

HIGH-THROUGHPUT QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI IN
SORGHUM: A ROADMAP TO BUILDING ROBUST DEEP LEARNING–BASED TOOLS
FOR AUTOMATED IMAGE ANALYSIS

by

SHUFAN ZHANG

(Under the Direction of Jonathan Arnold and Thirimachos Bourlai)

ABSTRACT

Arbuscular mycorrhizal fungi (AMF) form one of the most ancient and widespread symbioses, enhancing plant nutrient acquisition in exchange for photosynthetically derived carbon. Accurately quantifying AMF colonization at scale, however, remains a major bottleneck. This dissertation presents a three-stage roadmap for developing robust, deep learning–based tools that enable high-throughput, automated analysis of AMF in sorghum (*Sorghum bicolor*) roots.

First, a pilot study combined Mask R-CNN with mixed linear models to segment individual fungal structures in a recombinant inbred sorghum population and to relate colonization levels to root niche and fungal structure allocation. The study demonstrated that deep learning can capture biologically meaningful AMF phenotypes.

Second, to overcome data scarcity in training deep learning models, I assembled MycorrhhiSEE, a 15 TB collection of ~137,500 whole-slide images (WSIs) from 5,500 sorghum plants spanning 337 genotypes and diverse field treatments. A spline-guided tiling algorithm transformed gigapixel WSIs into uniform patches. Eight bootstrap evaluations confirmed consistent spline interpolation across expert-rated image quality classes.

Third, building on MycorrhhiSEE, an integrated pipeline was developed featuring (i) an enhanced spline-guided tiling algorithm with quantitative tiling quality metrics, and a two-step CNN-based classification that (ii) first removes background tiles with 99.7 % accuracy and then (iii) distinguishes AMF colonized from non-colonized image tiles. A wide selection of ImageNet-pretrained architectures was benchmarked to identify the optimal classifiers. **DenseNet** and **ResNet50**-based classification classifier achieved >98% accuracy and superior generalization on both MycorrhhiSEE and the external AMFinder dataset.

Collectively, these contributions—from computer vision modeling to large-scale dataset curation and pipeline optimization—provide a practical framework for rapid, unbiased AMF phenotyping. The resulting tools enable scalable integration of imaging, genomic, and environmental data, advancing precision agriculture and ecological research on AMF to improve sorghum performance under diverse field conditions.

INDEX WORDS: Mycorrhizal fungi, Arbuscular mycorrhizal fungi, Computer vision, Image classification, Instance segmentation, Transfer learning, High-throughput imaging, Whole slide imaging.

HIGH-THROUGHPUT QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI IN
SORGHUM: A ROADMAP TO BUILDING ROBUST DEEP LEARNING-BASED TOOLS
FOR AUTOMATED IMAGE ANALYSIS

by

SHUFAN ZHANG

B.S., The University of Washington, USA, 2019

A Dissertation Submitted to the Graduate Faculty of
The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

Athens, Georgia

2025

© 2025

Shufan Zhang

All Rights Reserved

HIGH-THROUGHPUT QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI IN
SORGHUM: A ROADMAP TO BUILDING ROBUST DEEP LEARNING-BASED TOOLS
FOR AUTOMATED IMAGE ANALYSIS

by

SHUFAN ZHANG

Major Professor: Jonathan Arnold
Thirimachos Bourlai

Committee: Daniel Promislow
Jason Wallace
Kaixiong Ye
Katrien Devos

Electronic Version Approved:

Ron Walcott

Dean of the Graduate School

The University of Georgia

August 2025

ACKNOWLEDGEMENTS

I am deeply grateful to everyone who helped bring this work to fruition. First, I thank the entire Arnold Lab—Jia-Hwei Cheong, Amanda Marie Bouffier, Yingqing Guo, Lauren Stupp, Isabella Wilson, Anna Prestel, William Lantz, Stephen Eastin, and Jake Spittle—for their daily encouragement, technical expertise, and unwavering camaraderie. I am equally indebted to my committee and mentors—Jonathan Arnold, Thirimachos Bourlai, Daniel Promislow, Abhyuday Mandal, Jason Wallace, Kaixiong Ye, Katrien Devos, Jan Mrázek, and Travis Glenn—whose insight and guidance continually nurtured me to be an independent scientist.

This project also benefited from the generosity of our collaborators in the Johnson, Devos, Bennetzen, and Chung labs, whose perspectives enriched every stage of the work. I extend heartfelt thanks to the dedicated Institute of Bioinformatic and its staff—April King Mosley, Sandra Getz, and Jara Usherwood—for helping me navigate the complexity of the PhD and life, and to the team at the J. Phil Campbell Sr. Research and Education Center—JD Hale and Eric Elsner—whose expert field management made the five successful field seasons.

Finally, I gratefully acknowledge funding from the U.S. Department of Energy, without which this research would not have been possible. To each of you and my family, please accept my sincere appreciation; your support has been both a professional boon and a personal inspiration.

GENERATIVE AI USE DISCLOSURE

The author of this dissertation acknowledges the use of generative AI for the following purposes:

1. Develop outline— During the initial planning phase, the author engaged generative AI as a conversation partner posing questions from the perspectives of a novice reader and a field expert. This iterative exchange revealed content gaps and strengthened the logical flow of the writing plan. The AI then produced a formal outline based on the final plan, which was reviewed and approved by the doctoral committee.
2. Critique writing style—AI was prompted to perform the role of a writing coach to comment on clarity, conciseness, and coherence, and suggest alternative phrasings to improve scientific writing style. All text was subsequently rewritten or edited by the author.
3. Provide formatting assistance—AI was used to generate a Microsoft word template in accordance with university dissertation formatting guidelines.

No AI-generated content was accepted without manual review, editing, and verification by the author to ensure technical accuracy and scholarly integrity. All data, analyses, data interpretation, and scientific conclusions remain entirely the author's work. This use of generative AI was reviewed and authorized by the doctoral advisory committee.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	IV
GENERATIVE AI USE DISCLOSURE	V
LIST OF TABLES.....	VIII
LIST OF FIGURES	X
CHAPTER	
1 INTRODUCTION.....	1
1.1 PREFACE.....	1
1.1.1 Motivation and Research Context.....	1
1.1.2 Dissertation Objectives and Contributions	3
1.1.3 Dissertation Overview	4
1.2 ARBUSCULAR MYCORRHIZAL FUNGI (AMF).....	6
1.2.1 Evolutionary History of AMF.....	6
1.3 SORGHUM.....	8
1.3.1 Agricultural Relevance of Sorghum	8
1.3.2 Genetic Diversity and Genomic Resources	9
1.4 SORGHUM-AMF SYMBIOSIS.....	10

1.5 INTUITIVE INTRODUCTION TO DEEP LEARNING	11
1.5.1 Fundamentals of Deep Learning.....	11
1.5.2 Basics of CNN Architecture.....	13
2 COMPUTER VISION MODELS ENABLE MIXED LINEAR MODELING TO PREDICT ARBUSCULAR MYCORRHIZAL FUNGAL COLONIZATION USING FUNGAL MORPHOLOGY	18
3 MYCORRHISEE: A HIGH-RESOLUTION IMAGE DATASET FOR DEEP LEARNING BASED QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI & A NEW SPLINE-BASED IMAGE PROCESSING TECHNIQUE WITH APPLICATIONS TO WHOLE- SLIDE IMAGING OF PLANT ROOTS.....	56
4 HIGH-THROUGHPUT QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI COLONIZATION USING SPLINE-GUIDED TILING AND CROSS-DOMAIN TRANSFER LEARNING WITH WHOLE-SLIDE IMAGES	66
5 LIMITATIONS AND FUTURE DIRECTIONS	93
5.1 DATASET LIMITATIONS.....	93
5.2 TECHNICAL LIMITATIONS.....	95
5.3 METHODOLOGICAL LIMITATIONS.....	98
REFERENCES.....	102

LIST OF TABLES

	Page
Table 2.1 Performance of the Mask R-CNN compares favorably with published performance on other image datasets	25
Table 2.2 Mixed linear models are well predicted for percent colonization by arbuscule count, arbuscule size, region, and plant	36
Table 2.3 Mixed linear models are well predicted for count density by counts of hyphal structures, root regions, and sorghum accessions	39
Table 2.4 Multiple comparison test of fungal strucure count and size along root regions.	41
Table 2.5 Multiple comparison test of overall fungal colonization along root regions	42
Table 2.6 Annotation rules for masking AM fungal structures	51
Table 3.1 Overview of Experiments in the Dataset	59
Table 4.1 Comparison of convolutional neural networks used	77
Table 4.2 Tukey Post hoc test of root coverage and patch overlap between spline guided tiling algorithm versions	82
Table 4.3 Statistic properties of root coverage and patch overlap for two tile sizes	83
Table 4.4. Within domain classification accuracy for AMF colonization of single dataset models	
Within domain classification accuracy for AMF colonization of single dataset models	90

Table 4.5. Cross-domain classification accuracy for AMF colonization of single dataset models.....	90
Table 4.6. Classification accuracy for AMF colonization of combined dataset models.....	90
Table 4.7. Classification accuracy for AMF colonization of single dataset models on combined test set.....	90

LIST OF FIGURES

	Page
Figure 1.1 Visualize feature map after convolution operation.....	14
Figure 1.2 Importance of nonlinearity in deep learning models.....	15
Figure 2.1 Instance segmentation on the test set images of RIL plants	27
Figure 2.2 Distribution of confidence scores from the best performing Mask R-CNN model tested on Georgia in-house image dataset	28
Figure 2.3 Correlations between fungal structure counts and size	30
Figure 2.4 Experimental design entails hierarchical sampling from a RIL population	31
Figure 2.5 Percent colonization is well predicted by Model 6	38
Figure 2.6 AMF colonization and structure abundance vary with root depth	43
Figure 2.7 The density plots of AMF percent colonization, count density and proportion of nutrient exchange (Proportion NE)	44
Figure 2.8 Trends in relative abundance of AM fungal structures when the total AMF count density is assumed to be the same across sibling plants.....	44
Figure 3.1 Examples of whole slide scanning images, individual root segments in an image, and zoomed in fungal structures	58
Figure 3.2 A demonstration of the sample processing and image acquisition workflow	60
Figure 3.3 The histogram illustrates the quality score distribution of 1,780 WSIs	61

Figure 3.4 Demonstration of our spline-based method vs. standard sliding window method.....	62
Figure 3.5 Violin plots of spline interpolation qualities	63
Figure 4.1 Workflow of proposed approach to count AM colonized tiles using an enhanced spline-guided tiling algorithm, a data processing step, and two classifiers.....	70
Figure 4.2 Demonstration of high variability in root area in AMFinder tile images	76
Figure 4.3 Examples of ZEISS AxioScan 7 WSIs of root segments labeled as excellent, moderate, bad based on root conformation and image quality	79
Figure 4.4 Visualization of patch overlap and root coverage using the spline-guided tiling algorithm on ZEISS AxioScan 7 images	80
Figure 4.5 Comparing tiling quality of the algorithm on AMFinder and MycorrhizSEE datasets	83
Figure 4.6 ROC curves demonstrating the in- and cross-domain performances of classification classifiers.....	91

CHAPTER

1

INTRODUCTION

1.1 Preface

1.1.1 Motivation and Research Context

The Kingdom *Fungi* is among the most diverse groups of organisms on earth (Wijayawardene et al., 2024). Mycorrhizal fungi—derived from the Greek *myco* (fungus) and *rhiza* (root)—, form symbiotic association with the roots of over 80% of vascular plants (Brundrett & Tedersoo, 2018). Through their extensive hyphal networks, mycorrhizal fungi provide plants access to otherwise unavailable soil nutrients, in exchange for up to 30% photosynthetic carbon (Hawkins et al., 2023). It is estimated that at least 50,000 fungal species from the phyla Glomeromycota, Ascomycota, Basidiomycota and Mucoromycota form mycorrhizal associations (van der Heijden et al., 2015). Based on their morphology and function, four principal types of mycorrhizal symbioses have been identified: arbuscular mycorrhiza (AM), ectomycorrhiza (ECM), orchid mycorrhiza (ORM), and ericoid mycorrhiza (ERM), with AM symbioses accounting for approximately 72% of all associations (Brundrett, 2009).

In the context of global challenges—such as climate change, ecosystem conservation, and the need for sustainable agriculture—AM symbioses provide a promising alternative to chemical fertilizers and pesticides for enhancing plant health and yield. The symbiotic relationship between plant roots and arbuscular mycorrhizal fungi (AMF) has been shown to enhance the drought, heat, salt, herbivory and pathogen resistance of plants in the most arid and barren soils on earth (M. J.

Pozo & C. Azcón-Aguilar, 2007) (J. M. Ruiz-Lozano et al., 2012). The extensive underground hyphal network of AMF facilitates the uptake of water, nitrogen, phosphorus, and other ions (Kakouridis et al., 2022; Smith & Read, 2010). However, the outcome of this resource exchange is highly context dependent. It is influenced by both biotic factors—such as the plants and fungal genetic make-up—and abiotic factors, including soil conditions, climate and agricultural management practices which can shift the interaction along the parasitism–mutualism continuum (Feddermann et al., 2010; Graham & Abbott, 2000; Hart & Reader, 2002; Hoeksema et al., 2018; Johnson et al., 1997b; Klein et al., 2022; Klironomos & John, 1999; Mensah et al., 2015; Munkvold et al., 2004; Taylor & Harrier, 2000).

How can we effectively harness AMF for agriculture? Greenhouse studies testing combinations of plants, AMF inoculants, and soils are unlikely to be predictive of mycorrhizal functions in field settings. Field applications of commercial inoculants, often formulated with few “super” species, have produced inconsistent plant growth responses (Koziol et al., 2024; Lutz et al., 2023). It has become increasingly evident that AM symbioses are not isolated, binary interactions between a single fungus and plant host. Rather, they function as complex adaptive systems, with plant fitness outcomes shaped by the interplay of host genotype, microbial community, and environmental context.

Sorghum bicolor, a drought-tolerant C4 crop, has been cultivated as food, feed, and fuel for centuries across the world (Nigam et al., 2025). It is rich in micronutrients such as iron and zinc, as well as dietary fiber, antioxidant nutrients, and starch. Half a billion residents in Africa, Asia, and other semi-arid areas relies on sorghum as one of the cheapest nutrient sources. In 2021, the U.S. alone planted 7.3 million acres of sorghum (Tubiello et al., 2023). In western countries, sorghum is increasingly recognized for its value as animal feed and, more recently, as a bioenergy

crop. Its fibrous stalks, a byproduct of sugar extraction, can be fermented into ethanol and bio-jet fuel (Mullet et al., 2014). Enhancing beneficial AM symbiosis in sorghum can further reduce fertilizer and irrigation demands, increasing its value as a sustainable bioenergy crop in the rapidly changing global climate (Frew, 2019).

1.1.2 Dissertation Objectives and Contributions

My PhD research, funded by the Department of Energy and conducted as part of the collaborative project titled “Systems analysis of the beneficial associations of sorghum with arbuscular mycorrhizal fungi studied with genetics, genomics, imaging and microbiomics”, focuses on developing tools to support the large-scale quantification of AMF colonization. My research aims to generate a large high-quality image dataset from field-derived root samples and to develop robust high-throughput deep learning-based solutions for the automated quantification of AMF colonization in root images.

Traditional microscopy-based methods for quantifying AMF are labor-intensive, time-consuming, and subject to human bias, which limits scalability and cross-study comparability (Kokkoris et al., 2019; McGonigle et al., 1990a). Progress to automate AMF quantification using deep learning has been slow due to the lack of large, labeled image datasets (Evangelisti et al., 2021a). My research addresses this gap by assembling a diverse, high quality, publicly available image dataset and training deep learning algorithms that are robust across imaging platforms and protocols. This work is essential for producing consistent colonization phenotypes that can be integrated with metagenomic, transcriptomic, and microbiomic data to construct predictive systems models (Torres et al., 2025). These models will help discover the functional linkages between sorghum genotypes and their associated microbiomes, ultimately enhancing our ability to predict and optimize plant performance under diverse environmental conditions.

1.1.3 Dissertation Overview

The overarching goal of this dissertation is to create an automated, high-throughput solution for accurately quantifying arbuscular mycorrhizal fungi (AMF) colonization in sorghum root images. In Chapter 1, I provide necessary biology background on AM symbiosis and a gentle introduction to deep learning to prepare the readers for the following research chapters. Chapter 2 begins with a small pilot dataset collected during the first field season of the DOE project in 2021. We introduced a multi-scale modeling framework that uses deep learning (Mask R-CNN) (He et al., 2020) and mixed linear models (MLMs) (Searle et al., 2009) to segment and classify individual AM fungal structures, and to model variation in AMF colonization among root regions and samples of *Sorghum bicolor*. The challenges and experiences in image data curation and training the initial deep learning model helped form the methodological basis for the subsequent chapters.

Chapter 3 addresses the data bottleneck identified in Chapter 2. To scale up, a high-throughput imaging workflow for harvesting, clearing, staining and imaging field grown sorghum root samples was developed by incorporating whole slide imaging (WSI) technology. This chapter describes the collaborative effort over two years to curate a 15-Terabyte image dataset, MycorrhiSEE (S. Zhang, T. Bourlai, et al., 2024). The dataset comprises ~137,500 multi-gigapixel WSIs of fine roots from 337 sorghum genotypes in the Bioenergy Association Panel (BAP) (Brenton et al., 2016), annotated with image quality and colonization labels. Additionally, we present a supporting image preprocessing software that transform whole slide images into tile images suitable for deep learning algorithms (S. Zhang, W. Lantz, et al., 2024). This resource adds new value to BAP by providing a mycorrhizal phenotype. New AM symbiotic genes can be discovered to advance the genetic engineering of sorghum as a bioenergy crop. MycorrhiSEE also

provides a new challenge to the AI community to develop new computer vision tools for the image analysis of mycorrhizal fungi.

Chapter 4 builds on the MycorrhizSEE dataset. Recognizing the wide variability in image quality, we decompose the challenge of developing a generalized DL tool to segment individual AM fungal structures into incremental steps. We begin with training an image classification model to count the proportion of AMF colonized tiles in WSIs using high quality tiles only. A range of convolutional neural network (CNN) architectures, pretrained on large public image databases, is evaluated for their classification performance on our specialized dataset of root images. Those CNNs serve as baselines for future efforts to iteratively train with mixed-quality data.

Chapter 5 discusses the limitations of this dissertation and outlines future directions toward creating an optimal deep learning tool for AMF quantification. We emphasize the need for automatic image quality assessment for large-scale image datasets. The integration of image augmentation and domain adaptation techniques during supervised model training can further improve CNN performance as the number of labeled images increases. Self-supervised learning offers the potential to further reduce the reliance on manual annotations and to enable the shift from image classification to accurate instance segmentation of fungal structures. The chapter concludes with key lessons learned throughout this dissertation and presents a roadmap for successfully applying deep learning to automate image analysis in fields that are underexplored or demand scalable, high-throughput solutions.

1.2 Arbuscular Mycorrhizal Fungi (AMF)

1.2.1 Evolutionary History of AMF

The evolution of arbuscular mycorrhizal fungi (AMF) has played a critical role in shaping terrestrial ecosystems. The symbiotic relationship between AMF and early land plants may have facilitated the transition from aquatic to terrestrial ecosystems and the evolution of vascular and root systems (Brundrett, 2009; Field et al., 2012b). Fossil evidence of AM-like structures has been discovered in the rhizomes of *Aglaophyton major*, an Early Devonian plant that predates the evolution of true roots by approximately 30 million years (Brundrett, 2002; Taylor & Osborn, 1996; Taylor et al., 1995). Molecular phylogenies of symbiotic genes further support the notion that the algal ancestors of land plants were pre-adapted for AM symbiosis (Delaux et al., 2015).

Phylogenetic analyses based on ribosomal DNA and gene loci indicate that *Glomeromycota*, the phylum containing AMF, has a monophyletic origin dating back to over 450 million years ago (James et al., 2006; Schüßler et al., 2001; Wijayawardene et al., 2024). This contrasts with other mycorrhizal types, such as ectomycorrhizal and orchid mycorrhizas, which have arisen multiple times independently, suggesting parallel or convergent evolutionary origins. Regardless of its long evolutionary history, *Glomeromycota* exhibits strikingly low species-level diversity, with an estimated 300 to 1,600 species globally, far fewer than the 20,000 ectomycorrhizal species (Kivlin et al., 2011; Kõljalg et al., 2013; Öpik et al., 2013; TEDERSOO et al., 2012; van der Heijden et al., 2015). Remarkably, AMF form symbiotic associations with more than 20,000 plant species, including both lower plants (e.g., liverworts and hornworts) and higher plants (e.g., shrubs, trees, cereals, and herbs) (Genre et al., 2020; Humphreys et al., 2010; van der Heijden et al., 2015). This accounts for approximately 72% of all mycorrhizal associations in flowering plants (Brundrett, 2009). AMF are widely distributed, inhabiting environments from the sub-polar regions to the

tropical rain forest, and from the deserts to even some aquatic ecosystems (Davison et al., 2015; Rosendahl, 2008).

The functional diversity of AMF with limited species richness has led to several evolutionary hypotheses. One explanation is the presence of high intraspecific genetic heterogeneity. For example, *Rhizophagus irregularis* isolates only share ~50% of their genes. High copy number variation and an abundance of transposable elements in *R. irregularis* may enhance adaptability to environmental variations (Chen et al., 2018). More than 150,000 accessory genes have been identified across only six *R. irregularis* isolates, suggesting that local adaptation and functional diversity can evolve without speciation (Chen et al., 2018).

Another hypothesis proposes multilevel selection involving plant hosts, AMF and their associated microbiome (Johnson & Marín, 2024). AMF often engage in cross-feeding relationships with the rhizosphere (root) and hyphosphere (hypha) bacteria (e.g., P-solubilizers, N-fixers, and Fe-providers) and can host hundreds of endobacterial taxa. These microbial consortia may be selected with the plant host as functional teams, as the teams confer the highest fitness advantage under certain environmental conditions. Evidence for this speculation includes studies showing that local AMF isolates provide greater growth benefits to native plant hosts compared to non-local combinations (Banerjee et al., 2018; Lutz et al., 2023).

These hypotheses underscore the complexity of AM symbiosis and the number of variables that need to be considered to understand the ecological and evolutionary drivers of AMF function. Currently, there are expanding efforts to link AMF taxa and genotypes with their functional traits across environmental conditions (Antunes et al., 2025; Corradi et al., 2024; Mathieu et al., 2018). As these studies grow in scale, there is an increasing demand for high throughput plant phenotyping and imaging-based quantification of AMF colonization, in conjunction with genomic

and microbiome profiling using Next Generation Sequencing (NGS) These integrative approaches will keep advancing our understanding of AMF functional diversity and its role in shaping plant fitness and ecosystem resilience.

1.3 Sorghum

1.3.1 Agricultural Relevance of Sorghum

Sorghum [*Sorghum bicolor* (L.) Moench] is a stress resistant C4 grass native to Africa and currently ranks as the fifth most cultivated cereal crop worldwide, following rice, wheat, maize and barley. Sorghum is highly diverse, with cultivars broadly classified into three categories: grain sorghum, biomass sorghum, and sweet sorghum (Silva et al., 2021). Grain sorghum serves as a staple crop that provides food security for more than ¼ of the world. It is a vital source of starch and micronutrients in arid and semi-arid regions. Biomass and sweet sorghum, collectively termed bioenergy sorghum, are cultivated for their rapid accumulation of structural and nonstructural carbohydrates (Brenton et al., 2016). Their sugary milk can be fermented into bioethanol. The fibrous residuals can be converted into biojet fuel or used as animal feed (Wu et al., 2010).

Sorghum's C4 photosynthetic pathway contributes to its exceptional water- and nitrogen-efficiencies (Enciso et al., 2015; Gardner et al., 1994; Weissmann & Brutnell, 2012). This makes sorghum particularly attractive for sustainable bioenergy applications and climate resilient agriculture. Unlike C3 cereal crops, which have fibrous and highly branched root systems with high nutrient acquisition capacity, sorghum plants have a thicker cortex and depend more heavily on AMF for nutrient uptake (HETRICK et al., 1988; Liu et al., 2021; Wilson & Hartnett, 1998). As a result, enhancing mutualistic AMF symbiosis in sorghum offers a promising strategy to improve plant resilience and productivity even in soils affected by salinity or heavy metal

pollutions (J. M. Ruiz-Lozano et al., 2012) (Chandrasekaran et al., 2016; Dhawi et al., 2016; Riaz et al., 2021).

1.3.2 Genetic Diversity and Genomic Resources

Sorghum was the first C₄ grass to have its genome sequenced (BTx623) using shotgun sequencing. This advancement established a foundation for sorghum genetic studies and the development of genomic resources. Its small diploid genome ($2n = 20$; ~730Mb) makes it a simpler model for functional genomic research than other C₄ cereal crops (Paterson et al., 2009).

One of the key genetic resources is the Bioenergy Association Panel (BAP), which includes a total of 390 sorghum accessions comprising 238 high biomass sorghum and 152 sweet sorghum lines from the National Plant Germplasm System (NPGS). The BAP represents historical and commercially relevant lines from all five major sorghum races (bicolor, caudatum, durra, guinea, and kafir) and three continents (Africa, Asia, and the Americas). These accessions display extensive variation in key traits, including stalk height, photoperiod sensitivity, development speed, and anthracnose resistance (Brenton et al., 2016).

The original genotyping of BAP accessions identified a set of 232,303 single nucleotide polymorphisms (SNPs) using Genotype by Sequencing (GBS). More recently, whole genome sequencing (WGS) of 365 accessions has yielded ~19.7 million SNPs and ~2.7 million indels. A filtered set of ~5.48 million high quality SNPs has facilitated more accurate and comprehensive analyses of population structure and genetic diversity (Kumar et al., 2024). Genome-wide association studies (GWAS) utilize association mapping panels like BAP to identify molecular markers associated with quantitative traits in plants (Zhu et al., 2008). SNPs associated with a variety of phenotypic traits have been identified in sorghum, including traits related to plant architecture (Hu et al., 2019; Kumar et al., 2023; Morris et al., 2013; Zhao et al., 2016), agronomy

(J. Lucas Boatwright et al., 2022; Boyles et al., 2017; Brown et al., 2006; Chopra et al., 2017; Li et al., 2018; Mace et al., 2013; Rhodes et al., 2014), bioenergy (Boyles et al., 2019; Brenton et al., 2016; Murray et al., 2009; Souza et al., 2021), and biomass and its compositional traits (J Lucas Boatwright et al., 2022; Brenton et al., 2016; Brown et al., 2008; Kumar et al., 2024; Murray et al., 2009; Zhang et al., 2015).

Additional genomic resources include the Sorghum Association Panel (SAP), Nested Association Mapping (NAM) populations, and recombinant inbred lines (RIL) with well-characterized genetic markers and diverse phenotypes (Boatwright et al., 2021; J. Lucas Boatwright et al., 2022; Govindarajulu et al., 2021). Genomics databases, including Phytozome (Goodstein et al., 2012), Gramene (Ware et al., 2002), SorghumBase (Gladman et al., 2022), and the Sorghum QTL Atlas (Mace et al., 2019), enable data integration and easy access.

1.4 Sorghum-AMF Symbiosis

AMF are important symbionts in sorghum that improve nutrient acquisition, water-use efficiency, and stress tolerance. The establishment and function of AM symbiosis in sorghum are influenced by fertilization, cover cropping, intercropping, tillage and land use forms, and irrigation (Abdelhalim et al., 2020; Birhane et al., 2018; Cobb et al., 2017; Egboka et al., 2022; Figueiredo de Oliveira et al., 2025; Moura et al., 2022; Tuheteru et al., 2020; Walder et al., 2015; Wipf Heidi et al., 2021).

Several studies have reported that AMF can regulate stress response and water regulation pathways in sorghum (Putri et al., 2023; Symanczik et al., 2020; Varoquaux et al., 2019), as well as the expression of phosphate transporters (PTs) and ammonium transporters (AMTs) (Koegel et al., 2013; Walder et al., 2015). AMF colonization has been linked to increases in sorghum height, biomass, grain production and micronutrient content under drought, salinity stress and phosphorus

deficient conditions (Bagayoko et al., 2000; Birhanu et al., 2024; Chandra et al., 2022; Cobb et al., 2016; Nakmee et al., 2016).

However, those benefits are often sorghum genotype dependent (Watts-Williams et al., 2022). Also, temporal variation in AMF community composition in sorghum roots suggests that timing plays a key role in the efficacy of AMF-based interventions (Gao et al., 2020). Despite its relevance, there is a limited number of systems-level studies examining AMF symbiosis in genetically diverse sorghum populations using multi-omic approaches (Kaur et al., 2022; Sawers et al., 2017; S. J. Watts-Williams et al., 2019). Notably, the AMF colonization of BAP accessions has not yet been systematically characterized. This gap highlights the need for integrative studies that consider host genetic diversity, root microbiome dynamics, and functional AMF responses—particularly in the context of breeding more symbiosis-responsive cereal crops (Sawers et al., 2018).

1.5 Intuitive Introduction to Deep Learning

1.5.1 Fundamentals of Deep Learning

Machine Learning (ML) represents algorithms that allow computers to identify patterns from data and make decisions or predictions on new unseen data without explicit programming commands. Traditional ML algorithms require manual extraction of relevant features from unstructured data, including text, images, audios, etc. Algorithms like support vector machines (SVMs), random forest, and k-nearest neighbors have been tested effective for classification tasks (Deisenroth et al., 2020).

Deep learning (DL) refers to a specialized subset of ML that utilizes neural networks to directly extract feature representations from unstructured data. Inspired by human brain's neural structure, neural networks consist of multiple layers of interconnected nodes (neurons), hence

named deep learning. Deep learning has demonstrated remarkable success in image analysis, also known as computer vision (Younesi et al., 2024). DL algorithm can perform various vision tasks.

Ordered with increasing complexity:

- Classification determines whether a specific object (e.g., a balloon) appears in an image, answering the question "What is in this image?"
- Object detection recognizes all instances of an object within an image and locates them with bounding boxes, addressing both “what” and “where” questions.
- Semantic segmentation identifies all pixels belonging to a certain class, but it does not distinguish between individual instances.
- Instance segmentation combines object detection with semantic segmentation to identify individual object instances at the pixel level.

Convolutional neural networks (CNNs) are particularly suited for computer vision tasks.

The theoretical foundations of CNN were established in 1988. In 1994, Yann LeCun introduced LeNet-5, a pioneering architecture with practical implementation in handwritten digit recognition. LeNet-5 incorporated the three essential components still found in modern CNNs: convolutional layers, pooling layers, and non-linear activation functions. Limited by computation power at the time, LeNet-5 only has five trainable layers: three convolutional layers and two pooling layers (Lecun et al., 1998).

In 2010, Graphics Processing Units (GPUs) was used for training CNNs for the first time, increasing the depth to nine layers. The release of large-scale labeled datasets, particularly ImageNet further accelerated the evolution of CNNs (Deng et al., 2009). In 2010, Krizhevsky et al. developed AlexNet and won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a large margin, compared to the traditional computer vision methods (Krizhevsky et al., 2017).

Since then, the depth and width of CNNs has kept expanding with increasing sophisticated designs. By 2015, ResNet have reached over 100 layers and achieved higher accuracy than humans in computer vision tasks in natural image domain (He et al., 2016). Many of the foundational CNN models from the era continue to be widely applied and adapted to new computer vision tasks.

In the sections that follow, I will provide an intuitive introduction to the fundamentals of CNN architectures and their key components, including input representation, output format, convolutional and pooling layers, fully connected layers, kernels, activation functions, loss functions, backpropagation, and optimization techniques. Then, I will review key foundation models and their architectural innovations, followed by an introduction to learning paradigms and advanced strategies such as transfer learning and domain adaptation.

1.5.2 Basics of CNN Architecture

Input representation

Images in CNNs are represented as numerical tensors. A color image has three channels—red, green, and blue (RGB). Each channel is represented as a 2D matrix of pixel values covering an image. When the color channels are stacked, the image becomes a 3D tensor with shape height $\times$ width $\times$ channels, where each pixel value encodes color intensity. For example, a 28×28 RGB image is represented as a $28 \times 28 \times 3$ tensor. The total number of possible colors is $256^3 = 16,777,216$. Black corresponds to (0, 0, 0), and white to (255, 255, 255). Often times, we standardize the input image so that the pixel values range from 0 to 1.

Kernels and convolutional layers

In CNNs, neurons in convolutional layers are powerful feature extractors, also called filters or kernels. A kernel is a small matrix (e.g., 3×3 or 5×5) that slides across the input tensor and

computes a dot product at each location. The values of the dot products are saved to a new matrix known as the feature map. The size of the resulting feature map depends on three parameters:

1. **Kernel size:** The spatial dimensions of the filter (e.g., 3×3 , 5×5)
2. **Stride:** The step size of the sliding kernel (e.g., a stride of 1 moves the kernel one pixel at a time)
3. **Padding:** Additional zero pixels added to the input borders (e.g. a padding of 1 changes the dimension of input matrix from 28×28 to 30×30)

For example, applying a 3×3 kernel with stride 1 and no padding to a 28×28 input yields a 26×26 feature map. To maintain the input dimension, a padding of 1 is required. For multi-channel inputs, the convolution operation is performed independently on each channel and produces three feature maps. CNNs typically employ multiple kernels per layer, with more kernels in deeper layer, increasing the network width and learning increasingly abstract features. Visualization of trained CNNs like AlexNet revealed that lower-level kernels capture simple features (e.g., edges, corners), while deeper layers capture higher-order patterns (e.g., textures, shapes). Feature maps highlight the presence of learned features in the original image (**Fig. 1.1**).



Figure 1.1 Visualize Feature Map after Convolution Operation. The original lion image is shown on the left. The feature map after convolution operation with edge detection kernel is shown on the right.

Pooling layer

A pooling layer comes after convolutional layers to downsample the feature maps. The goal of the pooling layers is to retain high level feature representations in the feature maps within as low spatial dimension as possible, which reduces the computational requirements. They also help mitigate overfitting by making the network less sensitive to exact position of the feature in the input image. The most common method, max pooling, slides a window (e.g., 2×2) across the feature map and retains the maximum value within the region. For example, applying 2×2 max pooling to a 28×28 feature map produces a 14×14 output, halving the spatial resolution. Average pooling, which retains the mean value of each region, is an alternative used in some architectures.

Activation function and forward propagation

Activation functions introduce non-linearity, enabling the network to learn complex patterns (**Fig. 1.2**). During forward propagation, the convolutional operation is mathematically described as:

$$Z = W * A + b \text{ (Equation 1)}$$

Where W is the input tensor, A is the kernel, b is the bias term, $*$ denotes the convolution operation.

The activation function g is then applied to produce the activated output:

$$A' = g(Z) \text{ (Equation 2)}$$

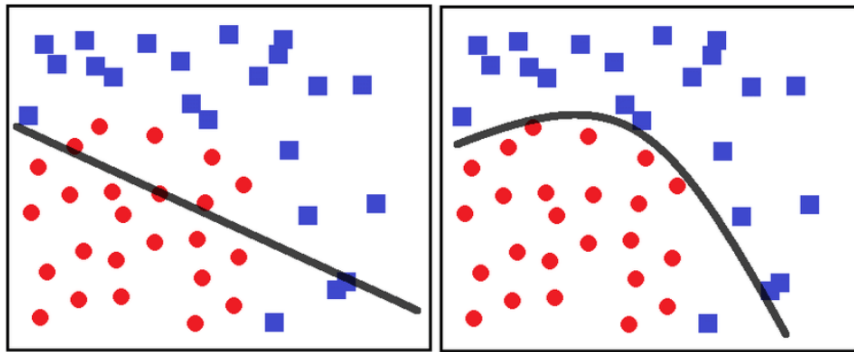


Figure 1.2 Importance of Nonlinearity in Deep Learning Models. This simple example scenario demonstrates that a non-linear activation function provides more flexibility and better separation of data than a linear activation function.

The output A' serves as the input to the next layer. The Rectified Linear Unit (ReLU), defined as $f(x) = \max(0, x)$, is a commonly used activation function in DL for a faster and more stable convergence during training (Nair & Hinton, 2010).

Fully connected layers and Output representation

So far, we have discussed how convolutional layers, pooling layers and activation functions extract hierarchical features from image data. The final stage in a CNN architecture involves converting these learned feature representations into task specific predictions.

Fully connected (FC) layers are derived from the multilayer perceptron (MLP) and are now commonly used as the classifier in CNNs for generating predictions. Unlike convolutional layers—where neurons are kernels and inputs are tensors—neurons in fully connected layers are scalar nodes that hold a single activation value and take vectors as inputs. Despite structural differences, neurons in both layers compute dot products, and their mathematical formulations are functionally equivalent (Equation 1 & 2).

Before passing data into the fully connected layers, the feature maps are flattened from multi-dimensional tensors into one-dimensional vectors. These vectors serve as input to the FC layers, where each neuron is connected to every activation in the previous layer. The dense connections of the FC layers enable the global combination of all features extracted from previous steps.

The architecture of the output layer depends on the computer vision task requirements. For image classification, the output layer typically uses a Softmax activation function to transform the output vector into a probability distribution across predefined classes. Object Detection employs two parallel FC branches: one for classifying objects, and another for predicting bounding box coordinates, often using regression-based loss functions. Semantic and Instance Segmentation may

combine FC layers with fully convolutional networks (FCNs) to produce class labels, bounding boxes, and pixel-level masks for detected objects.

Now, we understand the key components of CNNs and how CNNs extract features and make predictions in forward propagation. This provides the basics for understanding the state-of-the-art DL architectures that we selected to use in Chapters 2 and 4.

CHAPTER

2

COMPUTER VISION MODELS ENABLE MIXED LINEAR MODELING TO PREDICT ARBUSCULAR MYCORRHIZAL FUNGAL COLONIZATION USING FUNGAL MORPHOLOGY¹

¹Zhang, S., Y. Wu, M. Skaro, J-H. Cheong, A. Bouffier-Landrum, I. Torres, Y. Guo, L. Stupp, B. Lincoln, A. Prestel, C. Felt, S. Spann, A. Mandal, N. Johnson, & J. Arnold. 2024. *Nature Scientific Reports*. <https://doi.org/10.1038/s41598-024-61181-5>. Reprinted here with permission of the publisher.

Abstract

The presence of Arbuscular Mycorrhizal Fungi (AMF) in vascular land plant roots is one of the most ancient of symbioses supporting nitrogen and phosphorus exchange for photosynthetically derived carbon. Here we provide a multi-scale modeling approach to predict AMF colonization of a worldwide crop from a Recombinant Inbred Line (RIL) population derived from *Sorghum bicolor* and *Sorghum propinquum*. The high-throughput phenotyping methods of fungal structures here rely on a Mask Region-based Convolutional Neural Network (Mask R-CNN) in computer vision for pixel-wise fungal structure segmentations and mixed linear models to explore the relations of AMF colonization, root niche, and fungal structure allocation. Models proposed capture over 95% of the variation in AMF colonization as a function of root niche and relative abundance of fungal structures in each plant. Arbuscule allocation is a significant predictor of AMF colonization among sibling plants. Arbuscules and extraradical hyphae implicated in nutrient exchange predict highest AMF colonization in the top root section. Our work demonstrates that deep learning can be used by the community for the high-throughput phenotyping of AMF in plant roots. Mixed linear modeling provides a framework for testing hypotheses about AMF colonization phenotypes as a function of root niche and fungal structure allocations.

2.1 Introduction

Most vascular land plants have lived in symbiotic association with Arbuscular Mycorrhizal Fungi (AMF) for more than 400 million years (Bonfante & Genre, 2010). The plant provides carbon (C), and in return the AMF provide Nitrogen (N) and Phosphorus (P). This exchange of nutrients is central to tree diversity in forests worldwide (Zhong et al., 2021), determination of CO₂ in the atmosphere (Field et al., 2012a), and plant tolerance to drought, heat and pathogens (Gao et al., 2022; M. J. Pozo & C. Azcón-Aguilar, 2007; J. M. Ruiz-Lozano et al., 2012). The development of the AMF symbiosis is initiated by a single fungal hypha contacting a neighboring host root (Buee et al., 2000; Choi et al., 2018). Insertion of the epidermal layer by the prepenetration apparatus (Genre et al., 2008) is followed by intraradical hyphal growth. On reaching the inner cortex, branches arising from the intraradical hyphae could penetrate the cortical cell walls and form arbuscules known as the structure for nutrient transfer between symbionts (Pumplin & Harrison, 2009). Post-penetration development includes the differentiation of vesicles (Smith & Read, 2010) and spores (Marleau et al., 2011). Vesicles are nutrient storing structures for lipids and carbohydrates obtained from the plant host. AMF reproduce asexually using spores. The extensive extraradical hyphal network uptakes nitrogen, phosphorus, and other ions in soil that were otherwise inaccessible to the plant host. The various types of mycorrhizal structures differentiate from one continuum of fungal hyphae (Kokkoris et al., 2020) and can occur simultaneously in plant roots (Montero et al., 2019). Seminal work has shown that hyphal length, as well as spore counts and density, can vary significantly among conspecific AMF isolates, and that this variation has been shown to be correlated with differences in plant growth (Koch et al., 2006).

The internal development of the fungus is influenced by the plant genome (De Vita et al., 2018b; Michelle L. Pawlowski et al., 2020; Plouznikoff et al., 2019). An AMF species shows different morphological growth patterns, Arum- vs Paris-types, depending on the species of the plant partner in the association (Gerdemann, 1965; Jacquelinet-Jeanmougin & Gianinazzi-Pearson, 1983; Smith & Smith, 1997b). Large variation in AMF richness and abundance has been characterized in several plant populations, in an effort to identify symbiosis-associated genes (De Vita et al., 2018b; Johnson et al., 2022; Lehnert et al., 2017; M. L. Pawlowski et al., 2020; Plouznikoff et al., 2019; Stahlhut et al., 2021; S. J. Watts-Williams et al., 2019). Plant mutants were generated for biological validation of symbiosis genes (MacLean et al., 2017). DELLA proteins were revealed as master regulators that interact with the symbiosis signaling pathway, which provides a mechanism to integrate symbiosis with plant growth and development (Davière & Achard, 2013; Gallego-Bartolomé et al., 2012). For example, DELLA transcription and protein stabilization serves to restrain plant growth but to promote arbuscule development (Floss et al., 2013; Jiang et al., 2007). Direct evidence from the greenhouse highlighted that the functioning of colonization depends not only on the plant genotype but also on the identity of AMF genera/species/isolates (Stephanie J. Watts-Williams et al., 2019). The relative allocation to selfish versus non-selfish fungal structures (Johnson et al., 2003) also depends on the abiotic environmental conditions. Fertilization often reduces allocation to extraradical hyphae and arbuscules relative to other structures (Johnson et al., 2003). The genotypes of the organisms involved and the environmental conditions under which they interact determine the functioning of mycorrhizal association along the mutualistic-parasitic continuum (Feddermann et al., 2010; Graham & Abbott, 2000; Hart & Reader, 2002; Johnson et al., 1997b; Klironomos & John, 1999;

Mensah et al., 2015; Munkvold et al., 2004; Taylor & Harrier, 2000). A better understanding of the factors is needed.

The AMF research community is limited by a lack of cost efficient and high-throughput imaging methods to quantitate the abundance of AMF hyphal structures in roots. In 1990, McGonigle et al developed an unbiased approach for scoring AMF colonized root samples (McGonigle et al., 1990a). It is the gold standard until now, but it is laborious and demands skilled human scorers. Molecular quantification methods like AMF-specific phospholipid fatty acids (PLFA) approximate the amount of AM fungal biomass. DNA-based methods like quantitative real-time PCR (qPCR) allow quantification of specific AMF taxa in roots and soil. Amplicon sequencing allows the measurement of relative abundance of AMF taxa in root samples. A disadvantage of the PLFA- and DNA-based approaches is that they cannot measure colonization and morphology at the fungal structure level. Microscopy methods are synergistic by quantitating fungal structures and their morphology inside roots (McGonigle et al., 1990a; Trouvelot et al., 1985). Imaging, however, requires human scorers and the process is laborious and repetitive. Preparation and visual examination of 1,000 AMF slides with 20-30 root segments per slides takes an experienced researcher 2 months to complete. A computer vision model could potentially carry out this task in a few hours.

Machine learning has been applied to fungal image classification even with limited training data in *Neurospora crassa* (Krach et al., 2022; Krach et al., 2020). A deep learning-based software, AMFinder, was developed to automate the process of quantifying AMF colonized root images (Evangelisti et al., 2021a). The examples demonstrated computer vision as a powerful tool for high-throughput AMF phenotyping. Further improvements remain to quantitate the allocation to

AM fungal structures and their morphological phenotypes in the roots using a newly available instance segmentation method of computer vision model.

Instance segmentation using deep learning techniques, like Mask R-CNN (K. M. He et al., 2017), offers an opportunity for accurate and robust detection and per-pixel segmentation of different hyphal structures in root images. With the image analysis on the inferred segmentations, hyphal length/width, hyphal branching frequency, arbuscule length/width, vesicle size, spore size and other morphological traits can be automatically measured. These morphological traits can be correlated with various biological and physical processes of plants, such as photosynthesis, respiration, transpiration, and carbon and nutrient assimilation, which can be very useful for quantitative trait locus (QTL) mapping (Plouznikoff et al., 2019) and Genome-wide Association Studies (GWAS) (De Vita et al., 2018b) for symbiotic gene discovery.

Transfer learning is a technique that helps to transfer features learned from one dataset to another. The advancement of transfer learning benefits applications with limited annotated data. As of 2020, Mask R-CNN is one of the few deep learning architectures that can provide a generalist performance for instance image segmentation (He et al., 2018). Transfer learning-based application of Mask R-CNN have been adopted rapidly for imaging-based plant phenotyping in recent years (Ferguson et al., 2021).

We present a Mask R-CNN based image analysis method that provides the four previously unavailable advantages: (1) requires a minimal training data via transfer learning 2) achieves pixel level identification of multiple AM fungal structures via instance segmentation; 3) works on root samples colonized by a mixed populations of AMF in the field; 4) provides morphological measures on each category of AM fungal structure. We took the quantification and morphological measures from the image analysis to address fundamental questions about the AMF symbiosis: (1)

can a mathematical model be developed to predict AMF colonization; (2) does the allocation to AM fungal structures vary between plants; (3) are there differences in the niche within the root system, where AMF structures are found?

To understand AMF symbiosis as part of largescale systems biology studies, we developed a deep learning-based image analysis method to automatically measure AMF colonization intensity and fungal structure morphologies. The mixed linear model was used to provide a framework for testing hypotheses about AMF colonization and the variation in these morphometric measures. The result is a direct connection between the fungal structures present in each root sample and fungal colonization of the roots. This connection will permit the exploration of how AMF affect plant health through allocation to their structures.

Table 2.1 Performance of the Mask R-CNN compares favorably with published performance on other image datasets. The training set 1 is made of 767 in-house images on the Georgia samples. Training set 2 contains additional images from the AMFinder dataset (Evangelisti et al., 2021a; Tukey, 1949). The model performance is measured by average precision (AP) of each class, mean average precision (mAP) and mean average precision at intersection over union threshold of 0.5 (mAP50) (K. M. He et al., 2017)

Model	Training data	Testing data	AP							mAP50
			Root	Arbuscule	Extraradical hypha	Intraradical hypha	Vesicle	Spore	Non-AM	
A	In-house	In-house	46.2	29.6	6.7	7.4	21.9	55.8	13.7	47.5
						<u>mAP</u> 25.9				
B	Combined	In-house				<u>mAP</u> 21.4				39.7
C	Combined	Combined				<u>mAP</u> 29.6				50.2

2.2 Results

2.2.1 Performance of Mask R-CNN on AMF Image Segmentation

Our Mask R-CNN model can segment AMF colonized root images with satisfying performance (**Table 2.1**). The training images and annotations were generated by human scorers using the McGonigle method (McGonigle et al., 1990b) on a grid associated with the 192 root intersections per slide (See Materials and Methods). The pretrained Mask R-CNN model on the COCO dataset was loaded and trained on an in-house dataset with the default augmentation including image random flip and resize and a 0.7 confidence score threshold (**Table 2.1 Model A**) showing higher performance on our in-house testing images with 25.9 mean average precision (mAP) and 47.5 mean average precision at intersection over union (IoU) threshold of 0.5 (AP50) across classes, comparable to the performance of Mask R-CNN on other public datasets. For each class, the average precision (AP) captures both the precision (related to type I error) and recall (power=1-type II error) for IoU from 0.5 to 0.95 with a 0.05 step interval. Example results presented the agreement between model prediction and the ground truth (**Fig. 2.1**). The Mask R-CNN excelled at segmenting sorghum root and spore with AP values larger than 40 (**Table 2.1**).

Reasonable performance was achieved on arbuscule and vesicle with AP ranging from 20 to 30.

The model struggled with predicting instances of intraradical and extraradical hyphae.

Prediction

Ground

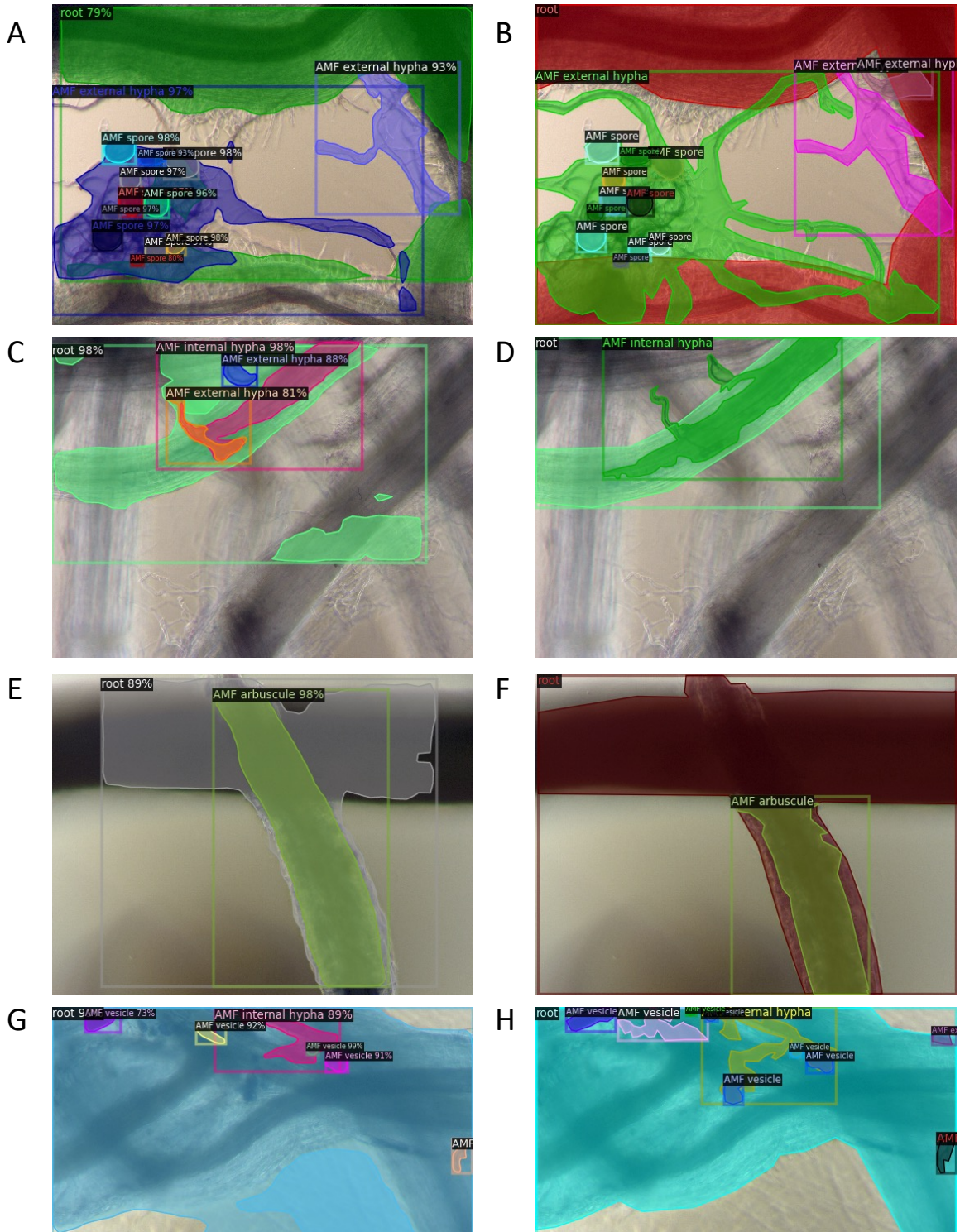


Figure 2.1 Instance segmentation on the test set images of RIL plants. The left column (A, C, E, G) shows Mask R-CNN predictions of all five AM fungal structures: arbuscule, external hypha, internal hypha, vesicle, and spore. Random color is used to fill individual masks. Classification with confidence score is labeled on the corner of a bounding box. The right column (B, D, F, H) displays the ground truth annotations from human scorers.

The confidence score distributions of high precision predictions of root, extraradical hypha, vesicle and spore were left-skewed, indicating high certainty on the assigned class labels (**Fig. 2.2**). Arbuscules had lower confidence scores in comparison to sorghum roots (**Fig. 2.2**) ($P < 0.00003$ Tukey Multiple Comparison applied to an ANOVA of angular transformation of confidence scores). The low AP value and the high confidence score of extraradical hypha suggest that the main challenge for the Mask R-CNN is the pixel level segmentation of extraradical hypha rather than instance classification. Given the low AP value and confidence score of intraradical hypha, this fungal structure was dropped from latter analyses. Difficulty in arbuscule classification could be driven by the observation that arbuscules present in both isolation and clusters in sorghum roots. The observed frequencies of fungal structures from Mask R-CNN predictions did not differ from the frequencies counted by human scorers on the testing images (p-value = 0.786 with Fisher's exact test). As the segmentation model produced satisfying results, we chose the best one (**Table**

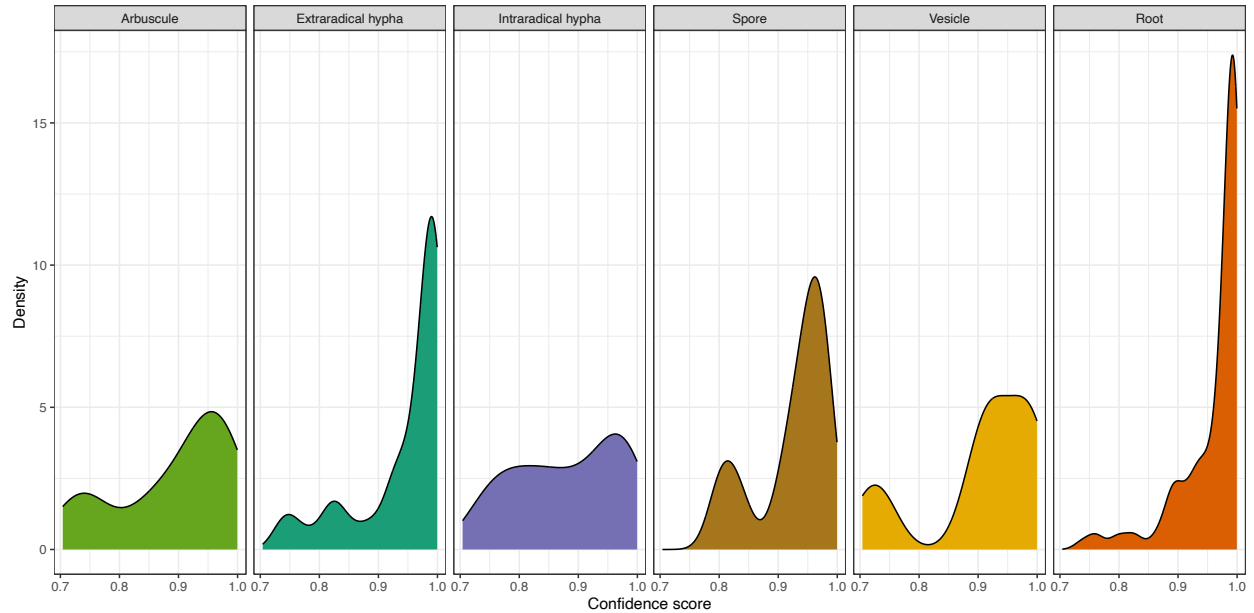


Figure 2.2 The best model trained using the in-house image dataset is used to do inference on the test set containing only in-house images of Georgia samples. Confidence scores are assigned to predicted instances during classification by the Mask R-CNN. A threshold of 0.7 was applied to select for high precision predictions only. Mask R-CNN has the highest confidence in labels assigned to predicted roots for showing a left skewed distribution.

2.1 Model A) for inference on a much larger in-house dataset (24,391 images) where images were collected in similar experimental settings.

The pretrained model was also trained and tested with a combined dataset including the AMFinder dataset² to illustrate how the model will be iteratively improved as more data are collected. The AMFinder dataset was made up of images collected from lab grown plants with a single AMF species inoculum and had patterns different from our in-house dataset collected from plants grown in soil from the field in Georgia, USA. With learning rates 0.001 and the default augmentation, we obtained a mAP of 29.6 and mAP50 of 50.2 in the combined test set (see Materials and Methods). This showed that our Mask R-CNN Model can be expanded and adapted to diverse conditions with different experimental settings as more representative training data are included.

We then tested whether adding the AMFinder images to the training dataset improved model performance on our in-house AMF colonized root images. The model performance was slightly decreased on the original test set. The best model trained on the combined image set had a mAP of 21.4 and mAP50 of 39.7. Whether adding new data with patterns different from the targeted conditions needs further testing with different training schedules and approaches.

2.2.2 From Image Segmentation to Measures of Fungal Morphology

The best performing computer vision model on in-house images presented above was applied to over 20,000 images of 108 root samples from the top, middle, bottom root regions of 12 sibling sorghum plants to generate pixel-wise segmentations of the five hyphal structures. From the segmentation results, the average quantity and size of fungal structures were computed for each root sample. Paired correlation analyses of the fungal structure morphological traits were examined first to identify whether fungal structures tend to co-occur in sorghum roots.

Positive associations dominated the frequencies of fungal structures. Higher occurrence of extraradical hyphae was consistently associated with higher occurrence of arbuscules. Vesicles and spores were positively correlated in sizes and counts. Larger number of vesicles and spores in a sample were suggestive of smaller arbuscules (Fig 2.3).

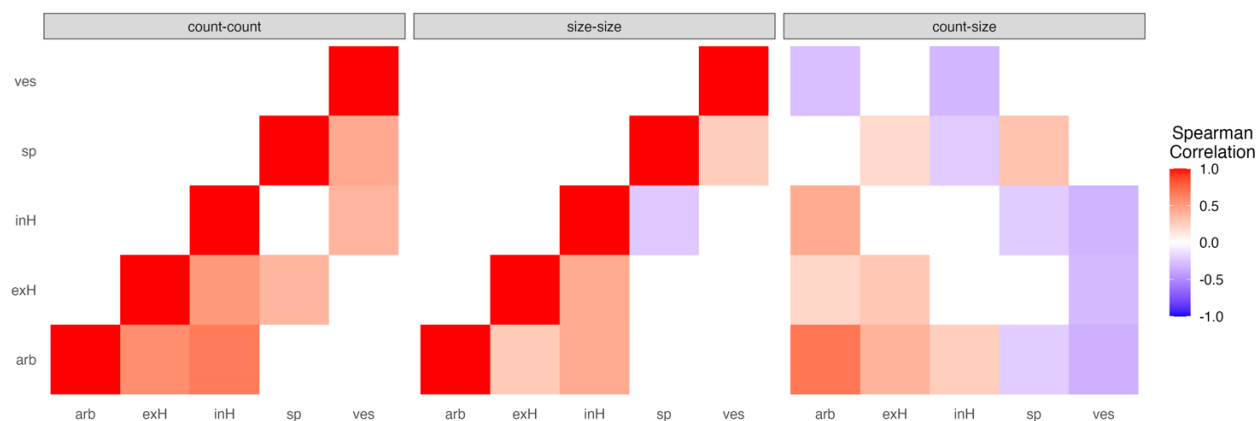


Figure 2.3 Significant correlations exist between structure counts and size. Shades of red indicate increasingly positive correlations. Shades of blue indicate decreasingly negative correlations. Fungal structures were abbreviated: arbuscule (arb), extraradical hypha (exH), intraradical hypha (inH), spores (sp), and vesicle (ves). For the correlations between fungal counts and sizes in the third panel, counts were arranged on the x-axis, sizes on the y-axis.

2.2.3 Using Mixed Linear Models to Predict Total AMF Abundance with Fungal Structure

Morphology

From the morphometric data of fungal structures, two measures of total AMF abundance were computed: the percentage root area occupied by AMF (percent colonization) and the density of AMF per root area (count density). Mixed linear model (MLM) analysis of the two phenotypes as response variables and the fungal structure morphology traits as predictors provides a means to test two hypotheses in the study of differential colonization by AMF. One major hypothesis is that AMF presents differential colonization between root sections. The second hypothesis is that allocation to AM fungal structures varies between plants. In this section, background on mixed linear modeling serves as an accessible introduction to how MLMs can be used to test these two hypotheses.

2.2.3.1 The Rationale for Mixed Effects

Mixed Linear Models (MLMs), also known as multilevel or hierarchical models, feature fixed and random effects (Searle et al., 2009) (**Box. 2.1**). Experimental treatments are typically modeled as fixed effects. Individual observations are grouped by random factors. Random factors, therefore, constitute the grouping level. Fixed factors are estimated as the mean effect for a particular factor level. In contrast, if the primary interest lies in estimating between-group variances, variables are modelled as random effects. The estimated values of random factors are shrunk towards the population mean.

The choice of using mixed effects to model AMF colonization is motivated by the experimental design. Our data are inherently hierarchical. AMF colonization was quantified in each of the three root regions within each sorghum plant, and three replicates were taken per root region (Fig. 2.4). The nested layers are plant, root depth, and replicate. The spatial scales between

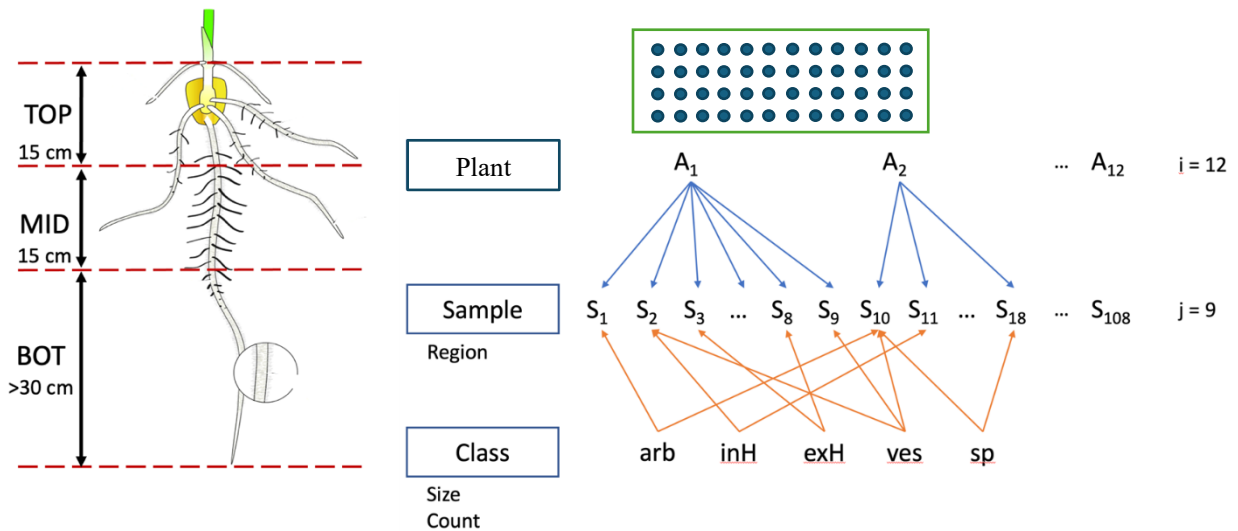


Figure 2.4 Experimental design entails hierarchical sampling from a RIL population. Twelve inbred lines were used in the experiment. One plant of each inbred line was sampled from the top (TOP), middle (MID), and bottom (BOT) of the root system. Each section was sampled three times. For each sample, 4 fields of view were imaged around each marker dot as shown on the schematic slide below, generating 192 root intersections per slide. Five fungal structures: arbuscule (arb), internal hypha (inH), external hypha (exH), vesicles (ves), and spore (sp) were segmented from each image. Root region is a variable recorded for each sample. Average size and count are computed for each class of instances found in a sample.

root samples and sorghum plants are biologically nested (Schielzeth & Nakagawa, 2013). Between-sample variance needs to be evaluated as a random factor. It would be wrong to treat root samples from the same sorghum plant as independent. The twelve sorghum plants are siblings randomly sampled from a RIL population. Between-plant variance is treated as a random factor as well. Root region and AMF structure level phenotypes are the fixed effects. The same model structure can be used to model the AMF count density.

2.2.3.2 Mixed Linear Model Assumptions

In MLM, both root samples and sibling plants are assumed to be randomly sampled from a population of samples and a population of sorghum recombinant inbred lines, respectively (Govindarajulu et al., 2021). It is assumed that the sampling errors (ϵ_{ij}) and the plant level random effects (u_{0i}) are independent and that the random effects and sampling errors have a normal distribution of zero mean and distinct constant variances (**Box. 2.1**). Another assumption is that the mean and variance of the response variable are not functionally dependent on each other across siblings. Transformation of the response variable is an effective method to remove the dependency. Choice of models are designed to address each of the three questions of the paper laid out at the end of the introduction.

2.2.3.3 The Null Model (Model 1)

The Null MLM Model does not include explanatory variables but only the mean overall AMF colonization and the plant level and sample level differences in colonization. The colonization by AMF of a root sample from a plant (y_{ij}) is equal to the mean colonization in the sorghum population (β_0) plus the plant level random difference from the population mean (u_{0i}) plus the sample level differences (ϵ_{ij}) (**Box. 2.1**). This simple Null Model shows that MLM partitions

Box 2.1. Mixed linear models (Searle et al., 2009) discover relations between AMF colonization and count density on the one hand and accession and fungal morphology on the other hand.

General Expression of MLMs

Response Intercept + Fixed Random Sampling

$$y = X\beta + ZU + \epsilon$$

Random Intercept Null Model (Model 1)

Percent Intercept Sorghum Sampling

$$\begin{aligned} y_{ij} &= \beta_0 + u_{0i} + \epsilon_{ij} \\ u_{0i} &\sim N(0, \sigma_u^2) \\ \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

Random Intercept Model with Experimental Design Variable (Model 2)

Percent Intercept Sorghum Region Sampling

$$\begin{aligned} y_{ij} &= \beta_0 + u_{0i} + \beta_1 x_{1ij} + \epsilon_{ij} \\ u_{0i} &\sim N(0, \sigma_u^2) \\ \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

Random Intercept Model with Structure Level Predictors (Model 5)

Percent Intercept Sorghum Region Arbuscule Arbuscule Sampling

$$\begin{aligned} y_{ij} &= \beta_0 + u_{0i} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \epsilon_{ij} \\ u_{0i} &\sim N(0, \sigma_u^2) \\ \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

Random Intercept and Slope Model with Structure Level Predictors (Model 6)

Percent Random intercept: Region Arbuscule Random slope: Sampling

$$\begin{aligned} y_{ij} &= \beta_0 + u_{0i} + \beta_{1ij} x_{1ij} + \beta_{2ij} x_{2ij} + (\beta_{3ij} + u_{1i}) x_{3ij} + \epsilon_{ij} \\ u_{0i} &\sim N(0, \sigma_{u_0}^2) \\ u_{1i} &\sim N(0, \sigma_{u_1}^2) \\ cov(u_{0i}, u_{1i}) &= \sigma_{u_0 u_1} \\ \epsilon_{ij} &\sim N(0, \sigma_\epsilon^2) \end{aligned}$$

the total variance in AMF colonization $\text{var}(y_{ij})$ into a variance between plants $\text{var}(u_{0i})$ and a variance between samples $\text{var}(\epsilon_{ij})$ as shown from **Box. 2.1 (Model 1)** (Juan, Basile, et al., 2005).

$$\text{var}(y_{ij}) = \text{var}(u_{0i}) + \text{var}(\epsilon_{ij}) = \sigma_u^2 + \sigma_\epsilon^2$$

The between plant variance $\text{var}(u_{0i})$ was estimated to be 0.0099 and two times that of the between sample variance $\text{var}(\epsilon_{ij})$ in **Table 2.2**. The proportion of the total percent colonization differences can be quantified at the plant level by computing the intraclass correlation (ICC) (Juan, Basile, et al., 2005).

$$ICC = \frac{\text{var}(u_{0i})}{\text{var}(u_{0i}) + \text{var}(\epsilon_{ij})} = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_\epsilon^2}$$

In **Table 2.2**, the ICC of Null Model was 0.647, which implies that 64.7% of differences in total percent colonization of AMF is at the plant level and could be controlled by the plant genome (Deng et al., 2021). An alternative hypothesis would be that this clustering of variance at the plant level might be attributable to the different composition of AMF structures, and this composition could be defined by the plant genome. We will come back to testing the alternative hypothesis in Models 3, 4, and 5.

2.2.3.4 Random Intercept Model with Experimental Design Variable (Model 2)

In Model 2, the Null Model (Model 1) is expanded by including the design variable, the root region where the root sample was obtained (**Fig. 2.4**), with fixed effect β_1 . Root region is a discrete predictor with three levels. The goal of the model is to investigate if AMF percent colonization differs between root regions and to determine the extent to which variance at plant level may change after taking into account differences in colonization in root regions. Proportional

change in variance (PCV) at different levels can be computed to evaluate the change using the following equation (Juan, Min, et al., 2005; Nakagawa & Schielzeth, 2013),

$$PCV_{plant} = \frac{var(u_{0i}) - var'(u_{0i})}{var(u_{0i})}$$

$$PCV_{sample} = \frac{var(\epsilon_{ij}) - var'(\epsilon_{ij})}{var(\epsilon_{ij})}$$

where $var(u_{0i})$ is the between plant variance in the Null Model and $var'(u_{0i})$ is the between plant variance in the new model. Comparing Model 2 to the Null Model 1, the PCV_{plant} was equal to -0.0221, and PCV_{sample} was equal to 0.3648 in Table 2. We concluded that 36.48% of sample variance within plants in the null model is attributed to differences in root regions. By adjusting for the root regions where the sample was obtained, another 2.21% of the variance in percent colonization by AMF was accounted for by plant differences in **Table 2.2**.

2.2.3.5 Random Intercept + Fixed Slope Model with AMF Structure Predictors (Model 3, 4, 5)

As mentioned earlier under the Null Model, an alternative hypothesis to the clustering of total percent colonization of AMF at the plant level is that the plant level differences in AMF abundance can be attributable to the different composition of AMF structures in the plants, which could be controlled by the plant genome (Deng et al., 2021; Merlo et al., 2005). To test the possibility, the three Models (3, 4, and 5) expand Model 2 by including arbuscule count and/or size as fixed effects determined to be significant by Lasso Regression (Groll & Tutz, 2014). By comparing Models 3, 4, and 5 to the Null Model, the changes in plant level variance can be measured using ICCs and PCVs after adding different AMF structure predictors to the model (**Table 2.2**). The interpretation of ICCs and PCVs leads to a major conclusion regarding the total AMF percent colonization and the composition of fungal structures, which is discussed in depth in a later section.

Table 2.2 Mixed linear models are well predicted for percent colonization by arbuscule count, arbuscule size, region, and plant. Proportional change in variance (PCV) is provided to measure the importance of random effects. Intraclass correlation (ICC) is used to implicate the variance between plants. Varied measures of fit are provided to assess model performance.

	Random				Intercept		Random Intercept + slope
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	
	Null Model	Model with one fixed effect	Model A with two fixed effects	Model B with two fixed effects	Model A with three fixed effects	Model B with three fixed effects	
Fixed							Effects
Intercept	0.3339	0.3360	0.3329	0.3307	0.3293	0.3410	
regionTOP	NA	0.0492	0.0354	0.0312	0.0241	0.0226	
regionBOT	NA	-0.0552	-0.0322	-.0213	-0.0100	-0.0087	
arb_count_scaled	NA	NA	NA	0.0885	0.0757	0.0937	
arb_size_scaled	NA	NA	0.0490	NA	0.0346	0.0304	
Variance of fixed effects	NA	0.0018	0.0040	0.0092	0.0109	0.0139	
Random							Effects
Variance between plants	0.0099	0.0101	0.0056	0.0015	0.0010	0.0009	
Variance between root samples	0.0054	0.0034	0.0024	0.0015	0.0009	0.0007	
Proportional	Change	in		Variance		(PCV)	
Between plants	NA	-0.0221	0.4355	0.8447	0.9003	0.9091	
Between root samples	NA	0.3648	0.5640	0.7298	0.8344	0.8713	
Intraclass	Correlation	(ICC)	or	Variance	Partition	Coefficient	(VPC)
Plants	0.6472	0.7470	0.7037	0.5133	0.5248	0.5645	
Model							Performance
Marginal	R ²	NA	0.1190	0.3345	0.7539	0.8525	0.8532
Conditional	R ²	0.6472	0.7771	0.8028	0.8802	0.9299	0.9574
AIC	-217.9263	-259.5121	-301.6817	-362.4751	-413.8518	-421.9404	
BIC	-209.8799	-246.1014	-285.5889	-346.3823	-395.0769	-397.8012	
Deviance	-223.9263	-269.5121	-313.6817	-374.4751	-427.8518	-439.9404	

The effect of average size of arbuscule (β_2) and the effect of arbuscule count (β_3) per sample are continuous variables that describe the association of the AMF structures with the total degree of root colonization by AMF (y_{ij})s. A positive estimate of β_3 , for example, indicates a positive linear relation between AMF colonization and average arbuscule size. A larger estimate of β_3 than β_4 means arbuscule size has a stronger effect on total AMF colonization per unit increase than arbuscule count. In all three models, the relations between fixed effects and overall AMF

colonization are considered to be the same in all sorghum plants. In other words, the slopes are fixed with respect to plant.

2.2.3.6 Random Intercept + Random Slope Model with AMF structure predictors (Model 6)

In Model 6, the effect of arbuscule count (β_3) on overall AMF colonization may differ between sorghum plants. For examples, in some plants with high AMF colonization in the roots, arbuscules may be the dominant hyphal structure but not in other plants. In Model 6, the regression coefficient of AMF colonization on arbuscule count varies at the plant level to capture this differential effect. By comparing Model 6 to Model 5, it is possible to determine whether the assumption of varying magnitude of association of arbuscule count and percent colonization between plants holds.

The total variance in AMF colonization $\text{var}(y_{ij})$ is still made up of two parts, a variance between plants $\text{var}(u_{0i}, u_{1i}x_{1ij})$ and a variance between samples $\text{var}(\epsilon_{ij})$. The variance between plants $\text{var}(u_{0i}, u_{1i}x_{1ij})$, however, partitions into a slope variance $\text{var}(u_{1i}x_{1ij})$, intercept variance $\text{var}(u_{0i})$ and their covariance $\text{cov}(u_{0i}, u_{1i}x_{1ij})$ (Subramanian et al., 2003). This variance is a quadratic function in arbuscule count:

$$\begin{aligned}\text{var}(y_{ij}) &= \text{var}(u_{0i}, u_{1i}x_{1ij}) + \text{var}(\epsilon_{ij}) \\ &= \text{var}(u_{0i}) + \text{var}(u_{1i}x_{1ij}) + 2\text{cov}(u_{0i}, u_{1i}x_{1ij}) + \text{var}(\epsilon_{ij}) \\ &= \sigma_{u_0}^2 + \sigma_{u_1}^2 x_{1ij}^2 + 2\sigma_{u_0 u_1} x_{1ij} + \sigma_{\epsilon}^2\end{aligned}$$

When there are random slopes in the model, the *Variance Partition Coefficient* (VPC), a function of arbuscule count, is calculated to measure the relationship of plant level variance to the total variance rather than an ICC (Subramanian et al., 2003):

$$VPC = \frac{\text{var}(u_{0i}, u_{1i}x_{1ij})}{\text{var}(u_{0i}, u_{1i}x_{1ij}) + \text{var}(\epsilon_{ij})} = \frac{\sigma_{u_0}^2 + \sigma_{u_1}^2 x_{1ij}^2 + 2\sigma_{u_0 u_1} x_{1ij}}{\sigma_{u_0}^2 + \sigma_{u_1}^2 x_{1ij}^2 + 2\sigma_{u_0 u_1} x_{1ij} + \sigma_{\epsilon}^2}$$

VPC is similar to ICC in terms of interpretation of the result.

2.2.4 Arbuscule Morphology is Predictive of Differential AMF Percent Root Colonization in Sibling Plants

Looking at how the plant level variance changes as predictors such as root region, arbuscule size, and count were added to the Null Model in Model 2, 3, 4, and 5, we observed that, in Model 6, 90.9% total variance in percent colonization is attributed to the plants ($PCV_{\text{plant}} = 0.9091$). An VPC of 56.6% suggested that even if variance at plant level shrunk, it still explained the majority of differences in percent colonization. This is possible because the sample level variance dropped with the plant level variance by 87.1% as arbuscule count and size were added as predictors.

Model 6 had the highest R-squared (Johnson, 2014; Nakagawa & Schielzeth, 2013) of 0.9574 and desired lowest information criteria and deviance. A model with random slopes for both arbuscule counts and size were fitted but not included in Table 2 as it was not significantly different

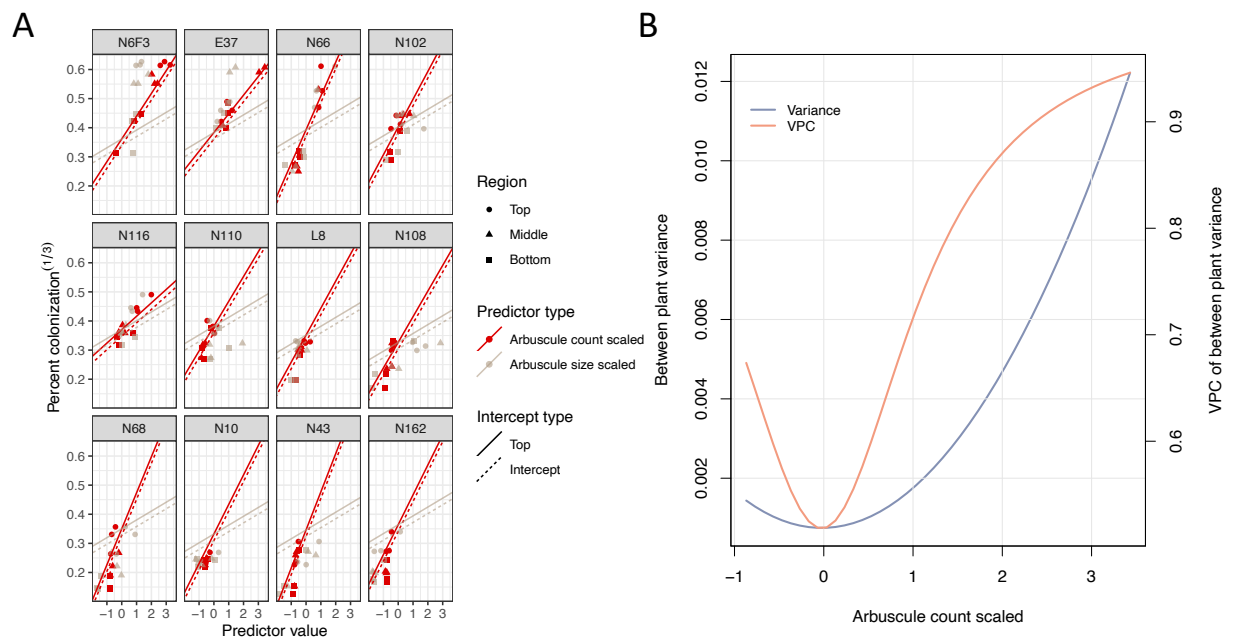


Figure 2.5 Percent colonization is well predicted by Model 6. (A) Random slopes (red) are fitted to scaled arbuscule count for each plant. Fixed slopes for arbuscule size (grey) are shown as reference for easy visualization of the changing slopes of arbuscule count. The solid line represents the intercept of the top root region, which predicts for higher AMF percent colonization than sample mean. (B) Variance between plants and its VPC are plotted as a function arbuscule count. Scaled arbuscule count is plot on the x-axis, with Variance between plants as the main y-axis, its VPC on the secondary y-axis.

from Model 6 in a likelihood ratio test. Confidence scores for arbuscule was added as a fixed effect to Model 6 and was tested nonsignificant using a likelihood ratio test. In **Fig. 2.5A**, the expected values of AMF colonization from Model 6 were plotted to visualize the fit of the model to the data and the varying slopes between plants. The plant level variance and its VPC is a function of arbuscule count in Model 6 (**Fig. 2.5B**). The clustering of plant level variance was stronger when more arbuscules were observed, also reflected by the increasing of VPC (**Fig. 2.5B**). For example, N6F3, E37, N66, N102, and N116 had higher slopes and hence higher arbuscule counts than the

Table 2.3 Mixed linear models are well predicted for count density by counts of hyphal structures, root regions, and sorghum accessions. Proportional change in variance (PCV) is provided to measure the importance of random effects. Intraclass correlation (ICC) is used to implicate the variance between accessions. Varied measures of fit are provided to assess model performance.

	Random intercept			Random intercept and slope
	Model 1	Model 2	Model 3	Model 4
	Null model	Model with one fixed effect	Model A with five fixed effects	Model B with five fixed effects
Fixed Effects x 10⁹				
Intercept	809770.357	793371.277	786651.710	790215.150
regionTOP	NA	114053.586	76643.165	86253.886
regionBOT	NA	-64856.345	-7287.224	-8284.854
arb count scaled	NA	NA	66137.475	60660.157
exH count scaled	NA	NA	60918.089	63380.104
sp count scaled	NA	NA	47678.393	39526.530
ves count scaled	NA	NA	97338.416	111460.129
Variance of fixed effects	NA	5.520	33.276	34.734
Random Effects x 10⁹				
Variance between Plants	20.689	21.322	7.049	6.012
Variance between root samples	27.674	21.979	7.642	6.264
Proportional Change in Variance (PCV)				
Between plants	NA	-0.031	0.659	0.709
Between root samples	NA	0.206	0.724	0.774
Intraclass Correlation (ICC) or Variance Partition Coefficient (VPC)				
Plant	0.428	0.492	0.480	0.490
Model Performance				
Marginal R ²	NA	0.113	0.694	0.692
Conditional R ²	0.428	0.550	0.841	0.875
AIC	-1543.514	-1563.654	-1674.505	-1678.541
BIC	-1535.467	-1550.243	-1650.366	-1649.037
Deviance	-1549.514	-1573.654	-1692.505	-1700.541

remaining accessions. The remaining plants, however, had lower arbuscule counts and similar total AMF percent colonization. It was impossible to distinguish the remaining plants by their arbuscule counts. At first sight it seemed strange that arbuscule count should be selected as a predictor when it had lower confidence scores than other traits (**Fig. 2.2**), but as shown in **Fig 2.5B**, there were substantial differences in arbuscule count for the first 5 accessions, and that was why arbuscule count was selected for inclusion in the mixed linear model to explain plant level differences.

2.2.5 The Plant Level Variance Component in Count is Stable across Models – Loss of Goodness of Fit in Random Effects is Offset by A Gain in Goodness of Fit to Fixed Effects of Fungal Structures

The same variable selection and model selection procedure for overall AMF percent colonization was applied to fit overall AMF count density as the response variable. The β vector of fixed effects was modified to include the appropriate AMF structure phenotypes as predictors. Lasso regression (Groll & Tutz, 2014) selected the count measures of all fungal structures as fixed effects in the mixed linear model to predict overall AMF count density in sorghum plants. Forward selection removed the number of intraradical hyphae as an explanatory variable. Root regions and the counts of arbuscules, vesicles, spores and extraradical hyphae were the five fixed effects in Model 3 and 4 for count density. Random slopes were added with respect to the four effects of fungal structures. The effect of extraradical hyphae was the only one that differed between the sorghum plants. A random slope was added to the effect of extraradical hyphae in Model 4. We added the confidence score of extraradical hypha as a fixed effect to count density Model 4 and found that it was not a significant variable.

The proportion of variance at the plant level remained stable as fixed effects were added to models (**Table 2.2**). In **Table 2.3**, the ICCs and VPC of between plant variance ranged from 0.428

Table 2.4 Tukey multiple comparison test with contrasts between region regions and their significance (p-values). The top, middle, and bottom root regions were abbreviated as TOP, MID and BOT.

Structure Count										
Contrast	<i>arb</i>		<i>exH</i>		<i>inH</i>		<i>sp</i>		<i>ves</i>	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
MID-TOP	-0.203	0.2538	-0.527	0.0015	-0.109	0.7834	-0.211	0.5379	0.192	0.4038
MID-BOT	0.383	0.0096	0.385	0.0274	-0.193	0.4663	0.205	0.5580	-0.018	0.9922
TOP-BOT	0.586	<0.0001	0.912	<0.0001	-0.084	0.8636	0.416	0.0957	-0.209	0.3397
Structure size										
Contrast	<i>arb</i>		<i>exH</i>		<i>inH</i>		<i>sp</i>		<i>ves</i>	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
MID-TOP	-0.281	0.2341	0.162	0.6823	0.229	0.4966	-0.144	0.7754	-0.170	0.7275
MID-BOT	0.469	0.0201	0.159	0.6942	-0.075	0.9274	0.308	0.3275	-0.243	0.5247
TOP-BOT	0.750	<0.0001	-0.004	0.9998	-0.304	0.2944	0.447	0.0922	-0.073	0.9434

to 0.490, which was less than a 7% difference in variance explained. The proportional changes in plant variance decreased by 70.9%, which was compensated for by a 77.4% drop in sampling variance. What variation in count density lost to the fixed effects was replaced by the improved fit of the model.

2.2.6 Differential AMF Colonization between Sorghum Root Regions

A significant improvement of Model 2 to the Null Model 1 for both total AMF colonization phenotypes supports that AMF colonization was different between root regions. PCV_{sample} was 0.365 and 0.205 respectively for AMF percent colonization and count density (Table 2.2 and 2.3). The positive signs of PCVs suggest that the sample variances within plants in the null models are attributed to differences in root regions. The top root region had the highest colonization by t-tests at the 0.05 significance level (Fig. 2.6 A and B). Arbuscule count was a predictor essential for the modeling of both phenotypes in the previous section. It is a reasonable speculation that arbuscule count is a main driving force in the positive correlation of the two total AMF colonization traits. The speculation is sustained by larger arbuscule size (Fig. 2.6C), higher of arbuscules and

extraradical hyphae in the top root region (**Fig. 2.6D**), tested significant using Tukey Multiple Comparison tests (**Table 2.4**).

Table 2.5 Multiple comparison test of AMF colonization along root regions.

Contrast	Percent colonization		Count density		Proportion nutrient exchange	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
MID-BOT	0.009	0.410	0.000032	0.345	0.408	0.011
MID-TOP	-0.023	0.005	-0.000077	0.004	-0.311	0.070
TOP-BOT	0.031	<0.0001	0.000109	<0.0001	0.719	<0.0001

Although the morphological traits of other AM fungal structures did not have significant effects, they could still contribute to differential colonization. If the sorghum plants were colonized by equal amounts of AMF, **Fig. 2.7 A and B** showed how the relative abundance of AM fungal structures in the roots could differ. When the twelve sorghum plants were ranked in decreasing order of AMF percent colonization from left to right, the relative amount of arbuscules and extraradical hyphae trended downward. The same observation held if the panel was divided by root regions.

To quantify the observation, one more phenotype was calculated, the amount of arbuscule and extraradical hyphae divided by the total AMF structures. It measures the proportion of nutrient exchange (PNE) structures (Johnson et al., 2003). After logit transformation, a mixed linear model was fitted to PNE with root region as the fixed effect and a plant level random effect. The intraclass correlation for PNE was 0.585. The conditional R-squared of the mixed linear model was 0.624. Tukey multiple comparison test showed that the top 0-15cm and the middle 15-30cm tested insignificant to each other, but both were tested significant to the bottom >30cm roots (**Table 2.5**). Percent colonization and count density had the middle 15-30cm root region tested insignificant against the bottom root regions. The levels of AMF colonization of the twelve sorghum plants were ranked differently with percent colonization, count density and PNE. Some similarity was found between percent colonization and PNE using Spearman correlation ($\rho=0.544$, $p<0.001$) There

was no correlation between proportion NE and AMF count density ($\rho=-0.089$, $p = 0.362$) (Fig. 2.8).

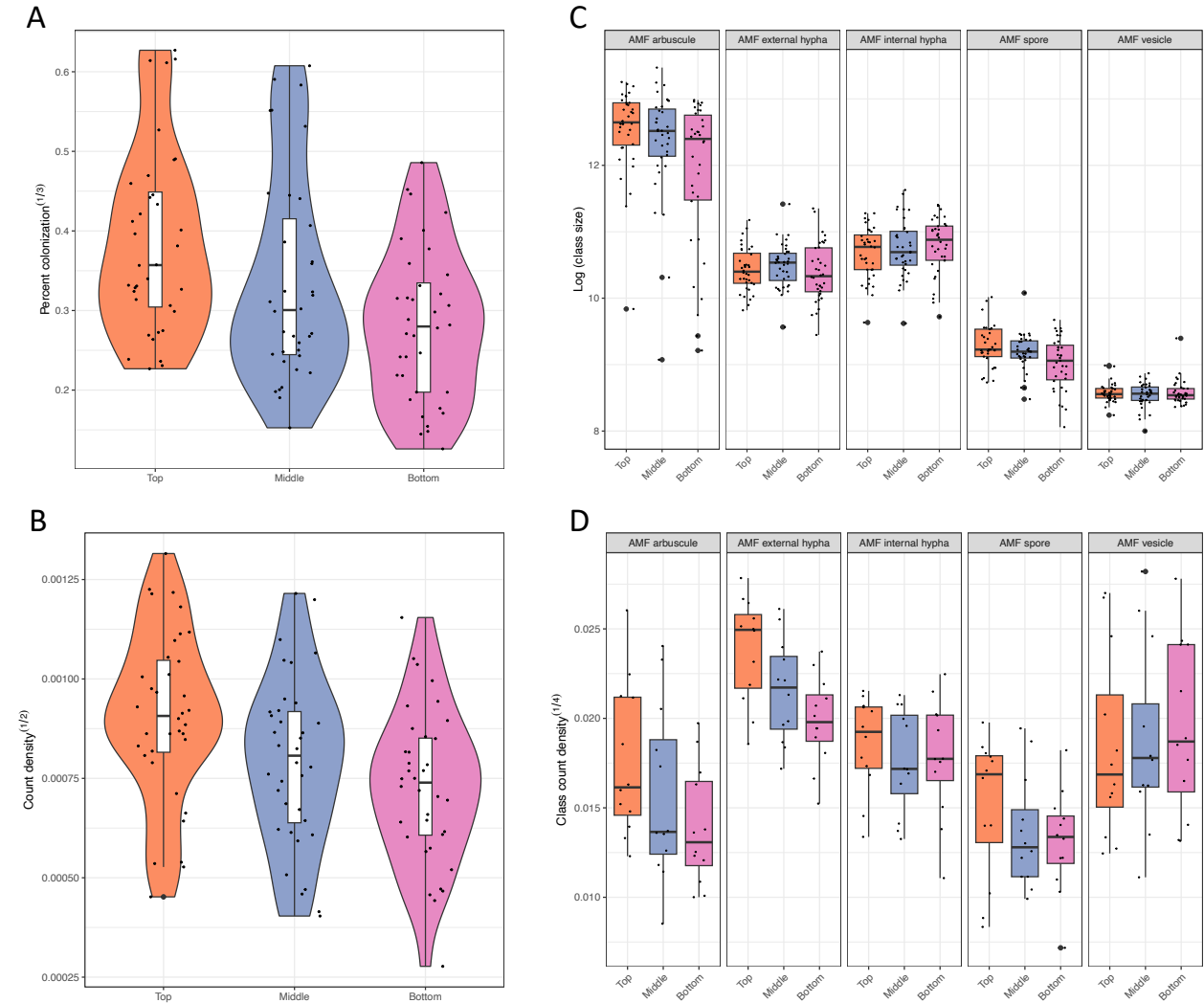


Figure 2.6 AMF colonization and structure abundance vary with root depth. A and B show the distribution of total AMF percent colonization and count density by the top, middle, bottom root regions. Colonization is the highest in the top root region. To examine the reason for high AMF colonization in the top root region, Box and Whisker plots are used to display the size and count density of each AM fungal structure by root regions in C and D.

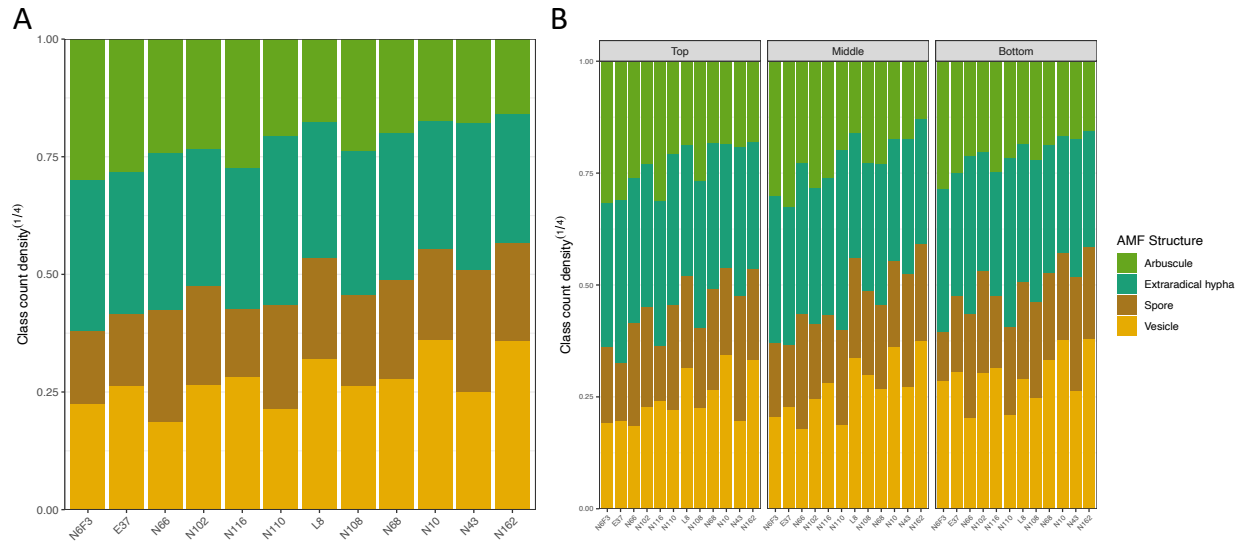


Figure 2.8 The stack bar plots reflect the relative abundance of AM fungal structures when the total AMF count density is assumed to be the same across sibling plants. Sibling plants are ordered in decreasing total AMF percent colonization. (A) The relative abundance of arbuscule and extraradical hypha decreases as the plant has more quantities of vesicle and spore. (B) The same relationship is observed in the top, middle and bottom root regions of sibling plants.

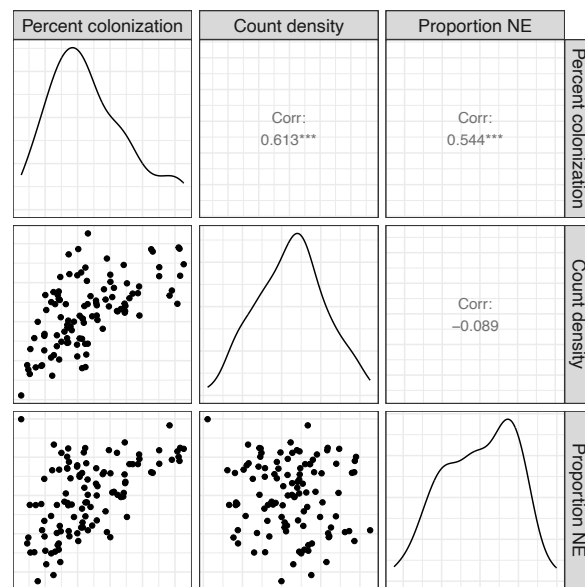


Figure 2.7 The density plots of AMF percent colonization, count density and proportion of nutrient exchange (Proportion NE) are arranged on the diagonal. The phenotypes were transformed to be normally distributed. Scatter plots of each pair of phenotypes are visualized in the lower panel. Pearson correlation values and significance are displayed in the upper panel. For example, percent colonization and proportion NE has a 0.544 positive correlation.

2.3 Discussion

2.3.1 Preferential AMF Colonization in Sorghum Roots

The 15cm of roots closest to the soil surface presented highest total AMF colonization and proportion of nutrient exchange fungal structures (**Fig. 2.4 and 2.6, Table 2.4 and 2.5**). The total AMF richness and colonization at 0-10 cm soil depth were shown to be higher than at deeper soil layers in *Pedicularis kansuensis* at a subalpine grassland ecosystem (Wang et al., 2018). Kabir et al found that total colonization, total hyphae density, and spore density were highest at a depth of 0-15 cm of soil in corn roots (Kabir et al., 1998). Top roots of sorghum produce secondary roots and fine root via branching. The growing tip is at the bottom of the root. One hypothesis is that the AMF preferentially colonize the actively branching sections of sorghum roots. An alternative hypothesis is that the top layer is rich in phosphorus, defining the niche for AMF colonization (Wang et al., 2018). One way to test the first hypothesis is to examine the variation in root morphology between inbred lines with the software DIRT (Das et al., 2015) to see whether or not root morphology has an impact on fungal structure counts. The latter hypothesis could also be tested by using nanodots or phosphorus labeling experiments to track the exchange in the top layer (Whiteside et al., 2019).

2.3.2 Plant Level Variance of AMF Colonization

In the null model, the total variance in AMF colonization was composed of the sampling variance and plant level variance. Expanding the null models transferred the plant level variance from the variance components to the fungal structure morphology traits. One hypothesis for the result is that sorghum plants of different genotypic backgrounds determine overall AMF abundance in roots by manipulating the developmental stage of AMF species after colonization. Alternatively, in lifestyle studies of single AMF species, researchers found that the composition and morphology

of their fungal structures vary. Sorghum lines may match with a community of AMF species in the soil that generates different compositions of AM fungal structures in roots (Martignoni et al., 2021).

Our experimental design does not permit direct testing of the amount of variance in AMF colonization that is under the genomic control of plant hosts, because there were no biological replicates of the RILs. If biological replicates were available, plant level variance would provide an estimate of population level variance of AMF colonization and the measurements, ICCs and VPCs, would capture the broad sense heritability including additive, dominant and epistatic effects. The results in **Table 2.2** are suggestive of performing a large scale Genome Wide Association Study (GWAS) and or Recombinant Inbred Lines (RILs) study in *Sorghum bicolor* to test whether or not sorghum genes play a role in AMF colonization (De Vita et al., 2018b; Plouznikoff et al., 2019).

2.3.3 Differential AMF Colonization between Sorghum Inbred Lines

Percent colonization was a new measurement of overall AMF abundance by taking the ratio of pixelwise segmentation of fungal structures and sorghum roots of our computer vision model. Count density was a similar measurement of AMF abundance to the output of McGonigle (McGonigle et al., 1990b) scoring method. The proportion of variance at plant level of these two phenotypes did not lose to the integration of fungal structure morphology traits into the models, which establishes them as favorable response variables for Genome-wide Association Studies of AMF colonization in sorghum populations. While both phenotypes were modeled with fungal structure morphology traits as fixed effects, the fixed effects differed in their ability to explain variability in the data. Two morphological traits of arbuscule were required to explain 95.7% variability in percent colonization suggestive of a mechanism for how the plant controls AMF

colonization. One example of a gene relevant to this mechanism encodes the protein DELLA (Floss et al., 2013; Yu et al., 2014).

Count density was modeled with four fungal structures. The final fit to predicting count density explained 87.2% of variability. One could argue that percent colonization is a better phenotype for GWAS for its simplicity, goodness of fit and higher plant level variance. The plant level variance of the phenotype would decrease as the number of plant replicates and sorghum inbred lines grow in scale. Count density should not be discarded as technologies provide newer features to quantify AMF abundance. Finally, the two phenotypes did rank the twelve sorghum inbred lines in different orders. Inbred line L8 had the highest AMF count density but ranked 7th for percent colonization. Ranks for most other inbred lines were comparable. It is recommended to use both phenotypes in future studies of AMF colonization using computer vision.

2.3.4 Role of Models across Scales to Understand the AMF Symbiosis with Plants

To understand the symbiosis of *Sorghum bicolor* with AMF it has been necessary here to develop models that operate at multiple scales (Johnson et al., 2006). First, computer vision models were developed that provided the high throughput data to describe how AMF colonize the plant at the individual plant level. A variety of measures were developed that then can be used in GWAS studies to test the role of the plant genome in shaping the AMF microbiome. To avoid the confounding effects of environmental field variables in a GWAS, it will important to use factorial designs (Fisher, 1935) and blocking to separate the effects of accession from environmental field variables, like Nitrogen and Phosphorus levels. These models operated at the individual level of plants in the study. The computer vision models provided a diverse array of measures to describe the colonization process described above (**Tables 2.2 and 2.3**). We are in the process of not only automating classification and segmentation of fungal structures but full automation of image

acquisition as well. A final limitation of the computer vision phenotyping method here is not having live cell images to follow the dynamics of the structures. The Kokkoris laboratory (Cornell et al., 2022) has been able to follow the dynamics of nuclei in AMF, and it would be useful to adopt a similar approach to other AMF structures to gain a time dimension on colonization. Live cell imaging will also contribute to understanding the prepenetration apparatus (Genre et al., 2008), arbuscule formation (Floss et al., 2013), and AMF inheritance (Marleau et al., 2011).

On top of the computer vision model for feature extraction, mixed linear models were added to discover relations between measures of overall AMF abundance with fungal structure morphology and root niche. This mixed linear modeling approach provides a framework for GWAS and QTL mapping by suggesting underlying mechanisms by which the plant genome controls the AMF community. For example, arbuscule count entered into the prediction. This relation suggests a mechanistic link to the arbuscules in how the plant genome controls the AMF microbiome. There are likely genes in both sides of the partnership controlling the development of these structures (Ivanov et al., 2019). Other layers to the modeling will need to be added to conceptualize our understanding of this ancient and fundamental symbiosis (Johnson et al., 2006).

2.4 Materials and Methods

2.4.1 The Georgia Dataset

Plant Cultivation:

Sorghum plants were derived from Recombinant Inbred Lines (RILs), a mapping population consisting of 191 F3:5 from a cross between an unnamed accession of *Sorghum propinquum* (William Rooney, Texas A & M University, College Station, TX) and inbred line TX7000 of *S. bicolor* (Govindarajulu et al., 2021). The seeds were kindly provided by Jeff Bennetzen, one of the PIs who constructed the RIL collection (Govindarajulu et al., 2021). No permissions are needed to use these Sorghum accessions. Three seeds from 15 RILs were planted on October 5, 2020, in steam sterilized Sungro garden soil in 2.5-gallon pots at the UGA Botany Greenhouse. Seedlings were grown on a 11-hour light cycle. Plants were fertilized with 1 tablespoon Osmocote. Individual seedling was transferred in 2.5-gallon pots filled with a 4:1 mix of steam sterilized turf and soil from Ironhorse Farm, Watkinsville, GA (**Table 2.1**) on day 15. Seedlings were grown to maturity on a 11-hour Light/Dark cycle with watering as needed. In addition, one commercial hybrid forage sorghum plant derived from Richardson, TX was harvested at Iron Horse Farm, GA on October 13, 2020, two grain sorghums of accession M72GB7 at Iron Horse Farm, GA on November 12, 2020, a Colby sorghum at UGA Botany Greenhouse. All the methods were carried out in accordance with relevant Institutional guidelines and regulations.

Root imaging:

Random samples of 0.25g of fine roots were taken from the whole roots of Richardson, M72GB7, Colby, E46-W, N88, E24, E46 for training images. The cleaned whole roots were cut into 1 cm pieces. Fine roots with intact cortex were randomly selected and weighed to get 3

cassettes of 0.25g of samples per plant. Root samples were cleared in 10% alkaline hydrogen peroxide solution for 2 hours and in 5% KOH overnight at room temperature. Fungal structures were stained using a modified Ink and Vinegar method (Vierheilig & Piché, 1998). Stained roots were spread and flattened on slides prior for imaging. Mounted root samples were imaged at 200X magnification with a Zeiss Primo Star compound microscope equipped with an Axiocam 105 color camera. Focusing was done locally and manually for every field of view during imaging to increase sharpness, but no post image acquisition processing was involved, such as adjusting contrast. McGonigle method was used to generate images at 192 root intersections. The root intersections were 0.5 cm equidistantly spread across a 75 x 25mm glass slide. The fungal structures at root intersections were manually scored and annotated for training the computer vision model.

To test for the differential AMF colonization in root regions and in sorghum plants, 12 RIL sorghums of E37, L8, N6-F3, N10, N43, N66, N68, N102, N108, N110, N116, and N162 were sampled from three root regions. The ‘TOP’ region was the first 15 cm of roots below soil surface. The ‘MID’ region is the next 15 cm below. The ‘BOT’ region was roots longer than 30 cm (**Fig. 2.4**). Aerial roots were excluded from sampling. From each region of a plant, 3 technical replicates of 0.25g of fine roots with intact cortex were randomly sampled. Each plant was represented with a total of 9 cassettes or 2.25g of root samples. A line has only 1 plant as biological replicate. The same clearing, staining and imaging procedures were applied.

2.4.2 The Cambridge Dataset

The publicly available Cambridge dataset (zenodo ID 10.5281/zenodo.5118948) included 15 whole slide scanning images acquired with a VHX-5000 digital microscope (Keyence, Milton Keynes, UK) set to $\times 200$ magnification. The images were downloaded from the zenodo data portal using the zenodo-get software method. The 15 whole slide images were in jpg format and 10389

× 5108 pixels and 96 pixels/inch in size. The original annotations were discarded. The same annotators and annotation standards for the Georgia dataset were used in reannotation to maintain uniformity. The annotated images were tiled and added to the Georgia dataset to create a secondary training set for more and even representation of each AMF class.

2.4.3 Image Annotation

The root image annotation was conducted using the VGG annotator tool (Dutta & Zisserman, 2019). The fungal and root structures were manually annotated using the polygon tool. One of seven class labels was assigned to a structure (**Table 2.1**). The annotation results were exported as a json file and csv table. All 746 jpg images in the Georgia dataset were segmented and annotated. We generated 3577 polygon annotations. A total of 14 out of 15 images were selected from the Cambridge dataset and produced 20588 annotations. The annotation criteria can be found in **Table 2.6**.

Table 2.6 Annotation rules for masking AM fungal structures.

Class	Annotation Rules for the Georgia Dataset
root	Plant root with intact cortex
extradical hypha (exH)	Filamentous structure outside the boundary of a plant root annotation
intraradical hypha (inH)	Filamentous structure within the boundary of a plant root annotation
spore (sp)	Circular structures with a solid outline, connected to AMF external hypha, and outside the boundary of plant root
vesicle (ves)	Circular or rectangular structures with a solid outline and within the boundary of plant root
arbuscule (arb)	Highly branched hypha with fuzzy outline and connected to intraradical hyphae within the boundary of plant root
others	Non-AM fungal structures

2.4.4 Data Cleaning

The Georgia and the Cambridge datasets were cleaned to produce similar input data. All segmentation shapes were approximated by polygons, including converting polyline to polygon directly and resampling points in circles to produce polygons. Some segmentation shapes, including point, rectangle and ellipse, were removed. Empty and undefined segmentations were also removed. Classes with few representative examples were merged into ‘others’. Class labels were made uniform in their vocabularies. The final class list included root, AMF internal hypha, AMF external hypha, AMF arbuscule, AMF vesicle, AMF spore and others. The cleaned Georgia dataset included 746 jpg images that are 2380×1740 pixels and 300 pixels/inch in size and 3577 annotations.

The Cambridge dataset needed additional processing steps. To have comparable input data in size, the 14 images were tiled and subsampled. The images were tiled into squares of 512×512 pixels and smaller images on the boundaries. The segmentations were subsampled to fit each tile. New segmentation polygons were produced at the intersection of the tiles and the original segmentations using Shapely (Gillies & others, 2007--). Polygons with self-intersection were dissected into smaller simple polygons. Points and LineStrings were ignored as subsampling results. Indices for segmentation and bounding box were recalculated relative to the new tiled image. Only tiled images with at least one segmentation annotation were kept. The quality of tiling and subsampling were checked by comparing segmentation in the raw images and the tiled small images visually. The resulting Cambridge dataset included 1379 tiled jpg images that are 512×512 pixels and 96 pixels/inch in size and 20558 annotations.

The Georgia dataset was separated into training, validation and testing sets at 8:1:1 ratio. The training set has 598 images and 2874 annotations. To increase the number of examples for

each fungal structure, the Cambridge dataset was divided in the same 8:1:1 ratio and merged to the previous Georgia training, validation and testing sets. The secondary training set is made up of 1105 images and 16417 annotations. The final prediction set was consisted of 24,391 root images from the ‘TOP’, ‘MID’, and ‘BOT’ regions of 12 sorghum RIL plants. Images with height or width less than 100 pixels in the prediction set were dropped.

2.4.5 Mask R-CNN Model training

Mask R-CNN was implemented in Detectron2 (K. M. He et al., 2017; Yuxin Wu et al., 2019) and is composed of the backbone, the region proposal network (RPN), and heads (K. M. He et al., 2017; Y Wu et al., 2019). The ResNet 50 and FPN (Feature Pyramid Network) backbone extracts feature map from images (K. M. He et al., 2017; T.-Y. Lin et al., 2014). RPN proposes candidate regions (Ren et al., 2015). Heads produce bounding box, mask, and class inferences. The Mask R-CNN model was pretrained on the COCO dataset with 3x schedule (K. M. He et al., 2017; T. Y. Lin et al., 2014; Yuxin Wu et al., 2019). The pretrained model was retrained on the first and secondary training sets for 50 epochs with batch size 2 and the default learning rate schedule.

Different hyperparameters were tested, and each combination was repeated three times with different random seeds. Learning rates of 0.001 and 0.002 were tested. The number of frozen or fine-tuned backbone modules was varied by changing the ‘FREEZE_AT’ parameter from 1 to 3. Two augmentation options were implemented. The default option included image random flip and resize, and the second option added random crop, rotation, and brightness adjustment as augmentation options. Other parameters were set to the defaults in Detectron2 configuration (Yuxin Wu et al., 2019).

Model performance and hyperparameters were evaluated based on mean Average Precision (mAP). The best fine-tuned model for defined hyperparameters was selected based on total loss in validation set during training (K. M. He et al., 2017).

The model quality metric mAP was calculated with varying confidence thresholds and averaged over all classes. In addition, AP50 was calculated at Intersection over union (IoU) level 50%, and AP was averaged over IoU levels from 50% to 95%. Score threshold for inference in test set was set to 0.7.

2.4.6 Mixed Linear Model Prediction And Statistical Analysis

The best model was used for the prediction set of 24,391 images. Other settings remained the same as training. Inferred segmentations in an image were cross-tabulated by class versus segmentation number and pixel number. For downstream statistical analysis on AMF colonization, three class level statistics were generated using the two outputs above.

Count density of an AMF structure was defined as its segmentation number divided by the root pixel number (count/pixel). Average class size of an AMF structure was its pixel number divided by its segmentation number (pixel/count). Percent colonization by an AMF structure was measured as its pixel number per root pixel (pixel/pixel). The three class level colonization statistics were calculated for every slide. A total of 648 entries was used for regression analysis to test for differential colonization in root regions and sorghum plants.

Mixed effect models in ‘lme4’ R package (Bates et al., 2015) were used for modeling the three class level AMF colonization statistics. ANOVA and t-tests were used to test for the significance of model parameters. Likelihood ratio test was used to test the significance of a model to a nested model.

2.4.7 Computational Resources

Model training and inference was implemented on sapelo2 at the Georgia Advanced Computing Resource Center (GACRC) with one p100 GPU, 4 CPUs, and 20 GB memory. GPUs were used for model training. CPUs were used for model inference. Codes are available in GitHub: https://github.com/Arnold-Lab/image_seg_sorghum_am.

2.5 Data Availability

Summary data and codes are available in GitHub: https://github.com/Arnold-Lab/image_seg_sorghum_am. The analyses and manuscript are available in RStudio with the exception of the tables, which were converted manually back to Word formatting from image formatting at the request of the publisher. The large collection of over 20,000 raw images is available upon reasonable request from a shared DropBox folder.

CHAPTER

3

MYCORRHISEE: A HIGH-RESOLUTION IMAGE DATASET FOR DEEP LEARNING
BASED QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI¹

&

A NEW SPLINE-BASED IMAGE PROCESSING TECHNIQUE WITH APPLICATIONS TO
WHOLE-SLIDE IMAGING OF PLANT ROOTS²

¹Zhang, S., T. Bourlai and J. Arnold. 2024. *IEEE International Conference on Big Data 2024*. 10.1109/BigData62323.2024.10825578. Reprinted here with permission of the publisher.

²Zhang, S., W. Lantz, I. Wilson, J. Spittle, S. Eastin, A. Mandal, J. Arnold and T. Bourlai. 2024. *IEEE SoutheastCon 2024*. 10.1109/SoutheastCon52093.2024.10500130. Reprinted here with permission of the publisher.

Abstract

Developing deep learning tools for automated quantification of arbuscular mycorrhizal fungi (AMF) necessitates large, high-quality datasets capturing the diversity of AMF-root interactions under varied environmental conditions. We present MycorrhiSEE, a 15-Terabyte (Tb), high-resolution image dataset comprising approximately 137,500 whole-slide images (WSIs) of root segments from 5,500 sorghum plants representing 337 genotypes, grown under diverse agronomic treatments. Images were acquired using the ZEISS AxioScan 7 high-throughput robotic microscope equipped with a 5-megapixel resolution color camera. This dataset provides unprecedented detail and diversity, offering a valuable resource for developing deep learning applications in AMF research. A new spline-based method was developed to transform WSIs into deep learning-ready image patches. We validated the robustness of the algorithm by analyzing eight bootstrap samples of a thousand images each classified as excellent, moderate, or bad quality by four independent experts. The goodness of fit, roughness, and irregularity of the splines were uniform across all quality levels, confirming our method's reliability for generating patches from gigapixel plant root images. These patches will be used as an input to deep learning algorithms capable of detecting and classifying mycorrhized root segments and types of fungal structures presented.

3.1 Introduction

Arbuscular mycorrhizal fungi (AMF) form symbiotic relationships with the roots of 70–90% of terrestrial plants (Smith & Read, 2010). In this mutualistic partnership, plants transfer up to 20% of their photosynthetically derived carbon to AMF (Johnson et al., 1997a), which in return facilitate the uptake of essential nutrients such as phosphorus, nitrogen, and other micronutrients (W. Wang et al., 2017). AMF significantly enhance plant nutrient absorption, improve soil structure, boost plant resilience to environmental stressors, and increase plant biomass and biodiversity (Baum et al., 2015; María J Pozo & Concepción Azcón-Aguilar, 2007; Juan Manuel Ruiz-Lozano et al., 2012; Zhong et al., 2021), making them critical for sustainable agriculture and ecological balance. To fully harness AMF as mutualistic biofertilizers, large-scale ecological and genomic studies are necessary to understand the complex interactions between AMF, plants, and their environments, as well as the genomic regulation of plant-AMF symbiosis (Klein et al., 2022). Imaging has been the primary method for studying the frequency and distribution of AM hyphal structures (SMITH & SMITH, 1997a). However, traditional imaging methods for quantifying AMF structures are time-consuming and prone to human bias, limiting scalability (McGonigle et al., 1990b). The development of deep learning-based computer vision tools for automated recognition

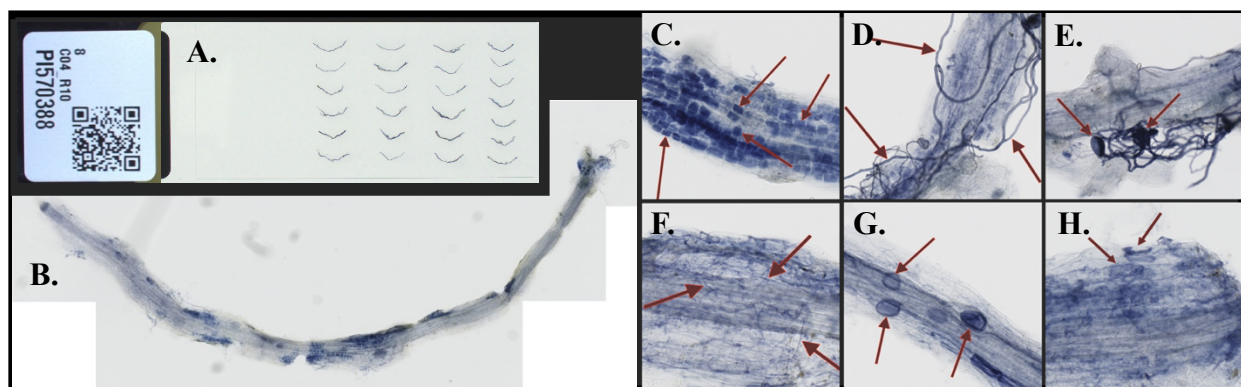


Figure 3.1 Left: A. Root segments arrange in U-shapes on a labeled microscope slide; B. WSI of a single root segment. Right: AM fungal structures were captured in high resolution as shown in the zoomed-in images with arrows pointing to C. arbuscules, D. extraradical hypha, E. spores, F. intraradical hypha, G. vesicles, H. coils.

and segmentation of AMF structures is promising but limited by the lack of comprehensive, annotated datasets that reflect the diversity of AMF-root associations (Evangelisti et al., 2021a; Shufan Zhang et al., 2024).

To address this need, we created MycorrhiSEE, a high-resolution, large-scale dataset of mycorrhized root images, collected utilizing Whole Slide Imaging technology (Shafi & Parwani, 2023). The dataset comprises ~137,500 multi-gigapixel images of fine roots from 5,500 sorghum plants representing 337 genotypes, grown under varied agronomical conditions (**Fig. 3.1**). To facilitate deep learning applications, we have developed a specialized tiling algorithm and created annotations that classify image tiles as colonized or non-colonized and by image quality (S. Zhang, W. Lantz, et al., 2024). The scale and quality of the MycorrhiSEE dataset provide an unprecedented level of detail and diversity, offering a rich resource for the AMF research community and for big data analytics.

3.2 Data Description

3.2.1 Sample Collection

Table 3.1 Overview of Experiments in the Dataset

Experiment	Year	# Plant Genotypes	# Samples	# Slides
Time-series experiment	2020	15	3 replicates 5 timepoints 225 samples	1028
Fertilizer treatments	2021	337	3 replicates 4 treatments 4044 samples	3682
Long-term monoculture	2020-2025	75-85	3 replicates 4 years 960 samples	790

Experiments were conducted in Watkinsville, Georgia to study AMF-plant interactions under varied conditions, including fertilizer treatments, fungal diseases, plant developmental stages, and

long-term monoculture effects (**Table 3.1**). The studies used the Bioenergy Association Panel (BAP), a genetically diverse population of *Sorghum bicolor* (Brenton et al., 2016).

Seedlings from the BAP accessions were germinated in a greenhouse and transplanted to the field after 2 weeks. Each plant was uniquely labeled with a QR code carrying genotype and

field position, a system retained throughout cultivation and image acquisition. At the time of harvest, plant age, flowering status, height, dry weight, tiller number, tiller width, disease load, and root dry weight were recorded. Plant roots were chilled on ice and transferred into 75% ethanol the same day for long-term storage at 4°C. From each plant, a 0.25 g sample of fine roots was randomly selected, and cleared and stained via a modified ink and vinegar protocol (Vierheilig & Piché, 1998). For imaging, 20-30 root segments of ~1 cm were mounted in a U-shape on slides to ensure alignment during stitching (**Fig. 3.1A**).

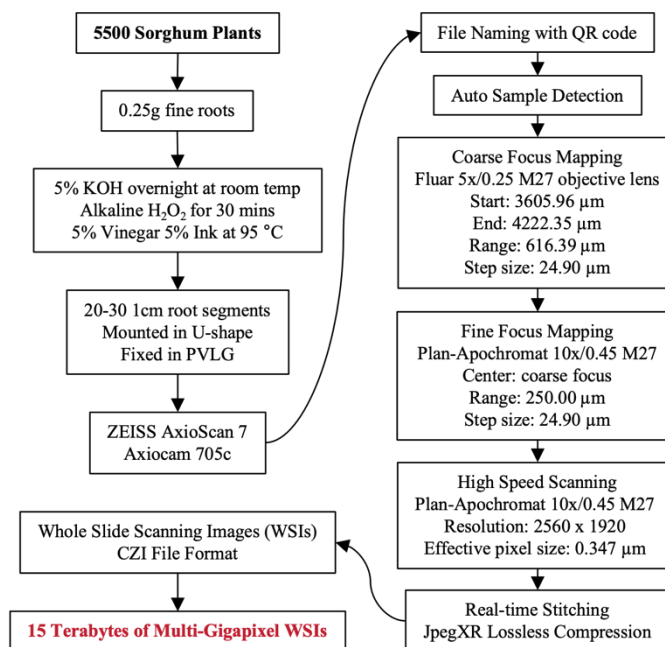


Figure 3.2 A demonstration of the sample processing and image acquisition workflow. The modified ink and vinegar protocol and microscope settings were summarized.

3.2.2 Image Acquisition

A ZEISS AxioScan 7 microscope, equipped with an AxioCam 705 color CMOS camera, was used to scan up to 100 slides per run. The microscope automatically identified QR codes for file naming and followed structured image acquisition steps optimized for Mycorrhisee root samples (**Fig. 3.2**). After automatic sample detection, adaptive coarse and fine focus mapping were

applied to maximize contrast across sample regions. Focus points were evenly distributed across the sample area based on size (mm²). Coarse focusing, performed with a Fluar 5×/0.25 M27 objective lens, covered a 616.39 µm range with a 24.90 µm step size. Fine focusing refined the depth within a 250 µm range centered on the coarse focus and used a 3.84 µm step size with a Plan-Apochromat 10×/0.45 M27 objective lens. During the final scan, image frames were captured at 2560 × 1920 pixel resolution 0.347 µm effective pixel size and stitched in real-time with 10% overlaps, producing WSIs with JpegXR lossless compression. WSIs and metadata were exported in ZEISS CZI format to ensure spatial integrity and data fidelity. The acquisition settings are reproducible by loading the scan profile saved in ZEISS csprof format.

3.2.3 Image Annotations and Quality Assessment

The resulting MycorrhizSEE dataset consists of 15 Tb of WSIs of fine root segments. To assess image quality, we randomly sampled 10,000 WSIs across 10 bootstraps, categorizing them as Excellent, Moderate, or Bad based on U-shape integrity, orientation, root branching, and clarity. The average quality distribution across bootstraps was 44.91% (±5.55) Excellent, 50.09% (±5.35) Moderate, and 5.80% (±1.25) Bad. For detailed quality assessment, a subset of 1,780 WSIs were evaluated for extra image quality factors, including U-shape deformity, dirtiness, opacity, and blurriness, and assigned severity scores on a scale of 1 to 7, where higher scores indicated lower

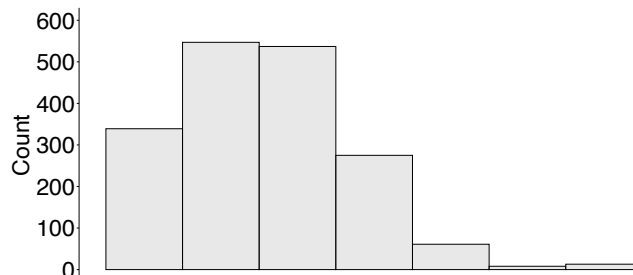


Figure 3.3 The histogram illustrates the quality score distribution of 1,780 WSIs. The scores, ranging from 1 (highest quality) to 7 (lowest quality). Scores of 2 and 3 dominate the distribution, with over 500 images in each category. Few WSIs reached the poorest quality categories (6 or 7), suggesting high image quality of the overall dataset.

image quality. The analysis revealed that the majority of WSIs fell within the higher quality range (Fig. 3.3). Furthermore, we selected high quality image tiles (701×701) with 75% or greater root coverage and created a balanced classification dataset consisting of 1,500 colonized and 1,500 non-colonized tiles.

3.3 Spline-guided Tiling Algorithm

To convert WSIs to direct inputs for deep learning-based computer vision algorithms, we developed a spline-guided tiling algorithm to remove blank backgrounds and generate high root coverage tiles of desired dimensions (S. Zhang, W. Lantz, et al., 2024). We utilized the Otsu's thresholding (Otsu, 1979) to isolate the significant root contours, selecting the largest to convert into a list of points. A quadratic polynomial was fitted to the points. The second derivative of the polynomial was computed to determine the concavity of the root segment. WSIs were rotated so that all root segments were convex. To satiate the injective condition of spline interpolation on a fixed coordinate frame, the smallest y value was selected for a multivalued x. Then, the unique shapes of contours were represented by a piecewise cubic function with six evenly distributed

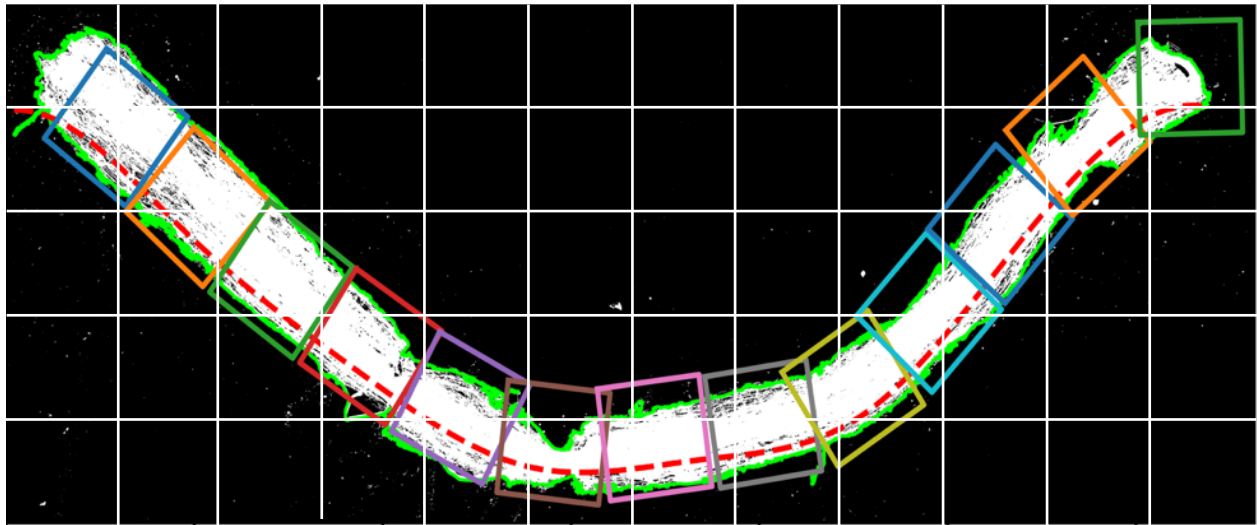


Figure 3.4 Demonstration of our spline-based method vs. standard sliding window method. Red spline guides the positioning of the colorful patches to trace the shape of the root segment. The white gridline divides the entire image into equal sized patches.

internal nodes (De Boor, 2001). We ensured a smooth interpolation by applying two continuous derivatives at each knot. Our algorithm computes uniformly spaced anchor points along the cubic spline, based on estimated arc length using the composite trapezoidal rule. The orientation and vertices of the patches were computed using the tangent and normal vectors at the anchor points, to position patches along the spline's curvature (**Fig. 3.4**).

The robustness of the algorithm was tested with 1000 WSIs bootstrapped 8 times. Four operators independently categorized WSIs into Excellent, Moderate, or Bad based on U-shape integrity, orientation, root branching, and clarity. Excellent images exhibited a complete, correctly oriented U-shape; Moderate images allowed for broken U-shapes and minor imperfections; Bad images included misshapen or highly branched roots and empty or out-of-focus images. Across bootstraps, the image quality distribution averaged 44.91 (± 5.55)%, 50.09 (± 5.35)% and 5.80 (± 1.25)% for Excellent, Moderate, and Bad, respectively. Our performance metrics for fit, roughness, and irregularity—residual sum of squares, third derivative average, and curvature—were stable across quality categories (**Fig. 3.5**).

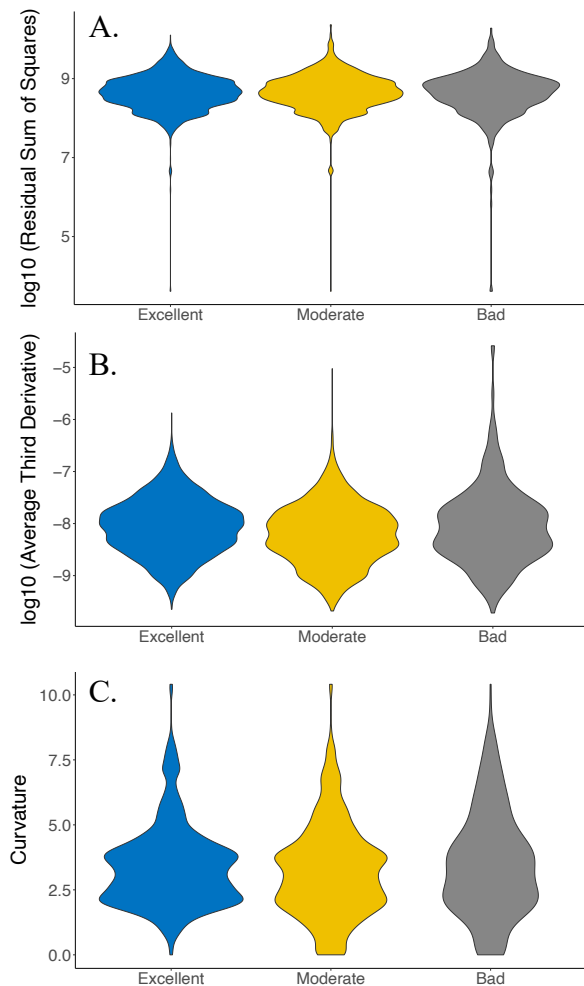


Figure 3.5 In violin plots, A) Residual sum of square of spline on log10 scale; B) Average third derivatives of spline on log10 scale; C) Curvature of spline

For 1000 bootstrap WSIs, we categorized the resulting 701×701 pixel image tiles by root coverage. Approximately 21.1% (8,879 images) had 75–100% root coverage, while the remaining tiles were distributed across 50–75% (9,898 images), 25–50% (6,764 images), and less than 25% root coverage (16,584 images). This distribution revealed underlying taxonomy groups within image tiles by root coverage.

3.4 Conclusion

This annotated dataset serves as a foundation for developing deep learning models for quantifying AMF colonization and evaluating root image quality. The MycorrhizSEE dataset will continue to expand in the coming years as root samples from ongoing and future experiments are imaged and added to the collection. This growth will further enhance the dataset’s diversity and utility for AMF research.

Additionally, we present a robust algorithm for generating patches from WSIs as DL inputs. To showcase its efficiency, we aim to compare it to the standard sliding-window method in producing informative patches. Enhancements to patch quality may involve analyzing the area of overlap with root segments and using local histogram equalization to improve tile image quality. We anticipate the broad applicability of our algorithm, extending to roots taken under contexts beyond mycorrhizal colonization research and analogous structures, such as roads, cracks, and blood vessels, in DL research.

Meanwhile, we are actively developing deep learning-based tools for more comprehensive image quality assessment, enabling finer categorization and filtering of WSIs based on multiple quality metrics. These tools will support more accurate and scalable analyses, maximizing the dataset’s potential for advancing big data-driven solutions for image-based AMF quantification.

3.5 Data Availability

The dataset and annotations are available upon request. The spline based tiling algorithm is available at <https://github.com/Arnold-Lab>. The experiment metadata associated with the WSIs will be disclosed after the experimental results are published.

CHAPTER

4

HIGH-THROUGHPUT QUANTIFICATION OF ARBUSCULAR MYCORRHIZAL FUNGI COLONIZATION USING SPLINE-GUIDED TILING AND CROSS-DOMAIN TRANSFER LEARNING WITH WHOLE-SLIDE IMAGES¹

¹Zhang, S., T. Bourlai and J. Arnold. 2025. To be submitted to *IEEE Access*.

Abstract

Quantifying arbuscular mycorrhizal fungi (AMF) colonization in plant roots through microscopy imaging poses significant challenges due to labor-intensive processes, human biases, and limited throughput. To address these challenges, we introduce an integrated deep learning pipeline with three key innovations. First, we developed an enhanced spline-guided tiling algorithm along with novel quality metrics to evaluate its effectiveness, demonstrating robust performance across two distinct datasets: MycorrhiSEE and AMFinder. Second, we proposed a two-step convolutional neural network (CNN) classification approach to quantify AMF colonization by counting colonized tiles within whole-slide images (WSIs). This stratified classification process simplifies training by initially removing easily identifiable background tiles using a ResNet50-based classifier, achieving an accuracy of 99.7%, and subsequently classifying tiles into colonized and non-colonized root regions. Third, we conducted comprehensive experiments evaluating the within- and cross-domain performance of various pretrained CNN architectures with differing depths, widths, and computational efficiencies. Our findings identify DenseNet and ResNet50 as the most suitable architectures, consistently outperforming baseline models and achieving accuracies exceeding 98%. Collectively, these innovations facilitate scalable, high-throughput, and accurate AMF colonization quantification, significantly advancing precision agriculture and ecological research by overcoming previous computational and generalization limitations.

4.1 Introduction

Precision agriculture enhances crop productivity and sustainability by tailoring agricultural practices to specific land conditions (Ayoub Shaikh et al., 2022). One promising but underutilized approach involves manipulating plant microbiomes to improve crop performance and resilience (Pace et al., 2025). Among microbial symbionts, arbuscular mycorrhizal fungi (AMF) play a crucial role in plant root systems, enhancing nutrient uptake, suppressing disease, and increasing stress tolerance, particularly in arid and nutrient-poor soils (Kakouridis et al., 2022; M. J. Pozo & C. Azcón-Aguilar, 2007; J. M. Ruiz-Lozano et al., 2012). Found in nearly all crops and soil types, AMF contribute to agricultural sustainability by improving soil structure and reducing greenhouse gas emissions (Field et al., 2012b; Hawkins et al., 2023; Wilson et al., 2009).

The global market for mycorrhizae-based biofertilizers, including AMF, is projected to grow from USD 1.29 billion in 2025 to USD 2.05 billion in 2030, doubling in five years (Intelligence, 2024). Despite this growth, there are no standardized guidelines for integrating AMF inoculations with conventional agricultural practices to maximize their benefits for plant and soil health (Rillig et al., 2019). The beneficial effects of mycorrhizal symbiosis are highly context dependent (Johnson et al., 1997b; Koch et al., 2017; Yang et al., 2017). The integration of knowledge on AMF symbiosis across diverse crops, accessions, and management practices remains challenging due to the absence of consistent and objective measures of AMF colonization, hindering cross-study comparisons and data integration (Antunes et al., 2025).

Quantifying AMF colonization in field-grown plants presents significant challenges due to the limitations of traditional imaging methods (McGonigle et al., 1990a; Trouvelot, 1986). These manual techniques are labor-intensive and prone to human bias (Füzy et al., 2015; Kokkoris et al., 2019). Advances in high-throughput imaging and deep learning (DL) offer a transformative

solution (Ferguson et al., 2021). Whole-slide scanning enables the rapid acquisition of large imaging datasets (S. Zhang, T. Bourlai, et al., 2024), while DL algorithms can automate and enhance AMF identification and quantification (Evangelisti et al., 2021b; Shufan Zhang et al., 2024).

However, applying deep neural networks to whole-slide images (WSIs) for AMF quantification remains a non-trivial task. WSIs are massive multi-gigapixel images where fungal colonization is sparse, requiring computationally efficient strategies to extract relevant features (Kong & Henao, 2022). Directly processing WSIs is infeasible due to GPU memory constraints, while excessive downscaling compromises spatial details critical for accurate analysis. Additionally, variability in imaging systems, sample preparation, magnification, sensor resolution, and staining techniques (e.g., differences between AMFinder and MycorrhizSEE datasets) leads to domain shifts, reducing model generalizability (Evangelisti et al., 2021b; S. Zhang, T. Bourlai, et al., 2024). Annotation inconsistency and imbalances further hinder the extraction of robust features across datasets.

To address these challenges, we present an integrated pipeline combining an enhanced spline-guided tiling algorithm (S. Zhang, W. Lantz, et al., 2024), a two-step classification approach, and cross-domain training with ImageNet-pretrained models to quantify AMF colonization in plant root WSIs. Our improvements to the spline-guided tiling algorithm include new quality metrics and modifications to enhance robustness across two whole slide imaging systems (Zeiss vs Keyence). This algorithm efficiently partitions WSIs into biologically relevant regions, excluding extraneous background to optimize computational resources.

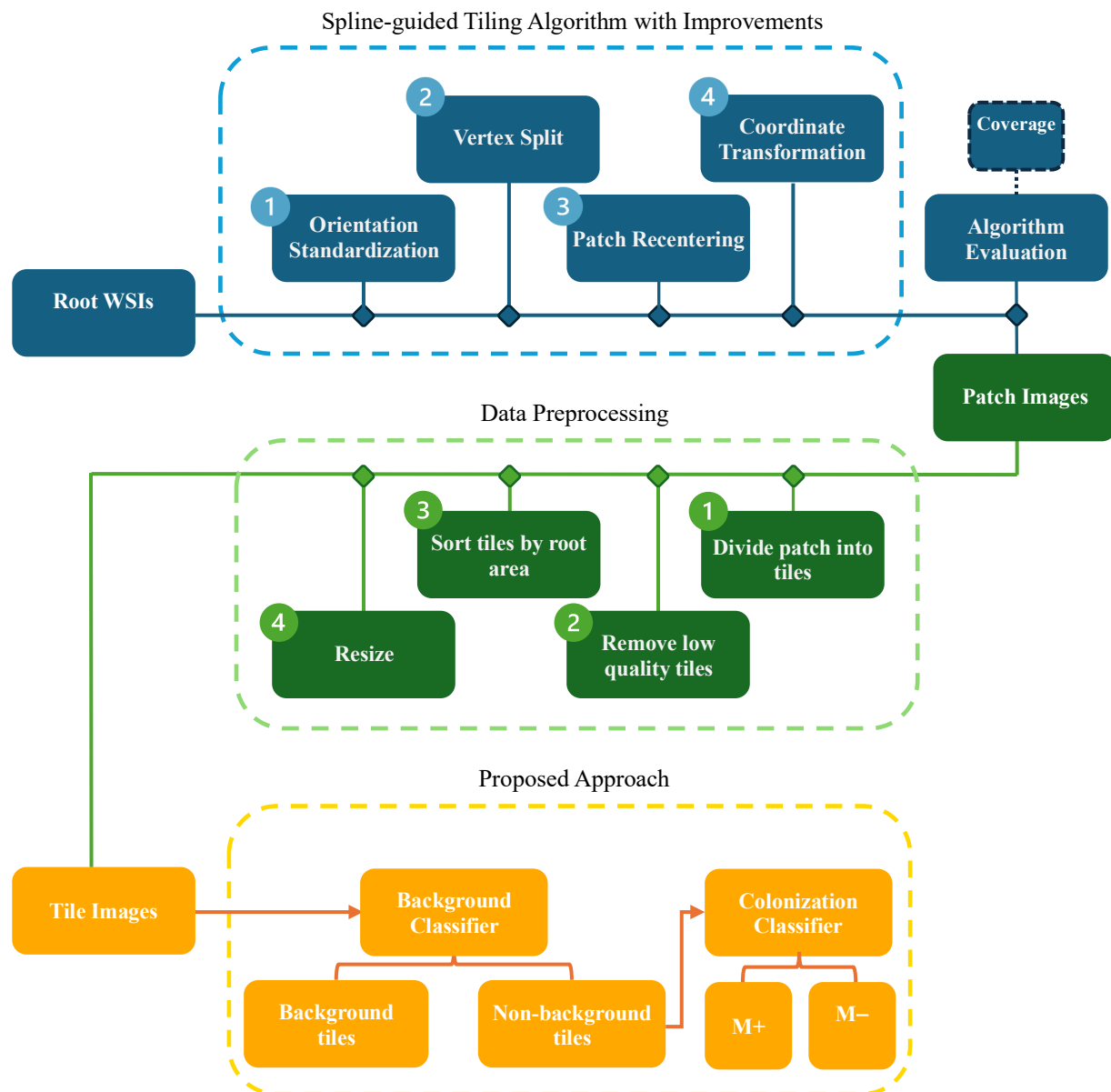


Figure 4.1 A) Workflow of proposed approach using an enhanced spline-guided tiling algorithm with improvements 1-4 to generate input tiles from root WSIs, a data processing step to filter out low quality tiles with low root coverage, and two classifiers to count AM colonized tiles.

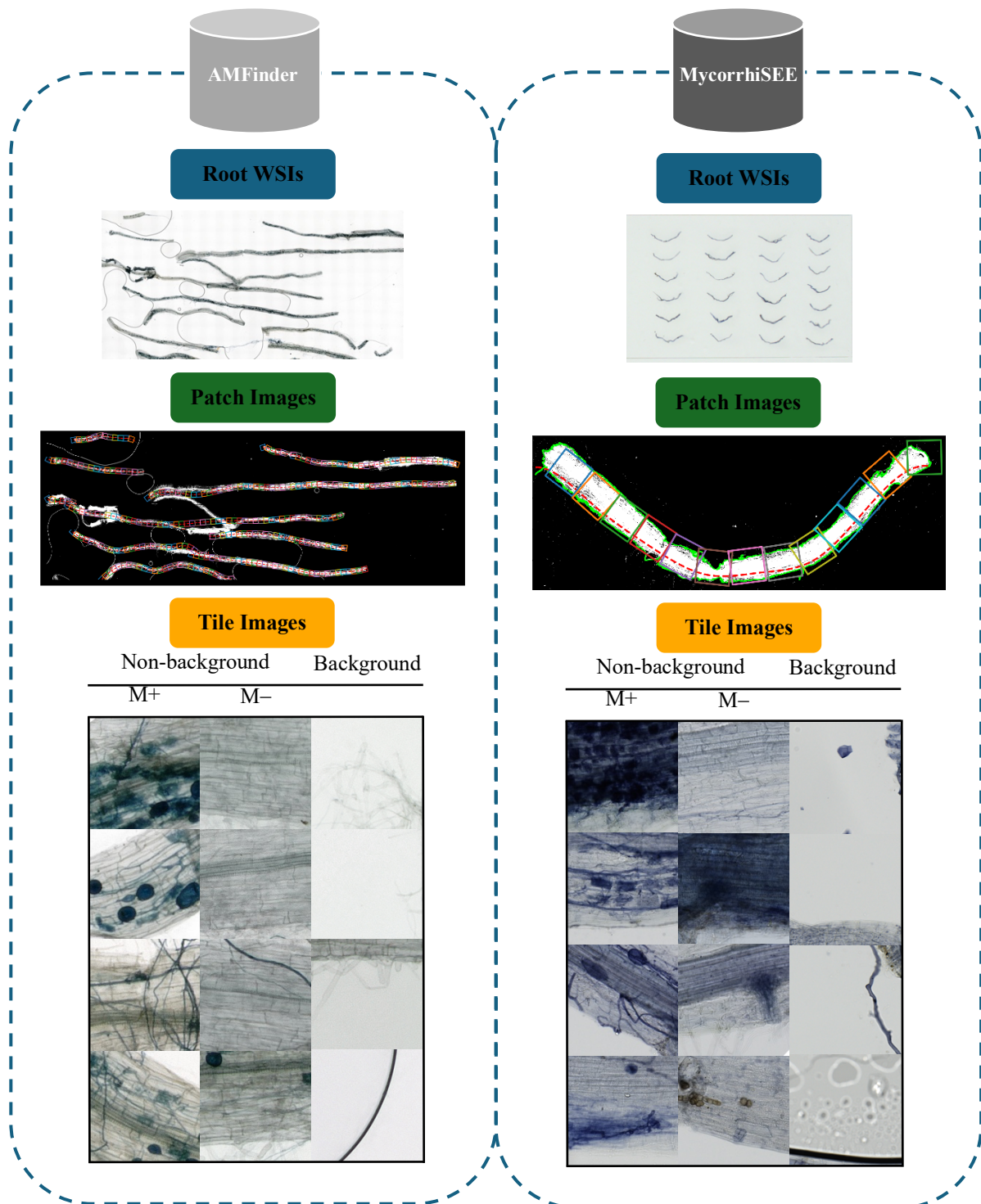


Figure 4.1 B) Examples of AMFinder and MycorrhhiSEE images in the proposed workflow demonstrated in Fig. 4.1A.

The two-step classification first employs a ResNet50-based classifier to filter out background tiles (Background Classifier), followed by fine-tuned deep convolutional neural networks (CNNs) to distinguish colonized (M+) and non-colonized (M−) root regions (Colonization Classifier). We evaluate a diverse set of ImageNet-pretrained architectures, including shallow and deep networks optimized for parameter propagation and training efficiency (Fig. 4.1A).

Our approach demonstrates that integrating whole-slide imaging with deep learning achieves high classification accuracy and balanced performance, providing a scalable, high-throughput tool for AMF quantification. Our findings highlight the importance of transfer learning and cross-domain training for robust mycorrhizal colonization classification across diverse imaging conditions. By systematically evaluating pretrained deep learning architectures, we provide critical insights into optimal model design for AMF quantification. These advancements enhance the precision and efficiency of AMF colonization analysis, supporting the broader application of high-throughput phenotyping in plant–microbe interaction research.

4.2 Related Work

4.2.1 Microscopy-Based AMF Characterization

Microscopy imaging has played a fundamental role in the study of arbuscular mycorrhizal fungi (AMF), facilitating research on their morphology, symbiosis, and ecological functions. Early AMF studies in the 1900s relied on staining techniques such as trypan blue (Phillips & Hayman, 1970) and the ink-and-vinegar method (Vierheilig et al., 1998) to visualize fungal structures, enabling detailed characterization of hyphae, appressoria, arbuscules, vesicles, and spores across species and genera using brightfield and dissection microscopes (Dodd et al., 2000; Friesse & Allen, 1991; Smith & Smith, 1997b; Smith & Read, 1997). Metabolic stains indicate nutrient exchange

between plant host and AMF (Boddington & Dodd, 1998, 1999; Vosatka & Dodd, 1998; Wanxiao Wang et al., 2017).

By the 2000s, microscopy remained important for measuring AMF colonization across a variety of plant genotypes and environmental conditions (Baon et al., 1993; Cockerton et al., 2020; De Vita et al., 2018a; Ellouze et al., 2015; Hetrick et al., 1992; Johnson et al., 2022; Kaeppler et al., 2000; M. L. Pawlowski et al., 2020; Plouznikoff et al., 2019; Sawers et al., 2017; Schultz et al., 2010; Stahlhut et al., 2021; Taylor et al., 2015; S. J. Watts-Williams et al., 2019). Fluorescence and live-cell imaging further advanced AMF research, enabling real-time observation of signaling pathways and metabolic exchanges in symbiosis (Ivanov & Harrison, 2014; Kobae & Hata, 2010; Kokkoris et al., 2020; MacLean et al., 2017; Oyarte Galvez et al., 2025). More recently, genomic advancements have been coupled with high-resolution imaging, linking genetic markers to AMF colonization patterns, morphological characteristics, and functional responses (Martin & van der Heijden, 2024; Säle et al., 2021).

Despite these advancements, scalability remains a major challenge, as conventional microscopy techniques are labor-intensive and lack the throughput necessary for large-scale studies. Standardized high throughput imaging techniques could further improve the accuracy and reproducibility of quantitative estimates, allowing researchers to measure hyphal, vesicular, and arbuscular root colonization more reliably, ultimately enhancing ecological models predicting AMF functionality (Antunes et al., 2025; Kokkoris et al., 2019). The increasing demand for high-throughput, automated imaging solutions has driven the adoption of deep learning-based computer vision methods to enhance the efficiency of AMF quantification.

4.2.2 Deep Learning for Image-Based Analysis

Convolutional Neural Networks (CNNs), a subset of deep learning (DL), have become the dominant approach for image classification, object detection, and segmentation in various domains (Dong et al., 2021; Noor & Ige, 2024; Younesi et al., 2024). A typical CNN architecture consists of convolution layers, pooling layers, and fully connected layers. Convolution acts as powerful feature extractors, enabling neural networks to learn spatial and hierarchical representations of image data, including edges, textures, and shapes (Krizhevsky et al., 2017).

The modern evolution of CNNs can be contributed to increased GPU processing power and the availability of large-scale labeled datasets such as ImageNet (Deng et al., 2009). Since 2012, CNN-based models have dominated the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), achieving human-level accuracy by 2015 with the introduction of ResNet (He et al., 2016). However, CNNs rely on large labeled datasets and substantial training, limiting their application in specialized domains where labeled data is scarce (Cheng et al., 2022; Iqra et al., 2024).

To address these limitations, transfer learning has emerged as a promising solution (Raina et al., 2007). By leveraging the pretrained CNNs trained on large datasets, transfer learning enables the models to be fine-tuned to new tasks with fewer labeled examples (Yosinski et al., 2014; Zhuang et al., 2019). This approach has been tested effective at fungal structure identification (Krach et al., 2022; Krach et al., 2020) and across numerous agricultural applications (Hossen et al., 2025), including plant phenotyping (Jiang & Li, 2020), pest detection (Chen et al., 2022), disease identification (Nazir et al., 2023), and yield estimation (Khaki et al., 2021).

4.2.3 CNNs for AMF Quantification

Efforts have been made to apply CNNs for automating imaging-based AMF quantification. Evangelisti et al developed the AMFinder pipeline, a semi-automated deep learning-based method that enables user-supervised validation of CNN-generated predictions (Evangelisti et al., 2021b). AMFinder works on whole-slide scanning images (WSIs) of stained root samples and splits them into smaller tiles for CNN classification. CNN1 distinguishes between colonized (M+), non-colonized (M-), and background tiles. CNN2 further classifies M+ tiles by AM fungal structures. Users can manually review and correct CNN predictions using AMFbrowser. AMFinder demonstrated that CNNs provide a feasible solution for the automation of AMF quantification, but it remains reliant on supervised training on a large volume of labeled data.

In another study, Mask R-CNN was trained for AMF segmentation using a transfer learning-based approach (K. He et al., 2017; Shufan Zhang et al., 2024). This method reduced labeling requirements, but imaging was done using traditional compound microscopy and manually moving of the camera across slides, which limited scalability.

While these studies point to the potential of CNNs and transfer learning for AMF quantification, challenges remain in developing scalable, high-throughput solutions capable of handling WSIs, domain variability, and large-scale studies. The integration of whole-slide imaging, CNN, and transfer learning presents a promising pathway toward automated, high-throughput AMF quantification.

4.3 Methodology

4.3.1 Image Preprocessing

To ensure comparable physical coverage between the AMFinder and MycorrhizSEE images (Fig. 1B), we calculated equivalent pixel regions for both imaging systems. The AMFinder WSIs

were captured with a VHX-5000 microscope at 200x magnification using a 1/1.8-inch (7.2 mm) CMOS sensor (1600 x 1200 resolution, 3.6 μm sensor pixel size, 0.95 μm actual pixel size) and cropped to 256×256 pixel image tiles. To match the same physical area in the MycorrhhiSEE dataset, taken by the ZEISS Axioscan 7 microscope (Carl Zeiss Microscopy, LLC, Thornwood, NY) at 10x magnification with an Axiocam 705 color camera (2464×2056 resolution, 3.45 μm sensor pixel size, 0.347 μm actual pixel size), a 701×701 pixel region was required.

For the initial arrangement of spline-guided patches in the MycorrhhiSEE dataset, a patch size of 1402×1402 pixels was used. These patches were further subdivided into four 701×701 tiles and resized to 256×256 pixels to match the dimensions of AMFinder image tiles.

4.3.2 Image Taxonomy

In analyzing the AMFinder and MycorrhhiSEE image tiles, we observed substantial variation in the amount of background pixels within each image, as demonstrated in **Fig. 4.2**. This variability posed significant challenges for consistent colonization labeling, as images with differing root area may represent fundamentally different taxonomic groups by root area or by visual characteristics, including colonization intensity. Failure to account for this underlying image

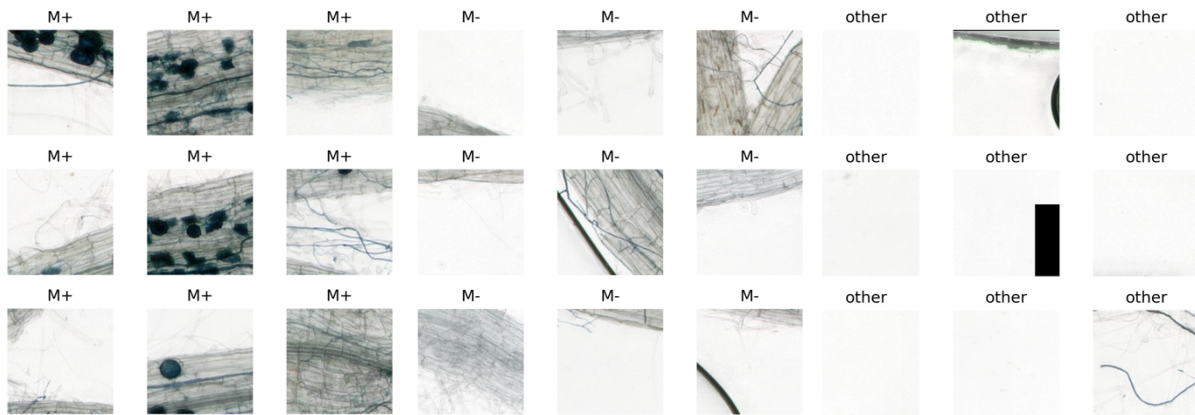


Figure 4.2 Demonstration of how variable the amount of root area per image is using AMFinder tile images. Labels are published in the AMFinder paper. Consistent colonization labeling cannot be achieved without accounting for image taxonomy related to the amount of background pixel per tile.

taxonomy could lead to imbalanced class representations at deeper levels, despite efforts to maintain a balanced dataset overall. To address this issue, we refined our training and testing datasets for AMF colonization classification by selecting only images with 75% or more visible root area (**Fig. 4.1B**).

Upon grouping the images by root area, we found that only 15.0% (2040 images) of AMFinder tiles contained 75-100% root coverage, while a larger portion fell into lower root area categories: 2714 images (50-75%), 3876 images (25-50%), and 4910 images (less than 25%). In the MycorrhizSEE dataset, 21.1% (8879 images) met the 75-100% root coverage threshold, with the remaining images distributed across 50-75% (9898 images), 25-50% (6764 images), and less than 25% root area (16584 images). These distributions highlight the significant presence of low root area images reinforces the importance of filtering based on root coverage to ensure a consistent training base for accurate model learning and robust performance across varying image taxonomies.

Table 4.1 Comparison of convolutional neural networks used. Convolution layers (conv). Fully connected layers (fc). Residual block (res).

Network	Year	Input Size	Depth	Parameters (millions)	Size	Convolutional Layers
VGG16	2014	224×224×3	16	138	528 MB	13 conv, & 3 fc
ResNet50	2015	224×224×3	50	25.6	98 MB	1 conv, 16 bottleneck res, & 1 fc
DenseNet121	2017	224×224×3	121	8.0	33 MB	4 conv, & 4 dense blocks
DenseNet169	2017	224×224×3	169	14.3	57 MB	4 conv, & 4 dense blocks
DenseNet201	2017	224×224×3	201	20.0	80 MB	4 conv, & 4 dense blocks
MobileNetV2	2018	224×224×3	53	3.5	13 MB	1 conv, & 19 bottleneck res.
EfficientNetV2-B0	2021	224×224×3	24	7.1	29 MB	2 conv, & 11 Fused-MBConv blocks
EfficientNetV2-B1	2021	240×240×3	26	8.1	31 MB	2 conv, & 12 Fused-MBConv blocks
AMFinder CNN1	2022	126×126×3	11	1.4	5.4MB	8 conv, & 3 fc

4.3.3 Dataset Preparation

For background classification, a tile was annotated as Background, if it had: 1) at least 75% image area to be background, blank, bubble or debris, 2) no blurring, excellent root clearing and fungal staining, and 3) less than 3 extraradical hyphae. The resulting dataset was made of 2000 Background and 2000 Non-background tiles with an equal representation of M+ and M- tiles from AMFinder and MycorrhizSEE datasets (**Fig. 4.1B**).

For colonization classification, we annotated tiles from AMFinder and MycorrhizSEE dataset using following criteria: 1) less than 25% background pixels, 2) no blurring, excellent root clearing and fungal staining, and 3) at least 5 hyphal structures to be colonized (M+), otherwise noncolonized (M-). The final dataset consisted of 1000 M+ and 787 M- tiles from AMFinder dataset and 1000 M+ and 1000 M- tiles from the MycorrhizSEE dataset. The AMFinder root samples were highly colonized, resulting in an insufficient number of M- tiles.

The datasets were divided into training, validation, and test sets using a 7:2:1 split. Tile images were resized to 224×224×3 for model input, except for EfficientNetV2B1 which requires an input size of 240×240×3 (**Table 4.1**). Model training utilized three datasets, AMFinder images, MycorrhizSEE images, and a Combined dataset (AMFinder + MycorrhizSEE). The training set size was determined to ensure accuracy within ± 0.02 of 0.928, requiring a minimum of 641 images for a 0.02 margin of error or 2,566 images for a 0.01 margin. Testing was conducted using three corresponding test sets: AMFinder, MycorrhizSEE, and Combined, allowing for robust evaluations within and across domains.

4.3.4 Pretrained ImageNet CNNs

Several ImageNet-pretrained models were evaluated for their ability to classify Background, M+, and M- tile images: VGG16, ResNet50, DenseNet121, DenseNet169,

DenseNet201, MobileNetV2, and EfficientNetV2 B0 and B1. AMFinder CNN1 architecture shares significant similarities with VGG16, featuring a sequential stack of 3x3 convolutional layers followed by max-pooling and fully connected (fc) layers; however, AMFinder has only 1.4 million parameters in its multiclass configuration, much lighter than VGG16's 138 million parameters. ResNet50, known for its bottleneck residual blocks (res), helps mitigate overfitting. MobileNetV2 and EfficientNetV2 prioritize efficiency through depthwise separable convolutions and compound scaling, while DenseNet models leverage dense connections to promote feature reuse and efficient parameter utilization (**Table 4.1**). This range of architectures allowed us to compare shallow and deep networks, as well as models optimized for different computational trade-offs.

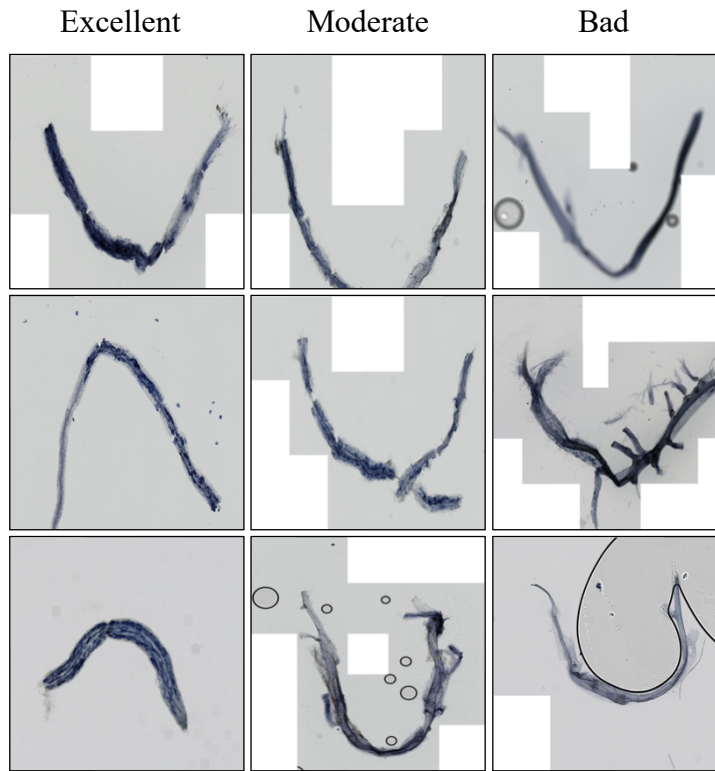


Figure 4.3 Examples of ZEISS AxioScan 7 WSIs of root segments. Images were labeled as excellent, moderate, bad based on root segment conformation and image quality. Excellent: complete curve with changing concavity. Moderate: disjoint or with minor imperfections (curving back, branching, air bubbles) Bad: misshapen or highly branched roots and empty or out-of-focus images. (see Methods for detailed description of labeling standards). White pixels are paddings for the background area around samples detected by the microscope. Images were resized for display.

4.3.5 Model Training

In this study, we adopted a two-step approach to classify tiles from root WSIs as M+, M-, or Background. First, we trained a Background Classifier. An ImageNet pretrained ResNet50 was trained to filter out Background tiles. A second Colonization Classifier was used to classify M+ and M- tiles from images predicted as Non-background. Several ImageNet pretrained models were finetuned for colonization classification task (**Fig. 4.1A**).

All models were trained using a batch size of 64 with the fully connected and the classification layers removed, and the backbone weights frozen. The optimization process employed Binary Cross-Entropy loss function and Adam optimizer with an initial learning rate (LR) of 0.0001 and a maximum of 50 epochs.

To enhance model convergence and prevent overfitting, two adaptive training strategies were implemented. The learning rate was dynamically reduced by a factor of 0.5 when the validation loss plateaued for 3 consecutive epochs, with a minimum allowable learning rate of 1×10^{-6} . Additionally, early stopping was triggered if no improvement in validation loss was observed for 5 consecutive epochs.

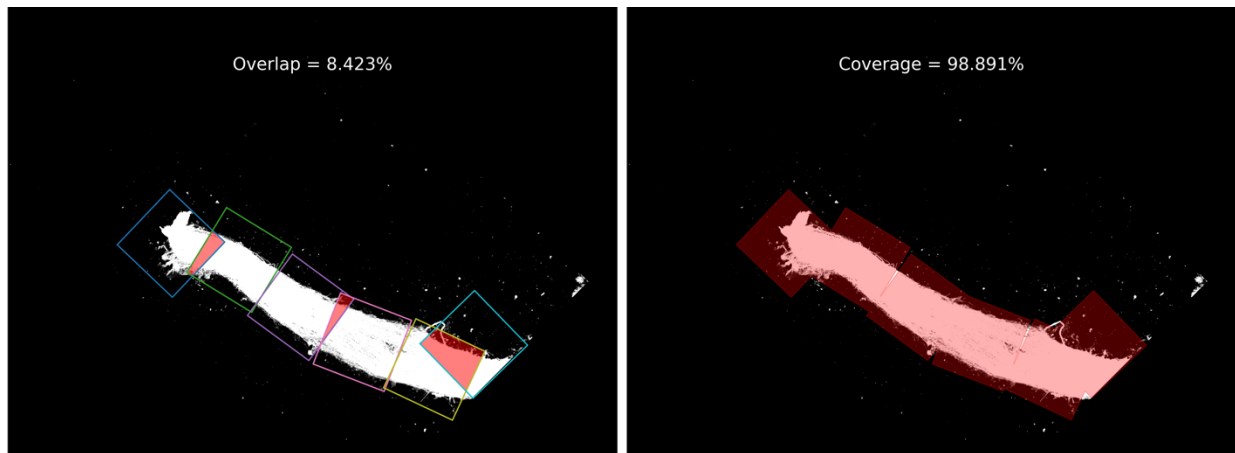


Figure 4.4 Visualization of patch overlap and root coverage using the spline-guided tiling algorithm on ZEISS AxioScan 7 images. The left panel illustrates the overlap between neighboring patches, which is 8.423%. The right panel shows the root coverage achieved, amounting to 98.891%. These metrics highlight the algorithm's ability to effectively cover the region of interest while managing overlap.

4.4.1 Experiment 1: Evaluation of the Improved Spline-Guided Tiling Algorithm on Mycorrhizal SEE Dataset

To evaluate the effectiveness of the original spline-guided tiling algorithm in generating patches, it was applied to 1005 Mycorrhizal SEE WSIs distributed evenly across quality categories (Excellent, Moderate, Bad) using patch size 1024×1024 . The Mycorrhizal SEE dataset consists of field grown *Sorghum bicolor* root WSIs imaged using ZEISS AxioScan 7 at 10x magnification. Image acquisition and quality categories were defined in previous publications (S. Zhang, T. Bourlai, et al., 2024) and summarized in **Fig. 4.3**. Key algorithm performance metrics included (**Fig. 4.4**):

- **Root Coverage:** Percentage of root area covered by patches.
- **Patch Overlap:** Degree of overlap between adjacent patches.
- **Patch Redundancy:** The amount of background patches out of all patches.

The original algorithm achieved average root coverage of 84.6% ($\pm 11.8\%$), 80.0% ($\pm 14.2\%$), and 72.3% ($\pm 19.4\%$) for Excellent, Moderate, and Bad WSIs (**Fig. 2**), respectively. Patch overlap across varying quality remained around 4% (3.8% Excellent, 3.70% Moderate, 4.40% Bad). It was estimated that over 90% WSIs in the Mycorrhizal SEE dataset had Excellent and Moderate quality. There was a significant data loss due to low patch coverage. To minimize data loss, four incremental improvements were implemented:

- **Improvement 1: Orientation Standardization**

To enhance spline interpolation, we fitted a polynomial to root contours, used the second derivative to estimate concavity, and rotated each image to achieve uniform orientation.

- **Improvement 2: Patch Recentering**

The original algorithm's injective condition retained only the highest y-value among points sharing the same x-value, causing the spline to drift from the root midline. Patches were recentered along normal vectors to realign with the midline. The first two adjustments improved overall root coverage by 5.2% ($p < 0.001$) (Table 4.2 Contrast 2-0).

- **Improvement 3: Handling Disjoint Root Segments**

Disjoint root segments in Moderate quality WSIs, partially recognized by the original algorithm, were addressed by splitting images at polynomial vertices and extracting the largest root contour from each half.

- **Improvement 4: Coordinate Frame Transformation**

Half-images were mirrored and rotated 45 degrees. This transformation resolved issues to assign patches to root sections with extreme curvature within a fixed coordinate frame.

Table 4.2 Tukey Post hoc test of root coverage and patch overlap between spline guided tiling algorithm versions. Contrast represents comparisons between two algorithm versions. Version 0: original algorithm, Version 1-4: improved algorithms with improvement 1 to 4 sequentially implemented.

Root Coverage						
Contrast	Excellent		Moderate		Bad	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
1-0	-0.017	0.948	-0.006	1.000	-0.011	0.999
2-1	0.057	<0.001	0.064	<0.001	0.068	<0.001
3-2	0.011	0.999	0.018	0.922	0.027	0.379
4-3	0.040	0.012	0.046	0.001	0.053	<0.001
2-0	0.040	0.012	0.059	<0.001	0.057	<0.001
3-0	0.051	<0.001	0.077	<0.001	0.084	<0.001
4-0	0.091	<0.001	0.123	<0.001	0.138	<0.001
Patch Overlap						
Contrast	Excellent		Moderate		Bad	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
1-0	-0.007	0.350	-0.004	0.994	-0.006	0.698
2-1	0.031	<0.001	0.026	<0.001	0.027	<0.001
3-2	0.005	0.838	0.011	0.004	0.014	<0.001
4-3	-0.001	1.000	-0.003	0.998	0.001	1.000
2-0	0.023	<0.001	0.023	<0.001	0.021	<0.001
3-0	0.029	<0.001	0.034	<0.001	0.036	<0.001
4-0	0.028	<0.001	0.031	<0.001	0.037	<0.001

The enhancements collectively improved root coverage by 9.1%, 12.3%, and 13.8% to 93.7%, 92.3% and 86.1% for Excellent, Moderate, and Bad images, respectively, while patch overlap increased modestly from 3.95% to 7.11%.

4.4.2 Experiment 2: Application of the Spline-Guided Tiling Algorithm to AMFinder Dataset

The algorithm was evaluated on AMFinder WSIs using two tile sizes (512×512 and 256×256) to test its adaptability to different imaging systems and influence of patch size selection

on algorithm performance.

AMFinder dataset consists of 7 large and 20 small WSIs of lab grown *Nicotiana benthamiana* roots captured with a VHX-5000 digital microscope (Keyence, Milton Keynes, UK) at 200x. Root coverage improved significantly with larger patch size 512×512 pixels ($p<0.001$), while overlap increased by 0.46%

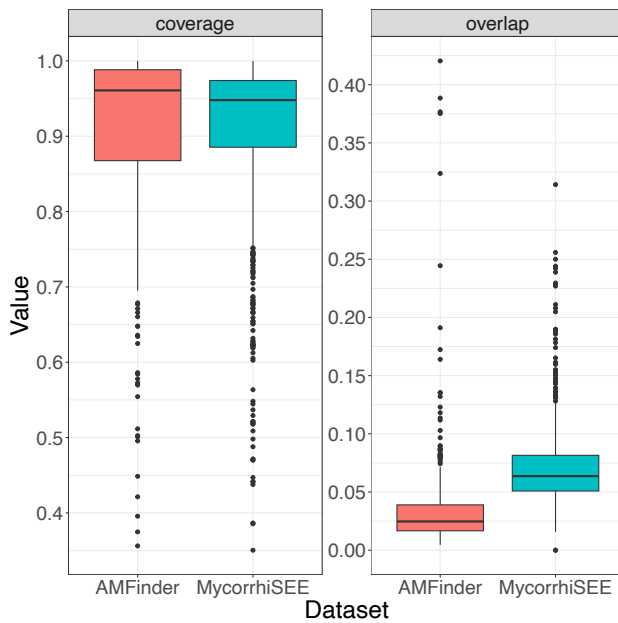


Figure 4.5 Comparing tiling quality of the algorithm on AMFinder and MycorrhizSEE datasets. The boxplots display the distribution of root coverage and tile overlap for each dataset. The spline guided tiling algorithm demonstrated robust performance

($p<0.001$) (Table 4.3). Using patch size 512×512, the average root coverage of the AMFinder dataset

Table 4.3 Statistic properties of root coverage and patch overlap for two tile sizes.

Patch size	coverage		overlap	
	mean	sd	mean	sd
256 x 256	0.751	0.192	0.0312	0.0405
512 x 512	0.907	0.124	0.0358	0.0502

was 90.7%, comparable to the 90.8% coverage across quality categories for MycorrhizSEE dataset (**Fig. 4.5**), demonstrating robust performance even on extensively entangled roots in the AMFinder WSIs. Comparison to the regular gridline-based tiling method demonstrated the algorithm's superior efficiency in reducing patch redundancy. The number of background tiles (256×256 pixels) decreased from 146,769 to 9,499 (human confirmed). Tile size 256×256 was used to generate tiles from AMFinder WSIs for the classification experiments below.

4.4.3 Experiment 3: Evaluate Proposed Two-Step Classification Approach

A ResNet50-based CNN was first trained to filter out Background tiles (Background Classifier), achieving an accuracy of 0.997. Subsequently, several ImageNet-pretrained models—VGG16, ResNet50, MobileNetV2, DenseNet121, DenseNet169, DenseNet201, and EfficientNetV2—were trained to perform classification of colonized (M+) and non-colonized (M-) image tiles (Colonization Classifiers). A dataset combination experiment was conducted to evaluate the performance of the Colonization Classifiers against AMFinder CNN1 models for their ability to correctly identify AMF colonized tiles (M+ tiles).

The AMFinder and MycorrhizSEE datasets used in the experiment differ in imaging systems, sample preparation, root maturity, fungal morphology, and background texture (**Fig. 4.1B**). These variations offered a valuable scenario for evaluating domain dependency, model robustness, and the efficiency of cross-domain learning. To fully explore these capabilities, the experiment involved training the models on three distinct data configurations:

1. **AMFinder-only**: Training on the AMFinder dataset.
2. **MycorrhizSEE-only**: Training on the MycorrhizSEE dataset.
3. **Combined dataset**: Training on a heterogeneous dataset that merged AMFinder and MycorrhizSEE images.

The experimental design evaluated the models' performance across three key scenarios:

- 1 **Within-domain classification:** Training and testing on the same dataset (e.g., AMFinder-only).
- 2 **Cross-domain generalization:** Training on one dataset and testing on another (e.g., AMFinder-trained model tested on Mycorrhizal images).
- 3 **Domain adaptation:** Training on the Combined dataset and testing across all test sets (AMFinder, Mycorrhizal, and Combined).

4.4.3.1 Within Domain Performances of Colonization Classifiers

Performance of AMFinder CNN1 Models

AMFinder CNN1 models were evaluated on tiles from the AMFinder dataset to establish baseline performance. Despite its larger training set (AMFinder paper appendix), the model demonstrated poor accuracy in multiple scenarios. Trained without augmentation, AMFinder CNN1v1 achieved only 53.3% overall accuracy on the AMFinder test set, with a pronounced bias toward the noncolonized class (25% versus 88.75%) (**Table 4.4**), indicating limited generalization even within the same dataset (Fig. 6A). While training with augmentation improved accuracy to 77.2%, the results of AMFinder CNN1v2 retained high misclassification rates for positive and negative examples (Table 4). Class imbalance and inconsistent labeling during training could have contributed to the poor performance of AMFinder CNN1 models. Visualization of original training images from the AMFinder dataset revealed mis-labeled tiles (**Fig. 4.2**). More robust feature extraction methods and training strategies are needed to reliably classify AMF colonization.

Performance of Pretrained Models

Within-domain evaluations revealed that the ImageNet-pretrained architectures outperformed the baseline AMFinder CNN1 models. When trained and tested on the AMFinder

dataset, the finetuned networks consistently achieved overall accuracies exceeding 98%, with DenseNet169 and DenseNet201 reaching approximately 99.44% accuracy and exhibiting near-perfect per-class performance (**Fig. 4.6A**). When trained and tested on Mycorrhizal dataset, architectures such as ResNet50 and DenseNet variants achieved overall accuracies in the range of 93–97.5% with balanced per-class accuracies (**Table 4.4**). These findings underscore the efficacy of leveraging ImageNet pretraining to enhance model classification performance within specific domains.

4.4.3.2 Cross Domain Performances of Single Dataset Models

The cross-domain evaluations revealed a significant impact of domain shift on model performance (comparing Fig. 4.6A and Fig. 4.6B). When models were trained on the AMFinder dataset and tested on the Mycorrhizal set, overall accuracies ranged from 54.5% to 69.5%, with pronounced class imbalances (**Table 4.5**). For example, ResNet50 achieved an overall accuracy of approximately 69.5% with a stark disparity between the colonized (40%) and noncolonized (99%) classes (**Table 4.5**). Similar trends were observed for DenseNet121 and DenseNet201, where high noncolonized accuracies were countered by very low colonized accuracies (30% and 9%, respectively) (**Table 4.5**). In contrast, when the training was conducted on the Mycorrhizal dataset with subsequent testing on AMFinder, overall performance improved, with ResNet50 and EfficientNetV2B0 reaching 83.3% and 81.1% accuracy, respectively (**Table 4.5**). However, this configuration also exhibited notable imbalances; several architectures, such as DenseNet121, DenseNet169, DenseNet201, and VGG16, attained perfect or near-perfect accuracies for the colonized class while substantially underperforming on the noncolonized class (with accuracies as low as 19%–41%) (**Table 4.5**). These findings underscore the challenges of cross-domain generalization with single dataset training in AMF colonization classification. Notably, the

baseline AMFinder CNN1 models continued to underperform in cross-domain settings (**Fig. 4.6B**), with both variants exhibiting a pronounced bias (**Table 4.5**).

4.4.3.3 Domain Adaptation of Combined Dataset Models

Building on the previously discussed cross-domain performance, training with the diverse combined dataset of AMFinder and MycorrhizSEE markedly improved generalization across models (comparing Fig. 4.6C to Fig. 4.6B). In this configuration, the networks not only achieved near-perfect performance when tested on the AMFinder dataset—with overall accuracies exceeding 98% and balanced per-class accuracies—but also demonstrated substantial enhancements on the MycorrhizSEE test set, where overall accuracies ranged from approximately 87% to 96.5% (**Table 4.6**). By integrating imaging conditions and colonization patterns from both datasets during training, the models effectively mitigated the biases observed in the earlier cross-domain experiments.

In addition, further evaluations on a combined test set provide additional insights into the benefits of cross-domain learning. Models trained on the AMFinder dataset exhibited moderate overall accuracies on the combined test set—ranging from approximately 74% (e.g., DenseNet201) to 82% (e.g., ResNet50 and DenseNet169) (**Table 4.7**). In contrast, networks trained on the MycorrhizSEE dataset generally achieved higher overall accuracies on the combined test set (up to 91% with ResNet50), with consistently higher colonized accuracies (**Table 4.7**). Models trained on MycorrhizSEE surpass those on AMFinder when tested on a combined dataset, indicating MycorrhizSEE dataset's potential for domain generalization (**Fig. 4.6D**). These findings, in conjunction with our earlier cross-domain evaluations, highlight that training on a single domain does not sufficiently capture the variability inherent in heterogeneous imaging data, thereby impairing the model's ability to generalize. Conversely, the enhanced performance observed with

diverse training data suggests that integrating multiple datasets enables the models to learn more robust, domain-invariant features, ultimately leading to improved and more balanced cross-domain performance.

4.5 Conclusion

4.5.1 Overall Accuracy of Two-Classifier Approach

The proposed two-step classification pipeline (Background Classifier + Colonization Classifier) exhibited consistently strong performance across the AMFinder, MycorrhhiSEE, and combined test sets. In the first step, the Background Classifier achieved near-perfect accuracy of 0.997 with ResNet50. In the second step, the Colonization Classifiers demonstrated excellent results. DenseNet121 trained on combined data delivered 100% accuracy on the AMFinder test set, while ResNet50 trained specifically on MycorrhhiSEE showing the best overall accuracies of 0.975 on the MycorrhhiSEE test data. These findings confirm that the proposed two-step approach can reliably segregates background tiles and automate the analysis of mycorrhizal colonization.

4.5.2 Best Model Architecture for AMF Colonization Classification

The complete performance analysis indicates that modern deep convolutional architectures, particularly DenseNet and ResNet50 variants, markedly outperform the baseline AMFinder CNN1 models across both the AMFinder and MycorrhhiSEE datasets. Additionally, EfficientNet architectures, while competitive, did not consistently match the performance of DenseNet or ResNet50 models, likely due to differences in reduced network depth and design trade-offs between model complexity and parameter efficiency. These findings imply that, for the AMF colonization classification problem, architectural depth and connectivity—whether through dense or residual linkages—play critical roles in achieving high accuracy and balanced class predictions.

The comprehensive evaluation of deep convolutional architectures for AMF colonization classification demonstrates that ImageNet-pretrained networks significantly outperform the baseline AMFinder CNN1 models across both within-domain and cross-domain scenarios. Cross-domain analyses further revealed that models trained solely on a single domain suffered from notable performance drops and class imbalances when exposed to external datasets. Cross-domain training enabled the networks to learn domain-invariant features and achieve more balanced performance across disparate test sets. These findings underscore the importance of incorporating heterogeneous data during model training to enhance generalization in real-world applications of DL in quantification of AMF.

In summary, the results suggest that future customized architectures for AMF colonization classification should prioritize robust feature propagation and multi-scale representation through enhanced depth and connectivity. The integration of diverse training datasets is also essential to mitigate domain shift and ensure consistent, high-accuracy performance in varied imaging conditions. These insights provide a clear direction for further research and development of tailored deep learning models to support precise and reliable AMF colonization diagnostics.

Table 4.4. Within Domain Classification Accuracy for AMF Colonization of Single Dataset Models

Network	Test	Acc.	M+ Acc.	M- Acc.
AMFinder CNN1v1	AMFinder	0.533	0.250	0.888
AMFinder CNN1v2	AMFinder	0.772	1	0.488
VGG16	AMFinder	0.983	0.988	0.980
	MycorrhhiSEE	0.940	0.960	0.920
ResNet50	AMFinder	0.978	0.975	0.980
	MycorrhhiSEE	0.975	0.980	0.970
DenseNet121	AMFinder	0.989	0.975	1
	MycorrhhiSEE	0.955	0.950	0.960
DenseNet169	AMFinder	0.994	1	0.990
	MycorrhhiSEE	0.965	1	0.930
DenseNet201	AMFinder	0.994	0.988	1
	MycorrhhiSEE	0.955	0.9900	0.920
MobileNetV2	AMFinder	0.978	0.963	0.990
	MycorrhhiSEE	0.930	0.930	0.930
EfficientNet V2-B0	AMFinder	0.972	0.975	0.970
	MycorrhhiSEE	0.935	0.970	0.900
EfficientNet V2-B1	AMFinder	0.967	0.975	0.960
	MycorrhhiSEE	0.940	0.950	0.930

Table 4.6. Classification Accuracy for AMF Colonization of Combined Dataset Models

Network	Test	Acc.	M+ Acc.	M- Acc.
VGG16	AMFinder	0.989	0.988	0.990
	MycorrhhiSEE	0.944	0.917	0.972
ResNet50	AMFinder	0.989	0.988	0.990
	MycorrhhiSEE	0.975	0.983	0.967
DenseNet121	AMFinder	1	1	1
	MycorrhhiSEE	0.972	0.967	0.978
DenseNet169	AMFinder	0.972	0.988	0.960
	MycorrhhiSEE	0.950	0.972	0.928
DenseNet201	AMFinder	0.994	1	0.990
	MycorrhhiSEE	0.969	0.989	0.950
MobileNetV2	AMFinder	0.972	0.963	0.980
	MycorrhhiSEE	0.917	0.883	0.950
EfficientNet V2-B0	AMFinder	0.983	0.975	0.990
	MycorrhhiSEE	0.947	0.933	0.961
EfficientNet V2-B1	AMFinder	0.972	0.988	0.960
	MycorrhhiSEE	0.956	0.967	0.944

Table 4.5. Cross-Domain Classification Accuracy for AMF Colonization of Single Dataset Models

Network	Test	Acc.	M+ Acc.	M- Acc.
AMFinder CNN1v1	MycorrhhiSEE	0.500	0.0400	0.960
AMFinder CNN1v2	MycorrhhiSEE	0.630	0.8800	0.380
VGG16	AMFinder	0.672	1	0.410
	MycorrhhiSEE	0.625	0.530	0.720
ResNet50	AMFinder	0.833	0.988	0.710
	MycorrhhiSEE	0.695	0.400	0.990
DenseNet121	AMFinder	0.744	0.988	0.550
	MycorrhhiSEE	0.645	0.300	0.990
DenseNet169	AMFinder	0.661	0.963	0.420
	MycorrhhiSEE	0.690	0.480	0.900
DenseNet201	AMFinder	0.550	1	0.190
	MycorrhhiSEE	0.545	0.090	1
MobileNetV2	AMFinder	0.772	0.550	0.950
	MycorrhhiSEE	0.685	0.470	0.900
EfficientNet V2-B0	AMFinder	0.811	0.713	0.890
	MycorrhhiSEE	0.695	0.500	0.890
EfficientNet V2-B1	AMFinder	0.744	0.750	0.740
	MycorrhhiSEE	0.645	0.720	0.570

Table 4.7. Classification Accuracy for AMF Colonization of Single Dataset Models on Combined Test Set

Network	Train	Acc.	M+ Acc.	M- Acc.
VGG16	AMFinder	0.783	0.733	0.833
	MycorrhhiSEE	0.836	0.978	0.694
ResNet50	AMFinder	0.822	0.656	0.989
	MycorrhhiSEE	0.914	0.983	0.844
DenseNet121	AMFinder	0.797	0.600	0.994
	MycorrhhiSEE	0.869	0.967	0.772
DenseNet169	AMFinder	0.825	0.711	0.939
	MycorrhhiSEE	0.844	0.983	0.706
DenseNet201	AMFinder	0.744	0.489	1
	MycorrhhiSEE	0.797	0.994	0.600
MobileNetV2	AMFinder	0.814	0.689	0.939
	MycorrhhiSEE	0.850	0.761	0.939
EfficientNet V2-B0	AMFinder	0.817	0.711	0.922
	MycorrhhiSEE	0.875	0.856	0.894
EfficientNet V2-B1	AMFinder	0.789	0.833	0.744
	MycorrhhiSEE	0.850	0.861	0.839

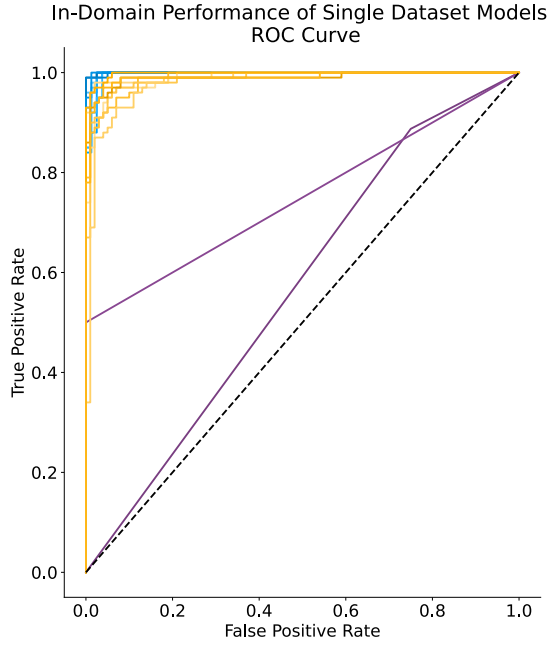


Figure 4.6A. Single dataset models are trained with either AMFinder or MycorrhhiSEE dataset and tested with images from the same dataset to evaluate their in-domain performance. Blue lines: Models trained and tested with AMFinder images. Yellow lines: Models trained and tested with MycorrhhiSEE images. Purple lines: Baseline performance of AMFinder classifiers CNN1v1 and CNN1v2 on AMFinder images.

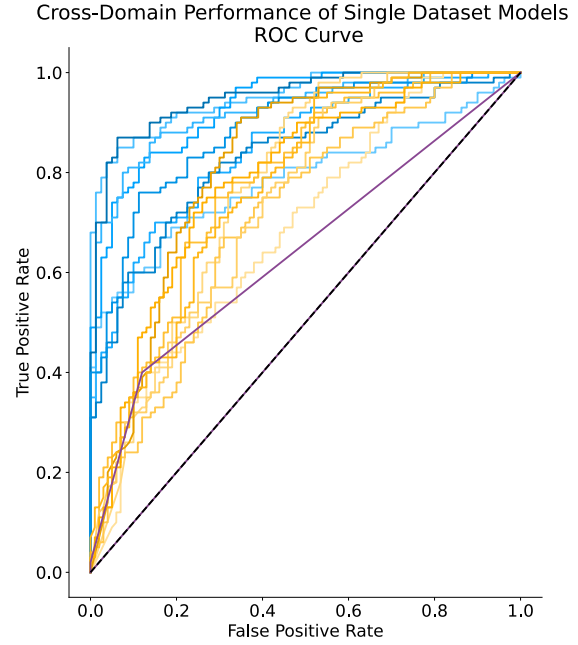


Figure 4.6B. Single dataset models are trained with either AMFinder or MycorrhhiSEE dataset tested with images from the same dataset to evaluate their in-domain performances. Blue lines: Models trained with MycorrhhiSEE images and tested on AMFinder images. Yellow lines: Models trained with AMFinder images and tested on MycorrhhiSEE images. Purple lines: Baseline performance of AMFinder classifiers CNN1v1 and CNN1v2 on MycorrhhiSEE images.

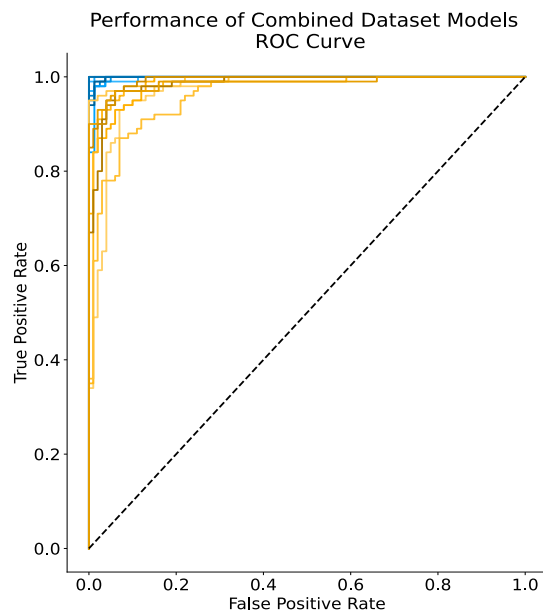


Figure 4.6C. Combined dataset models are trained with AMFinder and MycorrhiSEE images. Blue lines: Models tested on AMFinder images. Yellow lines: Models tested on MycorrhiSEE images. Comparing Fig. 4.6C to 6B, we can conclude that combined dataset models generalize better than single dataset models.

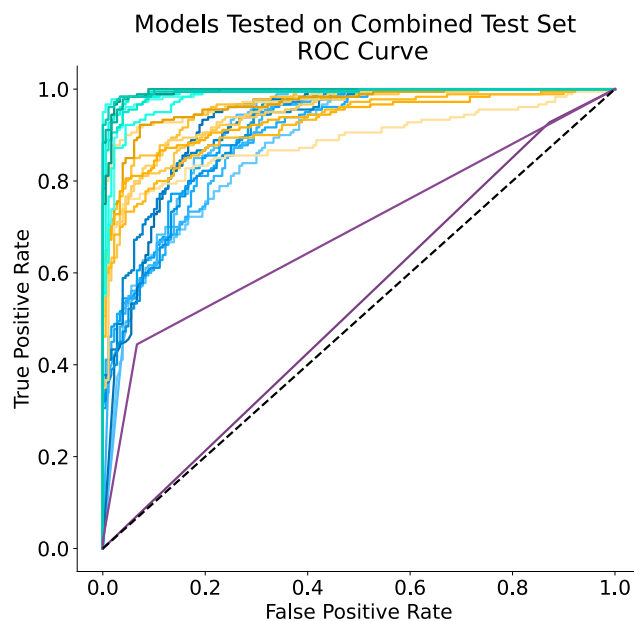


Figure 4.6D. Single and combined dataset models are tested on the combined test set. Teal lines: Combined models. Yellow lines: MycorrhiSEE models. Blue lines: AMFinder models. Purple lines: Baseline AMFinder classifiers CNN1v1 and CNN1v2. Pretrained models significantly outperform baseline AMFinder classifiers. Model performance increases when trained with diverse images with the combined models achieving the best results.

CHAPTER

5

LIMITATIONS AND FUTURE DIRECTIONS

5.1 Dataset Limitations

Limitation 1: Imbalanced Representation of AM Fungal Structures (Chapter 2)

In Chapter 2, one critical limitation was the imbalanced representation of AM fungal structures within the dataset used for training and testing the Mask R-CNN model. Due to limited personnel available at the beginning of the project, we only managed to image and annotate 165 root samples over a year, with the least represented fungal structure class having only 135 examples. Such imbalance likely compromised the model's ability to learn effectively across all classes. Although we used loss functions designed to mitigate class imbalance, these adjustments alone might not have fully compensated for the uneven representation.

As an immediate solution, we supplemented our training data by annotating additional root images from a publicly available dataset, AMFinder. In hindsight, a better approach would have been to prioritize class balance, randomly sampling annotations to match the number of examples of the least represented class. An iterative strategy could then be employed, progressively adding more annotations only if initial model training results were inadequate. Additionally, we could have explored image augmentation techniques specifically targeting underrepresented classes before merging our in-house data with the external AMFinder images.

Limitation 2: Unknown Variability in Image Quality Across the MycorrhiSEE Dataset (Chapter 3)

Another significant limitation discussed in Chapter 3 was the unknown variability in image quality across the extensive MycorrhiSEE dataset. In contrast to Chapter 2, which involved a carefully curated dataset of 10,944 manually collected images (2380×1740 pixels) from 57 root samples, the MycorrhiSEE dataset was composed of gigapixel WSIs from approximately 5,500 root samples collected over five years and multiple experiments. As datasets scale in size, maintaining uniform root clearing and staining quality becomes inherently challenging, inevitably leading to variability in image quality (S. Zhang, T. Bourlai, et al., 2024).

We are implementing an imaging profile to ensure that the robotic microscope accounts for varying sample quality and searches for optimal imaging depth for each root segment. This adaptive control has its limits. Platform variability is likely to be an issue as well. We currently have research technicians reviewing all completed WSIs when imaging finished and redo samples with blurry images. The variability in sample quality, however, cannot be evaluated at the time of imaging. The unknown distribution of these quality categories complicated efforts to assemble a truly representative training dataset, potentially impacting the accuracy and generalizability of deep learning models.

To address these limitations, we are currently developing machine learning and deep learning models capable of automatically detecting and categorizing low-quality images. With these tools, we aim either to filter out low-quality images or to strategically balance the representation of each quality category in both training and testing sets, thereby improving model robustness and generalization. Image quality enhancement is also worth exploring to minimize data loss. Another application of the image quality assessment ML and DL models is quality

control on the root samples. By screening for poorly cleared or stained root samples, we could identify problematic batches, repeat the sample preparation step, and improve image quality (Naus, 1975, 2015).

Limitation 3: Annotation Variability and Human Subjectivity (Chapters 2-4)

A consistent limitation across Chapters 2 to 4 was the inherent variability and subjectivity of human annotations. Annotation errors, typically ranging from 5% to 10%, introduced inconsistencies into the dataset. Recognizing this issue, we implemented strategies to mitigate variability. First, we assigned annotation tasks to groups of two or three annotators, enabling cross-checking and consensus-building. Additionally, before large-scale annotation, we conducted practice rounds in which all annotators labeled the same several hundred images. Results from these sessions were carefully reviewed and any discrepancies corrected, ensuring alignment among annotators. Clearly documented labeling rules were circulated and periodically reinforced among annotators throughout the annotation process. Furthermore, if inconsistencies were discovered during model training and testing, labels were corrected promptly, and the affected models were retrained. Despite these rigorous measures, the subjective nature of annotations remains a challenge, underscoring the ongoing need for clear guidelines, continuous quality control, and possibly automated annotation assistance tools in future studies.

5.2 Technical Limitations

Limitation 1: Computational Constraints Posed by WSIs (Chapters 3 and 4)

A major technical limitation encountered during this research is the computational constraints associated with processing multi-gigapixel WSIs. Due to GPU memory limitations, WSIs were not directly utilized as inputs for deep learning algorithms, as the size of these images far exceeds the memory available on standard GPUs. To address this constraint, we developed and

utilized strategic image patching methods, including spline-based techniques presented in Chapters 3 and 4. The spline-based tiling algorithm effectively segments WSIs into smaller, manageable patches, carefully excluding non-informative background regions with no root segments. These smaller patches significantly reduce the computational load, enabling their direct use as inputs to deep learning models without compromising spatial resolution. This approach allows models to leverage full resolution image data while maintaining feasibility within existing hardware constraints.

Limitation 2: Challenges in Segmenting Small, Intricate Fungal Structures (Chapter 2)

Another substantial challenge, discussed extensively in Chapter 2, relates to accurately segmenting small and intricate fungal structures such as extraradical hyphae and arbuscules. Segmentation delineates the boundaries of fungal structures in root samples. Detecting and segmenting these structures are inherently problematic for several reasons. First, the low resolution of small fungal structures provides fewer pixels to represent their morphological details, making them difficult to distinguish from the background of the root cortex, for example. Additionally, small objects often lack distinctive features or sufficient contextual information to facilitate precise identification. Human scorers utilize the contextual information in identifying AM fungal structures, such as the connections of arbuscules to intraradical hyphae or extraradical hyphae being outside the root cortex and distinguishing them from other mycorrhizal fungi. Variability in lighting and staining conditions further exacerbates this problem, causing inconsistencies in fungal structure classification and complicating detection of fungal structures.

To mitigate these challenges, we divided WSIs into smaller patches to enhance AMF classification and segmentation accuracy at the expense of losing contextual information. Patching significantly reduced the relative size differences between fungal structures and the overall image

dimensions. In Chapter 2, during the annotation process, annotators maintained tight segmentation margins around AM fungal structures and deliberately excluded ambiguous examples to ensure high-quality ground-truth labels. In Chapters 3 and 4, our spline-guided tiling algorithm, efficiently removed redundant background regions, while each patch provided sufficient contextual information about the root cortex and the relevant fungal structures. We carefully selected patch sizes to retain adequate context, as AM fungal structures never occur in isolation or in the root stele.

Future research should explore new DL architectures and methodological improvements to address this limitation. Transformer-based detectors and attention mechanisms also show significant potential in accurately capturing fine-scale details (Cheng et al., 2022). Additionally, advanced learning methodologies such as Knowledge Distillation (Nabavi et al., 2024), Self-Supervised Learning (W. Zhou et al., 2024), and Reinforcement Learning (Fang et al., 2024) represent promising directions to further improve model performance, robustness, and adaptability when dealing with small and intricate fungal structures.

Limitation 3: Data Storage and Transfer (Chapters 3 and 4)

The substantial size of the MycorrhizSEE dataset, discussed in Chapters 3 and 4, introduces additional technical limitations regarding data storage and transfer. Currently encompassing approximately 15 Terabytes of image data and continually expanding, the sheer scale of this dataset significantly complicates data management practices. Moving, backing up, and sharing such a large dataset are increasingly challenging. Cloud-based computing solutions, often recommended for high-performance processing tasks, are impractical in this scenario since few, if any, supercomputing facilities permit uploading datasets of this magnitude, although the Amazon Web Service (AWS) remains a possibility (Skaro et al., 2022).

Learning from this limitation, it would have been beneficial to adopt high-performance Solid State Drives (SSDs) earlier in the project for data storage. SSDs offer notably faster read-write speeds (500-3,500 MB/s) compared to traditional Hard Disk Drives (HDDs) (30-150 MB/s), significantly accelerating data retrieval and improving efficiency during intensive image processing tasks and deep learning model training. Moving forward, strategic investments in local computational infrastructure—such as dedicated servers equipped with high-performance GPUs, CPUs, and large-capacity SSD storage arrays—will likely be necessary. Such infrastructure would alleviate current bottlenecks associated with data handling, enabling more streamlined and efficient data processing and model development workflows in future large-scale image-based deep learning projects.

5.3 Methodological Limitations

Limitation 1: Limitations of Transfer Learning

One critical methodological limitation encountered throughout this project, pertains to the constraints inherent in transfer learning approaches. While leveraging pretrained CNNs trained on large, general-purpose datasets (e.g., ImageNet and COCO) provided an efficient starting point for developing AM fungal detection models, challenges arose regarding model generalization across diverse plant genotypes, varying agronomic conditions, distinct imaging platforms, and differing staining and image acquisition protocols. These domain-specific factors introduced significant variability that was not entirely captured by models pretrained on generalized image datasets. Consequently, the reliance on extensive annotated datasets tailored to each new imaging scenario has posed scalability challenges, restricting broader adoption of deep learning approaches within AMF research.

Future efforts should include rigorous fine-tuning of pretrained CNNs for specialized applications on AMF quantification, systematically varying the width and depth of fully connected layers, gradual unfreezing of the CNN backbone, and hyperparameter finetuning using such tools as the Maximally Informative Next Experiment (MINE) (Torres et al., 2025) and Bayesian optimization (Nogueira, 2014). Additionally, architectures that leverage attention mechanisms or hybridize CNN with transformer models could markedly improve generalization by capturing long-range dependencies and context more effectively (Cheng et al., 2022; Fang et al., 2024). Self-supervised learning techniques represent another promising avenue, significantly alleviating the dependence on large volumes of labeled data by leveraging unlabeled data to derive feature representations transferable across diverse AMF image analysis tasks (Zhou et al., 2024). Furthermore, exploring multi-frame feature fusion and cross-modal approaches may substantially enhance detection and segmentation of small fungal structures in complex backgrounds (Jiang et al., 2024). For example, combining fluorescent microscopy images with traditional brightfield images may synergistically improve model robustness and accuracy. Another possible solution to an expanded training dataset is using Interactive Semi-Automatic Annotation Tool integrated with Segment Anything (ISAT-SAM) to speed up the annotation process (Kirillov et al., 2023).

Limitation 2: Insufficient Evaluation of Data Augmentation Techniques (Chapters 2 and 3)

Another methodological constraint recognized in Chapters 2 and 3 involves the limited exploration and systematic evaluation of data augmentation techniques. Data augmentation holds significant promise for addressing class imbalance issues and facilitating effective merging of datasets from multiple imaging domains. Although we employed basic augmentation methods (such as flipping and rotation), these techniques alone might not sufficiently address the variability

across domains, nor do they fully mitigate the effects of imbalanced representation among fungal structure classes.

Moving forward, comprehensive evaluation and validation of more sophisticated augmentation methods, such as color jittering, elastic deformation, synthetic minority oversampling techniques (SMOTE), and advanced generative approaches (e.g., generative adversarial networks or GANs), may provide substantial improvements. Quantitative assessments of augmentation methods should involve rigorous experimentation comparing model performance metrics before and after the augmentation steps. Systematic evaluation of domain-specific augmentation effectiveness will ensure that future model training and domain integration efforts benefit fully from augmentation, leading to more robust and generalizable models.

Limitation 3: Model Interpretability Issues

Another methodological limitation is the difficulty in interpreting deep learning models, which significantly hinders biological understanding of the derived features and decision-making processes (Chen et al., 2023). While CNNs and Mask R-CNN models exhibited excellent performance in detecting and quantifying AM fungal structures, the inherent “black-box” nature of these deep learning architectures presents a challenge for biological interpretability. Understanding the specific image features driving model predictions is crucial, particularly for validating biological hypotheses and ensuring meaningful scientific conclusions.

To enhance interpretability, we suggest adopting visualization techniques for filters and feature maps learned by deep neural networks. Methods such as guided backpropagation, gradient-weighted class activation mapping (Grad-CAM), and saliency mapping can visually highlight the specific regions and image features critical to model predictions. Visualizing intermediate convolutional filters and feature maps may provide insights into what features the network

prioritizes when identifying specific fungal structures. Future research should systematically integrate these visualization tools into the model evaluation workflow, bridging the gap between computational models and biological insights. Improved interpretability can increase trust in deep learning models among biologists, fostering broader acceptance and more meaningful biological interpretation of deep learning-derived results.

REFERENCES

- Abdelhalim, T., Ramia, J., & Joergensen, R. G. (2020). Arbuscular mycorrhizal dependency and phosphorus responsiveness of released, landrace and wild Sudanese sorghum genotypes. *Archives of Agronomy and Soil Science*, 66(5), 706-716.
- Antunes, P. M., Stürmer, S. L., Bever, J. D., Chagnon, P.-L., Chaudhary, V. B., Deveautour, C., Fahey, C., Kokkoris, V., Lekberg, Y., Powell, J. R., Aguilar-Trigueros, C. A., & Zhang, H. (2025). Enhancing consistency in arbuscular mycorrhizal trait-based research to improve predictions of function. *Mycorrhiza*, 35(2), 14.
- Ayoub Shaikh, T., Rasool, T., & Rasheed Lone, F. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture*, 198, 107119.
- Bagayoko, M., George, E., RÖMheld, V., & Buerkert, A. (2000). Effects of mycorrhizae and phosphorus on growth and nutrient uptake of millet, cowpea and sorghum on a West African soil. *The Journal of Agricultural Science*, 135(4), 399-407.
- Banerjee, S., Schlaeppi, K., & van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology*, 16(9), 567-576.
- Baon, J. B., Smith, S. E., & Alston, A. M. (1993). Mycorrhizal responses of barley cultivars differing in P efficiency. *Plant and Soil*, 157(1), 97-105.

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1 - 48.
- Baum, C., El-Tohamy, W., & Gruda, N. (2015). Increasing the productivity and product quality of vegetable crops using arbuscular mycorrhizal fungi: a review. *Scientia horticultruae*, 187, 131-141.
- Birhane, E., Gebremeskel, K., Tadesse, T., Hailemariam, M., Hadgu, K. M., Norgrove, L., & Negussie, A. (2018). Integrating Faidherbia albida trees into a sorghum field reduces striga infestation and improves mycorrhiza spore density and colonization. *Agroforestry Systems*, 92(3), 643-653.
- Birhanu, M. W., Negussie, Z. T., Birhanu, M. W., Negussie, Z. T., & Salimonti, A. (2024). Arbuscular Mycorrhizal Fungi symbiosis in sustainable production of sorghum (*Sorghum bicolor* L. Moench) under drought stress: an emerging biofertilizer in dryland areas. *Advances in agriculture.*, 2024(1).
- Boatwright, J. L., Brenton, Z. W., Boyles, R. E., Sapkota, S., Myers, M. T., Jordan, K. E., Dale, S. M., Shakoor, N., Cooper, E. A., Morris, G. P., & Kresovich, S. (2021). Genetic characterization of a *Sorghum bicolor* multiparent mapping population emphasizing carbon-partitioning dynamics. *G3: Genes, Genomes, Genetics (Bethesda)*, 11(4).
- Boatwright, J. L., Sapkota, S., Jin, H., Schnable, J. C., Brenton, Z., Boyles, R., & Kresovich, S. (2022). Sorghum Association Panel whole-genome sequencing establishes cornerstone resource for dissecting genomic diversity. *The Plant Journal*, 111(3), 888-904.
- Boatwright, J. L., Sapkota, S., Myers, M., Kumar, N., Cox, A., Jordan, K. E., & Kresovich, S. (2022). Dissecting the genetic architecture of carbon partitioning in sorghum using multiscale phenotypes. *Frontiers in Plant Science*, 13, 790005.

- Boddington, C. L., & Dodd, J. C. (1998). A comparison of the development and metabolic activity of mycorrhizas formed by arbuscular mycorrhizal fungi from different genera on two tropical forage legumes. *Mycorrhiza*, 8(3), 149-157.
- Boddington, C. L., & Dodd, J. C. (1999). Evidence that differences in phosphate metabolism in mycorrhizas formed by species of *Glomus* and *Gigaspora* might be related to their life-cycle strategies. *New Phytologist*, 142(3), 531-538.
- Bonfante, P., & Genre, A. (2010). Mechanisms underlying beneficial plant–fungus interactions in mycorrhizal symbiosis. *Nature communications*, 1(1), 1-11.
- Boyles, R. E., Brenton, Z. W., & Kresovich, S. (2019). Genetic and genomic resources of sorghum to connect genotype with phenotype in contrasting environments. *The Plant Journal*, 97(1), 19-39.
- Boyles, R. E., Pfeiffer, B. K., Cooper, E. A., Zielinski, K. J., Myers, M. T., Rooney, W. L., & Kresovich, S. (2017). Quantitative trait loci mapping of agronomic and yield traits in two grain sorghum biparental families. *Crop Science*, 57(5), 2443-2456.
- Brenton, Z. W., Cooper, E. A., Myers, M. T., Boyles, R. E., Shakoor, N., Zielinski, K. J., Rauh, B. L., Bridges, W. C., Morris, G. P., & Kresovich, S. (2016). A genomic resource for the development, improvement, and exploitation of sorghum for bioenergy. *Genetics*, 204(1), 21-33.
- Brown, P., Klein, P., Bortiri, E., Acharya, C., Rooney, W., & Kresovich, S. (2006). Inheritance of inflorescence architecture in sorghum. *Theoretical and Applied Genetics*, 113, 931-942.
- Brown, P. J., Rooney, W. L., Franks, C., & Kresovich, S. (2008). Efficient mapping of plant height quantitative trait loci in a sorghum association population with introgressed dwarfing genes. *Genetics*, 180(1), 629-637.

- Brundrett, M. C. (2002). Coevolution of roots and mycorrhizas of land plants. *New Phytologist*, 154(2), 275-304.
- Brundrett, M. C. (2009). Mycorrhizal associations and other means of nutrition of vascular plants: understanding the global diversity of host plants by resolving conflicting information and developing reliable means of diagnosis. *Plant and Soil*, 320(1), 37-77.
- Brundrett, M. C., & Tedersoo, L. (2018). Evolutionary history of mycorrhizal symbioses and global host plant diversity. *New Phytologist*, 220(4), 1108-1115.
- Buee, M., Rossignol, M., Jauneau, A., Ranjeva, R., & Bécard, G. (2000). The pre-symbiotic growth of arbuscular mycorrhizal fungi is induced by a branching factor partially purified from plant root exudates. *Molecular Plant-Microbe Interactions*, 13(6), 693-698.
- Chandra, P., Singh, A., Prajapat, K., Rai, A. K., & Yadav, R. K. (2022). Native arbuscular mycorrhizal fungi improve growth, biomass yield, and phosphorus nutrition of sorghum in saline and sodic soils of the semi-arid region. *Environmental and Experimental Botany*, 201, 104982.
- Chandrasekaran, M., Kim, K., Krishnamoorthy, R., Walitang, D., Sundaram, S., Joe, M. M., Selvakumar, G., Hu, S., Oh, S.-H., & Sa, T. (2016). Mycorrhizal symbiotic efficiency on C3 and C4 plants under salinity stress – A Meta-Analysis. *Frontiers in Microbiology*, 7.
- Chen, D., Lu, Y., Li, Z., & Young, S. (2022). Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Computers and Electronics in Agriculture*, 198, 107091.
- Chen, E. C. H., Morin, E., Beaudet, D., Noel, J., Yildirim, G., Ndikumana, S., Charron, P., St-Onge, C., Giorgi, J., Krüger, M., Marton, T., Ropars, J., Grigoriev, I. V., Hainaut, M., Henrissat, B., Roux, C., Martin, F., & Corradi, N. (2018). High intraspecific genome diversity in the

- model arbuscular mycorrhizal symbiont *Rhizophagus irregularis*. *New Phytologist*, 220(4), 1161-1171.
- Chen, H., Covert, I. C., Lundberg, S. M., & Lee, S.-I. (2023). Algorithms to estimate Shapley value feature attributions. *Nature Machine Intelligence*, 5(6), 590-601.
- Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention Mask Transformer for Universal Image Segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Choi, J., Summers, W., & Paszkowski, U. (2018). Mechanisms underlying establishment of arbuscular mycorrhizal symbioses. *Annual Review of Phytopathology*, 56, 135-160.
- Chopra, R., Burow, G., Burke, J. J., Gladman, N., & Xin, Z. (2017). Genome-wide association analysis of seedling traits in diverse Sorghum germplasm under thermal stress. *BMC Plant Biology*, 17(1), 12.
- Cobb, A. B., Wilson, G. W. T., Goad, C. L., Bean, S. R., Kaufman, R. C., Herald, T. J., & Wilson, J. D. (2016). The role of arbuscular mycorrhizal fungi in grain production and nutrition of sorghum genotypes: Enhancing sustainability through plant-microbial partnership. *Agriculture, Ecosystems & Environment*, 233, 432-440.
- Cobb, A. B., Wilson, G. W. T., Goad, C. L., Bean, S. R., Tesso, T. T., & Wilson, J. D. (2017). Assessing the influence of farm fertility amendments, field management, and sorghum genotypes on soil microbial communities and grain quality. *Applied Soil Ecology*, 119, 367-374.
- Cockerton, H. M., Li, B., Stavridou, E., Johnson, A., Karlström, A., Armitage, A. D., Martinez-Crucis, A., Galiano-Arjona, L., Harrison, N., Barber-Pérez, N., Cobo-Medina, M., & Harrison, R. J. (2020). Genetic and phenotypic associations between root architecture,

- arbuscular mycorrhizal fungi colonisation and low phosphate tolerance in strawberry (*Fragaria × ananassa*). *BMC Plant Biology*, 20(1), 154.
- Cornell, C., Kokkoris, V., Turcu, B., Dettman, J., Stefani, F., & Corradi, N. (2022). The arbuscular mycorrhizal fungus *Rhizophagus irregularis* harmonizes nuclear dynamics in the presence of distinct abiotic factors. *Fungal Genetics and Biology*, 158, 103639.
- Corradi, N., Antunes, P.M. and Magurno, F. (2025), A call for reform: implementing genome-based approaches for species classification in Glomeromycotina. *New Phytologist*, 247: 50-54.
- Das, A., Schneider, H., BurrIDGE, J., Ascanio, A. K. M., Wojciechowski, T., Topp, C. N., Lynch, J. P., Weitz, J. S., & Bucksch, A. (2015). Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics. *Plant methods*, 11(1), 1-12.
- Davière, J.-M., & Achard, P. (2013). Gibberellin signaling in plants. *Development*, 140(6), 1147-1151.
- Davison, J., Moora, M., Öpik, M., Adholeya, A., Ainsaar, L., Bâ, A., Burla, S., Diedhiou, A. G., Hiiesalu, I., Jairus, T., Johnson, N. C., Kane, A., Koorem, K., Kochar, M., Ndiaye, C., Pärtel, M., Reier, Ü., Saks, Ü., Singh, R., Zobel, M. (2015). Global assessment of arbuscular mycorrhizal fungus diversity reveals very low endemism [Journal Article]. *Science (Washington)*, 349(6251), 970-973.
- De Boor, C. (2001). A practical guide to splines: with 32 figures (Rev. ed.). *Springer*.
- De Vita, P., Avio, L., Sbrana, C., Laidò, G., Marone, D., Mastrangelo, A. M., Cattivelli, L., & Giovannetti, M. (2018a). Genetic markers associated to arbuscular mycorrhizal colonization in durum wheat. *Scientific Reports*, 8(1), 10612.

- De Vita, P., Avio, L., Sbrana, C., Laidò, G., Marone, D., Mastrangelo, A. M., Cattivelli, L., & Giovannetti, M. (2018b). Genetic markers associated to arbuscular mycorrhizal colonization in durum wheat. *Scientific Reports*, 8(1), 1-12.
- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for machine learning / Marc Peter Deisenroth, University College London ; A. Aldo Faisal, Imperial College London ; Cheng Soon Ong, Data61, CSIRO. In: Cambridge University Press.
- Delaux, P.-M., Radhakrishnan, G. V., Jayaraman, D., Cheema, J., Malbreil, M., Volkening, J. D., Sekimoto, H., Nishiyama, T., Melkonian, M., Pokorny, L., Rothfels, C. J., Sederoff, H. W., Stevenson, D. W., Surek, B., Zhang, Y., Sussman, M. R., Dunand, C., Morris, R. J., Roux, C., Ané, J.-M. (2015). Algal ancestor of land plants was preadapted for symbiosis. *Proceedings of the National Academy of Sciences*, 112(43), 13390-13395.
- Deng, J., Dong, W., Socher, R., Li, L. J., Kai, L., & Li, F.-F. (2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition,
- Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *The ISME Journal*, 15(11), 3181-3194.
- Dhawi, F., Datta, R., & Ramakrishna, W. (2016). Mycorrhiza and heavy metal resistant bacteria enhance growth, nutrient uptake and alter metabolic profile of sorghum grown in marginal soil. *Chemosphere*, 157, 33-41.
- Dodd, J. C., Boddington, C. L., Rodriguez, A., Gonzalez-Chavez, C., & Mansur, I. (2000). Mycelium of Arbuscular Mycorrhizal fungi (AMF) from different genera: form, function and detection. *Plant and Soil*, 226(2), 131-151.

- Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, 40, 100379.
- Dutta, A., & Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. *Proceedings of the 27th Acm International Conference on Multimedia (Mm'19)*, 2276-2279.
- Egboka, N., Nnamdi, I., Chukwu, E., Okoli, N., & Afangide, A. (2022). Impact of land use types on spore abundance of arbuscular mycorrhizal fungi in the humid tropical rainforest, southeast nigeria. *Nigeria Agricultural Journal*, 53(1), 242-247.
- Ellouze, W., Hamel, C., Bouzid, S., & St-Arnaud, M. (2015). Root endophytes modify the negative effects of chickpea on the emergence of durum wheat. *Applied Soil Ecology*, 96, 201-210.
- Enciso, J., Jifon, J., Ribera, L., Zapata, S., & Ganjegunte, G. (2015). Yield, water use efficiency and economic analysis of energy sorghum in South Texas. *Biomass and Bioenergy*, 81, 339-344.
- Evangelisti, E., Turner, C., McDowell, A., Shenhav, L., Yunusov, T., Gavrin, A., Servante, E. K., Quan, C., & Schornack, S. (2021). Deep learning-based quantification of arbuscular mycorrhizal fungi in plant roots. *New Phytologist*, 232(5), 2207-2219.
- Fang, F., Liang, W., Cheng, Y., Xu, Q., & Lim, J. H. (2024). Enhancing representation learning with spatial transformation and early convolution for reinforcement learning-based small object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(1), 315-328.
- Feddermann, N., Finlay, R., Boller, T., & Elfstrand, M. (2010). Functional diversity in arbuscular mycorrhiza – the role of gene expression, phosphorous nutrition and symbiotic efficiency. *Fungal Ecology*, 3(1), 1-8.

- Ferguson, J. N., Fernandes, S. B., Monier, B., Miller, N. D., Allen, D., Dmitrieva, A., Schmuker, P., Lozano, R., Valluru, R., Buckler, E. S., Gore, M. A., Brown, P. J., Spalding, E. P., & Leakey, A. D. B. (2021). Machine learning-enabled phenotyping for GWAS and TWAS of WUE traits in 869 field-grown sorghum accessions. *Plant Physiology*, 187(3), 1481-1500.
- Field, K. J., Cameron, D. D., Leake, J. R., Tille, S., Bidartondo, M. I., & Beerling, D. J. (2012a). Contrasting arbuscular mycorrhizal responses of vascular and non-vascular plants to a simulated Palaeozoic CO₂ decline. *Nature Communications*, 3(1), 1-8.
- Field, K. J., Cameron, D. D., Leake, J. R., Tille, S., Bidartondo, M. I., & Beerling, D. J. (2012b). Contrasting arbuscular mycorrhizal responses of vascular and non-vascular plants to a simulated Palaeozoic CO₂ decline. *Nature Communications*, 3(1), 835.
- Figueiredo de Oliveira, I., Ferreira Simeone, M. L., Gomes de Paula Lana, U., de Carvalho Guimarães, C., & Morais de Sousa Tinôco, S. (2025). Enhancing sorghum growth: influence of arbuscular mycorrhizal fungi and sorgoleone. *Microorganisms*, 13(2).
- Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.
- Floss, D. S., Levy, J. G., Lévesque-Tremblay, V., Pumplin, N., & Harrison, M. J. (2013). DELLA proteins regulate arbuscule formation in arbuscular mycorrhizal symbiosis. *Proceedings of the National Academy of Sciences*, 110(51), E5025-E5034.
- Frew, A. (2019). Arbuscular mycorrhizal fungal diversity increases growth and phosphorus uptake in C₃ and C₄ crop plants. *Soil Biology and Biochemistry*, 135, 248-250.
- Friese, C. F., & Allen, M. F. (1991). The spread of VA mycorrhizal fungal hyphae in the soil: inoculum types and external hyphal architecture. *Mycologia*, 83(4), 409-418.

- Füzy, A., Biró, I., Kovács, R., & Takács, T. (2015). Estimation of AM fungal colonization—Comparability and reliability of classical methods. *Acta Microbiologica et Immunologica Hungarica*, 62(4), 435-451.
- Gallego-Bartolomé, J., Minguet, E. G., Grau-Enguix, F., Abbas, M., Locascio, A., Thomas, S. G., Alabadí, D., & Blázquez, M. A. (2012). Molecular mechanism for the interaction between gibberellin and brassinosteroid signaling pathways in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 109(33), 13446-13451.
- Gao, C., Montoya, L., Xu, L., Madera, M., Hollingsworth, J., Purdom, E., Singan, V., Vogel, J., Hutmacher, R. B., Dahlberg, J. A., Coleman-Derr, D., Lemaux, P. G., & Taylor, J. W. (2020). Fungal community assembly in drought-stressed sorghum shows stochasticity, selection, and universal ecological dynamics. *Nature Communications*, 11(1), 34.
- Gao, C., Xu, L., Montoya, L., Madera, M., Hollingsworth, J., Chen, L., Purdom, E., Singan, V., Vogel, J., Hutmacher, R. B., Dahlberg, J. A., Coleman-Derr, D., Lemaux, P. G., & Taylor, J. W. (2022). Co-occurrence networks reveal more complexity than community composition in resistance and resilience of microbial communities. *Nature Communications*, 13(1), 3867.
- Gardner, J., Maranville, J., & Paparozzi, E. (1994). Nitrogen use efficiency among diverse sorghum cultivars. *Crop Science*, 34(3), 728-733.
- Genre, A., Chabaud, M., Faccio, A., Barker, D. G., & Bonfante, P. (2008). Prepenetration apparatus assembly precedes and predicts the colonization patterns of arbuscular mycorrhizal fungi within the root cortex of both *Medicago truncatula* and *Daucus carota*. *The Plant Cell*, 20(5), 1407-1420.

- Genre, A., Lanfranco, L., Perotto, S., & Bonfante, P. (2020). Unique and common traits in mycorrhizal symbioses. *Nature Reviews Microbiology*, 18(11), 649-660.
- Gerdemann, J. W. (1965). Vesicular-arbuscular mycorrhizae formed on Maize and Tuliptree by *Endogone fasciculata*. *Mycologia*, 57(4), 562-575.
- Gillies, S., & others. (2007). Shapely: manipulation and analysis of geometric objects. <https://github.com/shapely/shapely>
- Gladman, N., Olson, A., Wei, S., Chougule, K., Lu, Z., Tello-Ruiz, M., Meijjs, I., Van Buren, P., Jiao, Y., & Wang, B. (2022). SorghumBase: a web-based portal for sorghum genetic information and community advancement. *Planta*, 255(2), 35.
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., & Putnam, N. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic acids research*, 40(D1), D1178-D1186.
- Govindarajulu, R., Hostetler, A. N., Xiao, Y., Chaluvadi, S. R., Mauro-Herrera, M., Siddoway, M. L., Whipple, C., Bennetzen, J. L., Devos, K. M., Doust, A. N., & Hawkins, J. S. (2021). Integration of high-density genetic mapping with transcriptome analysis uncovers numerous agronomic QTL and reveals candidate genes for the control of tillering in sorghum. *G3: Genes, Genomes, Genetics*, 11(2), jkab024.
- Graham, J. H., & Abbott, L. K. (2000). Wheat responses to aggressive and non-aggressive arbuscular mycorrhizal fungi. *Plant and Soil*, 220(1), 207-218.
- Groll, A., & Tutz, G. (2014). Variable selection for generalized linear mixed models by L1-penalized estimation. *Statistics and Computing*, 24(2), 137-154.
- Hart, M. M., & Reader, R. J. (2002). Does percent root length colonization and soil hyphal length reflect the extent of colonization for all AMF? *Mycorrhiza*, 12(6), 297-301.

- Hawkins, H. J., Cargill, R. I. M., Van Nuland, M. E., Hagen, S. C., Field, K. J., Sheldrake, M., Soudzilovskaia, N. A., & Kiers, E. T. (2023). Mycorrhizal mycelium as a global carbon pool. *Curr Biol*, 33(11), R560-r573.
- He, K., Gkioxari, G., Doll'ar, P., & Girshick, R. (2018). Mask R-CNN. *arXiv pre-print server*. arxiv:1703.06870
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV),
- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2020). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 386-397.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition,
- He, K. M., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2980-2988.
- Hetrick, B. A. D., Leslie, J. F., Wilson, G. T., & Kitt, D. G. (1988). Physical and topological assessment of effects of a vesicular–arbuscular mycorrhizal fungus on root architecture of big bluestem. *New Phytologist*, 110(1), 85-96.
- Hetrick, B. A. D., Wilson, G. W. T., & Cox, T. S. (1992). Mycorrhizal dependence of modern wheat varieties, landraces, and ancestors. *Canadian Journal of Botany*, 70(10), 2032-2040.
- Hoeksema, J. D., Bever, J. D., Chakraborty, S., Chaudhary, V. B., Gardes, M., Gehring, C. A., Hart, M. M., Housworth, E. A., Kaonongbua, W., Klironomos, J. N., Lajeunesse, M. J., Meadow, J., Milligan, B. G., Piculell, B. J., Pringle, A., Rúa, M. A., Umbanhowar, J., Viechtbauer, W., Wang, Y.-W., Zee, P. C. (2018). Evolutionary history of plant hosts and fungal

- symbionts predicts the strength of mycorrhizal mutualism. *Communications Biology*, 1(1), 116.
- Hossen, M. I., Awrangjeb, M., Pan, S., & Mamun, A. A. (2025). Transfer learning in agriculture: a review. *Artificial Intelligence Review*, 58(4), 97.
- Hu, Z., Olatoye, M. O., Marla, S., & Morris, G. P. (2019). An integrated genotyping-by-sequencing polymorphism map for over 10,000 sorghum genotypes. *The Plant Genome*, 12(1), 180044.
- Humphreys, C. P., Franks, P. J., Rees, M., Bidartondo, M. I., Leake, J. R., & Beerling, D. J. (2010). Mutualistic mycorrhiza-like symbiosis in the most ancient group of land plants. *Nature Communications*, 1(1), 103.
- Intelligence, M. (2024). Mycorrhiza biofertilizers market size & share analysis – growth trends & forecasts (2024–2029). <https://www.mordorintelligence.com/industry-reports/mycorrhiza-based-biofertilizer-market>.
- Iqra, Giri, K. J., & Javed, M. (2024). Small object detection in diverse application landscapes: a survey. *Multimedia Tools and Applications*, 83(41), 88645-88680.
- Ivanov, S., Austin, J., Berg, R. H., & Harrison, M. J. (2019). Extensive membrane systems at the host–arbuscular mycorrhizal fungus interface. *Nature Plants*, 5(2), 194-203.
- Ivanov, S., & Harrison, M. J. (2014). A set of fluorescent protein-based markers expressed from constitutive and arbuscular mycorrhiza-inducible promoters to label organelles, membranes and cytoskeletal elements in *Medicago truncatula*. *The Plant Journal*, 80(6), 1151-1163.

- Jacquelinet-Jeanmougin, S., & Gianinazzi-Pearson, V. (1983). Endomycorrhizas in the Gentianaceae. I. The Fungi Associated with *Gentiana lutea* L. *New Phytologist*, 95(4), 663-666.
- James, T. Y., Kauff, F., Schoch, C. L., Matheny, P. B., Hofstetter, V., Cox, C. J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., Lumbsch, H. T., Rauhut, A., Reeb, V., Arnold, A. E., Amtoft, A., Stajich, J. E., Hosaka, K., Sung, G.-H., Johnson, D., Vilgalys, R. (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature*, 443(7113), 818-822.
- Jiang, C., Gao, X., Liao, L., Harberd, N. P., & Fu, X. (2007). Phosphate Starvation Root Architecture and Anthocyanin Accumulation Responses Are Modulated by the Gibberellin-DELLA Signaling Pathway in Arabidopsis. *Plant Physiology*, 145(4), 1460-1470.
- Jiang, L., Yuan, B., Du, J., Chen, B., Xie, H., Tian, J., & Yuan, Z. (2024). MFFSODNet: Multiscale Feature Fusion Small Object Detection Network for UAV Aerial Images. *IEEE Transactions on Instrumentation and Measurement*, 73, 1-14.
- Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A Review. *Plant Phenomics*, 2020, 4152816.
- Johnson, A. C., Pendergast, T. H., Chaluvadi, S., Bennetzen, J. L., & Devos, K. M. (2022). Identification of microRNAs responsive to arbuscular mycorrhizal fungi in *Panicum virgatum* (switchgrass). *BMC Genomics*, 23(1), 688.
- Johnson, N. C., Graham, J.-H., & Smith, F. A. (1997). Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *New Phytologist*, 135(4), 575-585.
- Johnson, N. C., Hoeksema, J. D., Bever, J. D., Chaudhary, V. B., Gehring, C., Klironomos, J., Koide, R., Miller, R. M., Moore, J., Moutoglou, P., Schwartz, M., Simard, S., Swenson, W.,

- Umbanhowar, J., Wilson, G., & Zabinski, C. (2006). From Lilliput to Brobdingnag: extending models of mycorrhizal function across scales. *BioScience*, 56(11), 889-900.
- Johnson, N. C., & Marín, C. (2024). Functional team selection: a framework for local adaptation in plants and their belowground microbiomes. *EcoEvoRxiv*.
- Johnson, N. C., Rowland, D. L., Corkidi, L., Egerton-Warburton, L. M., & Allen, E. B. (2003). Nitrogen enrichment alters mycorrhizal allocation at five mesic to semiarid grasslands. *Ecology*, 84(7), 1895-1908.
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944-946.
- Juan, M., Basile, C., Min, Y., John, L., & Lennart, R. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology: linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology and Community Health*, 59(6), 443.
- Juan, M., Min, Y., Basile, C., John, L., & Lennart, R. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*, 59(9), 729.
- Kabir, Z., O'Halloran, I. P., Widden, P., & Hamel, C. (1998). Vertical distribution of arbuscular mycorrhizal fungi under corn (*Zea mays* L.) in no-till and conventional tillage systems. *Mycorrhiza*, 8(1), 53-55.
- Kaeppeler, S. M., Parke, J. L., Mueller, S. M., Senior, L., Stuber, C., & Tracy, W. F. (2000). Variation among maize inbred lines and detection of quantitative trait loci for growth at low phosphorus and responsiveness to arbuscular mycorrhizal fungi. *Crop Science*, 40(2), 358-364.

- Kakouridis, A., Hagen, J. A., Kan, M. P., Mambelli, S., Feldman, L. J., Herman, D. J., Weber, P. K., Pett-Ridge, J., & Firestone, M. K. (2022). Routes to roots: direct evidence of water transport by arbuscular mycorrhizal fungi to host plants. *New Phytologist*, 236(1), 210-221.
- Kaur, S., Campbell, B. J., & Suseela, V. (2022). Root metabolome of plant–arbuscular mycorrhizal symbiosis mirrors the mutualistic or parasitic mycorrhizal phenotype. *New Phytologist*, 234(2), 672-687.
- Khaki, S., Pham, H., & Wang, L. (2021). Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Scientific Reports*, 11(1), 11132.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). Segment Anything. 2023 IEEE/CVF International Conference on Computer Vision (ICCV).
- Kivlin, S. N., Hawkes, C. V., & Treseder, K. K. (2011). Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil Biology and Biochemistry*, 43(11), 2294-2303.
- Klein, M., Stewart, J. D., Porter, S. S., Weedon, J. T., & Kiers, E. T. (2022). Evolution of manipulative microbial behaviors in the rhizosphere. *Evolutionary Applications*, 15(10), 1521-1536.
- Klironomos, & John. (1999). Host-specificity and functional diversity among arbuscular mycorrhizal fungi.
- Kobae, Y., & Hata, S. (2010). Dynamics of periarbuscular membranes visualized with a fluorescent phosphate transporter in arbuscular mycorrhizal roots of rice. *Plant and Cell Physiology*, 51(3), 341-353.

- Koch, A. M., Antunes, P. M., Maherali, H., Hart, M. M., & Klironomos, J. N. (2017). Evolutionary asymmetry in the arbuscular mycorrhizal symbiosis: conservatism in fungal morphology does not predict host plant growth. *New Phytologist*, 214(3), 1330-1337.
- Koch, A. M., Croll, D., & Sanders, I. R. (2006). Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology Letters*, 9(2), 103-110.
- Koegel, S., Ait Lahmidi, N., Arnould, C., Chatagnier, O., Walder, F., Ineichen, K., Boller, T., Wipf, D., Wiemken, A., & Courty, P.-E. (2013). The family of ammonium transporters (AMT) in *orghum bicolor*: two AMT members are induced locally, but not systemically in roots colonized by arbuscular mycorrhizal fungi. *New Phytologist*, 198(3), 853-865.
- Kokkoris, V., Pogiatis, A., & Hart, M. M. (2019). Contrasting common measures of arbuscular mycorrhizal fungal root colonization. *Journal of Microbiological Methods*, 167, 105727.
- Kokkoris, V., Stefani, F., Dalpé, Y., Dettman, J., & Corradi, N. (2020). Nuclear dynamics in the arbuscular mycorrhizal fungi. *Trends in Plant Science*, 25(8), 765-778.
- Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Dueñas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, 22(21), 5271-5277.
- Kong, F., & Henao, R. (2022, 18-24 June 2022). Efficient Classification of Very Large Images with Tiny Objects. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- Koziol, L., McKenna, T. P., & Bever, J. D. (2024). Meta-analysis reveals globally sourced commercial mycorrhizal inoculants fall short. *New Phytologist*.

- Krach, E. K., Skaro, M., Wu, Y., & Arnold, J. (2022). Characterizing the gene-environment interaction underlying natural morphological variation in *Neurospora crassa* conidiophores using high-throughput phenomics and transcriptomics. *G3: Genes, Genomes, Genetics (Bethesda)*, 12(4).
- Krach, E. K., Wu, Y., Skaro, M., Mao, L., & Arnold, J. (2020). Wild isolates of *Neurospora crassa* reveal three conidiophore architectural phenotypes. *Microorganisms*, 8(11).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6), 84–90.
- Kumar, N., Boatwright, J. L., Boyles, R. E., Brenton, Z. W., & Kresovich, S. (2024). Identification of pleiotropic loci mediating structural and non-structural carbohydrate accumulation within the sorghum bioenergy association panel using high-throughput markers. *Frontiers in Plant Science*, 15.
- Kumar, N., Boatwright, J. L., Sapkota, S., Brenton, Z. W., Ballén-Taborda, C., Myers, M. T., Cox, W. A., Jordan, K. E., Kresovich, S., & Boyles, R. E. (2023). Discovering useful genetic variation in the seed parent gene pool for sorghum improvement. *Frontiers in Genetics*, 14, 1221148.
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Lehnert, H., Serfling, A., Enders, M., Friedt, W., & Ordon, F. (2017). Genetics of mycorrhizal symbiosis in winter wheat (*Triticum aestivum*). *New Phytologist*, 215(2), 779-791.
- Li, J., Tang, W., Zhang, Y.-W., Chen, K.-N., Wang, C., Liu, Y., Zhan, Q., Wang, C., Wang, S.-B., & Xie, S.-Q. (2018). Genome-wide association studies for five forage quality-related traits in sorghum (*Sorghum bicolor* L.). *Frontiers in Plant Science*, 9, 1146.

- Lin, T.-Y., Maire, M., Belongie, S. J., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision,
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *Computer Vision - Eccc 2014, Pt V*, 8693, 740-755.
- Liu, H., Wu, Y., Xu, H., Ai, Z., Zhang, J., Liu, G., & Xue, S. (2021). Mechanistic understanding of interspecific interaction between a C4 grass and a C3 legume via arbuscular mycorrhizal fungi, as influenced by soil phosphorus availability using a ¹³C and ¹⁵N dual-labelled organic patch. *The Plant Journal*, 108(1), 183-196.
- Lutz, S., Bodenhausen, N., Hess, J., Valzano-Held, A., Waelchli, J., Deslandes-Hérolde, G., Schlaeppli, K., & van der Heijden, M. G. A. (2023). Soil microbiome indicators can predict crop growth response to large-scale inoculation with arbuscular mycorrhizal fungi. *Nature Microbiology*, 8(12), 2277-2289.
- Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., Hassall, M., Hathorn, A., & Jordan, D. (2019). The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theoretical and Applied Genetics*, 132, 751-766.
- Mace, E. S., Hunt, C. H., & Jordan, D. R. (2013). Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theoretical and Applied Genetics*, 126(5), 1377-1395.
- MacLean, A. M., Bravo, A., & Harrison, M. J. (2017). Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *The Plant Cell*, 29(10), 2319-2335.

- Marleau, J., Dalpé, Y., St-Arnaud, M., & Hijri, M. (2011). Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. *BMC evolutionary biology*, *11*(1), 1-11.
- Martignoni, M. M., Garnier, J., Zhang, X., Rosa, D., Kokkoris, V., Tyson, R. C., & Hart, M. M. (2021). Co-inoculation with arbuscular mycorrhizal fungi differing in carbon sink strength induces a synergistic effect in plant growth. *Journal of Theoretical Biology*, *531*, 110859.
- Martin, F. M., & van der Heijden, M. G. A. (2024). The mycorrhizal symbiosis: research frontiers in genomics, ecology, and agricultural application. *New Phytologist*, *242*(4), 1486-1506.
- Mathieu, S., Cusant, L., Roux, C., & Corradi, N. (2018). Arbuscular mycorrhizal fungi: intraspecific diversity and pangenomes. *New Phytologist*, *220*(4), 1129-1134.
- McGonigle, T. P., Miller, M. H., Evans, D. G., Fairchild, G. L., & Swan, J. A. (1990a). A new method which gives an objective measure of colonization of roots by vesicular-arbuscular mycorrhizal fungi. *New Phytologist*, *115*(3), 495-501.
- McGonigle, T. P., Miller, M. H., Evans, D. G., Fairchild, G. L., & Swan, J. A. (1990b). A new method which gives an objective measure of colonization of roots by vesicular—arbuscular mycorrhizal fungi. *New Phytologist*, *115*(3), 495-501.
- Mensah, J. A., Koch, A. M., Antunes, P. M., Kiers, E. T., Hart, M., & Bücking, H. (2015). High functional diversity within species of arbuscular mycorrhizal fungi is associated with differences in phosphate and nitrogen uptake and fungal phosphate metabolism. *Mycorrhiza*, *25*(7), 533-546.
- Merlo, J., Yang, M., Chaix, B., Lynch, J., & Råstam, L. (2005). A brief conceptual tutorial on multilevel analysis in social epidemiology: investigating contextual phenomena in different groups of people. *J Epidemiol Community Health*, *59*(9), 729-736.

- Montero, H., Choi, J., & Paszkowski, U. (2019). Arbuscular mycorrhizal phenotyping: the dos and don'ts. *New Phytologist*, 221(3), 1182-1186.
- Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., Riera-Lizarazu, O., Brown, P. J., Acharya, C. B., Mitchell, S. E., Harriman, J., Glaubitz, J. C., Buckler, E. S., & Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proceedings of the National Academy of Sciences*, 110(2), 453-458.
- Moura, J. B. d., Ramos, M. L. G., Konrad, M. L. d. F., Saggin Júnior, O. J., Ribeiro Junior, W. Q., Carvalho, A. M. d., Santos, L. d. M., Souza, R. F. d., Lopes Filho, L. C., & Dutra e Silva, S. (2022). Arbuscular mycorrhizal fungi in the soil using cover crops with and without nitrogen addition. *Frontiers in Plant Science*, 13.
- Mullet, J., Morishige, D., McCormick, R., Truong, S., Hilley, J., McKinley, B., Anderson, R., Olson, S. N., & Rooney, W. (2014). Energy Sorghum—a genetic model for the design of C4 grass bioenergy crops. *Journal of Experimental Botany*, 65(13), 3479-3489.
- Munkvold, L., Kjølner, R., Vestberg, M., Rosendahl, S., & Jakobsen, I. (2004). High functional diversity within species of arbuscular mycorrhizal fungi. *New Phytologist*, 164(2), 357-364.
- Murray, S. C., Rooney, W. L., Hamblin, M. T., Mitchell, S. E., & Kresovich, S. (2009). Sweet sorghum genetic diversity and association mapping for brix and height. *The Plant Genome*, 2(1).
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel.

- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133-142.
- Nakmee, P. S., Techapinyawat, S., & Ngamprasit, S. (2016). Comparative potentials of native arbuscular mycorrhizal fungi to improve nutrient uptake and biomass of *Sorghum bicolor* Linn. *Agriculture and Natural Resources*, 50(3), 173-178.
- Naus, J. I. (1975). Data quality control and editing. Joseph I. Naus. In: M. Dekker.
- Naus, J. I. (2015). Editing statistical data. In *Wiley StatsRef: Statistics Reference Online* (pp. 1-10).
- Nazir, M. I., Ullah, M. O., & Akter, A. (2023). An efficient deep transfer learning based apple leaf disease classification. 2023 26th International Conference on Computer and Information Technology (ICCIT).
- Nigam, D., Devkar, V., Dhiman, P., Shakoor, S., Liu, D., Patil, G. B., & Jiao, Y. (2025). Emerging frontiers in sorghum genetic engineering. *The Plant Journal*, 121(4), e17244.
- Nogueira, F. (2014). Bayesian Optimization: Open source constrained global optimization tool for Python. <https://github.com/bayesian-optimization/BayesianOptimization>
- Noor, M. H. M., & Ige, A. O. (2024). A Survey on State-of-the-art Deep Learning Applications and Challenges. *arXiv:2403.17561*.
- Öpik, M., Zobel, M., Cantero, J. J., Davison, J., Facelli, J. M., Hiiesalu, I., Jairus, T., Kalwij, J. M., Koorem, K., Leal, M. E., Liira, J., Metsis, M., Neshataeva, V., Paal, J., Phosri, C., Pölm, S., Reier, Ü., Saks, Ü., Schimann, H., . . . Moora, M. (2013). Global sampling of plant roots expands the described molecular diversity of arbuscular mycorrhizal fungi. *Mycorrhiza*, 23(5), 411-430.
- Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.

- Oyarte Galvez, L., Bisot, C., Bourrienne, P., Cargill, R., Klein, M., van Son, M., van Krugten, J., Caldas, V., Clerc, T., Lin, K.-K., Kahane, F., van Staalduine, S., Stewart, J. D., Terry, V., Turcu, B., van Otterdijk, S., Babu, A., Kamp, M., Seynen, M., . . . Shimizu, T. S. (2025). A travelling-wave strategy for plant–fungal trade. *Nature*, 639(8053), 172-180.
- Pace, R., Schiano Di Cola, V., Monti, M. M., Affinito, A., Cuomo, S., Loreto, F., & Ruocco, M. (2025). Artificial intelligence in soil microbiome analysis: a potential application in predicting and enhancing soil health—a review. *Discover Applied Sciences*, 7(2), 85.
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A. K., Chapman, J., Feltus, F. A., Gowik, U., Rokhsar, D. S. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, 457(7229), 551-556.
- Pawlowski, M. L., Vuong, T. D., Valliyodan, B., Nguyen, H. T., & Hartman, G. L. (2020). Whole-genome resequencing identifies quantitative trait loci associated with mycorrhizal colonization of soybean. *Theor Appl Genet*, 133(2), 409-417.
- Pawlowski, M. L., Vuong, T. D., Valliyodan, B., Nguyen, H. T., & Hartman, G. L. (2020). Whole-genome resequencing identifies quantitative trait loci associated with mycorrhizal colonization of soybean. *Theoretical and Applied Genetics*, 133, 409-417.
- Phillips, J. M., & Hayman, D. S. (1970). Improved procedures for clearing roots and staining parasitic and vesicular-arbuscular mycorrhizal fungi for rapid assessment of infection. *Transactions of the British Mycological Society*, 55(1), 158-IN118.
- Plouznikoff, K., Asins, M. J., de Boulois, H. D., Carbonell, E. A., & Declerck, S. (2019). Genetic analysis of tomato root colonization by arbuscular mycorrhizal fungi. *Annals of Botany*, 124(6), 933-946.

- Pozo, M. J., & Azcón-Aguilar, C. (2007). Unraveling mycorrhiza-induced resistance. *Current Opinion in Plant Biology*, 10(4), 393-398.
- Pumplin, N., & Harrison, M. J. (2009). Live-cell imaging reveals periarbuscular membrane domains and organelle location in *Medicago truncatula* roots during arbuscular mycorrhizal symbiosis. *Plant Physiology*, 151(2), 809-819.
- Putri, D. A. L. P., Widyastuti, R., Idris, I., Ikhwan, A. Z. N., Nugroho, S., Sudiana, I. M., Kanti, A., Purnaningsih, I., Ochiai, K., Kobayashi, M., & Wijaya, L. (2023). Unraveling the mechanisms of drought tolerance enhancement in *Sorghum bicolor* through *Glomus mosseae* inoculation: Insights from comparative analysis of Super 2 and Konawe Selatan accessions. *South African Journal of Botany*, 161, 293-304.
- Raina, R., Battle, A., Lee, H., Packer, B., & Ng, A. Y. (2007). Self-taught learning: transfer learning from unlabeled data. Proceedings of the 24th international conference on Machine learning, Corvallis, Oregon, USA.
- Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39.
- Rhodes, D. H., Hoffmann Jr, L., Rooney, W. L., Ramu, P., Morris, G. P., & Kresovich, S. (2014). Genome-wide association study of grain polyphenol concentrations in global sorghum [*Sorghum bicolor* (L.) Moench] germplasm. *Journal of agricultural and food chemistry*, 62(45), 10916-10927.
- Riaz, M., Kamran, M., Fang, Y., Wang, Q., Cao, H., Yang, G., Deng, L., Wang, Y., Zhou, Y., Anastopoulos, I., & Wang, X. (2021). Arbuscular mycorrhizal fungi-induced mitigation of

- heavy metal phytotoxicity in metal contaminated soils: A critical review. *Journal of Hazardous Materials*, 402, 123919.
- Rillig, M. C., Aguilar-Trigueros, C. A., Camenzind, T., Cavagnaro, T. R., Degruene, F., Hohmann, P., Lammel, D. R., Mansour, I., Roy, J., van der Heijden, M. G. A., & Yang, G. (2019). Why farmers should manage the arbuscular mycorrhizal symbiosis. *New Phytologist*, 222(3), 1171-1175.
- Rosendahl, S. (2008). Communities, populations and individuals of arbuscular mycorrhizal fungi. *New Phytologist*, 178(2), 253-266.
- Ruiz-Lozano, J. M., Porcel, R., Azcon, C., & Aroca, R. (2012). Regulation by arbuscular mycorrhizae of the integrated physiological response to salinity in plants: new challenges in physiological and molecular studies. *Journal of Experimental Botany*, 63(11), 4033-4044.
- Ruiz-Lozano, J. M., Porcel, R., Azcón, C., & Aroca, R. (2012). Regulation by arbuscular mycorrhizae of the integrated physiological response to salinity in plants: new challenges in physiological and molecular studies. *J Exp Bot*, 63(11), 4033-4044.
- Säle, V., Palenzuela, J., Azcón-Aguilar, C., Sánchez-Castro, I., da Silva, G. A., Seitz, B., Sieverding, E., van der Heijden, M. G. A., & Oehl, F. (2021). Ancient lineages of arbuscular mycorrhizal fungi provide little plant benefit. *Mycorrhiza*, 31(5), 559-576.
- Sawers, R. J., Ramírez-Flores, M. R., Olalde-Portugal, V., & Paszkowski, U. (2018). The impact of domestication and crop improvement on arbuscular mycorrhizal symbiosis in cereals: insights from genetics and genomics. *New Phytologist*, 220(4), 1135-1140.
- Sawers, R. J. H., Svane, S. F., Quan, C., Grønlund, M., Wozniak, B., Gebreselassie, M.-N., González-Muñoz, E., Chávez Montes, R. A., Baxter, I., Goudet, J., Jakobsen, I., &

- Paszkowski, U. (2017). Phosphorus acquisition efficiency in arbuscular mycorrhizal maize is correlated with the abundance of root-external hyphae and the accumulation of transcripts encoding PHT1 phosphate transporters. *New Phytologist*, 214(2), 632-643.
- Schiellzeth, H., & Nakagawa, S. (2013). Nested by design: model fitting and interpretation in a mixed model era. *Methods in Ecology and Evolution*, 4(1), 14-24.
- Schultz, C. J., Kochian, L. V., & Harrison, M. J. (2010). Genetic variation for root architecture, nutrient uptake and mycorrhizal colonisation in *Medicago truncatula* accessions. *Plant and Soil*, 336(1), 113-128.
- Schüßler, A., Schwarzott, D., & Walker, C. (2001). A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycological Research*, 105(12), 1413-1421.
- Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components*. John Wiley & Sons.
- Shafi, S., & Parwani, A. V. (2023). Artificial intelligence in diagnostic pathology. *Diagn Pathol*, 18(1), 109.
- Silva, T. N., Thomas, J. B., Dahlberg, J., Rhee, S. Y., & Mortimer, J. C. (2021). Progress and challenges in sorghum biotechnology, a multipurpose feedstock for the bioeconomy. *Journal of Experimental Botany*, 73(3), 646-664.
- SMITH, F. A., & SMITH, S. E. (1997a). Structural diversity in (vesicular)–arbuscular mycorrhizal symbioses. *New Phytologist*, 137(3), 373-388.
- Smith, F. A., & Smith, S. E. (1997b). Tansley Review No. 96. Structural Diversity in (Vesicular)-Arbuscular Mycorrhizal Symbioses. *New Phytologist*, 137(3), 373-388.
- Smith, S. E., & Read, D. J. (1997). *Mycorrhizal Symbiosis* (2nd Edition ed.). Academic Press, New York.
- Smith, S. E., & Read, D. J. (2010). *Mycorrhizal symbiosis*. Academic press.

- Souza, V. F. d., Pereira, G. d. S., Pastina, M. M., Parrella, R. A. d. C., Simeone, M. L. F., Barros, B. d. A., Noda, R. W., da Costa E Silva, L., Magalhães, J. V. d., & Schaffert, R. E. (2021). QTL mapping for bioenergy traits in sweet sorghum recombinant inbred lines. *G3: Genes, Genomes, Genetics*, *11*(11), jkab314.
- Stahlhut, K. N., Dowell, J. A., Temme, A. A., Burke, J. M., Goolsby, E. W., & Mason, C. M. (2021). Genetic control of arbuscular mycorrhizal colonization by *Rhizophagus intraradices* in *Helianthus annuus* (L.). *Mycorrhiza*, *31*(6), 723-734.
- Subramanian, S. V., Jones, K., & Duncan, C. (2003). Multilevel methods for public health research. In I. Kawachi & L. F. Berkman (Eds.), *Neighborhoods and Health* (pp. 0). Oxford University Press.
- Symanczik, S., Krützmann, J., Nehls, U., Boller, T., & Courty, P.-E. (2020). Expression of major intrinsic protein genes in *Sorghum bicolor* roots under water deficit depends on arbuscular mycorrhizal fungal species. *Soil Biology and Biochemistry*, *140*, 107643.
- Taylor, A., Pereira, N., Thomas, B., Pink, D. A. C., Jones, J. E., & Bending, G. D. (2015). Growth and nutritional responses to arbuscular mycorrhizal fungi are dependent on onion genotype and fungal species. *Biology and Fertility of Soils*, *51*(7), 801-813.
- Taylor, J., & Harrier, L. (2000). A comparison of nine species of arbuscular mycorrhizal fungi on the development and nutrition of micropropagated *Rubus idaeus* L. cv. Glen Prosen (Red Raspberry). *Plant and Soil*, *225*(1), 53-61.
- Taylor, T. N., & Osborn, J. M. (1996). The importance of fungi in shaping the paleoecosystem. *Review of Palaeobotany and Palynology*, *90*(3), 249-262.
- Taylor, T. N., W., R., H., H., & and Kerp, H. (1995). Fossil arbuscular mycorrhizae from the Early Devonian. *Mycologia*, *87*(4), 560-573.

- Tedersoo, L., Bahram, M., Toots, M., Diédhiou, A. G., Henkel, T. W., Kjoller, R., Morris, M. H., Nara, K., Nouhra, E., Peay, K. G., Pölme, S., Ryberg, M., Smith, M. E., & Kõljalg, U. (2012). Towards global patterns in the diversity and community structure of ectomycorrhizal fungi. *Molecular Ecology*, 21(17), 4160-4170.
- Torres, I., Zhang, S., Bouffier, A., Skaro, M., Wu, Y., Stupp, L., Arnold, J., Chung, A., & Schuttler, H.-B. (2025). MINE: a new way to design genetics experiments for discovery. *Oxford University Press*.
- Trouvelot, A. (1986). Mesure du taux de mycorrhization d'un système racinaire. Recherche de méthodes d'estimation ayant une signification fonctionnelle. *Physiological and genetical aspects of mycorrhizae*, 217-221.
- Trouvelot, A., Kough, J., & Gianinazzi-Pearson, V. (1985). Measurement of the VA mycorrhization rate of a root system. Research and estimation methods of functional significance. *Physiological and genetic aspects of mycorrhizae, Dijon*, 217-221.
- Tubiello, F. N., Conchedda, G., Casse, L., Pengyu, H., Zhongxin, C., De Santis, G., Fritz, S., & Muchoney, D. (2023). Measuring the world's cropland area. *Nature Food*, 4(1), 30-32.
- Tuheteru, F. D., Husna, H., Albasri, A., Arif, A., Kramadibrata, K., & Soka, G. (2020). Composition and diversity of arbuscular mycorrhizal fungi spore associated with different land-use types in tropical gold mine. *Journal of Degraded and Mining Lands Management*.
- Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, 99-114.
- van der Heijden, M. G. A., Martin, F. M., Selosse, M.-A., & Sanders, I. R. (2015). Mycorrhizal ecology and evolution: the past, the present, and the future. *New Phytologist*, 205(4), 1406-1423.

- Varoquaux, N., Cole, B., Gao, C., Pierroz, G., Baker, C. R., Patel, D., Madera, M., Jeffers, T., Hollingsworth, J., Sievert, J., Yoshinaga, Y., Owiti, J. A., Singan, V. R., DeGraaf, S., Xu, L., Blow, M. J., Harrison, M. J., Visel, A., Jansson, C., Purdom, E. (2019). Transcriptomic analysis of field-droughted sorghum from seedling to maturity reveals biotic and metabolic responses. *Proceedings of the National Academy of Sciences*, 116(52), 27124-27132.
- Vierheilig, H., Coughlan, A. P., Wyss, U., & Piché, Y. (1998). Ink and Vinegar, a Simple Staining Technique for Arbuscular-Mycorrhizal Fungi. *Applied and Environmental Microbiology*, 64(12), 5004-5007.
- Vierheilig, H., & Piché, Y. (1998). A modified procedure for staining arbuscular mycorrhizal fungi in roots. *Zeitschrift für Pflanzenernährung und Bodenkunde*, 161(5), 601-602.
- Vosatka, M., & Dodd, J. C. (1998). The role of different arbuscular mycorrhizal fungi in the growth of *Calamagrostis villosa* and *Deschampsia flexuosa*, in experiments with simulated acid rain. *Plant and Soil*, 200(2), 251-263.
- Walder, F., Brulé, D., Koegel, S., Wiemken, A., Boller, T., & Courty, P.-E. (2015). Plant phosphorus acquisition in a common mycorrhizal network: regulation of phosphate transporter genes of the Pht1 family in sorghum and flax. *New Phytologist*, 205(4), 1632-1645.
- Wang, W., Shi, J., Xie, Q., Jiang, Y., Yu, N., & Wang, E. (2017). Nutrient exchange and regulation in arbuscular mycorrhizal symbiosis. *Molecular Plant*, 10(9), 1147-1158.
- Wang, X., Sui, X., Liu, Y., Xiang, L., Zhang, T., Fu, J., Li, A., & Yang, P. (2018). N-P fertilization did not reduce AMF abundance or diversity but alter AMF composition in an alpine grassland infested by a root hemiparasitic plant. *Plant Diversity*, 40(3), 117-126.

- Ware, D., Jaiswal, P., Ni, J., Pan, X., Chang, K., Clark, K., Teytelman, L., Schmidt, S., Zhao, W., & Cartinhour, S. (2002). Gramene: a resource for comparative grass genomics. *Nucleic acids research*, 30(1), 103-105.
- Watts-Williams, S. J., Emmett, B. D., Levesque-Tremblay, V., MacLean, A. M., Sun, X., Satterlee, J. W., Fei, Z., & Harrison, M. J. (2019). Diverse *Sorghum bicolor* accessions show marked variation in growth and transcriptional responses to arbuscular mycorrhizal fungi. *Plant, Cell & Environment*, 42(5), 1758-1774.
- Watts-Williams, S. J., Gill, A. R., Jewell, N., Brien, C. J., Berger, B., Tran, B. T., Mace, E., Cruickshank, A. W., Jordan, D. R., & Garnett, T. (2022). Enhancement of sorghum grain yield and nutrition: A role for arbuscular mycorrhizal fungi regardless of soil phosphorus availability. *Plants, People, Planet*, 4(2), 143-156.
- Weissmann, S., & Brutnell, T. P. (2012). Engineering C4 photosynthetic regulatory networks. *Current Opinion in Biotechnology*, 23(3), 298-304.
- Whiteside, M. D., Werner, G. D. A., Caldas, V. E. A., van't Padje, A., Dupin, S. E., Elbers, B., Bakker, M., Wyatt, G. A. K., Klein, M., & Hink, M. A. (2019). Mycorrhizal fungi respond to resource inequality by moving phosphorus from rich to poor patches across networks. *Current Biology*, 29(12), 2043-2050.
- Wijayawardene, N. N., Hyde, K. D., Mikhailov, K. V., Péter, G., Aptroot, A., Pires-Zottarelli, C. L. A., Goto, B. T., Tokarev, Y. S., Haelewaters, D., Karunarathna, S. C., Kirk, P. M., de A. Santiago, A. L. C. M., Saxena, R. K., Schoutteten, N., Wimalasena, M. K., Aleoshin, V. V., Al-Hatmi, A. M. S., Ariyawansa, K. G. S. U., Assunção, A. R., Karpov, S. A. (2024). Classes and phyla of the kingdom Fungi. *Fungal Diversity*, 128(1), 1-165.

- Wilson, G. W. T., & Hartnett, D. C. (1998). Interspecific variation in plant responses to mycorrhizal colonization in tallgrass prairie. *American Journal of Botany*, 85(12), 1732-1738.
- Wilson, G. W. T., Rice, C. W., Rillig, M. C., Springer, A., & Hartnett, D. C. (2009). Soil aggregation and carbon sequestration are tightly correlated with the abundance of arbuscular mycorrhizal fungi: results from long-term field experiments. *Ecology Letters*, 12(5), 452-461.
- Wipf Heidi, M. L., Xu, L., Gao, C., Spinner Hannah, B., Taylor, J., Lemaux, P., Mitchell, J., & Coleman-Derr, D. (2021). Agricultural soil management practices differentially shape the bacterial and fungal microbiomes of *Sorghum bicolor*. *Applied and Environmental Microbiology*, 87(5), e02345-02320.
- Wu, X., Staggenborg, S., Propheter, J. L., Rooney, W. L., Yu, J., & Wang, D. (2010). Features of sweet sorghum juice and their performance in ethanol fermentation. *Industrial Crops and Products*, 31(1), 164-170.
- Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., & Girshick, R. (2019). Detectron2. <https://github.com/facebookresearch/detectron2>
- Yang, H., Zhang, Q., Koide, R. T., Hoeksema, J. D., Tang, J., Bian, X., Hu, S., & Chen, X. (2017). Taxonomic resolution is a determinant of biodiversity effects in arbuscular mycorrhizal fungal communities. *Journal of Ecology*, 105(1), 219-228.
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, Montreal, Canada.

- Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: applications, challenges, and future trends. *IEEE Access*, *12*, 41180-41218.
- Yu, N., Luo, D., Zhang, X., Liu, J., Wang, W., Jin, Y., Dong, W., Liu, J., Liu, H., & Yang, W. (2014). A DELLA protein complex controls the arbuscular mycorrhizal symbiosis in plants. *Cell Research*, *24*(1), 130-133.
- Zhang, D., Li, J., Compton, R. O., Robertson, J., Goff, V. H., Epps, E., Kong, W., Kim, C., & Paterson, A. H. (2015). Comparative genetics of seed size traits in divergent cereal lineages represented by sorghum (Panicoideae) and rice (Oryzoideae). *G3: Genes, Genomes, Genetics*, *5*(6), 1117-1128.
- Zhang, S., Bourlai, T., & Arnold, J. (2024, 15-18 Dec. 2024). Mycorrhisee: a high-resolution image dataset for deep learning based quantification of arbuscular mycorrhizal fungi. 2024 IEEE International Conference on Big Data (BigData),
- Zhang, S., Lantz, W., Wilson, I., Spittle, J., Eastin, S., Mandal, A., Arnold, J., & Bourlai, T. (2024, 15-24 March 2024). A new spline-based image processing technique with applications to whole-slide imaging of plant roots. SoutheastCon 2024,
- Zhang, S., Wu, Y., Skaro, M., Cheong, J.-H., Bouffier-Landrum, A., Torres, I., Guo, Y., Stupp, L., Lincoln, B., Prestel, A., Felt, C., Spann, S., Mandal, A., Johnson, N., & Arnold, J. (2024). Computer vision models enable mixed linear modeling to predict arbuscular mycorrhizal fungal colonization using fungal morphology. *Scientific Reports*, *14*(1), 10866.
- Zhao, J., Mantilla Perez, M. B., Hu, J., & Salas Fernandez, M. G. (2016). Genome-wide association study for nine plant architecture traits in Sorghum. *The Plant Genome*, *9*(2), plantgenome2015.2006.0044.

- Zhong, Y., Chu, C., Myers, J. A., Gilbert, G. S., Lutz, J. A., Stillhard, J., Zhu, K., Thompson, J., Baltzer, J. L., & He, F. (2021). Arbuscular mycorrhizal trees influence the latitudinal beta-diversity gradient of tree communities in forests worldwide. *Nature communications*, 12(1), 1-12.
- Zhou, W., Wang, X., Fan, Y., Yang, Y., Wen, Y., Li, Y., Xu, Y., Lin, Z., Chen, L., Yao, S., Zequn, L., & Wang, J. (2024). KDSMALL: A lightweight small object detection algorithm based on knowledge distillation. *Computer Communications*, 219, 271-281.
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, 1(1).
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2019). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109, 43-76.