

Technical University of Applied Sciences Würzburg-Schweinfurt (THWS)
Faculty of Computer Science and Business Information Systems

Master Thesis Proposal

Determination of Drug Efficacy on Pancreatic Tumor 3D Spheroidal Tissues

**submitted to the Technical University of Applied Sciences Würzburg-Schweinfurt
in the Faculty of Computer Science and Business Information Systems to
complete a course of studies in Master of Artificial intelligence**

BIBIN BABU

Submitted on: 13.05.2024

Initial examiner: Prof. Dr. Magda Gregorová
Secondary examiner: Prof. Dr. Jan Hansmann

Abstract (en)

Pancreatic tumor treatment is hindered by the intricate nature of tumors and their diverse microenvironments. This complexity necessitates an exploration into identifying optimal drug combinations and concentrations tailored to each patient's specific tumor characteristics. This thesis aims to assess drug efficacy by ranking these various drug combinations and concentrations. The ranking is based on features extracted from bright-field microscopy images of three-dimensional tumor tissue models using representation learning. The core challenge is to learn robust features that accurately characterize alterations in these tumor tissue models induced by drug application over time. This research seeks to develop a standardized and effective approach for evaluating drug efficacy, potentially improving treatment outcomes for pancreatic tumor patients.

Abstract (de)

Die Behandlung von Bauchspeicheldrüsentumoren wird durch die komplexe Natur der Tumore und ihre vielfältigen Mikroumgebungen behindert. Diese Komplexität erfordert eine Untersuchung zur Identifizierung optimaler Medikamentenkombinationen und -konzentrationen, die auf die spezifischen Tumoreigenschaften jedes Patienten zugeschnitten sind. Diese Masterarbeit zielt darauf ab, die Wirksamkeit von Medikamenten zu bewerten, indem sie diese verschiedenen Medikamentenkombinationen und -konzentrationen einstuft. Die Bewertung basiert auf Merkmalen, die aus Helligkeitsmikroskopiebildern dreidimensionaler Tumorgewebsmodelle mittels Repräsentationslernen extrahiert werden. Die zentrale Herausforderung besteht darin, robuste Merkmale zu erlernen, die Veränderungen in diesen Tumorgewebsmodellen genau charakterisieren, die durch die Anwendung von Medikamenten über Zeit induziert werden. Diese Forschung zielt darauf ab, einen standardisierten und effektiven Ansatz zur Bewertung der Medikamenteneffizienz zu entwickeln, der möglicherweise die Behandlungsergebnisse für Patienten mit Bauchspeicheldrüsentumoren verbessert.

1 Introduction

Pancreatic tumor presents a significant challenge in terms of treatment due to its heterogeneous nature and the mutations that occur during its progression within the human body. Clinicians rely on case studies, human trials, and their own expertise gained from past patient treatments to select drugs for new patients. However, this approach is often based on trial and error, with varying outcomes. Patients may experience either successful treatment or severe side effects such as hair loss and damage to other organs. Since each patient's tumor cells exhibit unique characteristics influenced by factors such as age and genetics, treatments that have worked for one patient may not be effective for another. Consequently, clinicians may need to change the prescribed drugs or try different combinations, which can lead to delays and increased risks for the patient, including mortality.

In light of these challenges, Researchers at Fraunhofer Translational Center for Regenerative Therapy TLZ-RT Wuerzburg, propose a vision for the future: cultivating multiple three-dimensional tumor tissue models for each patients in the lab using biopsy samples and studying the efficacy of drugs on these three-dimensional tumor tissue models first. *(Note: In this thesis, "3D tumor tissue models or tumor tissue models" refers to physical, lab-grown tissues and not computational or AI models.)* By conducting drug development experiments and analyses on these tissue models, they aim to find the optimum or best drug combination tailored to each patient's specific tumor characteristics. This approach can not only minimize direct side effects on human patients and reduce the time needed to select the most effective personalized treatment, thereby decreasing the risk of that patient's mortality, but also significantly reduce the cost and time of preclinical testing in the drug development process. Ultimately, these information obtained from drug efficacy assessment experiments can inform clinicians' decisions, enabling them to select the most effective drug combination before administering it to the patient.

As a proof of concept, The Fraunhofer TLZ-RT Wuerzburg laboratory utilized a modular dual-arm robot-based system [3], equipped with incubators and bioreactors under physiological conditions to study drug efficacy for the long-term culture of these three-dimensional tumor tissue models. One advantage of this platform is its ability to capture bright-field microscopy images of 3D tumor tissue models using a customized microscope setup integrated into the robotic platform, offering flexibility in image acquisition according to experimental needs.

Although the vision for the future is to simulate the identical interaction environment of drugs with tumor cells as it occurs in the human body, current technology has not yet achieved this. The current three-dimensional tumor tissue models developed in the lab do not fully resemble real pancreatic tumor cells found in the human body. These 3D tumor tissue models only contain pure tumor tissues, whereas real human pancreatic tumor cells exist within a complex microenvironment comprising tumor cells, blood vessels, other tissues, and various cell types. Despite this limitation, our work serves as a valuable starting point for studying drug efficacy in a controlled environment. Fortunately, if we are able to replicate human body tumor cells in the lab in the future, the techniques currently used to study the bright-field microscopy images will still be applicable, since they are grayscale images. However, the fact that bright-field microscopy images are two-dimensional limits the ability to perform a comprehensive analysis of the drug's impact on the entire 3D structure of the cultivated tumor tissue models.

Alternatives to bright-field microscopy images include 3D fluorescence microscopy and luminescent cytotoxicity assays. However, both methods are invasive. Fluorescent molecules tend to generate reactive chemical species under illumination, enhancing phototoxic effects. This chemical reaction with the 3D tumor tissue model may alter its structure, making it not suitable to isolate the drug's effect over time. Similarly, luminescent cytotoxicity assays result in a dead culture, rendering them unsuitable for longitudinal studies. Additionally, both methods require removing the well plate from the isolated culture environment for extended periods, making the samples susceptible to external environmental factors. For instance, in fluorescence microscopy, cells are particularly vulnerable to phototoxicity from short wavelength light. In contrast, bright-field microscopy images are non-invasive, allowing continuous culture and the possibility of creating time series of images to study dynamic changes. Therefore, we rely on bright-field microscopy images to study the time-evolutionary effects of drugs, only applying the alternative methods at the end of the experiment.

3D tumor tissue models were cultured in well plates containing 96 wells, each providing a nutrient medium that allows them to maintain their tissue-specific functions in vitro. Although each plate can yield 96 pure 3D tumor tissue models, the edge effect is accounted for, where outer wells may be exposed to variable conditions such as temperature fluctuations, increased evaporation rates, and other environmental factors. Consequently, we restrict our analysis to the 60 inner wells per plate, adhering to standard procedures to ensure consistent and reliable experimental data. These 60 wells are divided into sections to provide the three experimental groups (explained below). From day 1 to day 7, the 3D tumor tissue models develop and on day 7, drug is applied. To isolate the effects of the drugs, changes in 3D tumor tissue models deterioration are evaluated on day 3 post-drug application, which is day 10, in accordance with established medical protocols and previous research findings.

Based on the drug concentration applied to 3D tumor tissue models, the bright-field microscopy images we capture can be categorized into three main categories:

Images of

1. Control (0 percentage drug applied) - For clarity in the paper, we refer to this category as "untreated."
2. Single concentration (theoretically recommended single concentration of drug treatment)
3. Drug screening: different drug combinations and concentrations used for experimental study of drug efficacy, which may or may not result in the killing of surrounding non-tumor cells in the human body and potential side effects.

For clarity in the paper, we refer second and third categories as "treated."

We assess the efficacy of the drug by comparing the changes it induces in the untreated bright-field microscopy images over a period of time. The current methods to differentiate these changes involve studying the alterations from day 7 (after applying the drug) to Day 10. These changes are typically observed in three main parameters:

1. Size/Area
2. Circularity/Diameter/Perimeter
3. Pixel intensity or color change

These parameters serve as human-interpretable metrics for assessing the efficacy of the drug. However, there may be other hidden information or patterns within these bright-field microscopy images that are not human-interpretable. This potential can be explored using representation learning techniques. Additionally, this method can provide more standardization compared to manual assessment.

2 Objective

This thesis aims to assess drug efficacy by ranking different drug combinations and concentrations. The ranking is based on features extracted from bright-field microscopy images of three-dimensional tumor tissue models using representation learning. The primary challenge lies in learning the efficient features of alterations induced in these tumor tissue models by the impact of drug application over a period of time. We believe this assessment can potentially aid clinicians in making informed decisions to identify the optimal drug combination tailored to each patient's unique tumor characteristics.

3 Research questions

1. Can we learn latent features that effectively establish a ranking of drug efficacy from bright-field microscopy images, specifically features that capture the alterations induced in three-dimensional tumor tissue models by drug application over a period of time?
2. What methodologies and frameworks can be employed to extract and learn these hidden representations efficiently?
3. What could be reasonable metrics, such as L2 loss or cosine similarity, for supporting the relative assessment of drug efficacy?

4 Related works

Base neural network architecture for representation learning. Learning visual representations of medical images, such as X-rays (radiographic images) and bright-field microscopy images, is crucial for medical image understanding. However, progress in this area has been hindered by the heterogeneity and complexity of subtle features in these images, especially when they don't have labels. Existing work often relies on fine-tuning weights transferred from ImageNet pretraining (Wang et al., 2017 [10] ; Esteva et al., 2017 [4] ; Irvin et al., 2019 [6]), which is suboptimal due to the drastically different characteristics of medical images. Recent studies have shown promising results using unsupervised contrastive learning on natural images, but these methods have limited effectiveness on medical images because of their high inter-class similarity.

To address these challenges, researchers have proposed various innovative approaches. ConVIRT [12] offers an alternative unsupervised strategy for learning medical visual representations by exploiting naturally occurring paired descriptive text. This method introduces a new approach to pretraining medical image encoders using paired text data via a bidirectional contrastive objective between the two modalities. It is domain-agnostic and requires no additional expert input. However, given the absence of specific paired text data for our image dataset, ConVIRT does not offer a solution tailored to our specific problem.

The contrastive loss used in ConVIRT is derived from the SimCLR [2] self supervised learning framework. SimCLR learns representations by maximizing agreement between differently augmented views of the same data example via a contrastive loss in the latent space. The framework consists of a neural network base encoder that extracts representation vectors from augmented data examples. The framework allows for various choices of network architecture without any constraints. The authors opt for simplicity and adopt ResNet, introducing a learnable nonlinear transformation between the representation and the contrastive loss to substantially improve the quality of the learned representations. However, these methods require careful treatment of negative pairs, typically relying on large batch sizes to retrieve them. Additionally, their performance is highly dependent on the choice of image augmentations. BYOL (Bootstrap Your Own Latent) [5] addresses these limitations by using an architecture with online and target neural networks, which does not require negative pairs and is more robust to the choice of image augmentations compared to contrastive methods.

While SimCLR has achieved impressive success in the computer vision field, directly applying it to the time series domain often yields poor performance due to its data augmentation and feature extractor not being tailored to the temporal dependencies inherent in time series data. To address this limitation and to obtain high-quality representations of univariate time series, [11] proposed TimeCLR, a framework that combines the strengths of Dynamic Time Warping (DTW) and InceptionTime. Drawing inspiration from the DTW-based k-nearest neighbor classifier, they introduced DTW data augmentation. This technique generates phase shifts and amplitude changes targeted by DTW, preserving the time series structure and feature information. By integrating the advantages of DTW data augmentation and InceptionTime, TimeCLR method extends SimCLR and adapts it effectively to the time series domain. [9] conducted a comprehensive comparative analysis between contrastive and generative self-supervised learning methods for time series data, focusing specifically on SimCLR and MAE (Masked Autoencoder). They observed that, overall, MAE tends to converge more rapidly and delivers impressive performance, particularly when the fine-tuning dataset is relatively small (around 100 samples). However, in scenarios with larger datasets, SimCLR demonstrates a slight but consistent outperformance over its generative counterparts.

Another recent alternative study for self-supervised visual representation is DINO [1], which can be also interpreted as a form of self-distillation with no labels. DINO provides new properties to Vision Transformers that stand out compared to convolutional networks.

SupCon [8] extends the self-supervised batch contrastive approach to the fully-supervised setting, allowing us to effectively leverage label information. Clusters of points belonging to the same class are pulled together in embedding space, while simultaneously pushing apart clusters of samples from different classes. The drug applied to tumor samples could be used as labels for the bright-field microscopy images.

Coupling engineered features and learned representation. Due to its specificity, fluorescence microscopy has become a quintessential imaging tool in cell biology. However, photobleaching, phototoxicity and related artifacts continue to limit its utility. Recently, it has been shown that artificial intelligence (AI) can transform one form of contrast into another. Mikhail et al.[7] present phase imaging with computational specificity (PICS), a combination of quantitative phase imaging (QPI) and AI, which provides information about unlabeled live cells with high specificity. By applying the computed fluorescence maps back to the QPI data, they measured the growth of both nuclei and cytoplasm independently over many days without loss of viability. This work could provide valuable insights for coupling fluorescent data with learned representations.

5 Proposed methodologies

We propose two different methodologies to approach the objective. The general idea of the proposed methodologies is to leverage representation learning of bright-field microscopy images to develop a ranking/ordering scale (1 to n) for these images.

5.1 Methodology 1

Step 1: Create a latent space representation of each image using contrastive learning techniques such as SimCLR, masked autoencoder, or any other self supervised architectures such as DINO that can effectively help learn the efficient features of alterations induced in three-dimensional tumor tissue models by the impact of drug application over a period of time.

Step 2: Train a time series prediction model exclusively on the representations of untreated images from Day 7 to Day 10 to predict the representation of Day 10 image.

Step 3: Perform inference on the representations of test images, which include untreated, clinically recommended, and drug screening images.

Since the time series model is trained solely on the representations of untreated images, the inference loss/metric (i.e., the difference between the predicted and actual Day 10 image representations) will be very small for untreated images. Conversely, the inference loss/metric will increase for treated images as their representations deviate from those of untreated images. This inference loss/metric will be used as the feature for the ranking/order scale, where the initial images will start with untreated images that have very small inference loss/metric, and the scale will end with images having high inference loss/metric in ascending order. Determining a reasonable inference loss/metric will be one of the research problems to tackle.

5.2 Methodology 2

This ranking/ordering scale categorizes the images into two distinct groups, ordered as follows:

The first group includes treated images where no explosion of tissue has occurred due to the drug impact (i.e., the tissue still maintain complete circularity after drug application). These images are ranked/ordered from 1 to x based on the percentage of tumor deterioration in ascending order, as observed on day 10.

The second group comprises treated images where tissue explosion has occurred, assuming that the tumor is potentially 100 percentage killed in every image in this group. These images are ranked/ordered from $x+1$ to n based on the degree of tumor tissue explosion (from less exploded to highly exploded), which can be learned using classical computer vision techniques like boundary detection, representing the ends of the scale.

This percentage of tumor deterioration is determined through a comparative study of 3D fluorescence microscopy or luminescent cytotoxicity assays between 30 untreated samples and 30 samples treated with the specific drug combination from a half-divided well plate setup.

In order to couple the percentage of tumor deterioration data with bright-field microscopy images, first, use a self-supervised, unsupervised, or supervised base neural network such as as DINO[1], SimCLR[2], SupCon[8] or ConVIRT[12], to learn the latent space representation of all images. Subsequently, we will train a ranking model on top of this frozen base network to rank each image according to the corresponding percentage of tumor deterioration.

In this ranking scale, x and n are natural numbers that define the boundaries and transitions between the groups, ensuring a systematic and comprehensible ranking from untreated to completely exploded tumor images.

Advantage of training AI in a supervised setting is the ability to circumvent the need for invasive and expensive 3D fluorescent microscopy images or luminescent cytotoxicity assays, as these data are already trained to correlate with bright-field microscopy images. Consequently, when new bright-field microscopy images are introduced as test data, we can predict the percentage of tumor deterioration.

6 Conclusion

These methodological steps collectively form the framework for addressing critical research questions and challenges throughout the course of this thesis. By exploring the dataset, assessing models, and developing custom architecture, we aim to to assess drug efficacy by ranking different drug combinations and concentrations.

7 Proposed timeline



Figure 7.1: Proposed Master Thesis timeline

Bibliography

- [1] Mathilde Caron et al. *Emerging Properties in Self-Supervised Vision Transformers*. 2021.
- [2] Ting Chen et al. *A Simple Framework for Contrastive Learning of Visual Representations*. 2020.
- [3] Sofia Dembski et al. “Establishing and testing a robot-based platform to enable the automated production of nanoparticles in a flexible and modular way”. In: *Scientific Reports* (2023).
- [4] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *Nature* (2017).
- [5] Jean-Bastien Grill et al. *Bootstrap your own latent: A new approach to self-supervised Learning*. 2020.
- [6] Jeremy Irvin et al. *CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison*. 2019.
- [7] Mikhail E. Kandel et al. “Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments”. In: *Nature Communications* 11.1 (Dec. 2020).
- [8] Prannay Khosla et al. *Supervised Contrastive Learning*. 2021.
- [9] Ziyu Liu et al. *Self-Supervised Learning for Time Series: Contrastive or Generative?* 2024.
- [10] Xiaosong Wang et al. “ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 3462–3471.
- [11] Xinyu Yang, Zhenguo Zhang, and Rongyi Cui. “TimeCLR: A self-supervised contrastive learning framework for univariate time series representation”. In: *Knowledge-Based Systems* 245 (2022), p. 108606.
- [12] Yuhao Zhang et al. *Contrastive Learning of Medical Visual Representations from Paired Images and Text*. 2022.

BIBIN BABU, 13.05.2024