

FEDERAL UNIVERSITY OF UBERLÂNDIA – UFU
INSTITUTE OF AGRICULTURAL SCIENCES – ICIAG

ARTHUR FAUSTINO DE ARAUJO TEIXEIRA

ENHANCING SOYBEAN MATURITY CLASSIFICATION USING HIGH-THROUGHPUT
UAV RGB IMAGING AND CONVOLUTIONAL NEURAL NETWORKS

MONTE CARMELO – MG

BRAZIL

2025

ARTHUR FAUSTINO DE ARAUJO TEIXEIRA

ENHANCING SOYBEAN MATURITY CLASSIFICATION USING HIGH-THROUGHPUT
UAV RGB IMAGING AND CONVOLUTIONAL NEURAL NETWORKS

Thesis submitted to the Graduate Program in
Agriculture and Geospatial Information at the
Federal University of Uberlândia, Monte
Carmelo *campus*, as part of the requirements
for obtaining the Master's degree.

Concentration area: Development and
applications of methods in geospatial
information

Advisor: Prof. PhD. Murillo Guimarães
Carneiro

MONTE CARMELO – MG

BRAZIL

2025

International Cataloguing-in-Publication Data (CIP)
UFU Library System, MG, Brazil.

- T266e
2025
- Teixeira, Arthur Faustino de Araujo, 1995-
Enhancing soybean maturity classification using highthroughput UAV RGB
imaging and convolutional neural networks [electronic resource] / Arthur Faustino
de Araujo Teixeira. - 2025.
- Advisor: Murillo Guimarães Carneiro.
Dissertation (Master's) - Federal University of Uberlandia, Postgraduate
Program in Agriculture and Geospatial Information.
Access Mode: Internet.
Available at: <http://doi.org/10.14393/ufu.di.2025.5563>
Includes bibliography.
Includes details.
1. Agronomy. 2. Deep learning (Machine learning). 3. Unmanned aerial
vehicles. I. Carneiro, Murillo Guimarães, 1988-, (Orient.). II. Federal University
of Uberlandia. Postgraduate Program in Agriculture and Geospatial Information.
III. Title.

CDU: 631

Rejâne Maria da Silva
Librarian-Documentalist – CRB6/1925



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Agricultura e Informações
Geoespaciais
Rodovia LMG 746, Km 01, s/nº, Bloco 1AMC, Sala 1A202, Monte Carmelo-MG, CEP 38.500-000
Telefone: (34) 3810-1033 - ppgaig@iciag.ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Agricultura e Informações Geoespaciais				
Defesa de:	Dissertação de Mestrado Acadêmico				
Data:	27/08/2025	Hora de início:	14h00	Hora de encerramento:	16h40
Matrícula do Discente:	32312AIG004				
Nome do Discente:	Arthur Faustino de Araújo Teixeira				
Título do Trabalho:	Enhancing Soybean Maturity Classification Using High-Throughput UAV RGB Imaging And Convolutional Neural Networks				
Área de concentração:	Informações geoespaciais e tecnologias aplicadas à produção agrícola				
Linha de pesquisa:	Desenvolvimento e aplicações de métodos em informações geoespaciais				
Projeto de Pesquisa de vinculação:	Inteligência artificial aplicada à agricultura e informações geoespaciais				

Reuniu-se na sala virtual, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Agricultura e Informações Geoespaciais, composta pelo prof. Dr. Murillo Guimarães Carneiro (Universidade Federal de Uberlândia-UFU), orientador do candidato, prof. Dr. Zhao Liang (Universidade de São Paulo - USP), e Dr. Filipe Inácio Matias (Syngenta Seeds).

Iniciando os trabalhos o presidente da mesa, Dr. Murillo Guimarães Carneiro, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu ao discente a palavra para a exposição do seu trabalho. A duração da apresentação do discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Murillo Guimarães Carneiro, Professor(a) do Magistério Superior**, em 27/08/2025, às 16:41, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Zhao Liang, Usuário Externo**, em 27/08/2025, às 16:55, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Filipe Inácio Matias, Usuário Externo**, em 03/09/2025, às 09:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site

[https://www.sei.ufu.br/sei/controlador_externo.php?](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0)

[acao=documento_conferir&id_orgao_acesso_externo=0](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **6594276** e o código CRC **49DE2594**.

To my beloved wife Ana Luisa

ACKNOWLEDGEMENTS

I express my profound gratitude to God for life and guidance; to my beloved wife Ana Luisa, for her unconditional love and constant support that made this journey possible; to the memory of my mother Josirene Gomes de Araujo Teixeira, whose teachings remain as a foundation in my life; to my father Sérgio Faustino Teixeira, for his example of integrity and confidence in my potential; to my family, a safe harbor in all moments; to Professor Murillo Carneiro, for his valuable guidance and opportunities for academic growth; to my friends, for their understanding and encouragement; to Syngenta Seeds, for the institutional support, especially to my managers Diego Almeida and Daniel Gieseeman for their flexibility and encouragement; and to all who contributed directly or indirectly to my education and to the completion of this work, my sincere recognition.

BIOGRAPHY

Arthur Faustino de Araujo Teixeira is a dedicated professional with extensive experience in agricultural technology, automation, and software development. Arthur holds a Bachelor's degree in Mechanical Engineering from the University of São Paulo (2016-2020), where he consistently ranked first in his class. He completed his Master's degree in Agriculture and Geospatial Information from the Federal University of Uberlândia (2023-2025), graduating with highest honors. He also earned a Post-Graduate Specialization in Automation and Control of Industrial and Agroindustrial Processes from the State University of Campinas (UNICAMP) (2021-2022).

Currently serving as a Global Field Systems Engineer at Syngenta Seeds (since October 2023), Arthur develops software solutions to improve real-time data streaming for agricultural operations from planting to harvesting. He previously worked as an Automation Specialist for Latin America at the same company (2022-2023), managing agricultural machinery automation projects and integrating embedded systems with automated data transmission to local servers.

Before joining Syngenta, Arthur worked at Jacto for two years (2020-2022), first as an Engineering Intern and later as a Process Engineer, where he was responsible for testing electronic components in agricultural machinery and automating test benches. His experience in the agricultural sector has given him valuable insights into the practical challenges of implementing technological solutions in farming operations.

Arthur has a strong research background, having worked on projects funded by FAPESP (São Paulo Research Foundation) and ANP (National Petroleum Agency). His research focused on multiphase flow measurements and oil-water flow dynamics, resulting in multiple publications and presentations at academic seminars. During his Master's program, he was a member of the AINET research group, where he developed machine learning solutions based on network architectures for agricultural applications, including automated crop monitoring and high-throughput phenotyping using computer vision and UAV imagery.

His Master's research focused on developing RMNet-E, an innovative convolutional neural network architecture for classifying soybean maturity using UAV RGB images, contributing to the advancement of deep learning applications in precision agriculture. This foundation in research methodology has proven invaluable in his current academic and professional pursuits.

Throughout his career, Arthur has demonstrated leadership abilities, serving as Coordinator of Mechanical and Mechatronic Engineering at EESC jr., Project Leader, and Financial Director for the Mechanical Engineering Week event. His collaborative approach and strong communication skills have enabled him to work effectively with cross-functional and international teams across Latin America.

Arthur is passionate about leveraging technology to enhance agricultural efficiency and sustainability, with a commitment to advancing smart farming practices through cutting-edge technological solutions. His technical expertise includes strategic project leadership, cross-functional team management, innovation strategy, data-driven decision making, Agile project management, automation and control engineering, process control, software development, precision agriculture, deep learning, computer vision, high-throughput phenotyping, machine learning applications in agriculture, and agricultural machinery automation. He is fluent in both Portuguese and English at a native or bilingual level and possesses an intermediate level of proficiency in Spanish.

Arthur's goal is to contribute to the future of smart farming through cutting-edge AI and precision agriculture solutions, with aspirations to advance into management roles where he can lead technological innovation in the agricultural sector.

SUMMARY

1 Introduction	2
2 Literature Review	5
2.1 Global Soybean Production and Breeding.....	5
2.2 Understanding Soybean Relative Maturity and Growth Stages	7
2.3 Genetic and Environmental Influences in Soybean Maturity.....	10
2.4 High-Throughput Phenotyping in Plant Breeding.....	15
2.5 Digital Image Analysis and Processing in Modern Agriculture	17
2.6 Machine Learning Fundamentals and Implementation in Agricultural Systems	20
2.6.1 Machine Learning Paradigms	21
2.6.2 Machine Learning Basics: Preprocessing and Evaluation.....	22
2.6.3 Machine Learning in Agriculture	25
2.7 Convolutional Neural Networks in Agricultural Applications	27
2.7.1 Convolutional Neural Networks.....	31
2.7.2 Other Deep Learning Techniques in Agriculture	36
2.8 Previous Studies in Soybean Maturity Classification	37
3 Materials and Methods	43
3.1 Experimental Design and Data Collection	43
3.2 Data Processing	49
3.3 Model Development and Training.....	52
3.4 Model Evaluation and Comparative Analysis	55
4 Results and Discussion	57
4.1 Single Location Analysis	57
4.1.1 Ablation Study	58
4.1.2 Comparative Results Against Traditional Machine Learning Techniques ...	60

4.1.3	Comparative Results Against Others Deep Learning Models	62
4.1.4	General Discussion	64
4.2	Cross-Location Analysis	69
4.2.1	Comparative Analysis of Base Model and Environmental Integration	71
4.2.2	Comparative Analysis Against Traditional Machine Learning Approaches	73
4.3	General Discussion	80
4.4	Limitations and Future Work.....	82
4.4.1	Current Technical and Operational Limitations	83
4.4.2	Environmental Coverage and Adaptation Challenges	83
4.4.3	Future Research Directions and Potential Improvements.....	84
5	Conclusions	86
	References	89

LIST OF FIGURES

FIGURE 1. Soybean phenological scale showing the complete developmental sequence from emergence through reproductive stages, with distinct vegetative (V) and reproductive (R) phases (Fehr; Caviness, 1977).....	8
FIGURE 2. Conceptual illustration of the complex relationship among environments, genotypes, and phenotypes, demonstrating how genetic expression is modified by environmental conditions to produce observable plant characteristics.....	11
FIGURE 3. Comprehensive image processing workflow for agricultural applications, showing the sequential steps from image acquisition through georeferencing, preprocessing, plot extraction, segmentation, feature extraction, and final analysis.....	18
FIGURE 4. Classification framework of machine learning systems according to the amount and type of supervision during training, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches.....	21
FIGURE 5. Canonical overfitting behavior graph showing the relationship between model complexity and error rates, with the characteristic U-shaped test error curve illustrating the transition from underfitting (high bias) to overfitting (high variance) as model complexity increases (Goodfellow; Bengio; Courville, 2016).....	24
FIGURE 6. Deep neural network modeling framework illustrating the three essential components: optimizer for model improvement, loss function for performance measurement, and monitoring metrics for evaluation during training and testing (Chollet, 2021).	29
FIGURE 7. Architectural illustrations of (a) a typical convolutional neural network showing the sequence of convolutional, pooling, and fully connected layers, and (b) the convolution operation with a 3×3 kernel, stride of 1, and no padding, demonstrating how feature maps are generated (He <i>et al.</i> , 2023).....	32
FIGURE 8. Dropout regularization technique visualization showing how randomly deactivating neurons during training creates an ensemble effect that improves model generalization by preventing co-adaptation of features (Goodfellow; Bengio; Courville, 2016).....	35
FIGURE 9. Temporal distribution of soybean maturity classes (Immature, Near-mature, and Mature) by days after planting in (A) Location 01 and (B) Location 02, showing the progression of maturity throughout the growing season.....	44
FIGURE 10. Temporal distribution of soybean maturity classes (Immature, Near-mature, and Mature) by days after planting in (A) Location 01 and (B) Location 02, showing the progression of maturity throughout the growing season. Percentages indicate the proportion of plots in each maturity class at specific timepoints, demonstrating the transition from immature plots to predominantly mature plots by late season, with a critical transition period occurring between 80 and 100 days after planting (DAP).	45
FIGURE 11. Comprehensive environmental parameter variations across flight dates organized by variable type in (A) Location 01 and (B) Location 02: (a) Soil moisture profiles at different depths (surface, root zone, and profile), (b) Meteorological variables including temperature, relative humidity, and specific humidity, (c) Solar radiation components including all-sky and clear-sky shortwave radiation, top-of-atmosphere radiation, direct normal irradiance, and diffuse radiation, and (d) Photosynthetically active radiation (PAR) and ultraviolet radiation variables (UVA and	

UVB). The temporal progression demonstrates distinct variability patterns across parameter groups: soil moisture shows high stability (CV 2-4%), meteorological variables exhibit moderate variation (CV 13-37%), solar radiation components display the highest variability (CV 5-43%), and PAR/UV variables show moderate fluctuation (CV 5-141%). Data encompasses 11 flight dates for Location 01 and 9 flight dates for Location 02, creating 20 distinct micro-environments across the growing season. 47

FIGURE 12. Comprehensive data preprocessing pipeline illustrating the sequential transformation of raw UAV imagery into standardized model inputs: (a) raw aerial orthomosaic of the experimental field; (b) overlay of plot boundary grid derived from GPS coordinates; (c) zoomed region showing detailed plot delineation; (d) integration of unique plot identifiers with boundary data; and (e) extracted individual plot image ready for standardization and model input. This systematic approach ensures consistent data preparation while preserving the spatial integrity of each experimental unit. 50

FIGURE 13. RMNet architecture diagram showing the sequential convolutional neural network design with two convolutional layers (64 and 128 filters), max-pooling operations, flattening, and dense layers optimized for soybean maturity classification. 53

FIGURE 14. Complete data analysis workflow diagram illustrating the systematic approach to model training, validation, and evaluation, including dataset partitioning, cross-validation strategy, and performance metric calculation. 54

FIGURE 15. Confusion matrices from four high-performing classification models: (a) Xception, (b) ResNet152V2, (c) Support Vector Machine, and (d) Ridge Regression, visualizing their classification performance across the three maturity classes: (0) Immature, (1) Near-mature, and (2) Mature. Values represent raw counts with percentages of total observations in parentheses. Note the varying performance patterns across different maturity classes, particularly in the critical Mature class. 65

FIGURE 16. Temporal progression of model performance metrics across the soybean growing season comparing RMNet (baseline model) and RMNet-E (environmentally-enhanced model): (a) Accuracy, (b) F1-score, (c) Precision, and (d) Recall. The x-axis represents Days After Planting (DAP), while the y-axis shows the respective performance metric values (0-1 scale). 72

FIGURE 17. Temporal evolution of model performance metrics comparing RMNet-E against traditional machine learning approaches (SVM, Ridge Regression, Decision Trees, Random Forest, Perceptron, LinearSVC, and kNN) throughout the soybean growing season: (a) Accuracy, (b) F1-Score, (c) Precision, and (d) Recall. The x-axis represents Days After Planting (DAP), while the y-axis shows performance metric values (0-1 scale). 75

FIGURE 18. Temporal evolution of mature class precision comparing RMNet-E against traditional machine learning models (SVM, Ridge Regression, Decision Trees, Random Forest, Perceptron, LinearSVC, and kNN) throughout the soybean growing season. The x-axis represents Days After Planting (DAP), while the y-axis shows precision values (0-1 scale) specifically for the mature class. 77

LIST OF TABLES

TABLE 1. Environmental parameters used in the study showing parameter abbreviations, descriptions, units of measurement, and temporal resolution for data collection.	46
TABLE 2. Classes definition for soybean maturity classification, including class labels (Immature, Near-mature, Mature), mathematical definitions based on relative maturity date (RM_{date}) and flight date (Fl_{date}), and descriptive explanations.	52
TABLE 3. Performance metrics of RMNet model variants showing accuracy, precision, recall, and F1-score results according to different combinations of convolutional and dense layers during architecture optimization.	58
TABLE 4. Statistical analysis results showing p-values from Levene's test for homogeneity of variances and ANOVA test for significant differences in accuracy, precision, recall, and F1-score across model configurations.	58
TABLE 5. Computational performance analysis showing processing time per epoch (in seconds) for different RMNet architectures according to the number of convolutional and dense layers...	59
TABLE 6. Confusion matrices for RMNet model variants with different numbers of convolutional and dense layers, showing classification performance across the three maturity classes: (0) Immature, (1) Near-mature, and (2) Mature. Values represent raw counts with percentages of total observations below. (continue).....	59
TABLE 7. Comparative performance metrics between RMNet and traditional machine learning models, showing accuracy, precision, recall, and F1-score for soybean maturity classification. .	60
TABLE 8. Computational efficiency comparison between RMNet and traditional machine learning models, showing training and prediction times in seconds.	61
TABLE 9. Comparative performance metrics between RMNet and state-of-the-art deep learning architectures, showing accuracy, precision, recall, and F1-score for soybean maturity classification.	63
TABLE 10. Computational efficiency comparison between RMNet and other deep learning models, showing processing time per epoch in seconds.	63
TABLE 11. Precision comparison for mature class detection across all tested models, highlighting RMNet's superior performance in identifying plots ready for harvest.	65
TABLE 12. Visual comparison of soybean maturity classification performance across different models (Xception, Support Vector Machine, and RMNet) against ground truth, showing both maturity maps and mature class detection.	67
TABLE 13. Mature plant detection performance during the critical harvest window (DAP 94-107) comparing traditional machine learning approaches (SVM, Ridge Regression, Random Forest, kNN) with RMNet-E. The table presents precision values specifically for the mature class and associated variance measures. Higher precision indicates better accuracy in identifying mature plants, while lower variance indicates greater consistency in predictions across different environmental conditions. Note that RMNet-E achieves both the highest precision and lowest variance, demonstrating superior reliability for harvest timing decisions in practical breeding applications.	79

TEIXEIRA, Arthur Faustino de Araujo. Enhancing Soybean Maturity Classification Using High-Throughput UAV RGB Imaging and Convolutional Neural Networks. 2025, 108 p. Dissertation (Master's in Agriculture and Geospatial Information) – Federal University of Uberlândia, Monte Carmelo Campus, Minas Gerais, Brazil. Available at: <<http://doi.org/10.14393/ufu.di.2025.5563>>.

ABSTRACT

This study introduces RMNet (Relative Maturity Neural Network) and its environmentally-enhanced variant RMNet-E, efficient convolutional neural network architectures for soybean maturity classification using UAV-based RGB imagery. RMNet achieves 98.2% precision for mature plot identification while processing $3.3\times$ to $20.4\times$ faster than comparable deep learning models, addressing critical bottlenecks in breeding programs that traditionally rely on labor-intensive manual evaluations. The research utilized over 52,000 experimental plots and 18,000 unique genotypes, representing one of the most genetically diverse datasets used in soybean maturity classification studies to date, across two locations, creating 20 distinct micro-environments to evaluate model transferability. RMNet demonstrated exceptional overall performance across all maturity classes with 90.4% accuracy, 93.5% precision, 96.8% recall, and 95.1% f1-score, significantly outperforming both traditional machine learning (Support Vector Machine) and deep learning approaches (Xception) in mature class detection. RMNet-E integrates meteorological, radiation, and soil parameters to enable cross-location deployment, maintaining consistent performance when transferred between training and testing environments. During the critical harvest window, RMNet-E achieved 99.3% precision for mature plant detection with the lowest prediction variance among all evaluated approaches, demonstrating superior reliability for harvest timing decisions across diverse environmental conditions. Comprehensive statistical validation using ANOVA and Tukey tests across 13 modeling approaches confirms RMNet's optimal balance of computational efficiency and classification accuracy. The cross-location validation simulates real-world deployment scenarios where models must transfer between different environmental conditions, addressing practical implementation challenges in breeding programs. This research provides a scalable solution for automated soybean maturity classification with proven transferability across diverse agricultural environments, enabling optimized harvest timing and accelerated genetic gain in breeding operations.

Keywords: Deep learning; High-throughput phenotyping; Precision agriculture; Soybean maturity classification; Unmanned aerial vehicles.

1 INTRODUCTION

Soybean (*Glycine max* L. Merrill) stands as one of the world's most significant agricultural commodities, projected to reach a global production of 420.87 million metric tons in 2024/2025. This remarkable growth trajectory, representing a 6% increase from the previous year and maintaining a 10-year compound average growth of 3%, underscores the crop's critical role in global food security and agricultural economics. Brazil leads global production at 40% (169 million tons), followed by the United States at 28% (118.84 million tons), highlighting the concentration of production among major agricultural nations.

The successful development of high-performing soybean varieties faces numerous challenges across diverse growing environments. Breeders must contend with variations in latitude, climate, altitude, soil conditions, and agricultural practices, while simultaneously addressing growth patterns, environmental stresses, and disease pressures. These complex interactions create substantial genotype-environment effects that significantly impact breeding program outcomes. The genetic diversity within breeding populations further complicates this challenge, as thousands of distinct genotypes with varying maturity characteristics must be evaluated simultaneously across multiple experimental plots. This research specifically addresses this complexity by incorporating over 12,000 unique genotypes across more than 50,000 experimental plots, representing one of the most genetically diverse datasets used in soybean maturity classification studies to date. Traditional phenotyping approaches, while foundational to breeding programs, present significant limitations in modern agricultural contexts. Manual assessment methods are typically costly, labor-intensive, and often destructive to plant materials. Moreover, the subjective nature of visual evaluations can introduce inconsistencies in data collection, particularly when scaling across large breeding programs. These limitations create bottlenecks in the breeding pipeline, potentially slowing genetic gain and reducing program efficiency. The heritability of maturity traits in the experimental populations tested in this study demonstrates the strong genetic basis for these characteristics, yet capturing this information through traditional methods remains challenging and inefficient, particularly when evaluating thousands of plots simultaneously.

High-throughput phenotyping (HTP) technologies, particularly when combined with deep learning approaches, offer promising solutions to these challenges. The integration of unmanned aerial vehicles (UAVs), advanced imaging systems, and sophisticated data analysis techniques provides opportunities for rapid, non-destructive, and objective phenotypic data collection.

However, significant gaps remain in current approaches to soybean maturity classification. Existing methods often rely on complex multispectral imagery (Zhou *et al.*, 2019), require extensive feature engineering (Hu *et al.*, 2023), or demonstrate limited transferability across different environments (Zhang *et al.*, 2022). Furthermore, many current models sacrifice computational efficiency for accuracy (Moeinizade *et al.*, 2022) or lack the precision needed for mature plot identification, which is critical for harvest timing decisions.

A particularly challenging aspect of high-throughput phenotyping for soybean maturity classification is the dynamic nature of field conditions across different locations and temporal sequences. This research specifically addresses this challenge by incorporating data from multiple locations captured across different days throughout the growing season, resulting in significant variations in environmental conditions, lighting, plant development stages, and image characteristics. This environmental variability, combined with the genetic diversity of the experimental populations, creates a complex classification problem that demands robust, adaptable modeling approaches. The temporal dynamics of plant development further complicate this challenge, as the visual characteristics of maturity stages evolve throughout the growing season, requiring models that can effectively capture these transitions.

This research addresses these limitations through the development and validation of Relative Maturity Neural Network (RMNet), a novel convolutional neural network (CNN) architecture specifically designed for soybean maturity classification. We hypothesize that a carefully optimized CNN architecture using only RGB imagery can achieve comparable or superior performance to more complex approaches while maintaining computational efficiency and environmental adaptability. Furthermore, we propose that the integration of environmental data can enhance model robustness across diverse growing conditions without significantly increasing computational requirements.

The study has four primary objectives:

- i. Develop an efficient and accurate deep learning model for soybean maturity classification using UAV-based RGB imagery.
- ii. Evaluate the model's performance across multiple locations and environmental conditions through the integration of environmental data.
- iii. Compare the model's effectiveness against both traditional machine learning approaches and contemporary deep learning architectures.

- iv. Demonstrate the practical applicability and computational efficiency of the proposed solution in real-world breeding scenarios.

This research contributes to the advancement of agricultural technology by providing a practical, efficient solution for automated soybean maturity classification. The development of RMNet represents a significant step toward modernizing breeding programs through the integration of high-throughput phenotyping and artificial intelligence, potentially accelerating genetic gain and improving efficiency in the breeding program. By addressing the limitations of existing approaches while accounting for the complex interactions between genetic diversity, environmental variability, and temporal dynamics, this work fills critical gaps in the current literature on high-throughput phenotyping for soybean breeding programs.

The dissertation is organized into five chapters. Following this introduction, Chapter 2 provides a comprehensive literature review covering global soybean production, maturity classification systems, high-throughput phenotyping technologies, and current applications of machine learning in agriculture. Chapter 3 details the materials and methods, including experimental design, data collection protocols, and model development procedures. Chapter 4 presents the results and discussion, examining model performance across different Locations and conditions while comparing it with existing approaches. Finally, Chapter 5 concludes the dissertation by summarizing key findings and suggesting future research directions.

2 LITERATURE REVIEW

This literature review provides a comprehensive examination of the key concepts, technologies, and methodologies underpinning soybean maturity classification using high-throughput phenotyping and deep learning approaches. Beginning with an overview of global soybean production and breeding challenges, the review progresses through the complexities of soybean maturity classification systems and the genetic and environmental factors influencing plant development. The chapter then explores the evolution of high-throughput phenotyping technologies and their application in modern agriculture, with particular focus on digital image analysis techniques and processing workflows. The final sections delve into machine learning fundamentals, neural network architectures, and their specific applications in agricultural contexts, highlighting both traditional approaches and advanced deep learning methodologies. Throughout the review, special attention is given to the intersection of these technologies with soybean maturity assessment, establishing the theoretical foundation for the research methodology presented in subsequent chapters.

2.1 Global Soybean Production and Breeding

Soybean (*Glycine max* L. Merrill) stands as one of the world's most significant agricultural commodities, projected to reach a global production of 420.87 million metric tons in 2024/2025, representing a 6% increase from the previous year and maintaining a 10-year compound average growth of 3%. The crop's remarkable expansion is primarily concentrated in a few key nations, with Brazil leading global production at 40% (169 million tons), followed by the United States at 28% (118.84 million tons), and Argentina at 12% (49 million tons) (USDA, 2025). This concentration of production among major players has been driven by increasing demand for both vegetable oil and protein meal, with soybeans now occupying approximately 6% of the world's arable land (Hamza *et al.*, 2024).

The crop's significance extends beyond traditional agricultural boundaries, playing a crucial role in global food security through its high nutritional value and versatility in food products. The remarkable production levels are attributed to both the expansion of cultivated areas and enhanced productivity through technological advancement, including genomic technologies and improved management practices. This progress has been particularly evident in the development of varieties adapted to various environmental conditions and stress factors, supported by extensive research

and development conducted by breeding companies worldwide (Gomes *et al.*, 2016; Hamza *et al.*, 2024).

This substantial growth in soybean production and its expanding role in global agriculture has been made possible through decades of focused breeding efforts. The evolution from traditional farming practices to modern, high-yielding cultivation systems represents the culmination of systematic breeding programs and genetic improvement initiatives.

Plant breeding aims to develop superior cultivars through genetic improvement, with the increase in performance achieved through artificial selection being termed genetic gain (Chawade *et al.*, 2019). The process involves germplasm collection, forming basic populations through hybridization, and selection to achieve superior genotypes (Sulistyo; Purwantoro; Sari, 2018). The concept of evolutionary plant breeding allows nature to select the best-adapted genotypes over time, minimizing the effort required for traditional selection and testing of individual genotypes (Reynolds; Braun, 2022). This genetic improvement results in valued characteristics through a process refined over thousands of years (Suza; Lamkey, 2023).

Maximizing soybean yields requires the application of diverse management strategies to fully exploit the genetic potential of existing cultivars. While accelerating the breeding cycle is one approach to enhance genetic gain, the inherently long growth cycles of soybean present a significant limitation to this strategy (Watson *et al.*, 2018). If certain characteristics could be predicted in advance, it would eliminate the need to wait until the end of the cycle to select the best materials within a population (Yuan *et al.*, 2019). The intricate genetic makeup of soybean necessitates the continued use of traditional breeding methodologies. These time-tested approaches remain crucial for assessing and improving yield potential in this complex crop (Christenson *et al.*, 2016).

The success of plant breeding and research programs hinges on accurate and precise phenotypic data collection. Researchers utilize advanced phenotyping techniques in conjunction with genetic tools to achieve their breeding objectives and research goals (Singh *et al.*, 2016). A persistent challenge in enhancing breeding program efficiency lies in the establishment of high-quality phenotypic data from field trials. The ability to collect reliable, comprehensive phenotypic information continues to be a limiting factor in advancing breeding efforts (Singh; Sethi; Singh, 2021). As the growing season concludes, breeding organizations face critical decisions. They must evaluate which varieties show promise for advancement to the next growing season or

commercialization, and which should be eliminated from further consideration (Moeinizade *et al.*, 2022).

Current breeding programs face constraints from limited resources, technologies, and crop reproductive biology. The challenge of designing a plant breeding program represents an engineering approach to plant science (Suza; Lamkey, 2023). These constraints highlight the need for innovative solutions in modern breeding programs.

These breeding efforts have been crucial in developing varieties that not only increase productivity but also address various agricultural challenges, from pest resistance to climate adaptation. Among these challenges, understanding and managing crop maturity emerges as a fundamental aspect of soybean breeding and production. The timing of maturity directly impacts yield potential, resource utilization, and adaptation to specific growing regions, making it a critical consideration in modern breeding programs.

2.2 Understanding Soybean Relative Maturity and Growth Stages

Maturity in agronomic crops is typically classified into two types: physiological maturity (cessation of reproductive growth) and harvest maturity (when the economically important organ reaches "ripeness") (Bruns, 2009). Relative maturity is a classification system that encompasses all environmental and genetic factors influencing a plant's developmental timeline, specifically measuring the period from planting to physiological maturity (Alliprandini *et al.*, 2009).

The relative maturity rating system serves as a standardized method for comparing different soybean cultivars, even when they haven't been directly evaluated in the same trials. This system is particularly valuable for breeding programs and farmers, as it provides a reliable framework for predicting crop development across various environmental conditions. When selecting cultivars, if all other characteristics are equal, early-maturity varieties are often preferred due to their shorter growing season, which can reduce risk exposure and allow for more flexible crop rotation strategies (Trevisan *et al.*, 2020).

The selection of ideal cultivars for specific regions requires careful consideration of local environmental conditions and agricultural practices. The optimal cultivar should maximize yield potential by fully utilizing the available growing season while avoiding delayed harvest, which can increase both risks and costs. This balance is particularly critical in regions with multiple cropping systems or where environmental stresses are common during specific periods of the growing season (Trevisan *et al.*, 2020).

To fully understand how maturity classifications translate into actual plant development, it's essential to examine the detailed growth stages that characterize the soybean lifecycle. These stages provide precise benchmarks for monitoring plant development and making management decisions throughout the growing season.

Understanding the growth stages of soybean plants is crucial for monitoring their development throughout the growing season. A widely adopted and comprehensive classification system, depicted in Figure 1, was introduced by Fehr and Caviness (1977). This system categorizes the soybean lifecycle into two main phases: vegetative (V) and reproductive (R). With the exception of VE (emergence) and VC (cotyledon) stages, each phase is further subdivided using numerical indices to denote specific development milestones.

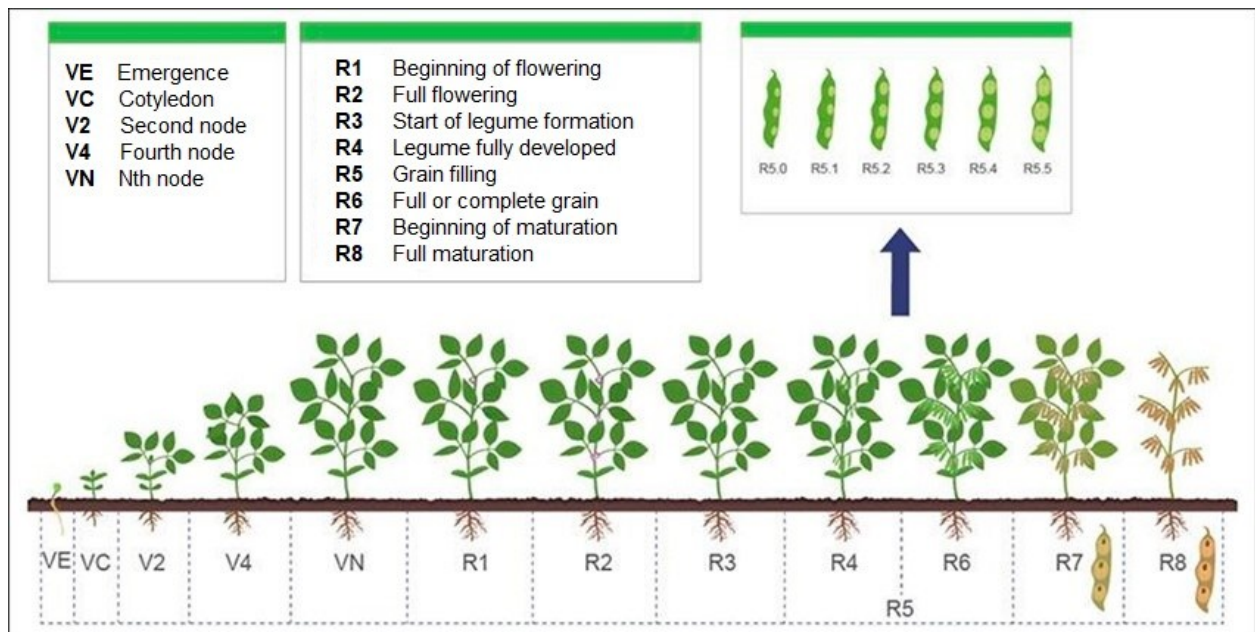


FIGURE 1. Soybean phenological scale showing the complete developmental sequence from emergence through reproductive stages, with distinct vegetative (V) and reproductive (R) phases (Fehr; Caviness, 1977).

The soybean's journey begins with stage VE, marked by the emergence of cotyledon leaves. This is followed by stage VC, characterized by fully expanded cotyledons and the initial separation of unifoliate leaf edges. Subsequent vegetative stages are denoted numerically (V1, V2, V3, etc.), corresponding to the number of nodes with fully developed leaves. The final vegetative stage, VN, represents the topmost node with a fully expanded leaf. As the plant transitions from maximum vegetative growth to reproductive development, it enters a series of eight stages. These can be

grouped into four distinct phases: flowering (R1-R2), pod formation (R3-R4), seed development (R5-R6), and plant maturation (R7-R8) (Neumaier *et al.*, 2000).

The onset of the R7 stage, or initial maturity, is characterized by pods turning gray, beige, or brown. As the plant approaches its final stages, leaf chlorophyll content diminishes, resulting in a yellow/gray coloration. This physiological change significantly alters the plant's spectral reflectance properties. Consequently, plant maturity can be assessed by analyzing specific wavelength bands that are responsive to canopy characteristics such as chlorophyll content, nitrogen levels, and water stress (Zhou *et al.*, 2019).

The understanding and accurate classification of soybean maturity represents a crucial intersection between breeding objectives and practical agriculture. For breeding programs, maturity classification serves as a fundamental tool for developing adapted varieties that optimize yield potential within specific environmental constraints. Breeders utilize this knowledge to develop comprehensive crossing strategies that effectively combine desired maturity traits with other important agronomic characteristics.

This approach enables the development of varieties specifically adapted to diverse geographical regions and cropping systems, while simultaneously creating cultivars capable of responding to changing climatic conditions. A key challenge in this process is balancing the trade-offs between early maturity and yield potential, as these traits often exhibit complex interactions. These considerations have become increasingly relevant as breeding programs work to address the challenges of climate change and the need for sustainable intensification of agricultural production.

Developing commercial soybean cultivars with varying maturity levels poses significant challenges for breeders worldwide. This complexity arises from the interplay of diverse factors including wide-ranging latitudes, climatic variations, altitude differences, soil type diversity, varied farming and planting methods, plant growth patterns, the presence or absence of long-juvenile traits, varying stress conditions, and disease pressures. These factors collectively contribute to substantial genotype-environment interactions (Alliprandini *et al.*, 2009).

Modern genomic analysis enables precise quantification of variation among individuals and populations. This technology supports quality control, maintenance of gene banks, distribution of genetic resources, identification of desired traits, development of improved varieties. These applications are crucial for addressing global food security challenges (Halewood *et al.*, 2018).

Soybean is a short-day species with varying sensitivity levels, enabling broad adaptability across latitudes (Bruns, 2009). Nine major genes control flowering and maturity timing, with flowering typically suppressed under long-daylength conditions (Watanabe; Harada; Abe, 2012). Harvest maturity varies between seasons due to environmental conditions, particularly temperature, relative humidity, and cloud cover (Bruns, 2009).

While maturity classification provides a practical framework for variety selection and breeding decisions, the underlying complexity of maturity traits stems from intricate interactions between genetic factors and environmental conditions. Understanding these interactions is crucial for developing adapted varieties and predicting crop performance across diverse growing conditions.

2.3 Genetic and Environmental Influences in Soybean Maturity

Plant phenotypes represent observable characteristics resulting from genotype-environment interactions and serve as indicators of plant performance under specific conditions. While a plant's genotype remains constant, its phenotype evolves over time, potentially enabling early-season traits to predict end-of-season outcomes such as yield, which could significantly accelerate crop breeding through earlier plant selection (Pasala; Pandey, 2020; Yuan *et al.*, 2019).

The primary abiotic factors influencing soybean phenology are temperature and photoperiod. Temperature generally exhibits a positive influence on crop development rate, while photoperiod modifies this temperature response. As a short-day plant, soybean's development rate decreases with longer daylength, which delays reproductive development (Setiyono *et al.*, 2007). Additionally, water availability plays a crucial role in flowering and maturation cycles, with temperature being closely related to altitude and geographical position (Ferreira, 2021).

The complexity of genotype-by-environment (G×E) interactions necessitate a deeper understanding of the phenome - the physical and biochemical traits of organisms as determined by their genetic makeup and environmental influences. The phenome represents a complex system with multiple organizational levels, from genes to populations, making it crucial for understanding plant responses under adverse conditions. While traditionally defined from a genetic perspective, modern phenomics considers the phenome as an intricate system that operates across various organizational levels, from molecular to population-wide responses (Lorence; Medina Jimenez, 2022).

The relationship between genotype and phenotype, illustrated in Figure 2, for quantitatively expressed traits is complicated by G×E. This phenomenon, defined as the differential behavior of genotypes across environments, assumes a critical role in phenotypic manifestation. When predominantly qualitative or crossed, G×E interactions can hinder the selection process and cultivar recommendations, as genotype rankings may change across assessment environments (Krause *et al.*, 2025; Tavares De Albuquerque *et al.*, 2022).

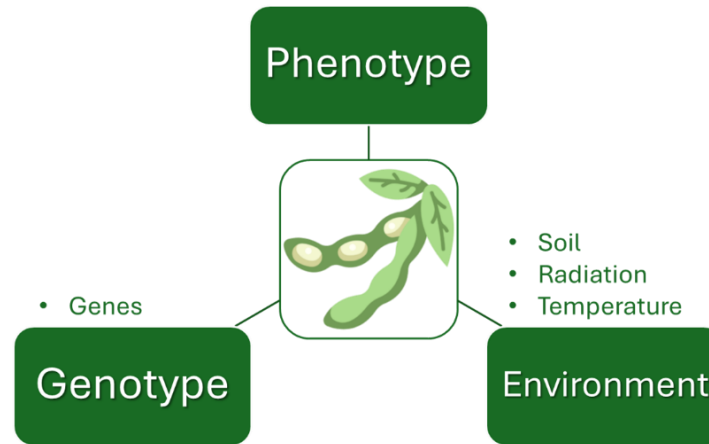


FIGURE 2. Conceptual illustration of the complex relationship among environments, genotypes, and phenotypes, demonstrating how genetic expression is modified by environmental conditions to produce observable plant characteristics.

Plant phenotyping, an emerging science linking genomics with plant ecophysiology and agronomy, involves comprehensive assessment of complex traits including development, growth, resistance, tolerance, physiology, architecture, yield, and ecology (Gill *et al.*, 2022; Xu; Li, 2022). The phenotypic value of an individual can be modeled using linear components, expressed as

$$P = \mu + G + E, \quad (1)$$

where P represents the phenotypic value, μ the overall mean, G the genotypic effect, and E the environmental effect (Suza; Lamkey, 2023).

Network modeling has become widespread in systems biology and integrative genomics, with plant breeders primarily relying on linear phenotypic models. The determination of fixed or random effects in these models influences variance component estimation and significantly affects overall phenotypic variability analysis. Fixed effects represent deliberately arranged components, while random effect models provide inferences for broader populations. (Alonso; Stepanova, 2015; Suza; Lamkey, 2023).

The source of phenotypic variability determines whether trait selection will result in a heritable response that can be passed to subsequent generations. Within a population, the phenotypic variance (V_P) equals the sum of genotypic variance (V_G) and environmental variance (V_E) (Suza; Lamkey, 2023).

Heritability, a fundamental concept in breeding, can be expressed in two ways. Broad-sense heritability

$$H = \frac{V_G}{V_P} \quad (2)$$

represents the ratio of total genetic variance to phenotypic variance. Narrow-sense heritability

$$h^2 = \frac{V_A}{V_P} \quad (3)$$

represents how phenotypes are determined by genes transmitted from parents, where V_A is the additive genetic variance referring to the deviation from mean phenotype due to specific allele inheritance. In field experiments, plant breeders often report heritability on an entry-mean basis using the formula

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \left(\frac{\sigma_{GE}^2}{E}\right) + \left(\frac{\sigma_\epsilon^2}{rE}\right)} \quad (4)$$

where σ_G^2 is genotypic variance, σ_{GE}^2 is genotype by environment variance, σ_ϵ^2 is error variance, r is replication number, and E represents environment number (Miller; Song; Li, 2023; Sulistyo; Purwantoro; Sari, 2018; Suza; Lamkey, 2023).

Heritability values provide crucial insights into the relative influence of genetic and environmental factors on phenotypic expression. A heritability value approaching 0 indicates that phenotypic variation is predominantly influenced by environmental factors, while a value approaching 1 suggests strong genetic control over the trait's expression. For practical breeding applications, heritability values are typically categorized into three ranges: low ($h^2 < 0.2$), medium ($0.2 < h^2 < 0.5$), and high ($h^2 > 0.5$). These classifications serve as valuable guidelines for plant breeders in developing selection strategies, as they indicate the degree to which trait expression will be consistently inherited across generations versus being modified by environmental conditions. Traits with higher heritability values generally respond more effectively to selection pressure, making them more reliable targets for breeding programs (Sulistyo; Purwantoro; Sari, 2018).

The relationship between replication number and heritability estimation is particularly evident in the denominator of the entry-mean heritability formula, where the experimental error is weighted by the number and replications. As the number of replications decreases, the error component increases proportionally, leading to a larger denominator and consequently lower heritability estimates. This effect becomes especially pronounced in multi-stage breeding programs, where early-generation materials typically have fewer replications due to limited seed availability or resource constraints. This mathematical relationship underscores the importance of considering replication structure when comparing heritability estimates across different stages of a breeding program or when making selection decisions based on these estimates.

Research has consistently demonstrated the critical importance of environmental considerations in breeding programs across different geographical regions. A comprehensive study in Ethiopia examining twenty soybean genotypes across six environments revealed highly significant environment, genotype, and G×E effects on grain yield, oil, and crude protein content. The study found that environmental variance and G×E captured the largest portion of total sum of squares, highlighting the substantial influence of environmental conditions on soybean performance under Ethiopian conditions (Gurmu; Mohammed; Alemaw, 2010).

In Mato Grosso State, Brazil, a six-year study across nine Locations with multiple sowing dates demonstrated that Locations were significantly more crucial than years for yield testing. The research revealed highly significant Location effects across all six years and three maturity groups, with significant Location × cultivar interactions suggesting challenges in selecting for wider genotype adaptation. The study concluded that breeding for broad adaptation in Mato Grosso State was an efficient strategy for cultivar development, recommending increased investment in testing Locations rather than years to enhance experimental precision (Tavares De Albuquerque *et al.*, 2022).

Further supporting these findings, research in Pakistan examining 96 soybean genotypes across different ecological regions demonstrated substantial genetic diversity for grain yield and identified environment-specific adaptations. This study identified 13 environment-specific genotypes and broadly adapted varieties which showed superior performance across multiple environments. Through advanced DNA technology and association mapping, the research identified 26 marker-trait associations for six agronomic traits, with significance levels ranging for plant height and for hundred-grain weight (Rani *et al.*, 2023).

These findings collectively emphasize the complexity of genotype-environment interactions and their crucial role in soybean breeding programs worldwide, highlighting the importance of multi-environment testing and the potential of combining traditional breeding approaches with modern genomic tools.

For modeling and interpretation of predictive results, reliable phenotypic data collection is essential. However, when screening hundreds or thousands of genotypes daily, there is a high probability of errors, especially in field conditions. Without adequate environmental characterization, even the generation of multidimensional information may not achieve consistent models. Genotypic sensitivity to macro- and microclimate conditions is recorded at each assessment; for example, spectral signatures of genotypes vary according to the environmental characteristics under which measurements are performed (Lorence; Medina Jimenez, 2022).

Modern phenotyping faces several challenges, particularly when screening large numbers of genotypes. While simple trait phenotyping can be accomplished by untrained personnel, complex trait assessment requires experienced professionals and is time-intensive. Visual phenotyping of large genotype sets often leads to measurement errors due to fatigue and subjective assessments (Reynolds; Braun, 2022). Digital phenotyping offers advantages through data reanalysis capabilities and objective quantitative analyses of large image sets (Sozzani; Benfey, 2011).

The integration of genetic and environmental influences in soybean breeding represents a complex challenge that requires sophisticated approaches to phenotyping and data analysis. As breeding programs continue to evolve, the combination of traditional field-based observations with modern digital phenotyping technologies offers promising opportunities for accelerating genetic gain. Understanding these interactions is crucial for developing resilient cultivars adapted to specific environmental conditions while maintaining desired agronomic traits. Future breeding efforts will likely depend increasingly on precise environmental characterization and high-throughput phenotyping methods to effectively select superior genotypes across diverse growing conditions (Chawade *et al.*, 2019).

The complexity of genetic and environmental interactions in soybean development necessitates advanced approaches to phenotype assessment. High-throughput phenotyping (HTP) has emerged as a transformative solution, addressing the limitations of traditional phenotyping

methods while enabling the evaluation of large populations under diverse environmental conditions.

2.4 High-Throughput Phenotyping in Plant Breeding

HTP has emerged as a transformative approach in plant breeding, addressing the limitations of traditional phenotyping methods. Conventional approaches, which rely on manual sampling, are typically costly, labor-intensive, and destructive, with inherent subjectivity in visual assessments compromising data accuracy and limiting throughput (Alabi *et al.*, 2022; Sankaran *et al.*, 2018; Xu; Li, 2022). This is particularly challenging in large-scale breeding programs where thousands of genotypes require evaluation, making manual field measurements impractical and time-consuming (Lorence; Medina Jimenez, 2022).

Field trials in plant breeding typically involve evaluating thousands of breeding lines and clones in small plots ranging from 1 to 10 meters in length. The major challenges include measuring numerous plots across multiple environments while optimizing resources, time, and data quality. High-throughput phenotyping directly addresses these challenges by increasing selection intensity and improving selection accuracy, ultimately accelerating genetic gain in breeding programs (Chawade *et al.*, 2019). This acceleration of genetic gain is crucial for developing improved varieties that can meet growing agricultural demands while adapting to changing environmental conditions.

The concept of HTP encompasses field-based platforms designed to deliver the necessary throughput for large-scale experiments while providing accurate trait performance assessment in real-world environments (Trevisan *et al.*, 2020). These technologies enable consistent, rapid, and extensive data acquisition beyond manual measurement capabilities, generating data on thousands of phenotypes per monitored unit (Alabi *et al.*, 2022; Chang *et al.*, 2021; Lopez-Cruz *et al.*, 2020).

Forward phenomics, a key application of HTP, uses advanced phenotyping tools to identify the most promising genotypes within large collections, enabling selection of elite lines with desirable traits. This approach is particularly valuable for phenotyping diverse plant populations, including germplasm collections, breeding populations, and mapping populations under varied growth conditions (Pasala; Pandey, 2020).

The rapid evolution of technologies for generating and analyzing large quantities of genotypic, phenotypic, and environmental data has revolutionized biological research. Halewood *et al.* (2018) emphasize that these technological advances, particularly in high-throughput assays

and genetic material synthesis, are transforming our understanding of plant biology and accelerating breeding progress. This technological revolution has enabled the development of increasingly sophisticated phenotyping platforms that combine multiple sensing technologies with advanced data processing capabilities.

Unmanned aerial vehicles (UAVs) have become particularly valuable in breeding programs requiring simultaneous evaluation of numerous genotypes (Johansen *et al.*, 2019). Equipped with various sensors including digital cameras, multispectral, hyperspectral, thermal infrared, and lidar devices, UAVs enable precise screening of hundreds of plots in a short timeframe (Ren *et al.*, 2023). Their ability to operate at low altitudes, collect data rapidly and accurately, and maintain functionality under varying atmospheric conditions makes them especially useful (Franchini *et al.*, 2018), though battery capacity limitations can restrict their utility for very large-scale operations (Gill *et al.*, 2022).

HTP integrates various methodologies to capture information related to plant performance, function, and structure. The development of RGB (red, green and blue) sensors and other imaging technologies has revolutionized plant phenotyping, allowing visualization of plant ontogeny, anatomy, physiology, and biochemistry throughout development stages (Lorence; Medina Jimenez, 2022). These systems can vary from ground-based imaging to aerial phenotyping and remote sensing, often incorporating artificial intelligence for data analysis (Gill *et al.*, 2022).

The integration of these technologies with breeding programs has particular relevance for soybean cultivation. For instance, physiological maturity date, a crucial selection criterion, can be more efficiently monitored through HTP systems. This is especially important as extended maturation can maximize yield potential, though delayed harvesting may impact seed quality (Vergara *et al.*, 2019). The interaction between maturity timing and latitude adds another layer of complexity, as early-maturing varieties from high latitudes may experience premature flowering in low-latitude regions, while late-maturing varieties may face frost risk when grown at high latitudes (Song *et al.*, 2019).

Recent years have witnessed significant advancement in technologies for generating and analyzing large quantities of genotypic, phenotypic, and environmental data (Halewood *et al.*, 2018). The emergence of intelligent agriculture has sparked increased interest in leveraging deep learning methods for soybean yield prediction, phenotype evaluation, and breeding research (Guo *et al.*, 2022). These developments, combined with the integration of remote sensing technologies

and phenotyping platforms, are enabling more efficient and accurate phenotyping with stable error across all genotypes, whether as single plants or in micro-plots (Reynolds; Braun, 2022).

The effectiveness of high-throughput phenotyping systems relies heavily on sophisticated image analysis techniques. The evolution of digital imaging and processing methods has revolutionized agricultural monitoring, enabling rapid and precise phenotypic assessment at unprecedented scales.

2.5 Digital Image Analysis and Processing in Modern Agriculture

The evolution of non-invasive, affordable sensors and imaging techniques has revolutionized plant phenomics, with remote sensing emerging as a cornerstone of agricultural monitoring (Gill *et al.*, 2022). Remote sensing, which involves deriving information about Earth's surfaces using electromagnetic radiation data collected from an overhead perspective, has its roots in early photography from the 1800s. By the mid-2000s, the convergence of various technologies - including remote sensing, Geographic Information Systems (GIS), and Global Positioning Systems (GPS) - created integrated systems capable of acquiring high-precision imagery and combining diverse data forms. This technological integration, coupled with the decentralization of information technology services and aerial imagery acquisition, has fostered innovation in specialized agricultural applications (Campbell; Wynne, 2011).

The fundamental components of digital image analysis in agriculture encompass three primary domains: spectral analysis, spatial analysis, and temporal analysis. Spectral analysis examines the interaction between electromagnetic radiation and plant tissues across different wavelengths, providing insights into crop health, nutrient status, and physiological conditions. Spatial analysis focuses on the geometric relationships and patterns within images, while temporal analysis tracks changes over time, crucial for monitoring crop development stages and stress responses (Campbell; Wynne, 2011).

The history of agricultural imaging dates back to the 1850s, when the first large-scale landscape photos were captured from balloons, establishing the foundation for modern analytical techniques. These early applications have evolved into sophisticated workflows involving spectral differentiation, radiometric differentiation, spatial differentiation, and geometric transformations (Lorence; Medina Jimenez, 2022).

Image standardization and calibration represent critical steps in agricultural image analysis. Radiometric calibration corrects for atmospheric effects and sensor variations, while geometric

calibration ensures spatial accuracy and proper alignment of multi-temporal images. Standard procedures include the use of calibration panels, ground control points, and automated image registration techniques. Quality assessment metrics typically include spatial resolution verification, radiometric accuracy, and geometric precision measurements (Wang; Myint, 2015).

Image composition and quality play crucial roles in agricultural analysis. Standard RGB cameras, the most widely used phenotyping sensors, capture images in three color channels, while some specialized applications, such as satellite imagery, may include additional frequencies like infrared (Géron, 2019). However, field conditions can affect image quality due to natural illumination variations (Xu; Li, 2022).

RGB image processing in agricultural applications follows a systematic workflow, as illustrated in Figure 3. The process begins with georeferencing, where raw image data is combined with metadata (geospatial information) to create accurate orthomosaic. Pre-processing follows, involving color space transformations such as RGB to HSV (Hue, Saturation, Value) or LAB color spaces, which enhance the separation of plant features from background elements. Plot extraction then identifies specific areas of interest within the orthomosaic, enabling focused analysis of individual experimental units. The segmentation phase isolates plant material from background elements, while feature extraction captures relevant characteristics including texture, shape, and spectral indices. Finally, post-processing techniques refine the extracted data and prepare it for analysis. This comprehensive workflow enables the generation of detailed phenotypic data for agricultural applications, with each step contributing to the overall accuracy and reliability of the analysis (Hamuda; Glavin; Jones, 2016).

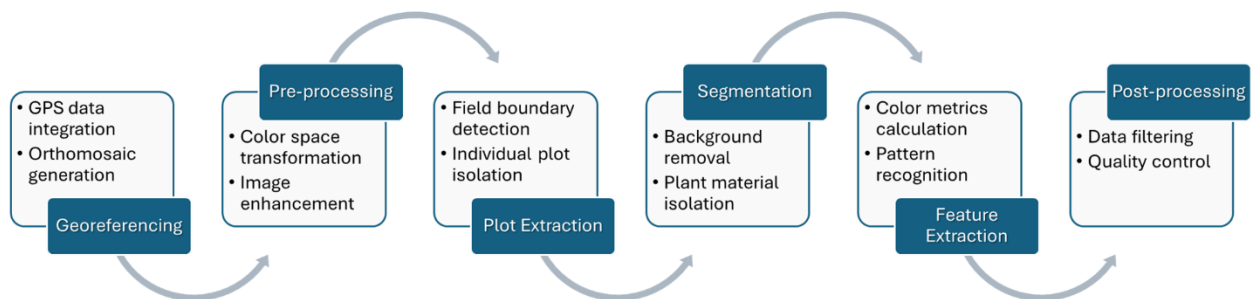


FIGURE 3. Comprehensive image processing workflow for agricultural applications, showing the sequential steps from image acquisition through georeferencing, preprocessing, plot extraction, segmentation, feature extraction, and final analysis.

Data management in agricultural image analysis presents unique challenges due to the volume, variety, and velocity of data generation. Modern phenotyping platforms can generate terabytes of data daily, requiring robust storage systems, efficient data compression techniques, and automated quality control procedures. Implementing standardized metadata schemas and automated data validation protocols helps ensure data integrity and accessibility (Kamilaris; Prenafeta-Boldú, 2018).

For phenotyping applications, proper georeferencing of sensor data is essential to ensure images can be accurately linked to specific plots during post-processing workflows (Lorence; Medina Jimenez, 2022). The implementation of machine learning methods, particularly with UAV-collected data, has enhanced data processing and prediction capabilities. High spatial resolution UAV-based remote sensing imagery, with resolutions between 0 and 10 cm, has become the primary data source for machine learning approaches, supporting both classification and regression analyses (Reynolds; Braun, 2022).

Feature classification in agricultural image analysis employs both traditional computer vision techniques and modern deep learning approaches. Spectral features include vegetation indices (such as NDVI, EVI, and SAVI), textural features derived from gray-level co-occurrence matrices, and shape-based features describing crop morphology. These features serve as inputs for various classification algorithms, enabling automated crop monitoring and phenotyping (Ram *et al.*, 2024).

One of the primary challenges in widespread HTP adoption is the analysis and interpretation of the vast amount of data generated. The field generates terabytes of data daily, necessitating automated processing through machine learning and deep learning programs (Gill *et al.*, 2022). This challenge extends to the development of specialized software tools for remote sensing data processing, extraction, and analysis, which has been identified as a potential bottleneck in generating high-quality phenotypic data. These tools must integrate various aspects, including crop growth performance assessment, phenological stage detection, agronomical yield component analysis, and pest and disease identification (Reynolds; Braun, 2022).

The integration of marker-assisted and genomic selection with image analysis has emerged as a powerful tool in modern plant breeding. High-throughput image analysis provides phenotypic data that, when combined with genetic information, enables more precise selection of desired traits.

This integration has particularly benefited the development of crops with improved stress tolerance and yield stability (Araus; Kefauver, 2018).

Quality assessment in agricultural image analysis follows a multi-tiered approach. At the acquisition level, factors such as spatial resolution, temporal frequency, and spectral fidelity are evaluated. During processing, quality metrics include segmentation accuracy, feature extraction reliability, and classification performance. The final analysis stage incorporates validation against ground-truth data and cross-validation with multiple datasets to ensure robust results (Detring *et al.*, 2024).

Recent advances in spectral analysis techniques have expanded beyond traditional RGB imaging. Hyperspectral imaging, which captures hundreds of narrow spectral bands, enables detailed chemical and physiological analysis of crops. Multi-temporal analysis combines images from different growth stages to track developmental changes and stress responses. These advanced techniques require sophisticated calibration procedures and standardized processing workflows to ensure data compatibility and reliability (Sarić *et al.*, 2022).

The vast amount of data generated through modern imaging techniques creates both opportunities and challenges in agricultural research (Hu *et al.*, 2023). Machine learning has emerged as a crucial tool for extracting meaningful patterns and insights from these complex datasets, enabling automated analysis and prediction in agricultural applications (Ren *et al.*, 2023).

2.6 Machine Learning Fundamentals and Implementation in Agricultural Systems

The rapid adoption of HTP platforms in agriculture has generated enormous volumes of diverse data, creating significant challenges in data management and interpretation. Machine learning (ML) and deep learning (DL) have emerged as powerful solutions for extracting useful information from these complex datasets, offering interdisciplinary approaches that combine probability, statistics, classification, regression, decision theory, data visualization, and neural networks (Gill *et al.*, 2022). At its core, ML is essentially applied statistics with increased emphasis on computer-based function estimation (Goodfellow; Bengio; Courville, 2016), representing a multidisciplinary approach that combines probability theory, decision theory, visualization, and optimization (Gill *et al.*, 2022).

Artificial Intelligence (AI), the broader field encompassing machine learning, represents the effort to automate intellectual tasks typically performed by humans (Chollet, 2021). Within this framework, machine learning requires three fundamental components: input data points, examples

of expected output, and a way to measure algorithm performance (Chollet, 2021). This approach enables the discovery of patterns in massive datasets by simultaneously analyzing multiple traits, rather than examining each feature in isolation (Gill *et al.*, 2022).

ML provides significant advantages to plant breeders, pathologists, physiologists, and agronomists by enabling the extraction and analysis of multiple parameters simultaneously. Its success lies in its ability to infer trends from data and generalize results through model training, efficiently identifying underlying structures, patterns, and relationships within datasets for classification and prediction tasks (Gill *et al.*, 2022). Machine learning is particularly valuable for problems requiring extensive fine-tuning, complex problem-solving, adaptation to fluctuating environments, and analysis of large datasets (Géron, 2019).

2.6.1 Machine Learning Paradigms

Machine learning systems can be classified into four major categories based on their supervision levels during training shown in Figure 4: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning (Géron, 2019). Among these, supervised and unsupervised learning represent the two primary approaches (Gill *et al.*, 2022).

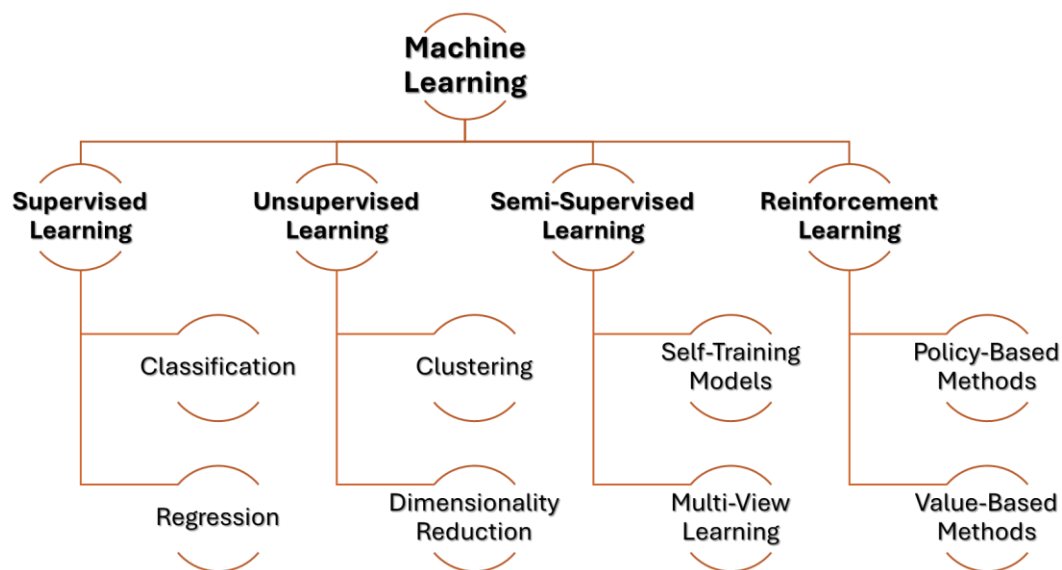


FIGURE 4. Classification framework of machine learning systems according to the amount and type of supervision during training, including supervised, unsupervised, semi-supervised, and reinforcement learning approaches.

In supervised learning, algorithms work with labeled datasets where each example is associated with a target outcome. The training process involves mapping relationships between

input and output variables, typically addressing either classification tasks (categorizing output variables) or regression problems (predicting real values) (Gill *et al.*, 2022). Key supervised learning algorithms include k-Nearest Neighbors, Linear Regression, Logistic Regression, Support Vector Machines (SVMs), Decision Trees, Random Forests, and Neural Networks (Géron, 2019).

Conversely, unsupervised learning algorithms work with unlabeled data, attempting to discover inherent patterns and structures within the dataset (Goodfellow; Bengio; Courville, 2016). This approach encompasses various techniques including clustering methods (such as K-Means, DBSCAN, and Hierarchical Cluster Analysis), anomaly detection, dimensionality reduction (including Principal Component Analysis and t-SNE), and association rule learning (Géron, 2019). Since labeling data is often time-consuming and costly, some algorithms can handle partially labeled data through semi-supervised learning approaches (Géron, 2019).

Reinforcement Learning represents a distinct category where an agent learns optimal behavior through interaction with its environment, receiving rewards or penalties based on its actions. The agent must develop a policy - a strategy for choosing actions in given situations - to maximize long-term rewards (Géron, 2019).

Machine learning systems can also be classified based on their learning method: batch learning requires training on all available data simultaneously, while online learning enables incremental learning from sequential data streams. The learning rate in online systems determines how quickly they adapt to new data, with higher rates enabling rapid adaptation but potentially forgetting older patterns, and lower rates providing more stability but slower learning (Géron, 2019).

2.6.2 Machine Learning Basics: Preprocessing and Evaluation

Data preprocessing represents a critical step in agricultural machine learning applications, particularly given the diverse nature of agricultural data. One of the most fundamental transformations required is feature scaling, as machine learning algorithms typically perform poorly when input numerical attributes have significantly different scales (Géron, 2019). This is especially relevant in agricultural applications where features may range from small-scale measurements like leaf thickness (millimeters) to large-scale measurements like field size (hectares).

Two primary approaches to feature scaling are commonly employed: min-max scaling and standardization. Min-max scaling, often referred to as normalization, transforms values to range

from 0 to 1 by subtracting the minimum value and dividing by the range (maximum minus minimum). This approach is particularly useful for agricultural image processing, where pixel values need to be normalized for consistent model performance. Standardization, alternatively, centers the data around zero by subtracting the mean and dividing by the standard deviation, resulting in unit variance. While standardization doesn't bound values to a specific range, it proves more robust against outliers, which are common in agricultural data due to environmental variations (Géron, 2019).

A critical aspect of machine learning is the model's ability to generalize - performing well on new, previously unseen inputs rather than just the training data (Goodfellow; Bengio; Courville, 2016). This creates a fundamental tension between optimization (improving performance on training data) and generalization (maintaining performance on new data) (Chollet, 2021).

A typical ML model implementation involves a careful calibration process where the model is trained on a designated training set and validated on a separate testing set (Gill *et al.*, 2022). To properly evaluate model performance, data is typically split into three sets: training, validation, and test sets. The validation set helps prevent information leaks and enables model tuning, while the test set provides the final evaluation of model performance before deployment (Chollet, 2021).

Two common validation approaches are simple holdout validation and k-fold validation. In holdout validation, a portion of data is set aside as the test set, with the remaining data used for training and validation. K-fold validation divides data into K equal partitions, training the model K times using different combinations of partitions, which is particularly useful when model performance shows significant variance based on the train-test split (Chollet, 2021).

Most ML algorithms require careful selection of hyperparameters - settings that control the learning algorithm's behavior but are not learned during training. These include learning rate, regularization constants, dropout ratio, and model architecture. The complexity of modern ML models makes selecting optimal hyperparameter combinations increasingly challenging (Cho; Hegde, 2019). Typically, about 80% of the training data is used for actual training, with 20% reserved for validation to guide hyperparameter selection (Goodfellow; Bengio; Courville, 2016).

Two central challenges in machine learning can be seen in Figure 5: underfitting and overfitting. Underfitting occurs when the model fails to achieve sufficiently low error rates on the training set, indicating it's too simple to capture the underlying data structure. Overfitting happens when the model performs well on training data but fails to generalize, showing a large gap between

training and test errors (Goodfellow; Bengio; Courville, 2016). To combat overfitting, regularization techniques can be employed, including early stopping - halting training when validation error begins to increase (Géron, 2019).

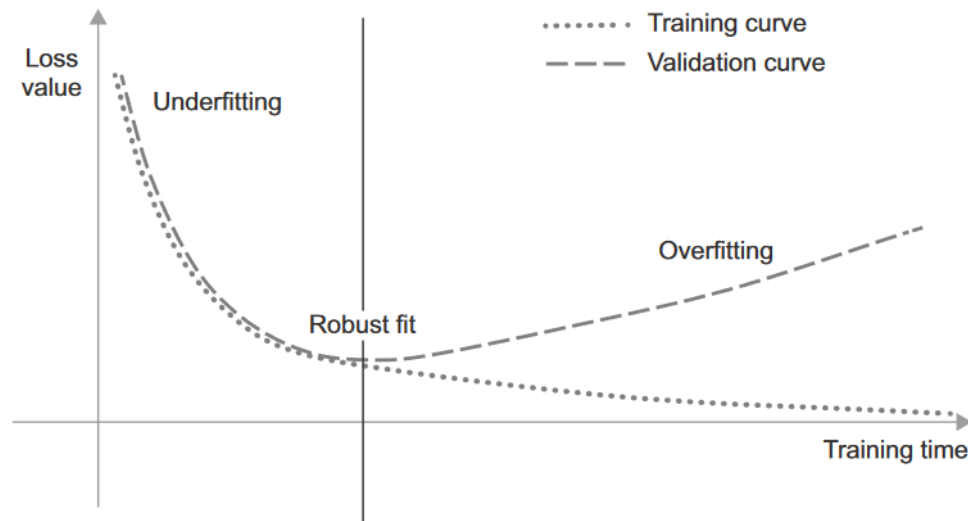


FIGURE 5. Canonical overfitting behavior graph showing the relationship between model complexity and error rates, with the characteristic U-shaped test error curve illustrating the transition from underfitting (high bias) to overfitting (high variance) as model complexity increases (Goodfellow; Bengio; Courville, 2016).

During the initial phases of model training, there is a direct relationship between model optimization and its ability to generalize: as the model's performance improves on the training dataset (shown by decreasing loss), it similarly improves on the test dataset. During this phase, the model is considered underfit, meaning it has not yet fully captured or learned the essential patterns within the training data, indicating room for improvement. However, as training continues and iterations increase, the model's ability to generalize reaches a plateau and subsequently deteriorates – this marks the onset of overfitting. At this point, the model begins to memorize specific characteristics or noise from the training data rather than learning meaningful patterns, making these learned features unreliable or meaningless when applied to new, unseen data (Chollet, 2021).

Several performance metrics are used to evaluate machine learning models, often differing from the cost functions used during training (Goodfellow; Bengio; Courville, 2016). For classification tasks, accuracy - the proportion of correctly classified examples - serves as a basic metric:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives. However, more nuanced metrics are often necessary for comprehensive evaluation.

Precision and recall are two fundamental metrics used in classification tasks. Precision measures the accuracy of positive predictions, calculated as the ratio of true positives to all positive predictions (Géron, 2019; Goodfellow; Bengio; Courville, 2016)

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

Recall, also known as sensitivity or true positive rate (TPR), measures the proportion of actual positive instances correctly identified (Géron, 2019; Goodfellow; Bengio; Courville, 2016)

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

To combine precision and recall into a single metric, the F1 score is commonly used. This score represents the harmonic mean of precision and recall:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

The F1 score particularly favors classifiers with balanced precision and recall. However, this balance isn't always desirable, as some applications may prioritize one metric over the other. This creates the precision/recall trade-off, where improving one metric typically comes at the cost of degrading the other (Géron, 2019; Goodfellow; Bengio; Courville, 2016).

The success of machine learning in agricultural systems ultimately depends on careful model selection, appropriate data preparation, proper validation procedures, and thorough performance evaluation. As agricultural data continues to grow in volume and complexity, machine learning techniques will become increasingly vital for extracting meaningful insights and supporting decision-making in agricultural research and practice.

2.6.3 Machine Learning in Agriculture

Agricultural machine learning applications face unique implementation challenges. Field data collection often involves dealing with missing values due to sensor failures, weather interference, or incomplete measurements. Additionally, agricultural data typically exhibits strong temporal and spatial correlations, requiring careful consideration during the train-test split to avoid data leakage (Kumar *et al.*, 2018).

In agricultural applications, feature scaling becomes particularly crucial when dealing with multiple data sources. For example, when combining drone imagery data (measured in reflectance values) with ground sensor data (measuring soil moisture, temperature, and nutrient levels), proper scaling ensures all features contribute proportionally to the model's learning process (Wang; Myint, 2015).

Recent research has demonstrated the versatility of machine learning approaches in soybean cultivation and assessment. Multiple regression models have shown promise for yield prediction, with Cubist and Random Forest achieving particularly strong results ($r^2 = 0.89$) against Extreme Gradient Boosting ($r^2 = 0.68$), Stochastic Gradient Boosting ($r^2 = 0.59$) and Support Vector Machine ($r^2 = 0.70$), when integrating UAV-derived spectral bands, canopy height, and texture features (Alabi *et al.*, 2022). Neural Network sensitivity analysis has successfully identified key meteorological factors influencing harvest dates (Niedbała *et al.*, 2022), while Random Forest, Neural Network, and Support Vector Machine algorithms have demonstrated high accuracy (>89%) in segmenting seed maturation stages using spectral imaging (Batista *et al.*, 2022).

For soybean maturity assessment specifically, several techniques have shown effectiveness. Multiple Linear Regression models using both wavebands ($r^2 = 0.43$) and vegetation indices ($r^2 = 0.50$) have provided valuable maturity predictions, with waveband approaches offering greater simplicity (Christenson *et al.*, 2016). Partial Least Square Regression utilizing fractional vegetation cover and leaf chlorophyll content from UAV-RGB imagery has achieved impressive accuracy (>98%) in maturity monitoring (Hu *et al.*, 2023). Non-Parametric Local Polynomial Regression has also demonstrated strong performance ($r^2 = 0.84$ – 0.97) when analyzing vegetation indices from UAV imagery (Volpato *et al.*, 2021).

The integration of maturity group information with remote sensing data has significantly enhanced soybean yield estimation through various regression techniques, including Partial Least Square Regression, Gaussian Process Regression, Random Forest Regression and Kernel Ridge Regression (Ren *et al.*, 2023). UAV-based multispectral imagery has proven practical for estimating maturity dates across hundreds of breeding lines using Partial Least Square Regression with cross-validation (Zhou *et al.*, 2019). Random Forest models have successfully performed binary maturity classification (accuracy >93%) using temporal multispectral data (Yu *et al.*, 2016).

More sophisticated approaches include Artificial Neural Networks for tropical soybean genotype classification by maturity group (Amaral *et al.*, 2022) and multilayer regression networks for predicting days to maturity ($r^2 = 0.66$) using multitemporal-multispectral datasets (Teodoro *et al.*, 2021).

Despite these advancements, gaps remain in the current state of the science. Most studies have focused on integrating multispectral imagery and traditional vegetation indices, with waveband models showing limitations due to their complexity. Although Partial Least Square Regression and other models have achieved high accuracy in monitoring soybean maturity, there is a need for more streamlined and efficient approaches.

While traditional machine learning approaches have demonstrated significant value in agricultural applications, the complexity of phenotypic data and the need for automated feature extraction have led to increased adoption of deep learning methods. Deep neural networks, particularly convolutional architectures, offer powerful solutions for processing and analyzing agricultural imagery.

2.7 Convolutional Neural Networks in Agricultural Applications

Deep learning represents a specialized branch of machine learning and artificial intelligence that has evolved significantly over recent decades, drawing heavily from our understanding of human brain functionality, statistics, and applied mathematics. While these systems can process vast amounts of data, they still fall considerably short of human brain capabilities (Altalak *et al.*, 2022). The field has seen remarkable growth due to advances in computing power, larger datasets, and improved training techniques (Goodfellow; Bengio; Courville, 2016).

The fundamental philosophy of deep learning is based on connectionism: while individual neurons or features may not exhibit intelligence independently, their collective behavior can demonstrate intelligent patterns. The effectiveness of neural networks has significantly improved since the 1980s, primarily due to the dramatic increase in network size and complexity (Goodfellow; Bengio; Courville, 2016).

Deep learning represents a new approach to learning data representations through successive layers of increasingly meaningful representations. The term "deep" refers to these multiple layers of representations rather than any deeper understanding achieved by the approach. The layered representations are learned through neural networks structured in literal layers stacked upon each other (Chollet, 2021).

Unlike traditional machine learning approaches that require extensive feature engineering - a laborious and expertise-dependent process - deep learning automates the feature extraction process. In shallow learning methods, input data typically needs to be transformed into one or two successive representation spaces through simple transformations. These approaches proved insufficient for complex problems that required more refined representations, forcing human experts to manually engineer appropriate layers of data representation. This feature engineering process was not only time-consuming but also heavily reliant on domain expertise and often resulted in suboptimal solutions. Deep learning, in contrast, automates this entire process by learning all necessary features in a single pass, eliminating the need for manual feature engineering. This automation has greatly simplified machine learning workflows, often replacing sophisticated multi-stage pipelines with single, end-to-end models (Chollet, 2021).

In agricultural applications, deep learning has demonstrated superior performance compared to conventional machine learning methods, particularly in plant phenotyping applications, where the automatic feature extraction capability has proven especially valuable in handling complex visual patterns and characteristics (Gill *et al.*, 2022). This shift from manual to automated feature extraction has democratized machine learning applications in agriculture, allowing researchers to focus more on problem definition and result interpretation rather than on the technical intricacies of feature design.

The learning process in deep learning involves mapping inputs (such as images) to specific targets (such as labels) through a series of data transformations. These transformations are stored in the layer's weights - essentially numerical values that parameterize the layer's operation. Learning occurs by finding optimal values for these weights across all network layers to correctly map inputs to their associated targets (Chollet, 2021).

To control and optimize this learning process, deep learning systems employ a loss function (also called objective or cost function) that measures the distance between the network's predictions and the desired outputs. This function provides a feedback signal that allows the network to adjust its weights incrementally through a process called backpropagation. Initially, the network's weights are randomly assigned, resulting in random transformations and high loss scores. As the network processes more examples, the weights are gradually adjusted to minimize the loss function, eventually resulting in a trained network capable of making accurate predictions (Chollet, 2021).

The deep neural networking modeling represented in Figure 6 involves three crucial components:

- 1) An optimizer - the mechanism for model self-improvement based on training data.
- 2) A loss function - the means of measuring performance on training data.
- 3) Monitoring metrics - such as accuracy measurements during training and testing.

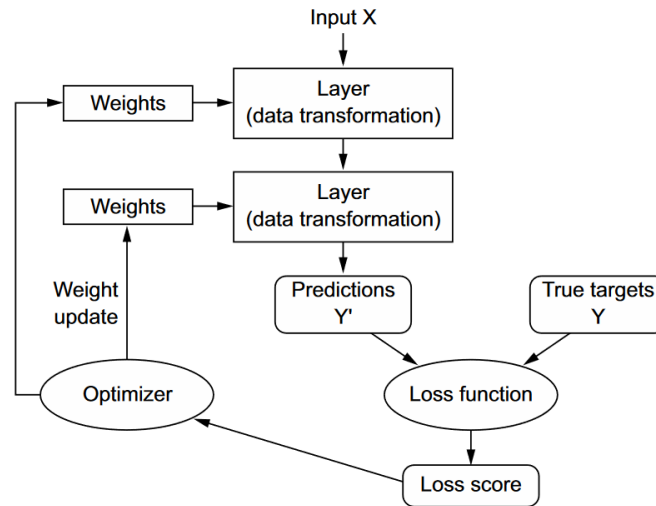


FIGURE 6. Deep neural network modeling framework illustrating the three essential components: optimizer for model improvement, loss function for performance measurement, and monitoring metrics for evaluation during training and testing (Chollet, 2021).

Deep learning models are typically overparameterized, meaning they have more degrees of freedom than necessary to fit the data. However, this overparameterization isn't problematic because training is interrupted before reaching minimum possible training loss. Finding the optimal point during training - the boundary between underfitting and overfitting - is crucial for improving generalization (Chollet, 2021).

Preprocessing requirements for deep learning are relatively minimal compared to traditional approaches. While images must be formatted to a standard size through cropping or scaling, other preprocessing steps are often unnecessary when working with large datasets and models. The system can learn to handle various types of data variability automatically (Chollet, 2021).

The size and complexity of neural networks make hardware infrastructure a critical consideration in deep learning applications. While traditional neural networks were trained using single-machine CPUs, modern deep learning typically requires more sophisticated computing

resources. Contemporary approaches primarily utilize GPU computing or networked CPU clusters, as neural networks involve large parameter buffers, activation values, and gradient values that must be updated during each training step (Goodfellow; Bengio; Courville, 2016).

GPUs offer significant advantages over CPUs due to their superior memory bandwidth, which becomes crucial when parameter buffers exceed traditional desktop computer cache sizes. Neural networks are particularly well-suited for GPU processing due to their minimal branching requirements and the independent nature of neuron processing within layers. This architectural compatibility allows neural networks to effectively leverage GPU parallelism (Goodfellow; Bengio; Courville, 2016).

In agricultural applications, particularly for deployment in field conditions, inference optimization becomes crucial. While training costs may be significant, they are typically one-time expenses, whereas inference costs affect each end-user interaction. For applications without personalization requirements, a single trained model can serve billions of users, making inference efficiency particularly important for resource-constrained end-users (Goodfellow; Bengio; Courville, 2016).

The development of deep learning architectures often follows an evolutionary rather than purely designed approach, with improvements emerging through iterative experimentation. This evolutionary nature requires careful validation through ablation studies - systematic investigations where system components are selectively removed or replaced to understand their contributions to overall performance (Chollet, 2021).

Ablation studies serve as a low-effort but effective method for understanding causality within deep learning systems. When a system combining components X, Y, and Z shows good performance, testing various combinations (X, Y, Z, X+Y, X+Z, Y+Z) helps identify the true sources of performance improvements. This approach is particularly valuable in agricultural applications, where understanding the contribution of different input features (such as spectral bands or temporal sequences) can inform both model optimization and data collection strategies (Chollet, 2021).

In the agricultural domain, numerous studies have explored the application of deep learning techniques. Reviews of these research efforts have consistently demonstrated that deep learning approaches outperform traditional machine learning methods in terms of accuracy. Comprehensive surveys of various deep neural network architectures used in agriculture have been conducted,

providing an overview of the current state-of-the-art technologies. These analyses have revealed that convolutional neural networks are the most frequently adopted foundational structure in agricultural applications (Altalak *et al.*, 2022; Gill *et al.*, 2022; Kamilaris; Prenafeta-Boldú, 2018; Mehta *et al.*, 2021; Ren; Kim; Jeong, 2020).

While CNNs remain dominant in agricultural image analysis, it is worth noting that the field continues to evolve with newer architectural approaches. Recent research has explored transformer architectures, originally designed for natural language processing but adapted to image recognition tasks through models like Vision Transformer (ViT) (Dosovitskiy *et al.*, 2021). Unlike CNNs, which process images through local convolutional filters with fixed receptive fields, transformers employ a self-attention mechanism that allows global modeling of relationships across the entire image. This capability has shown promise in agricultural applications requiring long-range spatial context or multi-modal data integration, particularly when pre-trained on large datasets (Wang *et al.*, 2024). However, transformers typically require more computational resources than CNNs and may underperform when training data is limited, making CNNs still the preferred choice for many practical agricultural applications, especially in resource-constrained environments. Hybrid approaches that combine CNN and transformer components have emerged as an area of ongoing research, potentially offering complementary benefits of both architectures.

2.7.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) represent a specialized type of neural network designed to process data with grid-like topologies, such as time-series data (1D grid) or image data (2D grid). These networks have achieved remarkable success in practical applications, particularly in computer vision tasks. The term "convolutional" refers to the mathematical operation of convolution, a specialized linear operation that replaces general matrix multiplication in at least one network layer (Goodfellow; Bengio; Courville, 2016).

CNNs have emerged as the most commonly used classification model in plant studies, with their primary advantage being the ability to utilize raw images as input without extensive preprocessing. These networks excel at learning complex problems rapidly through weight sharing and sophisticated model structures, enabling extensive parallelization (Pan; Yang, 2010). The architecture typically consists of three main components: convolutional layers, pooling layers, and fully connected layers as shown in Figure 7 (Ren; Kim; Jeong, 2020). Convolutional layers are primarily responsible for feature extraction, leveraging their ability to detect spatial relationships.

Pooling layers complement this process by reducing dimensionality, enhancing efficiency, and mitigating overfitting (Mendoza-Bernal; González-Vidal; Skarmeta, 2024).

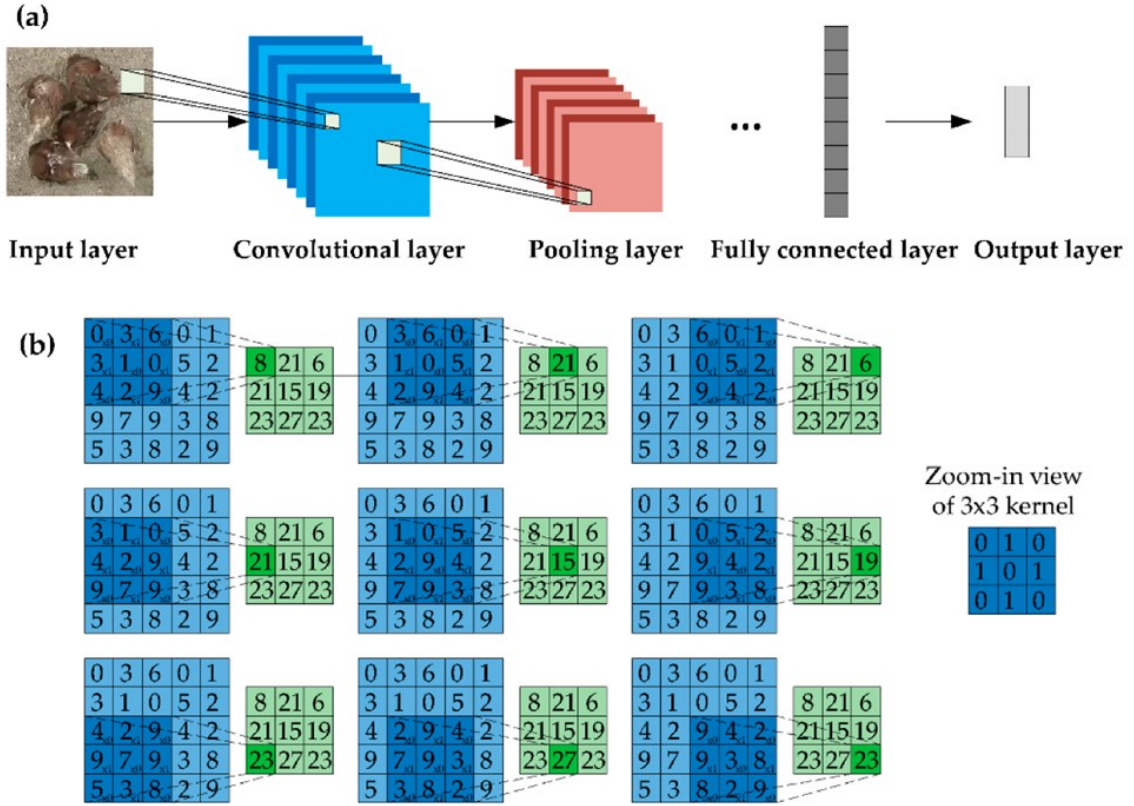


FIGURE 7. Architectural illustrations of (a) a typical convolutional neural network showing the sequence of convolutional, pooling, and fully connected layers, and (b) the convolution operation with a 3×3 kernel, stride of 1, and no padding, demonstrating how feature maps are generated (He *et al.*, 2023).

In the convolutional layers, neurons connect to small regions of the input image called receptive fields, rather than to every pixel. This design allows the network to focus on small, low-level features in the initial layers and progressively assemble them into larger, higher-level features in subsequent layers. The process involves sliding windows of defined sizes (typically 3×3 or 5×5) over the input feature map, extracting 3D patches at every possible Location (Chollet, 2021).

These networks implement parameter sharing, where the same parameters are used for more than one function in the model. This approach significantly reduces memory requirements and improves statistical efficiency compared to traditional neural networks. The convolution operation creates equivariance to translation, meaning that if the input changes, the output changes in the same way. This property is particularly useful for detecting features regardless of their position in the input data (Goodfellow; Bengio; Courville, 2016).

The pooling layers serve to subsample the input, reducing computational load, memory usage, and the number of parameters, thereby limiting overfitting risk. Unlike convolutional layers, pooling neurons have no weights and simply aggregate inputs using functions such as max or mean. The combination of convolutional and pooling layers progressively reduces the spatial dimensions while typically increasing the depth (number of feature maps) of the network (Géron, 2019).

The mathematical framework of CNNs involves specific calculations at each layer type. In the convolutional layer, the feature value at location (i, j) for the k^{th} feature map of the l^{th} layer is calculated through a combination of weights, input values, and activation functions. The most commonly used activation function in CNNs is *ReLU* (Rectified Linear Unit), which outputs the input directly if positive, or zero otherwise (Moreno-Revelo *et al.*, 2021).

After multiple convolutions, the fully connected layer integrates these highly abstracted features, which can be normalized to produce classification probabilities. While early layers commonly employ *ReLU* activation functions, the fully connected layers typically use *softmax* activation, which scales outputs between zero and one to indicate the probability of an input belonging to a specific class (Mendoza-Bernal; González-Vidal; Skarmeta, 2024).

The network's components can be mathematically described as follows:

- convolutional layer: the feature value at location (i, j) referring to the k^{th} feature map of the l^{th} layer, $z_{i,j,k}^l$ is calculated as:

$$z_i^l = \sigma \left(w_k^{lT} x_{i,j}^l + b_k^l \right), \quad (9)$$

in which $x_{i,j}^l$ is the input centered at position (i, j) referring to the l^{th} layer, w_k^{lT} and b_k^l are respectively the weight vector and the bias term of the k^{th} filter on the l^{th} layer, and σ is the activation function which in convolutional neural networks architecture is usually *ReLU* (Santos *et al.*, 2023):

$$ReLU(x) = \max\{0, x\}, \quad (10)$$

- pooling layer: each feature map of a pooling layer is connected to the corresponding feature map of the previous convolutional layer. The output generated by the pooling function for the feature map $z_{i,j,k}^l$ is represented by the expression:

$$p_{m,k} = \max_{n=1}^N (z_{i-1}^l \times q + n), \quad (11)$$

in which N is the cluster size; and q is the slide size, which determines the degree to which adjacent pool windows overlap (Santos *et al.*, 2023).

- fully connected (dense) layer: a fully connected layer takes all the neurons in the previous layer and connects them to every neuron in the current layer. Dropout technique is usually adopted in this layer, in which some neurons are randomly turned off, along their connections, during training only (Santos *et al.*, 2023). For classification problem, it is common to use as many neurons as there are classes to predict and the output $\sigma(z)_i$ has a *softmax* activation function for the i^{th} class represented in the following equation:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \quad (12)$$

in which K is the total number of classes.

The training process of CNNs is computationally intensive, with the most expensive part being feature learning. During supervised training with gradient descent, each step requires complete forward and backward propagation through the entire network (Ren; Kim; Jeong, 2020). The data used typically consists of several channels, each representing different quantities at specific points in space or time (Goodfellow; Bengio; Courville, 2016).

To combat overfitting, various regularization techniques are employed (Moreno-Revelo *et al.*, 2021). Dropout is one of the most effective and widely used methods, implementing a remarkably simple yet powerful regularization strategy. During training, each neuron in the network (except output neurons) has a probability p of being temporarily "dropped out" or deactivated at each training step. This dropout rate typically ranges between 20% and 50%. The technique effectively creates an ensemble of different neural networks by randomly removing various combinations of neurons during training, as each dropout configuration essentially produces a different sub-network of the original architecture. This approach is computationally inexpensive yet highly effective, as it forces the network to learn more robust features that are useful in conjunction with many different random subsets of the other neurons as shown in Figure 8. During testing, no units are dropped, but output values are scaled down by a factor equal to the dropout rate to maintain balance between training and inference phases. This introduction of noise helps break up coincidental patterns that the model might otherwise memorize, preventing complex co-adaptations between neurons where a feature detector is only helpful in the context of several

other specific feature detectors (Chollet, 2021; Géron, 2019; Goodfellow; Bengio; Courville, 2016).

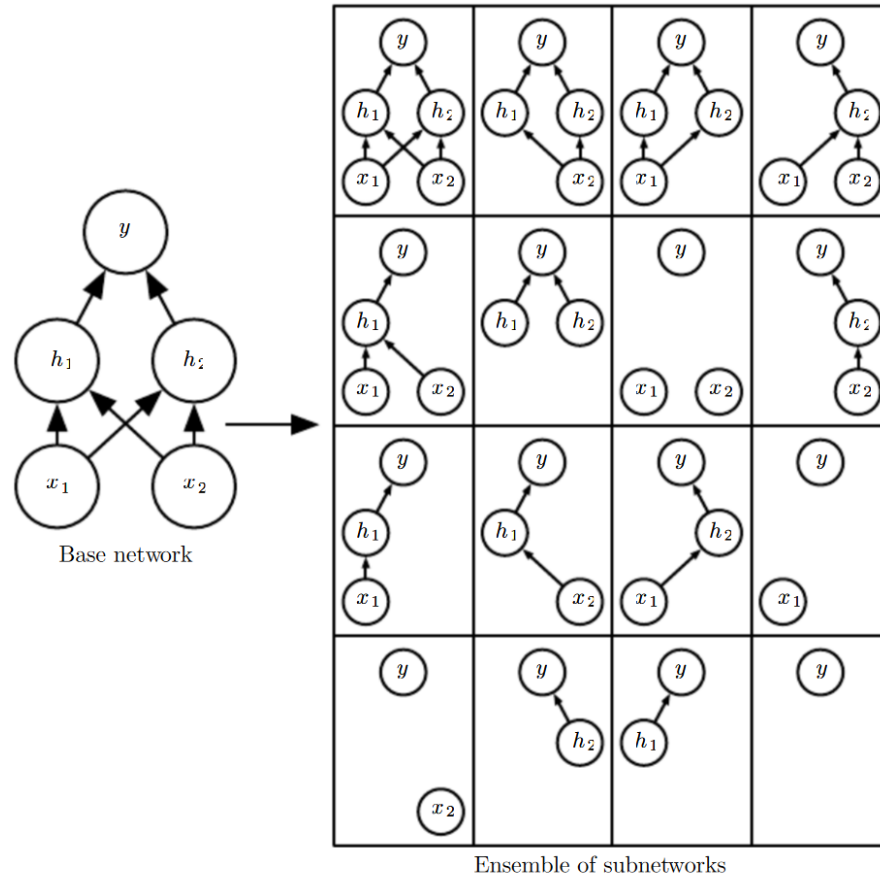


FIGURE 8. Dropout regularization technique visualization showing how randomly deactivating neurons during training creates an ensemble effect that improves model generalization by preventing co-adaptation of features (Goodfellow; Bengio; Courville, 2016).

Early stopping represents another crucial regularization technique, particularly effective for deep learning models that are inherently overparameterized. While training error typically continues to decrease throughout the training process, validation error often begins to increase after a certain point, indicating overfitting. Early stopping addresses this by monitoring the model's performance on a validation set and preserving the model parameters when validation metrics are at their best. The training process is terminated when the validation error shows no improvement for a predetermined number of iterations. This approach effectively finds the optimal point between underfitting and overfitting, ensuring the most generalizable model possible. Implementation typically involves saving the model state whenever validation performance improves and reverting

to the best-performing version once training concludes. This method has become one of the most widely used regularization techniques in deep learning, valued for both its effectiveness and simplicity (Chollet, 2021; Goodfellow; Bengio; Courville, 2016).

2.7.2 Other Deep Learning Techniques in Agriculture

Deep learning has emerged as a transformative approach for agricultural image analysis, with numerous applications specifically focused on soybean phenotyping and maturity assessment. A deep learning method within the Convolutional neural network was used to detect mature soybean stem nodes and identified soybean structural features through a novel directed search algorithm and RGB images - YOLOX had a maximum average accuracy of 94.36% for soybean stem nodes and scale markers (Guo *et al.*, 2022). A recognition method of soybean pods and estimation of pods weight per plant using RGB images were proposed based on improved YOLOv5 model reaching 91.7% with detection rate of 24.39 frames per millisecond (He *et al.*, 2023). Two complementary Convolutional neural networks were developed to predict the maturity date in an image-based high-throughput phenotyping system for evaluating soybean in breeding programs with an RMSE lower than 2.0 days in four out of five trials (Trevisan *et al.*, 2020). An UAV-based multimodal data fusion using RGB, multispectral and thermal sensors were used to estimate soybean grain yield within the framework of Deep Neural Network (Maimaitijiang *et al.*, 2020). Five deep learning architectures, Inception-v3, Resnet-50, VGG-16, VGG-19 and Xception, were evaluated for the classification of soybean pest images reaching accuracies of up to 93.82% (Tetila *et al.*, 2020). An automatic approach for estimating the relative maturity of soybeans using a time series of UAV images and an end-to-end hybrid model combining Convolutional neural networks and Long Short-Term Memory was tested on six different environments with less than 2 days of error for five environments (Moeinizade *et al.*, 2022). A new Convolutional neural network, called DS-SoybeanNet, was designed to improve the performance of UAV-based soybean maturity information monitoring showing overall accuracies for different fields of 99.19% and 86.26% (Zhang *et al.*, 2022).

While multimodal approaches have been explored, integrating diverse data types (RGB, multispectral, thermal) remains complex and computationally intensive. Streamlining these processes is necessary for practical applications. Moreover, models often show variability in performance across different environmental conditions, indicating a need for more robust models that can generalize well to various field conditions. High accuracy models often rely on complex

and resource-intensive methodologies, which may limit scalability and accessibility for broader agricultural use. By addressing these limitations and introducing innovative methodologies, this study seeks to advance the field of soybean maturity classification, contributing to improved agricultural practices and decision-making.

2.8 Previous Studies in Soybean Maturity Classification

This section examines key studies focused specifically on soybean maturity classification and prediction, highlighting their methodological approaches, achievements, and limitations that inform the current research. The present work builds upon these foundations while addressing several persistent challenges, particularly the need for environmental robustness and efficient deployment in breeding programs. Our approach leverages convolution-based deep learning architectures that can effectively process single-timepoint RGB imagery while maintaining high accuracy across diverse environments – a combination that previous methods have struggled to achieve simultaneously.

Soybean maturity classification represents a critical application of computer vision in agriculture. Trevisan *et al.* (2020) developed an RGB image-based high-throughput phenotyping system using UAV imagery with two complementary CNNs to predict maturity dates. Rather than classifying discrete maturity stages, their system was designed to directly predict the days until maturity from images taken at various timepoints throughout the growing season. Their system achieved RMSE values below 2.0 days in four out of five trials across diverse environments. However, they noted decreased performance when applied to environments with different lighting conditions or plant architectures, highlighting the challenge of environmental transferability.

Building on this foundation, Moeinizade *et al.* (2022) developed a temporal approach combining CNN feature extraction with LSTM networks to capture developmental patterns across multiple imaging dates. Their model achieved predictions within 2 days of ground-truth measurements across five of six environments tested. A key innovation was the integration of temporal information through the LSTM architecture, allowing their model to learn patterns of color change over time rather than relying on single-time point assessments. While their model performed exceptionally well in five environments (with MAE < 2 days), they observed reduced accuracy in one environment (MAE of approximately 2.5 days), which they attributed to outliers and the use of a different UAV platform. This highlights the ongoing challenge of creating models that generalize perfectly across diverse imaging conditions.

Zhang *et al.* (2022) focused on extracting both shallow and deep image features with their DS-SoybeanNet architecture. This approach was applied to classify soybean plots into four maturity categories: immature, near-mature, mature, and harvested. Testing across two independent soybean breeding fields, they achieved classification accuracies of 99.19% on their training field and 86.26% on an independent validation field. Their approach showed greater robustness to environmental variation compared to standard CNN architectures like MobileNetV2 and ResNet50, which exhibited substantial performance degradation when applied to the independent validation field. However, the authors acknowledged limitations in their approach, particularly the large model size which could limit deployment on resource-constrained devices and noted that further validation across additional fields and growing seasons would be necessary to fully establish the model's generalizability.

While Zhang's work focused on classifying current maturity states, other researchers have investigated predictive approaches that forecast maturity timing from earlier growth stages. Several researchers have explored the temporal dimension of soybean maturity. Yuan *et al.* (2019) developed an early prediction approach using color and texture features extracted from early-season canopy RGB imagery. Rather than tracking temporal changes in vegetation indices, their method focused on extracting rich feature sets from early-season images (V4/V5 growth stage) to predict end-season traits including maturity. Using the Cubist regression algorithm, they achieved promising results with $R^2 = 0.76$ and $RMSE = 3.70$ days across four locations. Texture features alone provided better predictions than color features, suggesting that canopy structural information carried more predictive power than simple color metrics. A key strength of their approach was enabling early-season predictions, though they acknowledged limitations including unexplained mechanisms, limited generalizability beyond their test environments in Nebraska, and potential inconsistencies due to lack of color calibration in their imaging protocol.

Zhou *et al.* (2019) focused on multispectral imagery during the maturity transition period (R7-R8), extracting 130 spectral features from five-band multispectral images and using PLSR to predict maturity dates with $R^2 = 0.81$ and $RMSE = 1.4$ days. They found that vegetation indices related to chlorophyll content and water stress were particularly valuable for capturing maturity transitions, and their method performed best using images from the beginning of maturity combined with change rates from mid-maturity. However, they acknowledged limitations in their approach, particularly related to imaging conditions due to environmental factors.

Volpato *et al.* (2021) tracked changes in vegetation indices over time using LOESS regression applied to time-series of GLI values derived from UAV imagery. They demonstrated that weekly imaging was sufficient to achieve high accuracy ($r = 0.84-0.97$), with minimal improvements from more frequent flights. Their approach showed robustness across different environments, with broad-sense heritability estimates for UAV-based dates equal to or higher than ground-based estimates in 40 of 53 field trials. While they emphasized the importance of environment-specific threshold calibration based on region, sensor hardware, and germplasm, they found their method relatively insensitive to small threshold variations once properly configured, indicating good cross-environmental robustness.

While the above studies primarily relied on RGB imagery or derived vegetation indices tracked over time, researchers have also explored whether incorporating additional sensing modalities beyond temporal information could further improve phenotyping accuracy. Maimaitijiang *et al.* (2020) investigated the integration of RGB, multispectral, and thermal data for soybean phenotyping, demonstrating that multi-modal data fusion significantly improved prediction accuracy. Their deep neural network approach achieved $R^2 = 0.72$ for yield prediction by leveraging complementary information: RGB data captured structural information, multispectral indices tracked physiological properties, and thermal imagery detected temperature variations related to plant water status. While showing promise for comprehensive phenotyping beyond maturity classification, the authors acknowledged that their approach required validation across diverse crops, genotypes, development stages, and environments, with the complexity of their multi-sensor system and data processing pipeline presenting significant barriers to adoption in production breeding programs.

Recent advances in deep learning have enabled more sophisticated approaches. Guo *et al.* (2022) used the YOLOX object detection network to identify stem nodes with a maximum average precision 94.36% and a novel directed search algorithm to extract structural features in mature soybeans. Their automated measurements of multiple phenotypes achieved Pearson correlation coefficients between 0.9084 and 0.9925 compared to manual measurements. While this work focused on stem-related phenotypes rather than maturity classification specifically, it demonstrated the potential of advanced deep learning approaches for high-throughput phenotyping of mature soybean plants.

He *et al.* (2023) enhanced the YOLOv5 architecture with a coordinate attention module and modified regression loss function to detect and count soybean pods, achieving 91.7% average precision while reducing computational requirements by 17%. Their yield prediction model based on pod phenotypic traits achieved $R^2 = 0.945$ for single pod weight estimation and a mean relative error of 0.122 for whole-plant yield estimation.

More recently, transformer architectures have emerged as powerful alternatives to traditional CNN-based approaches for agricultural image analysis. Unlike CNNs that process images using fixed receptive fields, transformers employ self-attention mechanisms that can dynamically model relationships between all parts of an image or sequence, potentially capturing more complex patterns in plant development and appearance. Bi *et al.* (2023) proposed a transformer-based approach for early soybean yield prediction using time-series images and seed information that reduced prediction error by over 40% by separately processing plant and soil image components and modeling temporal dependencies. However, their approach was limited by a dataset from a single location and season with only three time-series images per plot, restricting temporal analysis. The model also incorporated minimal seed information without considering other environmental or genetic factors, and calculated attention scores separately for images and time series rather than exploring cross-temporal patch relationships.

Beyond soybeans, Khan *et al.* (2023) developed a convolutional transformer for tomato maturity recognition that effectively classified ripeness stages across three datasets. While their hybrid architecture handled occlusion well, it struggled with extremely occluded or small tomatoes and sometimes generated false masks or pixel-level false positives. Their dataset's limited diversity in tomato varieties, growth stages, and lighting conditions potentially reduced generalizability, and its greenhouse-focused nature might not translate well to open-field applications without additional training data.

While these advanced sensing approaches and deep learning techniques offer improvements in accuracy, practical breeding applications must consider trade-offs in cost, complexity, and scalability. Multi-modal approaches show promise for improving robustness, but further work is needed to develop systems that are both accurate and practical for large-scale breeding operations.

Analysis of the current literature reveals several consistent challenges and research gaps in soybean maturity classification:

- a. **Environmental Transferability:** A persistent challenge across studies is the degradation in model performance when applied to new environments with different lighting conditions, plant architectures, or imaging platforms. While Volpato *et al.* (2021) demonstrated some cross-environmental robustness, most approaches like Trevisan *et al.* (2020) and Moeinizade *et al.* (2022) show significant performance drops in novel environments. This highlights the critical need for models that can generalize across diverse growing conditions without extensive recalibration.
- b. **Temporal Information Integration:** Current approaches vary widely in how they leverage temporal information. While Moeinizade *et al.* (2022) and Bi *et al.* (2023) incorporated time-series data through LSTM and transformer architectures respectively, they were limited by sparse temporal sampling. Yuan *et al.* (2019) took a different approach by extracting rich feature sets from early-season images. The optimal strategy for temporal information integration — balancing prediction accuracy with practical data collection requirements — remains an open question.
- c. **Feature Representation:** Studies have utilized diverse approaches to feature extraction, from handcrafted color and texture features (Yuan *et al.*, 2019) to CNN-based representations (Trevisan *et al.*, 2020) and transformer architectures (Bi *et al.*, 2023; Khan *et al.*, 2023). Zhang *et al.*'s (2022) DS-SoybeanNet demonstrated that combining shallow and deep features can improve robustness, but the ideal feature representation strategy for maturity classification remains unclear, particularly for achieving environmental generalization.
- d. **Practical Deployment Considerations:** Many high-performing approaches rely on specialized equipment (Maimaitijiang *et al.*, 2020), multiple imaging dates (Volpato *et al.*, 2021), or complex processing pipelines that limit practical implementation in large-scale breeding programs. The computational requirements of advanced models like Khan *et al.*'s (2023) convolutional transformer or Zhang *et al.*'s (2022) DS-SoybeanNet further complicate field deployment. Solutions that balance accuracy with operational simplicity could significantly increase adoption.
- e. **Experimental Validation vs. Performance:** Most deep learning approaches focus primarily on achieving high accuracy metrics without extensive experimental validation of the underlying biological mechanisms. While these models demonstrate impressive performance, the limited scope of experimental validation restricts understanding of the

visual features driving classification decisions and the biological processes underlying maturity progression. More comprehensive experimental frameworks could both improve model reliability and provide valuable insights into phenotypic mechanisms beyond simple classification.

- f. **Handling Spatial Heterogeneity:** Within-plot maturity variation represents a significant challenge rarely addressed in current literature. Most approaches aggregate plot-level features, potentially obscuring important spatial patterns in maturity progression. Automated methods for handling heterogeneous maturity patterns could improve both accuracy and utility for breeding applications where uniform maturity is a desirable trait.

The present research addresses these gaps by developing a computationally efficient architecture that maintains high accuracy across diverse environments while requiring only single-timepoint RGB imagery. By incorporating convolutional mechanisms that automatically focus on relevant plot regions and implementing domain adaptation techniques, this work aims to overcome the environmental transferability limitations identified in previous studies. Additionally, the proposed approach includes extensive and comprehensive experimental validation that provides insight into the visual features driving classification decisions, moving beyond the closed-system nature of many current solutions while maintaining the performance advantages of deep learning architectures.

3 MATERIALS AND METHODS

The experimental study was conducted during the 2023 growing season across two Locations in Iowa state, USA, utilizing advanced imaging technology and comprehensive environmental monitoring to develop and validate a deep learning approach for soybean maturity classification. The research infrastructure combined high-resolution aerial imagery, environmental sensors, and extensive genetic trials to create a robust framework for cross-location maturity assessment. This research was conducted using proprietary data from Syngenta Seeds breeding programs, with all experimental materials, genetic information, and resulting datasets remaining the exclusive property of Syngenta Seeds.

3.1 Experimental Design and Data Collection

The research was conducted at two experimental sites in Iowa, USA: a primary location in Slater (Location 01) and a secondary site in Alleman (Location 02) shown in Figure 9. Location 01 comprised 41,072 experimental plots, with dimensions of 10×2.5 feet per plot, strategically divided into training (33,184 plots) and testing (7,888 plots) fields. Location 02 contained 11,672 plots of identical dimensions, serving as an independent validation site for assessing model transferability across different environmental conditions. These experimental sites are part of Syngenta Seeds' proprietary breeding network, with all genetic materials representing confidential germplasm under development.

Data collection was performed using a DJI Phantom 4 unmanned aerial vehicle (UAV) equipped with an RGB camera. Flight operations were conducted between 11:00 and 13:00 local time to ensure consistent lighting conditions. The UAV operated at an altitude of 30 meters with 80% forward overlap and 70% side overlap, achieving a spatial resolution of 10mm. Flight campaigns at Location 01 spanned from 74 to 111 days after planting (DAP), resulting in 12 temporal datasets, while Location 02 was monitored from 75 to 107 DAP across 9 flight dates. All image data collection was conducted by Syngenta Seeds field operations team following standardized protocols for breeding program evaluations.

The experimental design incorporated multiple maturity groups and varying planting dates to capture the full spectrum of soybean development stages. Figure 10 illustrates the temporal distribution of soybean maturity classes across both locations, showing the progression from

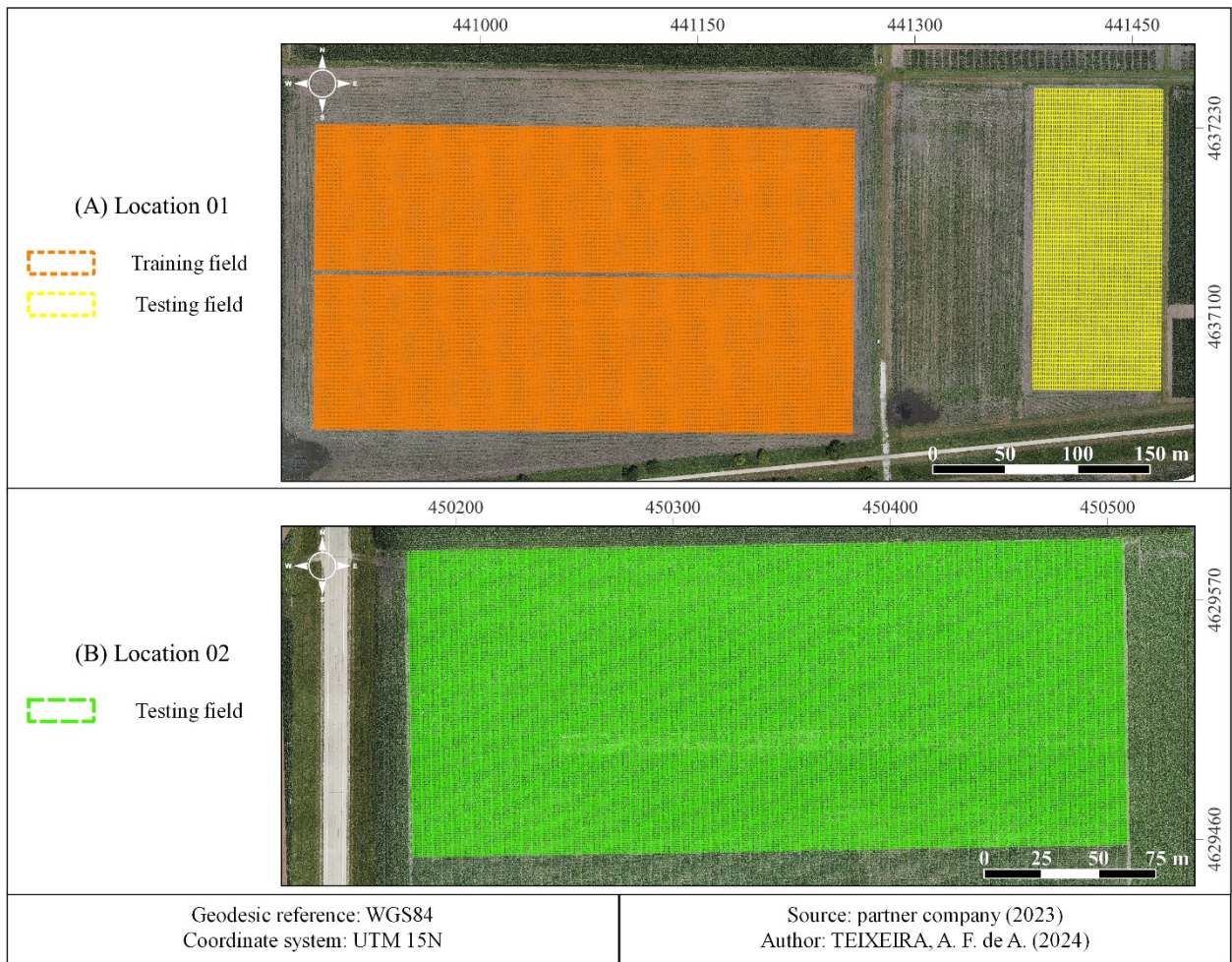


FIGURE 9. Temporal distribution of soybean maturity classes (Immature, Near-mature, and Mature) by days after planting in (A) Location 01 and (B) Location 02, showing the progression of maturity throughout the growing season.

predominantly immature plots to mature plots in late season. This visualization reveals critical transition periods between 80 and 100 DAP, where rapid shifts in maturity classification occur, presenting particular challenges for classification algorithms. The stacked bar charts quantify the percentage of plots in each maturity category throughout the growing season, highlighting the non-linear nature of the maturation process.

The study employed a multi-environment approach where each unique combination of location and flight date constituted a distinct environment, resulting in 20 discrete micro-environments (11 at Location 01 + 9 at Location 02). This design captured environmental variability through temporal variation within locations and spatial variation between locations,

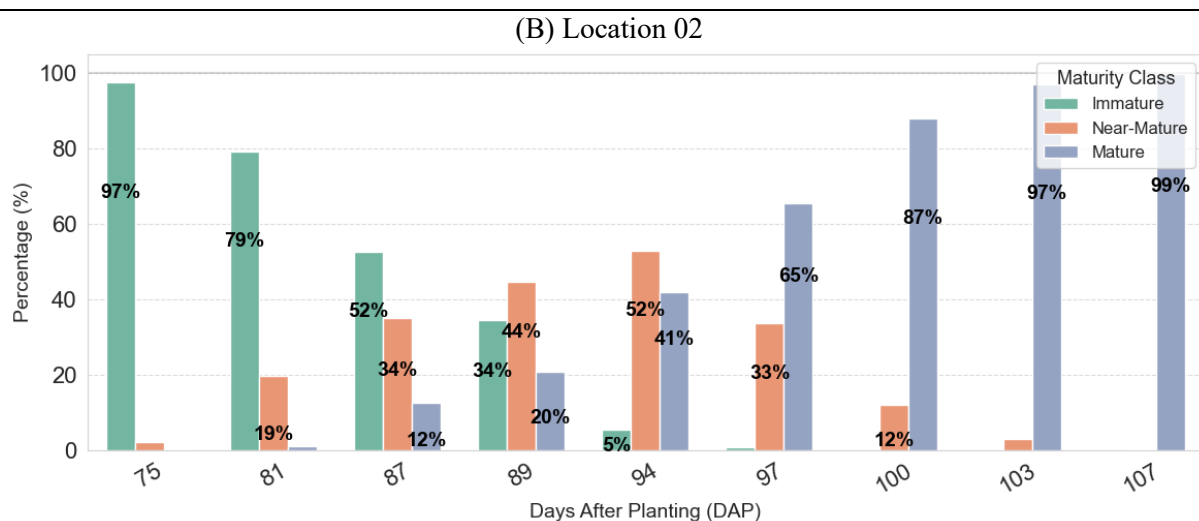
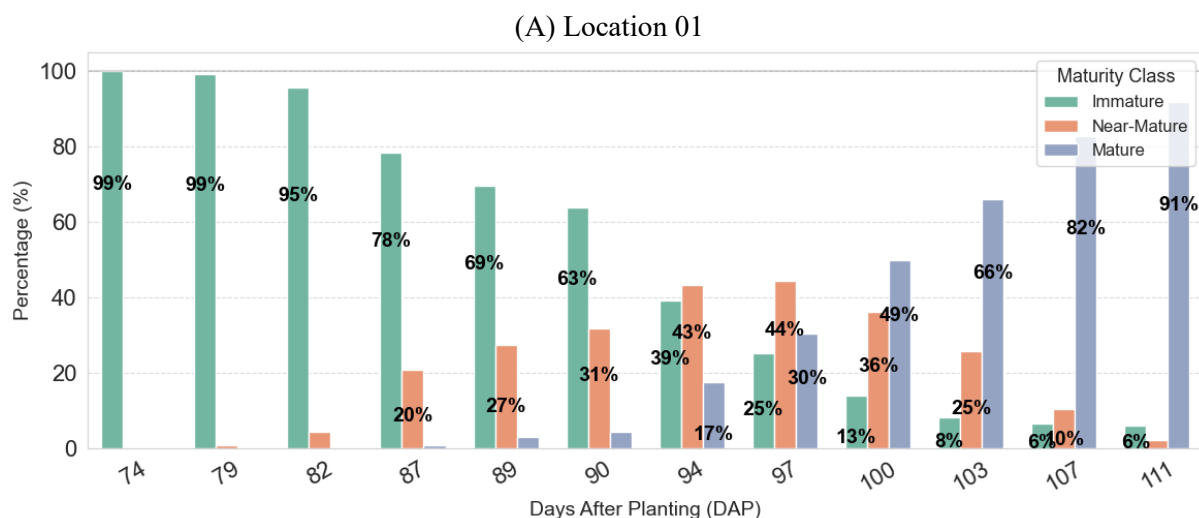


FIGURE 10. Temporal distribution of soybean maturity classes (Immature, Near-mature, and Mature) by days after planting in (A) Location 01 and (B) Location 02, showing the progression of maturity throughout the growing season. Percentages indicate the proportion of plots in each maturity class at specific timepoints, demonstrating the transition from immature plots to predominantly mature plots by late season, with a critical transition period occurring between 80 and 100 days after planting (DAP).

creating natural environmental gradients for robust model evaluation. Each environment was characterized by unique combinations of meteorological conditions, solar radiation patterns, and soil moisture levels during image acquisition.

Environmental data was collected from the NASA POWER satellite-derived database for both experimental locations, synchronized with UAV flight dates throughout the growing season. The dataset incorporated both hourly and daily measurements of environmental parameters (Table

1), providing comprehensive coverage of atmospheric and soil conditions, ranging from temperature and humidity to radiation metrics and soil moisture indices.

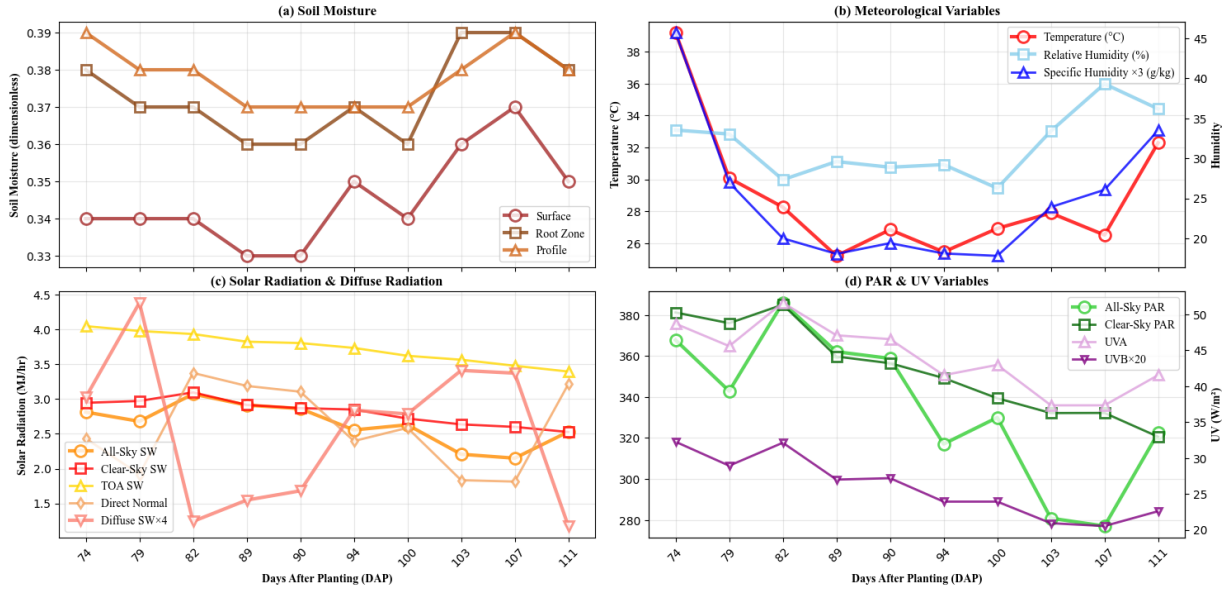
TABLE 1. Environmental parameters used in the study showing parameter abbreviations, descriptions, units of measurement, and temporal resolution for data collection.

Abbreviation	Description	Unit	Temporal Resolution
T2M	Temperature at 2 Meters	°C	Hourly
RH2M	Relative Humidity at 2 Meters	%	Hourly
QV2M	Specific Humidity at 2 Meters	g/kg	Hourly
ALLSKY_SFC_SW_DWN	Surface Shortwave Downward Irradiance	MJ/hr	Hourly
CLRSKY_SFC_SW_DWN	Clear Sky Surface Shortwave Downward Irradiance	MJ/hr	Hourly
ALLSKY_SFC_SW_DNI	Surface Shortwave Downward Direct Normal Irradiance	MJ/hr	Hourly
ALLSKY_SFC_SW_DIFF	Surface Shortwave Diffuse Irradiance	MJ/hr	Hourly
TOA_SW_DWN	Top-Of-Atmosphere Shortwave Downward Irradiance	MJ/hr	Hourly
ALLSKY_SFC_PAR_TOT	Surface PAR Total	W/m ²	Hourly
CLRSKY_SFC_PAR_TOT	Clear Sky Surface PAR Total	W/m ²	Hourly
ALLSKY_SFC_UVA	Surface UVA Irradiance	W/m ²	Hourly
ALLSKY_SFC_UVB	Surface UVB Irradiance	W/m ²	Hourly
ALLSKY_SFC_UV_INDEX	Surface UV Index	W/m ²	Hourly
GWETTOP	Surface Soil Wetness	dimensionless	Daily
GWETROOT	Root Zone Soil Wetness	dimensionless	Daily
GWETPROF	Profile Soil Moisture	dimensionless	Daily

The multi-environment approach was implemented through strategic flight scheduling across different dates, creating distinct environmental conditions for each imaging session. FIGURE 11 illustrates the environmental variability captured across different flight dates, showing substantial variation in key climatic parameters that created natural environmental gradients within the experimental design.

The environmental data integration process followed a two-step approach. First, daily measurements were merged with image data using the flight date as the linking parameter. Second, hourly measurements were processed by calculating mean values over a two-hour window centered on the typical flight time (11:00 a.m.-01:00 p.m. local time), ensuring temporal alignment with image acquisition. This integration strategy provided a comprehensive environmental context for each captured image while maintaining data quality through systematic handling of missing values.

(A) Location 01



(B) Location 02

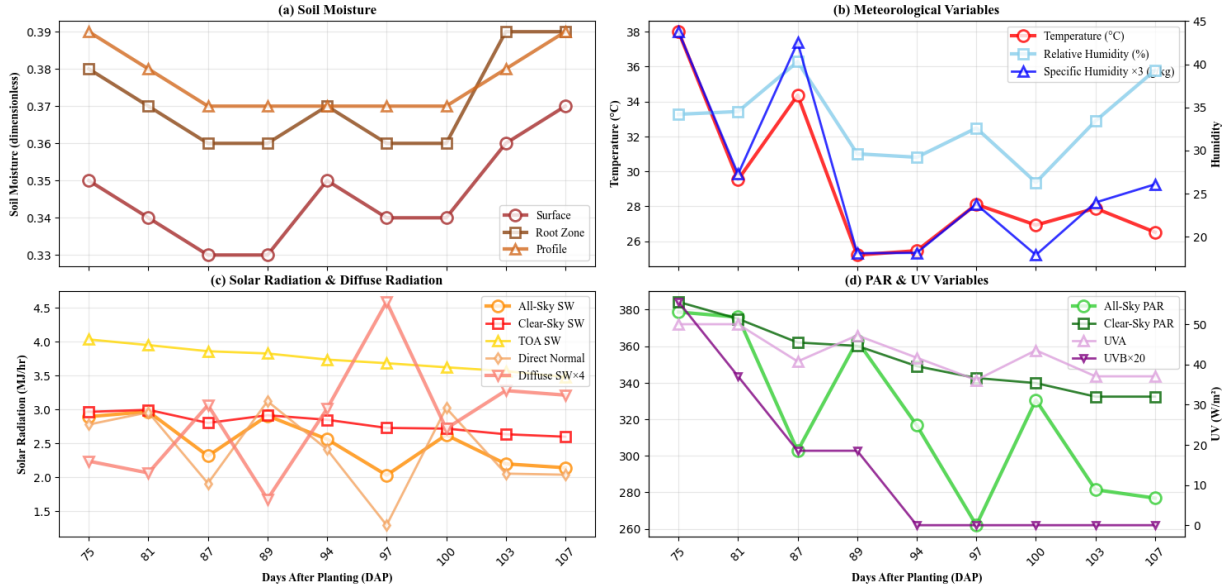


FIGURE 11. Comprehensive environmental parameter variations across flight dates organized by variable type in (A) Location 01 and (B) Location 02: (a) Soil moisture profiles at different depths (surface, root zone, and profile), (b) Meteorological variables including temperature, relative humidity, and specific humidity, (c) Solar radiation components including all-sky and clear-sky shortwave radiation, top-of-atmosphere radiation, direct normal irradiance, and diffuse radiation, and (d) Photosynthetically active radiation (PAR) and ultraviolet radiation variables (UVA and UVB). The temporal progression demonstrates distinct variability patterns across parameter groups: soil moisture shows high stability (CV 2-4%), meteorological variables exhibit moderate variation (CV 13-37%), solar radiation components display the highest variability (CV 5-43%), and PAR/UV variables show moderate fluctuation (CV 5-141%). Data encompasses 11 flight dates for Location 01 and 9 flight dates for Location 02, creating 20 distinct micro-environments across the growing season.

The experimental design captured significant environmental variation through multiple flight dates, with each flight representing a unique micro-environment characterized by distinct combinations of:

- Meteorological conditions: Temperature averaged 28.87°C (Location 01) and 29.11°C (Location 02) with moderate variability (CV ~14-15%), relative humidity showed similar patterns (31.68% vs 33.27%), while specific humidity exhibited the highest temporal variation (CV 36-37%) across both locations
- Solar radiation dynamics: All-sky shortwave radiation averaged 2.64 MJ/hr (Location 01) and 2.52 MJ/hr (Location 02), with diffuse radiation showing the highest variability (CV 30-43%) and direct normal irradiance exhibiting substantial fluctuation (CV 24-26%), creating diverse light conditions
- Soil moisture stability: Exceptional stability across all depths with surface wetness (0.35), root zone (0.37), and profile moisture (0.38) showing minimal variation between locations and low temporal variability (CV 2-4%), indicating uniform soil conditions
- PAR and UV exposure: Photosynthetically active radiation showed remarkable similarity between locations (334.49 vs 320.89 W/m²), UVA radiation remained consistent (44.03 vs 42.59 W/m²), while UVB exhibited high variability, particularly at Location 02, creating diverse photosynthetic and stress conditions

Cross-location environmental analysis demonstrated substantial variability within and between locations, validating the multi-environment experimental design. While mean environmental conditions appeared similar between locations, the temporal variability within each location created distinct micro-environments: specific humidity showed high temporal variation (CV 35.6% and 36.9%), diffuse solar radiation exhibited exceptional variability (CV 42.5% and 29.6%), and direct normal irradiance fluctuated considerably (CV 23.5% and 25.8%). Location 02 demonstrated greater environmental heterogeneity with higher coefficients of variation in most radiation parameters, including all-sky shortwave radiation (14.3% vs 11.2%) and PAR (13.7% vs 10.8%). The distinct flight dates across both locations captured environmental ranges spanning 14°C in temperature, 0.94 MJ/hr in solar radiation, and 116.67 W/m² in PAR, creating diverse growing conditions that challenged model robustness and ensured comprehensive evaluation across varying climatic scenarios during the critical maturation period.

This environmental heterogeneity was deliberately incorporated to enhance model robustness and ensure generalizability across diverse growing conditions. Each combination of location and flight date created a unique environmental signature, resulting in 20 distinct micro-environments rather than simply two locations, thereby providing a comprehensive evaluation framework for the deep learning model across varying climatic conditions during the critical maturation period.

The combined dataset merged plot-specific imagery with both instantaneous (hourly) and cumulative (daily) environmental parameters, creating a rich feature set for model training. This multi-temporal approach captured both the immediate environmental conditions during image acquisition and the broader environmental context affecting plant development, enabling the model to learn associations between environmental conditions and maturity patterns across diverse micro-environments.

3.2 Data Processing

The data acquisition process integrated multiple data streams, including aerial imagery, environmental measurements, and genetic information, to create a comprehensive dataset for maturity classification. The image collection protocol was standardized across both locations to ensure data consistency and quality.

The image preprocessing pipeline (Figure 12) consisted of five sequential steps designed to transform raw aerial imagery into standardized input data for model training:

1. Plot boundary definition was performed using high-precision GPS coordinates, creating individual plot polygons with 10-centimeter buffer zones to minimize edge effects (Figure 12a-12b).
2. Geospatial data integration merged plot boundaries with unique identifiers and trial metadata through spatial joining operations, enabling plot-level tracking throughout the analysis pipeline (Figure 12c-d).
3. Image extraction generated individual plot images by clipping the orthomosaic using the defined plot boundaries, isolating each experimental unit for independent analysis (Figure 12e).
4. Image standardization included resizing all plot images to 300×75 pixels while maintaining the original RGB color channels and aspect ratio, ensuring dimensional consistency across the dataset.

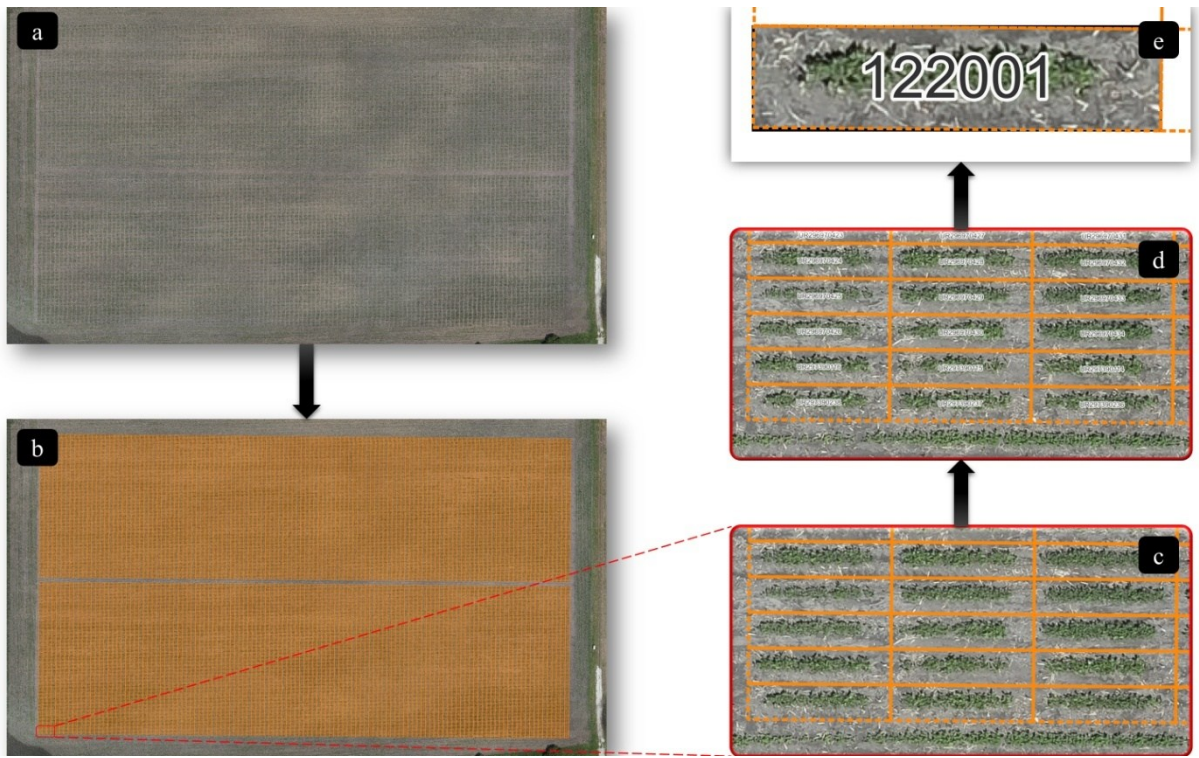


FIGURE 12. Comprehensive data preprocessing pipeline illustrating the sequential transformation of raw UAV imagery into standardized model inputs: (a) raw aerial orthomosaic of the experimental field; (b) overlay of plot boundary grid derived from GPS coordinates; (c) zoomed region showing detailed plot delineation; (d) integration of unique plot identifiers with boundary data; and (e) extracted individual plot image ready for standardization and model input. This systematic approach ensures consistent data preparation while preserving the spatial integrity of each experimental unit.

5. Data conversion transformed the processed images into NumPy arrays, applying normalization (0-1 range) to standardize pixel values across all samples and prepare them for neural network processing.

While no explicit data augmentation techniques were applied, the variety of capture dates provided natural variation in the dataset. These preprocessing steps ensured that the input data was consistently formatted, properly labeled, and well-organized for model training and evaluation. The approach of using multiple flight dates and balanced sampling helped to create a robust dataset capable of capturing the temporal aspects of soybean maturity.

This comprehensive preprocessing pipeline transformed the raw orthomosaic images into a structured dataset suitable for deep learning applications, while maintaining the integrity of the plot-specific information and addressing potential biases in class representation.

Genetic analysis was conducted independently for both experimental locations to assess the diversity and maturity patterns of the breeding populations. The experimental design encompassed

multiple developmental stages (from early to late maturity groups) with varying trial setups, creating a comprehensive evaluation framework. Location 01 featured a broader experimental scope with over 50 distinct trials, while Location 02 contained a more focused set of approximately 15 trials, each with specific objectives and designs. All genetic materials used in this study are proprietary breeding lines from Syngenta Seeds's soybean breeding program, with specific genotypic information and pedigrees remaining confidential.

At Location 01, the analysis encompassed 12,848 unique genotypes evaluated across 41,072 plots, yielding an average of 3.2 replications per genotype. The experimental window at Location 01 spanned 54 days, capturing a wide range of maturity expressions across diverse genetic backgrounds. Location 02 demonstrated complementary genetic diversity with 5,295 unique genotypes evaluated in 11,672 plots, with an average of 2.2 replications per genotype. The experimental window of 40 days at Location 02 captured sufficient variation in maturity timing to enable comprehensive genetic evaluation.

The contrasting experimental windows between locations reflect different breeding program objectives and environmental contexts. Location 01's 54-day window provided extensive temporal coverage for evaluating diverse maturity groups and their responses to varying seasonal conditions. Location 02's 40-day window offered concentrated evaluation during the critical maturation period, capturing genetic variation while minimizing late-season environmental confounding factors. Both approaches contribute valuable information for breeding decisions, with each location providing unique insights into genotype performance under different temporal and environmental scenarios.

Relative maturity date (RM_{date}) was determined through systematic field evaluations conducted by experienced plant breeders following standardized protocols. Expert evaluators assessed physiological maturity indicators including pod color, leaf senescence, and seed moisture content to determine the optimal harvest timing for each plot. These field-based maturity assessments were conducted regularly throughout the harvest window, ensuring accurate capture of genotype-specific maturity timing. The standardized evaluation protocols maintained consistency across both locations and different evaluation teams.

The substantial number of unique genotypes and diverse trial designs provided robust representation of genetic variation within the breeding populations. Environmental conditions influence genetic expression patterns differently at each location, creating complementary datasets

for comprehensive breeding evaluation. The consistent statistical significance across analyses confirmed the reliability of the genetic evaluations and the effectiveness of the experimental design in capturing meaningful genetic variation for maturity timing.

Class labels were assigned to each image based on the difference between the plot's relative maturity date and the flight date (Fl_{date}). Three distinct classes can be seen in Table 2: Immature, Near-mature and Mature.

TABLE 2. Classes definition for soybean maturity classification, including class labels (Immature, Near-mature, Mature), mathematical definitions based on relative maturity date (RM_{date}) and flight date (Fl_{date}), and descriptive explanations.

Class	Definition	Description
 (0) Immature	$(RM_{date} - Fl_{date}) > 7$	More than 7 days before maturity
 (1) Near-mature	$0 \leq (RM_{date} - Fl_{date}) < 7$	Within 7 days of maturity
 (2) Mature	$(RM_{date} - Fl_{date}) \leq 0$	At or past maturity date

The class distribution varied naturally across the growing season, with the proportion of mature plots increasing over time. To address this temporal variation, the training dataset was carefully balanced through stratified sampling, ensuring representative coverage of all maturity stages while maintaining the integrity of temporal sequences.

3.3 Model Development and Training

The model development phase focused on creating and optimizing two primary architectures: the base RMNet model for single-location analysis and its environmental data-enhanced variant, RMNet-E, for cross-location applications. Both architectures were designed to balance computational efficiency with classification accuracy.

The base architecture was developed as a sequential convolutional neural network specifically designed for soybean maturity classification. The network processes RGB images of dimension $100 \times 25 \times 3$ through a series of carefully structured layers. The initial convolutional layer employs 64 filters with a 5×5 kernel size and ReLU activation, followed by a 2×2 max-pooling layer for dimensionality reduction. A second convolutional layer with 128 filters (3×3 kernel size) and ReLU activation is followed by another 2×2 max-pooling layer. After flattening, the network

includes a dense layer of 128 neurons with ReLU activation, incorporating a dropout rate of 0.2 to prevent overfitting. The final dense layer contains three neurons with softmax activation, corresponding to the three maturity classes (Immature, Near-mature, and Mature).

For the cross-location analysis, this architecture was enhanced to incorporate environmental data, resulting in the RMNet-E model (Figure 13). This enhanced version features a dual-input design with two main branches: an image processing branch and an environmental data branch. The image branch maintains the core CNN structure with two convolutional layers (5×5 and 3×3 kernels) each followed by max-pooling operations. The environmental branch provides a dedicated input pathway for numerical environmental parameters including temperature, humidity, solar radiation, and soil moisture data. These branches are merged through concatenation before passing through dense layers, enabling the network to learn joint visual-environmental representations.

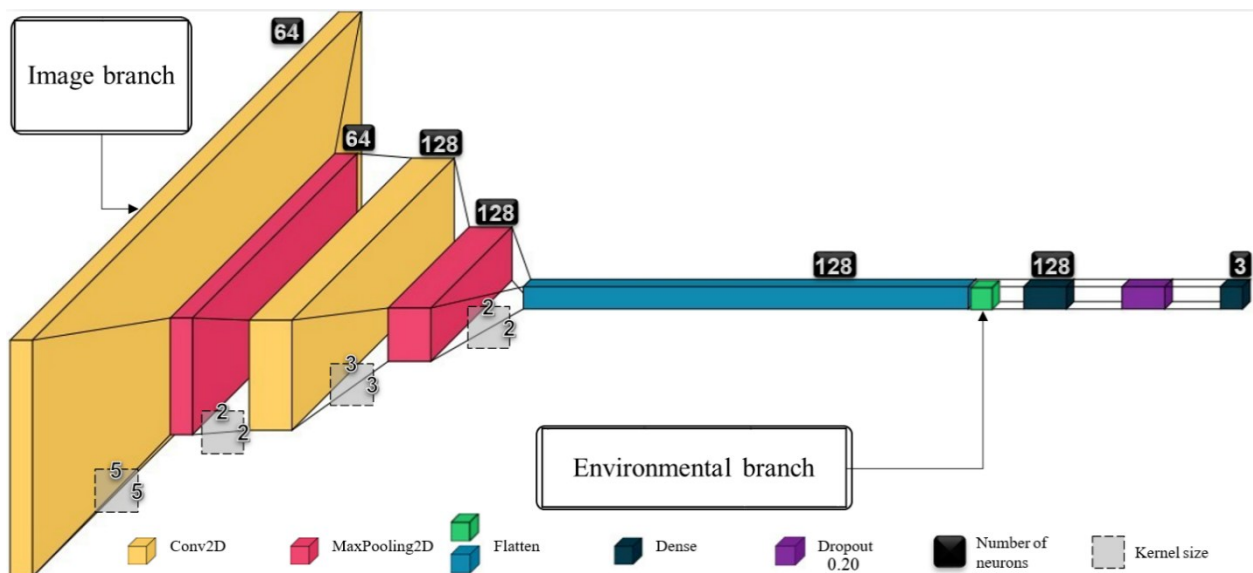


FIGURE 13. RMNet architecture diagram showing the sequential convolutional neural network design with two convolutional layers (64 and 128 filters), max-pooling operations, flattening, and dense layers optimized for soybean maturity classification.

The integration of environmental features was designed to improve model robustness across different growing conditions by capturing complex plant phenotype-environment interactions. This dual-input approach allows the model to simultaneously process spatial information from drone imagery and numerical environmental parameters, potentially enhancing classification accuracy across diverse field conditions and locations.

Training parameters were optimized using the Adam optimizer with a learning rate of 0.001 and Sparse Categorical Crossentropy as the loss function. An Early Stopping criterion was implemented, halting training if accuracy failed to improve over 20 consecutive epochs. Both architectures employed identical training configurations to ensure fair comparison between single-location and cross-location approaches.

The training dataset was divided into ten balanced groups, each containing 13,308 samples (4,436 per class), with an 80-20 split for training and validation within each group as described in Figure 14.

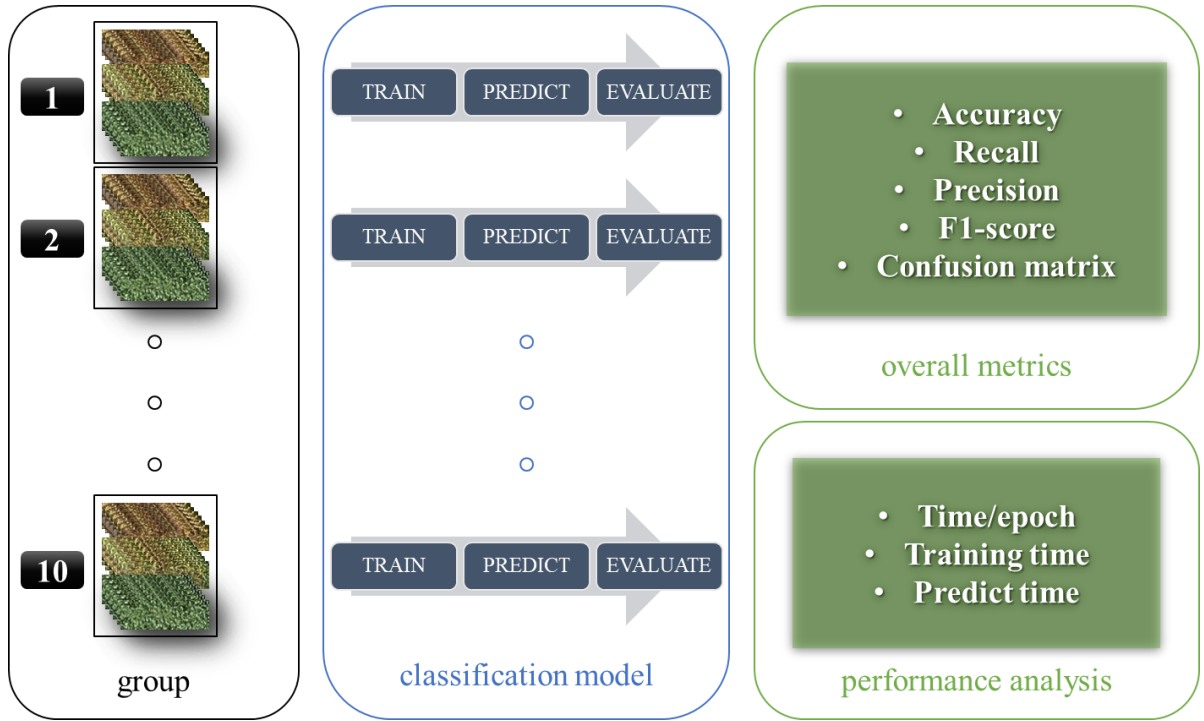


FIGURE 14. Complete data analysis workflow diagram illustrating the systematic approach to model training, validation, and evaluation, including dataset partitioning, cross-validation strategy, and performance metric calculation.

Model architecture optimization involved systematic evaluation of different configurations, varying the number of convolutional (two to three) and dense layers (one to four). Statistical validation using Levene's test confirmed homogeneity of variances across configurations ($p > 0.05$), while ANOVA testing assessed performance differences. The final architecture was selected based on optimal performance metrics and computational efficiency.

3.4 Model Evaluation and Comparative Analysis

A comprehensive evaluation framework was implemented to assess model performance across multiple dimensions. The primary metrics included accuracy, recall, precision, and F1-score, with particular emphasis on precision for the Mature class as a tiebreaker criterion. Performance analysis incorporated both computational efficiency metrics (training time per epoch for deep learning models and prediction time for machine learning models) and classification accuracy measures.

The RMNet architecture was benchmarked against both traditional machine learning and contemporary deep learning approaches. The machine learning comparison included seven established algorithms: Decision Tree, K-Nearest Neighbors, Linear Support Vector Machine, Multilayer Perceptron, Random Forest, Ridge Regression, and Support Vector Machine. These models were trained on flattened image arrays, where each RGB image was normalized (divided by the maximum pixel value) and reshaped into a one-dimensional vector. The 3D image data (height $\times$ width $\times$ channels) was transformed into a 2D array where each row represented a flattened image, preserving all pixel information while making it compatible with traditional machine learning algorithms. All models were trained using identical training and validation splits, with consistent random state initialization to ensure fair comparison.

For deep learning comparisons, six state-of-the-art architectures were evaluated: EfficientNetV2B0, MobileNetV2, ResNet50V2, ResNet101V2, ResNet152V2, and Xception. These models were implemented using the Keras applications module, with configurations optimized for the soybean maturity classification task. All models were initialized without their top layers to enable custom classification heads, used max pooling for feature aggregation, and maintained consistent input shape across architectures. The models utilized transfer learning, leveraging pre-trained weights from ImageNet, while the final classification layers were trained from scratch for the specific task. To ensure fair comparison, all models were trained using identical hardware configurations and data preprocessing pipelines, with standardized input dimensions and pixel value normalization procedures.

Model robustness was assessed through a comprehensive validation approach that combined k-fold cross-validation within each Location and independent testing across Locations. The models' generalization capability was evaluated using the separate test fields, which maintained the natural class distribution of the breeding program.

Performance was assessed using multiple evaluation metrics computed through TensorFlow's built-in classes. The primary metrics included accuracy, recall, precision, and f1-score, with confusion matrices generated to visualize the classification performance across all maturity groups, as illustrated in Figure 14. For deep learning models, performance was tracked using time/epoch parameters during training, while machine learning models were evaluated based on their training and prediction times in minutes.

The precision for the mature class ($Precision_{Mature}$) was calculated using the standard precision formula applied specifically to the mature category from the confusion matrix. For a three-class classification problem with classes 0 (Immature), 1 (Near-mature), and 2 (Mature), the precision for the mature class is defined as:

$$Precision_{Mature} = \frac{TP_{Mature}}{TP_{Mature} + FP_{Mature}} \quad (13)$$

where TP_{Mature} represents the number of plots correctly classified as mature and FP_{Mature} represents the number of plots incorrectly classified as mature when they actually belonged to classes 0 or 1. This metric specifically measures the model's reliability when it predicts a plot is ready for harvest—indicating what proportion of plots predicted as mature are actually mature. A high precision value (approaching 1.0) means that when the model identifies a plot as mature, it is very likely to be correct, minimizing the risk of premature harvest decisions. This class-specific precision was extracted directly from the confusion matrix for each model and served as the primary tiebreaker criterion when models showed similar overall performance metrics.

Statistical significance of performance differences between models was determined through a multi-stage analysis approach. Initially, Levene's test was employed to assess the equality of variances between models, followed by one-way ANOVA to evaluate differences in mean performance metrics. For cases where significant differences were detected ($p < 0.05$), post-hoc Tukey HSD tests were conducted to identify specific pairwise differences between models. This statistical framework was applied consistently across both single-Location and multi-Location analyses, with particular attention to model stability across different environmental conditions and temporal windows (Days After Planting), especially during critical transition periods between maturity.

4 RESULTS AND DISCUSSION

The development and evaluation of RMNet yielded comprehensive insights into the effectiveness of deep learning approaches for soybean maturity classification. This chapter presents the results of both single-location and cross-location analyses, examining the model's performance across different experimental conditions and in comparison with existing methodologies. The findings are organized into four main sections: single-location analysis, cross-location analysis, model robustness and generalization, and practical implications for agricultural implementation.

The single-location analysis begins with an ablation study that established the optimal network architecture, followed by comparative analyses against both traditional machine learning and contemporary deep learning approaches. The cross-location analysis explores the impact of environmental data integration and examines temporal performance patterns across different growth stages. Model robustness is assessed through cross-location validation, while practical implications are considered from the perspective of implementation requirements and constraints in breeding programs.

Throughout this chapter, statistical rigor is maintained through systematic application of Levene's tests and ANOVA, ensuring reliable interpretation of performance differences between models. Special attention is given to the precision in classifying mature plots, given its critical importance for harvest timing decisions in breeding programs.

4.1 Single Location Analysis

This section presents a comprehensive evaluation of the RMNet model's performance within a single location context. In this analysis, all available data from multiple flight dates were pooled together, treating each individual image as an independent sample with its corresponding maturity class label, without considering temporal sequences or flight-specific patterns.

The analysis is structured in four parts: first, an ablation study to determine the optimal network architecture; second, a comparison against traditional machine learning techniques; third, an evaluation against contemporary deep learning models; and finally, a general discussion synthesizing the key findings. Through this systematic approach, we establish RMNet's capabilities and advantages for soybean maturity classification, with particular emphasis on its precision in identifying mature plots – a critical factor for harvest timing decisions in breeding programs.

4.1.1 Ablation Study

Table 3 shows the metrics evaluated modifying dense layers from one to four and two or three convolutional layers. In Table 4, it is possible to check all input samples are from populations with equal variances, according to Levene test ($p\text{-value} > 0.05$), and there is no significant difference between accuracy, recall and f1-score; the model with three convolutional and three dense layers has presented a low significant difference for precision ($p\text{-value} < 0.05$) according to ANOVA test, but all the others are similar.

TABLE 3. Performance metrics of RMNet model variants showing accuracy, precision, recall, and F1-score results according to different combinations of convolutional and dense layers during architecture optimization.

Convolutional layers	Dense layers	Accuracy	Precision	Recall	F1-Score
2	1	0.9040	0.9353	0.9681	0.9511
	2	0.9116	0.9435	0.9631	0.9530
	3	0.9095	0.9506	0.9585	0.9543
	4	0.8702	0.9523	0.9360	0.9419
3	1	0.9087	0.9459	0.9578	0.9516
	2	0.9138	0.9459	0.9674	0.9564
	3	0.8886	0.9329	0.9657	0.9487
	4	0.8966	0.9429	0.9456	0.9440

TABLE 4. Statistical analysis results showing p-values from Levene's test for homogeneity of variances and ANOVA test for significant differences in accuracy, precision, recall, and F1-score across model configurations.

Test	Accuracy	Precision	Recall	F1-Score
Levene	0.4524	0.8529	0.8216	0.5641
ANOVA	0.4619	0.0298	0.3841	0.5493

Regarding performance, Table 5 clearly demonstrates that the model with two convolutional and one dense layer outperformed the others in terms of computational efficiency. This outcome was anticipated, given the reduced number of parameters and operations required by this streamlined architecture. Furthermore, as shown in Table 6, this model achieved superior precision in correctly classifying the *Mature* class (class 2), which is particularly important for harvest timing decisions. Based on this balance of efficiency and classification accuracy, we selected this network architecture for our final implementation, as illustrated in Figure 13. While all evaluated models demonstrated strong overall performance with accuracy metrics of

approximately 0.90, the selected architecture provides the optimal balance between computational efficiency and classification precision for the critical mature class.

TABLE 5. Computational performance analysis showing processing time per epoch (in seconds) for different RMNet architectures according to the number of convolutional and dense layers.

Convolutional layers	Dense layers	Time / epoch [s]
2	1	2.663
	2	2.709
	3	2.775
	4	2.842
3	1	2.733
	2	2.967
	3	3.017
	4	3.189

TABLE 6. Confusion matrices for RMNet model variants with different numbers of convolutional and dense layers, showing classification performance across the three maturity classes: (0) Immature, (1) Near-mature, and (2) Mature. Values represent raw counts with percentages of total observations below. (continue)

Convolutional layers		2			3		
Dense layers	1	True label \ Predicted label			True label \ Predicted label		
		0	1	2	0	1	2
		20301 36.8%	2105 3.8%	100 0.2%	20696 37.5%	1737 3.1%	74 0.1%
		985 1.8%	5744 10.4%	1669 3.0%	1266 2.3%	5989 10.8%	1144 2.1%
2	2	58 0.1%	379 0.7%	23872 43.2%	113 0.2%	705 1.3%	23492 42.5%
		0 1 2			0 1 2		
		Predicted label			Predicted label		
3	3	20613 37.3%	1838 3.3%	57 0.1%	20686 37.5%	1737 3.1%	84 0.2%
		1151 2.1%	6537 11.8%	711 1.3%	1006 1.8%	5974 10.8%	1419 2.6%
		56 0.1%	1071 1.9%	23183 42.0%	59 0.1%	457 0.8%	23794 43.1%
		0 1 2			0 1 2		
		Predicted label			Predicted label		
		20865 37.8%	1574 2.9%	68 0.1%	20216 36.6%	2269 4.1%	21 0.0%
		1282 2.3%	6280 11.4%	838 1.5%	850 1.5%	7222 13.1%	327 0.6%
		75 0.1%	1160 2.1%	23074 41.8%	271 0.5%	2414 4.4%	21625 39.2%
		0 1 2			0 1 2		
		Predicted label			Predicted label		

TABLE 6, Cont.

		Convolutional layers								
		2			3					
Dense layers										
4	True label	0	20946 37.9%	1476 2.7%	86 0.2%	True label	0	20630 37.4%	1830 3.3%	47 0.1%
		1	1356 2.5%	6107 11.1%	936 1.7%		1	1125 2.0%	6644 12.0%	630 1.1%
		2	737 1.3%	2574 4.7%	20998 38.0%		2	654 1.2%	1421 2.6%	22234 40.3%
			0	1	2			0	1	2
		Predicted label						Predicted label		

4.1.2 Comparative Results Against Traditional Machine Learning Techniques

The comparison between the proposed model and others traditional machine learning techniques led to the results: Support Vector Machine obtained the highest absolute value for accuracy (0.9220) – followed by RMNet (0.9040) and Linear Support Vector (0.8063) –, recall (0.9840) – followed by Multilayer Perceptron (0.9804) and RMNet (0.9681) – and f1-score (0.9609) – followed by RMNet (0.9511) and Linear Support Vector (0.9447) –, Ridge Regression obtained the highest absolute value for precision (0.9678) – followed by Support Vector Machine (0.9390) and RMNet (0.9353) –, K-Nearest Neighbors obtained the lowest absolute value for training time (0.55s) – followed by Multilayer Perceptron (8.15s) and Ridge Regression (16.93s) – and Decision Tree obtained the lowest absolute value for predict time (0.66s) – followed by Multilayer Perceptron (1.04s) and Linear Support Vector (1.12s) – as shown in Table 7 and Table 8.

TABLE 7. Comparative performance metrics between RMNet and traditional machine learning models, showing accuracy, precision, recall, and F1-score for soybean maturity classification.

Technique	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.6189	0.8467	0.7995	0.8209
K-Nearest Neighbors	0.7968	0.9111	0.9468	0.9278
Linear Support Vector	0.8063	0.9232	0.9676	0.9447
Multilayer Perceptron	0.7504	0.8915	0.9804	0.9327
Random Forest	0.5623	0.9275	0.8066	0.8629
Ridge Regression	0.6796	0.9678	0.6241	0.7490
Support Vector Machine	0.9220	0.9390	0.9840	0.9609
RMNet	0.9040	0.9353	0.9681	0.9511

TABLE 8. Computational efficiency comparison between RMNet and traditional machine learning models, showing training and prediction times in seconds.

Technique	Training time [s]	Predict time [s]
Decision Tree	108.95	0.66
K-Nearest Neighbors	0.55	186.33
Linear Support Vector	140.90	1.12
Multilayer Perceptron	8.15	1.04
Random Forest	106.07	2.06
Ridge Regression	16.93	8.10
Support Vector Machine	132.90	1066.20
RMNet	308.00	7.56

The computational efficiency patterns observed in Table 8 reflect the fundamental algorithmic differences between machine learning approaches. Models that exhibit fast training but slow prediction times, such as K-Nearest Neighbors (0.55s training, 186.33s prediction), represent learning algorithms that defer computation until prediction time. KNN stores all training data and performs distance calculations to all training samples for each prediction, resulting in minimal training overhead but significant prediction costs that scale linearly with dataset size. Conversely, models with longer training times but faster predictions, such as Decision Tree (108.95s training, 0.66s prediction) and Linear Support Vector (140.90s training, 1.12s prediction), represent eager learning algorithms that invest computational effort during training to build optimized decision structures. Decision trees construct hierarchical splitting rules during training, enabling rapid traversal during prediction through simple threshold comparisons. Support Vector Machine requires intensive optimization to find optimal hyperplanes during training but achieves fast predictions through simple dot product calculations with support vectors.

The extreme case of Support Vector Machine (132.90s training, 1066.20s prediction) likely reflects the complexity of the non-linear kernel computations required for this high-dimensional image classification task, where the model must compute kernel functions with numerous support vectors for each prediction. Ridge Regression shows moderate training time (16.93s) but higher prediction time (8.10s) due to matrix operations required for coefficient calculations. RMNet's computational profile (308.00s training, 7.56s prediction) reflects the typical deep learning pattern where extensive backpropagation and weight optimization during training enables efficient forward pass predictions. This analysis demonstrates that the choice between models should consider both

accuracy requirements and the specific computational constraints of the deployment environment, whether prioritizing training efficiency for model development or prediction speed for real-time applications.

The statistical analysis of Support Vector Machine and RMNet models for classifying soybean maturity stages reveals significant insights. Levene's test p-values for accuracy (0.63801) and f1-score (0.21824) both exceed 0.05, confirming homogeneity of variances. ANOVA p-values for accuracy (0.00213) and f1-score (0.00284) are well below 0.05, indicating statistically significant differences between the models. Support Vector Machine outperforms RMNet in both accuracy (0.9220 vs 0.9040) and f1-score (0.9609 vs 0.9511). While these differences are statistically significant and consistent, it is important to note that both models exhibit excellent performance, with metrics above 90%. These results indicate that while Support Vector Machine achieved the highest scores in accuracy, recall, and f1-score, RMNet consistently performed well across metrics. Additionally, RMNet's balance of precision and efficiency in training and prediction times highlights its potential for practical applications where speed and accuracy are essential. While Support Vector Machine shows a statistical edge, both models prove highly effective for the task, with the choice between them potentially influenced by other practical considerations beyond these key performance metrics.

4.1.3 Comparative Results Against Others Deep Learning Models

The comparison between the proposed model and others deep learning models led to the results: Xception obtained the highest absolute value for accuracy (0.9108) – followed by RMNet (0.9040) and ResNet152V2 (0.8991) –, precision (0.9452) – followed by ResNet50V2 (0.9390) and RMNet (0.9353) – and f1-score (0.9527) – followed by RMNet (0.9511) and ResNet101V2 (0.9455) –, ResNet152V2 obtained the highest absolute value for recall (0.9835) – followed by ResNet101V2 (0.9691) and RMNet (0.9681) – and RMNet obtained the lowest absolute value for time per epoch (2.663s) – followed by MobileNetV2 (8.665s) and EfficientNetV2B0 (15.909s) – as shown in Table 9 and Table 10.

The computational efficiency results in Table 10 reveal significant architectural differences among deep learning models and their impact on training speed. RMNet demonstrates exceptional computational efficiency at 2.663 seconds per epoch, representing a $3.3\times$ speedup compared to the second-fastest model (MobileNetV2 at 8.665s) and up to a $20.4\times$ speedup compared to the slowest

TABLE 9. Comparative performance metrics between RMNet and state-of-the-art deep learning architectures, showing accuracy, precision, recall, and F1-score for soybean maturity classification.

Technique	Accuracy	Precision	Recall	F1-Score
EfficientNetV2B0	0.3601	0.6493	0.6405	0.6377
MobileNetV2	0.8654	0.9220	0.9486	0.9326
ResNet50V2	0.8965	0.9390	0.9408	0.9381
ResNet101V2	0.8987	0.9252	0.9691	0.9455
ResNet152V2	0.8991	0.9110	0.9835	0.9453
Xception	0.9108	0.9452	0.9609	0.9527
RMNet	0.9040	0.9353	0.9681	0.9511

TABLE 10. Computational efficiency comparison between RMNet and other deep learning models, showing processing time per epoch in seconds.

Technique	Time / epoch [s]
EfficientNetV2B0	15.909
MobileNetV2	8.665
ResNet50V2	17.865
ResNet101V2	33.024
ResNet152V2	54.411
Xception	22.713
RMNet	2.663

model (ResNet152V2 at 54.411s). This dramatic efficiency advantage stems from RMNet's streamlined architecture with only two convolutional layers and minimal dense layers, specifically designed for the soybean maturity classification task.

The efficiency advantage of RMNet becomes particularly significant in practical breeding applications where models may need frequent retraining with new seasonal data or adaptation to different locations. With typical training requiring 50-100 epochs, RMNet's efficiency translates to training times of 2-4 minutes compared to 15-90 minutes for other architectures. This computational advantage, combined with statistically equivalent performance to more complex models, positions RMNet as an optimal solution for resource-constrained agricultural applications where rapid model deployment and iteration are essential.

The statistical comparison of Xception and RMNet models for classifying soybean maturity stages reveals no significant differences in overall accuracy or f1-score. Levene's test p-values for accuracy (0.7551) and f1-score (0.7372) both exceed 0.05, confirming homogeneity of variances.

ANOVA p-values for accuracy (0.3578) and f1-score (0.6261) are well above 0.05, indicating no statistically significant differences between the models. Xception slightly outperforms RMNet in both accuracy (0.9108 vs 0.904) and f1-score (0.9527 vs 0.9511), but these differences are not statistically significant. Both models exhibit excellent performance, with metrics above 0.90. Given the lack of significant differences in these key metrics, the choice between Xception and RMNet for soybean maturity stage classification may depend on other factors such as computational efficiency, model size, or specific application requirements. These results demonstrate that while RMNet and Xception achieved statistically equivalent scores in terms of accuracy, precision, recall and f1-score, the proposed technique offered a much lower computational time. In summary, RMNet showed an effective balance of predictive performance and efficiency, particularly excelling in processing speed. This makes it a viable option for scenarios where computational resources and time are critical considerations.

4.1.4 General Discussion

While RMNet did not achieve the highest overall metrics across all classes, it demonstrated superior performance in classifying the target *Mature* class, which is the most critical for harvest timing decisions. As shown in Figure 15, alternative models like Xception, ResNet152V2, Support Vector Machine, and Ridge Regression exhibit different error patterns across the three maturity classes.

Table 11 quantifies this performance difference, with RMNet achieving the highest precision in predicting the *Mature* class (0.9820), followed by ResNet101V2 (0.9785), ResNet152V2 (0.9740), and ResNet50V2 (0.9440). This class-specific excellence is particularly valuable in agricultural applications where the cost of misclassifying mature plots ready for harvest is significantly higher than other classification errors. These findings underscore RMNet's targeted effectiveness for precision agriculture applications where accurate mature crop detection directly impacts harvest timing decisions and yield optimization.

The superior performance of RMNet in mature class detection can be understood through soybean maturity physiology and the model's architectural design. During soybean maturation, plants undergo distinct physiological changes that manifest as observable color transitions captured by RGB imagery. The maturation process involves progressive senescence, characterized by chlorophyll degradation, water content reduction, and accumulation of secondary metabolites that

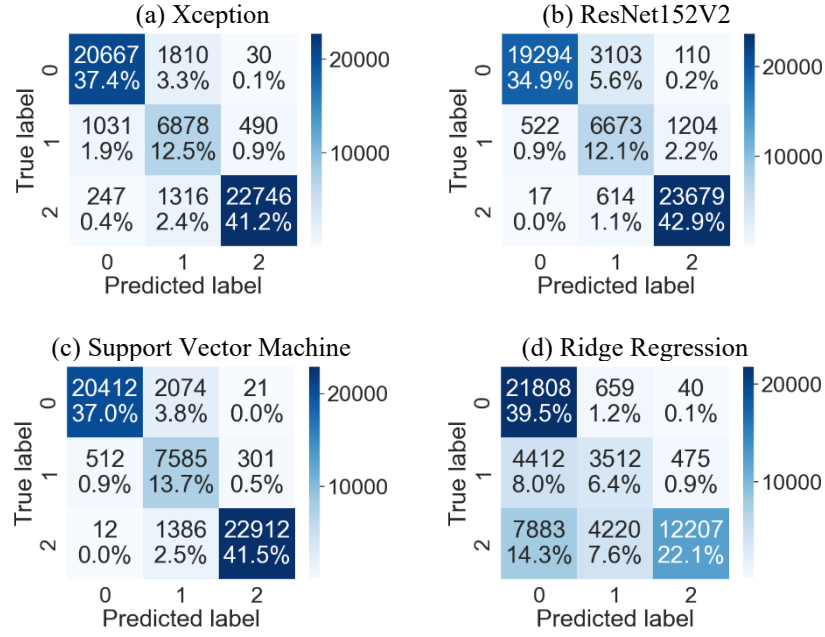


FIGURE 15. Confusion matrices from four high-performing classification models: (a) Xception, (b) ResNet152V2, (c) Support Vector Machine, and (d) Ridge Regression, visualizing their classification performance across the three maturity classes: (0) Immature, (1) Near-mature, and (2) Mature. Values represent raw counts with percentages of total observations in parentheses. Note the varying performance patterns across different maturity classes, particularly in the critical Mature class.

TABLE 11. Precision comparison for mature class detection across all tested models, highlighting RMNet's superior performance in identifying plots ready for harvest.

Technique	Precision (class <i>Mature</i>)
EfficientNetV2B0	0.1266
MobileNetV2	0.8855
ResNet50V2	0.9440
ResNet101V2	0.9785
ResNet152V2	0.9740
Xception	0.9357
Decision Tree	0.4779
K-Nearest Neighbors	0.9423
Linear Support Vector	0.7062
Multilayer Perceptron	0.6139
Random Forest	0.1209
Ridge Regression	0.5021
Support Vector Machine	0.9425
RMNet	0.9820

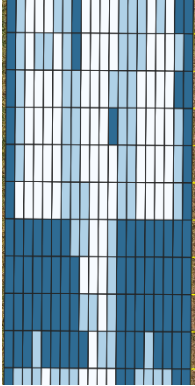
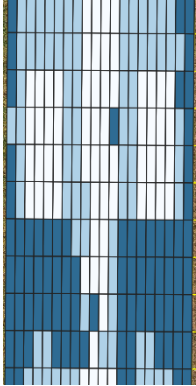
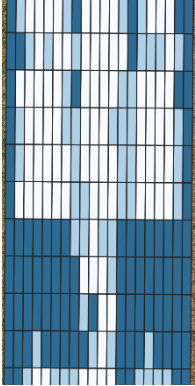
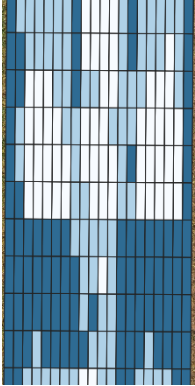
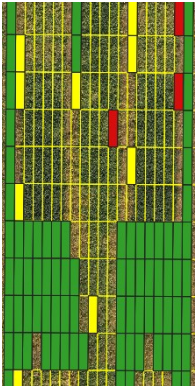
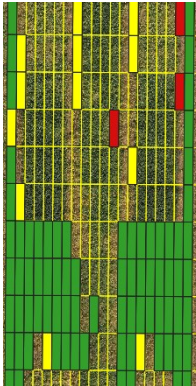
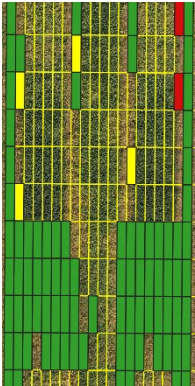
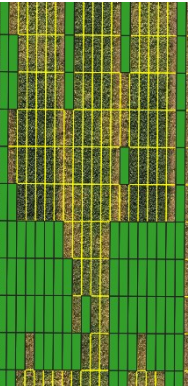
alter leaf and pod coloration (Fehr; Caviness, 1977). As soybeans transition from the R6 (full seed) to R8 (full maturity) stages, chlorophyll breakdown reveals underlying carotenoids and anthocyanins, creating the characteristic yellowing and browning patterns associated with physiological maturity.

Notably, RMNet outperforms both the best machine learning model, Support Vector Machine (0.9425), and the best deep learning model, Xception (0.9357). Statistical analysis confirms significant differences between RMNet and these top performers. For Support Vector Machine comparison, Levene's test ($p = 0.0523789$) and ANOVA ($p = 6.059e-05$) indicate a significant difference. Similarly, for Xception, Levene's test ($p = 0.05304346$) and ANOVA ($p = 0.01999$) demonstrate RMNet's superior performance. These results, combined with RMNet's previously noted computational efficiency, strongly support its effectiveness for precise mature soybean identification. The model's ability to balance high precision with efficient processing, while outperforming the best traditional machine learning and deep learning models, makes it an excellent choice for applications where both accuracy in identifying mature soybeans and computational resource management are crucial, reinforcing its value in agricultural image analysis tasks such as determining harvesting time or estimating yield.

From a practical breeding perspective, correctly identifying mature plots (ready for harvest) is significantly more critical than correctly classifying non-mature plots. Misclassifying a mature plot as immature or near-mature can result in delayed harvest, leading to seed shattering, quality degradation, and yield losses. Conversely, misclassifying an immature plot as mature may lead to premature harvest attempts, but this typically results in lower harvest efficiency rather than direct yield loss, as immature seeds are often left behind by harvesting equipment. Therefore, high precision in mature plot detection directly translates to optimized harvest timing and maximum yield recovery, making this metric the primary focus for practical agricultural applications.

Table 12 presents a portion of the orthomosaic from a test field conducted 135 days after planting, displaying the performance of Xception, Support Vector Machine and RMNet in soybean maturity classification. RMNet demonstrates impressive performance across different maturity stages, often aligning with the ground truth. It consistently outperforms traditional machine learning models, particularly in classifying *Mature* and *Near-mature* samples, and shows greater stability in predictions compared to Support Vector Machine. RMNet excels in handling subtle differences between maturity stages and maintains high precision for the *Mature* class, while

TABLE 12. Visual comparison of soybean maturity classification performance across different models (Xception, Support Vector Machine, and RMNet) against ground truth, showing both maturity maps and mature class detection.

	Xception	Support Vector Machine	RMNet (proposed)	Ground truth
Maturity map Immature Near-mature Mature				 
Class <i>Mature</i> True positive False negative False positive				

minimizing false classifications of *Immature* or *Near-mature* samples as *Mature*. Although some more complex deep learning models (like Xception) show slightly better overall performance, RMNet competes well against these advanced architectures. Its ability to generalize across different plots indicates robust performance in varying conditions. The model's simpler architecture offers an advantageous balance between accuracy and computational efficiency. In summary, RMNet emerges as a valuable tool for soybean maturity classification, providing a strong combination of accuracy, consistency, and efficiency, particularly when compared to traditional machine learning approaches and even holding its own against more complex neural networks.

Five deep learning architectures, Inception-v3, Resnet-50, VGG-16, VGG-19 and Xception, were evaluated for the classification of soybean pest images reaching accuracies of up to 0.9382, demonstrating that different architectures can achieve competitive performance in agricultural image classification tasks, with no significant difference in average performance among architectures using ANOVA. However, fine-tuning significantly improved classification rates compared to other methods (Tetila *et al.*, 2020), which reinforces the enhancement possibilities of RMNet through architectural optimization and transfer learning approaches.

In the study of Zhang *et al.* (2022), DS-SoybeanNet, a new convolutional neural network, was developed for soybean maturity classification across different training and test fields, alongside comparisons with ResNet50, MobileNetV2, Random Forest, and Support Vector Machine. While direct performance comparisons are not appropriate due to different experimental conditions, datasets, and evaluation protocols, this work demonstrates the feasibility of applying deep learning approaches to soybean maturity assessment. The study highlights the challenges of cross-field generalization and the importance of robust model architectures for agricultural applications, providing valuable insights that inform the development and evaluation of models like RMNet.

A binary prediction model for mature versus not-mature classification using time course multispectral images from a UAV-based HTP platform was developed using Random Forest (Yu *et al.*, 2016). While this approach demonstrated the potential of machine learning for soybean maturity monitoring, it was limited to binary classification. In contrast, RMNet addresses the more complex challenge of three-class discrimination (Immature, Near-mature, and Mature), providing more detailed maturity information that is valuable for breeding programs and harvest timing decisions. Both approaches

contribute to the growing body of knowledge in UAV-based crop monitoring, with each addressing different aspects of the maturity assessment challenge.

It is important to highlight the depth of this research. The study encompassed a comprehensive comparison of 13 different techniques, including both traditional machine learning and state-of-the-art deep learning models - this broad range of comparisons provides a thorough evaluation of RMNet's performance against diverse approaches. Furthermore, the research went beyond mere accuracy metrics, incorporating detailed computational time analyses for both training and prediction phases. This thorough approach provides valuable insights into the practical applicability of each model in real-world scenarios where processing speed is crucial.

The study's rigor is further evidenced by the application of statistical tests, including Levene's test and ANOVA, to validate the significance of performance differences between models. This statistical approach adds a layer of confidence to the conclusions drawn about RMNet's performance relative to other techniques.

A key strength of this research is its focused analysis on the *Mature* class, which is of primary interest in soybean breeding programs. By demonstrating RMNet's superior performance in identifying mature plots (precision of 98.20%), the study addresses a critical need in agricultural management and breeding practices. This targeted analysis, combined with the model's computational efficiency, positions RMNet as a particularly valuable tool for practical applications in the field.

The depth and breadth of this research, encompassing multiple techniques, large sample sizes, computational time analyses, statistical testing, and focused analysis on the target *Mature* class, collectively demonstrate the comprehensive nature of this study and its significant contribution to the field of soybean maturity classification using high-throughput phenotyping and deep learning techniques.

4.2 Cross-Location Analysis

Previous research in soybean maturity classification has primarily focused on single-location models, leaving significant gaps in understanding how these systems perform across diverse environments—a critical limitation for practical deployment in breeding programs. This section addresses this fundamental research gap by exploring a real-world scenario where soybean maturity is monitored at each flight date, using Location 01 as the training database and Location 02 as the deployment site. The analysis

addresses fundamental research gaps regarding model transferability and environmental adaptation, issues largely unaddressed in current literature.

The objective of this cross-location study is to simulate practical breeding program conditions where models trained at one location must be deployed across different environments. This represents a common challenge in agricultural applications, where models developed at research stations or specific breeding locations need to perform reliably when applied to diverse field conditions with varying soil types, microclimates, and management practices.

The experimental methodology employed a rigorous cross-location validation approach where all models were trained using the complete dataset from Location 01 (encompassing all available flight dates and maturity groups at that location). Each trained model was then systematically evaluated across all nine flight dates at Location 02, creating a comprehensive assessment of model transferability across temporal and spatial variations. This approach simulates real-world deployment scenarios where a single trained model must perform reliably across multiple time points and varying environmental conditions at a new location.

The integration of environmental data into the classification model marked a significant advancement in soybean maturity prediction methodology. This section examines the performance of two distinct models across different locations and growth stages: the baseline RMNet model and the environmentally-enhanced RMNet-E model, followed by an evaluation against traditional machine learning methods. The analysis reveals important insights into the value of incorporating environmental parameters in maturity classification and their impact on model performance across diverse growing conditions.

The cross-location experimental design creates 20 distinct micro-environments (11 at Location 01 + 9 at Location 02), where each unique combination of location and flight date constitutes a separate environmental condition. The training phase utilized all micro-environments from Location 01, while the evaluation phase systematically tested model performance across the nine distinct micro-environments at Location 02. This multi-environment approach captures both spatial variation between locations and temporal variation within locations, providing a comprehensive framework for evaluating model robustness across diverse agricultural conditions.

The integration of environmental data proved especially valuable during the critical maturation period (DAP 75-107), where the enhanced model demonstrated

superior capability in handling complex maturation patterns. This improvement manifested in reduced prediction variance, more accurate detection of transition onset, and better handling of edge cases during rapid maturation phases across the temporal sequence of flight dates.

The evaluation of RMNet-E's robustness and generalization capabilities across different locations and environmental conditions revealed significant insights into the model's practical applicability for real-world breeding programs. This cross-location validation approach provides critical insights into model transferability – a key requirement for practical deployment where training and application sites differ in environmental conditions, soil characteristics, and management practices. This section examines the model's performance through cross-location validation, environmental adaptation capabilities, and temporal stability analysis, providing a comprehensive assessment of its reliability in diverse agricultural settings.

4.2.1 Comparative Analysis of Base Model and Environmental Integration

The addition of environmental features to the baseline CNN architecture resulted in notable changes in model behavior and performance. While the baseline RMNet model demonstrated strong performance in single-location analysis, the integration of environmental data in RMNet-E introduced both advantages and challenges that varied across the growing season. In this cross-location validation scenario, where Location 01 served as the training environment and Location 02 as the deployment site, the environmental integration proved particularly valuable for handling the environmental differences between locations. Statistical analysis through ANOVA and Levene's tests across the nine flight dates from Location 02 revealed distinct patterns in model behavior during the critical maturation period.

The temporal analysis revealed distinct patterns in soybean maturity progression, captured by both models across the growing season. As shown in Figure 16, the classification performance metrics follow a characteristic pattern that aligns with the biological progression of soybean maturity observed in Figure 10 (B).

During the early transition period (DAP 75-89), statistical analysis revealed remarkable similarity between the two models. ANOVA tests showed no significant differences in mean performance across all metrics during this period, with p-values consistently above 0.05 for all metrics. This indicates that both models performed equivalently during the initial maturation phase when mature plants first appear.

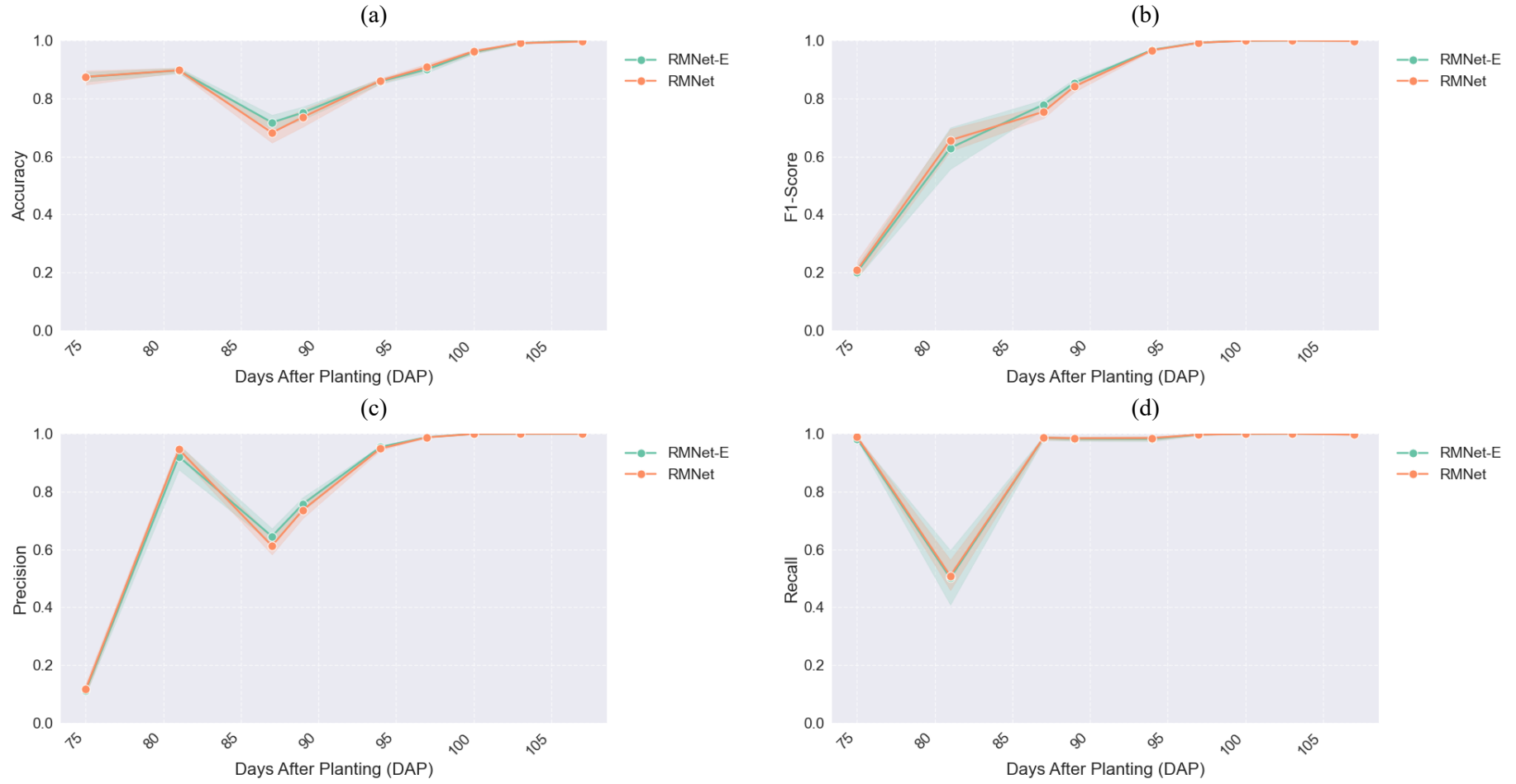


FIGURE 16. Temporal progression of model performance metrics across the soybean growing season comparing RMNet (baseline model) and RMNet-E (environmentally-enhanced model): (a) Accuracy, (b) F1-score, (c) Precision, and (d) Recall. The x-axis represents Days After Planting (DAP), while the y-axis shows the respective performance metric values (0-1 scale).

However, Levene's tests for variance equality revealed important differences in model stability during this transition period. At DAP 81, RMNet-E showed marginally significant improvements in prediction stability for precision, recall, and f1-score suggesting more consistent performance during this critical early maturation phase. By DAP 87 and 89, variance equality was restored, indicating that both models achieved similar stability levels.

The mid-season period (DAP 94-100) continued to show no significant differences between models in terms of mean performance, with ANOVA p-values remaining above 0.05 for all metrics. Levene's tests confirmed equal variances across all performance measures, indicating that both models maintained similar levels of consistency during this stable maturation phase.

Late season performance (DAP 103-107) revealed the most significant differences between the models. At DAP 107, statistical analysis showed significant differences in model performance and stability. ANOVA tests revealed significant differences in accuracy, recall, and f1-score, with RMNet-E demonstrating superior performance. Levene's tests confirmed significant differences in variance for those metrics indicating that RMNet-E not only performed better but also more consistently during the final maturation phase.

The integration of environmental features proved particularly valuable during the final maturation period, where RMNet-E demonstrated both superior mean performance and significantly lower variance in predictions. This enhanced stability manifested in more reliable detection of maturity transitions during the critical harvest timing window, which is particularly valuable for precision agriculture applications requiring consistent decision support.

Statistical validation through comprehensive ANOVA and Levene's tests across the entire maturation period revealed that while both models performed similarly during early and mid-season phases, RMNet-E's advantages became pronounced during the final maturation stage. The significant improvements in both accuracy and consistency at DAP 107 demonstrate the long-term benefits of environmental data integration for harvest timing decisions.

4.2.2 Comparative Analysis Against Traditional Machine Learning Approaches

While previous studies have compared deep learning models against traditional machine learning approaches in single-location settings, there remains a significant gap

in understanding how these methodologies perform across diverse environments throughout a growing season. This analysis addresses this research gap by evaluating model performance in a cross-location deployment scenario during the critical maturation period – a comparison notably absent in current literature but essential for practical breeding applications where models must transfer between different environmental conditions.

The cross-location validation approach simulates real-world conditions where breeding programs require models that can reliably perform across multiple sites with varying environmental characteristics. This represents a more stringent test of model robustness compared to single-location evaluations, as it challenges the model's ability to generalize beyond its training environment.

As visualized in Figure 17, the comparison between RMNet-E and traditional machine learning models (SVM, Ridge Regression, Decision Trees, Random Forest, Perceptron, LinearSVC, and kNN) revealed distinct performance patterns across the critical maturation period (DAP 75-107). Statistical analysis through ANOVA demonstrated significant differences in mean performance across all flight dates (for all metrics), with only one exception for precision at DAP 103, where models achieved more comparable precision levels. Levene's tests confirmed significant differences in variance equality across all time points and metrics, indicating substantial differences in both performance levels and prediction stability among others modeling approaches.

The Tukey analysis revealed specific pairwise comparisons between models, providing detailed insights into which approaches performed significantly better than others. At DAP 75, RMNet-E (mean accuracy: 0.874) demonstrated significant advantages over other machine learning models. Notably, RMNet-E and SVM achieved nearly identical mean accuracy, with no significant difference between them, indicating comparable top-tier performance.

The early maturation period (DAP 81-89) revealed increasingly pronounced differences between modeling approaches. At DAP 81, RMNet-E achieved the highest mean accuracy (0.897) but without significant differences from most traditional approaches, suggesting that early maturation stages present less discriminative challenges. However, by DAP 87, RMNet-E (0.717) significantly outperformed the models, while maintaining comparable performance to SVM.

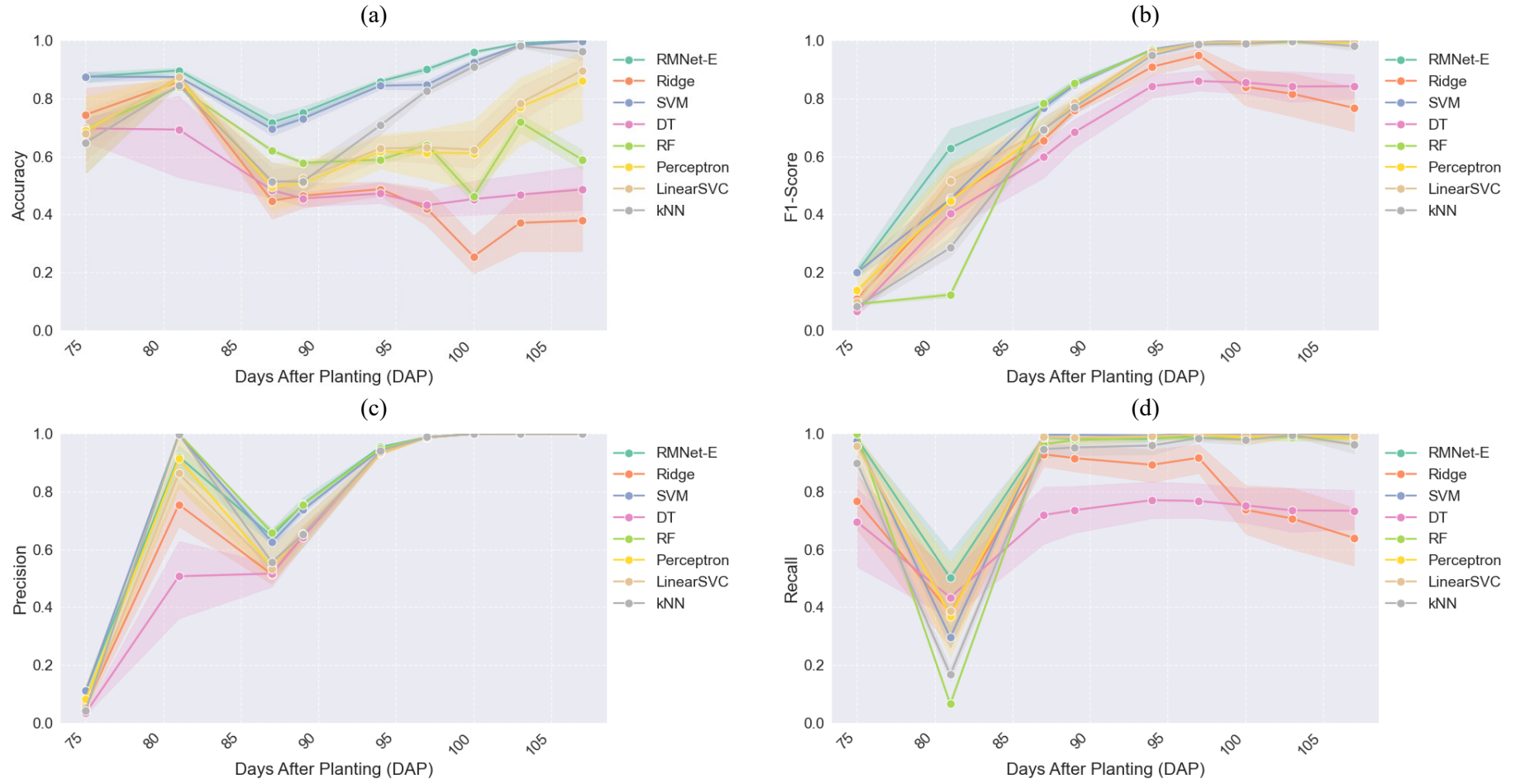


FIGURE 17. Temporal evolution of model performance metrics comparing RMNet-E against traditional machine learning approaches (SVM, Ridge Regression, Decision Trees, Random Forest, Perceptron, LinearSVC, and kNN) throughout the soybean growing season: (a) Accuracy, (b) F1-Score, (c) Precision, and (d) Recall. The x-axis represents Days After Planting (DAP), while the y-axis shows performance metric values (0-1 scale).

The mid-maturation period (DAP 89-94) demonstrated the most pronounced performance differentiation. At DAP 89, RMNet-E (0.751) significantly outperformed Ridge Regression, Decision Trees, and kNN, while maintaining statistical equivalence with SVM. By DAP 94, the performance gaps widened substantially, with RMNet-E (0.859) achieving significant advantages over all traditional approaches except SVM.

Late season analysis (DAP 97-107) revealed sustained performance advantages for RMNet-E. At DAP 97, RMNet-E (0.901) significantly outperformed all traditional approaches except SVM, including highly significant advantages over Ridge Regression, Decision Trees, and Random Forest. The final evaluation at DAP 107, RMNet-E achieved near-perfect accuracy (0.9997) and maintained significant advantages over most traditional approaches, while SVM remained the only approach achieving statistical equivalence.

Precision analysis revealed critical insights into class-specific performance capabilities. The precision improvements became more pronounced during mid-season, with DAP 81 and substantial performance gaps between approaches. By DAP 87-89, precision differences stabilized, but RMNet-E consistently maintained superior or equivalent performance compared to SVM while significantly outperforming most other traditional approaches. Recall analysis demonstrated RMNet-E's superior capability in correctly identifying positive cases across the maturation period, keeping the pattern of RMNet-E and SVM achieving top-tier performance continued throughout the evaluation period, with both approaches consistently outperforming other traditional methods.

F1-score analysis provided the most comprehensive assessment of overall model performance. RMNet-E consistently achieved the highest or near-highest f1-scores, with significant advantages over most traditional approaches while maintaining statistical equivalence with SVM at most time points.

The comprehensive Tukey analysis revealed a clear performance hierarchy across the eight modeling approaches. RMNet-E and SVM consistently formed the top performance tier, with statistical equivalence between them at most time points, while both significantly outperformed other traditional approaches. Random Forest and kNN typically occupied a middle performance tier, while Ridge Regression, Decision Trees, Perceptron, and LinearSVC showed more variable performance depending on the specific metric and time point.

The ability to accurately identify mature plants is particularly critical for optimizing harvest timing decisions in breeding programs. Figure 18 specifically

highlights the evolution of mature class precision across different models throughout the critical maturation period, revealing significant performance differences during key developmental stages when harvest timing decisions are most important.

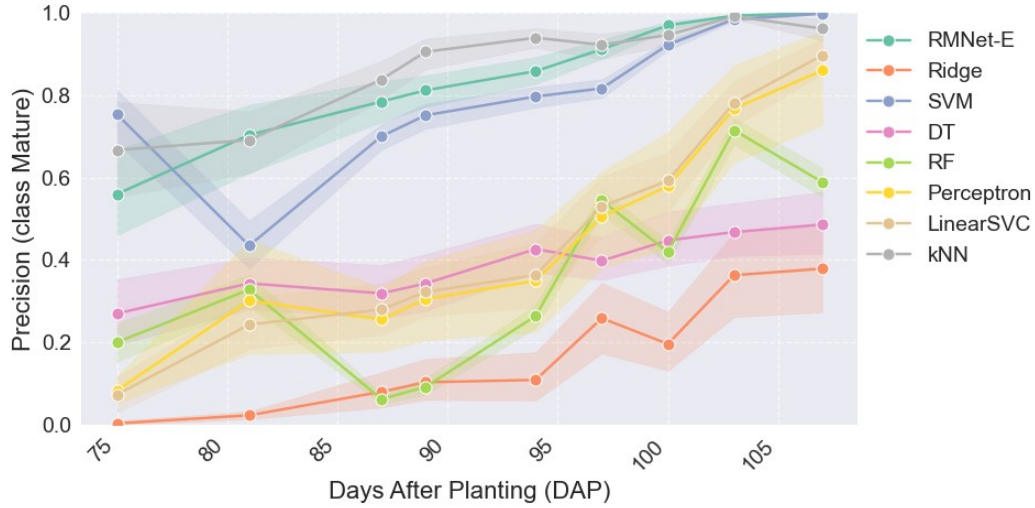


FIGURE 18. Temporal evolution of mature class precision comparing RMNet-E against traditional machine learning models (SVM, Ridge Regression, Decision Trees, Random Forest, Perceptron, LinearSVC, and kNN) throughout the soybean growing season. The x-axis represents Days After Planting (DAP), while the y-axis shows precision values (0-1 scale) specifically for the mature class.

Statistical analysis of mature class precision revealed highly significant differences between modeling approaches throughout the entire maturation period. ANOVA tests consistently achieved statistical significance across all nine flight dates. Levene's tests for variance equality also demonstrated significant differences in prediction stability, confirming that models not only performed at different levels but also exhibited substantially different consistency in mature plant identification.

The transition period (DAP 75-89) represents a critical window when mature plants first appear and begin increasing in frequency. At DAP 75, the performance hierarchy was clearly established. SVM achieved the highest mean precision (0.753), followed closely by kNN (0.667) and RMNet-E (0.560).

By DAP 81, the performance landscape shifted significantly. RMNet-E emerged as the top performer with mean precision of 0.703, significantly outperforming all other approaches including SVM, kNN, Decision Trees, Random Forest, LinearSVC, Perceptron, and Ridge Regression. This marked a critical turning point where RMNet-E's environmental integration began demonstrating clear advantages in mature plant detection.

The mid-transition period (DAP 87-89) showed the most dramatic performance differentiation. At DAP 87, kNN (0.838) achieved the highest mean precision, followed by RMNet-E (0.784) and SVM (0.700). While kNN demonstrated superior mean performance, RMNet-E showed comparable consistency. Tukey HSD analysis revealed that kNN significantly outperformed SVM, while RMNet-E maintained statistical equivalence with SVM.

At DAP 89, kNN achieved the highest mean precision (0.905), followed by RMNet-E (0.812) and SVM (0.751). Tukey analysis showed that kNN significantly outperformed RMNet-E, SVM, and all other traditional approaches. However, RMNet-E demonstrated superior stability with lower variance compared to kNN, indicating more consistent performance across different environmental conditions.

Late season analysis (DAP 94-107) revealed the full maturation of model capabilities during the critical harvest window when mature plant identification becomes most operationally important. This period was specifically selected based on soybean physiological development patterns observed in Location 02, where mature plants transitioned from representing 41.8% of the population at DAP 94 to 99.9% at DAP 107. This interval corresponds to the R7-R8 reproductive stages in soybean development, when pods reach physiological maturity and moisture content drops to optimal harvest levels. The rapid increase in mature plant frequency—from 4,882 mature plants (41.8%) at DAP 94 to 11,662 mature plants (99.9%) at DAP 107—represents the critical decision window for harvest timing in both breeding evaluation and commercial production.

As quantified in Table 13, RMNet-E demonstrated superior performance with the highest precision for mature plant detection (0.993) combined with the lowest variance (0.0052), indicating both accurate and highly consistent predictions. This physiologically-defined late season period provides the most relevant assessment of mature plant detection capabilities for practical breeding applications. During this interval, accurate mature plant identification directly impacts harvest efficiency, seed quality preservation, and yield optimization.

The temporal progression within this critical period demonstrates the practical importance of consistent mature plant detection. At DAP 94, when mature plants comprised 41.8% of the population, kNN reached peak performance (0.939) while RMNet-E achieved 0.858, both significantly outperforming SVM (0.797) and all other traditional approaches.

TABLE 13. Mature plant detection performance during the critical harvest window (DAP 94-107) comparing traditional machine learning approaches (SVM, Ridge Regression, Random Forest, kNN) with RMNet-E. The table presents precision values specifically for the mature class and associated variance measures. Higher precision indicates better accuracy in identifying mature plants, while lower variance indicates greater consistency in predictions across different environmental conditions. Note that RMNet-E achieves both the highest precision and lowest variance, demonstrating superior reliability for harvest timing decisions in practical breeding applications.

Technique	Precision (class <i>Mature</i>)	Variance
SVM	0.984	0.0068
Ridge	0.363	0.1759
RF	0.715	0.0559
kNN	0.991	0.0093
RMNet-E	0.993	0.0052

At DAP 97, corresponding to 65.6% mature plant frequency, RMNet-E achieved 0.911 precision, significantly outperforming all approaches except kNN (0.922) and SVM (0.816). This mid-harvest window represents optimal timing for many soybean varieties, where accurate mature plant identification directly impacts harvest efficiency and seed quality. By DAP 100, when mature plants dominated at 87.9% of the population, RMNet-E reached 0.969 precision with exceptional stability, significantly outperforming SVM (0.922) and all other traditional methods.

The final stages of the harvest window demonstrated peak model performance when mature plant identification becomes most critical for preventing over-maturity and associated yield losses. At DAP 103, with 97.0% mature plants, RMNet-E achieved near-perfect precision (0.993) with remarkable consistency, significantly outperforming all approaches except SVM (0.984) and kNN (0.991). The final evaluation at DAP 107, representing complete physiological maturity (99.9% mature plants) showed RMNet-E reaching perfect precision (0.999) with minimal variance, maintaining statistical equivalence with SVM (0.998) while significantly outperforming all other approaches.

Statistical validation through comprehensive ANOVA and Tukey HSD testing confirms that the observed differences in mature plant detection capabilities are statistically significant and consistent across the critical maturation period. These performance differences represent fundamental capabilities rather than random variation. Levene's tests consistently showing significant variance differences further confirm that models differ substantially in their reliability for mature plant identification across varying environmental conditions. Analysis of variance components revealed that RMNet-E maintained among the lowest prediction variance across all models,

particularly during critical transition periods, suggesting superior reliability in real-world applications.

This combination of high precision and low variance in mature plant detection is particularly valuable for practical breeding applications. In both breeding programs and production agriculture, the cost of misclassifying mature plants ready for harvest significantly impacts yield quantity, quality, and operational efficiency. The statistical evidence demonstrates that RMNet-E provides optimal balance of accuracy and consistency for mature plant identification, making it particularly suitable for automated harvest timing recommendations in field-scale implementation of maturity assessment systems.

These findings demonstrate that while traditional machine learning approaches can perform adequately in controlled single-location scenarios, they often struggle with the cross-location deployment challenge during the critical maturation period, particularly for mature plant detection when harvest timing decisions are most important. RMNet-E provides a more robust and reliable solution for soybean maturity classification across diverse environmental conditions during the critical harvest window. Its superior handling of cross-location deployment and consistent performance across varying micro-environments make it particularly suitable for practical agricultural applications where models trained at one location must be reliably deployed across multiple breeding sites during the critical maturation period. These results address critical knowledge gaps regarding model transferability and environmental adaptation, providing novel insights into cross-location performance that significantly advance the current state of knowledge in soybean maturity classification.

4.3 General Discussion

The comprehensive evaluation of RMNet and RMNet-E across both single-location and cross-location scenarios revealed significant insights into their technical capabilities and practical applicability for soybean maturity classification in diverse agricultural settings. This research addresses critical gaps in understanding model transferability and environmental adaptation – issues largely unaddressed in current literature but essential for practical deployment in breeding programs.

The single-location analysis established RMNet's effectiveness through systematic comparison against 13 different approaches, including both traditional machine learning and state-of-the-art deep learning models. The ablation study confirmed

that the streamlined architecture with two convolutional layers and one dense layer provided optimal balance between computational efficiency and classification performance. This represents a $3.3\times$ speedup compared to the second-fastest deep learning model and up to a $20.4\times$ speedup compared to complex architectures like ResNet152V2.

Most importantly, RMNet achieved superior performance in mature class detection, outperforming both the best traditional machine learning approach (SVM) and the best deep learning model (Xception). Statistical analysis confirmed these differences as significant, demonstrating RMNet's targeted effectiveness for the most critical classification task in harvest timing decisions. This class-specific excellence addresses the fundamental agricultural requirement where misclassifying mature plots ready for harvest carries significantly higher costs than other classification errors.

The cross-location analysis revealed the transformative impact of environmental data integration through RMNet-E. When deployed from Location 01 (training) to Location 02 (testing), the model demonstrated remarkable transferability across diverse micro-environments created by different combinations of location and flight dates. The integration of environmental features proved particularly valuable during critical periods, with statistical analysis showing significant improvements in both performance and stability.

During the early transition period, while both RMNet and RMNet-E showed equivalent mean performance, Levene's tests revealed that RMNet-E achieved superior prediction stability. The late season analysis demonstrated RMNet-E's most significant advantages during the critical harvest window – this period, corresponding to the R7-R8 reproductive stages when mature plants transitioned from 41.8% to 99.9% of the population, represents the most operationally critical window for harvest timing decisions.

The cross-location comparison against traditional machine learning approaches revealed distinct performance hierarchies throughout the growing season. Statistical analysis through ANOVA demonstrated significant differences in mean performance across all flight dates, with Levene's tests confirming substantial differences in prediction stability among modeling approaches. RMNet-E and SVM consistently formed the top performance tier, achieving statistical equivalence at most time points while significantly outperforming other traditional approaches. The mature class precision analysis revealed

RMNet-E's superior capabilities during the critical harvest window, achieving the highest precision combined with the lowest variance among all evaluated approaches.

The superior performance in mature class detection translates directly to optimized harvest timing and maximum yield recovery. Field validation demonstrates successful integration with existing agricultural management systems, with the model's ability to process UAV-captured images providing valuable decision support for field operations. The reduced infrastructure requirements compared to complex deep learning solutions ensure accessibility across operations of varying sizes while maintaining positive return on investment.

This research's methodological rigor is evidenced through systematic application of statistical tests, including Levene's tests and ANOVA, across all comparisons. The comprehensive evaluation encompassing 13 different techniques, detailed computational time analyses, and focused analysis on the critical mature class provides thorough validation of model capabilities. The cross-location validation approach simulates real-world conditions where models trained at one location must perform reliably across different environmental conditions—a critical requirement largely unaddressed in current literature.

The integration of environmental data represents a significant methodological advancement, demonstrating how auxiliary information can enhance model robustness and transferability. The temporal analysis across nine flight dates provides unprecedented insights into model behavior throughout the complete maturation cycle, revealing performance patterns aligned with soybean physiological development.

RMNet's robust performance across locations and conditions, coupled with its practical advantages in implementation, positions it as a reliable and valuable tool for soybean maturity classification in diverse agricultural settings. The model's ability to maintain high accuracy while adapting to local environmental conditions demonstrates its potential for broad-scale deployment in breeding programs and agricultural management systems, representing a significant advancement in agricultural technology application.

4.4 Limitations and Future Work

This section examines the current limitations of the RMNet architecture and explores promising directions for future research. While the model demonstrates strong

performance in soybean maturity classification, several constraints remain to be addressed, and emerging technologies offer significant opportunities for advancement.

4.4.1 Current Technical and Operational Limitations

While RMNet demonstrates strong performance in soybean maturity classification, several limitations warrant consideration. The model's reliance on RGB imagery, while practical for widespread deployment, potentially limits the capture of subtle spectral changes during maturation. Multispectral or hyperspectral data might provide additional valuable information for classification, particularly during early transition phases where RGB differences are less pronounced.

The current implementation faces challenges in extreme weather conditions, particularly during periods of heavy cloud cover or unusual lighting conditions. Although the RMNet-E variant shows improved stability through environmental data integration, the model's performance can still be affected by severe weather events that impact image quality. Analysis of classification accuracy during such conditions reveals increased variance in predictions, suggesting room for improvement in robust feature extraction under challenging environmental conditions.

Data availability and diversity presents another significant limitation, particularly regarding the diversity of environmental conditions represented in the training dataset. While the current study encompasses multiple locations and growing seasons, broader geographical and temporal coverage would enhance the model's generalization capabilities. The limited availability of standardized, high-quality labeled data across different regions and growing conditions constrains the model's adaptability to new environments without additional training.

4.4.2 Environmental Coverage and Adaptation Challenges

Environmental data integration, while beneficial, currently covers a limited range of environmental parameters. The present approach incorporates basic meteorological data but expanding the environmental feature set to include additional variables such as detailed soil conditions, microclimate variations, nutrient availability, and historical weather patterns could significantly enhance the model's predictive capabilities and environmental adaptation.

The model's performance across different soybean varieties and genetic backgrounds requires further investigation. While current results demonstrate good generalization within tested varieties, the genetic diversity of soybean germplasm

suggests potential challenges in maintaining consistent performance across all varieties. This limitation particularly affects breeding programs working with diverse genetic materials.

Cross-regional transferability remains partially validated, with current testing limited to specific geographical regions. The model's performance in significantly different climatic zones, soil types, or agricultural management systems requires comprehensive evaluation before broad-scale deployment can be recommended.

4.4.3 Future Research Directions and Potential Improvements

Several promising research directions emerge from the current limitations. First, the integration of advanced sensing technologies, particularly multispectral and thermal imaging, could enhance the model's ability to detect subtle phenological changes.

The development of more sophisticated environmental data integration methods represents another crucial research direction. Future work should explore advanced fusion techniques for combining multiple data streams, potentially incorporating temporal dynamics through recurrent neural network architectures. This could improve the model's ability to predict maturation trajectories rather than just current states.

Transfer learning techniques offer promising avenues for addressing the challenge of limited data availability in new environments. Research into efficient adaptation mechanisms could enable rapid model customization for different regions and varieties while maintaining core classification capabilities. Initial experiments with transfer learning show potential for reducing the required training data when adapting to new locations.

Transformer architectures represent a particularly promising direction for future research in soybean maturity classification. These models have demonstrated remarkable success in computer vision tasks by capturing long-range dependencies and global context. Unlike traditional CNNs that process images through local convolutional operations, transformers employ self-attention mechanisms that allow each part of the image to interact with all other parts, potentially capturing more complex relationships between visual features across the entire plot.

For agricultural applications specifically, transformers offer several advantages. Their ability to model global relationships could better capture the overall maturity state of a plot rather than focusing on individual plants. The self-attention mechanism could also help distinguish between relevant maturity indicators and irrelevant features such as

weeds or soil variations. Additionally, recent advances in vision transformers have demonstrated superior performance in handling variable lighting conditions and occlusions – common challenges in field imagery.

The integration of transformers with environmental data presents an especially compelling opportunity. Multi-modal transformer architectures could more effectively fuse image features with environmental parameters, capturing complex interactions between visual cues and growing conditions. This approach could significantly improve prediction stability during transition periods when visual cues alone may be ambiguous.

The integration of RMNet with other agricultural management systems presents opportunities for developing comprehensive crop management solutions. Future research should explore the synchronization of maturity classification with yield prediction, disease detection, and harvest optimization systems. This integration could provide valuable decision support tools for precision agriculture applications.

Automated model adaptation mechanisms represent another important research direction. Development of self-learning capabilities that allow the model to continuously improve its performance based on new data could enhance long-term reliability and reduce maintenance requirements. This could include automated feature selection and architecture optimization based on local conditions and requirements.

The expansion of environmental feature analysis to include more complex interactions and longer-term patterns could improve prediction accuracy during critical transition periods. Research into advanced environmental modeling techniques and their integration with deep learning architectures could provide more robust predictions under varying conditions.

These future research directions, combined with addressing current limitations, present a clear pathway for enhancing RMNet's capabilities and expanding its practical applications. The continued development of this technology promises to further improve the efficiency and reliability of soybean maturity classification in agricultural settings, ultimately contributing to more sustainable and productive farming practices.

5 CONCLUSIONS

This research successfully developed and validated RMNet, an innovative convolutional neural network architecture specifically designed for soybean maturity classification using UAV-based RGB imagery. The comprehensive study addressed all four primary objectives established at the outset, making significant contributions to the field of high-throughput phenotyping and agricultural automation while addressing critical gaps in cross-location model transferability.

Regarding the first objective to develop an efficient and accurate deep learning model for soybean maturity classification, RMNet demonstrated exceptional performance with an accuracy of 0.9040, precision of 0.9353, recall of 0.9681, and an f1-score of 0.9511. Most importantly, the model achieved the highest precision in mature plot classification (0.9820) among all 13 tested approaches, significantly outperforming both traditional machine learning methods and state-of-the-art deep learning architectures including Xception (0.9357) and Support Vector Machine (0.9425). This superior precision in mature classification, combined with remarkable computational efficiency (2.663 seconds per epoch – representing a $3.3\times$ to $20.4\times$ speedup over competing architectures), establishes RMNet as a particularly valuable tool for practical agricultural applications where both accuracy and processing speed are critical for harvest timing decisions.

For the second objective concerning model performance across locations and multiple environmental conditions, the environmental data-enhanced variant, RMNet-E, demonstrated exceptional cross-location transferability, addressing a fundamental research gap largely unaddressed in current literature. When deployed from Location 01 to Location 02, the model maintained consistent performance, successfully generalizing across 20 distinct micro-environments. The integration of environmental data proved particularly valuable during critical transition periods, where RMNet-E showed superior stability in handling complex plant phenotype-environment interactions. Statistical analysis revealed significant improvements in both performance and prediction consistency during the final maturation phase, with RMNet-E achieving the highest precision (0.993) combined with the lowest variance (0.0052) among all evaluated approaches during the critical harvest window.

In addressing the third objective of comparative evaluation, RMNet was benchmarked against seven traditional machine learning approaches and six

contemporary deep learning architectures through systematic statistical analysis using ANOVA and Tukey tests. While some models like Support Vector Machine and Xception showed competitive overall performance, RMNet's superior precision in mature plot detection and computational efficiency make it particularly suitable for large-scale breeding operations. The cross-location comparison revealed distinct performance hierarchies, with RMNet-E and SVM consistently forming the top performance tier while significantly outperforming other traditional approaches. The comprehensive evaluation framework, incorporating performance metrics, computational efficiency, and statistical validation, provides unprecedented insights into practical implementation considerations for agricultural phenotyping systems.

The fourth objective concerning practical applicability was successfully addressed through the RMNet's efficient architecture, which enables deployment on moderate computing hardware while maintaining high classification accuracy. The model's superior performance in mature class detection translates directly to optimized harvest timing and maximum yield recovery – critical factors in both breeding programs and production agriculture. Field validation confirmed successful integration with existing agricultural management systems, with the model's ability to process UAV-captured images providing valuable real-time decision support for field operations. The reduced infrastructure requirements compared to complex deep learning solutions ensure accessibility across operations of varying sizes while maintaining positive return on investment.

The scientific contributions of this research have been recognized through the article 'Convolutional Neural Networks for Soybean Maturity Classification Using UAV RGB Images', which covers the analysis discussed in section 4.1. Additionally, a second article titled 'A Multimodal Deep Neural Network for Cross-Location Soybean Maturity Classification Using UAV RGB Images and Environmental Data' focusing on the groundbreaking cross-location analysis and environmental data integration is currently in preparation and will be submitted following the defense of this dissertation, addressing the critical research gap in cross-location model transferability.

Future research directions should explore the integration of transformer architectures, which have shown remarkable potential in computer vision tasks. Unlike traditional CNNs that process images through local convolutional operations, transformers employ self-attention mechanisms that allow each part of the image to interact with all other parts, potentially capturing more complex relationships between visual features across entire plots.

For agricultural applications specifically, transformers offer several advantages, including better modeling of global relationships to capture overall maturity states, improved distinction between relevant maturity indicators and irrelevant features, and superior performance in handling variable lighting conditions and occlusions. The integration of transformers with environmental data presents an especially compelling opportunity, as multi-modal transformer architectures could more effectively fuse image features with environmental parameters to capture complex interactions between visual cues and growing conditions.

Additional enhancements could include the integration of multispectral or hyperspectral data to capture subtle phenological changes currently missed by RGB imagery alone, expanding the environmental feature set to include soil conditions and microclimate data, and developing more sophisticated data fusion techniques to improve predictive capabilities across diverse growing conditions. Transfer learning approaches also offer promising avenues for addressing the challenge of limited data availability in new environments, with initial experiments showing potential for reducing required training data when adapting to new locations.

This research represents a significant advancement in agricultural technology, offering a practical solution for automated soybean maturity classification that balances accuracy, efficiency, and scalability. The successful development and validation of RMNet and its environmentally-enhanced variant RMNet-E contribute to the broader goal of modernizing breeding programs and improving agricultural productivity through advanced phenotyping technologies. Particularly, RMNet-E's breakthrough in cross-location transferability and environmental adaptation addresses critical deployment challenges that have limited the practical application of agricultural AI systems, ultimately supporting more sustainable and productive farming practices through reliable, location-independent maturity assessment capabilities.

REFERENCES

- ALABI, Tunrayo R. *et al.* Estimation of soybean grain yield from multispectral high-resolution UAV data with machine learning models in West Africa. **Remote Sensing Applications: Society and Environment**, v. 27, p. 100782, Aug. 2022.
- ALLIPRANDINI, Luís Fernando *et al.* Understanding Soybean Maturity Groups in Brazil: Environment, Cultivar Classification, and Stability. **Crop Science**, v. 49, n. 3, p. 801–808, May 2009.
- ALONSO, Jose M.; STEPANOVA, Anna N. (EDS.). **Plant Functional Genomics: Methods and Protocols**. New York, NY: Springer New York, 2015. v. 1284
- ALTALAK, Maha *et al.* Smart Agriculture Applications Using Deep Learning Technologies: A Survey. **Applied Sciences**, v. 12, n. 12, p. 5919, 10 Jun. 2022.
- AMARAL, Lúgia De Oliveira *et al.* Artificial Neural Network for Discrimination and Classification of Tropical Soybean Genotypes of Different Relative Maturity Groups. **Frontiers in Plant Science**, v. 13, p. 814046, 12 Jul. 2022.
- ARAUS, José L.; KEFAUVER, Shawn C. Breeding to adapt agriculture to climate change: affordable phenotyping solutions. **Current Opinion in Plant Biology**, v. 45, p. 237–247, Oct. 2018.
- BATISTA, Thiago Barbosa *et al.* A Reliable Method to Recognize Soybean Seed Maturation Stages Based on Autofluorescence-Spectral Imaging Combined With Machine Learning Algorithms. **Frontiers in Plant Science**, v. 13, p. 914287, 14 Jun. 2022.
- BI, Luning *et al.* A transformer-based approach for early prediction of soybean yield using time-series images. **Frontiers in Plant Science**, v. 14, p. 1173036, 20 Jun. 2023.
- BRUNS, H. Arnold. A Survey of Factors Involved in Crop Maturity. **Agronomy Journal**, v. 101, n. 1, p. 60–66, Jan. 2009.
- CAMPBELL, James B.; WYNNE, Randolph H. **Introduction to remote sensing**. 5th ed. New York: Guilford Press, 2011.
- CHANG, Anjin *et al.* Unmanned Aircraft System- (UAS-) Based High-Throughput Phenotyping (HTP) for Tomato Yield Estimation. **Journal of Sensors**, v. 2021, n. 1, p. 8875606, Jan. 2021.
- CHAWADE, Aakash *et al.* High-Throughput Field-Phenotyping Tools for Plant Breeding and Precision Agriculture. **Agronomy**, v. 9, n. 5, p. 258, 22 May 2019.
- CHO, Minsu; HEGDE, Chinmay. Reducing the Search Space for Hyperparameter Optimization Using Group Sparsity. *In*: ICASSP 2019 - 2019 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING (ICASSP). **ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. Brighton, United Kingdom: IEEE, May 2019. Available at: <<https://ieeexplore.ieee.org/document/8682434/>>. Accessed on: 28 may. 2025

CHOLLET, François. **Deep Learning with Python, Second Edition**. 2. ed. New York: Manning Publications Co. LLC, 2021.

CHRISTENSON, Brent S. *et al.* Predicting Soybean Relative Maturity and Seed Yield Using Canopy Reflectance. **Crop Science**, v. 56, n. 2, p. 625–643, Mar. 2016.

DETRING, Justus *et al.* **Quality Assurance of Hyperspectral Imaging Systems for Neural Network supported Plant Phenotyping**. , 16 Jul. 2024. Available at: <<https://www.researchsquare.com/article/rs-4648326/v1>>. Accessed on: 5 jun. 2025

DOSOVITSKIY, Alexey *et al.* AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE. 2021.

FEHR, Walter R.; CAVINESS, Charles E. Stages of Soybean Development. Agriculture and Home Economics Experiment Station, Iowa State University of Science and Technology. n. Special Report 80, p. 11, 18 May 1977.

FERREIRA, Rogério Silva. DETERMINAÇÃO DE GRUPO DE MATURAÇÃO DE SOJA Glycine max L. Merrill. 2021.

FRANCHINI, Julio Cezar *et al.* Uso de imagens aéreas obtidas com drones em sistemas de produção de soja. **Embrapa Soja**, Embrapa Soja. n. 1^a, p. 39, Dec. 2018.

GÉRON, Aurélien. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow**. 2. ed. Canada: O'Reilly Media, 2019.

GILL, Taqdeer *et al.* A Comprehensive Review of High Throughput Phenotyping and Machine Learning for Plant Stress Phenotyping. **Phenomics**, v. 2, n. 3, p. 156–183, Jun. 2022.

GOMES, José Verleandson dos Santos *et al.* **CLASSIFICAÇÃO DO CICLO DA SOJA BASEADO NO GRUPO DE MATURIDADE RELATIVA.pdf**. , Sep. 2016.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. MIT Press, , 2016. Available at: <<http://www.deeplearningbook.org>>

GUO, Yixin *et al.* Automatic and Accurate Acquisition of Stem-Related Phenotypes of Mature Soybean Based on Deep Learning and Directed Search Algorithms. **Frontiers in Plant Science**, v. 13, p. 906751, 11 Jul. 2022.

GURMU, F.; MOHAMMED, H.; ALEMAW, G. Genotype X environment interactions and stability of soybean for grain yield and nutrition quality. **African Crop Science Journal**, v. 17, n. 2, 4 May 2010.

HALEWOOD, Michael *et al.* Plant genetic resources for food and agriculture: opportunities and challenges emerging from the science and information technology revolution. **New Phytologist**, v. 217, n. 4, p. 1407–1419, Mar. 2018.

HAMUDA, Esmael; GLAVIN, Martin; JONES, Edward. A survey of image processing techniques for plant extraction and segmentation in the field. **Computers and Electronics in Agriculture**, v. 125, p. 184–199, Jul. 2016.

HAMZA, Muhammad *et al.* Global Impact of Soybean Production: A Review. **Asian Journal of Biochemistry, Genetics and Molecular Biology**, v. 16, n. 2, p. 12–20, 30 Jan. 2024.

HE, Haotian *et al.* Recognition of soybean pods and yield prediction based on improved deep learning model. **Frontiers in Plant Science**, v. 13, p. 1096619, 13 Jan. 2023.

HU, Jingyu *et al.* UAV-Based Remote Sensing for Soybean FVC, LCC, and Maturity Monitoring. **Agriculture**, v. 13, n. 3, p. 692, 16 Mar. 2023.

JOHANSEN, Kasper *et al.* Unmanned Aerial Vehicle-Based Phenotyping Using Morphometric and Spectral Analysis Can Quantify Responses of Wild Tomato Plants to Salinity Stress. **Frontiers in Plant Science**, v. 10, p. 370, 29 Mar. 2019.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. A review of the use of convolutional neural networks in agriculture. **The Journal of Agricultural Science**, v. 156, n. 3, p. 312–322, Apr. 2018.

KHAN, Asim *et al.* Tomato maturity recognition with convolutional transformers. **Scientific Reports**, v. 13, n. 1, p. 22885, 21 Dec. 2023.

KRAUSE, Matheus D. *et al.* Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors. **Agronomy Journal**, v. 117, n. 1, p. e70023, Jan. 2025.

KUMAR, Santosh *et al.* Deep learning framework for recognition of cattle using muzzle point image pattern. **Measurement**, v. 116, p. 1–17, Feb. 2018.

LOPEZ-CRUZ, Marco *et al.* Regularized selection indices for breeding value prediction using hyper-spectral image data. **Scientific Reports**, v. 10, n. 1, p. 8195, 18 May 2020.

LORENCE, Argelia; MEDINA JIMENEZ, Karina (EDS.). **High-Throughput Plant Phenotyping: Methods and Protocols**. New York, NY: Springer US, 2022. v. 2539

MAIMAITIJANG, Maitiniyazi *et al.* Soybean yield prediction from UAV using multimodal data fusion and deep learning. **Remote Sensing of Environment**, v. 237, p. 111599, Feb. 2020.

MEHTA, Tanvi *et al.* Intensification of Agriculture using Deep Learning and Machine Learning: A Survey. *In*: 2021 10TH INTERNATIONAL CONFERENCE ON INTERNET OF EVERYTHING, MICROWAVE ENGINEERING, COMMUNICATION AND NETWORKS (IEMECON). **2021 10th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)**. Jaipur, India: IEEE, 1 Dec. 2021. Available at: <<https://ieeexplore.ieee.org/document/9689196/>>. Accessed on: 21 sep. 2023

MENDOZA-BERNAL, José; GONZÁLEZ-VIDAL, Aurora; SKARMETA, Antonio F. A Convolutional Neural Network approach for image-based anomaly detection in smart agriculture. **Expert Systems with Applications**, v. 247, p. 123210, Aug. 2024.

MILLER, Mark J.; SONG, Qijian; LI, Zenglu. Genomic selection of soybean (*Glycine max*) for genetic improvement of yield and seed composition in a breeding context. **The Plant Genome**, v. 16, n. 4, p. e20384, Dec. 2023.

MOEINIZADE, Saba *et al.* An applied deep learning approach for estimating soybean relative maturity from UAV imagery to aid plant breeding decisions. **Machine Learning with Applications**, v. 7, p. 100233, Mar. 2022.

MORENO-REVELO, Mónica Y. *et al.* Enhanced Convolutional-Neural-Network Architecture for Crop Classification. **Applied Sciences**, v. 11, n. 9, p. 4292, 10 May 2021.

NEUMAIER, Norman *et al.* Estádios de desenvolvimento da cultura de soja. **Embrapa**, Embrapa. 2000.

NIEDBAŁA, Gniewko *et al.* Application of Artificial Neural Network Sensitivity Analysis to Identify Key Determinants of Harvesting Date and Yield of Soybean (*Glycine max* [L.] Merrill) Cultivar Augusta. **Agriculture**, v. 12, n. 6, p. 754, 25 May 2022.

PAN, Sinno Jialin; YANG, Qiang. A Survey on Transfer Learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 10, p. 1345–1359, Oct. 2010.

PASALA, Ratnakumar; PANDEY, Brij Bihari. Plant phenomics: High-throughput technology for accelerating genomics. **Journal of Biosciences**, v. 45, n. 1, p. 111, Dec. 2020.

RAM, Billy G. *et al.* A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. **Computers and Electronics in Agriculture**, v. 222, p. 109037, Jul. 2024.

RANI, Reena *et al.* Analysis of genotype $\times$ environment interactions for agronomic traits of soybean (*Glycine max* [L.] Merr.) using association mapping. **Frontiers in Genetics**, v. 13, p. 1090994, 5 Jan. 2023.

REN, Chengjuan; KIM, Dae-Kyoo; JEONG, Dongwon. A Survey of Deep Learning in Agriculture: Techniques and Their Applications. **Journal of Information Processing Systems**, v. 16, n. 5, p. 1015–1033, 31 Oct. 2020.

REN, Pengting *et al.* Estimation of Soybean Yield by Combining Maturity Group Information and Unmanned Aerial Vehicle Multi-Sensor Data Using Machine Learning. **Remote Sensing**, v. 15, n. 17, p. 4286, 31 Aug. 2023.

REYNOLDS, Matthew P.; BRAUN, Hans-Joachim. Wheat Improvement. *In*: REYNOLDS, Matthew P.; BRAUN, Hans-Joachim (Eds.). **Wheat Improvement**. Cham: Springer International Publishing, 2022. p. 3–15.

SANKARAN, Sindhuja *et al.* High-throughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. **Computers and Electronics in Agriculture**, v. 151, p. 84–92, Aug. 2018.

SANTOS, Anisio P. *et al.* Convolutional Neural Networks for the Molecular Detection of COVID-19. *In*: NALDI, Murilo C.; BIANCHI, Reinaldo A. C. (Eds.). **Intelligent**

Systems. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 2023. v. 14196 p. 51–62.

SARIĆ, Rijad *et al.* Applications of hyperspectral imaging in plant phenotyping. **Trends in Plant Science**, v. 27, n. 3, p. 301–315, Mar. 2022.

SETIYONO, T. D. *et al.* Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. **Field Crops Research**, v. 100, n. 2–3, p. 257–271, Feb. 2007.

SINGH, Arti *et al.* Machine Learning for High-Throughput Stress Phenotyping in Plants. **Trends in Plant Science**, v. 21, n. 2, p. 110–124, Feb. 2016.

SINGH, Gurwinder; SETHI, Ganesh Kumar; SINGH, Sartajvir. Survey on Machine Learning and Deep Learning Techniques for Agriculture Land. **SN Computer Science**, v. 2, n. 6, p. 487, Nov. 2021.

SONG, Wenwen *et al.* Standard Cultivar Selection and Digital Quantification for Precise Classification of Maturity Groups in Soybean. **Crop Science**, v. 59, n. 5, p. 1997–2006, Sep. 2019.

SOZZANI, Rosangela; BENFEY, Philip N. High-throughput phenotyping of multicellular organisms: finding the link between genotype and phenotype. **Genome Biology**, v. 12, n. 3, p. 219, 2011.

SULISTYO, A.; PURWANTORO; SARI, K. P. Correlation, path analysis and heritability estimation for agronomic traits contribute to yield on soybean. **IOP Conference Series: Earth and Environmental Science**, v. 102, p. 012034, Jan. 2018.

SUZA, Walter; LAMKEY, Kendall. **Quantitative Genetics for Plant Breeding**. [S.l.]: Iowa State University Digital Press, 2023.

TAVARES DE ALBUQUERQUE, José Ricardo *et al.* Influence of genotype-environment interaction on soybean (*Glycine max* L.) genetic divergence under semiarid conditions. **Revista de la Facultad de Ciencias Agrarias UNCuyo**, v. 54, n. 1, p. 1–12, 4 Jul. 2022.

TEODORO, Paulo Eduardo *et al.* Predicting Days to Maturity, Plant Height, and Grain Yield in Soybean: A Machine and Deep Learning Approach Using Multispectral Data. **Remote Sensing**, v. 13, n. 22, p. 4632, 17 Nov. 2021.

TETILA, Everton Castelhão *et al.* Detection and classification of soybean pests using deep learning with UAV images. **Computers and Electronics in Agriculture**, v. 179, p. 105836, Dec. 2020.

TREVISAN, Rodrigo *et al.* High-Throughput Phenotyping of Soybean Maturity Using Time Series UAV Imagery and Convolutional Neural Networks. **Remote Sensing**, v. 12, n. 21, p. 3617, 4 Nov. 2020.

USDA, U. S. Departament of Agriculture Foreign Agriculture Service. **Production - Soybeans**. Available at: <<https://www.fas.usda.gov/data/production/commodity/2222000>>. Accessed on: 3 jun. 2025.

VERGARA, Rafael *et al.* Harvest delay, storage and physiological quality of soybean seeds. **Journal of Seed Science**, v. 41, n. 4, p. 506–513, Oct. 2019.

VOLPATO, Leonardo *et al.* Optimization of temporal UAS-based imagery analysis to estimate plant maturity date for soybean breeding. **The Plant Phenome Journal**, v. 4, n. 1, p. e20018, Jan. 2021.

WANG, Chuyuan; MYINT, Soe W. A Simplified Empirical Line Method of Radiometric Calibration for Small Unmanned Aircraft Systems-Based Remote Sensing. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 5, p. 1876–1885, May 2015.

WANG, Ruikun *et al.* Transformers for Remote Sensing: A Systematic Review and Analysis. **Sensors**, v. 24, n. 11, p. 3495, 29 May 2024.

WATANABE, Satoshi; HARADA, Kyuya; ABE, Jun. Genetic and molecular bases of photoperiod responses of flowering in soybean. **Breeding Science**, v. 61, n. 5, p. 531–543, 2012.

WATSON, Amy *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. **Nature Plants**, v. 4, n. 1, p. 23–29, 1 Jan. 2018.

XU, Rui; LI, Changying. A Review of High-Throughput Field Phenotyping Systems: Focusing on Ground Robots. **Plant Phenomics**, v. 2022, p. 9760269, 2022.

YU, Neil *et al.* Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. **Remote Sensing of Environment**, v. 187, p. 91–101, Dec. 2016.

YUAN, Wenan *et al.* Early Prediction of Soybean Traits through Color and Texture Features of Canopy RGB Imagery. **Scientific Reports**, v. 9, n. 1, p. 14089, 1 Oct. 2019.

ZHANG, Shanxin *et al.* Monitoring of Soybean Maturity Using UAV Remote Sensing and Deep Learning. **Agriculture**, v. 13, n. 1, p. 110, 30 Dec. 2022.

ZHOU, Jing *et al.* Estimation of the Maturity Date of Soybean Breeding Lines Using UAV-Based Multispectral Imagery. **Remote Sensing**, v. 11, n. 18, p. 2075, 4 Sep. 2019.