Deep Active Learning for Precision Agriculture: A Computational Approach

Mário de Araújo Carvalho

SERVIÇO DE PÓS-GRADUAÇÃO DA FACOM-UFMS

Data de Depósito:

Assinatura:______________________

Deep Active Learning for Precision Agriculture: A Computational Approach

Mário de Araújo Carvalho

Orientador: *Profº Drº Wesley Nunes Gonçalves*

Qualificação de doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Faculdade de Computação, mantido pela Universidade Federal de Mato Grosso do Sul, para o Exame de Qualificação de Doutorado, como parte dos requisitos para obtenção do título de Doutor em Ciência da Computação.

UFMS - Campo Grande
Março/2025

To God.

To my parents,
Welison and Maria.

In memoriam *of my maternal grandparents,*
Nonato and Esmerinda.

To my paternal grandparents,
Doralice and Valderico.

To my beloved wife,
Thalia.

To my friends.

Acknowledgments

Minha gratidão a Deus pelas bênçãos divinas que moldaram minha trajetória, guiando meus passos em cada etapa.

Agradeço profundamente à minha família e amigos, pilares que me sustentaram com amor e incentivo durante esta fase crucial da minha jornada.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), meu sincero reconhecimento pelo apoio financeiro que, através da bolsa de mestrado, impulsionou meu crescimento acadêmico e pessoal.

Abstract

Precision agriculture leverages remote sensing, autonomous vehicles, and real-time analytics to optimize inputs and enhance crop productivity across diverse environments. Integrating computer vision and machine learning further enables precise detection and targeted management of phytosanitary issues, improving efficiency and sustainability. Accurate weed species identification and classification are crucial for targeted interventions in precision agriculture; however, manually annotating large image datasets remains a significant bottleneck in this process. Recent deep active learning (DAL) methods have shown promise in reducing annotation effort while maintaining high performance in weed detection and classification tasks. In this study, we employ a dataset of 8,086 field-acquired images and start our pipeline with 100 labeled samples, iteratively querying batches of 10, 50, and 100 images using margin-based uncertainty sampling. We fine-tune a ResNet50 backbone, selected for its strong generalization ability, and rigorously evaluate its performance using F1-score, precision, and recall on an independent test set. Using MarginSampling with 100-image queries, the DAL strategy achieved an F1-score of 89.14%, close to the 91.34% obtained with the full dataset, while reducing the labeling requirements by approximately 86%. These results reinforce the viability of DAL as a scalable and cost-effective solution for real-time weed classification in precision agriculture. Future work will extend this framework to multimodal models using images and textual metadata, and this abstract represents the first step in a broader PhD investigation on DAL in precision agriculture.

Keywords: Deep Active Learning, Weed Classification, Precision Agriculture

Resumo

A agricultura de precisão utiliza sensoriamento remoto, veículos autônomos e análises em tempo real para otimizar insumos e aumentar a produtividade das culturas em diversos ambientes. A integração da visão computacional e do aprendizado de máquina permite ainda mais a detecção precisa e o gerenciamento direcionado de problemas fitossanitários, melhorando a eficiência e a sustentabilidade. A identificação e a classificação precisas de espécies de plantas daninhas são cruciais para intervenções direcionadas na agricultura de precisão; no entanto, a anotação manual de grandes conjuntos de dados de imagens continua sendo um gargalo significativo nesse processo. Métodos recentes de aprendizado ativo profundo (DAL) têm se mostrado promissores na redução do esforço de anotação, mantendo alto desempenho em tarefas de detecção e classificação de plantas daninhas. Neste estudo, empregamos um conjunto de dados de 8.086 imagens adquiridas em campo e iniciamos nosso pipeline com 100 amostras rotuladas, consultando iterativamente lotes de 10, 50 e 100 imagens usando amostragem de incerteza baseada em margem. Ajustamos um backbone ResNet50, selecionado por sua forte capacidade de generalização, e avaliamos rigorosamente seu desempenho usando pontuação F1, precisão e recall em um conjunto de teste independente. Utilizando MarginSampling com consultas de 100 imagens, a estratégia DAL alcançou uma pontuação F1 de 89,14%, próxima aos 91,34% obtidos com o conjunto de dados completo, reduzindo os requisitos de rotulagem em aproximadamente 86%. Esses resultados reforçam a viabilidade da DAL como uma solução escalável e econômica para a classificação de plantas daninhas em tempo real na agricultura de precisão. Trabalhos futuros estenderão essa estrutura a modelos multimodais usando imagens e metadados textuais, e este resumo representa o primeiro passo de uma investigação de doutorado mais ampla sobre DAL na agricultura de precisão.

Palavras-chave: Aprendizagem Ativa Profunda, Classificação de Ervas Daninhas, Agricultura de Precisão

Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence

ML Machine Learning

DL Deep Learning

AL Active Learning

DAL Deep Active Learning

CNN Convolutional Neural Networks

UAV Unmanned Aerial Vehicle

CV Computer Vision

SIS Semantic Image Segmentation

RGB Red, Green and Blue

TP True Positive

FN False Negative

FP False Positive

TN True Negative

Chapter 1

Introduction

1.1 Background and Motivation

Precision agriculture utilizes advanced technologies, including GPS, IoT sensors, and remote sensing, to optimize field-level management of crops. Precision agriculture faces challenges despite its potential to increase yield and resource-use efficiency, including heterogeneous data sources, environmental variability, and high implementation costs [Li et al., 2023; Oghaz et al., 2019].

Modern precision agriculture increasingly relies on deep learning (DL) techniques to automate decision-making and reduce human labor [Osco et al., 2021; Milioto et al., 2018]. This includes convolutional neural networks (CNNs) for classification, object-detection models for weed and pest localization, encoder-decoder architectures for semantic segmentation, and density-map estimators for counting applications. Despite their effectiveness, DL models typically require large volumes of manually labeled images to achieve high performance, rendering dataset creation both time-consuming and expensive. This requirement for extensive annotation poses a significant barrier to widespread adoption in agricultural settings [Li et al., 2023; Chiu et al., 2020].

Active learning (AL) offers a promising approach by selectively querying the most informative samples from an unlabeled dataset, thereby reducing the overall annotation effort needed to develop accurate models in precision agriculture contexts [van Marrewijk et al., 2024a; Zahidi and Cielniak, 2021]. Deep active learning (DAL), which integrates data sampling with model optimization, has improved performance compared to traditional acquisition active learning (AL) strategies. This method more effectively leverages model uncertainty and representation learning, facilitating faster convergence with fewer annotations [Ge et al., 2024; Sener and Savarese, 2017].

In image classification, DAL methods have been successfully applied to tasks such as crop-weed discrimination and row detection in plantations using UAV imagery. These methods achieve accuracy levels comparable to fully supervised approaches while requiring annotations for only a small portion of the data [Osco et al., 2021; Zahidi and Cielniak, 2021]. Furthermore, DAL strategies have been successfully implemented in semantic segmentation on extensive aerial datasets and field phenotyping studies to effectively delineate crops, weeds, and anomalies. Therefore, applying DAL to semantic image segmentation problems can increase annotation efficiency and improve segmentation accuracy under real-world conditions [Malambo et al., 2019; Chiu et al., 2020].

Despite recent advancements, the targeted application of Domain Adaptation Learning (DAL) specific to the dual functions of classification and semantic segmentation in precision agriculture remains insufficiently investigated. This gap in the literature underscores the need for the current study, which aims to develop and evaluate DAL frameworks that effectively address these shortcomings [Li et al., 2023; Ge et al., 2024].

1.2 Problem Definition

In precision agriculture, classifying weed species within agricultural imagery poses a significant challenge due to the considerable variability in plant appearance caused by differing growth stages, lighting conditions, and environmental factors. Furthermore, the subtle visual differences among various weed classes underscore the need for advanced feature extraction and classification techniques to ensure reliable discrimination between similar species [Goyal et al., 2025].

The annotation process for large-scale agricultural image datasets is labor-intensive, time-consuming, and costly, often requiring significant domain expertise to accurately differentiate between visually similar classes, such as crops and weeds [Wang and Stavness, 2025; Rawat et al., 2022]. Tasks that involve semantic segmentation, where each pixel must be labeled individually, are particularly challenging, prone to errors, and resource-intensive [Wang and Stavness, 2025; Marrewijk et al., 2024]. Furthermore, agricultural datasets frequently exhibit class imbalance, with many annotated images contributing little to model learning, which diminishes annotation efficiency and escalates costs [Wang and Stavness, 2025]. These challenges render manual annotation at scale increasingly unsustainable, often resulting in diminished dataset quality due to mislabeling and inconsistencies [Silva et al., 2024].

The limited availability of labeled data poses a significant challenge in precision agriculture classification tasks, which are often further complicated by substantial class imbalances and noisy or ambiguous samples. These issues can adversely affect model performance [Johnson and Khoshgoftaar, 2019]. To effectively tackle these challenges, it is essential to integrate advanced deep learning architectures with specialized active learning methodologies specifically designed to handle imbalanced and noisy datasets. Such comprehensive computational strategies have demonstrated promising results in enhancing the robustness and generalization capabilities of models when applied to various real-world agricultural environments [Ren et al., 2022].

1.3 Hypothesis

In this work, we have the following hypotheses:

1. **Deep Active Learning in Agriculture Applications Improves Accuracy with Less Labeled Data:** We hypothesize that utilizing Data-Active Learning (DAL) strategies will markedly enhance the accuracy of applications in precision agriculture when compared to conventional supervised learning methods, all while necessitating a reduced volume of labeled data. This improvement stems from the DAL's ability to intelligently identify the most informative samples for annotation, thereby directing labeling efforts toward the most challenging and ambiguous cases. In contrast, standard supervised learning often relies on random sampling, which may include a significant number of redundant or easily classified examples.

2. **Multimodal Models (Text+Images) Improve Deep Active Learning Methods** Integrating multimodal information, particularly the combination of text and images, can significantly improve the performance of Deep Active Learning (DAL) methods in image classification tasks. This improvement is anticipated due to the enhanced feature representations and contextual understanding that arise from merging visual and textual data. Such integration may lead to more effective selection of informative samples for annotation, as highlighted by recent advancements in vision-language models [Bang et al., 2024].

1.4 Objectives

The primary objective of this work is to evaluate DAL methods applied to RGB images in agricultural contexts, aiming to develop strategies that reduce the amount of labeled data required while maintaining classification accuracy. To achieve this overall goal, we have defined the following specific objectives:

1. **Dataset Creation and Annotation:** Construct datasets of high-resolution RGB images captured by UAVs for various precision agriculture applications - such as weed classification and anthill segmentation—ensuring representation across different environmental conditions, phenological stages, and spatial contexts. This process will involve field data collection and manual annotation tailored to the specific task requirements, such as classification labels or segmentation masks.

2. **Baseline Model Development and Evaluation:** Establish baseline performance by training and evaluating state-of-the-art CNN architectures on a fully labeled and randomly sampled subset of the dataset. Performance will be assessed using standard quantitative metrics, such as accuracy, precision, recall, and F1-score, to serve as a reference for comparisons with Deep Active Learning strategies.

3. **Implementation and Evaluation of Uncertainty-Based Sample Selection in DAL:** Implement and compare multiple uncertainty-driven acquisition functions within a Deep Active Learning (DAL) framework. CNN models will be iteratively trained using samples selected by each acquisition strategy, and their performance will be evaluated using the same metrics applied to the baseline, enabling a systematic comparison of sample selection efficiency.

4. **Labeling Effort Reduction and Analysis of Selected Samples:** Quantify the reduction in labeling effort achieved by the best-performing Deep Active Learning strategy relative to the baseline supervised approach. Additionally, analyze the characteristics of the selected samples through visualization, clustering, or dimensionality reduction techniques to better understand the selection behavior and identify challenging or informative patterns in the data.

5. **Development of a Multimodal Deep Active Learning Strategy:** Propose and evaluate a Deep Active Learning strategy that integrates multimodal data by combining RGB UAV imagery with textual or contextual metadata to guide the sample selection process. The approach aims

to leverage the complementary strengths of visual and textual information to enhance annotation efficiency and model performance in precision agriculture tasks.

1.5 Dissertation Text Organization

This section presents the organization of this thesis qualification proposal.

This thesis is organized into three main chapters. Chapter 1 presents the introductory sections of this qualification work, addressing the background and motivation in Section 1.1, the definition of the problem in Section 1.2, the hypotheses in Section 1.3, the general and specific objectives in Section 1.4 and the structure of the dissertation in Section 1.5.

Chapter 2 is a scientific paper developed during the implementation of the objectives of this work, where the main results of the first part of this thesis qualification work are presented. In this Chapter 2, it is possible to find the entire methodology of this scientific process used, the definition of the problem proposed in the chapter, related works, methodology, definition of the data sets used and the DAL techniques evaluated, ablation results, quantitative and qualitative results, in addition to the general discussion of the chapter and limitations found in the process. Finally, in Chapter 3 we present the conclusions of this qualification work to date, including a parallel between the objectives of this dissertation and the results obtained, future works and the execution schedule of the following activities to be developed until the completion of the final thesis.

Chapter 2

Deep Active Learning for Weed Species Classification in Images

This chapter presents the paper submitted to **IEEE Geoscience and Remote Sensing Letters.**

Abstract

Accurate weed identification is crucial for precision agriculture; however, it presents significant challenges due to visual variations influenced by growth stage, species, environment, and image capture techniques, which typically require extensive manual annotation. This study tackles the issue of annotation costs by investigating Deep Active Learning (DAL) for classifying weeds into multiple categories. We developed the DalMax framework, which incorporates a ResNet-50 convolutional neural network (CNN) trained on an agricultural image dataset and evaluates various active learning strategies, including MarginSampling, BALDDropout, EntropySampling, LeastConfidence, and others. Our approach commenced with 100 labeled images and iteratively enriched the dataset over 10 rounds by querying informative samples in batches of 10, 50, and 100. The performance of DAL models was evaluated both quantitatively and qualitatively based on the hyperparameters tested and the number of images collected per round. Margin sampling with a query size of 100 yielded an F1-score of 89.14%, approaching the 91.34% upper bound achieved by training on the complete dataset of 8,086 images. Confusion matrices and visual analyses demonstrated that DAL strategies consistently enhanced accuracy compared to random selection, although certain classes with subtle

distinctions remained difficult to classify. These findings underscore the potential of DAL, particularly MarginSampling, to significantly reduce annotation efforts without substantial loss in accuracy, attaining near upper-bound performance with approximately 86% fewer labeled images. This research emphasizes the viability of DAL as a scalable and cost-effective solution for real-time weed classification, thereby promoting sustainable agricultural practices.

2.1 Introduction

Precision agriculture is increasingly leveraging advanced image analysis techniques to enhance crop management and promote sustainable farming practices. A critical component of this approach is the accurate identification and classification of weed species, which is essential for optimizing herbicide application and minimizing environmental impact [Zhang et al., 2023]. These automated systems enhance decision-making processes, resulting in more efficient and environmentally friendly agricultural operations. Weed infestations pose a persistent and significant threat to global agriculture, resulting in substantial losses of crop yield. Traditional weed management methods, such as manual removal and extensive use of herbicides, are labor-intensive, environmentally harmful, and economically unsustainable. In contrast, precision agriculture provides a promising alternative by facilitating targeted weed control, thereby reducing chemical use and promoting overall crop health [Hasan et al., 2023].

Accurate identification and classification of weed species are essential for the success of precision agriculture. However, the visual similarities between weeds and crops, variations in growth stages, and diverse environmental conditions complicate this process. Conventional supervised learning models, which form the basis of many image classification tasks, require extensive labeled datasets; however, these are often limited in agricultural contexts due to the high costs and time associated with data annotation [Hasan et al., 2021].

Recent advancements in deep learning (DL) have revolutionized image classification tasks, achieving notable accuracy and robustness. Convolutional Neural Networks (CNNs) and other deep architectures have proven highly effective at extracting complex features from images, which is particularly advantageous for the intricate task of weed classification [Sishodia et al., 2020; Weiss et al., 2020]. Nonetheless, these models generally require large amounts of labeled data, presenting significant challenges due to the substantial costs and labor involved in manual annotation.

Deep active learning (DAL) effectively solves the data annotation bottleneck by selectively querying the most informative samples for labeling. This method reduces the labeling effort while enhancing model performance, especially in multiclass classification challenges where inter-class variations can be subtle and complex [Li et al., 2024a]. Consequently, integrating active learning with deep learning frameworks emerges as a promising strategy to improve the efficiency and accuracy of weed classification systems in practical agricultural applications.

Deep learning (DL) models have demonstrated significant potential in classifying plant images [Carvalho et al., 2022; Joshi et al., 2025]. However, their effectiveness largely depends on the availability of extensively annotated datasets, often scarce in the agricultural sector. This scarcity can hinder the model's ability to generalize across diverse conditions. Moreover, a notable research gap exists in the application of DAL strategies aimed at reducing reliance on large labeled datasets, particularly for weed classification [Adhinata et al., 2024]. While existing datasets, such as DeepWeeds [Olsen et al., 2019] and Weed25 [Wang et al., 2022], demonstrate the potential of DL in this field, they also underscore the urgent need for innovative methods that can reduce dependence on extensive labeled data.

Deep learning models that accurately classify weed species can be crucial in a global landscape where precision agriculture is becoming increasingly implemented to enhance productivity and sustainability. However, their scalability is frequently hindered by the need for large amounts of labeled data, which may not be easily accessible in many areas, especially in developing countries [Adhinata et al., 2024].

This study aims to develop a deep learning (DL) model for the multiclass classification of weed plant images using DAL methods. It seeks to integrate active learning strategies to reduce the dependence on large labeled datasets and to assess the model's performance across various environmental conditions and weed species. Ultimately, the primary objective of this research is to propose a computational approach that combines DL and DAL techniques for the efficient and accurate classification of multiple weed species. This approach aims to reduce reliance on extensive labeled datasets, thereby contributing to the broader accessibility and scalability of precision agriculture technologies worldwide.

2.1.1 Chapter Organization

The chapter is organized as follows. The section above clearly presents the background and motivation, defining the research problem and outlining the primary contributions. This establishes a comprehensive framework for the study, following practical guidelines for scientific communication in computer science [Shaw, 2003; Theisen et al., 2017].

Section 2.2 provides a review of the related literature on deep learning (DL) and active learning methodologies, with a focus on their applications to precision agriculture and weed classification tasks. This section synthesizes insights from recent research, identifying current advances and explicitly highlighting knowledge gaps targeted by our study [Gastel and Day, 2019].

The subsequent sections outline the methodology, experimental design, and results, offering a comprehensive discussion of the findings. The paper concludes by summarizing key results and suggesting directions for future research. This structured approach ensures logical flow, clarity, and coherence, consistent with established best practices in scientific writing [Gastel and Day, 2019].

2.2 Related Work

2.2.1 Deep Learning in Image Classification

Deep learning has fundamentally revolutionized image classification tasks, driving substantial advancements across various domains, including medical imaging, autonomous vehicles, and remote sensing [LeCun et al., 2015]. Convolutional Neural Networks (CNNs), in particular, have played a pivotal role, consistently establishing new benchmarks for accuracy and robustness in visual recognition tasks [Alom et al., 2019]. These models have significantly benefited from architectural innovations and the availability of large-scale datasets, resulting in enhanced capabilities for feature extraction, generalization, and reliable performance across diverse and challenging scenarios [Khan et al., 2020].

Recent advancements in convolutional neural network (CNN) architectures have significantly enhanced the capabilities of feature extraction and representation. Innovations such as residual connections, attention mechanisms, and depthwise separable convolutions have contributed to the development of more efficient and effective models [He et al., 2016; Dosovitskiy et al., 2020; Wang et al., 2025]. For instance, the integration of residual connections in ResNet architectures has addressed the vanishing gradient problem, enabling the training of deeper networks and improving performance in high-dimensional image datasets [He et al., 2016]. Similarly, attention mechanisms, as implemented in models such as SENet, have enabled networks to focus on the most informative features, thereby enhancing accuracy [Hu et al., 2018]. Furthermore, the use of depthwise separable convolutions in architectures such as MobileNet has reduced computational complexity while maintaining robust performance, making them suitable for deployment on resource-constrained devices [Howard et al., 2017].

Recent research has explored the integration of novel data augmentation and regularization techniques to address challenges related to data scarcity and class imbalance [Shorten and Khoshgoftaar, 2019]. Compelling evidence demonstrates that advanced augmentation strategies not only mitigate overfitting but also enhance the generalization capabilities of DL models across diverse application scenarios [Shorten and Khoshgoftaar, 2019].

In addition to architectural and data-centric enhancements, the development of efficient training algorithms has become a focal point in addressing the computational demands of deep neural networks. Notably, the advent of models such as MobileNet has demonstrated that it is possible to achieve high accuracy with significantly fewer parameters, making it suitable for deployment in resource-constrained environments [Howard et al., 2017]. Similarly, the introduction of EfficientNet has demonstrated the efficacy of compound scaling, which involves uniformly scaling network depth, width, and resolution to achieve superior performance with reduced computational resources [Tan and Le, 2019]. These advancements offer promising solutions for deploying high-performance image classification models in resource-constrained environments. Some works have introduced a novel training framework that leverages model improvement and transfer learning, offering a promising solution for deploying high-performance image classification models in resource-constrained environments [Yousif, 2024; Liu et al., 2023].

The classification task in this study employs the ResNet50 convolutional neural network, a widely recognized architecture known for its excellent performance and ability to mitigate the vanishing gradient problem through residual connections [He et al., 2016]. ResNet50 has demonstrated strong generalization capabilities across various image classification tasks, making it particularly suitable for evaluating the effectiveness of samples selected by deep active learning techniques [Alom et al., 2019]. By adopting ResNet50 as the classifier backbone, we ensure a robust and consistent evaluation of the selected weed images, facilitating accurate assessments of active learning strategies and their contribution to model accuracy and efficiency.

2.2.2 Active Learning and Deep Active Learning Techniques

Active learning has become a crucial strategy for reducing annotation costs and enhancing model performance, particularly in image classification tasks. Foundational approaches such as Least Confidence **(LeastConfidence)** and Margin Sampling **(MarginSampling)** have paved the way for modern active learning techniques [Lewis and Gale, 1994; Scheffer et al., 2001]. These methods assess sample uncertainty based on the model's prediction confidence, providing an effective baseline for further developments.

Entropy-based methods have also been widely adopted to quantify uncertainty in predictions. The classical work on entropy-based sampling, often referred to as **EntropySampling**, is well documented in the literature [Settles, 2009]. Extensions of this approach, such as **EntropySamplingDropout**, integrate dropout-based uncertainty estimation to enhance robustness in DL models, thereby offering improved decision-making in sample selection.

Cluster-based selection techniques aim to diversify the training set by selecting representative samples from the data distribution. Approaches like **KCenterGreedy** and **KMeansSampling**, as exemplified in the core-set methodology [Sener and Savarese, 2017], ensure that the chosen samples effectively cover the data manifold. Furthermore, adversarial strategies, known as the **Adversarial margin** method, leverage the sensitivity of models to adversarial examples to identify informative samples [Ducoffe and Precioso, 2018].

Recent advancements have focused on refining uncertainty estimation by incorporating dropout during inference. Techniques such as **Least Confidence Dropout** and **Margin Sampling Dropout** employ stochastic forward passes to obtain more reliable uncertainty measures. This refinement has been further validated by adaptive approaches that combine dropout with active learning, complementing previous Bayesian frameworks in the field [Gal and Ghahramani, 2015].

The integration of multiple active learning strategies—including LeastConfidence, MarginSampling, EntropySampling, BALDDropout (as in Deep Bayesian Active Learning with Dropout Estimation), KCenterGreedy, KMeansSampling, and the adversarial margin approach—has shown significant promise in reducing labeling efforts while enhancing classification accuracy. Scalable methods for multiclass image classification problems further support the advantages of these techniques for complex image classification tasks [Joshi et al., 2012; Ranganathan et al., 2017; Li et al., 2024b]. For clarity, the abbreviations used in our experiments are defined as follows: **BALDDropout** (Deep Bayesian Active Learning with Dropout Estimation), **EntropySampling** (entropy-based sampling), **EntropySamplingDropout** (entropy-based sampling with dropout estimation), **KCenterGreedy** (cluster-based selection using a greedy center approach), **KMeansSampling** (sample selection based on K-Means clustering), **LeastConfidence** (uncertainty sampling based on the least confident prediction), **LeastConfidenceDropout** (LeastConfidence with dropout estimation), **MarginSampling** (sampling based on prediction margins), **MarginSampling-Dropout** (MarginSampling with dropout estimation), and **RandomSampling** (random selection of samples as a baseline).

2.2.3 Applications in Agriculture and Weed Classification

Recent advances in DL have revolutionized the field of precision agriculture, particularly in automated weed detection and classification. Convolutional neural network structures can effectively distinguish weed species from crop images, demonstrating high accuracy even under variable field conditions [Abdulsalam and Aouf, 2020; Pai et al., 2024]. This approach minimizes manual work and also improves decision-making in crop management.

The integration of unmanned aerial vehicle (UAV) imaging with DL techniques has significantly advanced agricultural applications. For instance, a study utilized UAV-acquired RGB images to compare various DL segmentation models for weed detection, identifying that the UNet [Ronneberger et al., 2015] model with EfficientNetB0 [Tan and Le, 2019] as a backbone achieved high precision and recall rates, thereby enhancing the efficiency of weed management in agricultural settings [Pai et al., 2024]. Similarly, research has demonstrated that combining UAV technology with convolutional neural networks (CNNs) enables accurate and efficient weed detection, highlighting the potential for the scalable and practical deployment of DL methods in real-world farming scenarios [Zhang et al., 2024a].

Visual characteristics, such as color, texture, and shape, inherent in RGB images, are crucial for distinguishing between weeds and crops, providing significant advantages in the classification of invasive plants. The joint analysis of color and texture in RGB images can achieve considerable accuracy in weed identification [Mekhalfa and Yacef, 2021]. This integrated approach proves particularly useful in scenarios where the image is simply a traditional RGB image captured via a UAV.

The active learning strategy in the agricultural sector holds promise for optimizing resources and reducing costs associated with data annotation [van Marrewijk et al., 2024b]. In contrast to the practice of exhaustively labeling the entire dataset, active learning focuses on intelligently selecting the most informative samples for annotation [Settles, 2009]. This targeted approach not only minimizes the manual effort and costs associated with annotating large volumes of data but also ensures that the model learns more efficiently by focusing on the information most relevant to the task at hand, such as detecting crop diseases or identifying invasive plants [van Marrewijk et al., 2024b]. In this way, active learning can facilitate more effective resource management and informed decision-making in the agricultural sector.

2.3 Materials and Methods

2.3.1 Study Area and Data Acquisition

The images were captured using a UAV, which enabled the acquisition of high-resolution agricultural scenes and the creation of an orthomosaic. Different regions of Brazil were surveyed, covering 282 farms located in the states of São Paulo and Mato Grosso do Sul, Brazil. The exact locations of these regions have been omitted due to contractual privacy agreements. The orthomosaics were partitioned into non-overlapping patches of 128×128 pixels. Subsequently, specialists labeled only those patches containing weeds, classifying each into one of the C weed classes to ensure accurate annotations. All images were stored in RGB format. Consequently, a comprehensive dataset of 10,192 images was compiled and organized into two primary directories: an Unlabeled Pool set containing 8,086 images and a fixed Labeled Test set comprising 2,106 images. Five distinct weed species classes were identified: Gramineae spp., Ricinus communis, Brachiaria spp., Panicum maximum, and broad-leaved species. Figure 2.1 illustrates representative examples of these weeds under realistic environmental conditions.

In the context of DAL, the Pool set is intended to serve as a reservoir of unlabeled images from which the model actively selects the most informative samples for annotation during the experiment. To simulate the active learning process without incurring excessive manual labor, all images in the Pool set were pre-annotated. This pre-annotation enabled us to efficiently mimic the gradual annotation process that typically occurs as images are selected for training.

2.3.2 Image Annotation and Preprocessing

The annotation process was meticulously performed by a domain expert who manually reviewed each image and correctly assigned it to its respective class folder based on the weed species. This careful manual annotation ensured that the dataset maintained high labeling accuracy, which is critical for training robust deep active learning models. The clear segregation of images into class-specific directories has streamlined subsequent data handling and analysis. Consequently, during the experiments, although the Pool set was treated as unlabeled for active selection, the availability of pre-annotated images enabled rapid and accurate updates to the training base, facilitating an effective evaluation of the DAL methods. The distribution of images by class for both the Pool (annotated post-selection) and Test sets is detailed in Table 2.1.

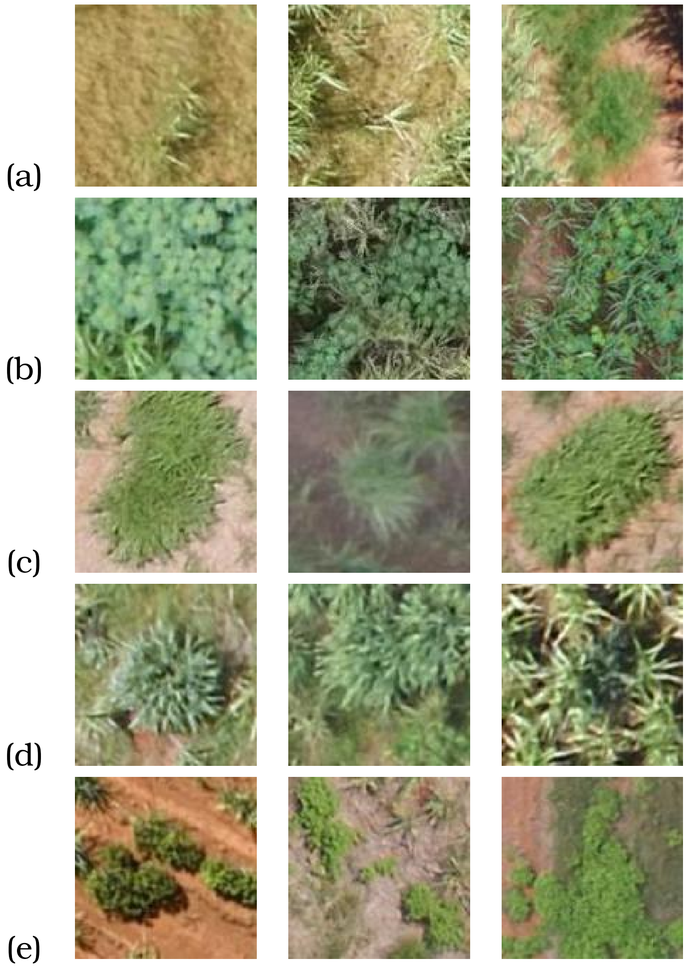


Figure 2.1: Representative sample images of each class in the dataset under realistic environmental conditions, namely: (a) Gramineae spp., (b) Ricinus communis, (c) Brachiaria spp, (d) Panicum maximum and (e) Broadleaf species. Source: Carvalho, 2024.

Dataset class	Size Pool set	Size Test set
Gramineae spp.	1,198	970
Ricinus communis	3,370	456
Brachiaria spp.	774	335
Panicum maximum	1,036	190
Broadleaf species	1,708	155
Total	8,086	2,106

Table 2.1: Distribution of classes in the annotated data set.

All images were normalized using the standard ImageNet weights [Deng et al., 2009]. Specifically, each image was scaled and normalized based on the mean and standard deviation computed from the ImageNet dataset [Deng et al., 2009]. This normalization step is a widely adopted practice in DL pipelines [Ioffe and Szegedy, 2015] that helps reduce the effects of varying illumination and color discrepancies across the dataset.

2.3.3 Proposed Methodology

The methodological framework of this study involved acquiring high-resolution imagery data using an unmanned aerial vehicle (UAV), followed by the specialized manual annotation of images to create a high-quality initial dataset. This dataset was integrated into the DalMax platform [Carvalho, 2024], an open-source deep active learning (DAL) framework implemented in Python 3[1]. DalMax enables flexible experimentation with a wide range of active learning strategies, offering structured control over parameters such as iteration count, query size, and initial labeled set size through configuration files [Pezoa et al., 2016].

The experimental protocol was divided into two phases. In the first phase, different DAL methods were evaluated using varied parameter settings. In the second phase, an image classification task was conducted to validate the effectiveness of the DAL strategies in selecting the most informative samples. For all experiments, the ResNet-50 architecture [He et al., 2016], pre-trained on ImageNet and fine-tuned for the weed classification task, was employed as the backbone model. Minor architectural modifications were made to efficiently accommodate the iterative learning process, allowing the network to integrate newly labeled samples over successive rounds.

Each experimental execution consisted of ten active learning rounds. Initially, 100 samples were randomly selected and annotated. In each subsequent round, a new batch of samples was queried and annotated, with query sizes of 100, 50, or 10 samples, depending on the experiment. At each round, the ResNet-50 model was retrained for ten epochs using the updated dataset and evaluated on a reserved test set. This progressive labeling strategy allowed for a dynamic assessment of model performance and sample informativeness.

To ensure robustness, three distinct experiments were conducted corresponding to the three query sizes, and each configuration was repeated three times with different random seeds. A total of 90 experimental runs were performed (10 DAL methods $\times$ 3 seeds $\times$ 3 query sizes). The primary performance metric was the F1-score, complemented by confusion matrix analyses to gain a deeper understanding of classification outcomes.

As the number of annotated samples increased with each round, the total varied according to the query size. For instance, with 100 samples per round, the final set comprised 1,100 annotated images. With 50 and 10 samples per round, the final totals were 600 and 200 images, respectively. In all cases, the annotated subset remained significantly smaller than the full dataset of 8,086 images. For baseline comparisons, random sampling was consistently included in all experiments.

[1] https://www.python.org/

Additionally, to establish an upper performance bound, a full training session was conducted using the entire pool of images. This complete training run provided a reference for maximum achievable accuracy, facilitating the evaluation of the performance gap between random selection and DAL strategies.

The DAL methods tested included RandomSampling, LeastConfidence, MarginSampling, EntropySampling, LeastConfidenceDropout, MarginSamplingDropout, EntropySamplingDropout, KMeansSampling, KCenterGreedy, and BALDDropout. The DalMax platform's flexibility enabled systematic comparisons across all these methods under standardized experimental conditions.

An overview of the complete methodological workflow, from UAV-based data acquisition to model evaluation, is depicted in Figure 2.2.

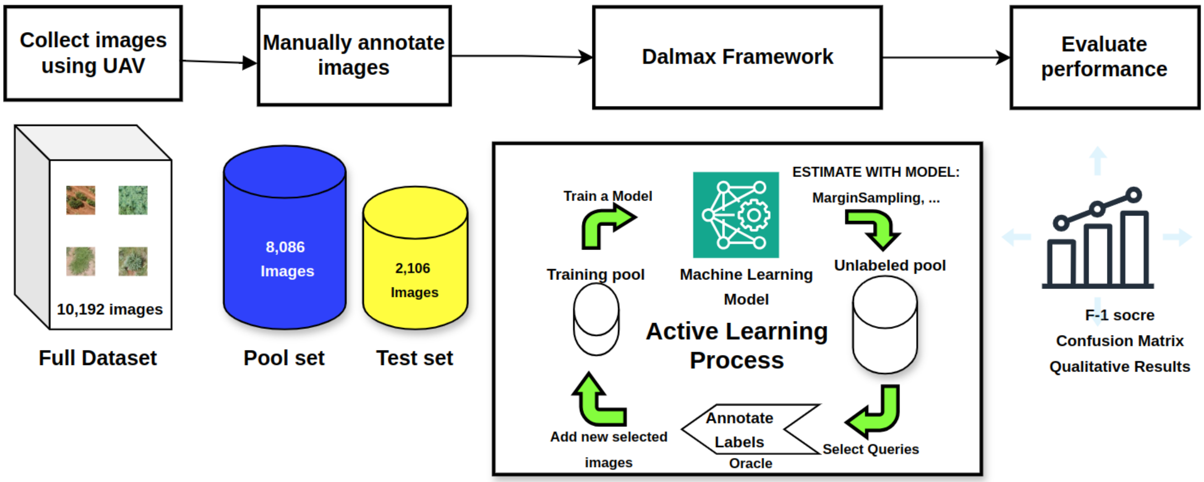


Figure 2.2: Overview of the methodological workflow employed in this study, including UAV-based data acquisition, preprocessing steps, manual annotation, integration into the DalMax framework, active learning model execution, and evaluation procedures.

2.3.4 Performance Evaluation

Model performance was evaluated using metrics derived from the confusion matrix: True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN) [Sokolova et al., 2009]. For this multi-class classification task, these metrics were calculated per class and then aggregated using a weighted average to account for class imbalance [Hossin and Sulaiman, 2015].

Precision, Recall and the F1-score were calculated as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2.2}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.3}$$

Due to the significant class imbalance in our dataset, the F1-score was prioritized over overall accuracy. Accuracy can be misleading in imbalanced scenarios, as it may be dominated by the performance of majority classes [Brodersen et al., 2010]. The F1-score, which considers both precision and recall, offers a more robust evaluation of model performance across all classes. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of performance, particularly valuable in imbalanced datasets [Chinchor, 1992; Powers, 2011]. Confusion matrices were also generated to analyze the patterns of misclassification.

2.3.5 Hardware Configuration and Execution Details

All experiments were performed on a high-performance workstation running Ubuntu 22.04.5 LTS, equipped with 128 GiB of RAM, an AMD Ryzen 9 5950x processor (16 cores × 32 threads), and two NVIDIA GeForce RTX 3080 GPUs (10 GB each). The experiments were executed using CUDA 12.4[2] with NVIDIA driver version 550.120 and implemented in PyTorch [Paszke et al., 2019][3], ensuring efficient model training and inference. Table 2.2 summarizes the hardware configuration used for all experiments.

Parameter	**Specification**
Operating System	Ubuntu 22.04.5 LTS
RAM	128.0 GiB
Processor	AMD® Ryzen 9 5950x (16-core × 32)
GPU	2 × NVIDIA GeForce RTX 3080 (10 GB each)
CUDA Version	12.4
Driver Version	550.120

Table 2.2: Detailed specifications of the high-performance computing environment used for training, evaluation, and testing of the DAL methods.

2.4 Experimental Results

2.4.1 Ablation Studies

Our ablation studies began by determining the top reference performance for the classification task. Training the model using the full dataset resulted in an F1-score of 0.9134 (±0.0084), calculated from the average of three runs

[2] https://developer.nvidia.com/cuda-toolkit
[3] https://pytorch.org/

with different random seeds. In this section, the ablation analysis focused on the individual impact of the number of images collected per round (*n_query*), evaluating the sizes of 10, 50, and 100 images.

Figure 2.3 illustrates the F1-score evolution across 10 rounds when 10 samples are queried per round. In this scenario, although the initial performance of the DAL methods was lower than that of RandomSampling, by the final round, several methods such as EntropySampling (F1-score of 0.6784 ± 0.0375), LeastConfidence (0.6602 ± 0.0073), and MarginSamplingDropout (0.6606 ± 0.0855) exceeded the baseline RandomSampling performance (0.6222 ± 0.0614). This plot underscores that even with a small query size, DAL strategies can gradually enhance their selection efficacy over successive rounds. However, since 10 samples represent a limited quantity, the effective improvement achieved each round remains marginal. Although EntropySampling, LeastConfidence, and MarginSamplingDropout surpassed the random baseline of 0.6222 (std ± 0.0614) F1-score, these values remained considerably lower than the upper bound performance of 0.9134 (std ± 0.0084) F1-score achieved by training on the complete dataset.

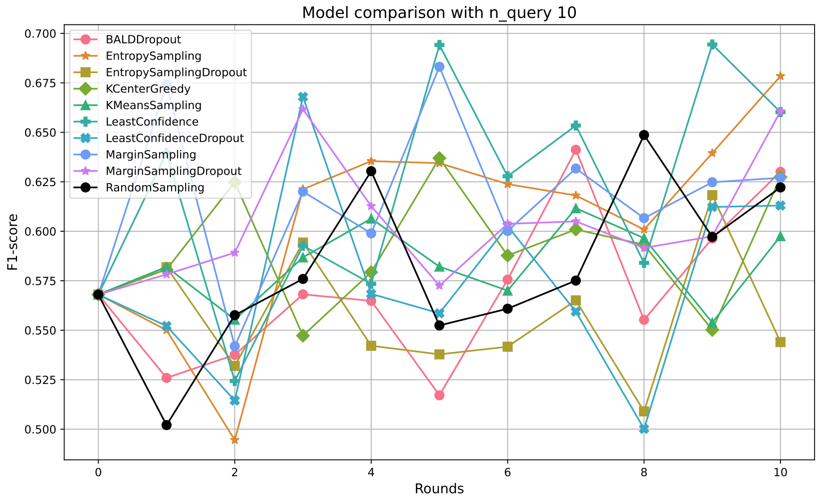


Figure 2.3: F1-score progression over 10 rounds for a query size of 10 samples per round.

Figure 2.4 presents the results for experiments with 50 queries per round. Here, the performance gap becomes more pronounced: the MarginSampling method achieved an F1-score of 0.8791 (± 0.0225) at the 10th round, significantly outperforming the RandomSampling baseline, which recorded an F1-score of 0.7562 (± 0.0460). Additional methods, including LeastConfidence (0.8336 ± 0.0178) and BALDDropout (0.8055 ± 0.0230), also demonstrated notable improvements. These results suggest that a moderate increase in the query size accelerates the convergence of DAL methods towards higher performance levels. Notably, the F1-score of 0.8791 achieved by MarginSampling, while remarkably close to the upper-bound performance of 0.9134 attained by training on the entire full dataset consisting of 8086 images, was accomplished with a significantly smaller subset of only 600 images (100 initial images + 50 images queried per round across 10 rounds). This highlights the efficiency of the active learning strategy in achieving near-optimal performance with substantially reduced annotation effort compared to a fully supervised approach.

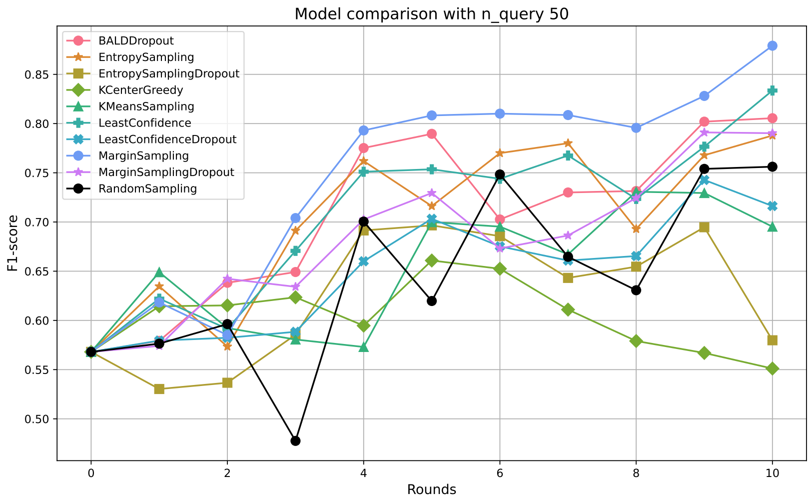


Figure 2.4: F1-score progression over 10 rounds for a query size of 50 samples per round.

The results with 100 queries per round, as shown in Figure 2.5, indicate that the advantages of active learning become even more substantial with larger query sizes. In this configuration, MarginSampling achieved the highest F1-score of 0.8914 (± 0.0191), followed by LeastConfidence at 0.8712 (± 0.0237) and EntropySampling at 0.8618 (± 0.0296), all outperforming the RandomSampling baseline, which reached 0.8365 (± 0.0279). These findings demonstrate that increasing the number of queried samples per round can significantly enhance the overall effectiveness of the active learning process. Notably, the peak F1-score of 0.8914 achieved by MarginSampling approached the upper-bound performance of 0.9134 obtained with the complete dataset of 8086 images while utilizing only 1100 images (100 initial images + 100 images queried per round across 10 rounds). This underscores the capacity of active learning with a larger query size to achieve high performance with a considerably smaller labeled dataset compared to full supervision. The F1-score achieved by MarginSampling (0.8914) was only 0.022 units lower than the upper-bound F1-score of 0.9134.

Overall, the ablation studies reveal that while DAL methods may initially lag behind RandomSampling, they eventually surpass it in performance, par-

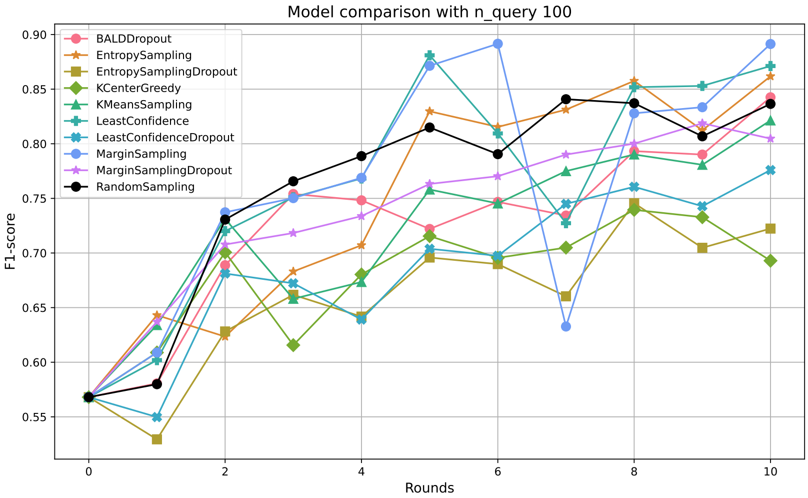


Figure 2.5: F1-score progression over 10 rounds for a query size of 100 samples per round.

ticularly as the query size increases. This trend suggests that, in practical scenarios, the strategic selection of informative samples can yield substantial gains in classification accuracy on imbalanced, multi-class datasets. The observed improvements in F1-score across different query sizes highlight the potential benefits of adopting DAL methods over a random selection approach.

2.4.2 Quantitative Results

Initially, it is essential to note that training the classifier on the entire pool of 8,086 images yielded an upper-bound F1-score of 0.9134(±0.0084) with 3 different seeds, which represents the maximum achievable performance and serves as a benchmark for our experiments. The following quantitative results compare the performance of various DAL methods under different query sizes, demonstrating how close the active learning strategies can approach this upper-bound while using significantly fewer labeled samples.

For experiments with $n\_query = 10$, Table 2.3 reports an average F1-score of 0.6222 (±0.0614) for the RandomSampling baseline after 10 rounds. The best-performing DAL method was EntropySampling, achieving an F1-score of 0.6784 (±0.0375). Other DAL methods, when arranged in increasing order of performance, include MarginSampling with 0.6270 (±0.0688), KCenterGreedy with 0.6274 (±0.0975), BALDDropout with 0.6301 (±0.1053), LeastConfidence with 0.6602 (±0.0073), and MarginSamplingDropout with 0.6606 (±0.0855). These results demonstrate that even with a small query size, resulting in a modestly sized cumulative labeled dataset, advanced active learning techniques, particularly EntropySampling, can substantially improve performance relative to random selection.

Method	F1-score
BALDDropout	0.6301 (±0.1053)
EntropySampling	**0.6784 (±0.0375)**
EntropySamplingDropout	0.5440 (±0.1169)
KCenterGreedy	0.6274 (±0.0975)
KMeansSampling	0.5975 (±0.0585)
LeastConfidence	0.6602 (±0.0073)
LeastConfidenceDropout	0.6130 (±0.0864)
MarginSampling	0.6270 (±0.0688)
MarginSamplingDropout	0.6606 (±0.0855)
RandomSampling	**0.6222 (±0.0614)**

Table 2.3: F1-score results for $n\_query = 10$ after 10 rounds (mean ± std).

In the case of $n\_query = 50$, as shown in Table 2.4, the RandomSampling baseline reached an F1-score of 0.7562 (±0.0460). Here, the performance gap widened with several DAL methods significantly outperforming the baseline. The best-performing DAL method was MarginSampling, which recorded an F1-score of 0.8791 (±0.0225). The remaining methods, listed in increasing order of performance, include EntropySampling with 0.7879 (±0.0829), MarginSamplingDropout with 0.7902 (±0.0126), BALDDropout with 0.8055 (±0.0230), and LeastConfidence with 0.8336 (±0.0178). This demonstrates that with an increased query size, the DAL methods can leverage a larger number of informative samples, resulting in a significant improvement in classification performance.

Method	F1-score
BALDDropout	0.8055 (±0.0230)
EntropySampling	0.7879 (±0.0829)
EntropySamplingDropout	0.5798 (±0.1443)
KCenterGreedy	0.5511 (±0.1752)
KMeansSampling	0.6952 (±0.0731)
LeastConfidence	0.8336 (±0.0178)
LeastConfidenceDropout	0.7162 (±0.0532)
MarginSampling	**0.8791 (±0.0225)**
MarginSamplingDropout	0.7902 (±0.0126)
RandomSampling	**0.7562 (±0.0460)**

Table 2.4: F1-score results for $n\_query = 50$ after 10 rounds (mean ± std).

For experiments with $n\_query = 100$, Table 2.5 indicates that the RandomSampling baseline achieved an F1-score of 0.8365 (±0.0279) after 10 rounds. In contrast, the best-performing DAL strategy was MarginSampling, which produced an F1-score of 0.8914 (±0.0191). The remaining methods, in increasing order of performance, include BALDDropout with 0.8426 (±0.0041), EntropySampling with 0.8618 (±0.0296), and LeastConfidence with 0.8712 (±0.0237). These results underscore that with a larger query size, active learning methods can nearly reach the performance upper-bound while requiring substantially fewer labeled images than the full dataset.

Method	F1-score
BALDDropout	0.8426 (±0.0041)
EntropySampling	0.8618 (±0.0296)
EntropySamplingDropout	0.7224 (±0.0354)
KCenterGreedy	0.6929 (±0.1527)
KMeansSampling	0.8214 (±0.0138)
LeastConfidence	0.8712 (±0.0237)
LeastConfidenceDropout	0.7759 (±0.0152)
MarginSampling	**0.8914 (±0.0191)**
MarginSamplingDropout	0.8047 (±0.0280)
RandomSampling	**0.8365 (±0.0279)**

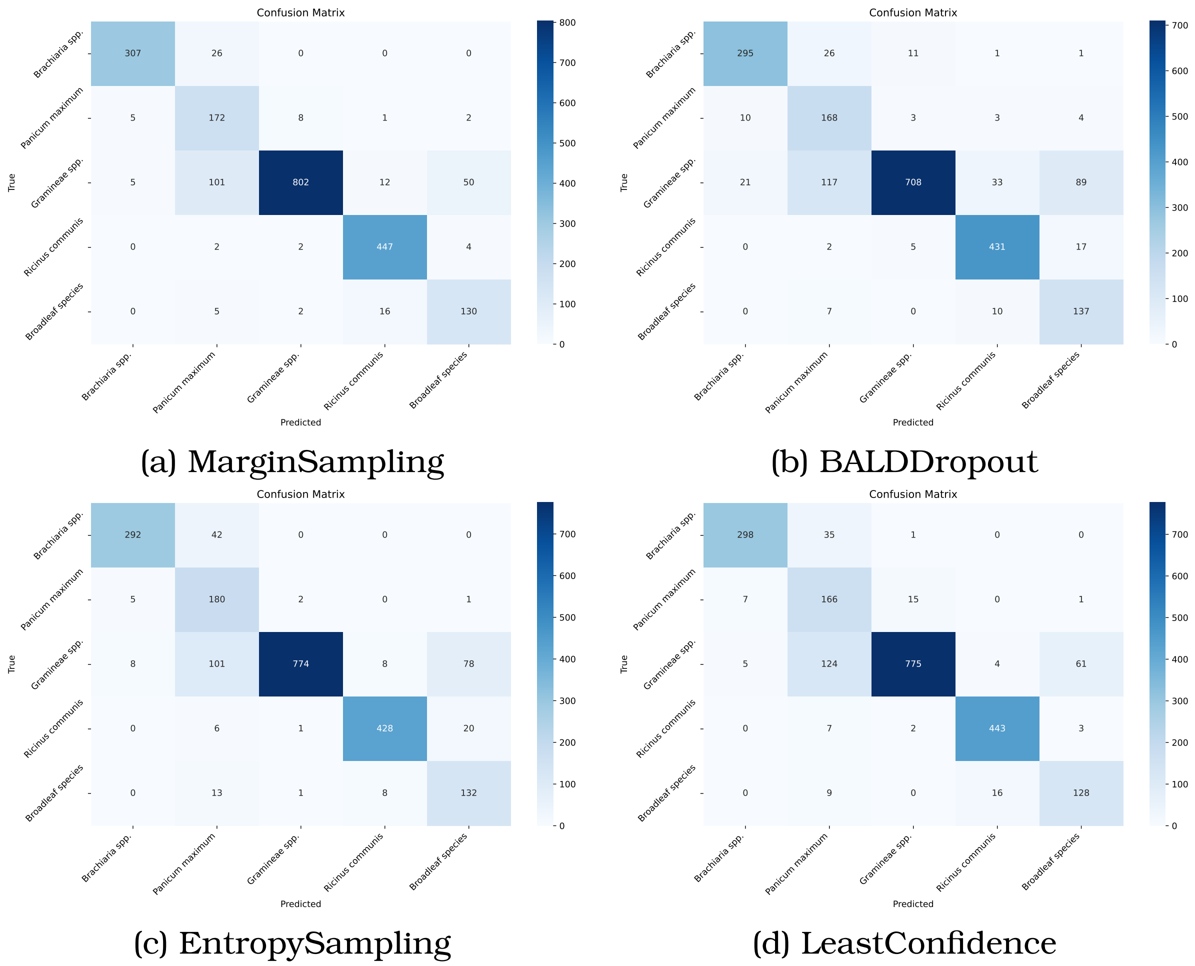
Table 2.5: F1-score results for $n\_query = 100$ after 10 rounds (mean ± std).

These quantitative findings demonstrate that the DAL methods can closely approach the upper-bound performance of 0.9134(±0.0084) F1-score while significantly reducing the labeling requirements. Although initial performance levels of all DAL models were below this upper-bound, several techniques eventually outperformed the RandomSampling baseline as the query size increased. This suggests that in practical settings, the adoption of advanced active learning strategies can yield considerable improvements in classification accuracy, reaching near-optimal performance with only a fraction of the data being labeled.

Using MarginSampling with $n\_query = 100$, the DAL method achieved an F1-score of 0.8914 (±0.0191) using only 1,100 labeled images, compared to a performance upper-bound of 0.9134(±0.0084) obtained with 8,086 labeled images. This represents a decrease of approximately 0.022 percentage points in the F1-score, or about a 2.468% reduction relative to the maximum performance. In parallel, the number of labeled images required was reduced by approximately 86.3%, demonstrating that MarginSampling can nearly match the performance of a fully labeled dataset while substantially reducing the labeling effort.

2.4.3 Weed Classification

This subsection discusses the confusion matrices of the four best DAL methods with $n\_query = 100$. The confusion matrices for the MarginSampling, BALDDropout, EntropySampling, and LeastConfidence methods, obtained after 10 query rounds of 100 samples per round (resulting in 1100 labeled images), are shown in Figure 2.6. We also present the confusion matrix of the upper bound experiment, shown in Figure 2.7, where the model was trained on the full dataset of 8086 images, providing a performance ceiling for comparison.



(a) MarginSampling (b) BALDDropout

(c) EntropySampling (d) LeastConfidence

Figure 2.6: Confusion matrices of the top four DAL methods with $n\_query = 100$: (a) MarginSampling, (b) BALDDropout, (c) EntropySampling, and (d) LeastConfidence.

MarginSampling: Figure 2.6 item a) shows the confusion matrix for the *MarginSampling* method. Brachiaria spp. was consistently identified with minimal confusion, as indicated by 307 correctly predicted samples. A moderate level of confusion is observed in the classification of Panicum maximum and Gramineae spp., where a portion of Panicum maximum was misclassified as Gramineae spp. (8 samples) and vice versa. Nevertheless, Ricinus communis and Broadleaf species maintain relatively high true-positive counts, demonstrating the effectiveness of *MarginSampling* in distinguishing between these weed classes.

BALDDropout: The confusion matrix for *BALDDropout* is illustrated in Figure 2.6 item b). Although Brachiaria spp. achieved 295 correctly classified samples, a noticeable amount of confusion arose between Brachiaria spp.,

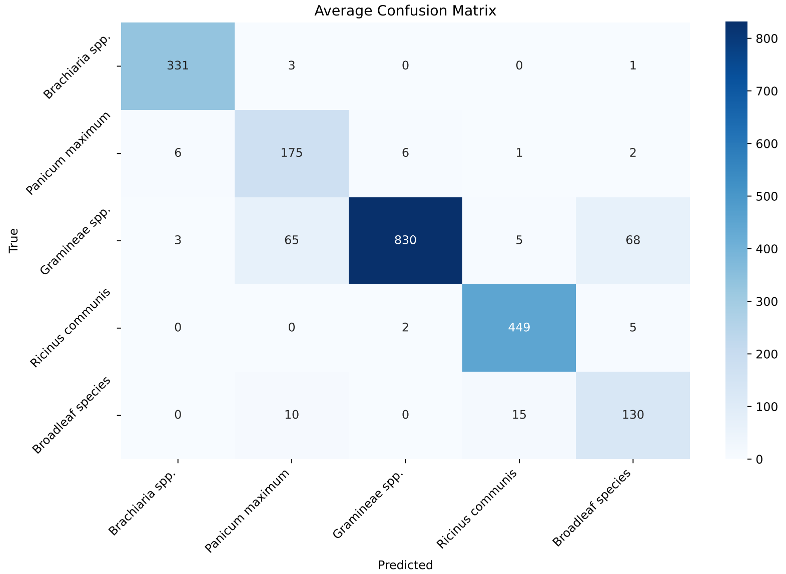


Figure 2.7: Confusion matrix for the Upper-bound experiment (8,086 images).

Panicum maximum, and Gramineae spp. (e.g., 26 samples of Brachiaria spp. misclassified as Panicum maximum and 21 Gramineae spp. misclassified as Brachiaria spp.). Despite these errors, *BALDDropout* generally performed well across the majority of classes, showcasing a strong ability to separate Ricinus communis and Broadleaf species.

EntropySampling: Figure 2.6 item c) presents the confusion matrix for *EntropySampling*. Brachiaria spp. and Panicum maximum exhibit relatively distinct boundaries, with 292 and 180 correct classifications, respectively. However, Gramineae spp. displays some confusion with Panicum maximum, reflected in the 101 samples that were incorrectly labeled. While Ricinus communis was accurately identified in 428 instances, a subset of Broadleaf species was misclassified under Gramineae spp., indicating a challenge in separating certain broadleaf weeds from grass-like species.

LeastConfidence: The confusion matrix for *LeastConfidence* in Figure 2.6 item d) demonstrates that this method correctly classified 298 instances of Brachiaria spp., 166 of Panicum maximum, and 775 of Gramineae spp. Some confusion is observed between Gramineae spp. and Panicum maximum (124 samples). However, Ricinus communis and Broadleaf species retain high levels of correct classification, highlighting the potential of *LeastConfidence* in differentiating visually distinct classes.

Upper-bound Experiment (8,086 Images): Figure 2.7 illustrates the confusion matrix resulting from the "upper-bound" experiment, in which the classifier was trained using the complete dataset containing all 8,086 Pool images. The classifier achieved an overall f1-score of 0.9134 (±0.0084), indicating high effectiveness. The diagonal elements of the matrix represent correct classifications (true positives), with notably strong results for *Gramineae spp.* (775 correctly classified instances) and *Ricinus communis* (443 correctly classified instances). However, the classifier showed some confusion between similar categories, particularly between *Gramineae spp.* and *Panicum maximum*, where 124 instances of *Gramineae spp.* were incorrectly predicted as *Panicum maximum*, highlighting morphological similarities between these groups.

Misclassifications were fewer but still noteworthy for *Brachiaria spp.*, which had 35 instances incorrectly classified as *Panicum maximum*, suggesting overlapping visual features. Likewise, broadleaf species displayed a modest level of confusion, with 16 instances misclassified as *Ricinus communis*. Despite these minor inaccuracies, the high accuracy across most categories underscores the robustness of the classifier when trained comprehensively, as demonstrated in this upper-bound scenario.

Comparative Discussion Comparing the four DAL methods with the upper-bound experiment highlights the trade-off between labeling effort and classification accuracy. Although each DAL strategy exhibits some misclassifications, the overall performance is notably high given the substantially reduced number of labeled images. *MarginSampling* and *LeastConfidence* generally show less confusion in Brachiaria spp. and Ricinus communis, while *BALD-Dropout* and *EntropySampling* display some overlap between grass-like classes (Gramineae spp. and Panicum maximum). In contrast, the upper-bound experiment demonstrates fewer misclassifications, as it leverages a far larger training set. Nonetheless, these results confirm that DAL methods can effectively approach close to the upper-bound accuracy levels with significantly fewer annotated samples, underscoring their utility for real-world weed classification tasks.

2.4.4 Qualitative Analysis and Visual Discussion

In this section, we focus on the MarginSampling method with $n\_query = 100$, which achieved the best performance among the DAL methods. We analyze qualitative examples by reviewing five visual cases where the model correctly classified the weed species and five cases where misclassifications occurred. This analysis aims to elucidate the strengths and limitations of the model, based on the error patterns observed in the confusion matrix.

Figure 2.8 shows five representative examples of correct classifications for all the classes in the dataset. These examples demonstrate the model's ability to discriminate between different classes of weeds accurately. The correct identification of Brachiaria spp. and Panicum maximum is particularly notable, given their visual similarity, indicating that the model has learned to distinguish subtle differences in their morphological characteristics. The high true-positive rates in these cases reflect the robust feature extraction capabilities of the MarginSampling method under the $n\_query = 100$ configuration.

Conversely, Figure 2.9 showcases five examples of misclassifications. Analysis of these error cases reveals that the model tends to struggle when discriminating between classes with subtle visual differences, particularly between Gramineae spp and Panicum maximum. In some instances, overlapping characteristics such as similar leaf textures and colors result in the model mislabeling an image. These cases indicate areas where the model's decision boundaries might be further refined through enhanced feature representation or additional training data.

Comparing these qualitative observations with the overall performance metrics, it is evident that while the MarginSampling method generally performs well, certain classes remain challenging to classify. The examples of correct predictions demonstrate that the model is highly effective when the visual cues are distinct and clear. In contrast, the misclassification examples highlight the difficulty in cases with ambiguous or overlapping features. This insight is valuable for guiding future improvements, as targeted enhancements in feature discrimination could help the model achieve even higher accuracy in real-world applications.

(a) (b) (c) (d) (e)

Figure 2.8: Examples of correct classifications obtained by the MarginSampling method with $n\_query = 100$. (a) Brachiaria spp., (b) Panicum maximum, (c) Gramineae spp., (d) Ricinus communis, and (e) Broadleaf species.

2.5 Discussion

2.5.1 General Discussion

The experimental results demonstrate that DAL methods effectively approach the upper-bound performance with considerably fewer labeled samples than a fully supervised approach. Quantitative metrics, such as F1-score,

(a) (b) (c) (d) (e)

Figure 2.9: Examples of incorrect classifications obtained by the MarginSampling method with $n\_query = 100$. (a) Correct: Brachiaria spp., Predicted: Panicum maximum; (b) Correct: Panicum maximum, Predicted: Gramineae spp.; (c) Correct: Gramineae spp., Predicted: Panicum maximum; (d) Correct: Ricinus communis, Predicted: Broadleaf species; (e) Correct: Broadleaf species, Predicted: Ricinus communis.

and qualitative analyses of confusion matrices and visual examples reveal that strategies like MarginSampling, BALDDropout, EntropySampling, and LeastConfidence consistently enhance classification performance by incrementally labeling informative samples. Notably, MarginSampling with $n\_query = 100$ achieved an F1-score of 0.8914, closely approaching the 0.9134 (±0.0084) upper bound of full supervision.

The progressive improvement across active learning rounds highlights the effectiveness of selective sampling in identifying the most valuable data points. Experiments with varying query sizes indicate that larger queries generally boost performance; however, even the smallest query size yielded substantial gains over random sampling. This robustness and adaptability to different labeling budgets highlight the scalability of DAL methods for scenarios with limited labeled data [Zhang et al., 2024b].

Qualitative analysis of correct and misclassified images reveals that while the models excel with distinct visual features, challenges arise with subtle inter-class differences. These detailed observations deepen our understanding of the model's decision boundaries and guide future improvements in active learning.

The superior performance of MarginSampling can likely be attributed to its theoretical foundation in uncertainty sampling. By selecting samples where the model exhibits the least confidence (i.e., the margin between the top two predicted probabilities is smallest), MarginSampling prioritizes instances that are most ambiguous and, thus, most informative for refining the decision boundary. This contrasts with methods like EntropySampling and LeastConfidence, which focus solely on the highest prediction probability, potentially overlooking valuable information present in the second-highest probability. BALDDropout, which aims to reduce the model's uncertainty about its predictions by considering the variance across multiple stochastic forward passes, also performed well, suggesting that modeling predictive uncertainty is crucial for effective sample selection. The consistent outperformance of MarginSampling indicates that focusing on the relative uncertainty between the most likely classes provides a more targeted approach to identifying informative samples, particularly in scenarios with visually similar classes where the distinction lies in subtle features. By actively seeking out these ambiguous cases, MarginSampling likely forces the model to learn more discriminative features, leading to better generalization and performance closer to the fully supervised upper bound [Gul et al., 2024].

2.5.2 Implications for Computational Approaches in Agriculture

The promising performance of DAL methods in weed classification presents significant implications for computational approaches in agriculture. The demonstrated ability of DAL to achieve close to the upper-bound performance with only a fraction of the labeled data not only reduces annotation costs but also accelerates the deployment of automated decision-making systems in precision agriculture [Kamilaris and Prenafeta-Boldú, 2018; Liakos et al., 2018]. In scenarios where timely identification of weed species is critical, these methods can improve crop management practices and support sustainable agricultural production by enabling rapid and accurate assessments in the field.

Moreover, integrating DAL with advanced imaging technologies and real-time data processing holds the potential to revolutionize agricultural monitoring. The ability to selectively label the most informative samples enables the efficient use of resources while maintaining high classification accuracy, which is crucial for precision agriculture applications. As modern agriculture increasingly adopts data-driven strategies, leveraging computational methods that minimize manual intervention will be key to optimizing resource utilization, reducing environmental impact, and ultimately enhancing overall farm productivity [Mohanty et al., 2016].

These advancements highlight the transformative potential of integrating Deep Active Learning (DAL) with computational agricultural systems. By optimizing the data annotation process and achieving near-optimal performance with minimal labeled data, DAL methods can play a significant role in advancing innovative agricultural technologies and furthering the digital transformation of precision agriculture.

2.6 Conclusion

Our study demonstrates that DAL methods can achieve performance close to the upper-bound in weed classification tasks while significantly reducing the need for extensive labeled datasets. By employing various DAL strategies, particularly MarginSampling with $n\_query = 100$, our experiments achieved an F1-score of 0.8914—nearly matching the 0.9134(±0.0084) performance upper-bound reached with full supervision on 8,086 images. The quantitative analyses, supported by detailed confusion matrices and qualitative visual evaluations, indicate that these active learning strategies effectively leverage informative samples to enhance classification accuracy in imbalanced, multi-class scenarios.

Furthermore, our experiments revealed that even with smaller query sizes, DAL methods progressively improve model performance, underscoring their potential for scalable implementation in real-world agricultural applications. The results suggest that by strategically selecting samples for annotation, it is possible to attain high performance with only a fraction of the data normally required for training, thereby reducing annotation costs and expediting model deployment in precision agriculture.

2.7 Limitations

Despite the promising results demonstrated by our DAL methods, several limitations must be acknowledged. The current study was conducted on a specific dataset, which may not fully capture the variability encountered in broader agricultural contexts. Furthermore, the performance of certain DAL strategies was affected by overlapping visual features among weed species, suggesting that even the most effective methods may struggle in cases of subtle inter-class differences. These limitations suggest that additional research is needed to enhance the feature representation for challenging cases [Settles, 2009; Oliver et al., 2018]. Exploring real-time active learning scenarios in dynamic agricultural settings will also be crucial for translating these findings into practical, scalable solutions for precision agriculture [Ranganathan et al., 2017].

2.8 Future Research Directions

Future research should focus on extending the evaluation of DAL methods to larger and more diverse agricultural datasets that better capture the inherent variability of field conditions [Kamilaris and Prenafeta-Boldú, 2018; Liakos et al., 2018]. Furthermore, there is a critical need for the development of real-time, adaptive, active learning frameworks that can operate in dynamic agricultural environments [Ren et al., 2021]. Multimodal DAL models may be a viable path for significant improvements in the proposed approach, as they combine both visual features and textual information [Yang et al., 2024; Bang et al., 2024]. Scaling up the DL task to test the effectiveness of DAL methods should be evaluated, for example, semantic image segmentation. Addressing these challenges is crucial to advancing the digital transformation of agriculture and realizing the full potential of computational approaches to enhance crop management and sustainability. Precision agriculture requires new techniques that can significantly reduce the annotated training set without loss of performance.

CHAPTER

3

Conclusions and Future Work

3.1 Conclusions

To date, research efforts have focused on the study of DAL techniques and the acquisition of the target dataset. Following the preparation of the dataset and the development of a robust coding infrastructure, empirical experiments were carried out, and a scientific paper was drafted, subsequently incorporated as Chapter 2. The principal objective of this phase was to implement and assess various DAL methods concerning the proposed task of classifying weed images. As demonstrated in Chapter 2, these methods exhibited strong performance, thereby highlighting the significant potential of DAL approaches to address the challenges inherent in our classification problem. In the second stage of the work, we intend to add another deep learning task and explore the efficiency of DAL methods in semantic image segmentation problems.

Recent studies emphasize that semantic segmentation demands extensive, costly, and time-consuming pixel-level manual annotations [Yujian et al., 2022; Chakravarthy et al., 2022; Csurka et al., 2023]. Zhang et al. [Zhang et al., 2025] introduce active learning techniques to efficiently train semantic segmentation models with limited annotated data, thereby alleviating the burden of ground-truth labeling. Reza et al. [Reza et al., 2025] highlight that obtaining large-scale, point-by-point manual annotations remains a significant challenge and propose a region-based active learning strategy. Furthermore, the prohibitive cost of obtaining dense pixel-level annotations for supervised training is identified as a central obstacle in semantic segmentation [Gao et al., 2023]. Preparing datasets for training segmentation deep neural network is inherently laborious, time-intensive, and expensive, as it necessitates both dense pixel-wise annotations and large volumes of labeled images to establish reliable ground truth and support large-scale unsupervised segmentation [Zhang et al., 2025; Reza et al., 2025; Gao et al., 2023; Luo et al., 2018].

Encouraged by the promising classification results achieved with conventional DAL techniques, the next phase of this work will explore more advanced deep active learning frameworks, particularly those that leverage multimodal models [Yang et al., 2024; Bang et al., 2024]. To this end, we will expand our current dataset to encompass a broader range of scenarios and rigorously test the limits of these novel models on larger data collections. Furthermore, recognizing the substantial annotation effort required for pixel-wise segmentation [Yujian et al., 2022], we will investigate the integration of specialized active learning strategies for segmentation, which have shown potential to significantly lower labeling costs and accelerate result generation [Mittal et al., 2024]. In this context, we aim to develop and apply multimodal DAL methods for both classification and semantic segmentation tasks, thereby enhancing the scope and robustness of our proposed solutions.

3.2 Future works

Integrating visual and textual data has shown promise in enhancing model performance in various computer vision tasks. Building upon the methodologies presented by Yang et al. [Yang et al., 2024] and Bang et al. [Bang et al., 2024], we propose a framework that combines active learning strategies with vision-language models to improve the efficiency and accuracy of agriculture precision systems.

- **Multimodal Dataset Collection:** Gather a diverse set of weed plant images across various growth stages, lighting conditions, and backgrounds, and pair each image with detailed textual descriptions (morphological characteristics, habitat information, etc.) to create a rich, complementary dataset.

- **Vision–Language Model Adaptation:** Employ a pre-trained language-vision model (e.g., CLIP) that encodes images and text into a shared embedding space and apply fast learning techniques to specialize in semantic weed species classification and segmentation tasks [Bang et al., 2024].

- **Active Learning Sample Selection:** Use the Plug and Play Active Learning (PPAL) strategy to combine uncertainty-based and diversity-based sampling methods, selecting the most informative multimodal samples for annotation to maximize performance gains per annotation cost [Yang et al., 2024].

- **Iterative Training and Refinement:** Fine-tune the adapted vision–language model on the selected samples (image + text), evaluate using accuracy, precision, recall, and F1-score, and iteratively update the model with new annotations-monitoring performance improvements to guide further sample selection.

3.3 Execution cronogram

During the first and second years of the doctoral program, all credits required for the qualification examination were completed. Table 3.1 lists the courses, seminars, and other academic activities undertaken during this period. The article developed in chapter 2 has been submitted to the journal **IEEE Geoscience and Remote Sensing Letters** and is currently being evaluated by the editors and reviewers[1].

Year–Semester	Course	Grade
2023-1	Analysis of Algorithms	C
2023-1	Teaching Internship II	–
2023-1	Artificial Intelligence	A
2023-2	Computer Architecture	A
2023-2	Teaching Internship III	–
2023-2	Directed Study in Methodology and Techniques of Computing	A
2023-2	Computing Seminars	A
2023-2	Special Topics	A
2024-1	Validation of Master's Degree Course credits	–
2024-1	Submission of an article to a journal	–

Table 3.1: Record of the courses and activities carried out within the program's framework in the first two years of the doctoral course.

To achieve the objectives outlined in Section 1.4, it is essential to implement the steps proposed in Section 3.2 over the next two years, the deadline set for the completion of this study, as detailed in the biannual schedule presented in Table 3.2.

[1] https://sibgrapi.sbc.org.br/2025/index.php/grsl/

* Activity/Stage	2025		2026	
	1º Sem.	2º Sem.	1º Sem.	2º Sem.
Qualification Exam and Revisions	•			
Dataset Expansion (Classification)	•	•		
Dataset Creation (Segmentation)	•	•		
Multimodal DAL Experiments		•	•	
Performance Evaluation and Refinement			•	
Manuscript Preparation (Classification)			•	
Manuscript Preparation (Segmentation)			•	
Final Thisis Writing				•

Table 3.2: Project development timeline leading up to the Thesis Defense.

Bibliography

Abdulsalam, M. e Aouf, N. (2020). Deep Weed Detector/Classifier Network for Precision Agriculture. In *2020 28th Mediterranean Conference on Control and Automation (MED)*, pages 1087–1092, Saint-Raphaël, France. IEEE. Cited on page 13.

Adhinata, F. D., Wahyono, e Sumiharto, R. (2024). A comprehensive survey on weed and crop classification using machine learning and deep learning. *Artificial Intelligence in Agriculture*, 13:45–63. Cited on page 9.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Van Essen, B. C., Awwal, A. A. S., e Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics*, 8(3):292. Cited on pages 10 and 11.

Bang, J., Ahn, S., e Lee, J.-G. (2024). Active Prompt Learning in Vision Language Models. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26994–27004, Seattle, WA, USA. IEEE. Cited on pages 3, 31, and 34.

Brodersen, K. H., Ong, C. S., Stephan, K. E., e Buhmann, J. M. (2010). The balanced accuracy and its posterior distribution. *2010 20th International Conference on Pattern Recognition*, pages 3121–3124. Cited on page 18.

Carvalho, M. d. A. (2024). Dalmax: Framework for deep active learning approaches. https://github.com/MarioCarvalhoBr/dalmax-framework-deep-active-learning-python. Available on GitHub. Cited on page 16.

Carvalho, M. D. A., Marcato, J., Martins, J. A. C., Zamboni, P., Costa, C. S., Siqueira, H. L., Araújo, M. S., Gonçalves, D. N., Furuya, D. E. G., Osco, L. P., Ramos, A. P. M., Li, J., De Castro, A. A., e Gonçalves, W. N. (2022). A deep learning-based mobile application for tree species mapping in RGB images. *International Journal of Applied Earth Observation and Geoinformation*, 114:103045. Cited on page 9.

Chakravarthy, A. D., Abeyrathna, D., Subramaniam, M., Chundi, P., e Gadhamshetty, V. (2022). Semantic image segmentation using scant pixel annotations. *Machine Learning and Knowledge Extraction*, 4(3):621–640. Cited on page 33.

Chinchor, N. (1992). Muc-4 evaluation metrics. In *Proceedings of the 4th conference on Message understanding*, pages 22–29. Cited on page 18.

Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R., Khachatrian, H., Karapetyan, H., Dozier, I., Rose, G., Wilson, D., Tudor, A., Hovakimyan, N., Huang, T. S., e Shi, H. (2020). Agriculture-Vision: A Large Aerial Image Database for Agricultural Pattern Analysis. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2825–2835, Seattle, WA, USA. IEEE. Cited on pages 1 and 2.

Csurka, G., Volpi, R., e Chidlovskii, B. (2023). Semantic image segmentation: Two decades of research. Cited on page 33.

Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li, e Li Fei-Fei (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, Miami, FL. IEEE. Cited on page 15.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., e Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR*, abs/2010.11929. Cited on page 10.

Ducoffe, M. e Precioso, F. (2018). Adversarial Active Learning for Deep Networks: a Margin Based Approach. Cited on page 12.

Gal, Y. e Ghahramani, Z. (2015). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. Cited on page 12.

Gao, S., Li, Z.-Y., Yang, M.-H., Cheng, M.-M., Han, J., e Torr, P. (2023). Large-scale unsupervised semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6):7457–7476. Cited on pages 33 and 34.

Gastel, B. e Day, R. (2019). *How to Write and Publish a Scientific Paper*. Greenwood, Westport, CT, 1st ed edition. Cited on page 10.

Ge, J., Zhang, Z., Phan, M. H., Zhang, B., Liu, A., e Zhao, Y. (2024). ESA: Annotation-Efficient Active Learning for Semantic Segmentation. Cited on page 2.

Goyal, R., Nath, A., e Niranjan, U. (2025). Weed detection using deep learning in complex and highly occluded potato field environment. *Crop Protection*, 187:106948. Cited on page 2.

Gul, S., Elmahallawy, M., Madria, S., e Tripathy, A. (2024). FisherMask: Enhancing Neural Network Labeling Efficiency in Image Classification Using Fisher Information. arXiv:2411.05752. Cited on page 29.

Hasan, A. S. M. M., Diepeveen, D., Laga, H., Jones, M. G., e Sohel, F. (2023). Image patch-based deep learning approach for crop and weed recognition. *Ecological Informatics*, 78:102361. Cited on page 8.

Hasan, A. S. M. M., Sohel, F., Diepeveen, D., Laga, H., e Jones, M. G. (2021). A survey of deep learning techniques for weed detection from images. *Computers and Electronics in Agriculture*, 184:106067. Cited on page 8.

He, K., Zhang, X., Ren, S., e Sun, J. (2016). Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA. IEEE. Cited on pages 10, 11, and 16.

Hossin, M. e Sulaiman, M. N. (2015). A review of evaluation metrics for data classification evaluations. *International Journal of data mining & knowledge management process*, 5(2):1. Cited on page 17.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., e Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. Cited on pages 10 and 11.

Hu, J., Shen, L., e Sun, G. (2018). Squeeze-and-Excitation Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, Salt Lake City, UT. IEEE. Cited on page 10.

Ioffe, S. e Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Bach, F. e Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France. PMLR. Cited on page 15.

Johnson, J. M. e Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):27. Cited on page 3.

Joshi, A. J., Porikli, F., e Papanikolopoulos, N. P. (2012). Scalable Active Learning for Multiclass Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2259–2273. Cited on page 12.

Joshi, R. C., Burget, R., e Dutta, M. K. (2025). AgriDeep-net: An advanced deep feature fusion-based technique for enhanced fine-grain image analytics in precision agriculture. *Ecological Informatics*, 86:103069. Cited on page 9.

Kamilaris, A. e Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, 147:70–90. Cited on pages 29 and 30.

Khan, A., Sohail, A., Zahoora, U., e Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 53(8):5455–5516. Cited on page 10.

LeCun, Y., Bengio, Y., e Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444. Cited on page 10.

Lewis, D. D. e Gale, W. A. (1994). A Sequential Algorithm for Training Text Classifiers. In Croft, B. W. e Van Rijsbergen, C. J., editors, *SIGIR '94*, pages 3–12. Springer London, London. Cited on page 11.

Li, D., Wang, Z., Chen, Y., Jiang, R., Ding, W., e Okumura, M. (2024a). A Survey on Deep Active Learning: Recent Advances and New Frontiers. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21. Cited on page 9.

Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., e Tan, X. (2023). Label-efficient learning in agriculture: A comprehensive review. *Computers and Electronics in Agriculture*, 215:108412. Cited on pages 1 and 2.

Li, X., Wang, X., Chen, X., Lu, Y., Fu, H., e Wu, Y. C. (2024b). Unlabeled data selection for active learning in image classification. *Scientific Reports*, 14(1):424. Cited on page 12.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., e Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8):2674. Cited on pages 29 and 30.

Liu, J., Chui, K. T., e Lee, L.-K. (2023). Enhancing the Accuracy of an Image Classification Model Using Cross-Modality Transfer Learning. *Electronics*, 12(15):3316. Cited on page 11.

Luo, Y., Wang, Z., Huang, Z., Yang, Y., e Zhao, C. (2018). Coarse-to-fine annotation enrichment for semantic segmentation learning. In *Proceedings of* *the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 237–246. ACM. Cited on page 34.

Malambo, L., Popescu, S., Ku, N.-W., Rooney, W., Zhou, T., e Moore, S. (2019). A Deep Learning Semantic Segmentation-Based Approach for Field-Level Sorghum Panicle Counting. *Remote Sensing*, 11(24):2939. Cited on page 2.

Marrewijk, B. M. v., Dandjinou, C., Rustia, D. J. A., Gonzalez, N. F., Diallo, B., Dias, J., Melki, P., e Blok, P. M. (2024). Active learning for efficient annotation in precision agriculture: a use-case on crop-weed semantic segmentation. arXiv:2404.02580. Cited on page 2.

Mekhalfa, F. e Yacef, F. (2021). Supervised learning for crop/weed classification based on color and texture features. Cited on page 13.

Milioto, A., Lottes, P., e Stachniss, C. (2018). Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2229–2235, Brisbane, QLD. IEEE. Cited on page 1.

Mittal, S., Niemeijer, J., Schäfer, J. P., e Brox, T. (2024). Best practices in active learning for semantic segmentation. In Köthe, U. e Rother, C., editors, *Pattern Recognition*, pages 427–442, Cham. Springer Nature Switzerland. Cited on page 34.

Mohanty, S. P., Hughes, D. P., e Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419. Cited on page 29.

Oghaz, M. M. D., Razaak, M., Kerdegari, H., Argyriou, V., e Remagnino, P. (2019). Scene and Environment Monitoring Using Aerial Imagery and Deep Learning. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 362–369, Santorini Island, Greece. IEEE. Cited on page 1.

Oliver, A., Odena, A., Raffel, C., Cubuk, E. D., e Goodfellow, I. J. (2018). Realistic evaluation of deep semi-supervised learning algorithms. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 3239–3250, Red Hook, NY, USA. Curran Associates Inc. Cited on page 30.

Olsen, A., Konovalov, D. A., Philippa, B., Ridd, P., Wood, J. C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., Calvert, B., Azghadi, M. R., e White, R. D. (2019). DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Scientific Reports*, 9(1):2058. Cited on page 9.

Osco, L. P., Dos Santos De Arruda, M., Gonçalves, D. N., Dias, A., Batistoti, J., De Souza, M., Gomes, F. D. G., Ramos, A. P. M., De Castro Jorge, L. A., Liesenberg, V., Li, J., Ma, L., Marcato, J., e Gonçalves, W. N. (2021). A CNN approach to simultaneously count plants and detect plantation-rows from UAV imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174:1–17. Cited on pages 1 and 2.

Pai, D. G., Kamath, R., e Balachandra, M. (2024). Deep Learning Techniques for Weed Detection in Agricultural Environments: A Comprehensive Review. *IEEE Access*, 12:113193–113214. Cited on page 13.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*. Cited on page 18.

Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., e Vrgoč, D. (2016). Foundations of json schema. In *Proceedings of the 25th International Conference on World Wide Web*, pages 263–273. International World Wide Web Conferences Steering Committee. Cited on page 16.

Powers, D. M. (2011). Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, 2(1):37–63. Cited on page 18.

Ranganathan, H., Venkateswara, H., Chakraborty, S., e Panchanathan, S. (2017). Deep active learning for image classification. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 3934–3938. Cited on pages 12 and 30.

Rawat, S., Chandra, A. L., Desai, S. V., Balasubramanian, V. N., Ninomiya, S., e Guo, W. (2022). How Useful Is Image-Based Active Learning for Plant Organ Segmentation? *Plant Phenomics*, 2022:9795275. Cited on page 2.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., e Wang, X. (2021). A survey of deep active learning. *ACM Comput. Surv.*, 54(9). Cited on page 30.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B., Chen, X., e Wang, X. (2022). A Survey of Deep Active Learning. *ACM Computing Surveys*, 54(9):1–40. Cited on page 3.

Reza, M. A., Manley, E., Chen, S., Chaudhary, S., e Elafros, J. (2025). Segbuilder: A semi-automatic annotation tool for segmentation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, page 8494–8503. IEEE. Cited on pages 33 and 34.

Ronneberger, O., Fischer, P., e Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. Cited on page 13.

Scheffer, T., Decomain, C., e Wrobel, S. (2001). Active Hidden Markov Models for Information Extraction. In Goos, G., Hartmanis, J., Van Leeuwen, J., Hoffmann, F., Hand, D. J., Adams, N., Fisher, D., e Guimaraes, G., editors, *Advances in Intelligent Data Analysis*, volume 2189, pages 309–318. Springer Berlin Heidelberg, Berlin, Heidelberg. Cited on page 11.

Sener, O. e Savarese, S. (2017). Active Learning for Convolutional Neural Networks: A Core-Set Approach. Cited on pages 2 and 12.

Settles, B. (2009). Active learning literature survey. *University of Wisconsin-Madison Department of Computer Sciences*. Cited on pages 12, 13, and 30.

Shaw, M. (2003). Writing good software engineering research papers. In *25th International Conference on Software Engineering, 2003. Proceedings.*, pages 726–736. Cited on page 9.

Shorten, C. e Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):60. Cited on page 11.

Silva, C., Costa, D., Costa, J., e Ribeiro, B. (2024). Data annotation quality in smart farming industry. *Production & Manufacturing Research*, 12(1):2377253. Cited on page 3.

Sishodia, R. P., Ray, R. L., e Singh, S. K. (2020). Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sensing*, 12(19):3136. Cited on page 8.

Sokolova, M., Japkowicz, N., e Szpakowicz, S. (2009). A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437. Cited on page 17.

Tan, M. e Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Cited on pages 11 and 13.

Theisen, C., Dunaiski, M., Williams, L., e Visser, W. (2017). Writing good software engineering research papers: Revisited. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*, pages 402–402. Cited on page 9.

van Marrewijk, B. M., Dandjinou, C., Rustia, D. J. A., Gonzalez, N. F., Diallo, B., Dias, J., Melki, P., e Blok, P. M. (2024a). Active learning for efficient annotation in precision agriculture: a use-case on crop-weed semantic segmentation. Cited on page 1.

van Marrewijk, B. M., Dandjinou, C., Rustia, D. J. A., Gonzalez, N. F., Diallo, B., Dias, J., Melki, P., e Blok, P. M. (2024b). Active learning for efficient annotation in precision agriculture: a use-case on crop-weed semantic segmentation. Cited on page 13.

Wang, P., Tang, Y., Luo, F., Wang, L., Li, C., Niu, Q., e Li, H. (2022). Weed25: A deep learning dataset for weed identification. *Frontiers in Plant Science*, 13:1053329. Cited on page 9.

Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., e Wang, L. (2025). Vision Transformers for Image Classification: A Comparative Survey. *Technologies*, 13(1):32. Cited on page 10.

Wang, Y. e Stavness, I. (2025). Dynamic-budget superpixel active learning for semantic segmentation. *Frontiers in Artificial Intelligence*, 7:1498956. Cited on page 2.

Weiss, M., Jacob, F., e Duveiller, G. (2020). Remote sensing for agricultural applications: A meta-review. *Remote Sensing of Environment*, 236:111402. Cited on page 8.

Yang, C., Huang, L., e Crowley, E. J. (2024). Plug and play active learning for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17784–17793. Cited on pages 31, 34, and 35.

Yousif, M. J. (2024). Enhancing The Accuracy of Image Classification Using Deep Learning and Preprocessing Methods. *Artificial Intelligence & Robotics Development Journal*. Cited on page 11.

Yujian, Y., Wu, Y., Yang, X., Liu, F., e Liao, Y. (2022). Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493:626–646. Cited on pages 33 and 34.

Zahidi, U. A. e Cielniak, G. (2021). Active Learning for Crop-Weed Discrimination by Image Classification from Convolutional Neural Network's Feature Pyramid Levels. In Vincze, M., Patten, T., Christensen, H. I., Nalpantidis, L., e Liu, M., editors, *Computer Vision Systems*, volume 12899, pages 245–257. Springer International Publishing, Cham. Cited on pages 1 and 2.

Zhang, J., Yu, F., Zhang, Q., Wang, M., Yu, J., e Tan, Y. (2024a). Advancements of UAV and Deep Learning Technologies for Weed Management in Farmland. *Agronomy*, 14(3):494. Cited on page 13.

Zhang, X., Zhang, W., Wei, X.-Y., Wu, J., Zhang, Z., Lei, Z., e Li, Q. (2024b). Generative Active Learning for Image Synthesis Personalization. arXiv:2403.14987. Cited on page 28.

Zhang, Y., Wang, M., Zhao, D., Liu, C., e Liu, Z. (2023). Early weed identification based on deep learning: A review. *Smart Agricultural Technology*, 3:100123. Cited on page 8.

Zhang, Y., Zhao, S., Gu, H., e Mazurowski, M. A. (2025). How to efficiently annotate images for best-performing deep learning-based segmentation models: An empirical study with weak and noisy annotations and segment anything model. *J Imaging Inform Med*. Cited on pages 33 and 34.