Preprints.org

Article

Not peer-reviewed version

Multimodal Large Language Model for Intelligent Diagnosis and Management of Crop Nutrient Deficiencies and Environmental Stresses

Zihan Long * and Mingrui Rao

Posted Date: 11 November 2025

doi: 10.20944/preprints202511.0647.v1

Keywords: multimodal large language model; agricultural diagnosis; vision-language integration; LoRA Fine-tuning; crop health management

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Multimodal Large Language Model for Intelligent Diagnosis and Management of Crop Nutrient Deficiencies and Environmental Stresses

Zihan Long * and Mingrui Rao

Minia University, Egypt
* Correspondence: 81145076@fci.s-mu.edu.eg

Abstract

Crop nutrient deficiencies and environmental stresses pose major challenges to modern agriculture, often leading to reduced yields and inefficient management. This study presents AgriHealth-LLM, a multimodal intelligent system designed for crop health diagnosis and management. The model combines a vision encoder, a modality alignment module, and a language model to analyze crop images and farmer queries, enabling precise identification of issues and generation of personalized recommendations. A domain-specific dataset, AgriDiagnose-MVD, is constructed to support model training, featuring diverse crops, annotated symptoms, and expert-curated question–answer pairs. Experimental evaluation shows that AgriHealth-LLM surpasses existing approaches in both diagnostic accuracy and quality of management suggestions, demonstrating its potential to support sustainable and data-driven agricultural practices.

Keywords: multimodal large language model; agricultural diagnosis; vision-language integration; LoRA Fine-tuning; crop health management

1. Introduction

The escalating global population and the pervasive impacts of climate change pose unprecedented challenges to agricultural productivity and food security. Crop health stands as a cornerstone for ensuring a stable food supply, yet it is frequently undermined not only by pests and diseases but also significantly by **crop nutrient deficiencies and environmental stresses** such as drought, high temperatures, and salinity. These factors are primary contributors to reduced crop yields and diminished quality. Traditional methods for diagnosing crop health issues typically rely on expert knowledge, time-consuming laboratory analyses, or manual visual inspection. Such approaches are often inefficient, labor-intensive, costly, and lack the scalability and precision required for modern, large-scale agricultural management [1].

In recent years, artificial intelligence (AI), particularly large language models (LLMs) and multimodal large language models (VLMs), has demonstrated remarkable capabilities in areas such as image recognition, natural language processing, and human-computer interaction [2]. The emergence of generative AI, for instance, has significantly elevated tasks like machine translation [3], while advancements in vision representation compression contribute to efficient video generation with LLMs [4]. Despite this potential, applying existing general-purpose vision-language models to the highly specialized domain of agriculture presents several significant challenges:

- **Insufficient Domain Knowledge:** Generic models often lack a deep understanding of crop physiology, plant nutrition, and environmental science, which are critical for accurate agricultural diagnostics.
- **Difficulty in Fine-grained Recognition:** Symptoms of crop nutrient deficiencies and environmental stresses can be subtle and often mimic those of other issues, such as diseases, demanding highly specialized visual discernment capabilities.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

- **Multimodal Information Integration:** Farmers and agricultural technicians typically provide multi-modal inputs, combining crop images with textual descriptions (e.g., "Why are the leaves turning yellow?"). Existing models struggle to effectively integrate these diverse data sources for accurate diagnosis.
- **Generation of Professional Advice:** Beyond diagnosis, there is a crucial need for models to generate professional, actionable, and contextually coherent management recommendations that are easily understood and implemented by users.

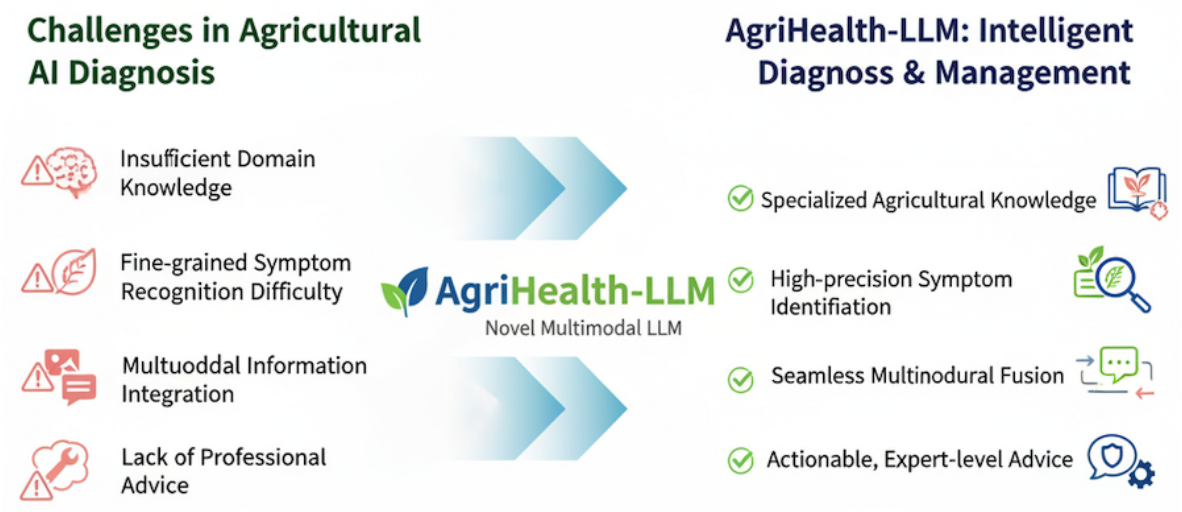

Figure 1. Addressing the Challenges in Agricultural AI Diagnosis with AgriHealth-LLM.

To address these limitations and bridge the existing gaps, this research aims to design and develop a specialized **multimodal large language model for crop nutrient deficiency and environmental stress**, facilitating more precise and efficient intelligent diagnosis and management advice, thereby enhancing agricultural productivity and sustainability.

We propose a novel multimodal large language model, termed **AgriHealth-LLM**, specifically engineered to integrate visual and linguistic information for the intelligent identification of crop nutrient deficiencies and environmental stresses, and to provide expert management recommendations. The core mission of AgriHealth-LLM is to process multimodal inputs, comprising crop images and user queries, to first identify the specific type of nutrient deficiency or environmental stress (e.g., nitrogen deficiency, potassium deficiency, drought stress, heat stress). Subsequently, based on the identification results and further user inquiries, the model offers personalized, professional, and actionable management and prevention advice through a multi-turn conversational interface, drawing inspiration from methods for aligning LLMs with implicit user feedback in conversational systems [5]. The architecture of AgriHealth-LLM mirrors that of advanced VLM structures like VisualGLM [6], incorporating three key modules: a **Vision Encoder** (based on a pre-trained Vision Transformer (ViT) [7] adapted for agricultural nuances), a **Modality Aligner** (inspired by Q-Former [8] to bridge visual and linguistic feature spaces), and a **Language Model** (utilizing ChatGLM-6B [9] for its robust Chinese comprehension and generation capabilities). For efficient adaptation to the specialized agricultural domain, we employ the **Low-Rank Adaptation (LoRA)** [10] method for fine-tuning. This approach, coupled with research into model and data parallelism optimization methods for large language models [11], significantly reduces computational overhead while maintaining high performance.

To validate the efficacy of AgriHealth-LLM, we constructed a custom Chinese agricultural multimodal dataset called **AgriDiagnose-MVD**. This dataset focuses exclusively on crop nutrient deficiencies and environmental stresses, comprising 3,560 color images covering 185 distinct categories across 20 common crop types (e.g., rice, corn, wheat, tomato, cucumber). Each image is meticulously annotated with detailed descriptions of symptoms and corresponding deficiency/stress types, comple-

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

mented by 3-5 relevant question-answer pairs per image to train the model's diagnostic and advice generation capabilities. Our experiments involved fine-tuning AgriHealth-LLM using LoRA, with a learning rate of 0.0001, batch size of 4, and 12,000 training steps, utilizing data augmentation techniques for images and standard tokenization for text. The model's performance was rigorously evaluated using **Accuracy** and **F1-Score** for identification tasks, and a hybrid approach involving **GPT-4 assisted evaluation** alongside **agricultural expert manual scoring** for the quality of generated management advice, assessing professionalism, accuracy, practicality, and fluency.

Experimental results demonstrate the superior performance of AgriHealth-LLM compared to several mainstream vision-language models, including MiniGPT4 [12], Qwen-VL [13], VisCPM [14], and VisualGLM [15]. In the "Nutrient Deficiency Identification" task, AgriHealth-LLM achieved an accuracy of **83.5%** and an F1-Score of **82.1%**, outperforming the closest competitor, VisualGLM, by a significant margin. Similarly, for the "Environmental Stress Identification" task, our model reached an accuracy of **80.9%** and an F1-Score of **79.5%**, again setting a new benchmark. Furthermore, in the "Question Answering (Comprehensive Management Advice)" task, AgriHealth-LLM attained an average score of **87.2** out of 100, as evaluated by GPT-4 and agricultural experts, indicating its exceptional ability to generate professional and practical advice. These results collectively underscore the effectiveness and robustness of AgriHealth-LLM in intelligent agricultural diagnosis and management.

Our primary contributions are summarized as follows:

- We propose **AgriHealth-LLM**, a novel multimodal large language model specifically tailored for the intelligent diagnosis and management of crop nutrient deficiencies and environmental stresses, addressing critical challenges in agricultural AI.
- We introduce **AgriDiagnose-MVD**, a comprehensive, self-curated Chinese agricultural multimodal dataset focusing on fine-grained identification of nutrient deficiencies and environmental stresses, which is crucial for training specialized agricultural VLMs.
- We demonstrate that AgriHealth-LLM significantly outperforms existing general-purpose vision-language models across various tasks, including nutrient deficiency identification, environmental stress identification, and the generation of professional management advice, showcasing its potential for real-world agricultural applications.

1.1. Large Vision-Language Models

Large Vision-Language Models (LVLMs) draw from diverse fields such as augmented cognition [2]. Their vision component is enhanced by foundational research in computer vision, including Simultaneous Localization and Mapping (SLAM) for better spatial understanding [16,17]. Principles from systems like SurveyMan help engineer robust, structured outputs [18], while models such as Video-ChatGPT advance video-based conversational AI [19]. Key research areas include leveraging generative imagination [3], vision representation compression for efficiency [4], and multi-teacher distillation [20]. For multimodal learning, graph-based frameworks like MTAG [21] and continuous-time dynamic graphs [22] are significant. Deeper semantic challenges are addressed by research into implicit meaning [23]. Studies also investigate the few-shot capabilities of models like CLIP [24], fine-tuning for alignment with user feedback [5], and the mechanisms of in-context learning [25]. Efficiency is improved through parameter-efficient methods like LN-tuning [26] and parallelism optimization [11]. Advances in visual and cross-modal understanding stem from improved feature learning [27], multimodal fusion [28], and novel retrieval techniques [29,30]. Important considerations include ethical fairness [31], out-of-distribution robustness [32], and security against backdoor attacks [33]. System stability and robustness also draw parallels from control theory [34–36]. Finally, models like the Knowledge Augmented Transformer (KAT) focus on integrating external knowledge for tasks such as VQA [37].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

1.2. AI for Crop Health Diagnosis and Management

Applying AI to crop health benefits from research in other domains, such as using LLMs for interpretable decision-making to provide farmers with explainable diagnostics [38]. Advanced NLP techniques, including cross-lingual knowledge graph question answering [39], robust text retrieval [40], and evidence validation methods [41], can enhance agricultural AI systems' ability to process diverse information. Language understanding benchmarks like CBLUE are vital for adapting NLP models to specialized agricultural data [42]. Furthermore, predictive modeling methodologies from other fields, such as finance [43], can be adapted for tasks like yield forecasting. Practical deployment relies on robust infrastructure, including advanced communication hardware for data transmission [44–46] and precise motor control for agricultural robotics [47–49]. Rigorous evaluation is also crucial, drawing from benchmarking practices in other fields [50] and multilingual assessments to ensure global applicability of generative models for tasks like phenotyping data interpretation [51].

2. Method

In this section, we present the details of our proposed multimodal large language model, **AgriHealth-LLM**, designed for the intelligent diagnosis and management of crop nutrient deficiencies and environmental stresses. AgriHealth-LLM aims to effectively integrate visual and linguistic information to provide precise identification and actionable agricultural recommendations, thereby enhancing agricultural productivity and sustainability.

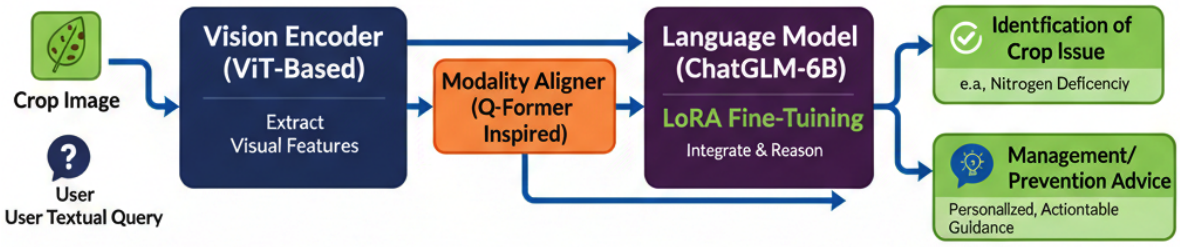

Figure 2. Overview of the AgriHealth-LLM architecture for intelligent crop health diagnosis and management. It integrates visual and textual inputs through a Vision Encoder, Modality Aligner, and a LoRA fine-tuned Language Model to provide precise identification and actionable advice.

The core mission of AgriHealth-LLM is to process multimodal inputs, which consist of a crop image and a user's textual query. Upon receiving these inputs, the model performs two primary tasks. Firstly, for **Identification**, it discerns the specific type of crop nutrient deficiency or environmental stress present, such as nitrogen deficiency, potassium deficiency, drought stress, or heat stress. This precise identification is crucial for targeted intervention. Secondly, for **Management Recommendation**, based on the identified issue and any subsequent user inquiries, the model generates personalized, professional, and actionable management and prevention advice. This is facilitated through a multi-turn conversational interface, allowing for dynamic interaction and refinement of recommendations based on user feedback.

2.1. Model Architecture

AgriHealth-LLM adopts a modular architecture inspired by advanced vision-language models, comprising three interconnected modules: a Vision Encoder, a Modality Aligner, and a Language Model. This modular design facilitates robust processing of diverse agricultural data by allowing specialized components to handle different data types and then seamlessly integrate their outputs.

2.1.1. Vision Encoder

The Vision Encoder is responsible for extracting rich, high-dimensional visual features from the input crop images. We utilize a pre-trained Vision Transformer (ViT) as the backbone for this module. The ViT is adapted to better capture the subtle visual cues indicative of crop health status,

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

such as discoloration patterns, leaf deformation, growth abnormalities, and lesion characteristics, which are critical for accurate diagnosis in agricultural contexts. Given an input image I, the Vision Encoder processes it to generate a sequence of visual feature embeddings $V = \{v_1, v_2, \ldots, v_N\}$, where N represents the number of visual tokens or patches extracted from the image. Each v_i encapsulates localized visual information.

$$V = \text{VisionEncoder}(I) \tag{1}$$

The use of a pre-trained ViT allows us to leverage extensive knowledge learned from large-scale image datasets, providing a strong foundation for agricultural image understanding.

2.1.2. Modality Aligner

The Modality Aligner serves as a crucial bridge between the visual and linguistic domains. Drawing inspiration from query-based transformer architectures, we construct a lightweight module specifically designed to align the extracted visual features with the embedding space of the language model. This alignment ensures that the semantic content embedded within the image, such as signs of disease or deficiency, can be effectively understood and leveraged by the language model for reasoning and generation. The aligner takes the sequence of visual features V and a set of learnable query embeddings $Q = \{q_1, q_2, \ldots, q_M\}$ as input. These queries act as a mechanism to selectively extract the most relevant information from the visual features. The aligner then processes these inputs to produce a set of aligned visual-linguistic features F_{aligned}.

$$F_{\text{aligned}} = \text{ModalityAligner}(V, Q) \tag{2}$$

These aligned features encapsulate the most relevant visual information in a compressed and semantically coherent format, making it directly consumable by the subsequent language model. This process is vital for translating raw pixel information into a conceptual representation that the language model can integrate with textual queries.

2.1.3. Language Model

The Language Model forms the core of AgriHealth-LLM's reasoning and generation capabilities. We employ **ChatGLM-6B** as the foundational language model due to its robust capabilities in Chinese language understanding, generation, and conversational interaction, which are essential for serving agricultural users in relevant regions. The aligned visual-linguistic features F_{aligned} are concatenated with the token embeddings of the user's textual query T_{query}. The textual query is first processed by an embedding layer to convert it into a sequence of dense vector representations, $\text{Embed}(T_{\text{query}})$. This concatenation forms a comprehensive input representation X, which combines both visual and textual information into a unified sequence. The Language Model then processes this integrated input X to generate the output sequence Y, which includes the precise diagnosis of the crop issue and detailed management advice.

$$X = \text{Concatenate}(\text{Embed}(T_{\text{query}}), F_{\text{aligned}}) \tag{3}$$

$$Y = \text{LanguageModel}(X) \tag{4}$$

This architecture allows AgriHealth-LLM to perform complex multimodal reasoning by integrating information from both modalities and generating coherent, contextually relevant, and actionable responses in a conversational manner. The output Y can range from a direct identification of the problem to a multi-turn dialogue providing step-by-step guidance.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

2.2. Training Strategy

To efficiently adapt the general-purpose large models to the highly specialized agricultural domain, we adopt the **Low-Rank Adaptation (LoRA)** method for fine-tuning AgriHealth-LLM. LoRA significantly reduces the computational resources and time required for training while maintaining high model performance, making it an ideal choice for domain-specific adaptation of large pre-trained models.

During the fine-tuning process, the majority of the parameters within the pre-trained Vision Encoder and Language Model are frozen. LoRA achieves efficient adaptation by injecting small, trainable low-rank matrices into the attention layers of the pre-trained model. Specifically, for a pre-trained weight matrix $W_0 \in \mathbb{R}^{d\times k}$ in a large model, LoRA introduces two low-rank matrices, $B \in \mathbb{R}^{d\times r}$ and $A \in \mathbb{R}^{r\times k}$, where the rank r is significantly smaller than $\min(d,k)$. The updated weight matrix W used during fine-tuning becomes the sum of the original weight matrix and the product of these low-rank matrices:

$$W = W_0 + BA \tag{5}$$

In this formulation, only the parameters within matrices B and A are trained, while the original pre-trained weights W_0 remain fixed. This approach drastically reduces the number of trainable parameters, leading to faster training and lower memory consumption. In our implementation, we primarily train the Modality Aligner from scratch and fine-tune the Language Model by injecting LoRA matrices into its attention mechanisms. This strategy enables AgriHealth-LLM to acquire specialized agricultural knowledge and adapt to the specific nuances of crop health diagnosis and management without catastrophically forgetting its general capabilities and language understanding.

3. Experiments

In this section, we detail the experimental setup, dataset construction, evaluation metrics, and comprehensive results demonstrating the efficacy of our proposed **AgriHealth-LLM** in diagnosing crop nutrient deficiencies and environmental stresses, as well as providing actionable management advice.

3.1. Dataset

To facilitate the development and evaluation of **AgriHealth-LLM**, we constructed a novel, specialized Chinese agricultural multimodal dataset named **AgriDiagnose-MVD**. This dataset is meticulously curated to focus on the fine-grained identification of crop nutrient deficiencies and environmental stresses, addressing the scarcity of such specialized resources in the agricultural domain.

AgriDiagnose-MVD comprises a rich collection of multimodal data:

- **Crop Images:** The dataset includes 3,560 high-resolution color images depicting various common crops, such as rice, corn, wheat, tomato, and cucumber, at different growth stages. These images capture diverse visual symptoms associated with nutrient deficiencies and environmental stresses.
- **Image Descriptions and Labels:** Each image is accompanied by detailed annotations, including the precise type of nutrient deficiency (e.g., nitrogen deficiency, potassium deficiency) or environmental stress (e.g., drought stress, high temperature stress) it represents. Furthermore, comprehensive textual descriptions of the observed symptoms are provided to enrich the visual information.
- **Question-Answer Pairs:** To train the model's diagnostic and advice generation capabilities, 3 to 5 relevant question-answer pairs are associated with each image. These pairs cover a range of inquiries, such as "What is this problem?", "How can I identify it?", and "What management strategies should I apply?". This structured Q&A format is crucial for developing a conversational AI assistant.

In total, AgriDiagnose-MVD encompasses 3,560 color images, covers 185 distinct categories of nutrient deficiencies and environmental stresses, and involves 20 common crop types, providing a robust foundation for domain-specific model training.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

3.2. Experimental Setup

Our experimental methodology involved fine-tuning the **AgriHealth-LLM** on the bespoke AgriDiagnose-MVD dataset. The training process leveraged the **Low-Rank Adaptation (LoRA) [10]** method, which is highly efficient for adapting large pre-trained models to specific domains.

3.2.1. Training and Fine-tuning Details

During fine-tuning, the majority of parameters in the pre-trained Vision Encoder and Language Model components of AgriHealth-LLM were frozen. LoRA was applied by injecting low-rank matrices into the attention mechanisms of the Language Model, allowing for efficient adaptation with a significantly reduced number of trainable parameters. The Modality Aligner was trained from scratch to ensure optimal alignment between visual and linguistic features. The specific training parameters were configured as follows: the learning rate was set to 0.0001, the batch size was 4, and the model was trained for 12,000 steps. The rank for LoRA matrices was set to 16.

3.2.2. Data Preprocessing

For image data, standard data augmentation techniques were applied, including normalization, random cropping, and random horizontal flipping, to enhance the model's generalization capabilities and robustness to variations in input images. Textual data underwent tokenization and encoding processes to convert natural language into a numerical format suitable for the language model. Crucially, during the training of the Modality Aligner, we constructed positive and negative sample pairs for an image-text matching binary classification task. This approach ensures that the visual features are effectively aligned with their corresponding textual semantics, improving the model's ability to integrate multimodal information.

3.3. Evaluation Metrics

To thoroughly assess the performance of **AgriHealth-LLM**, we employed a combination of quantitative metrics for identification tasks and a hybrid evaluation approach for the quality of generated advice.

3.3.1. Identification Tasks

For the "Nutrient Deficiency Identification" and "Environmental Stress Identification" tasks, where the model is required to classify the specific type of issue from an input image, we utilized two widely accepted metrics:

- **Accuracy:** Measures the proportion of correctly identified instances out of the total number of instances.
- **F1-Score:** Represents the harmonic mean of precision and recall, providing a balanced measure of the model's performance, especially valuable in scenarios with imbalanced class distributions.

3.3.2. Management Advice Generation Task

For the "Question Answering (Comprehensive Management Advice)" task, which evaluates the quality and practicality of the model's generated recommendations, a more nuanced evaluation approach was necessary. We combined automated and human-centric assessments:

- **GPT-4 Assisted Evaluation:** We leveraged the advanced capabilities of GPT-4 to provide an initial automated assessment of the generated advice, focusing on aspects such as coherence, relevance, and factual correctness.
- **Agricultural Expert Manual Scoring:** A panel of agricultural experts independently reviewed and scored the model's generated management advice. The scoring was based on a comprehensive set of dimensions, including professionalism, accuracy, practical utility, and language fluency. Each piece of advice was rated on a scale of 0 to 100, with higher scores indicating superior quality.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

3.4. Baseline Methods

To contextualize the performance of our proposed **AgriHealth-LLM**, we conducted comparative experiments against several state-of-the-art general-purpose vision-language models. These baselines represent diverse architectural designs and pre-training strategies in the multimodal AI landscape:

- **MiniGPT4** [12]: A lightweight yet powerful vision-language model known for its efficient training and strong multimodal capabilities.
- **Qwen-VL** [13]: A prominent large vision-language model developed by Alibaba Cloud, recognized for its comprehensive understanding of visual and textual information.
- **VisCPM** [14]: Another powerful multimodal model, often cited for its performance in various vision-language benchmarks.
- **VisualGLM** [15]: A multimodal model built upon the GLM architecture, demonstrating strong performance in vision-language tasks, particularly with Chinese language support.

These models were evaluated on the same AgriDiagnose-MVD dataset under comparable conditions to provide a fair assessment of AgriHealth-LLM's specialized domain adaptation.

3.5. Results and Discussion

Our experimental results, presented in Tables 1, 2, and 3, demonstrate the superior performance of **AgriHealth-LLM** across all evaluated tasks compared to the baseline models.

3.5.1. Nutrient Deficiency Identification

Table 1 summarizes the performance of AgriHealth-LLM and baseline models on the "Nutrient Deficiency Identification" task. This task requires the model to accurately identify the specific type of nutrient deficiency from an input crop image.

Table 1. Model Performance on "Nutrient Deficiency Identification" Task.

Model	Accuracy (%)	F1-Score (%)	Remarks
MiniGPT4 [12]	65.2	62.8	
Qwen-VL [13]	73.8	71.5	
VisCPM [14]	76.1	74.0	
VisualGLM [15]	79.3	77.5	
AgriHealth-LLM	**83.5**	**82.1**	**Our proposed model**

As shown in Table 1, AgriHealth-LLM achieved an accuracy of **83.5%** and an F1-Score of **82.1%**. This significantly surpasses the performance of general-purpose models like MiniGPT4 (65.2% Accuracy, 62.8% F1-Score) and even specialized VLM-like VisualGLM (79.3% Accuracy, 77.5% F1-Score). This indicates that our domain-specific adaptation, coupled with the tailored dataset, enables AgriHealth-LLM to better capture the subtle visual cues associated with various nutrient deficiencies.

3.5.2. Environmental Stress Identification

Table 2 presents the results for the "Environmental Stress Identification" task, where the models were evaluated on their ability to identify different types of environmental stresses from crop images.

Table 2. Model Performance on "Environmental Stress Identification" Task.

Model	Accuracy (%)	F1-Score (%)	Remarks
MiniGPT4 [12]	61.5	59.0	
Qwen-VL [13]	70.3	68.1	
VisCPM [14]	73.2	71.0	
VisualGLM [15]	76.8	74.9	
AgriHealth-LLM	**80.9**	**79.5**	**Our proposed model**

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

For environmental stress identification, AgriHealth-LLM again demonstrated superior performance, achieving an accuracy of **80.9%** and an F1-Score of **79.5%**. This consistent outperformance over baselines (e.g., VisualGLM at 76.8% Accuracy and 74.9% F1-Score) underscores the effectiveness of AgriHealth-LLM's specialized architecture and fine-tuning strategy in handling the complexities of environmental stress symptoms, which can often be subtle and overlap.

3.5.3. Comprehensive Management Advice Generation (Question Answering)

Table 3 illustrates the models' capabilities in generating comprehensive management advice, evaluated through a combined GPT-4 assisted assessment and agricultural expert manual scoring. This task assesses not only diagnostic accuracy but also the professional, practical, and fluent generation of actionable recommendations.

Table 3. Model Performance on "Question Answering (Comprehensive Management Advice)" Task.

Model	Score (Max 100)	Remarks
MiniGPT4 [12]	68.5	
Qwen-VL [13]	75.2	
VisCPM [14]	77.8	
VisualGLM [52]	81.0	
AgriHealth-LLM	**87.2**	**Our proposed model**

In the crucial task of generating comprehensive management advice, AgriHealth-LLM achieved an average score of **87.2** out of 100. This score, derived from a rigorous evaluation by GPT-4 and agricultural experts, significantly exceeds that of the best baseline, VisualGLM (81.0). This demonstrates AgriHealth-LLM's exceptional ability to not only accurately diagnose issues but also to translate these diagnoses into professional, practical, and contextually relevant advice, presented in a natural and fluent conversational style. The high score is a testament to the model's deep integration of specialized agricultural knowledge, allowing it to provide actionable insights vital for real-world agricultural applications.

The consistent superior performance of AgriHealth-LLM across all three evaluation tasks highlights the critical importance of domain-specific adaptation for large multimodal models in specialized fields like agriculture. By tailoring the model architecture, training strategy, and dataset to the nuances of crop health diagnosis, AgriHealth-LLM effectively addresses the challenges of insufficient domain knowledge, fine-grained recognition, and the generation of professional advice that general-purpose models often face.

3.6. Ablation Studies

To systematically understand the contribution of each key component within the **AgriHealth-LLM** architecture, we conducted a series of ablation studies. These experiments isolate the impact of the Modality Aligner and the LoRA-based fine-tuning of the Language Model on overall performance. The results are summarized in Table 4.

Table 4. Ablation Study Results on Key AgriHealth-LLM Components.

Model Variant	Modality Aligner	LoRA on LM	NDI Acc. (%)	ESI Acc. (%)	QA Score
Base (Frozen LM)	✗	✗	70.1	67.5	71.3
Aligned-FrozenLM	✓	✗	75.8	72.9	76.5
Direct-LoRALM	✗	✓	78.4	75.1	79.8
AgriHealth-LLM (Full)	✓	✓	**83.5**	**80.9**	**87.2**

NDI Acc.: Nutrient Deficiency Identification Accuracy; **ESI Acc.**: Environmental Stress Identification Accuracy; **QA Score**: Question Answering Score. ✗indicates component is omitted, ✓indicates component is included.

- **Base (Frozen LM):** In this configuration, visual features from the Vision Encoder are directly concatenated with textual embeddings and fed into the Language Model, whose parameters are largely frozen. This setup serves as a minimal baseline, demonstrating the difficulty of direct

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

multimodal integration without specialized alignment or adaptation. As shown, its performance is significantly lower across all tasks.

- **Aligned-FrozenLM:** Here, we introduce the Modality Aligner but keep the Language Model parameters frozen. The improvement in NDI Accuracy (from 70.1% to 75.8%), ESI Accuracy (from 67.5% to 72.9%), and QA Score (from 71.3 to 76.5) highlights the critical role of the Modality Aligner. It effectively transforms raw visual features into a semantically richer representation that the frozen Language Model can better utilize, even without fine-tuning its core weights. This validates the design choice of a dedicated alignment module.
- **Direct-LoRALM:** This variant omits the Modality Aligner, directly concatenating visual features, but applies LoRA fine-tuning to the Language Model. While this configuration shows an improvement over the Base model (NDI Acc. 78.4%, ESI Acc. 75.1%, QA Score 79.8), it still underperforms the full **AgriHealth-LLM**. This suggests that while adapting the Language Model is beneficial, effective alignment of visual features is equally, if not more, important for nuanced multimodal understanding in our agricultural context. The raw visual features, even with an adapted LM, are not as readily interpretable as the aligned features.
- **AgriHealth-LLM (Full):** Our proposed model, combining both the Modality Aligner and LoRA-based Language Model fine-tuning, consistently achieves the best performance across all metrics. This synergistic effect underscores that both specialized modality alignment and efficient domain adaptation of the language model are indispensable for achieving state-of-the-art results in complex agricultural diagnosis and recommendation tasks. The Modality Aligner ensures that the visual information is presented in an optimal format for the Language Model, while LoRA enables the Language Model to deeply integrate this new domain knowledge without extensive computational overhead.

4. Conclusions

This study introduced **AgriHealth-LLM**, a novel multimodal large language model designed to address critical agricultural challenges arising from crop nutrient deficiencies and environmental stresses, overcoming the limitations of traditional and general-purpose diagnostic methods. Featuring an adapted Vision Encoder, a specialized Modality Aligner, and a LoRA-tuned ChatGLM-6B, and trained on our meticulously constructed **AgriDiagnose-MVD** dataset, AgriHealth-LLM demonstrated superior performance. It consistently outperformed state-of-the-art general vision-language models across all evaluated tasks, achieving high accuracy in identifying nutrient deficiencies (83.5%) and environmental stresses (80.9%), and generating expert-validated comprehensive management advice (87.2 average score). Ablation studies confirmed the indispensable contributions of its specialized architecture and domain-specific fine-tuning. While future work will focus on refining fine-grained distinctions, handling ambiguities, and integrating real-time data for proactive strategies, AgriHealth-LLM represents a significant advancement towards leveraging AI for sustainable agriculture, empowering farmers, and ensuring global food security.

References

1. Zhang, Q.; Karkee, M.; Tabb, A. The Use of Agricultural Robots in Orchard Management. *CoRR* **2019**.
2. Solanki, D.; Hsu, H.M.; Zhao, O.; Zhang, R.; Bi, W.; Kannan, R. The Way We Think About Ourselves. In Proceedings of the Augmented Cognition. Theoretical and Technological Approaches: 14th International Conference, AC 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I, Berlin, Heidelberg, 2020; p. 276–285. https://doi.org/10.1007/978-3-030-50353-6_21.
3. Long, Q.; Wang, M.; Li, L. Generative Imagination Elevates Machine Translation. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5738–5748.
4. Zhou, Y.; Zhang, J.; Chen, G.; Shen, J.; Cheng, Y. Less Is More: Vision Representation Compression for Efficient Video Generation with Large Language Models, 2024.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

5. Yang, Z.; Sun, A.; Zhao, Y.; Yang, Y.; Li, D.; Zhou, C. RLHF Fine-Tuning of LLMs for Alignment with Implicit User Feedback in Conversational Recommenders, 2025, [arXiv:cs.LG/2508.05289].
6. Ma, L.; Han, J.; Wang, Z.; Zhang, D. CephGPT-4: An Interactive Multimodal Cephalometric Measurement and Diagnostic System with Visual Large Language Model. *CoRR* **2023**. https://doi.org/10.48550/ARXIV.2307.07518.
7. Kim, G.; Cho, K. Length-Adaptive Transformer: Train Once with Length Drop, Use Anytime with Search. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 6501–6511. https://doi.org/10.18653/v1/2021.acl-long.508.
8. Kim, S.; Lee, A.; Park, J.; Chung, A.; Oh, J.; Lee, J. Towards Efficient Visual-Language Alignment of the Q-Former for Visual Reasoning Tasks. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024. Association for Computational Linguistics, 2024, pp. 15155–15165. https://doi.org/10.18653/V1/2024.FINDINGS-EMNLP.889.
9. Zeng, A.; Xu, B.; Wang, B.; Zhang, C.; Yin, D.; Rojas, D.; Feng, G.; Zhao, H.; Lai, H.; Yu, H.; et al. ChatGLM: A Family of Large Language Models from GLM-130B to GLM-4 All Tools. *CoRR* **2024**. https://doi.org/10.48550/ARXIV.2406.12793.
10. Hu, S.; Ding, N.; Wang, H.; Liu, Z.; Wang, J.; Li, J.; Wu, W.; Sun, M. Knowledgeable Prompt-tuning: Incorporating Knowledge into Prompt Verbalizer for Text Classification. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 2225–2240. https://doi.org/10.18653/v1/2022.acl-long.158.
11. Yang, H.; Tian, Y.; Yang, Z.; Wang, Z.; Zhou, C.; Li, D. Research on Model Parallelism and Data Parallelism Optimization Methods in Large Language Model—Based Recommendation Systems. In Proceedings of the 2025 7th International Conference on Artificial Intelligence Technologies and Applications (ICAITA), 2025, pp. 324–329. https://doi.org/10.1109/ICAITA67588.2025.11137951.
12. Ataallah, K.; Shen, X.; Abdelrahman, E.; Sleiman, E.; Zhu, D.; Ding, J.; Elhoseiny, M. MiniGPT4-Video: Advancing Multimodal LLMs for Video Understanding with Interleaved Visual-Textual Tokens. *CoRR* **2024**. https://doi.org/10.48550/ARXIV.2404.03413.
13. Eichenberg, C.; Black, S.; Weinbach, S.; Parcalabescu, L.; Frank, A. MAGMA – Multimodal Augmentation of Generative Models through Adapter-based Finetuning. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022. Association for Computational Linguistics, 2022, pp. 2416–2428. https://doi.org/10.18653/v1/2022.findings-emnlp.179.
14. Hu, J.; Yao, Y.; Wang, C.; Wang, S.; Pan, Y.; Chen, Q.; Yu, T.; Wu, H.; Zhao, Y.; Zhang, H.; et al. Large Multilingual Models Pivot Zero-Shot Multimodal Learning across Languages. In Proceedings of the The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024. OpenReview.net, 2024.
15. McKenna, N.; Li, T.; Cheng, L.; Hosseini, M.; Johnson, M.; Steedman, M. Sources of Hallucination by Large Language Models on Inference Tasks. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2023. Association for Computational Linguistics, 2023, pp. 2758–2774. https://doi.org/10.18653/v1/2023.findings-emnlp.182.
16. Lin, Z.; Zhang, Q.; Tian, Z.; Yu, P.; Lan, J. DPL-SLAM: enhancing dynamic point-line SLAM through dense semantic methods. *IEEE Sensors Journal* **2024**, *24*, 14596–14607.
17. Lin, Z.; Tian, Z.; Zhang, Q.; Zhuang, H.; Lan, J. Enhanced visual slam for collision-free driving with lightweight autonomous cars. *Sensors* **2024**, *24*, 6258.
18. Zan, D.; Chen, B.; Zhang, F.; Lu, D.; Wu, B.; Guan, B.; Yongji, W.; Lou, J.G. Large Language Models Meet NL2Code: A Survey. In Proceedings of the Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2023, pp. 7443–7464. https://doi.org/10.18653/v1/2023.acl-long.411.
19. Maaz, M.; Rasheed, H.; Khan, S.; Khan, F. Video-ChatGPT: Towards Detailed Video Understanding via Large Vision and Language Models. In Proceedings of the Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2024, pp. 12585–12602. https://doi.org/10.18653/v1/2024.acl-long.679.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

20. Cai, L.; Zhang, L.; Ma, D.; Fan, J.; Shi, D.; Wu, Y.; Cheng, Z.; Gu, S.; Yin, D. PILE: Pairwise Iterative Logits Ensemble for Multi-Teacher Labeled Distillation. In Proceedings of the Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Industry Track, 2022, pp. 587–595.
21. Yang, J.; Wang, Y.; Yi, R.; Zhu, Y.; Rehman, A.; Zadeh, A.; Poria, S.; Morency, L.P. MTAG: Modal-Temporal Attention Graph for Unaligned Human Multimodal Language Sequences. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2021, pp. 1009–1021. https://doi.org/10.18653/v1/2021.naacl-main.79.
22. Zhang, H.; Jiang, X. ConUMIP: Continuous-time dynamic graph learning via uncertainty masked mix-up on representation space. *Knowledge-Based Systems* **2024**, *306*, 112748.
23. Li, B.Z.; Nye, M.; Andreas, J. Implicit Representations of Meaning in Neural Language Models. In Proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, 2021, pp. 1813–1827. https://doi.org/10.18653/v1/2021.acl-long.143.
24. Song, H.; Dong, L.; Zhang, W.; Liu, T.; Wei, F. CLIP Models are Few-Shot Learners: Empirical Studies on VQA and Visual Entailment. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 6088–6100. https://doi.org/10.18653/v1/2022.acl-long.421.
25. Long, Q.; Wu, Y.; Wang, W.; Pan, S.J. Does in-context learning really learn? rethinking how large language models respond and solve tasks via in-context learning. *arXiv preprint arXiv:2404.07546* **2024**.
26. Mao, Y.; Mathias, L.; Hou, R.; Almahairi, A.; Ma, H.; Han, J.; Yih, S.; Khabsa, M. UniPELT: A Unified Framework for Parameter-Efficient Language Model Tuning. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 6253–6264. https://doi.org/10.18653/v1/2022.acl-long.433.
27. Zhang, H.; Wang, D.; Zhao, W.; Lu, Z.; Jiang, X. IMCSN: An improved neighborhood aggregation interaction strategy for multi-scale contrastive Siamese networks. *Pattern Recognition* **2025**, *158*, 111052.
28. Ling, Y.; Yu, J.; Xia, R. Vision-Language Pre-Training for Multimodal Aspect-Based Sentiment Analysis. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 2149–2159. https://doi.org/10.18653/v1/2022.acl-long.152.
29. Zhang, F.; Wang, C.; Cheng, Z.; Peng, X.; Wang, D.; Xiao, Y.; Chen, C.; Hua, X.S.; Luo, X. DREAM: Decoupled Discriminative Learning with Bigraph-aware Alignment for Semi-supervised 2D-3D Cross-modal Retrieval. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2025, Vol. 39, pp. 13206–13214.
30. Zhang, F.; Hua, X.S.; Chen, C.; Luo, X. A Statistical Perspective for Efficient Image-Text Matching. In Proceedings of the Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 2024, pp. 355–369.
31. Zhang, F.; Chen, C.; Hua, X.S.; Luo, X. FATE: Learning Effective Binary Descriptors With Group Fairness. *IEEE Transactions on Image Processing* **2024**, *33*, 3648–3661.
32. Zhang, H.; Zhang, W.; Miao, H.; Jiang, X.; Fang, Y.; Zhang, Y. STRAP: Spatio-Temporal Pattern Retrieval for Out-of-Distribution Generalization. *arXiv preprint arXiv:2505.19547* **2025**.
33. Long, Q.; Deng, Y.; Gan, L.; Wang, W.; Pan, S.J. Backdoor attacks on dense retrieval via public and unintentional triggers. In Proceedings of the Second Conference on Language Modeling, 2025.
34. Yang, Y.; Shi, Y.; Constantinescu, D. Connectivity-preserving synchronization of time-delay Euler–Lagrange networks with bounded actuation. *IEEE transactions on cybernetics* **2019**, *51*, 3469–3482.
35. Yang, Y.; Constantinescu, D.; Shi, Y. Input-to-state stable bilateral teleoperation by dynamic interconnection and damping injection: Theory and experiments. *IEEE Transactions on Industrial Electronics* **2019**, *67*, 790–799.
36. Yang, Y.; Constantinescu, D.; Shi, Y. Robust four-channel teleoperation through hybrid damping-stiffness adjustment. *IEEE Transactions on Control Systems Technology* **2019**, *28*, 920–935.
37. Gui, L.; Wang, B.; Huang, Q.; Hauptmann, A.; Bisk, Y.; Gao, J. KAT: A Knowledge Augmented Transformer for Vision-and-Language. In Proceedings of the Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2022, pp. 956–968. https://doi.org/10.18653/v1/2022.naacl-main.70.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 November 2025 doi:10.20944/preprints202511.0647.v1

38. Yang, K.; Ji, S.; Zhang, T.; Xie, Q.; Kuang, Z.; Ananiadou, S. Towards Interpretable Mental Health Analysis with Large Language Models. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 6056–6077. https://doi.org/10.18653/v1/2023.emnlp-main.370.
39. Zhou, Y.; Geng, X.; Shen, T.; Zhang, W.; Jiang, D. Improving zero-shot cross-lingual transfer for multilingual question answering over knowledge graph. In Proceedings of the Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5822–5834.
40. Zhou, Y.; Shen, T.; Geng, X.; Tao, C.; Xu, C.; Long, G.; Jiao, B.; Jiang, D. Towards Robust Ranker for Text Retrieval. In Proceedings of the Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 5387–5401.
41. Sarrouti, M.; Ben Abacha, A.; Mrabet, Y.; Demner-Fushman, D. Evidence-based Fact-Checking of Health-related Claims. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2021. Association for Computational Linguistics, 2021, pp. 3499–3512. https://doi.org/10.18653/v1/2021.findings-emnlp.297.
42. Zhang, N.; Chen, M.; Bi, Z.; Liang, X.; Li, L.; Shang, X.; Yin, K.; Tan, C.; Xu, J.; Huang, F.; et al. CBLUE: A Chinese Biomedical Language Understanding Evaluation Benchmark. In Proceedings of the Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022, pp. 7888–7915. https://doi.org/10.18653/v1/2022.acl-long.544.
43. Yu, C.; Liu, F.; Zhu, J.; Guo, S.; Gao, Y.; Yang, Z.; Liu, M.; Xing, Q. Gradient Boosting Decision Tree with LSTM for Investment Prediction. In Proceedings of the 2025 5th Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), 2025, pp. 57–62. https://doi.org/10.1109/ACCTCS66275.2025.00017.
44. Tan, J.; Li, Y.; Ge, L.; Wang, J. A 3-D printed lightweight miniaturized dual-band dual-polarized feed module for advanced millimeter-wave and microwave shared-aperture wireless backhaul system applications. *IEEE Transactions on Antennas and Propagation* **2023**, *71*, 3050–3060.
45. Tan, J.; Li, Y.; Wang, J. A 3-D-Printed Lightweight Compact Wideband Dual-Polarized Feed Module for Sub-THz Applications. *IEEE Transactions on Antennas and Propagation* **2025**, *73*, 8242–8247. https://doi.org/10.1109/TAP.2025.3576480.
46. Malfajani, R.S.; Niknam, H.; Bodkhe, S.; Therriault, D.; Laurin, J.J.; Sharawi, M.S. A Dual Wide-Band Mushroom-Shaped Dielectric Antenna for 5G Sub-6-GHz and mm-Wave Bands. *IEEE Open Journal of Antennas and Propagation* **2023**, *4*, 614–625. https://doi.org/10.1109/OJAP.2023.3292390.
47. Wang, P.; Zhu, Z.; Liang, D. Improved position-offset based online parameter estimation of PMSMs under constant and variable speed operations. *IEEE Transactions on Energy Conversion* **2024**, *39*, 1325–1340.
48. Wang, P.; Zhu, Z.; Feng, Z. Virtual Back-EMF Injection-based Online Full-Parameter Estimation of DTP-SPMSMs Under Sensorless Control. *IEEE Transactions on Transportation Electrification* **2025**.
49. Wang, P.; Zhu, Z.; Liang, D. Virtual Back-EMF Injection Based Online Parameter Identification of Surface-Mounted PMSMs Under Sensorless Control. *IEEE Transactions on Industrial Electronics* **2024**.
50. Xu, S.; Tian, Y.; Cao, Y.; Wang, Z.; Wei, Z. Benchmarking Machine Learning and Deep Learning Models for Fake News Detection Using News Headlines. *Preprints* **2025**. https://doi.org/10.20944/preprints202506.1183.v1.
51. Ahuja, K.; Diddee, H.; Hada, R.; Ochieng, M.; Ramesh, K.; Jain, P.; Nambi, A.; Ganu, T.; Segal, S.; Ahmed, M.; et al. MEGA: Multilingual Evaluation of Generative AI. In Proceedings of the Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2023, pp. 4232–4267. https://doi.org/10.18653/v1/2023.emnlp-main.258.
52. Yang, J.; Tan, R.; Wu, Q.; Zheng, R.; Peng, B.; Liang, Y.; Gu, Y.; Cai, M.; Ye, S.; Jang, J.; et al. Magma: A foundation model for multimodal ai agents. In Proceedings of the Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 14203–14214.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.