

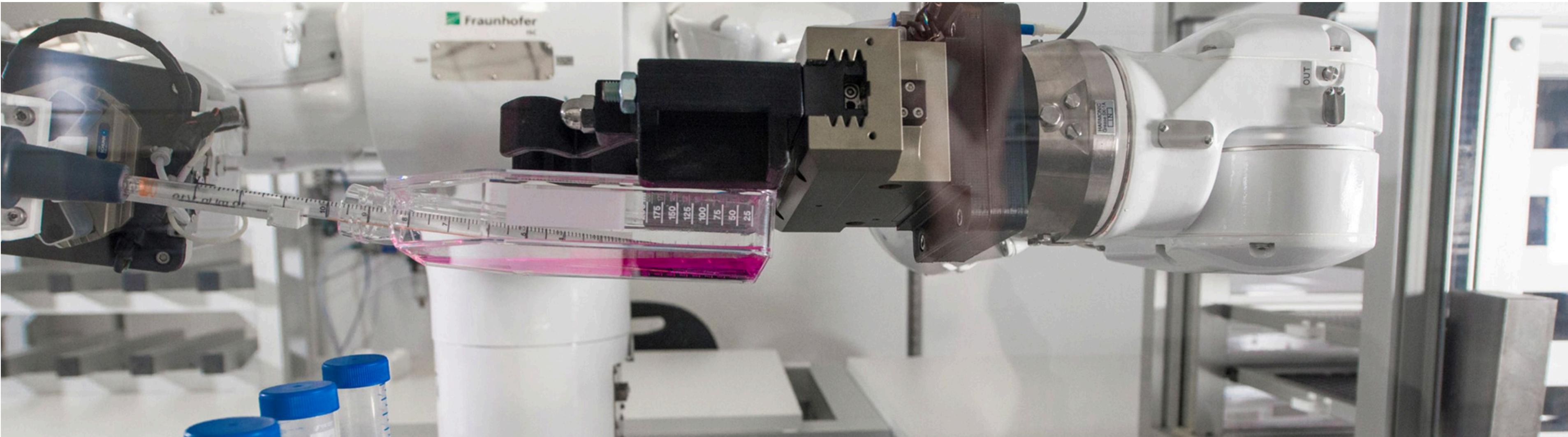
Determination of Drug Efficacy on Pancreatic Cancer 3D Spheroidal Tissues

Bibin Babu

15.04.2024

Supervisors

Prof. Dr. Magda Gregorová, Prof. Dr. Jan Hansmann



Motive

the current tumor model study developing in the lab can't say 100% resembles the reality of human body. because those only contains pure cancer cells where in reality it's a cluster of cancer cell+ blood vessels+ other tissues + different vessels (In future if we get to implement the reality still bright field images are applicable)

But eventough it's a good starting point.

current method :

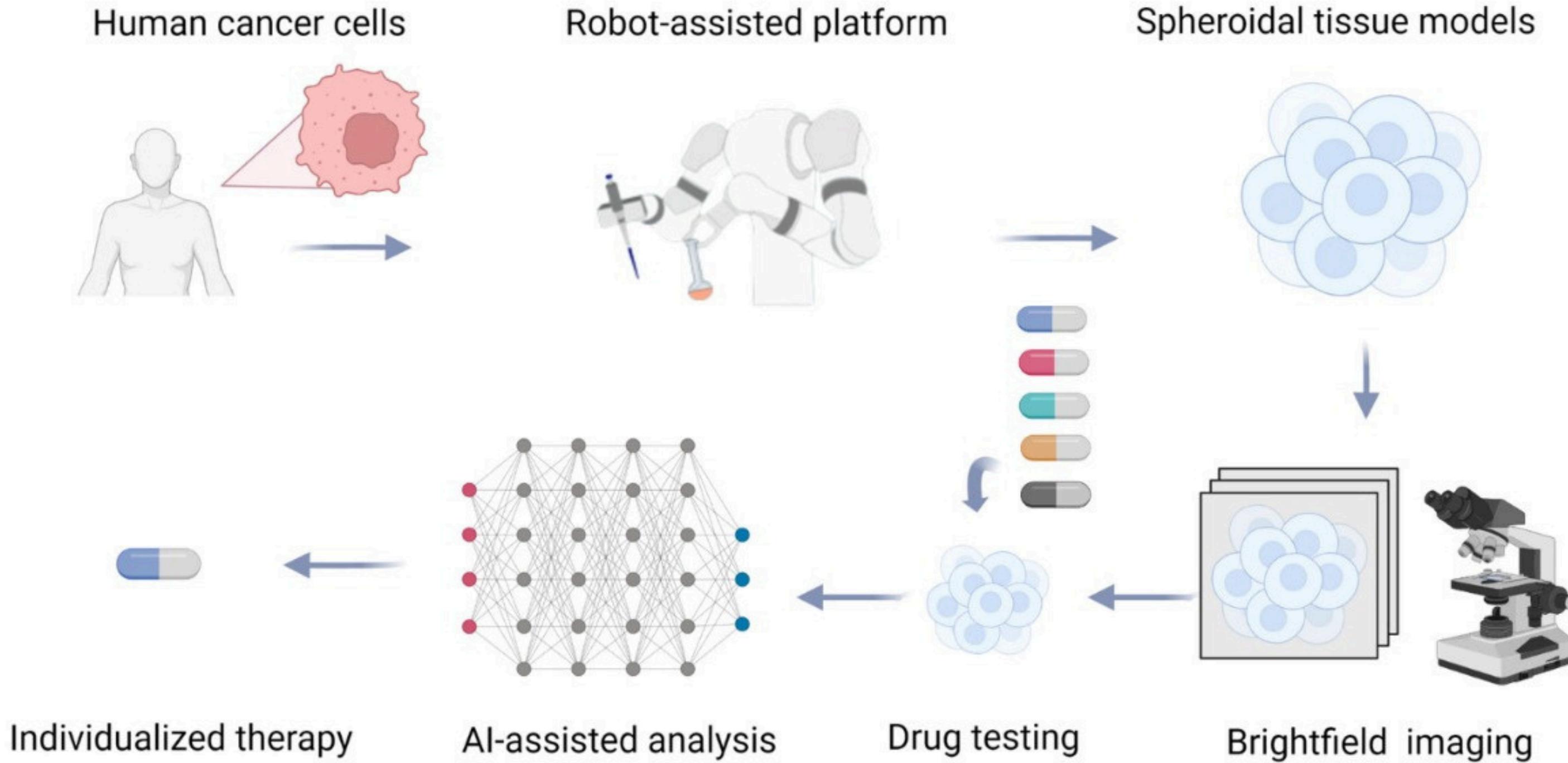
For cancer especially when it comes to pancreatic cancer cells there is no standard treatment due to it's variations and mutations. Doctor read case studies and use his own expertise to refer the drug to a patient. If the patient is lucky enough it works or give worse side effects like hair loss, bad effects on other body parts. etc. doctor changes the drug in that case and try different combo by that time there is a potential risk that the patient is already dead.

That's where we propose a vision for the future/ concept which can reduce direct effect on human patient by cultivating artificially at lab using biopsy we produce multiple tumor cells of one patient and study the effect of drugs. at the end we can use the data/information of which drug combination and select the most effective one before injecting directly to human patient.

the current tumor model study developing in the lab can't say 100% resembles the reality of human body. because those only contains pure cancer cells where in reality it's a cluster of cancer cell+ blood vessels+ other tissues + different vessels (In future if we get to implement the reality still bright field images are applicable)

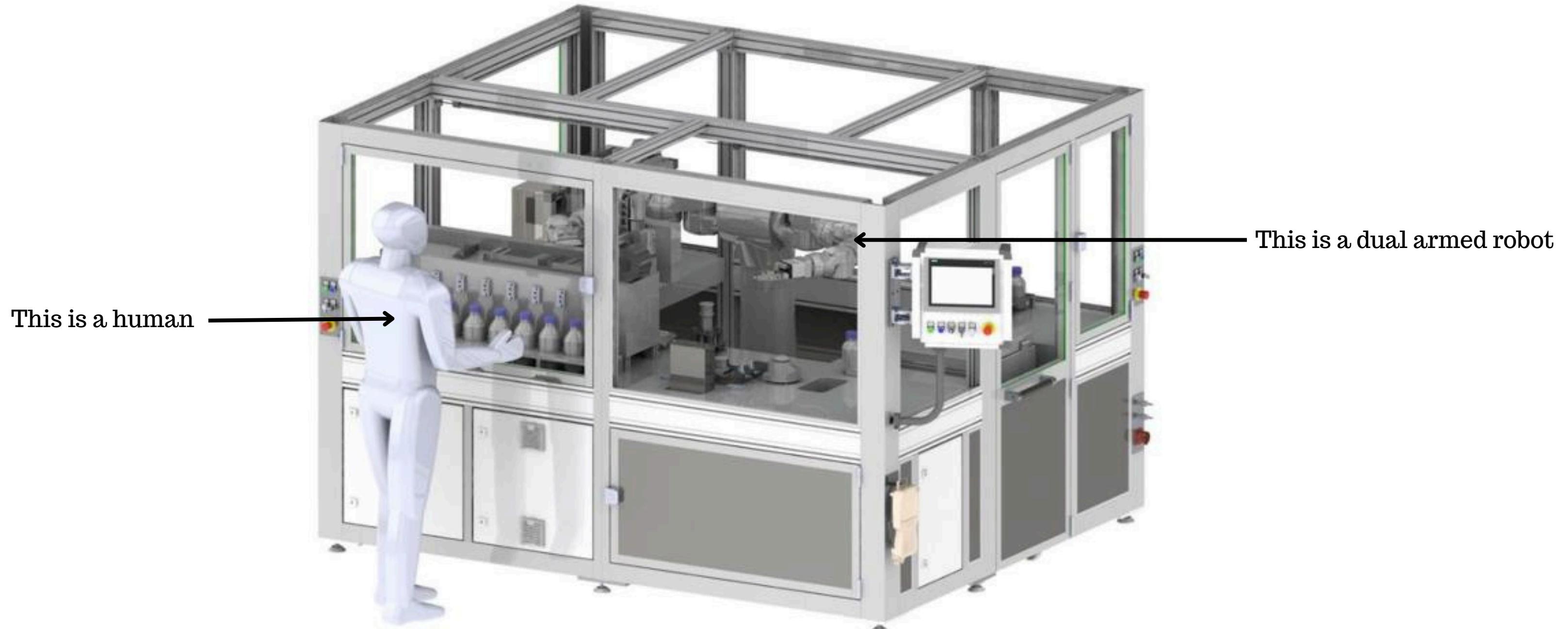
But eventough it's a good starting point.

Big picture

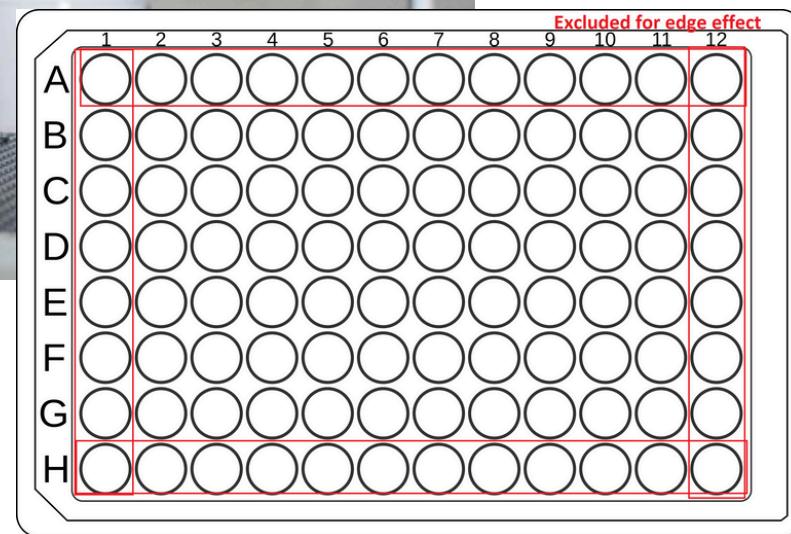
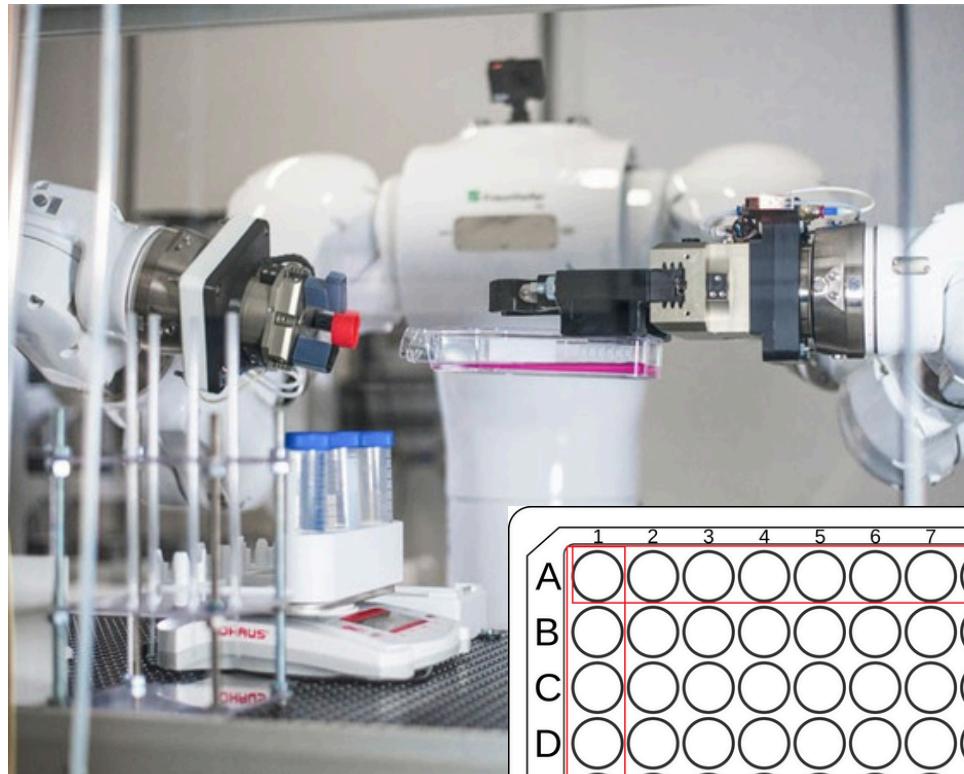


second slide to third slide transition : tool for this gonna be in next slide the setup

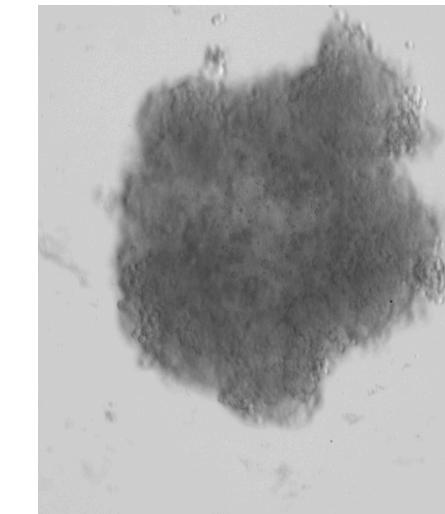
Setup



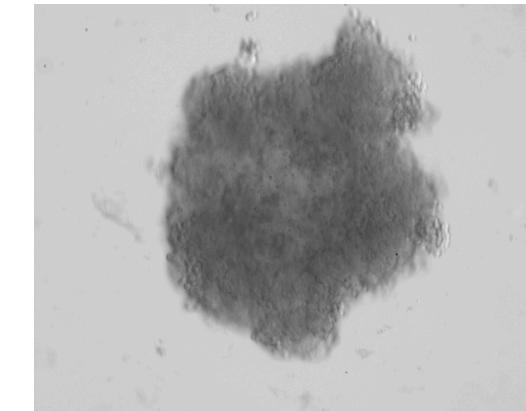
- Increased the synthesis accuracy and reproducibility
- Reduced the personnel time and costs by up to 75%



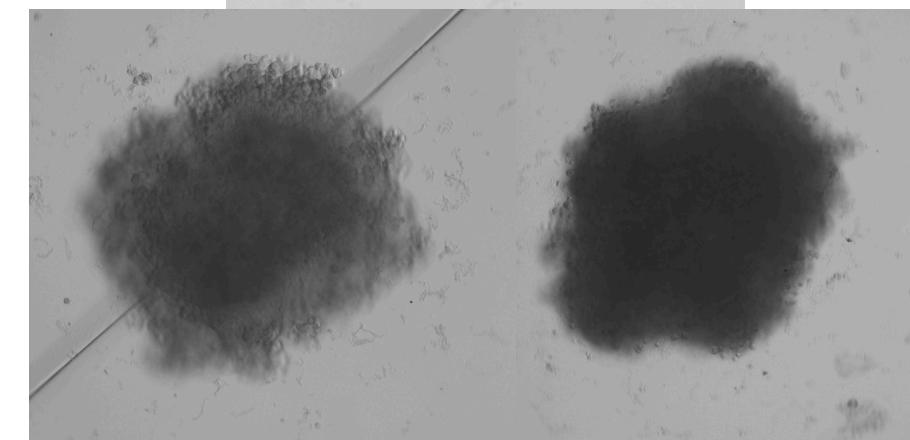
Day 1



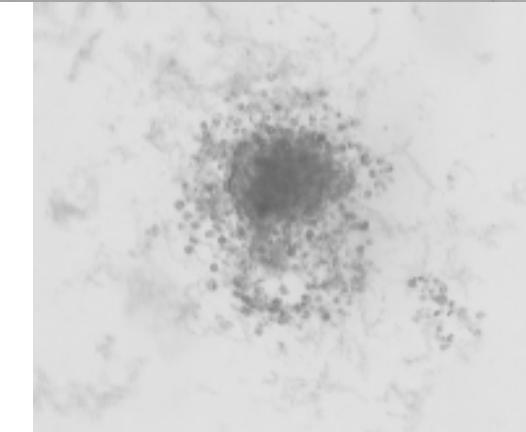
Day 3



Untreated (controlled)

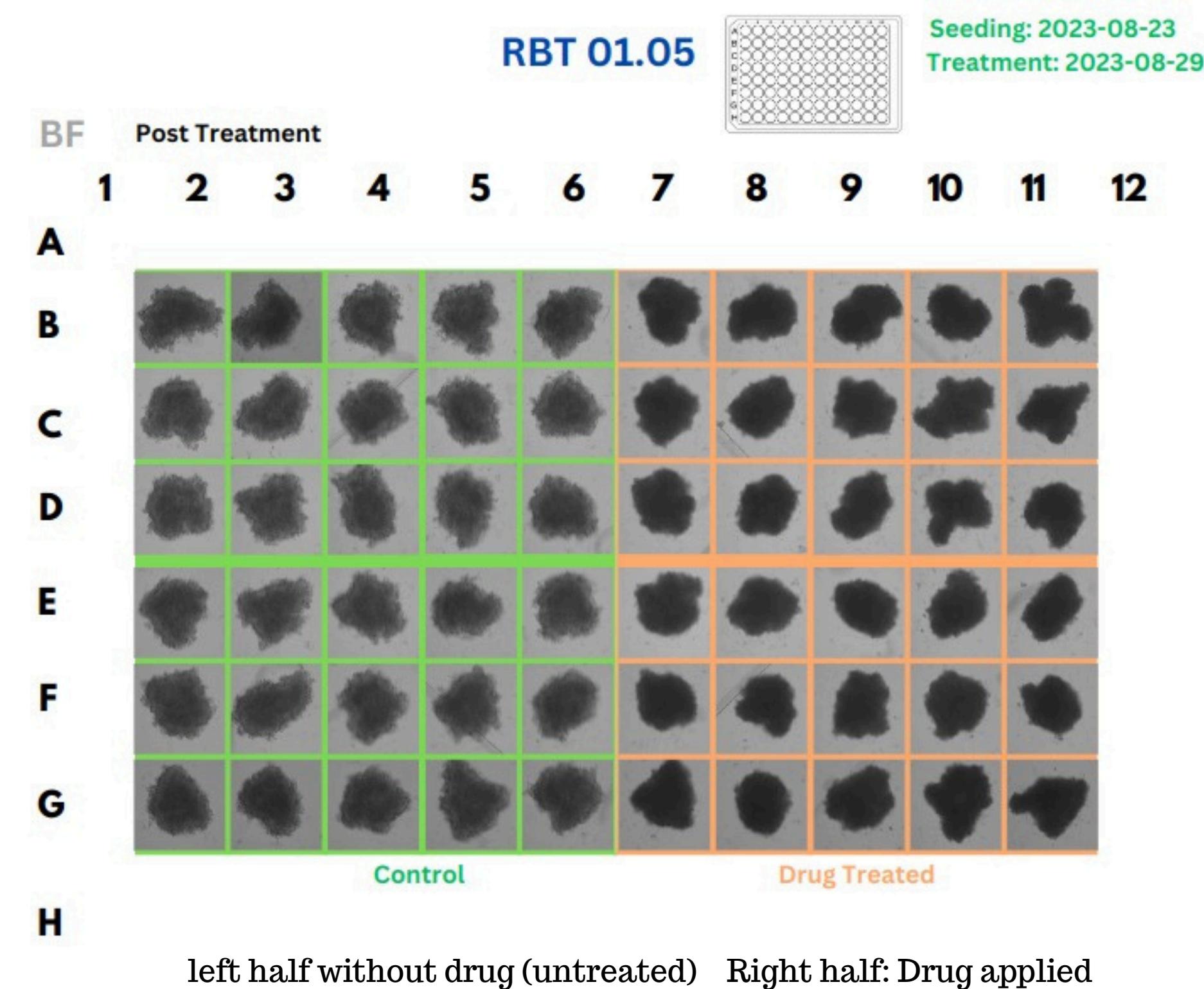


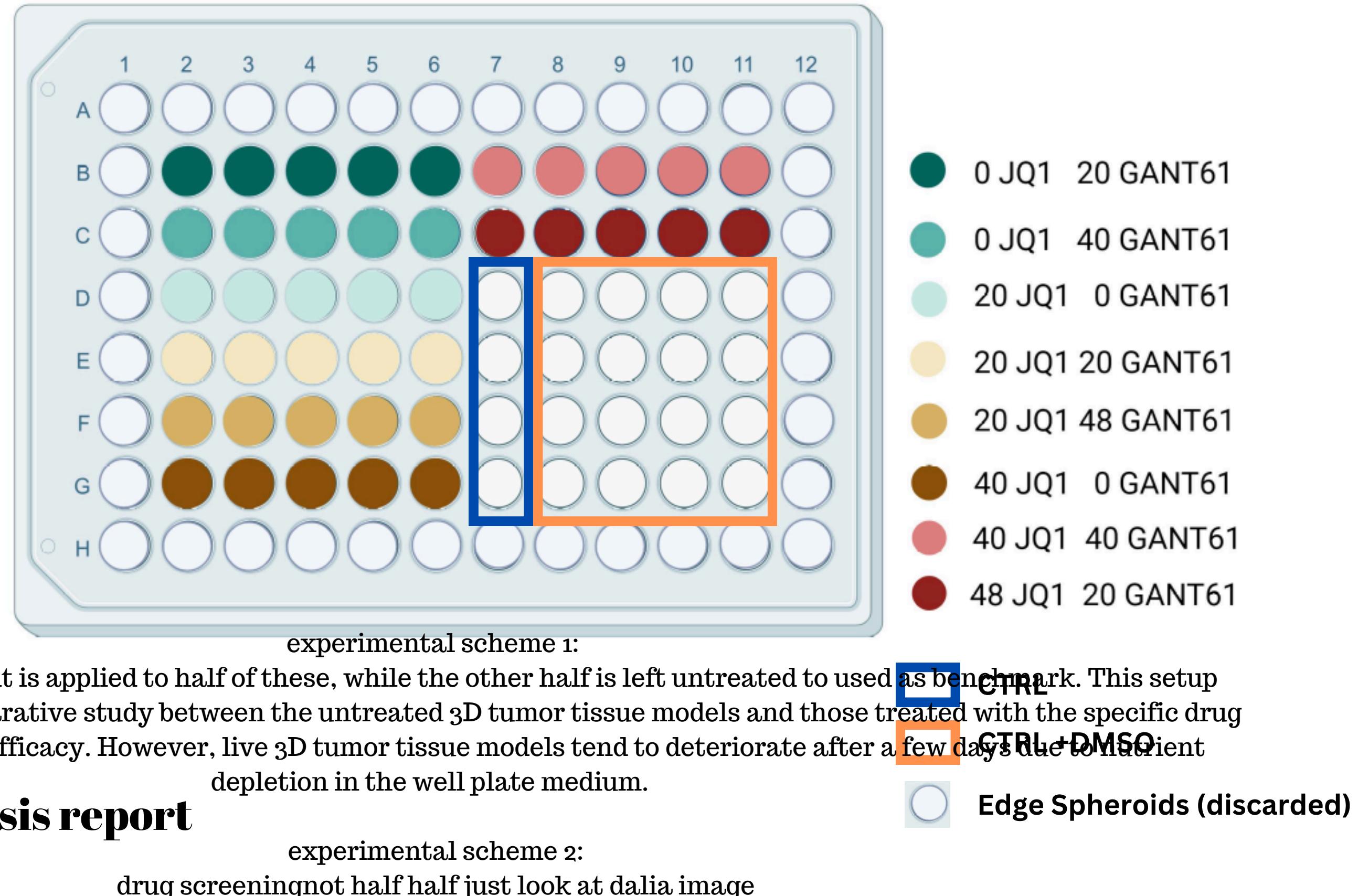
Treated (Clinically recommended with std concentration)
Eg: 20%



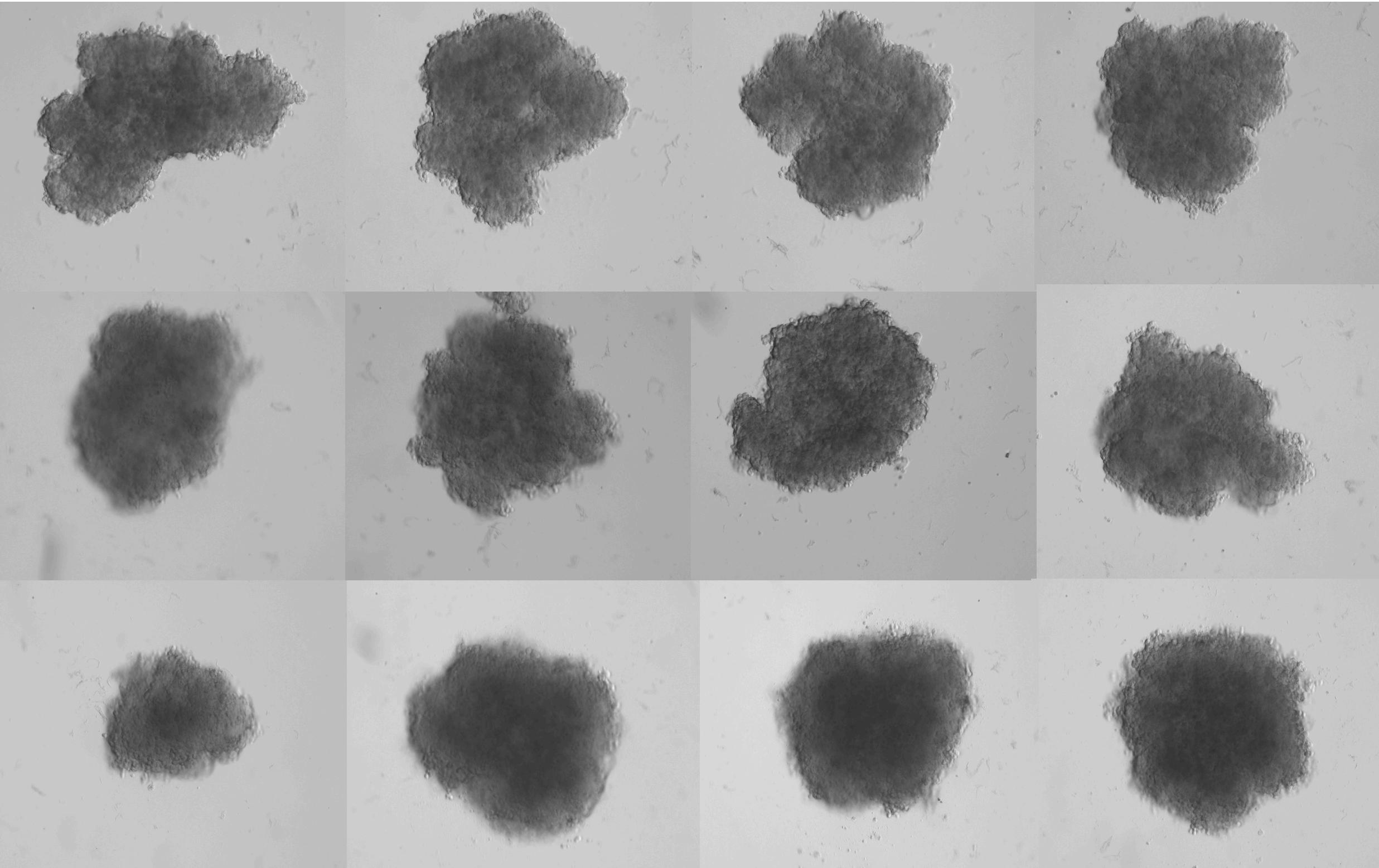
Drug screening
(trying different concentrations)
Eg: 100%



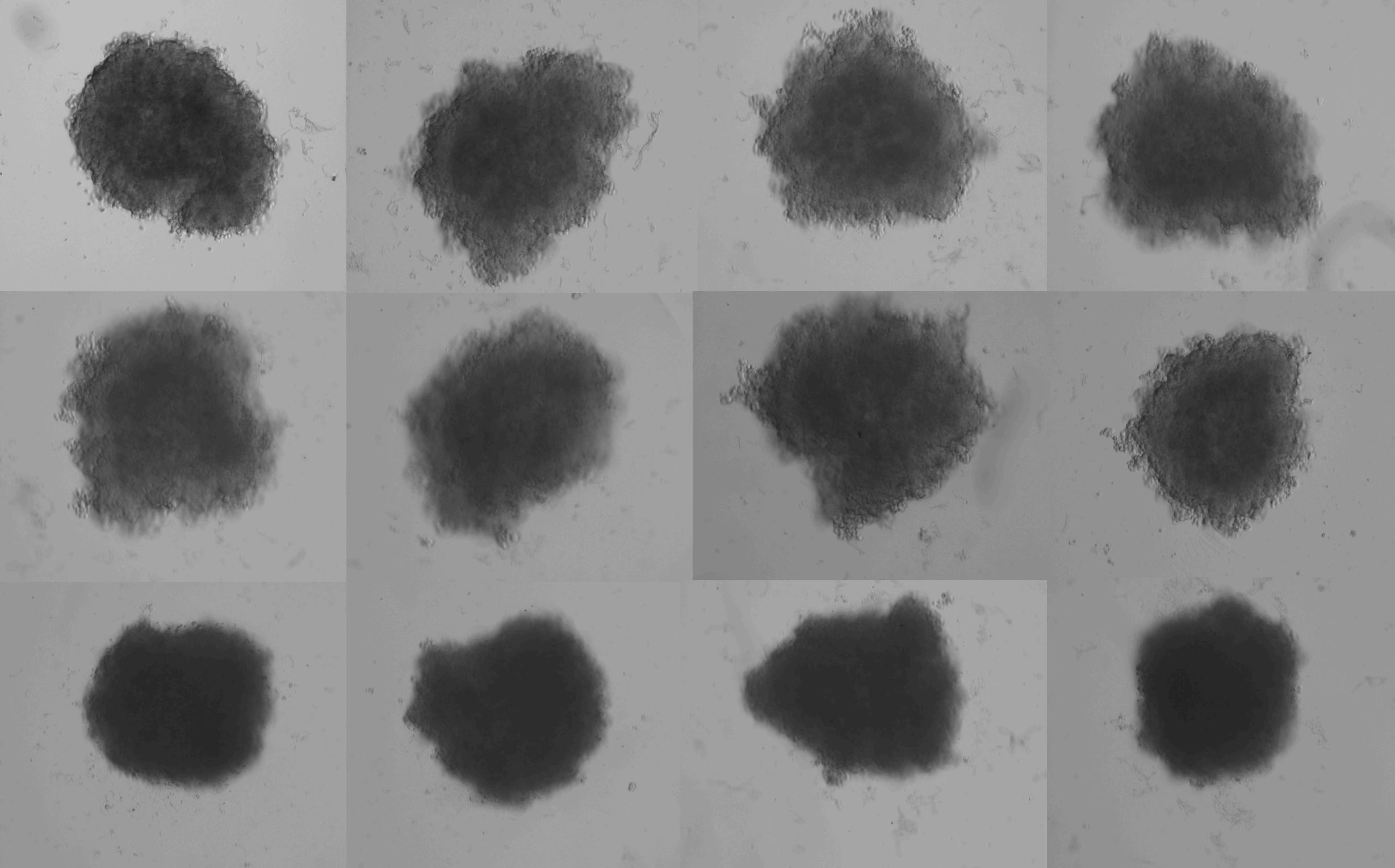




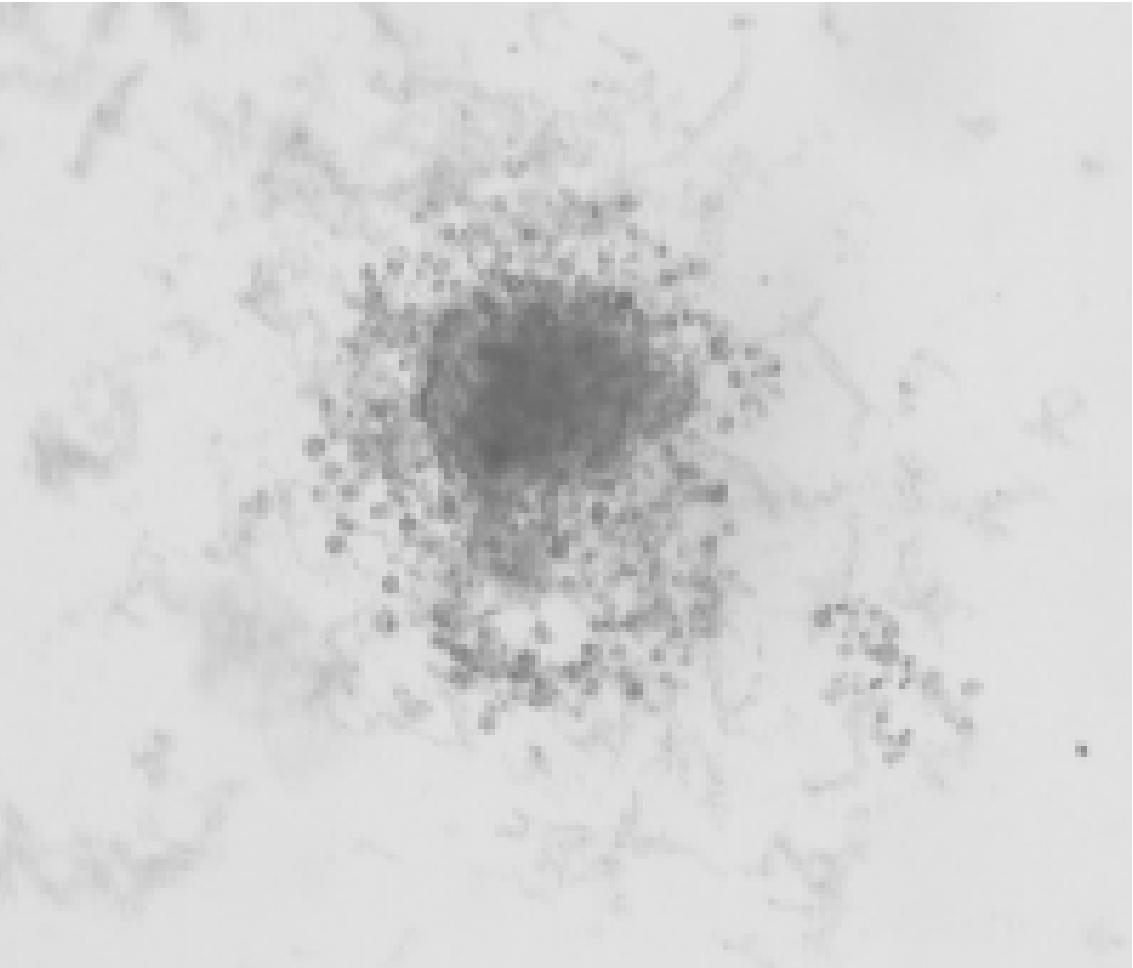
Untreated



Treated (recommended)



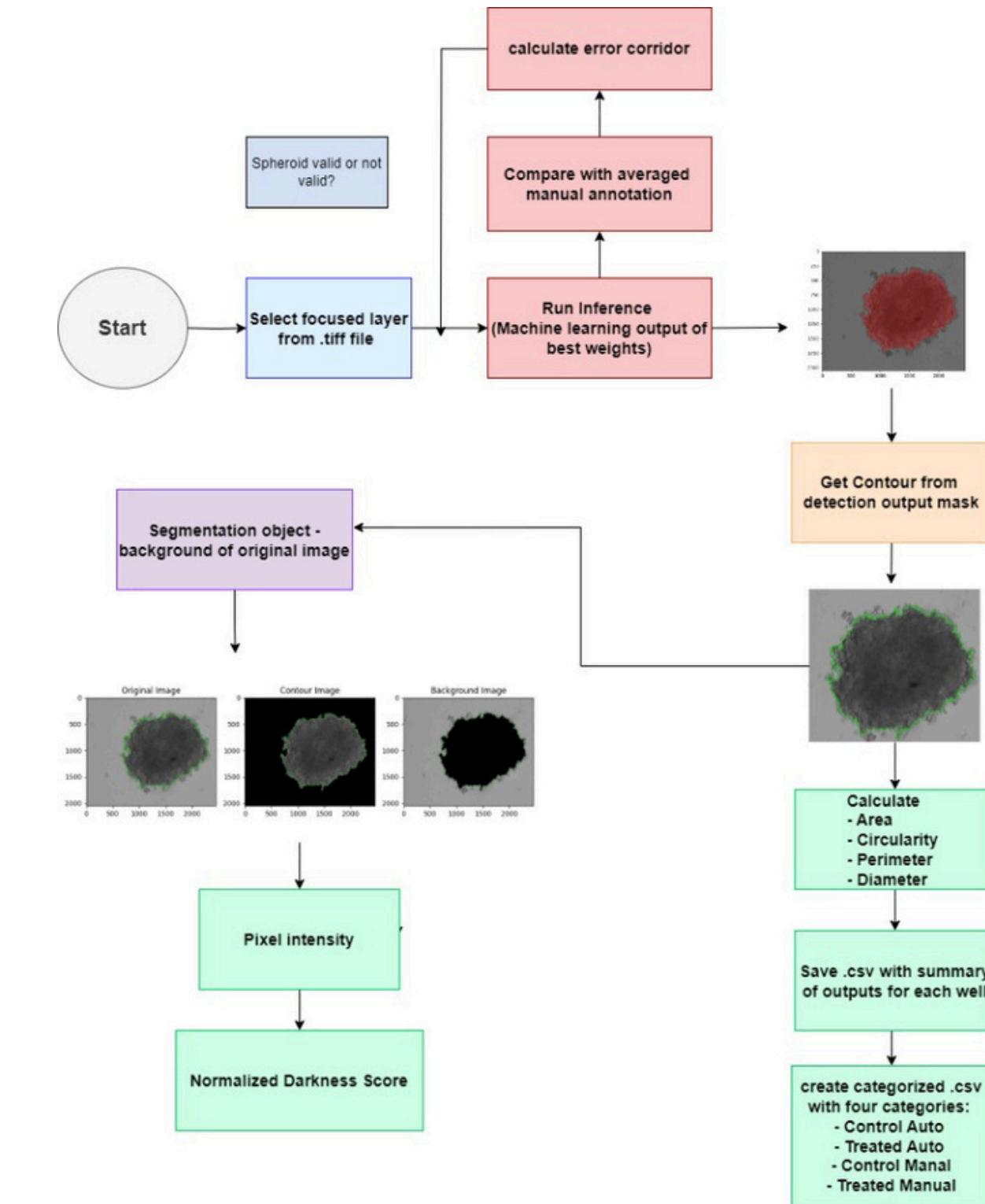
Drug screening



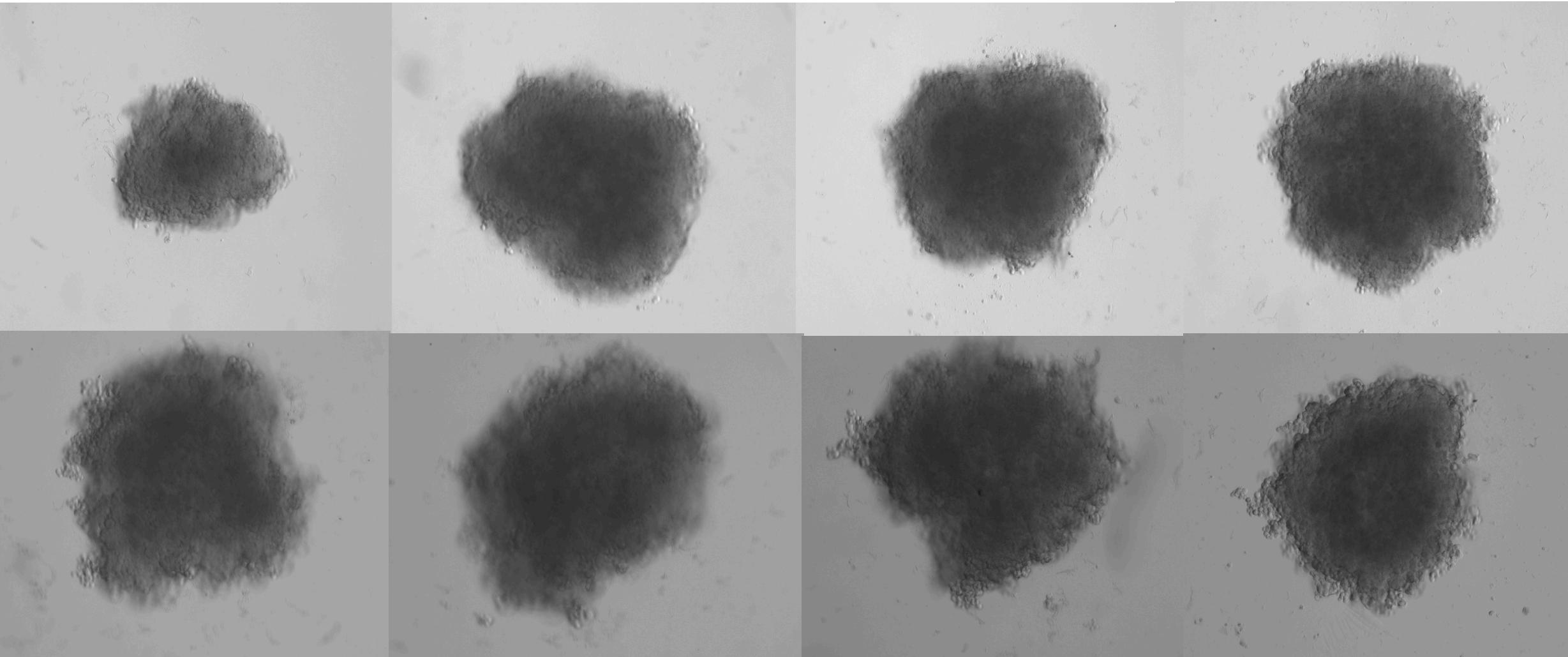
Current methods to differentiate the changes

- Size/Area
- Circularity/diameter/perimeter
- Pixel intensity change in the color

which metrics are human interpretation



Untreated



We tend to forget that this is a 2d Image of 3d spheroid.

One can assume that the cells are concentrated at the center hence thickness.

more darkness could be healed/killed but not we are not sure.

Hence human interpretation is difficult/not enough.

surprise

not human-interpretable (eg: debris amount)

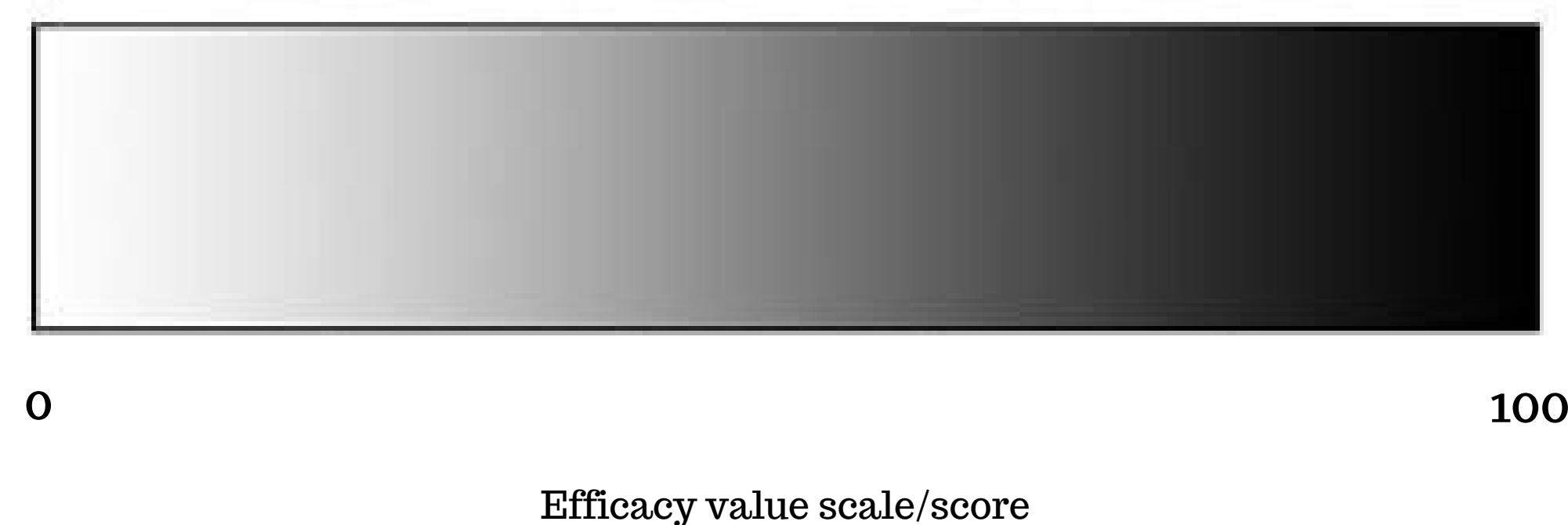
What if we can find drug efficacy using other hidden information/patterns in the time evolution of treatment with different drug combinations?

AIM

Ranking: should we give 5 star to exploder or to best drug efficiency?

A) give 5 to exploder because we don't know exactly which drug conc have best optimum

Create a value scale of drug efficacy with different drug combinations



In the end , we are not interested in just how the final day cancer cell images looks like. we are interested in how the cancer cell image evolves over time. or how much the cancer cell deteriorate from day 0 after applying drug.

Proposed Solution

Combination of

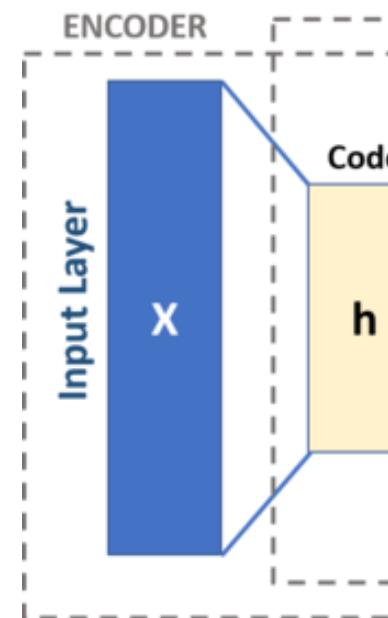
1. Contrastive learning method
2. Time series prediction model

Time series prediction model

Step 1:

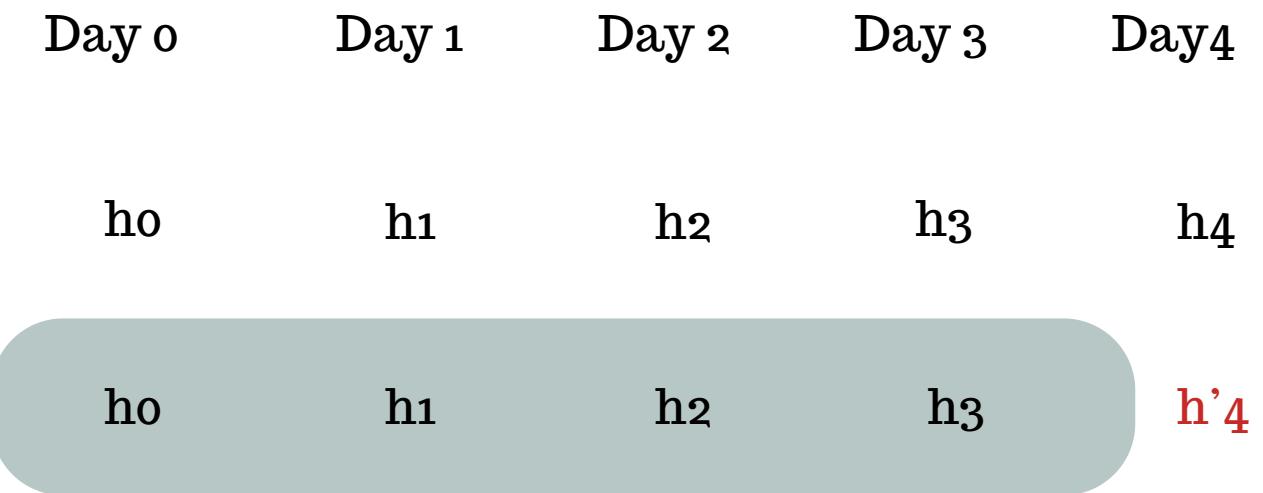
Create a latent space representation of every image by training them using Autoencoders

$$h = f(x)$$

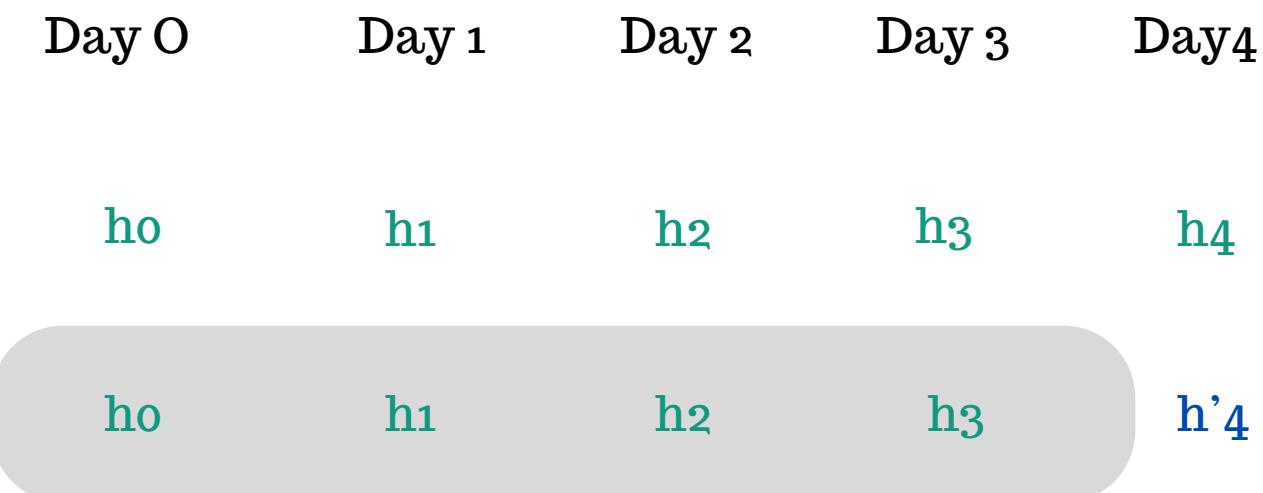


Time series prediction model (continuation)

Step 2 : Train time series prediction model only on untreated images.

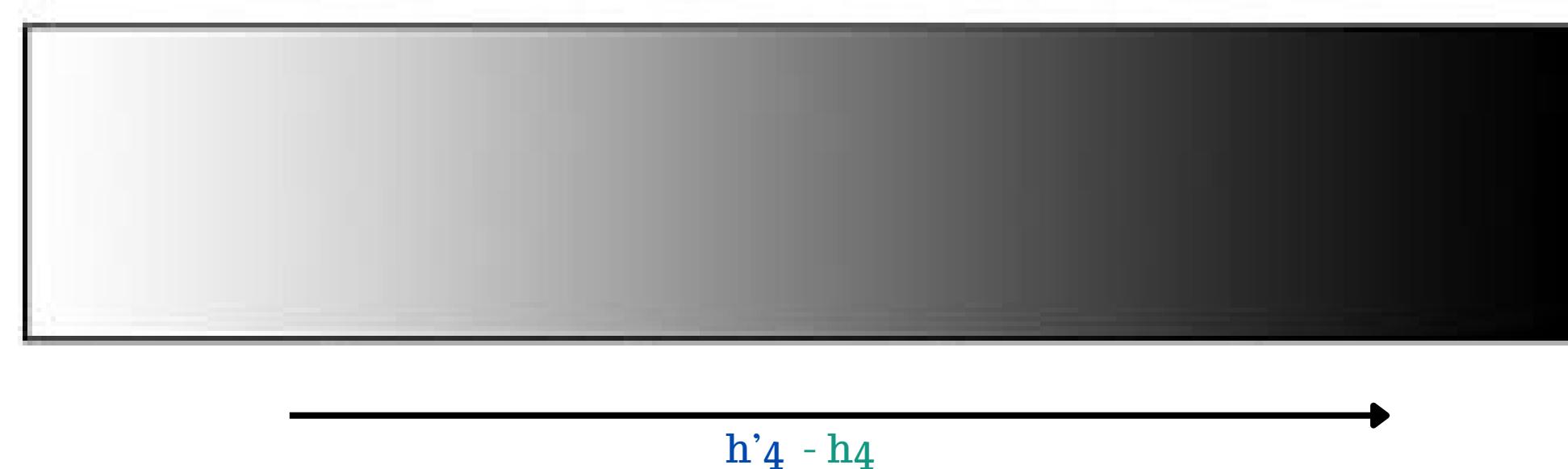


Step 3 : Inference on test images including untreated/clinically recommended/drug screening images



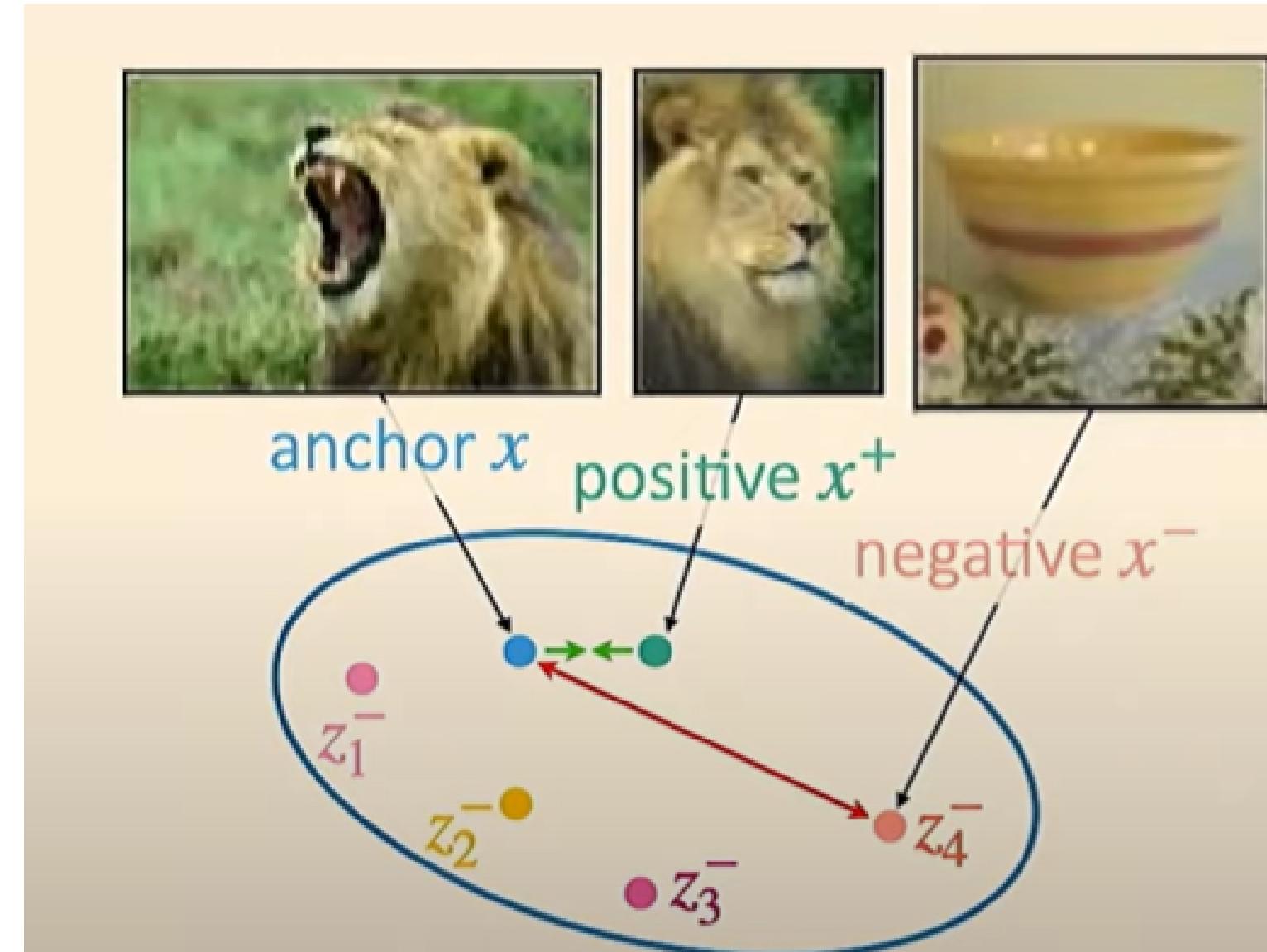
AIM

Create a value scale of drug efficiency with different drug combinations

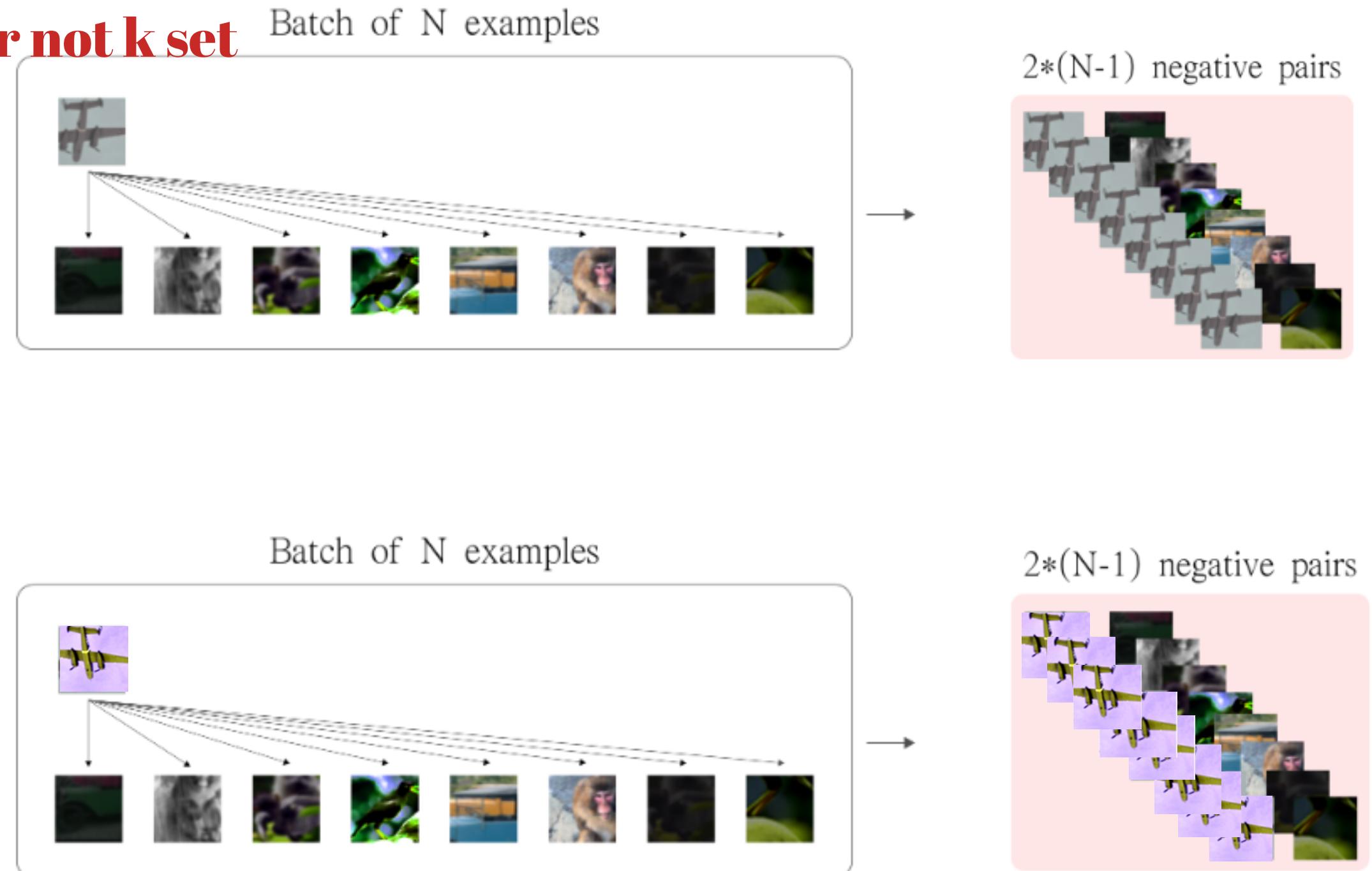
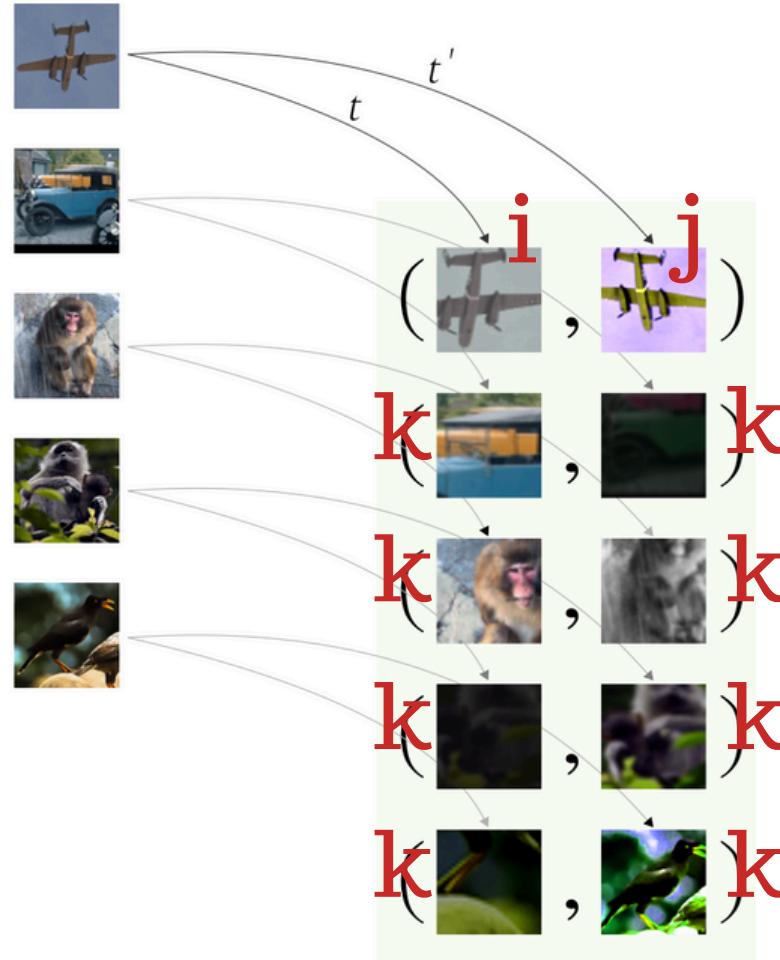


Use this as informed decision to find optimum drug efficient combination

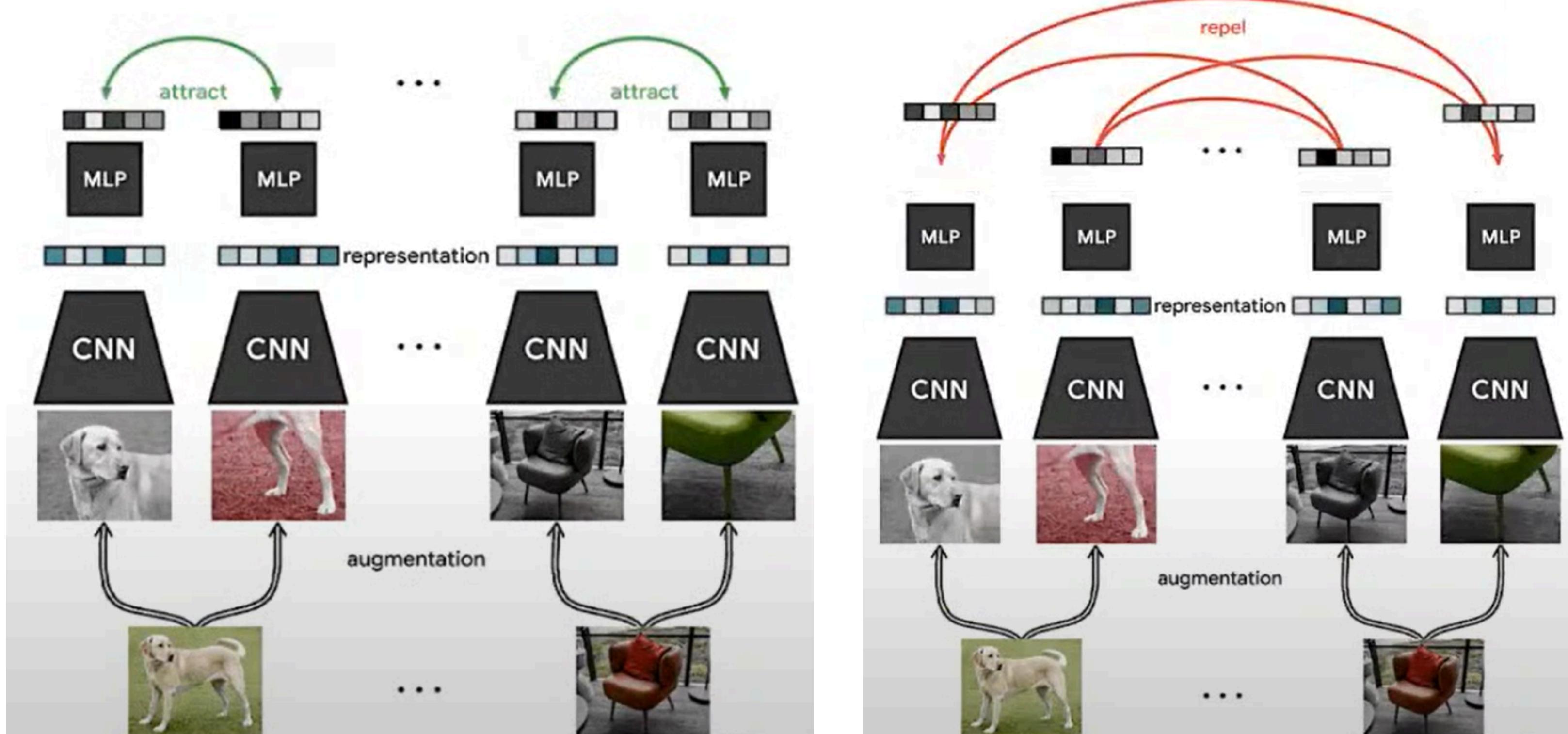
Contrastive learning



May wrong image check code implemetation for confirming j is included or not k set



SimCLR



The output features of the two augmented images are then trained to be close to each other, while all other images in that batch should be as different as possible. This way, the model has to learn to recognize the content of the image that remains unchanged under the data augmentations, such as objects which we usually care about in supervised tasks.

SimCLR

- Task: select/identify positives out of negatives
- Advantage: Doesn't restricted to one specific architecture in encoder (ie. instead of resnet50 we can use other architectures. Eg: transformers?)
- need more negative samples if we are using simCLR (in BOYL, we don't use negatives)
- We can use multi layer images/cell as data augmentation.
- no of class = total no of samples in minibatch
- increase model size, training epoch, larger batch size better classification accuracy

Small batch sizes work well too with good hparams tuning

Original SimCLR was developed with large batch size, so the hyper-params were not optimized for smaller ones in the above batch size study.

With proper tuning on learning rate, temperature, and deeper projection head, the difference in batch sizes is pretty small.

Table D.3: Linear evaluation accuracy (top-1) of ResNet-50 trained with different losses on ImageNet (with 3-layer projection head).

Loss	Epoch	100	200	400	800
		Batch size	512	66.6	68.4
NT-Xent	1024	66.8	68.9	70.1	70.9
	2048	66.8	69.1	70.4	71.3

SimCLR

- “If we don't have color then distorting intensities instead of colors” - Simon Kornblith (Author of Simclr paper)

Self-Supervised Learning for Time Series: Contrastive or Generative?

- To summarize, when working with datasets that contain a small proportion of labeled data, where SSL pre-training is crucial, we recommend employing MAE for human activity data and SimCLR for ECG data. The inconsistent results between different data types may be attributed to variations in data characteristics. The ECG patterns present in this arrhythmia dataset are relatively simple and easily distinguishable, as evidenced by several baseline experiments where even SVM achieves relatively good performance

SimCLR

ℓ_2 norm?	τ	Entropy	Contrastive acc.	Top 1
Yes	0.05	1.0	90.5	59.7
	0.1	4.5	87.8	64.4
	0.5	8.2	68.2	60.7
	1	8.3	59.1	58.0
No	10	0.5	91.7	57.2
	100	0.5	92.1	57.0

if this loss fn trying to get more accuracy in classification rather than contrastive loss (pushing and pulling) then other contrastive loss fn maybe better than simclr to do better loss minimization

SimCLR

Are we interested in Contrastive loss or linear classification accuracy?
because Without ℓ_2 normalization, the contrastive task accuracy is higher,
but the resulting representation is worse under linear evaluation as in table.

if we are interested in contrastive loss then which one should we consider with projection head
or without projection head?

Magda said since the space may be small so somehow they have to get compacted



randomly selected 10 classes by our best ResNet-50 (top-1 linear evaluation 69.3%)

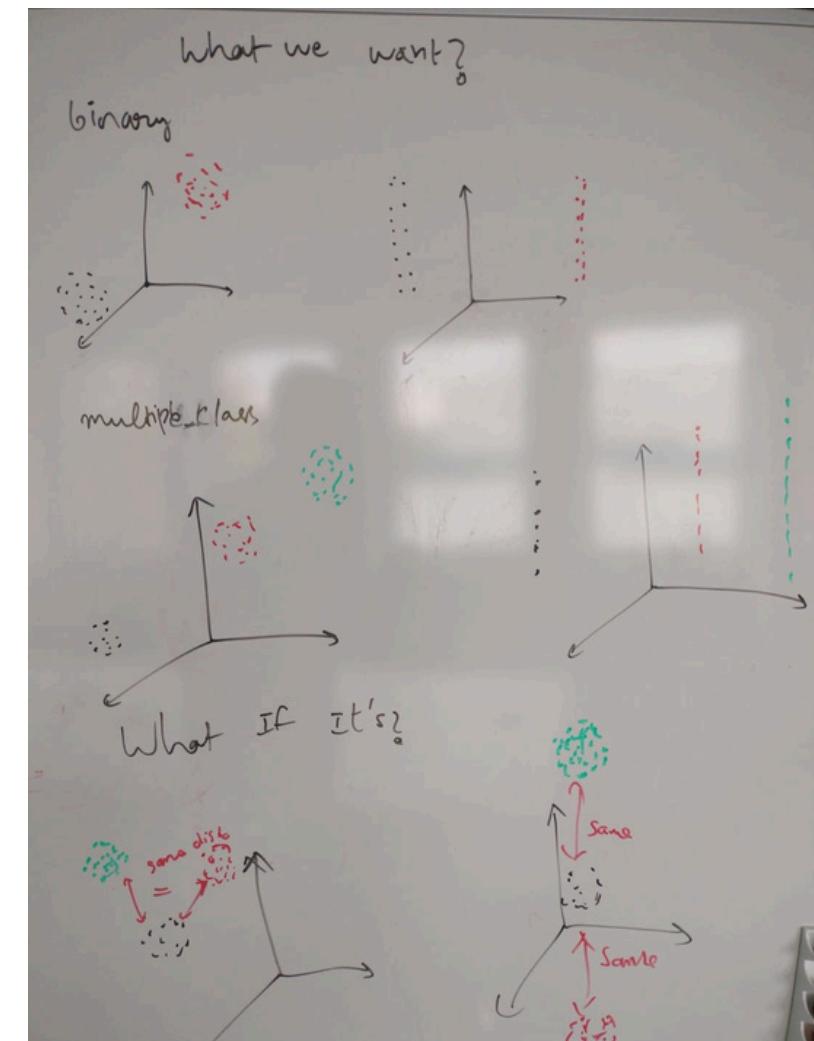
(a) h

(b) $z = g(h)$

Classes represented by h are better separated compared to z . does that mean we get representation with far way
distances untreated

ℓ_2 norm?	τ	Entropy	Contrastive acc.	Top 1
Yes	0.05	1.0	90.5	59.7
	0.1	4.5	87.8	64.4
	0.5	8.2	68.2	60.7
	1	8.3	59.1	58.0
No	10	0.5	91.7	57.2
	100	0.5	92.1	57.0

logically its better to consider contrastive loss,
since its directly related to cosine similarity ie
closeness of 2 vectors samples in the space. hence
we should consider with projection head.



SimCLR

- This paper give importance to downstream task (linear classification) accuracy where maybe our focus is in the contrastive loss/accuracy.
- Question their results for contrastive loss VS linear classifier (downstream tasks)
- Hence it's better that we calculate contrastive loss values for each setup.
- Example 1: normalisation table 5
- Example 2: “even if in the final layer we're pushing things apart we kind of figure that this earlier representation might not have pushed apart the things that really are semantically similar” - Simon Kornblith

SimCLR

Why NT XENT loss better than logistic and marginal loss?

Name	Negative loss function	Gradient w.r.t. \mathbf{u}
NT-Xent	$\mathbf{u}^T \mathbf{v}^+ / \tau - \log \sum_{\mathbf{v} \in \{\mathbf{v}^+, \mathbf{v}^-\}} \exp(\mathbf{u}^T \mathbf{v} / \tau)$	$(1 - \frac{\exp(\mathbf{u}^T \mathbf{v}^+ / \tau)}{Z(\mathbf{u})}) / \tau \mathbf{v}^+ - \sum_{\mathbf{v}^-} \frac{\exp(\mathbf{u}^T \mathbf{v}^- / \tau)}{Z(\mathbf{u})} / \tau \mathbf{v}^-$
NT-Logistic	$\log \sigma(\mathbf{u}^T \mathbf{v}^+ / \tau) + \log \sigma(-\mathbf{u}^T \mathbf{v}^- / \tau)$	$(\sigma(-\mathbf{u}^T \mathbf{v}^+ / \tau)) / \tau \mathbf{v}^+ - \sigma(\mathbf{u}^T \mathbf{v}^- / \tau) / \tau \mathbf{v}^-$
Margin Triplet	$-\max(\mathbf{u}^T \mathbf{v}^- - \mathbf{u}^T \mathbf{v}^+ + m, 0)$	$\mathbf{v}^+ - \mathbf{v}^- \text{ if } \mathbf{u}^T \mathbf{v}^+ - \mathbf{u}^T \mathbf{v}^- < m \text{ else } \mathbf{0}$

Looking at the gradient, we observe

- 1) ℓ_2 normalization (i.e. cosine similarity) along with temperature effectively weights different examples, and an appropriate temperature can help the model learn from hard negatives; and
- 2) unlike cross-entropy, other objective functions do not weigh the negatives by their relative hardness. As a result, one must apply semi-hard negative mining (Schroff et al., 2015) for these loss functions: instead of computing the gradient over all loss terms, one can compute the gradient using semi-hard negative terms (i.e., those that are within the loss margin and closest in distance, but farther than positive examples).

SimCLR

- How does the augmentations of dogs come closer in latent space representation while in the loss function they are trying to minimise the cosine similarity i. e maximise the distance between them in latent space representation?
- Need to check the implementation part of SimCLR to understand to check whether the above happening? or all dogs embeddigs are closer or not in real.
- why can't we use supervised/semi supervised sampling from the beginning instead of unsupervised sampling. we somehow modify/change loss function to pull the all positive samples Human annotating negative samples [supervised manner] to the original image and push all negative samples (Human annotating negative samples [supervised manner])
- Think about what are the constants and variants in the loss function equation to make sure that z_i , z_j , z_k vectors changes or simply what happens each step, what are the paramets updating (look at gradient descent to understand) how/or what does it effect to loss function change

SimCLR

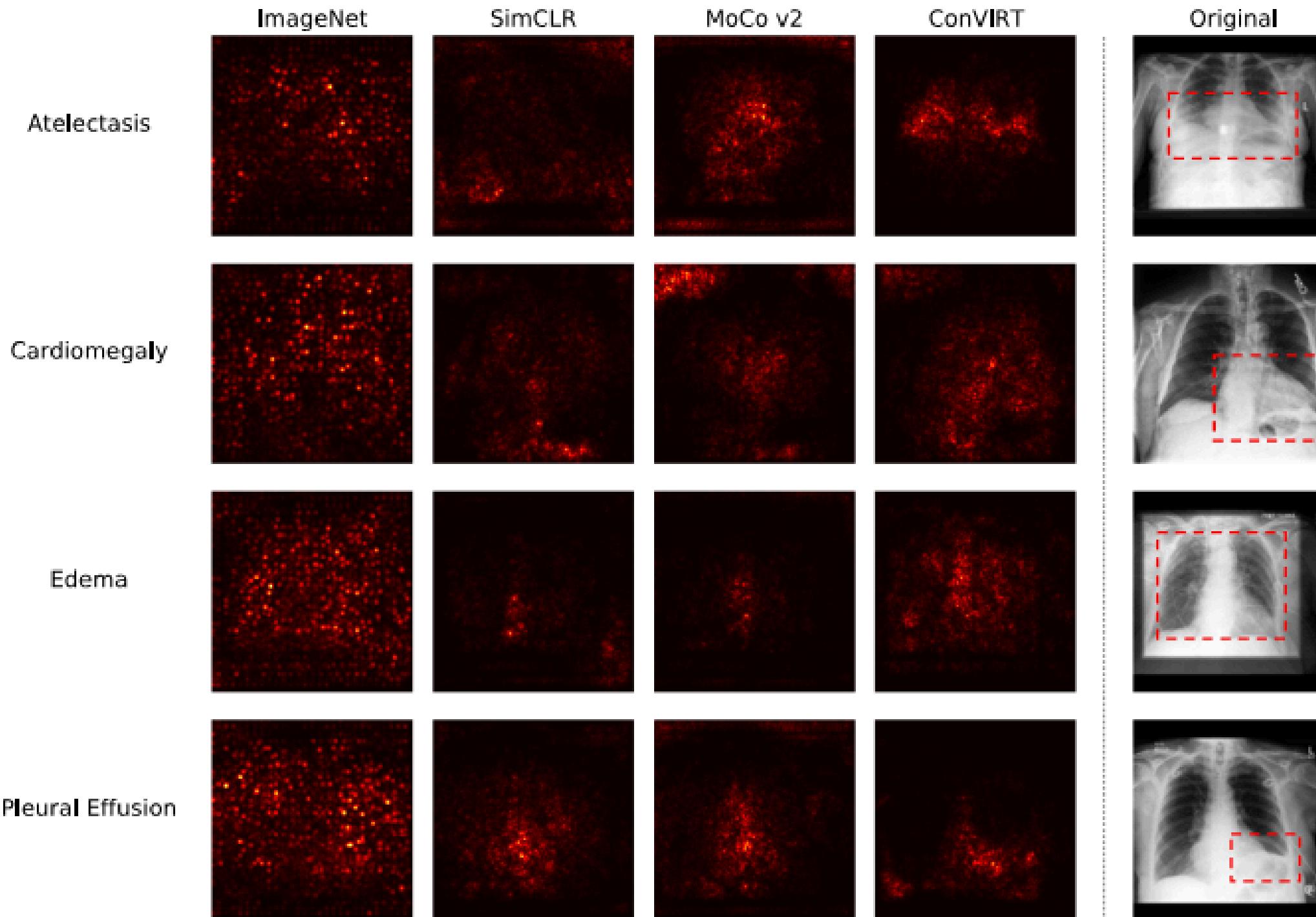
- can we make a loss function where it only pull together if 2 representations are similar and only push apart if 2 representations are dissimilar?
- we give all untreated one as positive samples. so x_i and x_j will be untreated. x_k will be treated one.
- in order to decouple the negatives and batch size for the above solution we can use Moco method maybe

SimCLR

- What does that mean by semi supervised/ supervised simclr ? does that mean instead of random positive and negative sampling are we giving what are positives and negatives or is it supervised/semi supervised after encoding part ie after taking h representaiton then used it for classification?

SimCLR

when the dataset is large in imagenet hence only few positives are pushing apart.



Contrastive Learning of Medical Visual Representations from Paired Images and Text
<https://arxiv.org/pdf/2010.00747.pdf>

Idea: Transfer Convirt medical image model weights for our problem.

pretraining encoders with the paired text data via a bidirectional contrastive objective between the two modalities (text and image)

in our case we only have text for 0, 100 which are untreated and exploded correspondingly. BUT IF GET GROUND TRUTH by lucifer then we have text for each image case. (Until then we have no use)

architecnture is same as simclr , do some transformation get h apply nonlinear projection head then get final g.

Supervised labeling (SupCon)

- we can separate positives as untreated + their augmentations and all others negatives + their augmentations
- In that case, while doing negatives, we should treat everything without grouping.
- question is should we separate all treated into multiple classes? maybe clinically recommended + drug screening?
- But what if treated but still look similar to untreated because of no effect of drug so we want them to attract that's where supervised breaks. or untreated cancer cell A looks similar to treated Cancer cell B (after days of treatment)
- that's where something in the loss function like below idea works better:
- Is there anyway in loss function where we can push depend on the contrastive value (because at the moment it pushes every negatives equally in latent space)
- if it's more similar to the positive then push little if its more non similar push more (involve some parameter to do this)
A) the above thought is stupid as dumb because even though the encoding to both augmentations of the same images, cosine similarity maybe not close because of the neural network transformation. that's why we need to train architecture to give similar image cosine similarity close.
 - so first start with self supervised then go to 2 class (untreated + treated) then go to more classes)

SimCLR

SimCLR

SimCLR

SimCLR

Further readings

Solutions to requiring a large batch size

1. Store representations from previous batches (“momentum contrast”)

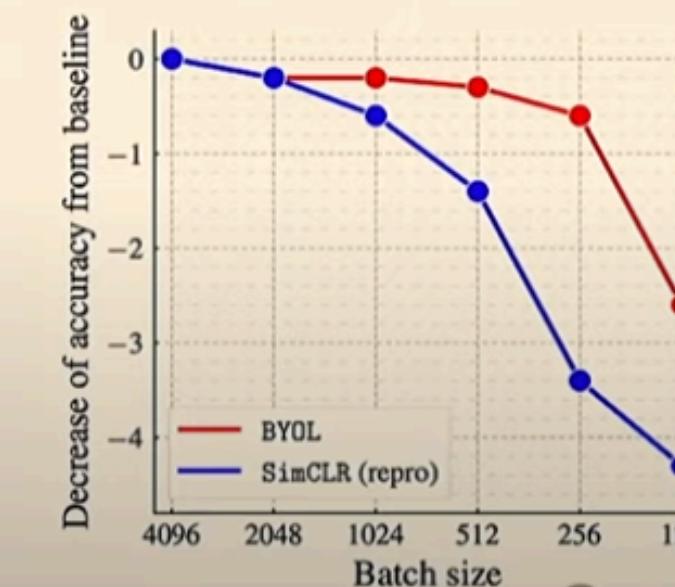
He, Fan, Wu, Xie, Girshick. MoCo. CVPR 2020

- Good results with mini batch size of 256

2. Predict representation of same image under different augmentation (“BYOL”)

Grill*, Strub*, Altché*, Tallec* Richemond*, et al. BYOL. NeurIPS 2020

- No negatives required!
- More resilient to batch size



1) use memory bank for getting larger batch size

2nd paper Give the intuition not learning from negative samples but only from positive samples then do classification
(similar to autoencoder time series prediction model where we learn NN for untreated and use that model to differentiate treated images)

Further readings

Contrastive learning beyond augmentations

We don't have good engineered augmentations for many applications!

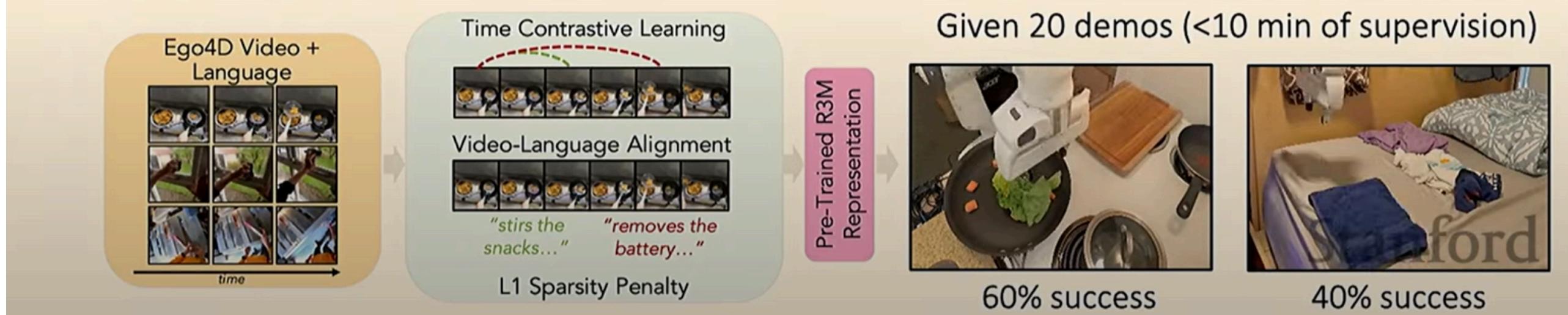
1. *Learn* the augmentations in adversarial manner (but perturbations bounded to ℓ_1 sphere)

Tamkin, Wu, Goodman. Viewmaker Networks. ICLR 2021

- > competitive with SimCLR on image data
- > good results on speech & sensor data

2. *Time-contrastive learning* on *videos* effective for robotics pre-training

Nair, Rajeswaran, Kumar, Finn, Gupta. R3M. CoRL 2022.



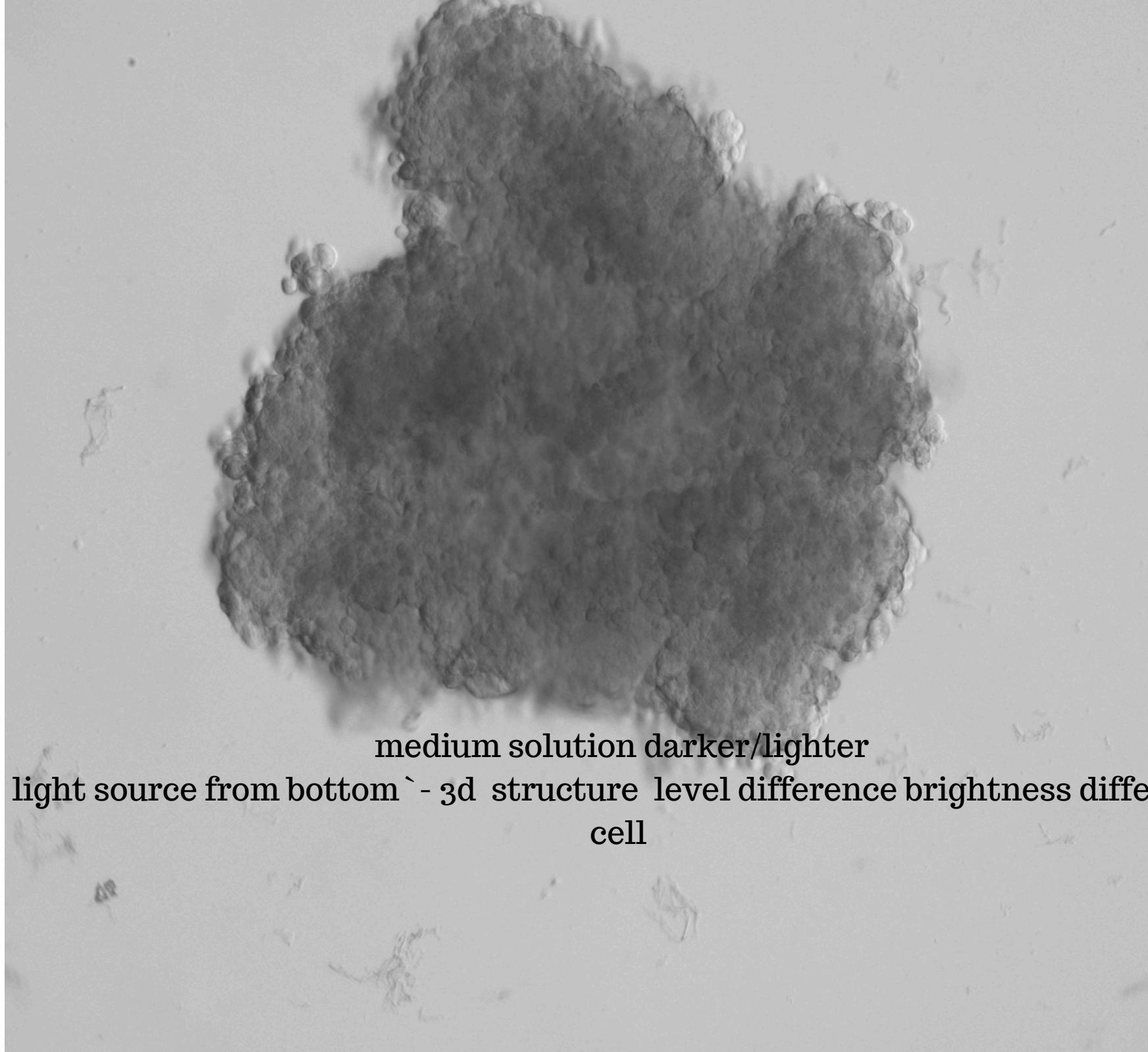
Further readings

- SimCLR V2 : Big Self-Supervised Models are Strong Semi-Supervised Learners
- DINO (Self supervised + Transformer) : Emerging Properties in Self-Supervised Vision Transformers
- Supervised Contrastive Learning 2004 : if we consider giving labels as drug 0%, drug A 20%, drug A+B 30,40% . then try to train a latent space representation from where we find the distance. <https://arxiv.org/abs/2004.11362>
-

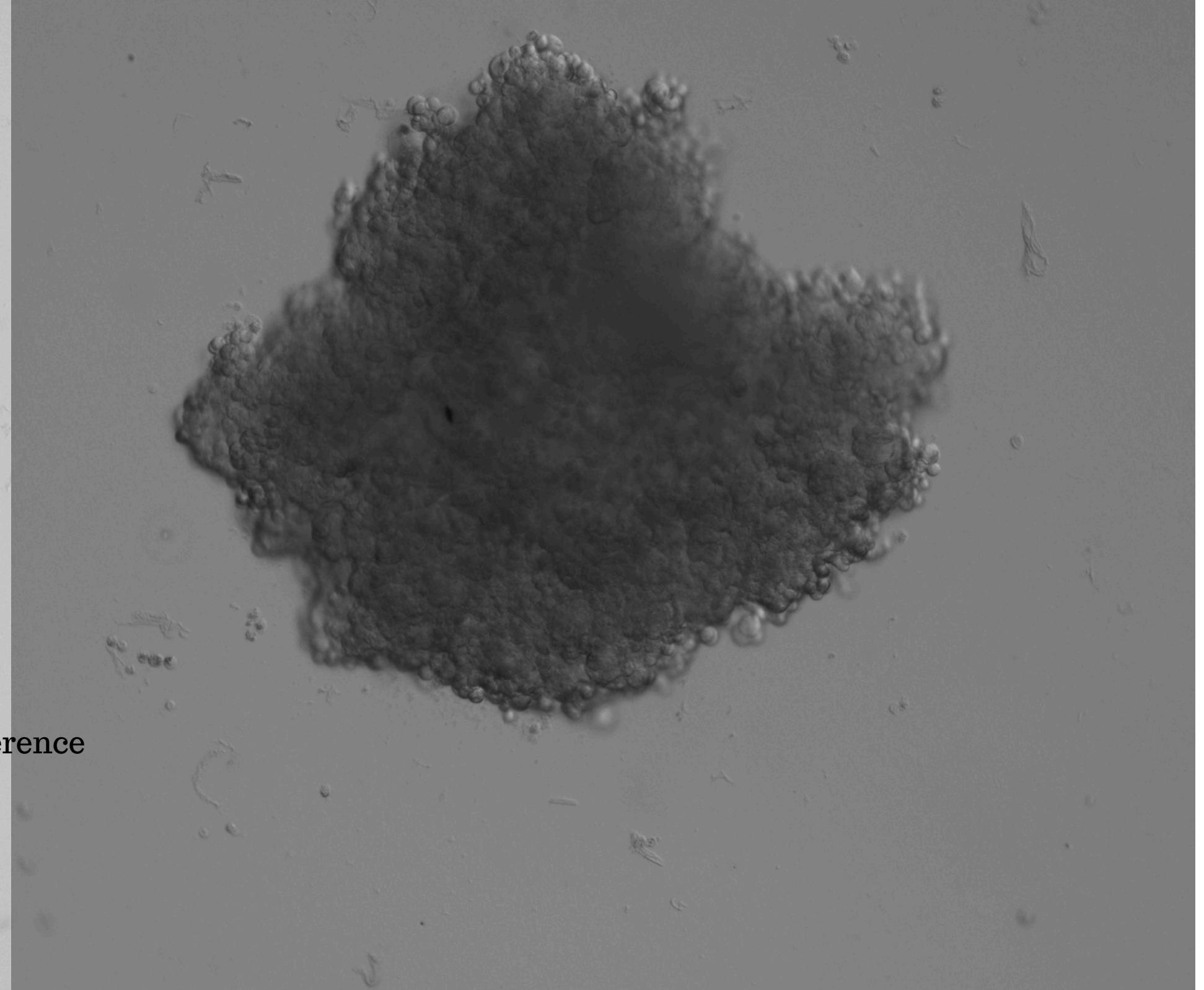
List of open questions

Brightness change + outer cell object changes within same setup

Should we remove the surroundings to make surrounding unique?



medium solution darker/lighter
light source from bottom ^ - 3d structure level difference brightness difference
cell



How to evaluate the value scale ?

we can use something similar to linear evaluation by simclr guys upon our encoder. groundtruth as fluorescent + lucifer solution)

1. we will set up fluorescent dye solution (not a good approach because of some kind of bias) or/+ lucifer method proposed by Philipp.

Question) If we can obtain ground truth efficacy from the 'Lucifer' method suggested by Philipp, then why do we need to perform AI with brightfield images?

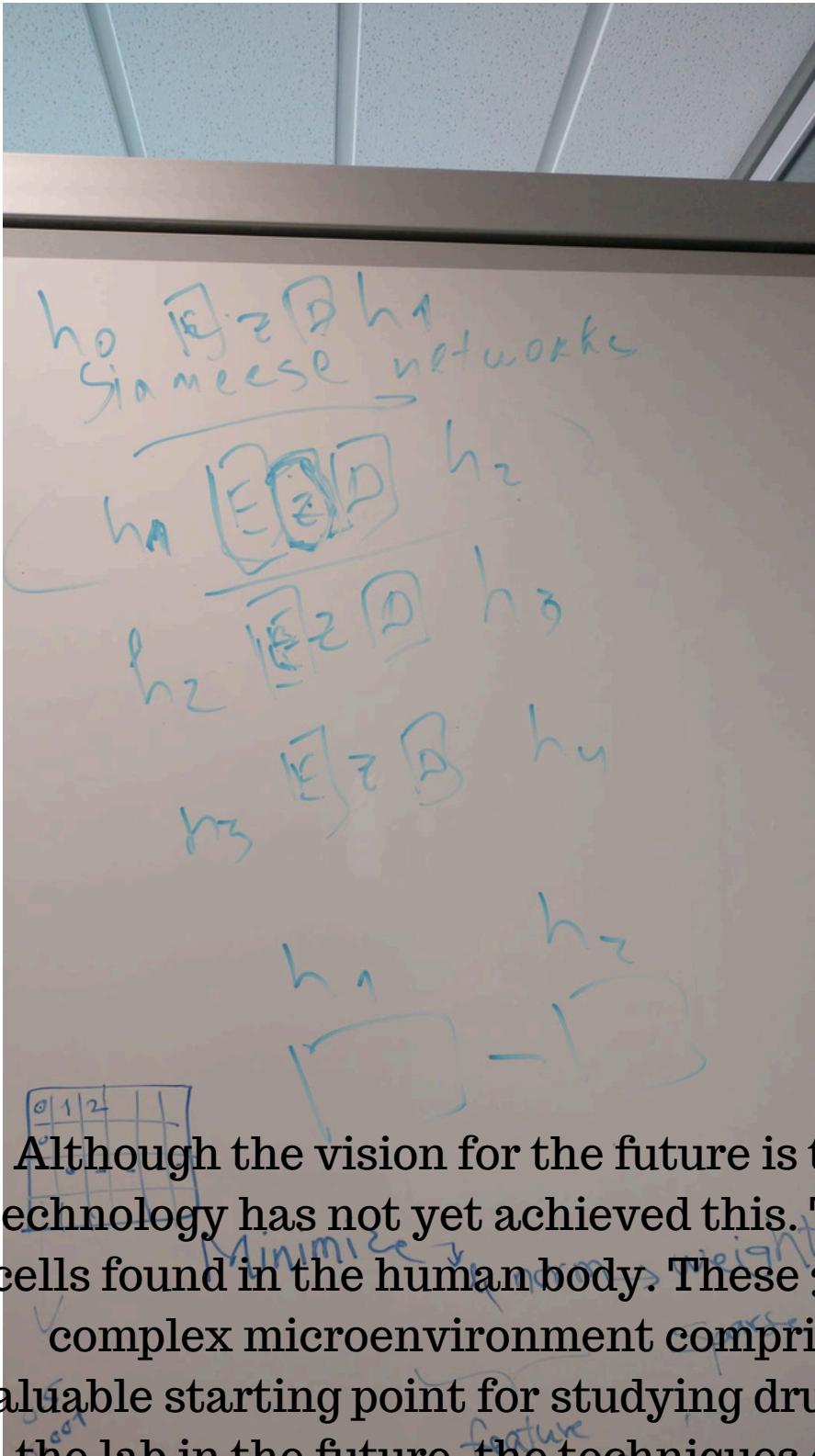
A) "the reason to use the microscopic images for the assessment of the spheroids is that it is noninvasive. We can still keep the culture going and for example make time series to also study the dynamics. The assay Philipp mentioned is invasive and results in a dead culture."

Fluorescent problem: we can't see inside/ center part what happened whether its green or red?

lucifer only gives amount of live cells at the end what is the use if we don't know the original amount at day 0.
same with fluorescent dye solution if can't have day 0 then it doesn't make sense like above reason.

Maybe integrated data driven information to doctor, ie size, color, our value scale + fluorescent + lucifer

Future work / Open questions/ out of scope of this master thesis



Learn how fast the drug efficacy by analysing day 8 9 10
evaluation of the ranking by luminar assay or fluroscent data

Can we learn next cycle prediction of drug effects by using time series to generate (generative contrastive learning MAE) day 30 for next chemotherapy situation and maybe based on that change the concentration of drug to 20%?
siameese networks for contrasting but this need labeling from stanford prof lecture.

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 cancer tissues, whereas real human pancreatic cancer cells exist within a complex microenvironment comprising cancer 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 identically 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.

TO fill up the 80 pages

experimental scheme 1:

On day 7, the drug treatment is applied to half of these, while the other half is left untreated to used as benchmark. This setup enables us to conduct a comparative study between the untreated 3D tumor tissue models and those treated with the specific drug combination to assess it's efficacy. However, live 3D tumor tissue models tend to deteriorate after a few days due to nutrient depletion in the well plate medium.

experimental scheme 2:

drug screening not half half just look at dalia image

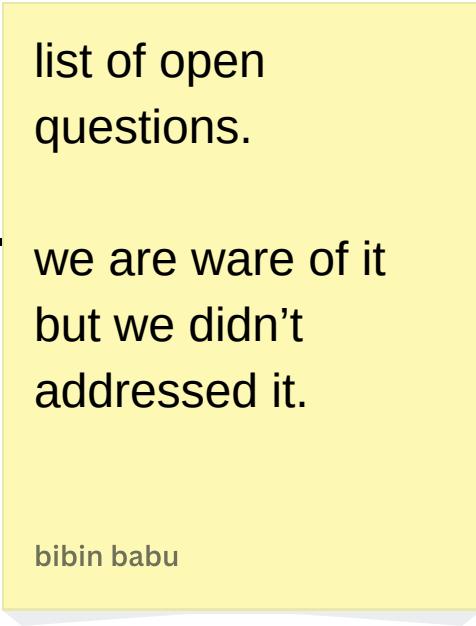
Thank
you



contrastive learning

Representation learning

Recunstruction



list of open
questions.

we are ware of it
but we didn't
addressed it.

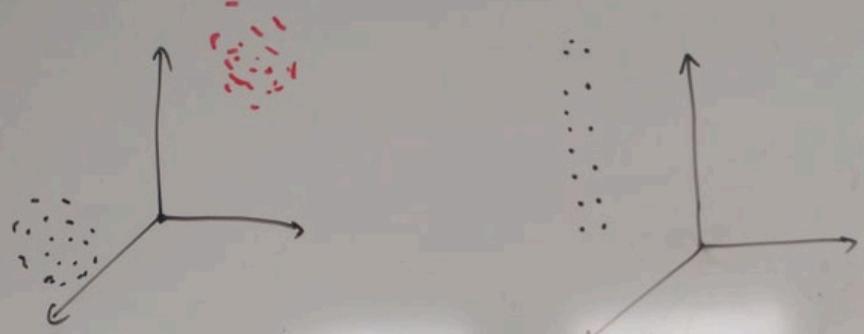
bibin babu

AIM

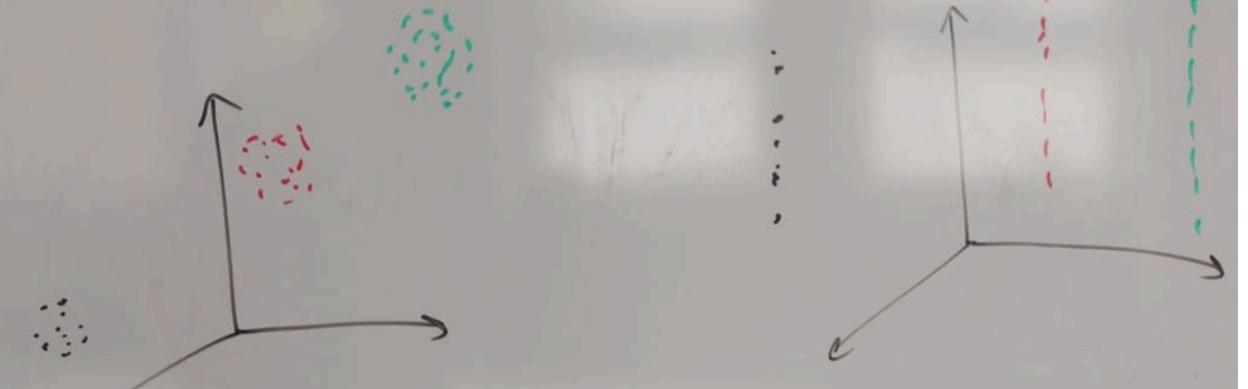
Explore any other hidden information from the images to make a informed decision

What we want?

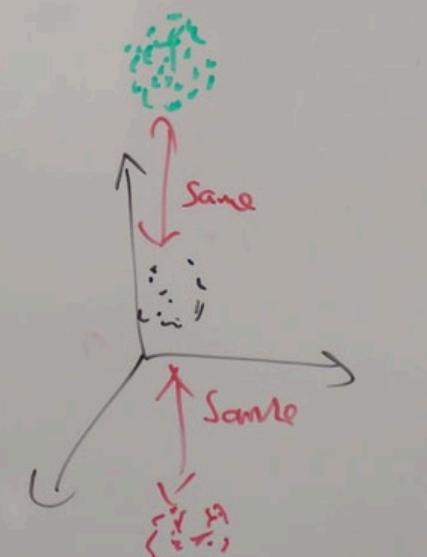
binary



multiple classes

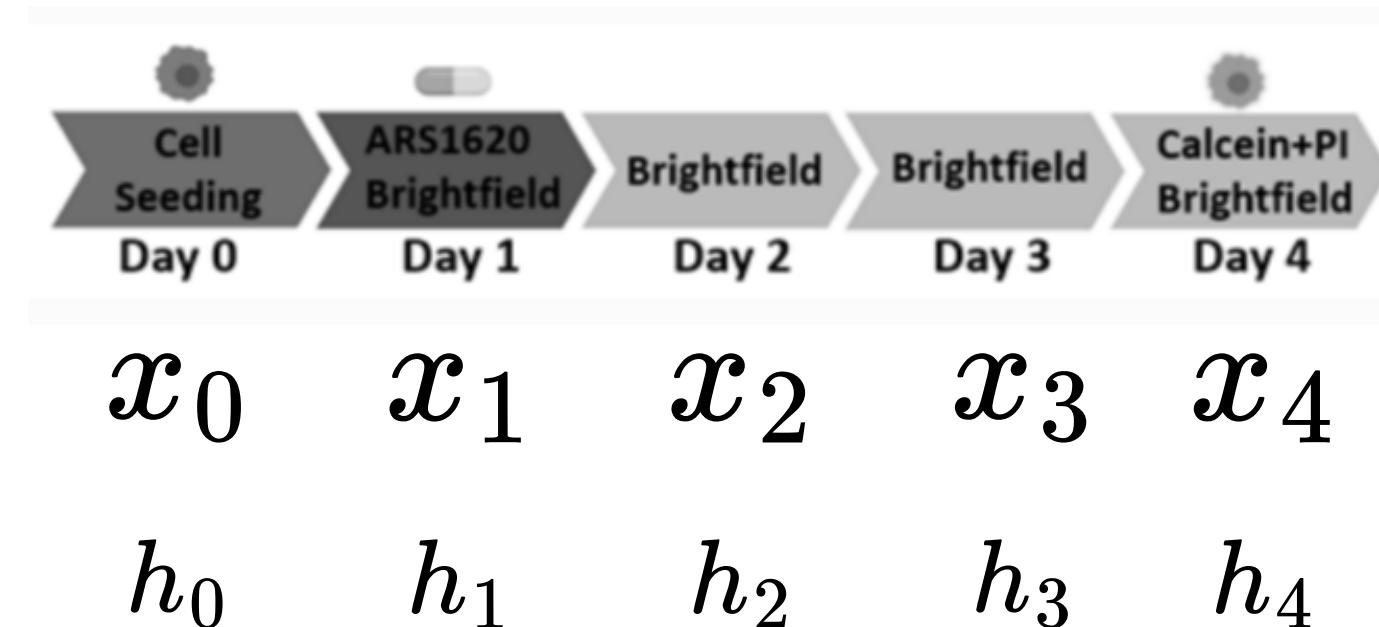


What if it's?



Proposed solution

Step 2: Train time series prediction model only on untreated images.

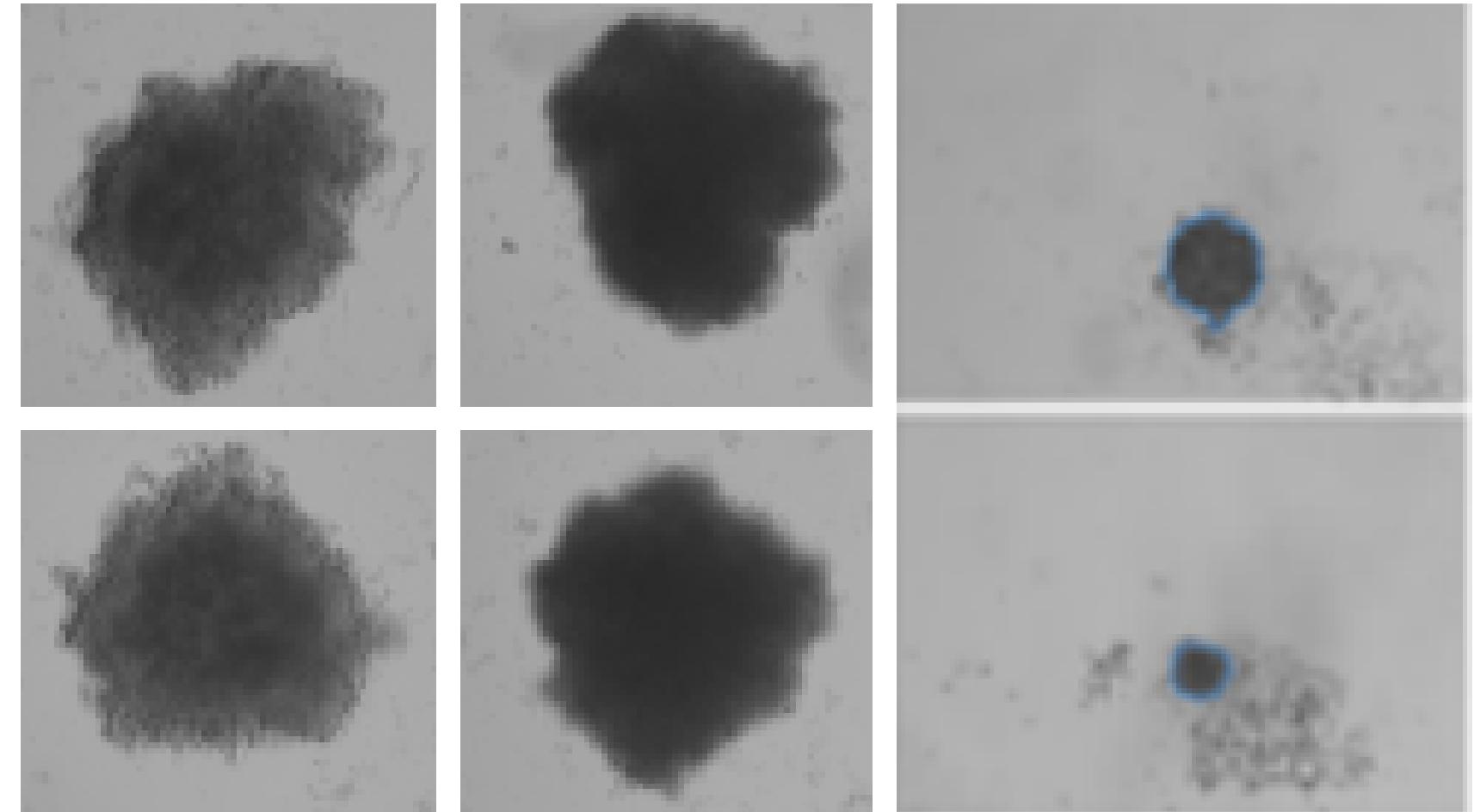


Step 3: Inference on test images including untreated/clinically recommended/drug screening images

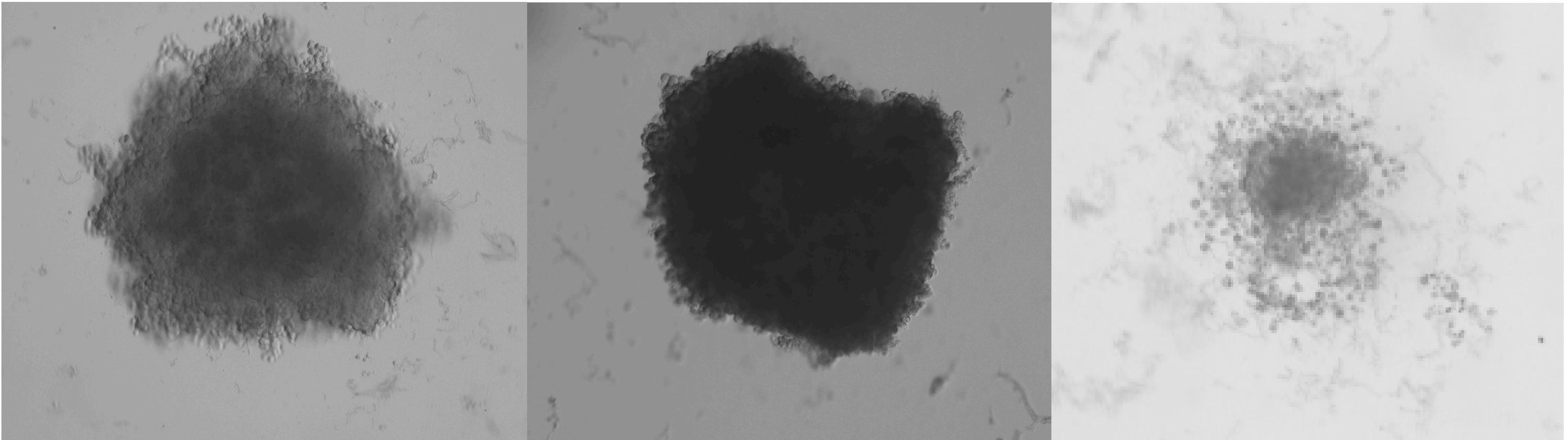


Current using method to differentiate the changes

- Size/Area
- Circularity/diameter/perimeter
- Pixel intensity change in the color



Treated



Untreated

