

Appendix

In Section A, we first provide additional implementation details for 1) LLM fine-tuning, 2) visual translator, and 3) clinical report summarizer. Then in Section B, we give additional discussion about our approach and dataset. In Section C, we provide the template (to GPT-4) for LLM-based summarizer.

A Implementation Details

LLM Fine-Tuning. We leverage the TRL-Transformer Reinforcement Learning GitHub repo to fine-tune Llama2. We utilized the TRL library incorporating 4-bit QLoRA for efficient training. Fine-tuning was conducted with a batch size of 4 per device with maximum 10,000 steps. Learning rate was $2e-4$ with a constant scheduler and maintained a maximum sequence length of 512. All other parameters align with those specified in the TRL GitHub repo.

Visual Translator. Our implementation is based on the Wizaron/instance-segmentation-pytorch GitHub repo. We utilized a ResNet50 backbone for our instance segmentation network, training separate models for lower (553 images) and upper (486 images) eyelids on CRC data annotated with gland masks. Each model was trained on 256x256 resized images for 300 epochs using a batch size of 8 and a learning rate of 1.0, employing the Adadelat optimizer with a weight decay of $1e-3$. Other parameters align with those specified in the GitHub repo.

LLM-Based Clinical Report Summarizer. The summarizer uses the GPT-4 API in the Erol444/gpt4-openai-api GitHub repo, with a unique seed to control uncertainty. We input raw morphology data along with a structured task template (Section C) to produce clinical reports via GPT-4.

B Discussions

Does MGD Imply DE? The presence of Meibomian Gland Dysfunction (MGD) does not necessarily imply evaporative Dry Eye (DE). While MGD is a common factor in evaporative DE, it is not always the case (Galor, 2014). MGD, DE, and blepharitis are distinct conditions with overlapping symptoms. In our model, we use independent labels for these conditions based on the TFOS 2017 DEWS II Definition and Classification Report (Craig et al., 2017). DE is marked by tear film instability, ocular surface damage, and symptoms. MGD involves ductal stenosis and gland secretion quality, while blepharitis is identified by eyelid margin inflammation, debris, and collarettes. Differentiating these conditions in our model enhances diagnostic accuracy and treatment efficacy.

Is our Dataset Diverse Enough? It is important to note that our datasets come from diverse study populations. Data from the CRC and DREAM (a major clinical trial with 11 meibography sites across the US) are combined. Our distributions closely align with US Census statistics for age, sex, and race. Additionally, our dataset encompasses a wide range of disease severities. Efforts are ongoing to obtain additional data from much younger and older populations, male subjects, and individuals of African ethnicity.

C Template for LLM-based Summarizer

Clinical Report Summarizer Template (to GPT-4):

Task Description:

You are an intelligent medical summary generator. Your task is to generate a clinical report summary for the raw clinical metadata mentioned in the caption. I will provide you with medical data obtained from meibography images of patients and generate concise summaries. Please generate a human readable summary with Q&A format by setting subject's demography, and MG morphology as the Question while the ocular surface disease as the Answer.

Supporting Examples:

Here's an example: {"42_2_R": {"gender": "Male", "age": 30, "race": "Asian", "TMH": 0.28, "NIK BUT": 12.33, "MG_Morph": {"avg_length": 4.878048, "avg_width": 0.463366, "avg_contrast": 13.573304, "avg_tortuosity": 0.277598}, "Dry Eye": "Yes", "Meibomian Gland Dysfunction": "No", "Blepharitis": "Yes"}}

You should output something like:

###Human: Subject 42_2_R and right eye. The person is a male with an age of 30, and the race is Asian. The Tear Meniscus Height (TMH) is 0.28mm, The Non-Invasive Keratograph Tear Breakup Time (NIK BUT) is 12.33 sec. The meibomian gland morphology has average length of 4.87mm, average width of 0.46mm, avg contrast is 13.57, and average tortuosity is 0.28.

###Assistant: The Dry Eye (DE) condition for this subject is Yes, and The Meibomian Gland Dysfunction (MGD) is No, and the Blepharitis is also Yes.

This is the patient 42_2_R, 2 means OS2 category, 42 is patient ID, R is right eye. MG_Morph means meibomian gland morphology features. Please remove values after second decimal place. Please write in one paragraph as a clinical report summary. This summary could be an input data to fine tune large language model.

Prompting the Clinical Metadata

Could you give a clinical report summary of the data?
{your input metadata}