Language
Understanding
$\lambda(\mu)$ Laboratory

Language Understanding Laboratory, Myanmar

Summer Internship 2025 Report

Beyond Flavor: A Multimodal Benchmark and Deep Evaluation of Myanmar Food, Recipes, and Culture

Submitted by

Image2Text Team

Under the supervision of

Professor Ye Kyaw Thu
LU Lab., Myanmar
NECTEC, NSTDA Pathum Thani, Thailand
yekyaw.thu@nectec.or.th

With mentorship from

Eaint Kay Khaing Kyaw
King Mongkut's Institute of Technology Ladkrabang

Team Members

Pyae Linn
University of Information Technology
Yangon, Myanmar

Lynn Myat Bhone
University of Computer Studies
Yangon, Myanmar

Shin Thant Phyo
University of Information Technology
Yangon, Myanmar

Thet Hmue Khin
University of Information Technology
Yangon, Myanmar

Khaing Hsu Yee
University of Technology(Yatanarpon Cyber City)
Pyin Oo Lwin, Myanmar

Hay Man Soe
University of Technology (Yadanarpon Cyber City)
Pyin Oo Lwin, Myanmar

Abstract

This report summarizes the activities and findings of a summer internship at the Language Understanding Laboratory, focusing on the creation of a visual question answering (VQA) benchmark for Myanmar traditional food. Over the internship, we collected diverse food images, designed multiple-choice question templates, and developed automated JSON generation workflows. Leading models such as Gemini, Qwen, Pixtral, and Gemma were evaluated on the dataset. The resulting benchmark supports multiple question categories—ingredient recognition, cultural context, and comparative reasoning—providing a foundation for future research in low-resource and culturally specific VQA tasks.

Food is a central part of cultural identity, and Myanmar cuisine reflects regional diversity, traditions, and festivals. However, there has been no multimodal benchmark for evaluating AI systems on this cultural knowledge. To address this gap, we introduce **MyanFoodQA**, the first multimodal evaluation benchmark on Myanmar food culture. It includes single-image and multi-image multiple-choice questions, as well as text-only reasoning tasks. Questions cover ingredients, preparation, symbolic meaning, and cultural events. Data was collected from **personal photos and web sources**, and annotated by **native Burmese speakers** to ensure cultural accuracy.

We evaluated the dataset in a zero-shot setting, finding that models perform well on **text-only questions** but struggle with **image-based and multi-image reasoning**. These results highlight current limitations in vision-language models for cultural reasoning. MyanFoodQA thus serves as a first benchmark for Myanmar food culture and a valuable step toward advancing multimodal AI for underrepresented languages.

Contents

Chapter 1

Objective

The primary objective of this research is to develop and evaluate a comprehensive multimodal benchmark, **MyanFoodQA**, designed specifically for Myanmar traditional food. This benchmark addresses the existing gap in datasets that integrate visual, textual, and cultural reasoning, particularly for low-resource languages like Myanmar. By curating a diverse set of food images and corresponding culturally relevant questions, this study aims to provide a robust dataset for assessing the capabilities of Vision-Language Models (VLMs) and Large Language Models (LLMs) in food-related tasks.

The research specifically seeks to:

1. **Evaluate Model Performance**: Assess the effectiveness of state-of-the-art VLMs and LLMs in answering food-related questions that require both cultural understanding and image-based reasoning.

2. **Identify Model Limitations**: Identify the strengths and weaknesses of current models, particularly in integrating visual inputs with culturally specific knowledge.

3. **Advance Multimodal AI Development**: Contribute to the development of multimodal AI systems that are capable of understanding and reasoning over both visual and cultural aspects, thereby fostering more inclusive AI technologies that represent diverse cultural contexts.

4. **Provide a Foundation for Future Research**: Offer a publicly available dataset that serves as a benchmark for future work in the field of multimodal AI for low-resource languages, with the long-term goal of improving AI's ability to understand culturally grounded knowledge.

Chapter 2

Introduction

Food is not just about eating—it is also about culture, traditions, and daily practices. Understanding food in an AI system is challenging because it requires more than just recognizing images; it also needs cultural knowledge. Our project is inspired by the *FoodieQA dataset* (Jacovi et al., 2023)[1], which studied Chinese food culture through multimodal questions and answers. Following their approach, we create a similar dataset for Myanmar traditional food, with culturally relevant questions and annotations. This adds to the growing resources that explore food and culture in AI.

Most existing food datasets focus on simple tasks such as recognizing dishes, detecting ingredients, or linking food to recipes. However, very few datasets cover fine-grained cultural reasoning, such as why a dish is eaten during certain festivals or how it is prepared in traditional ways. This problem is even more noticeable for **low-resource languages like Burmese**, where there are almost no datasets that connect language, images, and culture.

To fill this gap, we developed **MyanFoodQA**, a dataset designed to support multimodal question answering about Myanmar food. The dataset includes three types of tasks: (1) single-image reasoning, (2) multi-image reasoning, and (3) text-only reasoning. All the data were carefully collected and annotated by **native Burmese speakers**, ensuring that cultural details are correct. The process included gathering images, writing cultural notes, and checking that the annotations reflected Myanmar traditions accurately.

We found that all evaluated models can understand a little about culture and perform well on general knowledge, such as eating styles, types of food, and taste. However, they struggle with deeper cultural questions. Detailed results from further evaluation will be presented in the next chapters.

Chapter 3

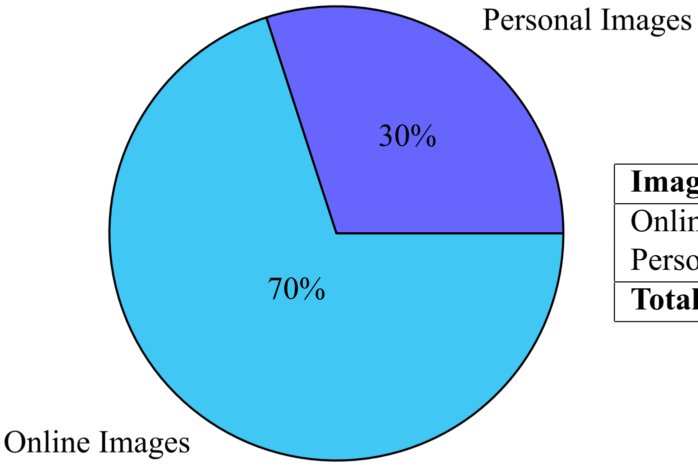
Work Done

3.1 Dataset Creation

Introduce **Beyond Flavor**, Food Dataset which construct a comprehensive data set on traditional Myanmar food, we first select **20** types of food that are widely popular throughout the country. These types of food were classified into five main groups: soups, snacks, beverages, meals, and salads. Once the categories were defined, we initiated the image collection process.

3.1.1 Image Collection

To keep our benchmark images new and avoid any overlap with the training data of existing vision-language models, we created a Google Form and distributed it to native Myanmar citizens, allowing them to upload photographs of traditional food they had personally taken. In addition to crowd-sourcing images from the public, our team also gathered images from the Internet and captured our own photos to enrich the data set. A total of six team members participated in this collection effort, ensuring that images were obtained for all twenty food categories. We got a total of 988 personal images that have neither been downloaded nor uploaded to the Web[9] [11] or social networks[10]. After manually filtering and cleaning, **713** images left and 285 images deleted. Also, from the Internet, we collected a total of **1,750** images. Our team collected **2463** images for our dataset.



Images	After filtering
Online	1750
Personal	713
Total	**2463**

Figure 3.1: Distribution of online and personal images after filtering.

3.1.2 Cultural Information Gathering

Beyond image collection, we made a comprehensive effort to gather in-depth cultural information for each food item included in the dataset. For each dish, we document its official and local names, place of origin, historical background, cultural significance, traditional preparation techniques, ingredients, and any associated customs or rituals[8]. This information provides a better understanding of each food item, highlighting not just what it looks like, but also the stories, traditions, and practices tied to it. Data collection was carried out by four additional team members. This detailed approach allows our data set to serve not only as a visual repository for research and machine learning applications but also as a valuable resource for cultural preservation, culinary education, and cross-cultural studies. By combining visual and cultural information, the data set captures the holistic essence of traditional Myanmar food and provides a foundation for future studies that require contextually rich data.

3.1.3 Making question templates

To systematically evaluate the models using our Myanmar traditional food dataset, we developed a complete set of template questions. These templates are designed to test different capabilities of the models, including understanding of images, distinguishing between correct and incorrect options, and reasoning based on textual descriptions. By structuring the evaluation in this way, we ensure that the models are tested not only on their ability to recognize food visually, but also on their understanding of preparation methods, ingredients, and cultural context.

For single-image questions, we initially created 32 templates that cover common question types for each food category. An example question is:

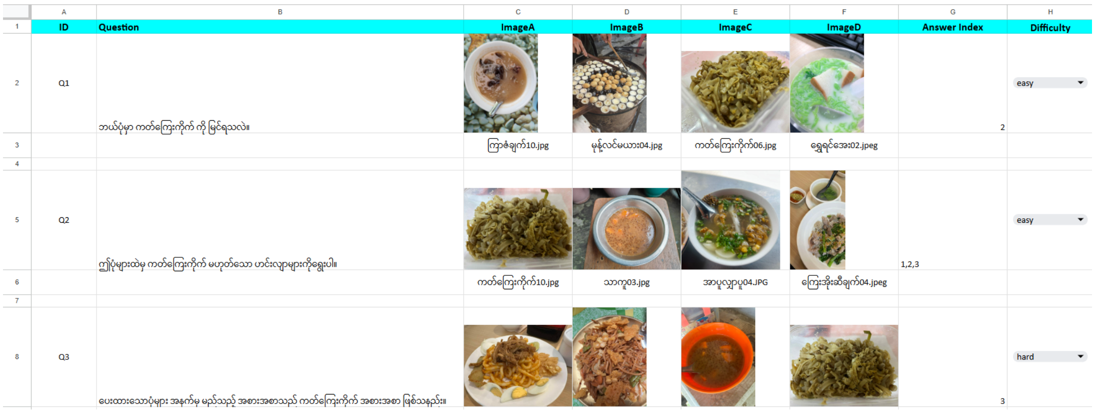
Myanmar: ``ပုံတွင်ပြထားသော အစားအစာကို ချက်ပြုတ်ရာတွင် (......) အစား ပြောင်းလဲ အသုံးပြုနိုင်သော ပစ္စည်း ကို ရွေးပါ။''

English: "Instead of....., choose the item that can be used to cook the food shown in the photo."

Two reviewers independently examined all initial templates and classified each as 'good', 'bad', or 'uncertain.' After removing the bad templates, **25** high-quality templates were retained. An enhancement process then produced additional questions that preserved the meaning of the original templates, resulting in a final set of 55 single image question templates.

For multi-image questions, we generated four answer choices for each question: one correct image and three randomly selected incorrect images. This setup is designed to test the performance of the model in distinguishing correct answers from visually similar alternatives.

For text-only questions, we created templates following the same approach as the single-image questions, but these questions are tailored for models that can only process textual input. In this case, the name of the food was explicitly included in the question to provide sufficient context for evaluation.



	A	B	C	D	E	F	G	H
1	ID	Question	ImageA	ImageB	ImageC	ImageD	Answer Index	Difficulty
2	Q1	ဘယ်ပုံမှာ ကတ်ကြေးကိုက် ကို မြင်ရသလဲ။					2	easy
3			ကြာဇံချက်10.jpg	မုန့်လင်မယား04.jpg	ကတ်ကြေးကိုက်06.jpg	ရွှေရင်အေး02.jpeg		
4								
5	Q2	ဤပုံများထဲမှ ကတ်ကြေးကိုက် မဟုတ်သော ဟင်းလျာများကိုရွေးပါ။					1,2,3	easy
6			ကတ်ကြေးကိုက်10.jpg	သာကူ03.jpg	အာပူလျှာပူ04.JPG	ကြေးအိုးဆီချက်04.jpeg		
7								
8	Q3	ပေးထားသောပုံများ အနက်မှ မည်သည့် အစားအစာသည် ကတ်ကြေးကိုက် အစားအစာ ဖြစ်သနည်း။					3	hard

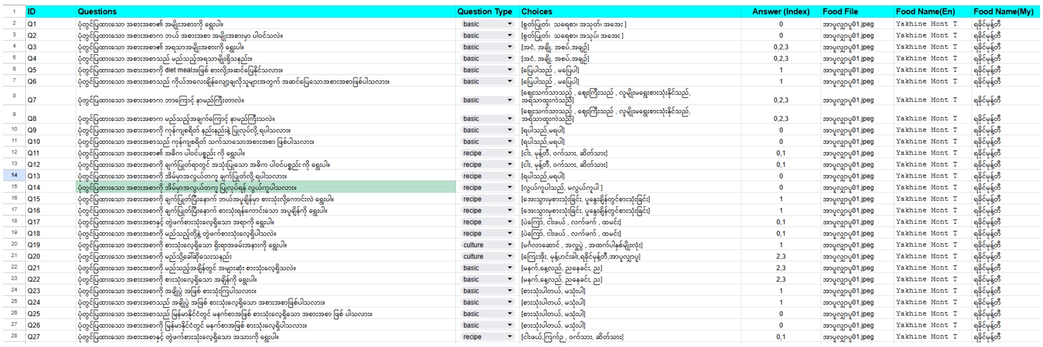
Figure 3.2: Beyond Flavor Sample Multi Images Questions (19 Food Types*15 Questions)

Question Type	Number of Questions
Single Image	1100
Text Only	1100
Multi-Image	285
Total	**2485**

Table 3.1: Summary of MyanFoodQA question types.

Food Type	Single Image Questions
ကိုက်ကြေးကိုက် (Kat Kyay Kite)	55
မုန့်လင်မယား (Mont Lin Ma Yarr)	55
နန်းကြီးသုပ် (Nan G Thote)	55
အုန်းနို့ခေါက်ဆွဲ (Coconut Noodle)	55
ခေါက်ဆွဲသုပ် (Noodle Salad)	55
ကောက်ညှင်းပေါင်း (Kout Nyin Paung)	55
ထမနဲ (Hta Ma Nae)	55
ရွှေရင်အေး (Shwe Yin Aye)	55
ဝက်သားတုတ်ထိုး (Wat Thar Tote Htoe)	55
မုန့်ဖက်ထုပ် (Mont Phat Htote)	55
မုန့်လက်ဆောင်း (Mont Let Saung)	55
သာကူ (Thar Ku)	55
ရခိုင်မုန့်တီ (R Pu Shar Pu)	55
ထမင်းပေါင်း (Htamin Paung)	55
ကြေးအိုးဆီချက် (Kyay Ohh See Chat)	55
တို့ဟူးနွေး (Tofu Nwe)	55
ကြာဇံချက် (Kyar San Chat)	55
မုန့်ဟင်းခါး (Mohinga)	55
လက်ဖက်သုပ် (Lat Phat Thote)	55
ရှမ်းခေါက်ဆွဲ (Shan Noodle)	55
Total	**1100**

Table 3.2: Number of questions per food type in MyanFoodQA. Each type has 55 questions, for a total of 1100 questions.



1	ID	Questions	Question Type	Choices	Answer (Index)	Food File	Food Name(En)	Food Name(My)
2	Q1	ပုံတွင်ပြထားသော အစားအစာ၏ အမျိုးအစားကို ရွေးပါ။	basic	[စွတ်ပြုတ်၊ သရေစာ၊ အသုတ်၊ အအေး]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
3	Q2	ပုံတွင်ပြထားသော အစားအစာက ဘယ် အစားအစာ အမျိုးအစားမှာ ပါဝင်သလဲ။	basic	[စွတ်ပြုတ်၊ သရေစာ၊ အသုပ်၊ အအေး]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
4	Q3	ပုံတွင်ပြထားသော အစားအစာ၏ အရသာအမျိုးအစားကို ရွေးပါ။	basic	[အငံ, အချို, အစပ်,အချဉ်]	0,2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
5	Q4	ပုံတွင်ပြထားသော အစားအစာသည် မည်သည့်အရသာမျိုးရှိသနည်း။	basic	[အငံ, အချို, အစပ်,အချဉ်]	0,2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
6	Q5	ပုံတွင်ပြထားသော အစားအစာကို diet mealအဖြစ် စားလို့အဆင်ပြေနိုင်သလား။	basic	[ပြေပါသည် , မပြေပါ]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
7	Q6	ပုံတွင်ပြထားသော အစားအစာသည် ကိုယ်အလေးချိန်လျှော့ချလိုသူများအတွက် အဆင်ပြေသောအစားအစာဖြစ်ပါသလား။	basic	[ပြေပါသည် , မပြေပါ]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
8	Q7	ပုံတွင်ပြထားသော အစားအစာက ဘာကြောင့် နာမည်ကြီးတာလဲ။	basic	[ဈေးသက်သာသည့် , ဈေးကြီးသည် , လူမျိုးမရွေးစားသုံးနိုင်သည်, အရသာထူးကဲသည်]	0,2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
9	Q8	ပုံတွင်ပြထားသော အစားအစာက မည်သည့်အချက်ကြောင့် နာမည်ကြီးသလဲ။	basic	[ဈေးသက်သာသည့် , ဈေးကြီးသည် , လူမျိုးမရွေးစားသုံးနိုင်သည်, အရသာထူးကဲသည်]	0,2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
10	Q9	ပုံတွင်ပြထားသော အစားအစာကို ကုန်ကျစရိတ် နည်းနည်းနဲ့ ပြုလုပ်လို့ ရပါသလား။	basic	[ရပါသည်,မရပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
11	Q10	ပုံတွင်ပြထားသော အစားအစာသည် ကုန်ကျစရိတ် သက်သာသောအစားအစာ ဖြစ်ပါသလား။	basic	[ရပါသည်,မရပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
12	Q11	ပုံတွင်ပြထားသော အစားအစာ၏ အဓိက ပါဝင်ပစ္စည်း ကို ရွေးပါ။	recipe	[ငါး, မုန့်တီ, ဝက်သား, ဆိတ်သား]	0,1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
13	Q12	ပုံတွင်ပြထားသော အစားအစာကို ချက်ပြုတ်ရာတွင် အသုံးပြုသော အဓိက ပါဝင်ပစ္စည်း ကို ရွေးပါ။	recipe	[ငါး, မုန့်တီ, ဝက်သား, ဆိတ်သား]	0,1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
14	Q13	ပုံတွင်ပြထားသော အစားအစာကို အိမ်မှာအလွယ်တကူ ချက်ပြုတ်လို့ ရပါသလား။	recipe	[ရပါသည်,မရပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
15	Q14	ပုံတွင်ပြထားသော အစားအစာကို အိမ်မှာအလွယ်တကူ ပြုလုပ်ရန် လွယ်ကူပါသလား။	recipe	[လွယ်ကူပါသည်, မလွယ်ကူပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
16	Q15	ပုံတွင်ပြထားသော အစားအစာကို ချက်ပြုတ်ပြီးနောက် ဘယ်အပူချိန်မှာ စားသုံးလို့ကောင်းလဲ ရွေးပါ။	recipe	[အေးသွားမှစားသုံးခြင်း, ပူနွေးချိန်တွင်စားသုံးခြင်း]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
17	Q16	ပုံတွင်ပြထားသော အစားအစာကို ချက်ပြုတ်ပြီးနောက် စားသုံးရန်ကောင်းသော အပူချိန်ကို ရွေးပါ။	recipe	[အေးသွားမှစားသုံးခြင်း, ပူနွေးချိန်တွင်စားသုံးခြင်း]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
18	Q17	ပုံတွင်ပြထားသော အစားအစာနှင့် တွဲဖက်စားသုံးလေ့ရှိသော အရာကို ရွေးပါ။	recipe	[ပဲကြော်, ငါးဖယ် , လက်ဖက် , ထမင်း]	0,1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
19	Q18	ပုံတွင်ပြထားသော အစားအစာကို မည်သည့်တို့နဲ့ တွဲဖက်စားသုံးလေ့ရှိပါသလဲ။	recipe	[ပဲကြော်, ငါးဖယ် , လက်ဖက် , ထမင်း]	0,1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
20	Q19	ပုံတွင်ပြထားသော အစားအစာကို စားသုံးလေ့ရှိသော ရိုးရာအခမ်းအနားကို ရွေးပါ။	culture	[မင်္ဂလာဆောင် , အလှူပွဲ , အထက်ပါနှစ်မျိုးလုံး]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
21	Q20	ပုံတွင်ပြထားသော အစားအစာကို မည်သို့ခေါ်ဆိုသေးသနည်း	culture	[ကြေးအိုး, မုန့်ဟင်းခါး,ရခိုင်မုန့်တီ,အာပူလျှာပူ]	2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
22	Q21	ပုံတွင်ပြထားသော အစားအစာကို မည်သည့်အချိန်တွင် အများဆုံး စားသုံးလေ့ရှိသလဲ။	basic	[မနက်,နေ့လည်, ညနေခင်း, ည]	2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
23	Q22	ပုံတွင်ပြထားသော အစားအစာကို စားသုံးလေ့ရှိသော အချိန်ကို ရွေးပါ။	basic	[မနက်,နေ့လည်, ညနေခင်း, ည]	2,3	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
24	Q23	ပုံတွင်ပြထားသော အစားအစာကို အချိုပွဲ အဖြစ် စားသုံးကြပါသလား။	basic	[စားသုံးပါတယ်, မသုံးပါ]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
25	Q24	ပုံတွင်ပြထားသော အစားအစာသည် အချိုပွဲ အဖြစ် စားသုံးလေ့ရှိသော အစားအစာဖြစ်ပါသလား။	basic	[စားသုံးပါတယ်, မသုံးပါ]	1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
26	Q25	ပုံတွင်ပြထားသော အစားအစာသည် မြန်မာနိုင်ငံတွင် မနက်စာအဖြစ် စားသုံးလေ့ရှိသော အစားအစာ ဖြစ် ပါသလား။	basic	[စားသုံးပါတယ်, မသုံးပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
27	Q26	ပုံတွင်ပြထားသော အစားအစာကို မြန်မာနိုင်ငံတွင် မနက်စာအဖြစ် စားသုံးလေ့ရှိပါသလား။	basic	[စားသုံးပါတယ်, မသုံးပါ]	0	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ
28	Q27	ပုံတွင်ပြထားသော အစားအစာနှင့် တွဲဖက်စားသုံးလေ့ရှိသော အသားကို ရွေးပါ။	recipe	[ငါးဖယ်,ကြက်ဥ , ဝက်သား, ဆိတ်သား]	0,1	အာပူလျှာပူ01.jpeg	Yakhine Mont T	ရခိုင်မုန့်တီ

Figure 3.3: Beyond Flavor Sample Single Image Questions (20 Food Types*55 Questions)

Food Type	Multi Iamge Questions
ကိုက်ကြေးကိုက် (Kat Kyay Kite)	15
မုန့်လင်မယား (Mont Lin Ma Yarr)	15
နန်းကြီးသုပ် (Nan G Thote)	15
အုန်းနို့ခေါက်ဆွဲ (Coconut Noodle)	15
ခေါက်ဆွဲသုပ် (Noodle Salad)	15
ကောက်ညှင်းပေါင်း (Kout Nyin Paung)	15
ထမနဲ (Hta Ma Nae)	15
ရွှေရင်အေး (Shwe Yin Aye)	15
ဝက်သားတုတ်ထိုး (Wat Thar Tote Htoe)	15
မုန့်ဖက်ထုပ် (Mont Phat Htote)	15
မုန့်လက်ဆောင်း (Mont Let Saung)	15
သာကူ (Thar Ku)	15
ရခိုင်မုန့်တီ (R Pu Shar Pu)	15
ထမင်းပေါင်း (Htamin Paung)	15
ကြေးအိုးဆီချက် (Kyay Ohh See Chat)	15
တို့ဟူးနွေး (Tofu Nwe)	15
ကြာဇံချက် (Kyar San Chat)	15
မုန့်ဟင်းခါး (Mohinga)	15
လက်ဖက်သုပ် (Lat Phat Thote)	15
Total	**285**

Table 3.3: Number of questions per food type in MyanFoodQA. Each type has 15 questions, for a total of 285 questions.

3.1.4 Data Cleaning

Two reviewers independently reexamined all the questions and assigned labels of ’good’, ’bad’, or ’not sure’. Questions flagged as bad were removed, while those marked as unclear were revised and corrected. The final set of approved questions and their corresponding answers was stored in JSON format, which serves as the standardized input for our experiments.

3.2 Model Testing and Evaluation

3.2.1 Model Testing

To evaluate the effectiveness of our dataset, we conducted zero-shot testing with a variety of state-of-the-art Large Language Models (LLMs) and Vision-Language Models (VLMs). For image-based tasks, we selected models such as [4]Qwen2.5-VL-7B, [2]Gemma-3-12B, [3]Pixtral-12B-2409, and [5]Gemini 1.5 Flash, which are capable of reasoning over visual inputs. For text-only cultural questions, we tested advanced LLMs which can only accept text such as [6]llama3.2: latest(3b), and [7]mistral:latest(7b).We used GPT-4o, Claude, Gemini, Qwen, and DeepSeek in zero-shot mode.

We applied our evaluation on three types of tasks:

- **Single-Image VQA**: Each question provided one food image and four answer choices. The model was required to infer details such as cooking methods, ingredients, or cultural significance of the dish.
- **Multi-Image VQA**: Each question presented four candidate images, with the model identifying the correct dish based on cultural or recipe-based clues. This task required multi-hop reasoning (e.g., first identifying the dish, then linking it to cultural attributes).
- **Text-Only QA**: Each question involved cultural or recipe-related reasoning based only on text (food names and cultural notes). This setup tested the models’ knowledge without visual inputs.

3.2.2 Prompt Selection

To evaluate the performance of the model, we created the three prompts using ’eng_version’ and ’myanmar_version’. The prompts that we used are:

Prompt 0 (Multiple Images)

Please carefully examine ALL the provided photos and identify the different foods shown in each image. Consider all images together to answer the question below.Choose the best answer from options {choice_range}.
Question: {question}
Choices: {choice_text}
Answer:

Prompt 0 (Single Image)

Please carefully look at the provided photo and identify the food shown. Based on this, answer the question below. Note:There may be multiple correct answers. Select all that apply. Choose from options {choice_range}. Question: {question} Choices: {choice_text} Answer:

Prompt 1 (Multiple Images)

ပုံများအားလုံးကိုသေချာစွာကြည့်ပြီးမေးခွန်းကိုဖြေပါ။ ရွေးချယ်မှုများထဲမှ အကောင်းဆုံးအဖြေကိုရွေးပါ။ Question: {question} Choices: {choice_text} Answer:

Prompt 1 (Single Image)

ပုံကိုကြည့်ပြီးမေးခွန်းကိုဖြေပါ။ မှတ်ချက်–အဖြေများစွာမှန်ကန်နိုင်ပါတယ်။ သင့်လျော်သမျှရွေးပါ။ ရွေးချယ်မှုများထဲမှရွေးပါ။
Question: {question} Choices: {choice_text} Answer:

Prompt 2 (Multiple Images)

Step 1: Look at each image and identify what food is shown. Step 2: Consider the relationship between all foods shown in the images. Step 3: Based on your analysis, answer the question. Step 4: Choose the best answer from options {choice_range}.
Question: {question} Available Options: {choice_text} Your Answer:

Prompt 2 (Single Image)

Step 1: Carefully examine the image and identify the food. Step 2: Consider the food's characteristics, ingredients, or preparation method. Step 3: Based

> on your analysis, answer the question. Note: Multiple answers may be correct. Select all that apply. Step 4: Choose from options {choice_range}. Question: {question} Available Options: {choice_text} Your Answer:

Prompt Comparison Results

- **Prompt 0**: Produced the best accuracy and detailed explanations (correct reasoning and correct answers).
- **Prompt 1**: Moderate accuracy. Often gave correct answers but lacked explanations.
- **Prompt 2**: Poorer performance overall. Models frequently produced wrong or incomplete answers.

Among the three prompting strategies we evaluated, **Prompt 0 consistently achieved the best performance**. It not only produced the correct answers but also provided clear explanations that reflected reasoning about the food images and cultural knowledge. In contrast, Prompt 1 achieved moderate accuracy but often returned only the final choice without reasoning, while Prompt 2 gave weaker and less reliable outputs. Since our goal is to evaluate both answer accuracy and reasoning ability in Myanmar food culture, we selected **Prompt 0 as the standard prompt** for the remainder of our experiments.

3.2.3 Hint-Based Evaluation

To further test cultural reasoning, we introduced a **hint-based evaluation** strategy. In this setup, models were given additional information such as the food name (e.g., "Mohinga") alongside the image. This reduced the difficulty of recognition and allowed us to observe whether models could improve on cultural or recipe-based reasoning once identification was resolved.

Our findings show that hint-based testing significantly increased accuracy for vision-language models in both single-image and multi-image VQA. For example, when the food name was revealed, models such as Pixtral and Gemini improved by more than 20% on average, particularly in tasks involving lesser-known regional foods. This suggests that the primary bottleneck for current VLMs lies in food recognition, while their cultural and recipe reasoning ability is stronger once the dish is identified.

Chapter 4

Evaluation Results on Myanmar Food Understanding with LLMs

In this chapter, we present the evaluation results of Large Language Models (LLMs) and Vision-Language Models (VLMs) on the task of understanding Myanmar food from both images and text. The experiments are divided into five sections: (1) Single Image Results, (2) Multi-Image Results, (3) Comparison of Single and Multi-Image Settings, (4) Text-Only Models, and (5) Comparison between Text-Only and Vision-Language Models. We also highlight the effect of providing additional hints (e.g., food names) that significantly improve performance on certain dishes such as *Mont Phat Htote*.

4.1 Single Image Results

We evaluate the performance of VLMs when given a single food image as input. The results demonstrate that there are considerable differences among the models in their ability to recognize Myanmar food items. Gemma3:12b achieved an accuracy of 43.38%, while Pixtral-12b-2409 performed better at 53.37%. Qwen 2.5VL:7b lagged behind with 30.74%, indicating that its visual reasoning ability may be less suited for this task. Gemini 1.5 Flash outperformed all other models with 64.27%, showing strong capabilities in image-based recognition. These results suggest that single-image recognition is a challenging but feasible task, with certain models already achieving promising performance. However, the moderate accuracy levels also indicate that understanding Myanmar food requires more contextual cues than just single visual inputs, especially given the wide diversity of dishes and variations in presentation.

Model	Accuracy (%)
Gemma3:12b	43.38
Pixtral-12b-2409	53.37
Qwen2.5VL:7b	30.74
Gemini 1.5 Flash	64.27

Table 4.1: Single Image Evaluation Results

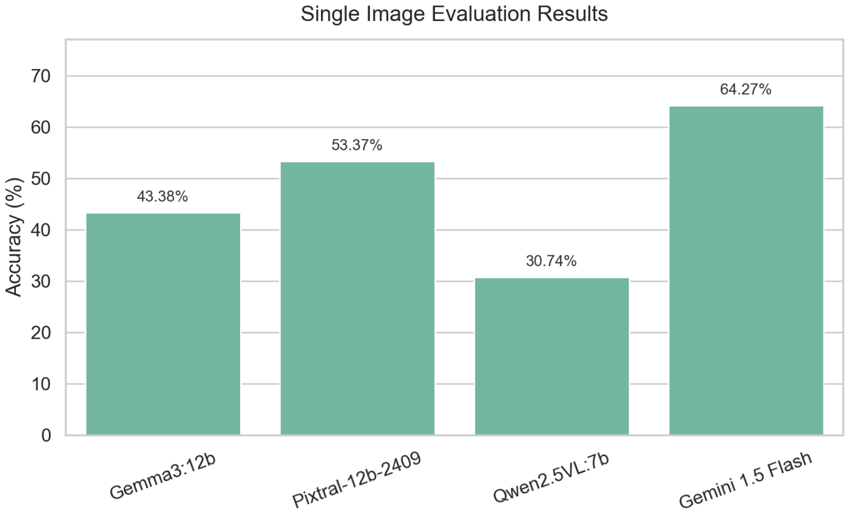


Figure 4.1: Impact of Single Image Inputs on Recognition Accuracy

4.2 Multi-Image Results

We then test models with multiple food images to assess their ability to generalize across different visual representations. This setup provides models with more visual context, allowing them to capture different perspectives of the same dish. Interestingly, the performance patterns shifted compared to the single-image setting. Gemma3:12b improved slightly to 44.89%, showing that multiple images helped it extract more consistent features. Pixtral-12b-2409, however, dropped to 37.58%, suggesting that the additional visual input introduced noise rather than clarity. Qwen2.5VL:7b benefited the most, jumping significantly to 65.30%, which highlights its strength in aggregating multiple visual cues. Gemini 1.5 Flash, on the other hand, dropped sharply to 36.76%, possibly due to overfitting or confusion when faced with multiple perspectives. These results illustrate that not all models handle multi-image inputs equally well, and designing architectures capable of effectively combining visual cues is a crucial direction for future research.

Model	**Accuracy (%)**
Gemma3:12b	44.89
Pixtral-12b-2409	37.58
Qwen2.5VL:7b	65.30
Gemini 1.5 Flash	36.76

Table 4.2: Multi-Image Evaluation Results

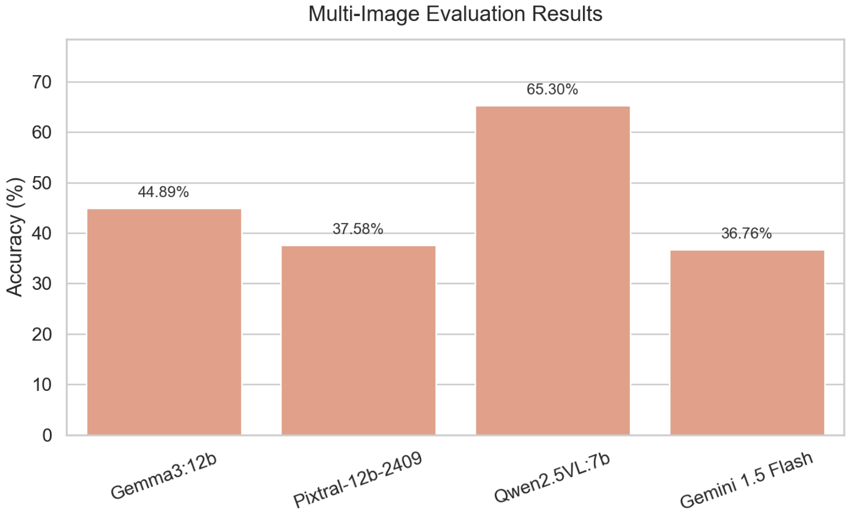


Figure 4.2: Impact of Multi-Image Inputs on Recognition Accuracy

4.3 Comparison of Single vs. Multi-Image

This section compares the relative performance difference between single-image and multi-image settings. The comparison highlights interesting trade-offs in how different models process food recognition tasks when provided with varying amounts of visual context. Gemma3:12b shows only a small improvement from 43.38% to 44.89%, meaning that additional images provide limited value for this model. Pixtral-12b-2409 performs worse when using multiple images (53.37% vs. 37.58%), suggesting difficulty in handling multi-view aggregation. By contrast, Qwen2.5VL:7b benefits greatly from multi-image inputs, jumping from 30.74% to 65.30%, almost doubling its accuracy. Gemini 1.5 Flash exhibits the opposite trend, dropping from 64.27% with single images to 36.76% with multi-images, raising questions about its robustness in complex visual contexts. These mixed results show that multi-image evaluation is not universally beneficial. Some models gain significant advantages, while others struggle with integrating multiple sources of visual information, underlining the importance of architecture design in VLM research.

Model	Single Image (%)	Multi Image (%)
Gemma3:12b	43.38	44.89
Pixtral-12b-2409	53.37	37.58
Qwen2.5VL:7b	30.74	65.30
Gemini 1.5 Flash	64.27	36.76

Table 4.3: Accuracy Differences Between Single and Multi-Image Evaluations

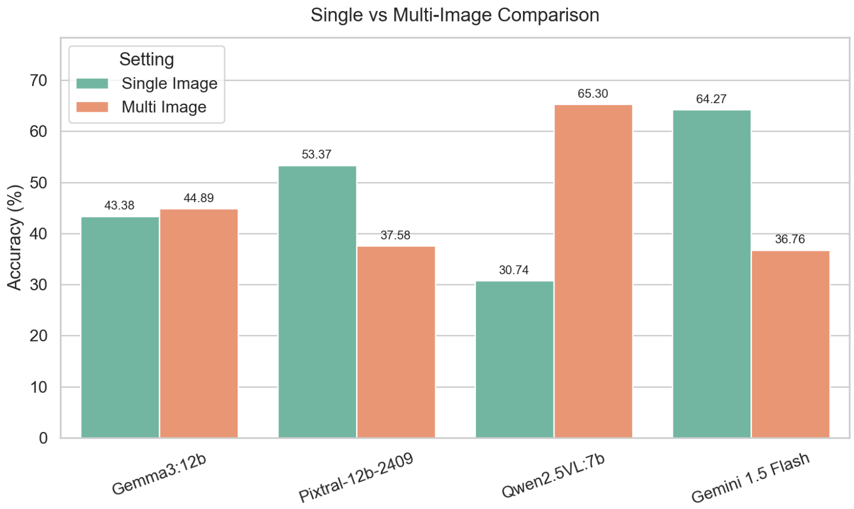


Figure 4.3: Impact of Single vs. Multi-Image Inputs on Model Accuracy

4.4 Text-Only Models

We also evaluate text-only LLMs on food recognition tasks, without image inputs. This evaluation setting is significantly more challenging because the models must rely solely on textual descriptions or food names without any visual context. LLaMA3.2:latest (3B) achieved 37.77%, while Mistral:latest (7B) performed better at 43.04%. Although these scores are lower than the top-performing VLMs, they show that text-only models can still capture some knowledge about Myanmar cuisine from their training data. The relatively close performance between these models suggests that scaling up model size brings moderate improvements, but cannot compensate for the lack of visual grounding. In food recognition tasks, cultural context and regional variations play a large role, and without images, these models may struggle to differentiate between dishes with similar names or overlapping ingredients. Nevertheless, text-only evaluations provide a useful baseline and highlight the value of multimodal integration for more accurate recognition.

Model	**Accuracy (%)**
LLaMA3.2:latest (3B)	37.77
Mistral:latest (7B)	43.04

Table 4.4: Text-Only Model Results

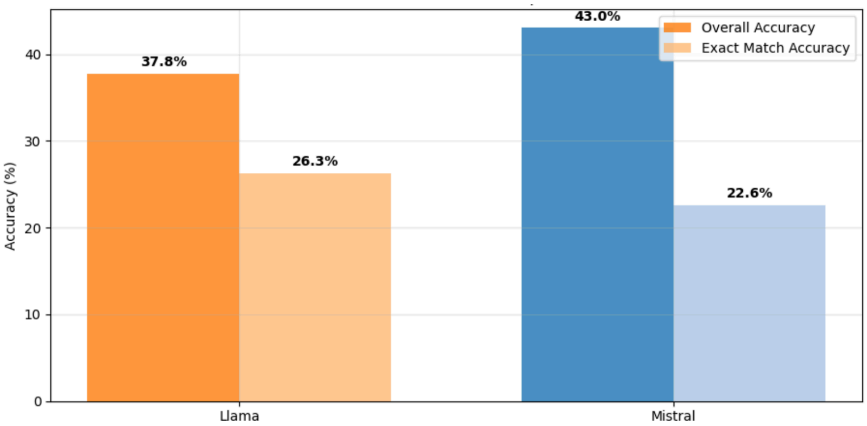


Figure 4.4: LLaMA vs Mistral Text-Only Model Comparison

4.5 Comparison of Text-Only and Vision-Language Models

Finally, we compare text-only language models with vision–language models (VLMs) using single-image accuracies to assess the value of visual grounding. On average, VLMs achieve higher performance than text-only models (47.9% vs. 40.4%), indicating that image features contribute meaningful information beyond linguistic priors. Nonetheless, performance varies substantially across VLMs: Gemini 1.5 Flash attains 64.27%, whereas Qwen2.5VL:7b reaches 30.74%, overlapping with the text-only range (e.g., Mistral:latest at 43.04%). This variation suggests that the benefits of multimodality depend on model design and training, not merely the presence of images. Overall, the aggregate advantage of VLMs supports the importance of visual context in food recognition, while the dispersion underscores the need for robust multimodal architectures to realize consistent gains.

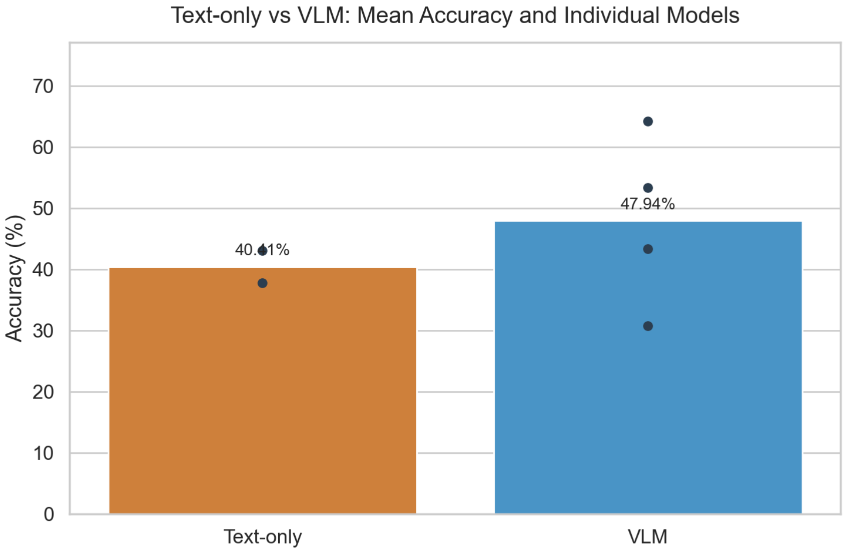


Figure 4.5: Text-Only vs Vision-Language Models Comparison

Chapter 5

Analysis and Discussion

This chapter presents a comprehensive analysis of the experimental results obtained from evaluating various vision-language models (VLMs) and text-only models on the task of Myanmar food recognition. The performance is assessed across single-image, multi-image, and text-only scenarios to benchmark the models' understanding of fine-grained, culturally-specific attributes. The discussion contextualizes these findings, highlighting strengths, weaknesses, and implications for future research.

5.1 Overall Model Performance Comparison

The evaluation reveals significant disparities in model capabilities, with performance heavily dependent on the task modality. A total of 1100 questions were evaluated to obtain these results.

5.1.1 Single Image Visual Question Answering

As shown in Table 5.1, the closed-weight API model, Gemini 1.5 Flash, significantly outperformed all open-weight models. Pixtral-12b-2409 emerged as the strongest open-weight model for this task, demonstrating competitive performance. Gemma3:12b performed respectably, while Qwen2.5VL:7b struggled considerably with single-image recognition.

Table 5.1: Single Image VQA Accuracy (%). Gemini 1.5 Flash demonstrated superior performance in identifying dishes from a single visual input.

Model	Accuracy
Gemini 1.5 Flash	**64.27**
Pixtral-12b-2409	53.37
Gemma3:12b	43.38
Qwen2.5VL:7b	30.74

Table 5.2: Multi-Image VQA Accuracy (%). Qwen2.5VL:7b showed remarkable strength in comparative analysis across multiple images.

Model	Accuracy
Qwen2.5VL:7b	**65.30**
Gemma3:12b	44.89
Pixtral-12b-2409	37.58
Gemini 1.5 Flash	36.76

5.1.2 Multi-Image Visual Question Answering

A strikingly different ranking emerged in the multi-image VQA task (Table 5.2). Qwen2.5VL:7b excelled, achieving the highest accuracy and demonstrating a particular aptitude for complex visual comparison. Conversely, the performance of both Gemini 1.5 Flash and Pixtral-12b-2409 dropped dramatically compared to their single-image results, indicating their strengths may lie more in direct recognition than in comparative reasoning. Gemma3:12b showed the most consistent performance across both vision tasks.

5.1.3 Text-Only Question Answering

Table 5.3: Text-Only Model Accuracy (%). Mistral (7b) outperformed the smaller Llama model on culinary knowledge probing.

Model	Accuracy
Mistral:latest (7b)	**43.04**
Llama3.2:latest (3b)	37.77

The text-only results (Table 5.3) show that larger model size (7b vs. 3b parameters) correlated with better performance on culinary knowledge tasks. The overall lower accuracy compared to the best VLM tasks underscores the critical value of visual information for this domain.

5.2 Comparative Analysis of Modalities and Techniques

5.2.1 Vision-Language Models vs. Text-Only Models

A core finding is the definitive advantage of VLMs over text-only LLMs for culinary understanding. The best VLM (Gemini 1.5 Flash, 64.27%) outperformed the best text-only model (Mistral, 43.04%) by over 21 percentage points on single-image tasks. This gap confirms that visual cues—such as color, texture, presentation, and ingredients—are indispensable for accurate food recognition and cultural reasoning.

5.2.2 Impact of Providing Hints

Table 5.4: Impact of Providing Dish Name Hints on Single-Image VQA Accuracy (%).

Model	Without Hint	With Hint	Improvement
Pixtral-12b-2409	53.37	62.00	+8.63
Qwen2.5VL:7b	30.74	37.87	+7.13
Gemma3:12b	43.38	46.78	+3.40
Gemini 1.5 Flash	64.27	66.08	+1.81

Providing the dish name as a textual hint alongside the image consistently improved model performance (Table 5.4). The improvement was most dramatic for Pixtral-12b-2409 and Qwen2.5VL:7b, suggesting that their primary bottleneck is initial dish identification rather than cultural knowledge retrieval. For the higher-performing Gemini 1.5 Flash, the hint provided less marginal gain, indicating it is already highly proficient at dish recognition. A notable example is the performance on *Mont Phat Htote*, where Gemma's accuracy improved by 17.31 points from 29.81% without giving hint to 47.12% with a hint.

5.3 Fine-Grained Analysis of Model Capabilities

5.3.1 Performance by Food Type: VLMs

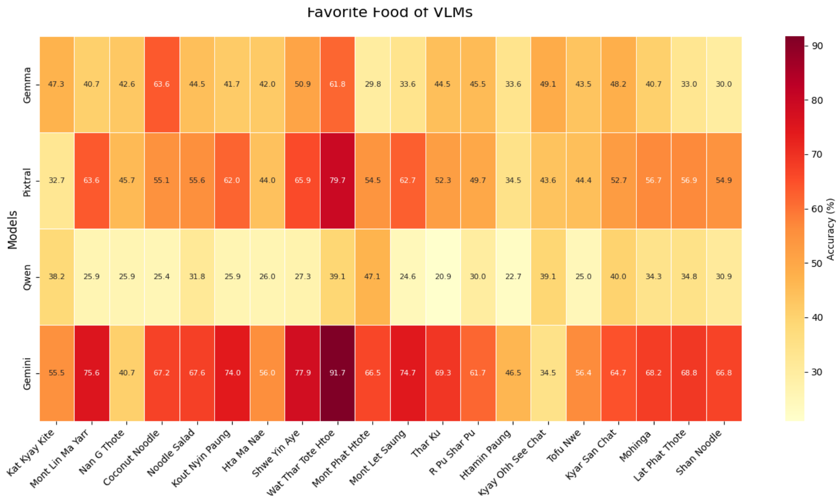


Figure 5.1: Heatmap of VLM performance accuracy across various Myanmar dishes. Warmer colors indicate higher accuracy. Gemini 1.5 Flash shows consistently high performance, while each model demonstrates specific strengths on particular dishes.

The heatmap visualization in Figure 5.1 reveals distinct patterns in VLM performance across different Myanmar dishes:

- **Gemini 1.5 Flash** demonstrates the most consistent high performance across nearly all dishes, particularly excelling at *Wat Thar Tote Htoe* (91.70%) and *Shwe Yin Aye* (77.91%).
- **Pixtral-12b-2409** shows strong performance on several traditional dishes including *Mont Lin Ma Yarr* (63.58%), *Kout Nyin Paung* (62.04%), and *Shwe Yin Aye* (65.91%).
- **Qwen2.5VL:7b** demonstrates particular strength on *Mont Phat Htote* (47.12%), where it achieves its highest performance.
- **Gemma3:12b** performs well on *Coconut Noodle* (63.64%) and shows moderate performance across most categories.

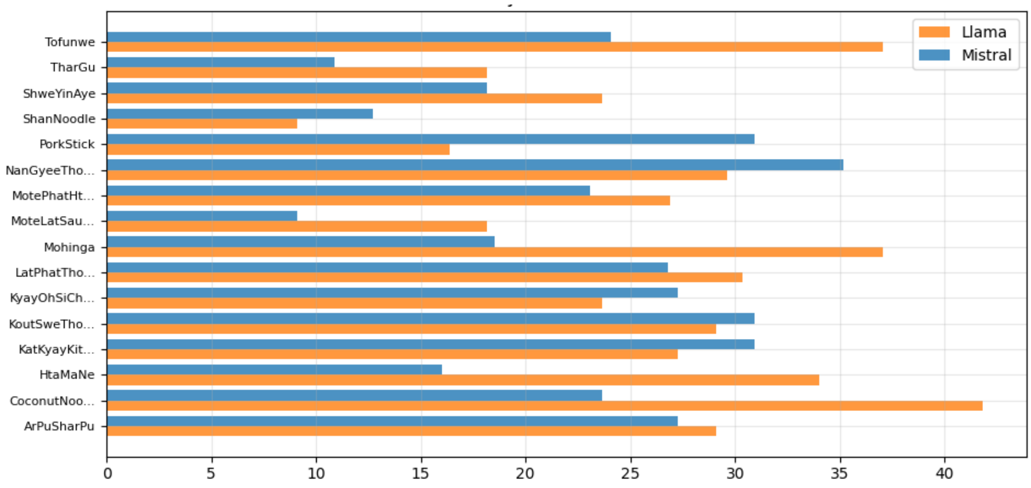


Figure 5.2: Performance of text-only LLMs across Myanmar dishes. Both Llama3.2 (3B) and Mistral (7B) remain significantly below VLM performance, with maximum accuracy below 42%. While Mistral shows advantages on some dishes (e.g., Mont Phat Htote, Mohinga), Llama performs better on others (e.g., TharGu, Coconut Noodle). Overall, both models reveal the limitations of text-only approaches for food recognition tasks.

5.3.2 Performance by Food Type:Text-Only Models

Figure 5.2 presents the performance of text-only models across Myanmar food categories, providing a crucial comparison to the VLM results:

- **Severe Performance Limitation:** Both text-only models show low accuracy (all below 42%), confirming that purely text-based approaches are not effective for food recognition tasks where visual cues are critical.
- **Mixed Outcomes Between Models:** Mistral (7B) achieves higher scores on complex dishes such as *Mont Phat Htote* and *Mohinga*, while Llama3.2 (3B) performs better on *TharGu*, *Coconut Noodle*, and *Hta Ma Ne*. This indicates complementary strengths rather than one model consistently outperforming the other.
- **Knowledge Bias:** The varying performance suggests that text-only models may rely on how frequently certain food names or descriptions appeared in their training data, rather than capturing deeper cultural or visual features.

5.3.3 Knowledge Domain Analysis: Radar Chart

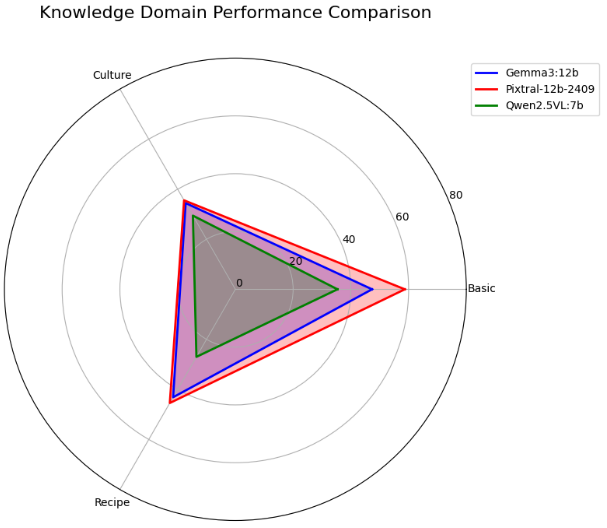


Figure 5.3: Radar chart comparing model performance across Basic, Cultural, and Recipe knowledge domains. The chart visualizes the specialized strengths of each model in different aspects of food understanding.

The radar chart (Figure 5.3) provides a visual comparison of model capabilities across different knowledge domains, complemented by the quantitative data in Table 5.5.

Table 5.5: Model Performance by Knowledge Domain (% Accuracy). Pixtral-12b-2409 led in all three domains.

Model	Basic	Culture	Recipe
Pixtral-12b-2409	**58.70**	**35.55**	**45.43**
Gemma3:12b	47.35	34.44	43.14
Qwen2.5VL:7b	35.37	29.44	27.00

As shown in both the radar chart and table, Pixtral-12b-2409 proved to be the most knowledgeable model, leading in all three domains: Basic (identification), Cultural (origins, traditions), and Recipe (ingredients, cooking method). A key finding is that **Cultural knowledge** is the most challenging domain for all models. This suggests that while models can learn to identify dishes and their components, grasping the deeper cultural context and significance associated with food remains a formidable challenge.

5.4 Discussion and Implications

The results of this analysis lead to several important conclusions:

- **Task-Dependent Performance:** The optimal model choice is highly task-dependent. Gemini 1.5 Flash is best for single-image recognition, Qwen2.5VL:7b excels at multi-image comparison, and Pixtral-12b-2409 possesses the deepest culinary knowledge.
- **The Multimodal Advantage is Clear:** VLMs significantly outperform text-only models, firmly establishing that visual understanding is critical for real-world culinary applications.
- **Hints Mitigate Recognition Bottlenecks:** Providing dish names significantly boosts performance for most models, identifying visual recognition as a key barrier for open-weight models.
- **Cultural Understanding is the Next Frontier:** Models are becoming proficient at *recognizing* food but still struggle with *understanding* its cultural context. This represents a significant area for future development.
- **Data Bias is Apparent:** Performance variability across dishes suggests that current models are biased towards foods well-represented in their training data. Curating diverse and culturally representative datasets is crucial.

This chapter has detailed the strengths and limitations of current models in understanding Myanmar's food culture. The findings underscore the complexity of the domain and provide a benchmark for future research aimed at developing more culturally-aware and robust multimodal AI systems.

Chapter 6

Future Work

While the dataset and evaluation strategy have made significant progress in showcasing the effectiveness of combining visual and cultural information for traditional Myanmar food, several avenues for future work remain to further improve the dataset, expand its use, and refine model performance.

Dataset Expansion: The current dataset includes 20 food types, but there is a goal to expand it to include more regional and lesser-known traditional dishes from Myanmar. This expansion will provide a more comprehensive representation of Myanmar's culinary heritage, enriching both the dataset and the model's understanding of diverse food types.

Incorporating Multilingual Support: Although a Myanmar-language version of the prompts has been created, expanding the dataset and model evaluation to include additional languages such as English and other ethnic languages spoken in Myanmar would increase accessibility and broaden the scope of model applications.

Refinement of Vision-Language Models: While large language models (LLMs) performed well on text-only questions, Vision-Language Models (VLMs) showed limitations in image-based reasoning. Future work could explore the integration of more specialized vision models trained specifically on Myanmar food imagery, which would improve recognition accuracy and reduce errors in classification tasks.

Cultural Contextualization: The cultural information gathered for each food item currently provides valuable context for understanding Myanmar cuisine. Further work could focus on creating a more structured format for this information, including user-generated content such as stories or historical contexts, to enhance the richness and depth of the dataset.

Model Fine-Tuning: Further fine-tuning of VLMs using the dataset could lead to improved reasoning abilities for both cultural and recipe-based questions. Leveraging additional training data from Myanmar's culinary history, customs, and preparation techniques could enhance the models' contextual knowledge.

Exploration of Interactive Applications: With the dataset and improved models, the creation of interactive applications could serve as educational tools for learning about Myanmar's traditional foods. This could take the form of a mobile app or website that helps users explore food history, recipes, and cultural significance while interacting with visual recognition tools.

Chapter 7

Conclusion

In this internship,we focused on developing a multimodal benchmark for Myanmar traditional food, addressing the lack of datasets that combine visual, textual, and cultural reasoning for low-resource languages. Through the creation of **MyanFoodQA**, we collected diverse food images, designed multiple-choice question formats, and ensured cultural accuracy with the help of native Burmese annotators. The dataset covers a wide range of categories, including ingredients, preparation methods, symbolic meanings, and festival-related practices, making it a rich resource for studying the intersection of food and culture.

Evaluation of several leading **vision-language models (VLMs)** and **large language models (LLMs)** showed that while text-based cultural reasoning is handled relatively well, performance drops significantly on image-based and multi-image tasks. This gap highlights the challenges of teaching AI systems to understand culturally grounded visual knowledge, and suggests that current models are still limited in their ability to integrate cultural context from multimodal data.

Overall, this work represents the first step toward building a reusable benchmark for Myanmar food culture and contributes to the broader effort of developing **multimodal AI for underrepresented languages**. Future directions include fine-tuning models on MyanFoodQA, expanding the dataset with more categories and languages, and releasing resources to the research community. By continuing this line of work, we aim to improve AI's ability to capture cultural diversity and support inclusive, globally relevant applications.

Acknowledgment

Sincere gratitude is extended to **Prof. Ye Kyaw Thu**, whose guidance and unwavering support were instrumental throughout the research process. Prof. Ye Kyaw Thu not only provided the opportunity to join the Language Understanding Lab but also offered invaluable insights and mentorship that greatly contributed to the success of this project. We deeply appreciate Prof. Ye Kyaw Thu's dedication to fostering an environment of academic excellence and research innovation.

Special thanks are due to the members of the Language Understanding Lab for providing essential resources and support during the project. Their collaborative efforts and technical assistance were crucial to the successful completion of the research.

Heartfelt thanks go to Ma Eaint Kay Khaing Kyaw for her invaluable mentorship and supervision throughout the research. Her critical guidance ensured the project stayed on track, and her expertise played a significant role in refining the research methodology and results.

Appreciation is also directed to **Ko Than Zaw Toe** for his invaluable contribution to the dataset collection, preprocessing, and the dedicated efforts put into ensuring the quality of the data used in this research.

Appreciation is also directed to **Zay Yar Htet** and **Myat Thin Thin Kyi** for his invaluable contribution to the dataset collection into ensuring the quality of the data used in this research.

The dedication and contributions of the team members are deeply appreciated. Their hard work and commitment were crucial to the successful completion of this project. Thanks are due to each of them for giving their best time and effort to data collection, image processing, and the overall progress of the research.

Lastly, sincere gratitude is extended to all individuals who participated in the data collection via Google Forms, submitting and providing images that played a critical role in the development of the dataset.

References

[1] Z. Liu, X. Wang, Y. Chen, et al., “Foodie QA: A Large-Scale Question Answering Dataset for Food Images,” arXiv:2406.11030 [cs.CL], 2024. Foodie QA paper airXiv

[2] Gemma Team, “Gemma 3,” Kaggle, 2025. Gemma 3 Report (Available: https://www.kaggle.com/datasets/Gemma3/Gemma3Report)

[3] Mistral AI, “Pixtral 12B,” arXiv preprint arXiv:2410.07073, 2024. Available: https://arxiv.org/abs/2410.07073

[4] Alibaba Group, “Qwen 2.5 VL Technical Report,” arXiv preprint arXiv:2502.13923 [cs.CV], 2025. Available: https://doi.org/10.48550/arXiv.2502.13923

[5] Gemini Team (Google), “Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context,” arXiv preprint arXiv:2403.05530 [cs.CL], 2024. Available: https://doi.org/10.48550/arXiv.2403.05530

[6] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample, “LLaMA: Open and Efficient Foundation Language Models,” arXiv preprint arXiv:2302.13971 [cs.CL], 2023. Available: https://doi.org/10.48550/arXiv.2302.13971

[7] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. de las Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier, L. R. Lavaud, M.-A. Lachaux, P. Stock, T. Le Scao, T. Lavril, T. Wang, T. Lacroix, W. El Sayed, “Mistral 7B,” arXiv preprint arXiv:2310.06825 [cs.CL], 2023. Available: https://doi.org/10.48550/arXiv.2310.06825

[8] Daw Yin Yin Aye, *Recipes and Home Businesses 1000*, Daw Khin Yee, January 1996.

[9] Wikipedia contributors, *Wikipedia*, Available at: `https://www.wikipedia.org/`, Accessed: September 6, 2025.

[10] Facebook, ”Page title or post description,” *Facebook*, Available at: `https://www.facebook.com/...`, Accessed: September 6, 2025.

[11] MyFood Myanmar, “Home | MyFood Myanmar,” MyFood Myanmar, Myanmar, Accessed: September 6, 2025. Available: `https://myfoodmyanmar.com/`