Strategies for Fine-Tuning Geneformer to Predict the Exposure Level of Cancer Cells to Treatments

Student: *Octavian-Teodor Dragon* (O.Dragon@student.tudelft.nl); Supervisors: *Marcel J.T. Reinders, Niek Brouwer*

TU Delft

1. Background

Studying **gene interactions** within a cell can lead to a better understanding of the cell's behavior in different circumstances **[1].** This is a step towards *effective (personalized) treatments for diseases,* but gathering cell samples for diseases where there is not much data is **time-consuming** and **costly** → What if we could apply knowledge from a related domain with a larger quantity of data?

- **Transfer learning:** a versatile approach used to apply knowledge from a domain with lots of data to a similar domain with limited data **[2].** In Machine Learning, this can be done through various fine-tuning strategies **[3]**, i.e. non-iterative strategies → known for their simplicity & being less computationally demanding **[2].**
- **Non-iterative TL**: fine-tuning (FT) strategies where a model's hyper-parameters are adjusted without exhaustive exploration of parameter search space **[2].**

Non-iterative FT strategies have been compared in the context of *classifying medical images* **[4]**, but not for predicting the **sensitivity of cancer cells exposed to treatments.**

2. Research Question

What different strategies can be employed to fine-tune Geneformer so that it is able to correctly identify a sample taken from a cancer cell line as either untreated/not sensitive to a treatment, or treated and sensitive to the treatment?

Topics to be covered to answer the research question:

- Predictive accuracy of evaluated FT strategies;
- Complexity of each fine-tuning strategy;

3. Method & Materials

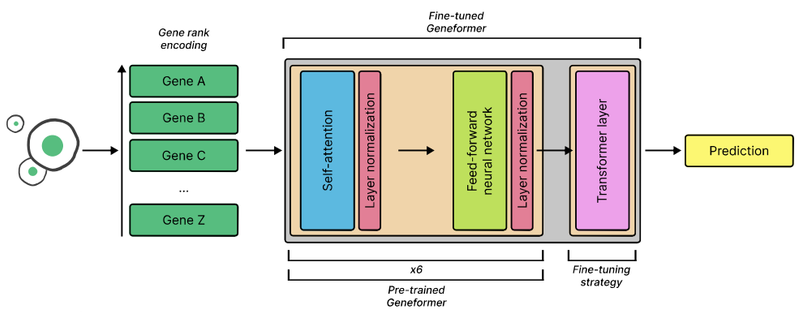

Figure 1

- **Experiment:** classifying the nutlin-3A dosage to which cancer cells were exposed.
- Literature review performed to compile a list of popular fine-tuning strategies → Strategies chosen for assessment: Selective Fine-tuning, Linear Probing, Gradual Unfreeze (last/all)(LP-FT), Full Fine-tuning.
- Dataset: **sciplex2 [5]** - human lung adenocarcinoma cells exposed to BMS345541, dexamethasone, Nutlin-3A, SAHA or DMSO vehicle control.
- **Input:** samples from sciplex2 → cancer cells exposed to Nutlin-3A with dosages between 0μM and 125μM.

4. Experiment & Results

Geneformer *(Figure 1)* has been run using the following parameters: max learning rate, 5 × 10−5; weight decay, 0.001; batch size, 12; no. of layers frozen, depending on the strategy.

- **Task:** (1) pre-process the sample inputs through tokenization and modifying some attribute labels; (2) split the samples into an eval set and a test set with a ratio of 8:2; (3) fine-tune Geneformer on the eval set; (4) predict the nutlin-3A dosage to which samples in the test set were exposed, measure accuracy.
- **Measurements:** confusion matrix, mean accuracy, F1-score.

I). Single-nuclei profiling of human dilated and hypertrophic cardiomyopathy: reproduced to establish a baseline performance. Results obtained were in accordance with the ones presented in the original manuscript: out-of-sample accuracy of **87.31%**, F1 score **0.85.**

II). Two-fold classification - subset of sciplex2 with human lung adenocarcinoma cells exposed to nutlin-3A *(Figure 2)*:

- Dosages**:** 0μM, 25μM;
- **Results**: average out-of-sample accuracy of **96.25%**, F1 score **0.94.**

III). Four-class classification *(Figure 3)* **–** same dataset as in ***II).*** but:

- Dosages: 0.25μM, 2.5μM, 25μM, 125μM;
- **Results**: average out-of-sample accuracy of **68.05%**, F1 score **0.65.**

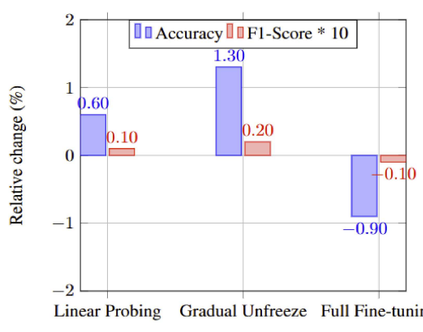

Figure 2

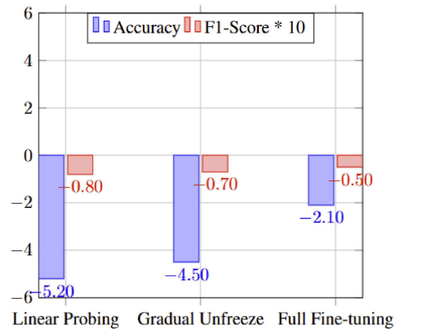

Figure 3

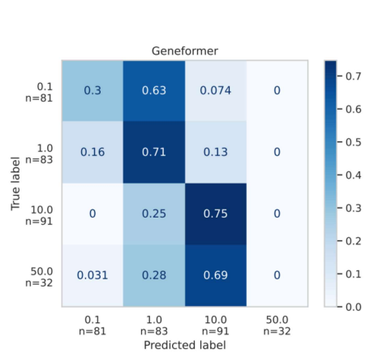

Figure 4

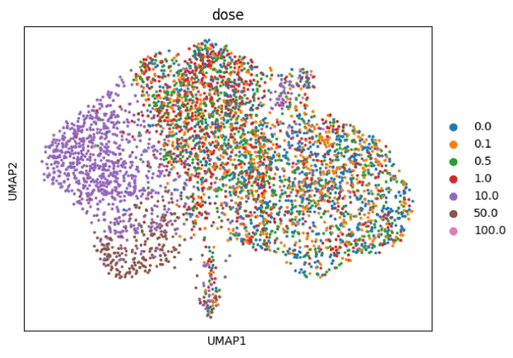

Figure 5

5. Discussion

Different FT methods are suitable for different tasks, similar to medical imaging and other domains **[2].** This depends on the task and how knowledge is being stored within the weights.

Linear Probing is good for simple tasks and tasks where the source and the target domain share a higher degree of similarity.

Contrary to the hypothesis that **Full Fine-tuning** would lead to a model which is better suited for the downstream task, it resulted in worse accuracy scores for both scenarios.

Gradual Unfreeze can be used for a more control over how much the layers are fine-tuned (no. of epochs), at the cost of more complexity.

The confusion matrices from the strategies for the four-class task show a clustering of the data into two groups: (1) cells exposed to dosages < 25μM and (2) cells exposed to ≥ 25μM *(Figures 4,5)*.

6. Conclusions & Further Work

- There is no optimal solution. Different methods are suitable for different tasks **[2].**
- Only non-iterative fine-tuning strategies were explored → Expanding on the research by also exploring iterative TL approaches.
- A limited set of samples was used: human lung adenocarcinoma cells exposed to nutlin-3A. A more comprehensive evaluation should be done in future works to assess the model's ability to perform on a more general task → sciplex3, more cell lines and compounds.

7. References

[1] Swen Jesse J, et al. Translating pharmacogenomics: Challenges on the road to the clinic. PLoS Medicine, 4:1317–1324, 2007. **[2]** Davila Ana, et al. Comparison of fine-tuning strategies for transfer learning in medical image classification. Image and Vision Computing, 146:105012, 2024. **[3]** Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10):1345–1359, 2010. **[4]** Hee E. Kim, Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E. Maros, and Thomas Ganslandt. Transfer learning for medical image classification: a iterature review. BMC Medical Imaging, 22(69), 2022. **[5]** Srivatsan Sanjay R., McFaline-Figueroa Jos´e L., Ramani Vijay, Saunders Lauren, Cao Junyue, Packer Jonathan, Pliner Hannah A., Jackson Dana L., Daza Riza M., Christiansen Lena, Zhang Fan, Steemers Frank, Shendure Jay, and Trapnell Cole. Massively multiplex chemical transcriptomics at single-cell resolution. Science, pages 45–51, 2019.