# **Sorghum Cultivar Identification**

Veronica Thompson

Colorado State University Global

MIS 581: Capstone

Dr. Orenthio Goodwin

5/8/2022

## Abstract

Large-scale plant breeding programs for sorghum, an important cereal crop, will enable increased production of both food and bioenergy. Analysis of growing plants is an important component of a breeding program. When the breeding program is conducted on a large scale such analysis can best be accomplished using automated processing of sensing data. An example of automated analysis is cultivar identification from photos, which can be used to verify the integrity of propagation experiments. This project evaluates sorghum cultivar identification employing deep learning methods of computer vision. Candidate models include models completely trained using the sorghum images and pretrained models employing transfer learning. All models were compared to a baseline model consisting of a simple convolution neural network, using classification accuracy as the metric. For each candidate model variations including larger input images and data augmentation were examined. All models exhibited significant overfitting when trained without data augmentation. Most models showed increased accuracy when trained with larger images. The most successful model contains three VGG-inspired convolution blocks and employs data augmentation.


*Keywords:* Convolution neural network, deep learning, transfer learning, image classification

**Contents**

## Sorghum Cultivar Identification

Creating improved varieties of important crops is an important way to ensure sufficient sources of food and renewable energy supplies, especially given the challenges to agriculture posed by global warming. A key component of rapid plant breeding programs is the ability to evaluate plant characteristics on a large scale, especially from image sensing data (Minervini et al., 2015). Identifying plant cultivar from image data is an example the sort of plant evaluation which is desirable. Currently, there is no image classification model which is sufficiently accurate for large-scale sorghum cultivar identification. This work investigates possible image classification systems to be used to identify sorghum cultivars from plant images.

## Background

Sorghum is an important grain crop for both food and bio-fuel production with diverse uses, including livestock food, human food, and biofuel production (Ren et al., 2021). United States farmers planted 5.8 million acres of Sorghum in 2020, mostly in Kansas, Texas, Colorado, Oklahoma, and South Dakota (National Sorghum Producers, 2022). This increased by 24% to 7.3 million acres in 2021 (Grainnet, 2021). With better drought tolerance than other grains such as corn, Sorghum is likely to continue to increase in importance due to climate changes causes by global warming, increasing need for food for a growing global population, and increasing need for renewable sources of energy.

The Transportation Energy Resources from Renewable Agriculture Phenotyping Reference Platform (TERRA-REF) project's goal is to develop methods to evaluate sorghum cultivars using remote sensing and to use the evaluation methods to rapidly breed plants with higher yield (LeBauer et al., 2021). A dataset of sorghum plant images is available as part of the Sorghum-100 Cultivar Identification contest on Kaggle (Kaggle, 2022). The goal of the contest is to develop a machine learning algorithm to classify images by sorghum cultivar.

**Objectives**

An effective sorghum cultivar identification system will allow large scale verification of plant variety (Ren et al., 2021). This is useful to detect errors during planting, such as incorrectly labeled seeds or malfunctioning planters. Planting errors occur surprisingly often, and add uncertainly to plant variety trials (Kaggle, 2022).

**Overview**

Multiple deep learning image classification systems were used to identify sorghum by cultivar. The classification systems were trained on images of sorghum plants in mid-stage growth and compared based on their accuracy in identifying cultivars in previously unseen images.

**Research Question and Hypotheses**

This project created multiple classification models to answer the question: Which sort of deep learning, computer vision model is most promising for sorghum cultivar identification? Such a model could be used to check for planting errors in large-scale tests of sorghum varietals and could be extended for other uses in sorghum plant breeding programs.

As explained by Ren et al. (2021), knowledge learned from the cultivar identification system may be useful for the creation of models which identify additional plant attributes. For example, extended models may help scientists understand differences between cultivars, especially which show the most promise for use as bio-energy sources. They may be able to predict end of season yield from early season photographs, which will allow resources to be diverted from plants which slow little promise. This might occur in both breeding and production settings and could result in significant cost savings. Or they might determine if a plant is under stress, allowing a farmer to apply extra water before irreparable crop damage was incurred. Robust models for cultivar identification will provide a strong foundation for these other applications of computer vision to sorghum plant breeding and production.

**Hypotheses**

Multiple classification methods were tested on the Sorghum-100 dataset. For each classification model, the null hypothesis, $H_o$, is that there is no improvement in classification accuracy over the baseline classification model. The alternative hypothesis, $H_{A,}$ is that the candidate classification model does have better accuracy than the baseline model. Models are compared on accuracy rate on the validation set of data.

## Literature Review

Image classification has been a goal of computer vision researchers for many years. Prior to 2012, most researchers believed that image classification required experts to define the image features which were most important to distinguish between the images in the dataset (Krizhevsky et al., 2017). This changed in 2012, when a neural network image classification system significantly outperformed hand-crafted models in the annual ImageNet Large Scale Visual Recognition Challenge, or ILSVRC (Brownlee, 2020).

**Convolution Neural Networks**

The image classification system which won the 2012 ILSVRC was a deep learning convolution neural network (CNN) (Krizhevsky et al., 2017). Several well-known deep learning CNN architectures are currently considered state of the art for image classification (Picek, 2022). A CNN can extract image features and determine which features are important to the classification task at hand rather than requiring the features be preidentified (Maitra et al., 2015). As CNNs require fewer connections between layers than do standard feedforward neural networks, they require significantly less training time (Krizhevsky et al., 2017).

**Transfer Learning**

Transfer learning in machine learning is the use of a pre-trained model on a new but related problem (Brownlee, 2020). The model is pre-trained on a very large, standard dataset, and then finetuned on the smaller target dataset. Employing transfer learning allows effective models to be developed with significantly lower training resources. It can result in models with higher accuracy than the same model architecture trained from scratch on the target dataset. Pretrained versions of both deep multiple CNN models are available and were investigated as possible classification models for the sorghum cultivar classification (Brownlee, 2020).

**Fine-Grained Visual Categorization**

Fine-grained visual categorization (FGVC) describes a subset of computer vision image classification problems. In a dataset of images for FGVC, there is a small amount of variation between the classes and a large amount of variation within each class (Picek et al., 2022). The Sorghum-100 cultivar identification is a FGVC problem, as there are only small differences in appearance between cultivars and a large variation between images of the same cultivar due to differing lighted conditions and plant positions within the photos.

**Data Augmentation**

Data augmentation, or creating additional data samples from the existing set, is commonly used in machine learning (Brownlee, 2020). It is especially useful when there is insufficient data available to train the model without overfitting. Accuracy on the training dataset which is significantly higher than that on the validation dataset is evidence of overfitting. For image classification, common data augmentation methods include adding slight shifts from side to side or up and down, or horizontal or vertical flips of the image.

## Research Design

**Methodology**

Experiments were conducted to determine the effectiveness of a variety of image classification models, with the models compared with the metric of accuracy. This places the study in the quantitative tradition, where a theory and hypotheses are created, data is gathered and analyzed, and the analysis is used to draw conclusions. (O'Leary, 2021).

For a machine learning project, such as this image classification project, a hypothesis takes the form of a candidate model for mapping the input images to output categories (Brownlee, 2020). The set of hypotheses are the chosen models and model configurations used to map the input to the output. The hypotheses are tested by comparing each model's accuracy to a baseline model: those models which do not make more accurate predictions than the baseline model are disproven as improved models for classifying the images.

**Method**

The dataset employed is of sorghum plant images, available as part of the Sorghum-100 Cultivar Identification contest on Kaggle (Kaggle, 2022). The dataset includes 48,106 images of plants from 100 cultivars, with the 22194 images in the training data set labeled by cultivar. The images were captured in June 2017 as part of the TERRA-REF project, using an RGB camera mounted on an automated gantry system located over a test field at the University of Arizona, with the camera located 2 meters above the plant canopy (Ren et al., 2021). The plants pictured in the images are in the phase of their growth cycle where they are large but not yet falling over. The training dataset consists of two components, a set of 22,194 image files and a csv file. The csv file matches each image filename to a cultivar identifier. A data dictionary for the csv file is shown in Table 1. The image files each contain 1024 by 1024 pixels of RGB data, with pixels in the range 0 to 255.

**Table 1**

*Data Dictionary for Sorghum 100 Dataset CSV File*

| Variable name | Definition | Data type | Example |
|---|---|---|---|
| Image | Image file name | String | “2017-06-16__12-24-20-930.png“ |
| Cultivar | Cultivar identifier | String | “PI_257599“ |

Multiple deep learning models were evaluated for use on the Sorghum-100 dataset. The baseline classification system consists of two convolution layers followed by a pooling layer and a fully connected output layer. The baseline model is inspired by the first layer of the VGG model (Brownlee, 2020). One candidate model was a three-layer version of the VGG model, trained from scratch on the Sorghum-100 dataset. The other candidate models, VGG16 and ResNet50, were trained using transfer learning methods (He et al., 2016; Simonyan & Zisserman, 2015). The labeled portion of the dataset was partitioned into a training and validation dataset with 20% of the images in the validation set. Both baseline models and those in the hypotheses set were tested using the previously unseen validation dataset, using accuracy of classification as the metric. Models lacking improvements in accuracy over the baseline model would be disproven as possible classification models.

The Python programming language were used to implement the candidate sorghum classification systems. Several libraries were also required. Scikit-learn was used to partition images into training and validation sets (Pedregosa et al., 2011). The OpenCV library was used to resize images (Bradski, 2000). Keras was used for image pre-processing steps such as normalizing pixel values and data augmentation such as image flips and shifts (Brownlee, 2020; Chollet, 2015). Keras was also used to build the neural network models, and to access pre-defined and pre-trained models used for transfer learning.

**Limitations**

The test dataset contains between 150 and 300 images per cultivar. The use of more images per class is associated with better classification performance (Shahinfar et al., 2020). The number of images available may not be sufficient to train a high-accuracy model. Also, the images have significant differences due to lighting conditions which may increase the difficulty of creating a robust model (Kaggle, 2022). Finally, successful transfer learning requires that the images used for pre-training be similar to the target classification images (Hassani et al., 2021). As the sorghum images are rather different than the images found in the pre-training datasets such as ImageNet, transfer learning may not be a successful strategy.

**Ethical Considerations**

The Sorghum-100 dataset contains images which were collected for the purpose of research. No humans are pictured, so there are no issues of subject consent. The images are owned by TERRA-REF, which has the authority to give permission for their use. The dataset creators have provided the images for non-commercial uses including both Kaggle contest submission and academic research. As this project is academic and not commercial, it falls well within the use guidelines for the dataset.

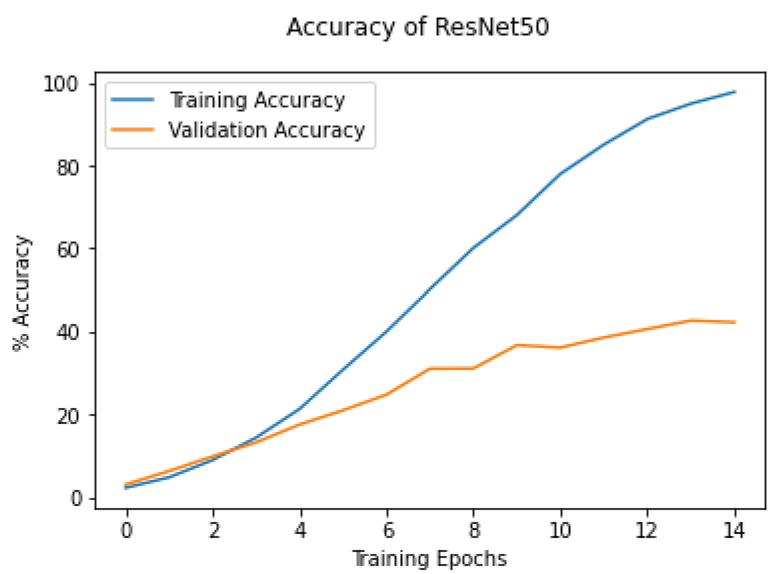
## Results

Almost all model variations outperformed the baseline image classification model. The only exception was the 1-block VGG-style model when trained on images of 224x224 pixels. All models were trained until no increase in accuracy was observed with further training. Every model except the baseline model showed increased accuracy when trained on larger images; training the baseline model with larger images resulted in earlier overfitting and lower accuracy. The most successful tested model was three VGG blocks combined with data augmentation which was trained from scratch. The accuracy results are summarized in Table 2.

Every model exhibited overfitting when trained using the original dataset. An example of the overfitting observed is shown in Figure 1, where the training and validation accuracy for the ResNet50 model is shown. In this case, the model was trained using images with 224x224 pixels and no data augmentation. Even though the ResNet50 models was the most successful model when trained without data augmentation, it still shows evidence of overfitting. Overfitting is indicated by the large difference between training and validation accuracy, with training accuracy approaching 100% at the end of the training period. Overfitting occurred to a significant degree on every model trained with the original dataset with no augmentation. It was successfully addressed through data augmentation. When employed, data augmentation consisted of random combinations of horizontal and vertical flips as well as horizontal and vertical shifts of up to 10% of the image pixels.

**Figure 1**

*Example of Overfitting*




**Baseline Model**

The baseline model consists of one block of layers inspired by the VGG architecture (Simonyan & Zisserman, 2015). Specifically, it includes two convolution layers with 64 filters of 3x3 pixels, followed by a pooling layer and a fully connected classification layer (Brownlee, 2020). Tested variations include input images of 224x224 pixels and data augmentation. The baseline model did not benefit from an increase in the size of the input images. Validation accuracy was actually lower with 224x224 pixel images than with 128x128 pixel images. This was true whether or not data augmentation was employed.

**VGG with 3 Blocks**

The next model tested was similar to the baseline mode, this time with three blocks of VGG-inspired convolution and pooling layers. This was the most successful model tested. When trained using 224x224 pixel images and combined with data augmentation, 65.0% accuracy was achieved on the validation dataset.

**VGG16**

A pretrained version of the VGG16 model was tested, in an application of transfer learning architecture (Simonyan & Zisserman, 2015). The feature identification layers of the model were used with no additional training. A fully connected classification layer was added to the pretrained portion of the model, and only this portion of the model trained on the Sorghum image data. Lack of time prevented testing this model with data augmentation enabled.

**ResNet50**

The pre-trained ResNet50 model was also tested (He et al., 2016). Again, a fully connected classification layer was added to the pretrained feature identification layers and only the classification layer received training specific to the Sorghum images. This was the most successful model without data augmentation, with a 42.3% accuracy rate on the validation data. However, this was the only model which did not have a large increase in accuracy when data augmentation was employed.

**Table 2**

*Accuracy Results for Evaluated Models*

| Model | Data Augmentation | Image Resolution | Validation Accuracy |
|---|---|---|---|
| 1 VGG Block (Baseline) | No | 128x128 | 4.9% |
| 1 VGG Block | No | 224x224 | 3.7% |
| 1 VGG Block | Yes | 128x128 | 35.9% |
| 1 VGG Block | Yes | 224x224 | 28.6% |
| 3 VGG Blocks | No | 128x128 | 8.0% |
| 3 VGG Blocks | Yes | 128x128 | 54.0% |
| **3 VGG Blocks** | **Yes** | **224x224** | **65.0%** |
| VGG16 | No | 128x128 | 15.5% |
| VGG16 | No | 224x224 | 18.7% |
| ResNet50 | No | 128x128 | 33.6% |
| ResNet50 | No | 224x224 | 42.3% |
| ResNet50 | Yes | 224x224 | 44.9% |

**Hypotheses**

Most models tested had increased accuracy over the baseline model. The only exception was the baseline model when trained on images of 224x224 pixels. The null hypothesis that no better model than the baseline exists is rejected. There are multiple models which provide better classification accuracy than the baseline model.

## Recommendations

None of the models provided sufficient accuracy for use to identify planting errors in a sorghum breeding program. Future work might focus on the two most successful models. The model using three VGG blocks might achieve an increase in performance through the addition of an additional block of convolution layers. Dropout might also improve model performance (Brownlee, 2020). Ren et al. (2021) achieved improved accuracy on the Sorghum-100 dataset by replacing max pooling layers with average pooling layers in their model, which is an additional idea to try with the three VGG block model. The ResNet50 model might benefit from fine-tuning of the model parameters for the sorghum data. This could be achieved by allowing the feature identification layer weights to be updated using the sorghum images. Inspired by the accuracy improvements achieved by increased image resolution and the work of Ren et al. (2021), training with full-resolution tiles of the images rather than scaled versions of the images is another approach that might improve the accuracy results.

## Conclusion

Image classification models were tested on images of sorghum plants to determine which model can most accurately distinguish cultivars. Important machine learning techniques were employed, including deep learning convolution neural networks, vision transformers, and transfer learning. The most successful model contained 3 convolution layers inspired by the VGG architecture. It was completely trained for the sorghum dataset with no application of transfer learning. With data augmentation, this model achieved 65.0% accuracy on the validation dataset.

**References**


Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brownlee, J. (2020). *Deep learning for computer vision: Image classification, object detection, and face recognition in Python*. Machine Learning Mastery.

Chollet, F., & others. (2015). Keras. https://keras.io

Grainnet. (2021, September 13). *USDA report boosts sorghum acreage to 7.3 million, most since 2015*. https://www.grainnet.com/article/246916/usda-report-boosts-sorghum-acreage-to-7-3-million-most-since-2015

Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., & Shi, H. (2021). Escaping the big data paradigm with compact transformers. https://doi.org/10.48550/arXiv.2104.05704

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2016.90

Kaggle. (2022). *Sorghum -100 cultivar identification - FGVC 9*. https://www.kaggle.com/competitions/sorghum-id-fgvc-9/overview

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90. https://doi.org/10.1145/3065386

LeBauer, D., Burnette, M., Fahlgren, N., Kooper, R., McHenry, K., & Stylianou, A. (2021). What does TERRA-REF's high resolution, multi sensor plant sensing public domain data offer the computer vision community? *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. https://doi.org/10.1109/iccvw54120.2021.00162

Maitra, D. S., Bhattacharya, U., & Parui, S. K. (2015). CNN based common approach to handwritten character recognition of multiple scripts. *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*. https://doi.org/10.1109/icdar.2015.7333916

Minervini, M., Scharr, H., & Tsaftaris, S. A. (2015). Image analysis: The new bottleneck in plant Phenotyping [Applications corner]. *IEEE Signal Processing Magazine*, *32*(4), 126-131. https://doi.org/10.1109/msp.2015.2405111

National Sorghum Producers. (2022). *Sorghum 101*. https://sorghumgrowers.com/sorghum-101/

O'Leary, Z. (2021). *The essential guide to doing your research project* (4th ed.). Sage Publications.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Picek, L., Sulc, M., Matas, J., Jeppesen, T. S., Heilmann-Clausen, J., Lassoe, T., & Froslev, T. (2022). Danish fungi 2020 – Not just another image recognition dataset. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/wacv51458.2022.00334

Ren, C., Dulay, J., Rolwes, G., Pauli, D., Shakoor, N., & Stylianou, A. (2021). Multi-resolution outlier pooling for sorghum classification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://doi.org/10.1109/cvprw53098.2021.00328

Shahinfar, S., Meek, P., & Falzon, G. (2020). “How many images do I need?” Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring. *Ecological Informatics*, *57*. https://doi.org/10.1016/j.ecoinf.2020.101085

Simonyan, K., & Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition*. In *ICLR*. https://arxiv.org/abs/1409.1556

TERRA-REF. (2022). The University of Arizona. https://terraref.org/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *NIPS'17: Proceedings of the 31st International Conference on*

*Neural Information Processing System* (pp. 6000-6010).

https://dl.acm.org/doi/10.5555/3295222.3295349