Table 3. Like the ROI results, the SFFS-SVM model performed the worst of the three methods at the location level, with

Table 3. Selected information on location-wise estimates of residue using two classification methods and one regression method. Location-wise estimates were based on 50 region of interest images.

Feature Selection		Outliers (>10% absolute error)		Training Location Model Fit		Testing Location Model Fit	
Method[1]	**Strategy**	**Training Locations**	**Testing Locations**	$r^{2\,(2)}$	**MAE[2]**	$r^{2\,(2)}$	**MAE[2]**
RFE-SVM	Classification	86, 76	22, 5, 12	0.97	3.8	0.93	4.9
SFFS-SVM	Classification	87, 79 76, 35	22, 13, 23, 5	0.97	3.7	0.91	5.9
FRFS	Regression	51, 35	14, 6	0.98	3.1	0.94	5.2

[1] SFFS-SVM=sequential forward feature selection support vector machine; RFE-SVM=recursive feature elimination support vector machine; FRFS=forward regression feature selection. [2] r^2 and MAE (mean absolute error) calculated based on the deviation of predicted values from the 1:1 line for predicted versus ground truth.

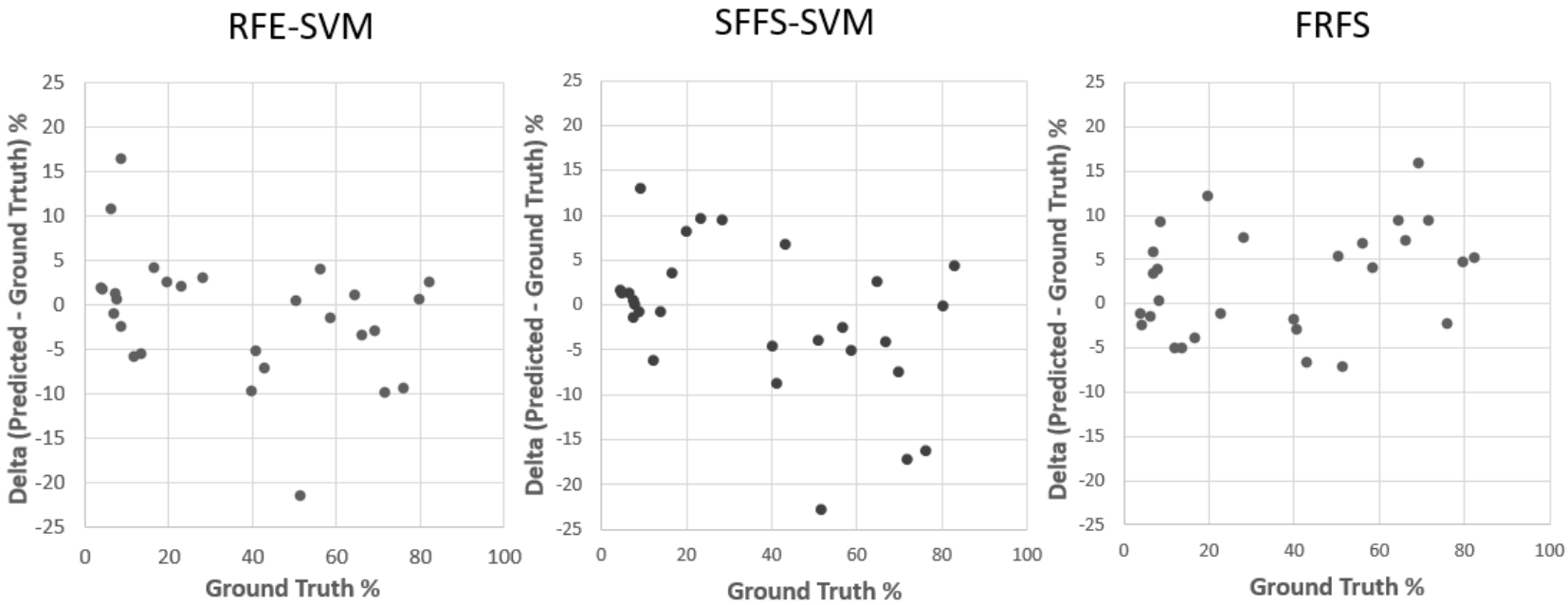

Figure 7. Delta of predicted minus ground truth versus ground truth of location-wise data for three feature selection strategy: recursive feature elimination using support vector machine (RFE-SVM), sequential forward feature selection using SVM (SFFS-SVM) and forward regression feature selection (FRFS).

the lowest r^2, highest MAE, and the most outliers in the testing dataset. It is hard to decide which is superior between the RFE-SVM and FRFS models. Based on testing r^2, FRFS was better, while based on testing MAE RFE-SVM was better (Table 3). The RFE-SVM model had three outliers, one > 20, whereas the FRFS model had two outliers, both < 20 (Fig. 7). The best-fit slope comparing predicted location values and the ground truth had a slope greater than one (slope = 1.07; slope > 1.00 $P = 0.1$) which implied the FRFS model was over-estimating residue. This also had the effect of increasing both r^2 and MAE score which were calculated based on the difference from the 1:1 line (Table 3). Without this apparent bias, the regression fit statistics would have been superior.

There was commonality of two outliers among the two classification pathways (locations 22 and 5) but there was no commonality among the classification and regression pathway outliers (Table 3). Reviewing the content of images from the outlier locations, all outliers in both classification and regression models had images that included shadow and/or weeds. However, when we compared the five locations with the highest overestimate of residue with the five locations with the lowest -under-estimation of residue, shadow appeared with similar frequency in both sets of images for both RFE-SVM and FRFS models. As an example, RFE-SVM had three out of five shadow related locations that highly overestimated the residue and five out of five shadow related locations that highly underestimated residue and three out five shadow related locations that had low-bias for residue estimation. This implies that shadow was not uniquely associated with over- or under-estimates of residue. Similar observations were seen for weeds. Hence, we were not able to attribute shadow or weeds as the cause of outliers, nor

have we identified any other predictor of outliers for both RFE-SVM and FRFS location results.

The presence of a systematic over-estimation of residue cover in the regression model suggested a failure of the regression model developed using 2018 data to accurately predict the 2019 testing data. When we compared model fit of predominant soybean residue locations in the training set (n = 23 of 60 locations) versus model fit in the testing data set (n = 19 of 28 locations) the slope of the training data was 0.97 compared to the slope of the testing data which was 1.08 (data not shown). This difference in slope was not observed with the RFE-SVM soybean data. The testing data set had a higher representation of soybean locations than the training dataset (68% vs. 38%; Fig. 1) and the range of residue cover was more extensive in the testing data set (training set 3 to 27%, testing 4 to 72%; Fig. 1). When we developed a final FRFS model using all images and the seven selected features, the best-fit slope comparing predicted location values and the ground truth improved (slope = 1.04; slope > 1.00 $P = 0.20$; MAE = 4.3) and differences between soybean training and testing fit were reduced. These results suggest a possible explanation of the over-estimation of residue in the 2019 data is that the FRFS model over predicted soybean residue cover in 2019 which was mostly addressed by adjusting model parameters using 2018 and 2019 data combined.

To summarize, both RFE-SVM and FRFS performed similarly. Shadow and weeds were associated with outlier locations, but the problem is not systematic, with classification and regression models identifying different outliers, emphasizing the differences among the models.

PERFORMANCE OF ITERATIVE FEATURE SELECTION

Removing selected features and repeating feature selection had little impact on model performance when using RFE-SVM (Tables 4 and 5). Image-wise and location-wise testing indicators suggested better performance for the models from iterations two and three with iteration three reporting the best results. The optimum number of features used, based on the 0.005 threshold, varied by iteration (Table 4). Surprisingly, the second iteration feature set performed better than the first iteration with a lesser number of features. The larger number of features used in the third iteration likely explains some of its comparative success to the previous iterations. The superior performance of iteration 2 and iteration 3 confirms that there are multiple feature sets generated from VGGNet16 features that perform similarly or superiorly using RFE-SVM.

There was no clear pattern of correlation among the features selected in the three iterations (Fig. 8). The highest correlation (0.89) was between fifth feature of first iteration and eighth feature of third iteration. Generally, the pattern of high correlation ($r > 0.8$) seemed random. This implies that the models from each iteration were not using redundant features.

In summary, there was little evidence of reduced performance with iteration and little evidence of correlation among features selected by each iteration. This suggests there were multiple sets of unique features derived from the VGGNet16 capable of estimating residue when using the RFE-SVM feature selection strategy.

Similar to RFE-SVM, SFFS-SVM testing scores for ROI image-wise results did not decrease with iteration (Table 4). With SFFS-SVM, the number of selected features decreased with each iteration. Consistent with a forward selection strategy, the first

Table 4. Selected information from three iterations of model results and performance for two classification-based model selection methods and one regression-based model selection method using the region of interest (ROI) image dataset.

					Model Performance Scores		
Method[1]	**Iteration**	**Data set[2]**	**Number selected features**	**Selected Feature ID numbers[3]**	**Validation**	**Training**	**Testing**
RFE-SVM	1	1472	7	785(4), 457(4), 764(4), 713(4), 705(4), 868(4), 198(3)	0.92	0.92	0.85
RFE-SVM	2	1465	4	320(3), 759(4), 735(4), 1036(5)	0.90	0.90	0.87
RFE-SVM	3	1461	12	1139(5), 1256(5), 545(4), 462(4), 353(3), 509(4), 739(4), 482(4), 483(4), 546(4), 933(4), 867(4)	0.92	0.92	0.87
SFFS-SVM	1	1472	10	646(4), 764(4), 482(4), 814(4), 735(4), 386(3), 1357(5), 636(4), 1060(5), 868(4)	0.92	0.92	0.81
SFFS-SVM	2	1462	5	980(5), 759(4), 793(4), 1036(5), 1206(5)	0.91	0.91	0.81
SFFS-SVM	3	1457	4	320(3), 462(4), 112(2), 492(4)	0.90	0.91	0.82
FRFS	1	1472	7	481(4), 462(4), 263(3), 784, 18(1), 1254(5), 669(4)	0.88[4]	0.88[4]	0.87[4]
FRFS	2	1465	7	980(5), 948(4), 759(4), 1088(5), 170(2), 735(4), 808(4)	0.88[4]	0.88[4]	0.84[4]
FRFS	3	1458	6	646(4), 482(4), 713(4), 683(4), 686(4), 1165(5)	0.88[4]	0.88[4]	0.83[4]

[1] SFFS-SVM=sequential forward feature selection support vector machine; RFE-SVM=recursive feature elimination support vector machine; FRFS=forward regression feature selection. [2] Data set is the number of features used for that analysis (equals the previous iteration minus features selected in the previous iteration). [3] Number in parenthesis is the max pooling layer associated with the feature. [4] Model performance scores for FRFS calculated by converting region of interest residue estimates into the appropriate three-class level and comparing with the ground truth class.

Table 5. Selected information three iterations of location-wise estimates of residue using two classification methods and one regression method. Location-wise estimates were based on 50 region of interest images.

Feature Selection		Details		Outliers (>10% absolute error)		Training Location Model Fit		Testing Location Model Fit	
		Iteration	Feature Number						
Method[1]	Strategy			Training Locations	Testing Locations	$r^{2\,(2)}$	MAE[2]	$r^{2\,(2)}$	MAE[2]
RFE-SVM	Classification	1	7	86, 76	22, 5, 12	0.97	3.8	0.93	4.9
		2	4	79, 55, 35, 87, 86, 76	19, 2, 5, 14	0.97	4.3	0.96	4.1
		3	12	76, 79, 86, 87	22, 23	0.97	3.9	0.95	4.1
SFFS-SVM	Classification	1	10	87, 79, 76, 35	22, 13, 23, 5	0.97	3.7	0.91	5.9
		2	5	79, 76, 55, 87, 86, 59	21, 20, 19, 28, 6, 5	0.96	4.2	0.90	6.4
		3	4	87, 35, 86, 59	5, 22, 21, 17, 8, 14, 3, 6	0.97	4.0	0.88	6.7
FRFS	Regression	1	7	51, 35	14, 6	0.98	3.1	0.94	5.2
		2	7	51, 59, 70, 55	8, 1, 14, 6, 19	0.97	3.6	0.93	5.8
		3	6	49, 51, 76, 72, 35	8, 2, 9, 22, 24, 17, 26, 14, 1, 25	0.97	3.6	0.87	7.6

[1] SFFS-SVM=sequential forward feature selection support vector machine; RFE-SVM=recursive feature elimination support vector machine; FRFS=forward regression feature selection. [2] r^2 and MAE (mean absolute error) calculated based on the deviation of predicted values from the 1:1 line for predicted versus ground truth

selected feature was highly correlated among the three iterations (Fig. 9). Subsequent selected features suggested a random pattern of features with typically low correlation. Similar to the results with RFE-SVM, this suggested multiple sets of unique features derived from the VGNNet16 capable of estimating residue. In contrast to RFE-SVM, location-wise testing performance decreased with each iteration (Table 5). The performance of SFFS-SVM also was worse than RFE-SVM for all iterations (Table 5).

In contrast to the two classification strategies, the testing performance of FRFS decreased with each iteration for both ROI image scores and location-wise model fit statistics (Tables 4 and 5). There was less variation in the number of selected features (seven features for iterations 1 and 2, six features for iteration 3). Consequently, the reduced performance with iteration is clearly due to lower performing features being selected in the second and third iteration. Similar to the behavior with the other forward selection method, SFFS-SVM, the first selected feature in each iteration were highly correlated (Fig. 10). And consistent with the both classification selection methods, there was little evidence of a correlation among the subsequent selected features.

There was little consistency in testing data set outliers among selection strategies and between iterations (Table 5). There was a small subset of locations that appeared as an outlier in at least one iteration of all three selection methods. This set included locations 14, 19, and 22. In general, a similar pattern emerged as was previously discussed with the first iteration comparison. Shadow and/or weed was present in locations that overestimated, underestimated, or had no bias. In the iterative models, shadow and weeds were not uniquely associated with outlier locations. The lack of consistency in outliers

among iterations also supports the conclusion that the iterative models are finding unique combinations of feature characteristics to obtain an estimate of residue.

A concern with transfer learning strategies derived from an extensive feature set such as VGGNet16 is that there are many redundant features (Mingbao Lin et al., 2021; Zuo, Chen, Shi, & Sun, 2020) making it difficult to find a superior unique feature set (Jovic, Brkic, & Bogunovic, 2015; Yu & Liu, 2004). Based on the lack of correlation among features selected in the iteration analysis, it does not appear that redundant features were a common element of the iterative models (Figs. 8, 9, and 10). Beyond the correlations among first features selected by forward selection methods (SFFS-SVM and FRFS) feature correlation seemed random. The low correlation might be a product of the averaging step in feature extraction (Fig.3).

Another criticism of transfer learning strategies derived from an extensive feature set such as VGGNet16 is that the large number of features increases the chance of deriving an over-fitted model based on random chance (Ying, 2019). We took three steps that limited this concern. The reduction of features using averaging (Fig. 3) decreased the number of features by about three orders of magnitude. We also focused on creating a large data set, larger than any other study using RGB images, to estimate residue for training. The number of ROI images used for training the model was nearly double the number of features extracted from the CNN. Finally, we used a completely independent testing data set taken from different locations and in a different year representing different environmental conditions. The decrease in fit between training and testing for the ROI images using the FRFS strategy decreased 4.2% (Fig. 5) and location testing scores were similarly close for both RFE-SVM and FRFS (Table 2). This supports

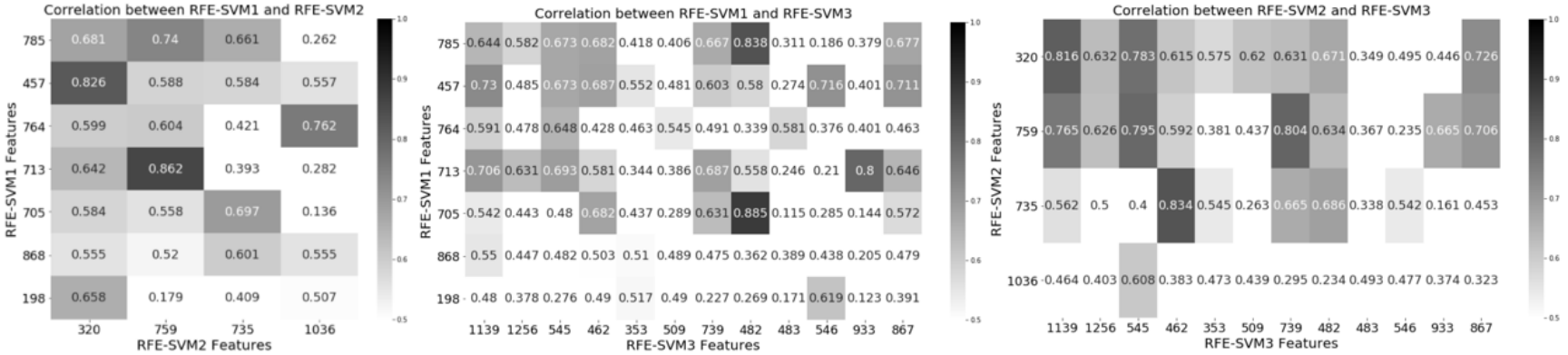

Figure 8. Correlation among model iteration selected feature used for RFE-SVM strategy. The first selected feature is at the left of the *x*-axis and top of the *y*-axis.

Figure 9. Correlation among model iteration selected feature used for SFFS-SVM strategy. The first selected feature is at the left of the *x*-axis and top of the *y*-axis.

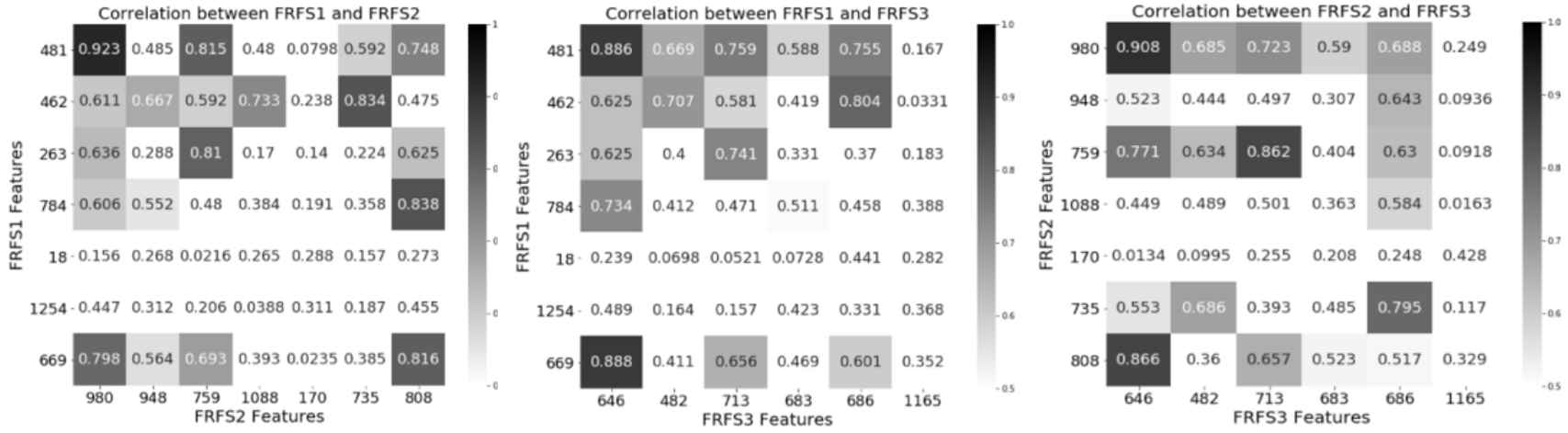

Figure 10. Correlation among model iteration selected feature used for FRFS strategy. The first selected feature is at the left of the *x*-axis and top of the *y*-axis.

that over-fitting was not a significant problem using the selection criteria we used in this project.

The averaging step in feature extraction is equivalent to global average pooling (GAP, Fig. 3) commonly used in CNN's (Li & Wang, 2020; Min Lin, Chen, & Yan, 2014). The concern with this step is it may reduce sensitivity of the resulting feature due to the averaging process. The success of the resulting models using RFE-SVM and FRFS implies that a lack of features sensitive to residue cover was not an issue. Future work could test other methods to pool data to reduce the number of features.

In summary, we concluded that the strategy employed to combine traditional machine learning with transfer learning features from VGGNet16 using a GAP strategy was successful for modeling residue cover with little evidence of overfitting.

Performance of Regression versus Classification Strategies

We hypothesized that regression may have advantages over the classification method we tested because it was better adapted to the ground truth as a continuous variable. However, looking at the performance of classification based on RFE-SVM versus the results of regression using FRFS, there was no clear evidence that one performed better than the other (Tables 4 and 5). Testing results with RFE-SVM improved with each iteration for both ROI and location-wise analyses. In the first iteration, FRFS performed better but degraded with each iteration. One reason for this may be that regression always selects the remaining feature most correlated with the ground truth for the first selected feature. This apparently was limiting opportunities to identify successful alternative feature sets in iterations two and three. In contrast, RFE-SVM, as a backwards feature

selection method, resulted in three feature sets that were not well correlated with each other (Fig. 8) but performed similarly in testing (Table 5).

An alternative explanation is that the FRFS approach was better able to define a superior model. Evidence for this was that the location-wise testing score would have been superior were it not for a bias to over-estimate soybean residue in the testing dataset. This feature set was successful enough to identify this difference in 2018 training versus 2019 testing data. Additionally, the fact that RFE-SVM selected multiple models with similar testing results could be a sign of weakness in the approach. We are concerned that the three-class strategy that was employed was not the optimum classification strategy. One possible interpretation of our results is that a better implementation of classification with RFE-SVM could improve results. Future work should evaluate if changes in the classification strategy such as increasing the number of classes and/or adjusting the classification methodology to account for the known relationships among the classes could improve the outcome with RFE-SVM.

We also should note that for our implementation of RFE-SVM, at the location scale, the Bayesian model that integrated data from 50 images also could have accounted for some inherent biases from the classification system. The regression approach did not benefit from this second level of model fitting. Finally, we note there was nothing significant that can be interpreted from the role of outliers in regression versus classification methods.

COMPARISON WITH OTHER APPROACHES

Upadhyay et al., (2022) used the same data set of images to develop, train and test a three-class model selecting from a set of 70 features that included local binary pattern features (LBP), global texture features, shape features, and color features. For the image-wise analysis, they obtained a three-class testing accuracy score of 0.81 which was less than what was obtained with the transfer learning feature sets for RFE-SVM and FRFS (Table 2). Location-wise results were also worse than all three models developed with transfer learning features (r^2 = 0.89, MAE = 6.5, with seven outliers; compare to Table 3). Finally, using the 70 features used by Upadhyay et al. (2022), we used the FRFS methodology to select features and develop a new model. The resulting three-class classification score was 0.83, which was superior to their RFE-SVM result. Location-wise model fit of the FRFS model was marginally worse (r^2 = 0.88, MAE = 5.9, with 3 outliers) primarily due to a single outlier (location 23, delta = 33). The ability of the transfer learning feature set to out-perform the manually selected feature set in all cases implied the 70 features used by Upadhyay et al. (2022) lacked needed information and there is potential to identify existing and/or develop new features that could improve residue estimation.

A neural network is considered a black box and the specific nature of the features is difficult to understand. We investigated correlations between known selected features from Upadhyay et al. (2022) and transfer learning selected features by RFE-SVM, SFFS-SVM and FRFS (Fig. 11). The first selected feature of both SFFS-SVM (481) and FRFS (646) were highly correlated (r > 0.8) with LBP texture bins 14, 19, 20 and 21 with bin 21 the highest correlation. These bins are associated with edge (14) and light corner

Table 6. Selected information on region-of-interest and location-wise estimates of residue based on Upadhyay et al. (2021) features.

Feature Selection		Classification Scores	Number of Features	Outliers (>10% absolute error)	Testing Location Model Fit	
Method[1]	**Strategy**	**Testing**		**Testing Locations**	$r^{2\,(2)}$	**MAE[2]**
RFE-SVM	Classification	0.81	13	19, 13, 5, 20, 21, 14, 8	0.89	6.5
FRFS	Regression	0.83[3]	8	23, 22, 21	0.88	5.9

[1] RFE-SVM=recursive feature elimination support vector machine results reported by Upadhyay et al. (2021) and FRFS=forward regression feature selection. [2] r^2 and MAE (mean absolute error) calculated based on the deviation of predicted values from the 1:1 line for predicted versus ground truth. [3] Model performance scores for FRFS calculated by converting region of interest residue estimates into the appropriate three-class level and comparing with the ground truth class.

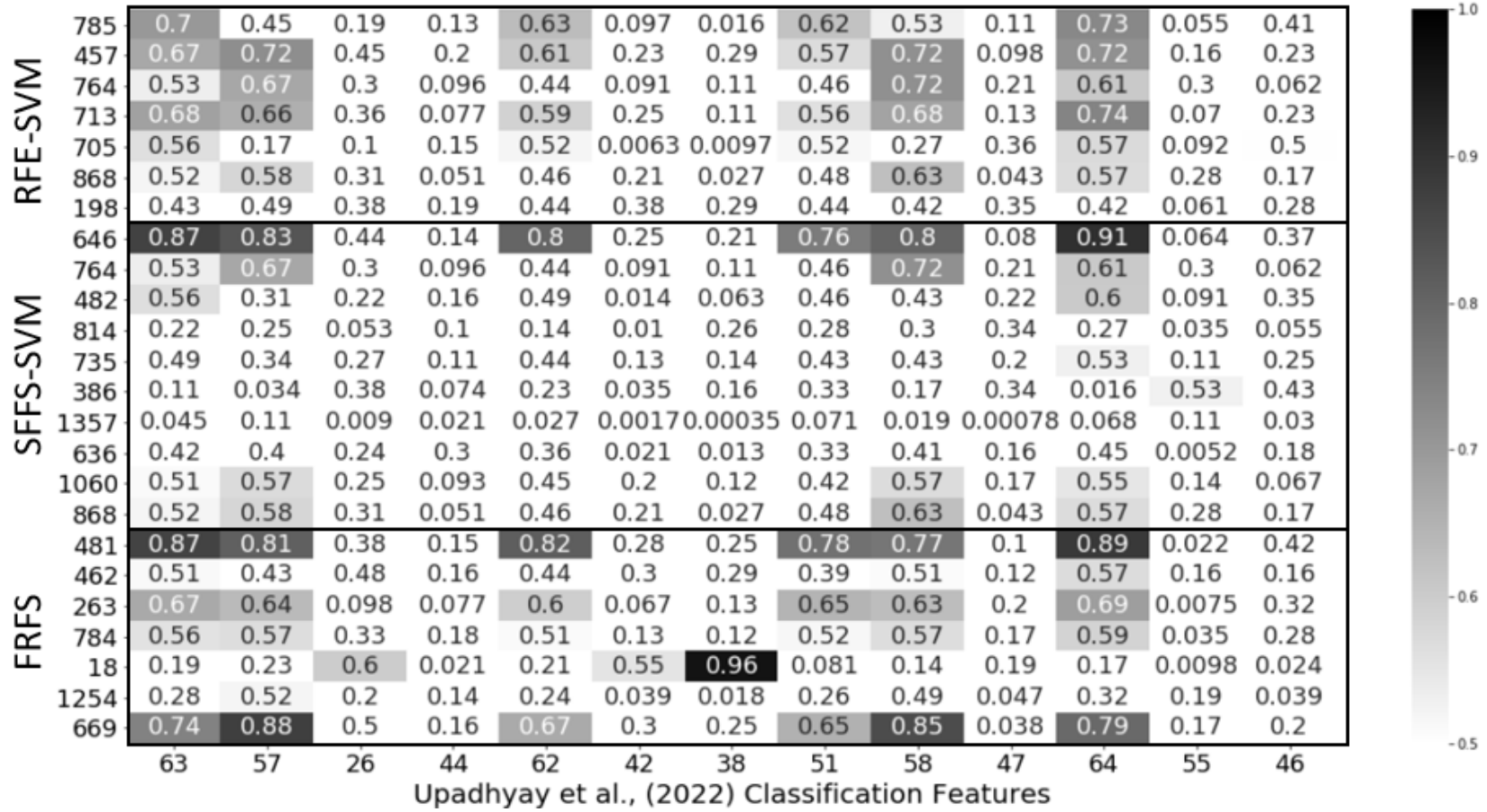

Figure. 11 Correlations between features selected by three methods (RFE-SVM, SFFS-SVM and FRFS) from a transfer-learning derived feature set (*y*-axis) versus features selected by Upadhyay et al., (2022) using RFE-SVM. The first selected feature is at the left of the *x*-axis. Black outlines separate the RFE-SVM, SFFS-SVM and FRFS from top to bottom. The first feature selected by each method is at the top of each of the three sections of the *y*-axis.

feature in a dark area (bins 19. 20, and 21). The highest observed correlation was FRFS feature 18 with the standard deviation of A, from the LAB color space (r = 0.96; Fig. 11). There were five features from Upadhyay et al. (2022) that did not correlate well with any of the selected transfer learning features. These included two features based on standard deviation of a color space feature (STD of B in LAB and STD of S in HSV) and three LBP features associated with dark or mostly dark areas. For FRFS, two selected features had low correlation with all Upadhyay et al. (2022) features. For RFE-SVM, there was no correlation > 0.75 indicating little commonality in the feature sets. These results further support the conclusion that each model had unique characteristics limiting the ability to derive insight about transfer learning features. There is some evidence that edges and light corners are a component of some transfer learning features, but not dark areas.

The results of this research compare favorably with previous published research estimating of residue cover using high-resolution RGB images which reported performance r^2 between 0.75 and 0.90 (Bauer & Strauss, 2014; Kavoosi et al., 2020; Laamrani et al., 2018; Riegler-Nurscher et al., 2018; Upadhyay et al., 2022). The data set used in this research was more extensive than most previously reported work based on 4,400 ROI images from 88 locations spanning two years.

We recommend more exploration of transfer learning strategies and transfer learning features based on the superiority of the result in this paper. The correlation of some transfer learning features with LBP texture features suggest that further exploration of LBP features may lead to improvements in residue prediction. Further analysis is also needed on the opportunity to generalize results for other GSD.

Conclusions

A transfer learning strategy using features extracted from VGGNet16 pre-trained on ImageNet provided superior results for residue cover prediction compared to previous research, including research using the same data set evaluating 70 manually extracted known features (Upadhyay et al., 2022). We obtained similar success using a regression strategy (FRFS) and a classification strategy (RFE-SVM), but a forward selection classification strategy (SFFS-SVM) was less successful. This outcome among models was true using both the transfer learning feature set and the 70-known features set from the previous research. Future work should consider improvements to the regression approach such as evaluating a backwards selection process; and improvements to the classification approach such as more classes when using RFE-SVM.

There was little commonality among models derived from each strategy. There was no consistent pattern of feature correlation among models and outliers among the models did not follow a consistent pattern. An iterative analysis of feature selection also documented that each feature set was unique among the three iterations. This suggested redundancy was not an issue with the transfer learning feature set. The success of the transfer learning method suggests that the GAP strategy to reduce the number of features extracted from the CNN was also successful at reducing redundancy associated with the CNN feature set.

Implementation of a model derived from transfer learning is more difficult than implementing a model derived from known manually extracted features. However, the success of the transfer learning model documented that there is an opportunity to improve

upon the features developed to date by more conventional methods. There is a clear opportunity to find better known features to help with a residue problem.

The success of this research suggested there are opportunities to improve outcomes from the transfer learning approach. There may be an opportunity to learn more by comparing how residue cover model perform using other CNNs with the same extraction strategies and comparing how a residue cover model performs with extraction of existing VGGNet16 model using different pre-trained dataset. Another potential opportunity is to investigate alternative ways to the GAP approach to reduce features such as global max pooling and univariate feature selection.

This research confirms the utility of high-resolution RGB imagery to quantify residue cover. Expanding access to low-cost high-resolution RGB images through technologies such as smart phones and UAVs suggests this will be an important strategy to document compliance with residue requirements in agricultural systems.

References

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., … Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8, 53. Retrieved from https://doi.org/10.1186/s40537-021-00444-8

Bauer, T., & Strauss, P. (2014). A rule-based image analysis approach for calculating residues and vegetation cover under field conditions. *Catena*, 113, 363–369. Retrieved from https://doi.org/10.1016/j.catena.2013.08.022

Bronick, C. J., & Lal, R. (2005). Soil structure and management: A review. *Geoderma*, 124(1–2), 3–22. Retrieved from https://doi.org/10.1016/j.geoderma.2004.03.005

Burkov, A. (2019). *The Hundred-Page Machine Learning Book*. Andriy Burkov.

Cherubin, M. R., Oliveira, D. M. D. S., Feigl, B. J., Pimentel, L. G., Lisboa, I. P., Gmach, M. R., … Cerri, C. C. (2018). Crop residue harvest for bioenergy production and its implications on soil functioning and plant growth: A review. *Scientia Agricola*, 75(3), 255–272. Retrieved from https://doi.org/10.1590/1678-992x-2016-0459

Chi, J., & Crawford, M. M. (2014). Spectral Unmixing-Based Crop Residue Estimation Using Hyperspectral Remote Sensing data: A Case Study at Purdue University. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6), 2531–2539. Retrieved from https://doi.org/10.1109/JSTARS.2014.2319585

Chollet, F. (2015). Keras. Github. Retrieved from https://github.com/fchollet/keras

Dabney, S. M., Yoder, D. C., Vieira, D. A. N., & Bingner, R. L. (2011). Enhancing RUSLE to include runoff-driven phenomena. *Hydrological Processes*, 25(9), 1373–1390. Retrieved from https://doi.org/10.1002/hyp.7897

Daughtry, C. S. T., Doraiswamy, P. C., Hunt, E. R., Stern, A. J., McMurtrey, J. E., & Prueger, J. H. (2006). Remote sensing of crop residue cover and soil tillage intensity. *Soil and Tillage Research*, 91(1–2), 101–108. Retrieved from

https://doi.org/10.1016/j.still.2005.11.013

Dickey, E. C., Shelton, D. P., Meyer, G. E., & Fairbanks, K. T. (1989). Determining Crop Residue Cover with Electronic Image Analysis. *Biological Systems Engineering*. Retrieved from https://doi.org/10.1139/f79-068

Ferri, F. J., Pudil, P., Hatef, M., & Kittler, J. (1994). Comparative study of techniques for large-scale feature selection. *Machine Intelligence and Pattern Recognition*, 16, 403–413. Retrieved from https://doi.org/10.1016/B978-0-444-81892-8.50040-7

Flanagan, D., Gilley, J., & Franti, T. (2007). Water Erosion Prediction Project (WEPP) Development History, Model Capabilities, and Future Enhancements. *Transactions of the ASABE*, 50(5), 1603–1612.

Gausman, H. W., Gerbermann, A. H., Wiegand, C. L., Leamer, R. W., Rodriguez, R. R., & Noriega, J. R. (1975). Reflectance Differences Between Crop Residues and Bare Soils. *Soil Science Society of America Journal*, 39(4), 752–755. Retrieved from https://doi.org/10.2136/sssaj1975.03615995003900040043x

Glaser, L. K. Provisions of the Food Security Act of 1985 (1985). Retrieved from https://www.ers.usda.gov/webdocs/publications/41995/15133_aib498_1_.pdf?v=153.7

Hively, W. D., Shermeyer, J., Lamb, B. T., Daughtry, C. T., Quemada, M., & Keppler, J. (2019). Mapping crop residue by combining landsat and worldview-3 satellite imagery. *Remote Sensing*, 11(16), 1857. Retrieved from https://doi.org/10.3390/rs11161857

Isabelle, G., Jason, W., Stephen, B., & Vladimir, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46, 389–422. Retrieved from http://link.springer.com/article/10.1023/A:1012487302797

Jovic, A., Brkic, K., & Bogunovic, N. (2015). A review of feature selection methods with applications. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1200–1205. Retrieved from https://doi.org/10.1109/MIPRO.2015.7160458

Kavoosi, Z., Raoufat, M. H., Dehghani, M., Abdolabbas, J., Kazemeini, S. A., & Nazemossadat, M. J. (2020). Feasibility of satellite and drone images for monitoring soil residue cover. *Journal of the Saudi Society of Agricultural Sciences*, 19(1), 56–64. Retrieved from https://doi.org/10.1016/j.jssas.2018.06.001

Kaya, A., Keceli, A. S., Catal, C., Yalic, H. Y., Temucin, H., & Tekinerdogan, B. (2019). Analysis of transfer learning for deep neural network based plant classification models. *Computers and Electronics in Agriculture*, 158, 20–29. Retrieved from https://doi.org/10.1016/j.compag.2019.01.041

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., … Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, 87–90. Retrieved from https://doi.org/10.3233/978-1-61499-649-1-87

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26, 159–190. Retrieved from https://doi.org/10.1007/s10462-007-9052-3

Laamrani, A., Joosse, P., & Feisthauer, N. (2017). Determining the number of measurements required to estimate crop residue cover by different methods. *Journal of Soil and Water Conservation*, 72(5), 471–479. Retrieved from https://doi.org/10.2489/jswc.72.5.471

Laamrani, Ahmed, Lara, R. P., Berg, A. A., Branson, D., & Joosse, P. (2018). Using a mobile device “app” and proximal remote sensing technologies to assess soil cover fractions on agricultural fields. *Sensors*, 18(3), 708. Retrieved from https://doi.org/10.3390/s18030708

Laflen, J. M., Amemiya, M., & Hintz, E. A. (1981). Measuring crop residue cover. *Journal of Soil and Water Conservation*, 36(6), 341–343.

Li, Y., & Wang, K. (2020). Modified convolutional neural network with global average

pooling for intelligent fault diagnosis of industrial gearbox. *Maintenance and Reliability*, 22(1), 63–72.

Lin, Min, Chen, Q., & Yan, S. (2014). Network in network, 1–10. Retrieved from https://doi.org/10.48550/arXiv.1312.4400

Lin, Mingbao, Ji, R., Zhang, Y., Zhang, B., Wu, Y., & Tian, Y. (2021). Channel pruning via automatic structure search. *IJCAI International Joint Conference on Artificial Intelligence*, 94, 673–679. Retrieved from https://doi.org/10.24963/ijcai.2020/94

Lory, J. A., Upadhyay, P., Lagaunne, T. A. P., Spinka, C., Miller, R., Davis, G., & DeSouza, G. N. (2021). Capability of high-resolution RGB imagery to accurately document residue in row-crop fields. *Journal of Soil and Water Conservation*, 76(5), 403–413. Retrieved from https://doi.org/10.2489/jswc.2021.00193

Morrison Jnr, J. E., Huang Chi Hua, Lightle, D. T., & Daughtry, C. S. T. (1993). Residue measurement techniques. *Journal of Soil and Water Conservation*, 48(6), 478–483.

Nagi, J., Ducatelle, F., Di Caro, G. A., Cireşan, D., Meier, U., Giusti, A., … Gambardella, L. M. (2011). Max-pooling convolutional neural networks for vision-based hand gesture recognition. *2011 IEEE International Conference on Signal and Image Processing Applications, (ICSIPA)*, 342–347. Retrieved from https://doi.org/10.1109/ICSIPA.2011.6144164

Najafi, P., Feizizadeh, B., & Navid, H. (2021). A comparative approach of fuzzy object based image analysis and machine learning techniques which are applied to crop residue cover mapping by using sentinel-2 satellite and uav imagery. *Remote Sensing*, 13(5), 937. Retrieved from https://doi.org/10.3390/rs13050937

Pacheco, A., & McNairn, H. (2010). Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. *Remote Sensing of Environment*, 114(10), 2219–2228. Retrieved from https://doi.org/10.1016/j.rse.2010.04.024

Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Blondel, M., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830. Retrieved from http://scikit-

learn.sourceforge.net.

Pouyanfar, S., Sadiq, S., Yan, Y., Tao, Y., Reyes, M. P., Shyu, M., … Iyengar, S. S. (2019). A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Computing Surveys*, 51(5), 92. Retrieved from https://doi.org/10.1145/3234150

Quemada, M., Hively, W. D., Daughtry, C. S. T., Lamb, B. T., & Shermeyer, J. (2018). Improved crop residue cover estimates obtained by coupling spectral indices for residue and moisture. *Remote Sensing of Environment*, 206, 33–44. Retrieved from https://doi.org/10.1016/j.rse.2017.12.012

Quemada, Miguel, & Daughtry, C. S. T. (2016). Spectral indices to improve crop residue cover estimation under varying moisture conditions. *Remote Sensing*, 8(8), 660. Retrieved from https://doi.org/10.3390/rs8080660

Ranaivoson, L., Naudin, K., Ripoche, A., Affholder, F., Rabeharisoa, L., & Corbeels, M. (2017). Agro-ecological functions of crop residues under conservation agriculture. A review. *Agronomy for Sustainable Development*, 37. Retrieved from https://doi.org/10.1007/s13593-017-0432-z

Richards, B. K., Walter, M. F., & Muck, R. E. (1984). Variation in line of crop residue cover. *Journal of Soil and Water Conservation*, 39(1), 60–61.

Riegler-Nurscher, P., Prankl, J., Bauer, T., Strauss, P., & Prankl, H. (2018). A machine learning approach for pixel wise classification of residue and vegetation cover under field conditions. *Biosystems Engineering*, 169, 188–198. Retrieved from https://doi.org/10.1016/j.biosystemseng.2018.02.011

Searle, A. S., & Bitnere, K. (2017). Review of the impact of crop residue management on soil organic carbon in Europe. *International Council on Clean Transportation*. Retrieved from https://theicct.org/publications/impact-of-crop-residue-mgmt-EU

Segaran, T. (2007). *Programming Collective Intelligence: Building Smart Web 2.0 Applications*. O'Reilly Media, Inc.

Shaheen, F., Verma, B., & Asafuddoula, M. (2016). Impact of Automatic Feature

Extraction in Deep Learning Architecture. *2016 International Conference on Digital Image Computing: Techniques and Applications, (DICTA) 2016*, 1–8. Retrieved from https://doi.org/10.1109/DICTA.2016.7797053

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–14. Retrieved from https://doi.org/10.48550/arXiv.1409.155

Singh, B.-, & Rengel, Z. (2007). *The Role of Crop Residues in Improving Soil Fertility*. *Nutrient Cycling in Terrestrial Ecosystems*. Retrieved from https://doi.org/10.1007/978-3-540-68027-7_7

Upadhyay, P. C., Lory, J. A., DeSouza, G. N., Lagaunne, T. A. P., & Spinka, C. M. (2022). Classification of Crop Residue Cover in High Resolution RGB Images Using Machine Learning. *Journal of the ASABE*, 65(1), 75–86. Retrieved from https://doi.org/10.13031/ja.14572

USDA-NRCS. (2011). National Agronomy Manual, 250.

Vasko, K., Toivonen, H. T. T., & Korhola, A. (2000). A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction. *Journal of Paleolimnology*, 24, 243–250. Retrieved from https://doi.org/10.1023/A:1008180500301

Weiss, K., Khoshgoftaar, T. M., & Wang, D. D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1). Retrieved from https://doi.org/10.1186/s40537-016-0043-6

Weltz, M. A., Huang, C. H., Newingham, B. A., Tatarko, J., Nouwakpo, S. K., & Tsegaye, T. (2020). A strategic plan for future USDA agricultural research service erosion research and model development. *Journal of Soil and Water Conservation*, 75(6), 137A-143A. Retrieved from https://doi.org/10.2489/JSWC.2020.0805A

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168(2). Retrieved from https://doi.org/10.1088/1742-

6596/1168/2/022022

Yu, L., & Liu, H. (2004). Redundancy based feature selection for microarray data. *KDD-2004 - Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 737–742. Retrieved from https://doi.org/10.1145/1014052.1014149

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, 818–833.

Zhang, Q. shi, & Zhu, S. chun. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology and Electronic Engineering*, 19, 27–39. Retrieved from https://doi.org/10.1631/FITEE.1700808

Zhang, Y., Tino, P., Leonardis, A., & Tang, K. (2021). A Survey on Neural Network Interpretability. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(5), 726–742. Retrieved from https://doi.org/10.1109/TETCI.2021.3100641

Zheng, B., Campbell, J. B., & de Beurs, K. M. (2012). Remote sensing of crop residue cover using multi-temporal Landsat imagery. *Remote Sensing of Environment*, 117, 177–183. Retrieved from https://doi.org/10.1016/j.rse.2011.09.016

Zheng, B., Campbell, J. B., Serbin, G., & Galbraith, J. M. (2014). Remote sensing of crop residue and tillage practices: Present capabilities and future prospects. *Soil and Tillage Research*, 138, 26–34. Retrieved from https://doi.org/10.1016/j.still.2013.12.009

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., … He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43–76. Retrieved from https://doi.org/10.1109/JPROC.2020.3004555

Zuo, Y., Chen, B., Shi, T., & Sun, M. (2020). Filter Pruning without Damaging Networks Capacity. *IEEE Access*, 8, 90924–90930. Retrieved from https://doi.org/10.1109/ACCESS.2020.2993932