AI-Powered Dermatological Assistant: Bridging Healthcare Gaps Through Multimodal Intelligence

Aranya Saha Tanvir Ahmed Khan Ismam Nur Swapnil Mohammad Ariful Haque

Department of Electrical and Electronic Engineering
Bangladesh University of Engineering and Technology

Project Page

Global Healthcare Challenge & AI

The Critical Gap: Over 2.6 billion people worldwide lack access to specialized dermatological care, leading to preventable suffering and advanced disease progression in underserved communities.

Geographic Disparities:

- Rural and remote areas: Limited specialist availability
- Developing regions: High consultation costs and travel barriers
- Emergency settings: Need for rapid preliminary diagnosis
- Resource-constrained facilities: Lack of diagnostic infrastructure

Technology Limitations:

- Existing AI systems lack medical specialization
- General-purpose models (ChatGPT, etc.) insufficient for clinical accuracy
- Traditional telemedicine limited by human expert availability
- Current solutions not optimized for deployment in low-resource settings

Our Innovation: An intelligent dermatological assistant that democratizes expert-level skin diagnosis through advanced multimodal AI, making specialized care accessible globally through any smartphone or basic computing device.

Breakthrough AI Framework

We developed a sophisticated multimodal intelligence system combining AI techniques:

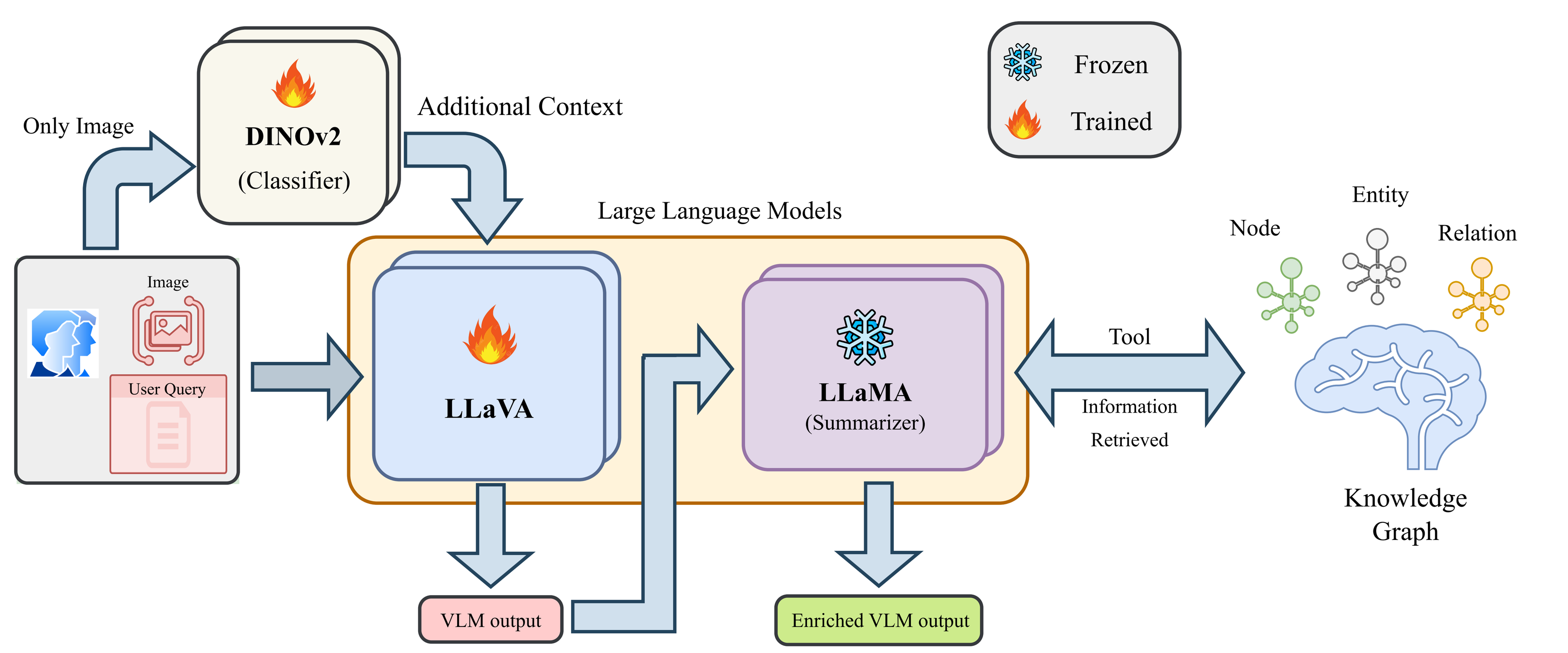

Figure 1. Our Overall Proposed Framework

- The user submits a query with an input image. DINOv2 processes the image to predict a disease label. If the softmax probability exceeds a set threshold, the label is converted into text and passed as context to the Vision-Language Model (VLM). In this case, DINOv2 acts as an Auxiliary Classifier to help the Vision Language Model.
- The VLM (fine-tuned and compressed LLaVA) uses the query, image, and context to generate an initial response.
- This response is sent to LLaMA, which accesses a medical knowledge graph to perform retrieval-augmented generation (RAG) and produce a final enriched answer.
- DINOv2 handles disease classification; LLaVA is used for visual question answering; LLaMA manages RAG and summarization.
- DINOv2 and LLaVA are fine-tuned, and LLaVA is compressed for efficient deployment, LLaMA is used as a pre-trained model

Overall Training Pipeline

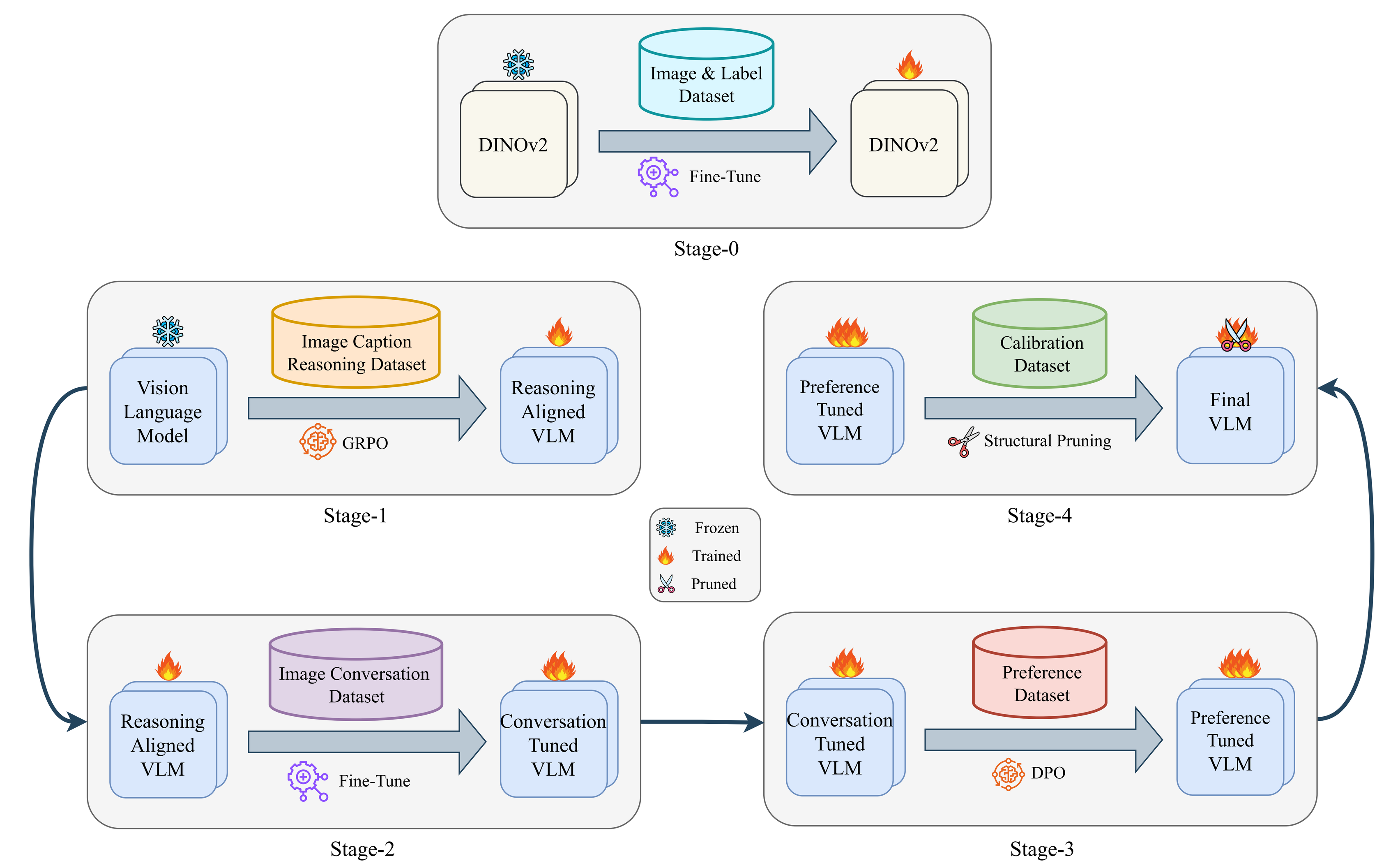

Figure 2. Full Training Pipeline of DINOv2 and LLaVA (VLM)

Our systematic approach consists of four sequential enhancement stages:

- **Stage-0 – Auxiliary Classifier Training:** DINOv2 fine-tuned as an auxiliary classifier for dermatology using an image-label dataset.
- **Stage-1 – Medical Reasoning Training:** A large language and vision assistant fine-tuned for medical reasoning using an image-caption dataset and GRPO optimization.
- **Stage-2 – Patient Communication Enhancement:** Fine-tuned on an image-conversation dataset for natural patient interactions.
- **Stage-3 – Interaction Optimization:** Direct Policy Optimization (DPO) using a preference dataset to improve response quality.
- **Stage-4 – Resource Optimization:** Structural pruning with a calibration dataset to enable deployment on resource-limited hardware.

Clinical-Grade Performance

Our progressive enhancement approach delivered remarkable diagnostic capabilities:

Intelligence Level	Diagnostic Accuracy	Patient Interaction Score
General AI Baseline	30.77%	3/10
+ Medical Reasoning	79.49%	4/10
+ Patient Communication	64.10%	6/10
+ Interaction Optimization	64.10%	8/10
+ Auxiliary Classifier	82.05%	8/10
+ Knowledge Graph	82.05%	9/10
+ Resource-Optimization	76.92%	8/10

Table 1. Progressive enhancement toward clinical excellence

- The model progressively improves quality, while remaining efficient for deployment.

Example Input-Output

Model Input Prompt

Question: Describe the image and find out the disease.

Model Output Examples

Dermatological Analysis:

1. **Color and Pigmentation:** The image shows reddened areas on the face, which could be indicative of Rosacea.
2. **Texture and Surface Characteristics:** The reddened areas appear smooth and without visible scales or ulcers, which is consistent with Rosacea.
3. **Shape and Border:** The reddened areas have irregular borders, which is characteristic of Rosacea.

<answer>

Rosacea

</answer>

Figure 3. An example input-output pair from our system for the disease Rosacea.

Key Achievement

Our final VLM achieved **76.92% diagnostic accuracy** (vs. 30.77% baseline) and a **conversation score of 8/10**, reaching **82.05% accuracy** and **9/10** before compression—marking a breakthrough in multimodal dermatological AI, with compression improving efficiency.

Local & Global Impact: Scaling Potential from Bangladesh to the World

Immediate Healthcare Applications (Global & Local):

- **Primary Care Support:** AI-assisted preliminary screening in underserved areas, including rural Bangladesh
- **Telemedicine Enhancement:** Remote consultations enhanced by AI, supporting telehealth growth in Bangladesh and beyond
- **Medical Education:** Interactive training tool for medical students and healthcare workers in local institutions
- **Emergency Response:** Fast diagnostic support during natural disasters and public health emergencies, especially in vulnerable regions

Scaling Vision:

- **Specialty Expansion:** Extend to fields like ophthalmology, cardiology, and radiology
- **Clinical Validation:** Collaborate with local hospitals and global institutions for real-world trials
- **Continuous Learning:** Employ federated learning from both global users and Bangladeshi deployment feedback
- **Healthcare Integration:** Seamless integration with electronic health records (EHRs) and optimization of clinical workflows

Transformative Potential: This intelligent assistant proves that AI can deliver clinically relevant, compassionate diagnostic support. It holds immense potential to democratize expert-level healthcare — from district hospitals in Bangladesh to remote clinics worldwide — ultimately saving lives through accessible, intelligent medical assistance.