LLM-Driven Descriptive Analysis of Metaphase Cell Images Containing ecDNA

Archit Pimple   Andrew Yin   Mohit Sridhar   Mentors: Utkrisht Rajkumar   Thiago Mosqueiro

apimple@ucsd.edu   anyin@ucsd.edu   msridhar@ucsd.edu   utkrisht96@gmail.com   tmosqueiro@gmail.com

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Background

- Extra Chromosomal DNA (ecDNA) exists outside chromosomes in a cell, often in circular form. In humans, ecDNA is particularly relevant to cancer research because it can drive tumor growth.
- **Limitations of Current Analytical Tools for ecDNA:** Existing tools primarily focus on identifying ecDNA and conducting basic quantitative measurements but do not provide structural insights or analyze metaphase spreads for chromosomal abnormalities or genetic diseases. This gap in research tools limits the ability of scientists to fully interpret their data and understand the role of ecDNA in various biological processes.

Objectives

Our project proposes using LLMs as virtual pathologists to generate detailed insights from metaphase images containing ecDNA.

- **Incorporating Large Language Models (LLMs) for Descriptive Analysis:** By integrating existing tools like ecSeg performing semantic segmentation, LLMs will enhance the interpretation of ecDNA-related structures and provide meaningful descriptions that could assist researchers in diagnosing and analyzing genetic abnormalities.
- **Expected Output:** This project will generate a report using prompt-engineered LLMs to analyze metaphase images with ecDNA, comparing results to expert-annotated ground truth data. Findings will be presented in a research paper and demonstrated through a model showcasing LLM performance, assessed using scoring metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

Methods

N-shot Learning
N-images with varied counts are used as examples for the model.

Multi-Layered Prompting
Multiple questions are asked about the image before asking it about ecDNA.

Temperature Adjustment
Temperature affects the randomness of the LLM response, so adjusting this affects determinism.

Aggregating Results Over Multiple Trials
A baseline prompt is used as question to ask over a certain number of trials, and the generated results are averaged.

Context Messages
Model is provided with different system/context messages using the baseline prompt template to find best suited context.

Ground Truth Data

Our ground truth data is generated using the dataset provided by ecSeg. We use the cropped train and test images and conduct image segmentation to identify the number of nuclei, chromosomes, and ecDNA.
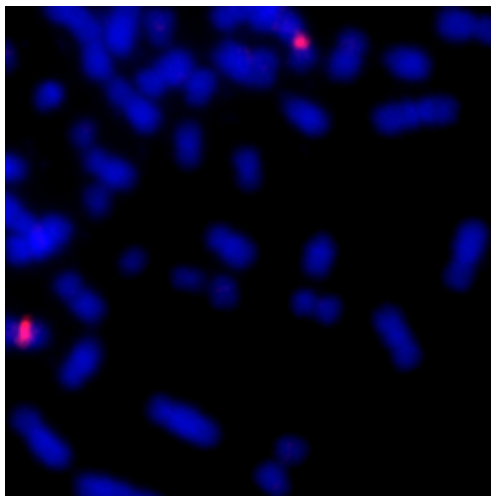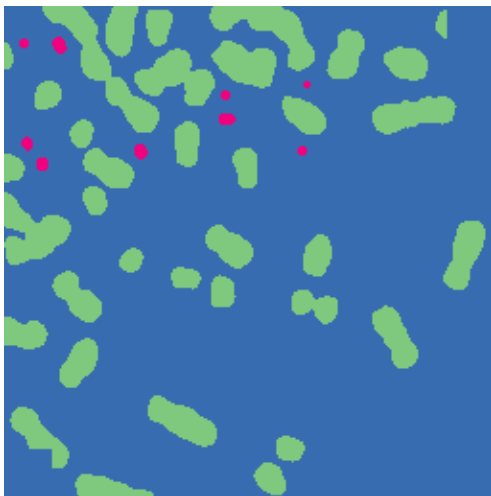

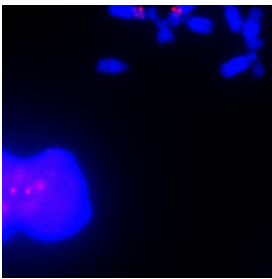Figure 1. Original Image   Figure 2. Segmented Image

LLM Generated Output

The following output from each model is generated using the test image provided above, which has 9 ecDNA. The outputs are generated using the 3-shot Learning Method:
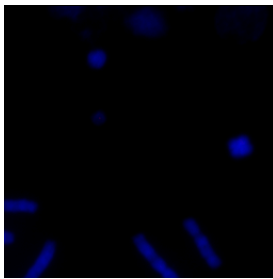
Shared Context: *"You are a pathologist analyzing metaphase cell images. These images contain three main structures: ecDNA, nuclei, and chromosomes. Your goal is to focus on connected component analysis in the image."*
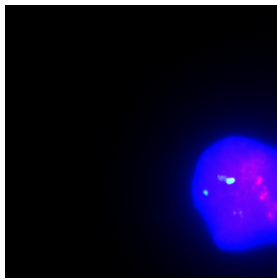
"This is a cell image with 0 ecDNA."   *"This is a cell image with 1 ecDNA."*   *"This is a cell image with 7 ecDNA."*

Shared Prompt: *"How many ecDNA are in the image? Return an integer number."*

Pixtral-12B

- Raw answer: "The analysis indicates that there are **10** distinct ecDNA structures."
- Parsed answer: **10**

Qwen2-VL

- Raw answer: '\n "ecDNA": 7\n '
- Parsed answer: **7**

MiniCPMv2

- Raw answer: There are **10** instances of ecDNA.
- Parsed answer: **10**

Results

Pixtral-12B

Experiment	MAE	RMSE
Baseline	10.71	18.64
3-shot Learning	12.97	23.43
Multi-Layer Prompting	12.37	22.36
0.2 Temperature	12.67	25.30
Result Aggregation	12.18	22.43

Qwen2-VL

Experiment	MAE	RMSE
Baseline	34.94	18.12
3-shot Learning	33.27	17.69
Multi-Layer Prompting	37.03	18.77
0.2 Temperature	29.93	17.05
Result Aggregation	34.80	18.38

MiniCPMv2

Experiment	MAE	RMSE
Baseline	10.44	24.67
3-shot Learning	12.21	24.57
Multi-Layer Prompting	154.46	265.65
0.2 Temperature	9.80	22.33
Result Aggregation	10.58	25.17

Limitations

- **Counting Struggles:** While effective in general descriptive image analysis, LLMs still struggle with specifically counting problems without the assistance of additional models.
- **High Computational Cost:** The majority of LLMs advanced enough for our task require high computational demands and time which makes batch predictions difficult without access to more computational power.

Conclusion

- **Results Summary:** The results show that Pixtral-12B is the best-performing model with respect to MAE and RMSE. Trailed by MiniCPMv2 and then Qwen2-VL, this difference is likely due to the larger context and parameter size of Pixtral allowing it to handle complex queries and produce accurate answers.
- **Looking Ahead:** In the future, we hope to implement fine-tuning to enhance the accuracy of our best-performing model and reduce dependence on pre-training. Additionally, we hope to eventually use datasets annotated by experts to minimize faulty data.