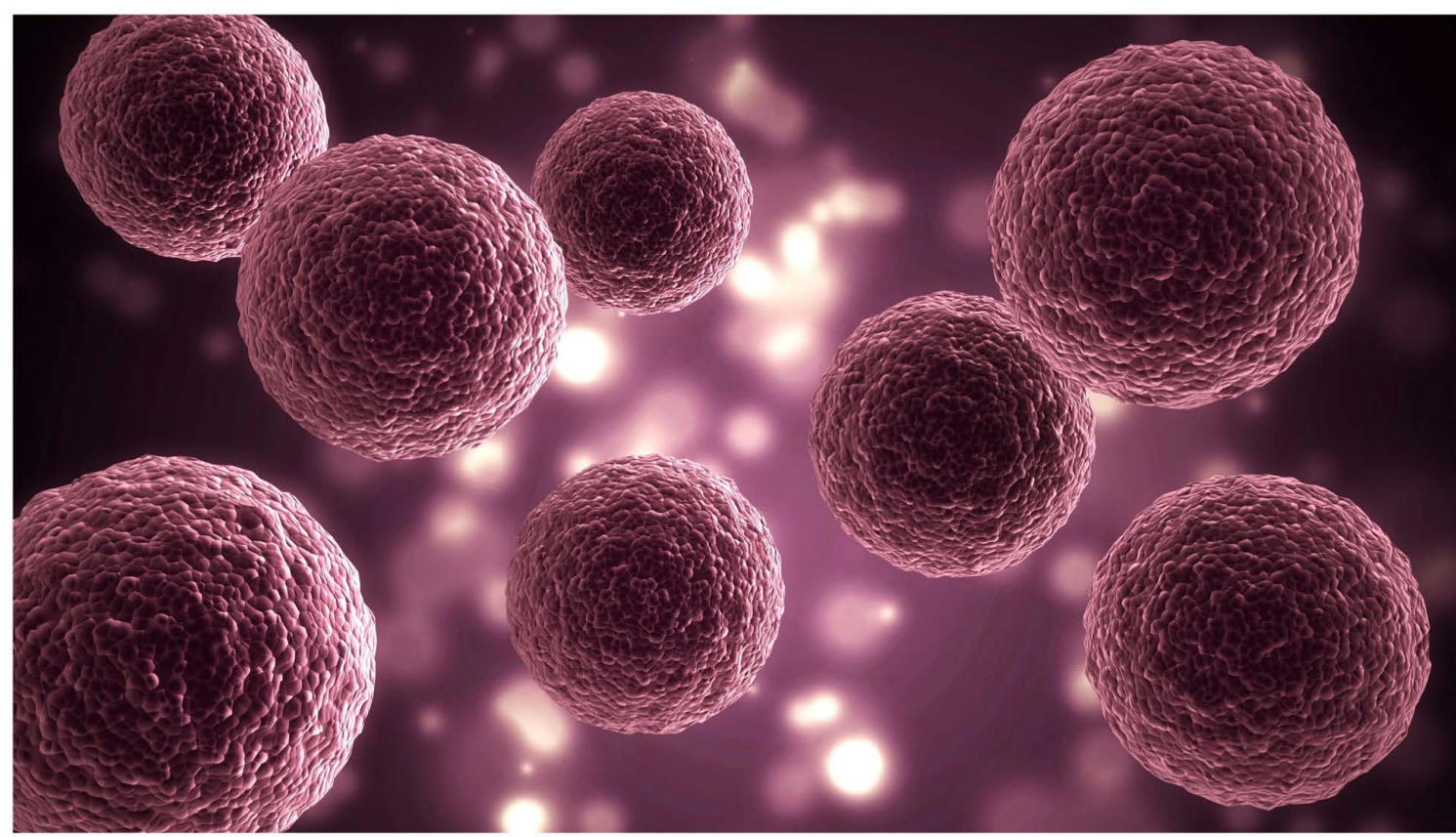


DeepFake for Life Sciences

Francesca Luongo, Evangelos Ntavelis, Iason Kastanis, Philipp A. E. Schmid
 CSEM SA, Switzerland

Spheroids are three-dimensional cellular aggregates and one of the most common and versatile way to culture cells in 3D. In order to scale laboratory tests, automated processes are needed, including robust classification. The spheroids need to be sorted with a high accuracy as misclassification can pollute the entire batch of healthy spheroids. The automation often uses deep learning algorithms to sort healthy and unhealthy spheroids. Nevertheless, the spheroids grown under normal conditions exhibit a high-class imbalance. It is estimated that only 1-10% of unhealthy spheroids are contained in one cell culture batch. This imbalance can bias the classification algorithm. Classical data warping augmentations and weighted loss are state-of-the-art methods when dealing with class imbalance. However, they are limited to the extent they can increase the performance of the classifier. Thus, the underrepresented class is oversampled with synthetic images generated with Generative Adversarial Networks (GAN). The results show significant improvement of the classification performance of an imbalanced dataset with this novel method of data augmentation.

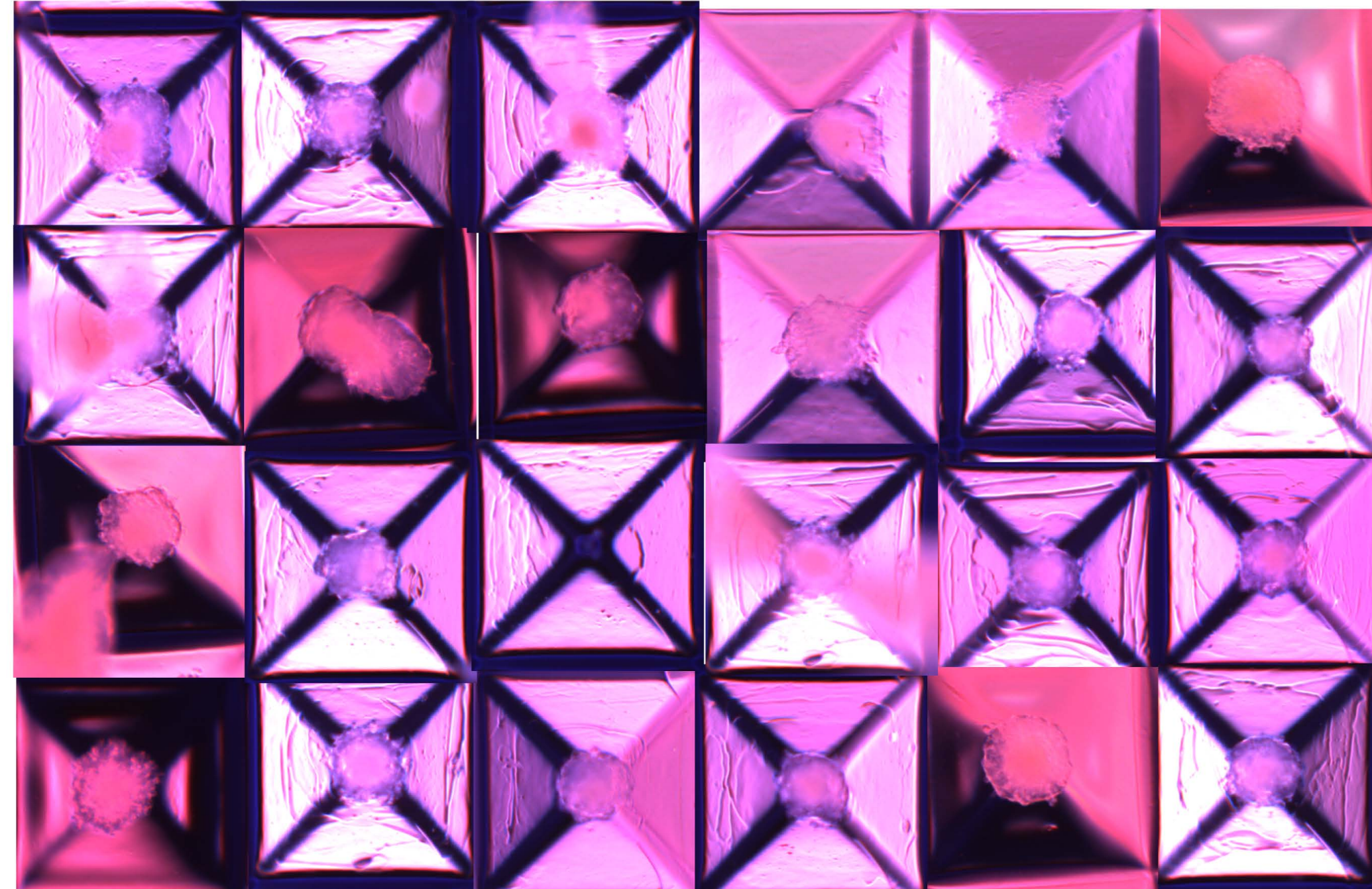
Automated Spheroid Classification



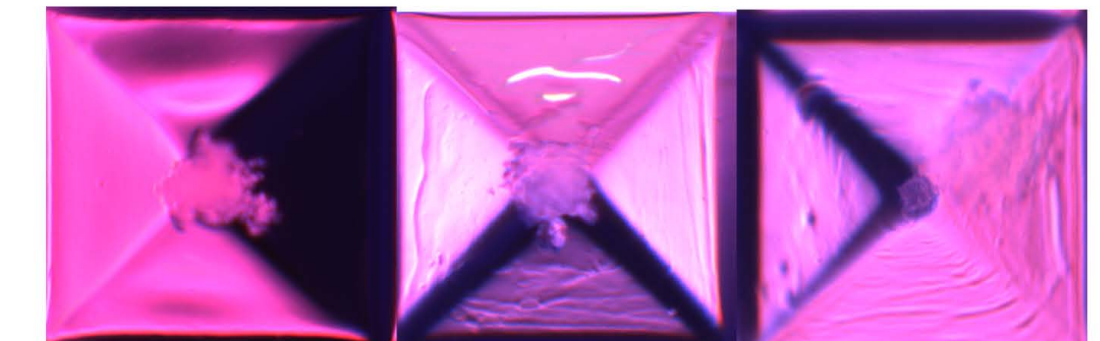
- Spheroids are models for organoids, disease and cancer:

+ 3D cellular culture closer to in vivo tissues

- speed is crucial for healthy spheroids viability



• Healthy spheroids



• Unhealthy spheroids

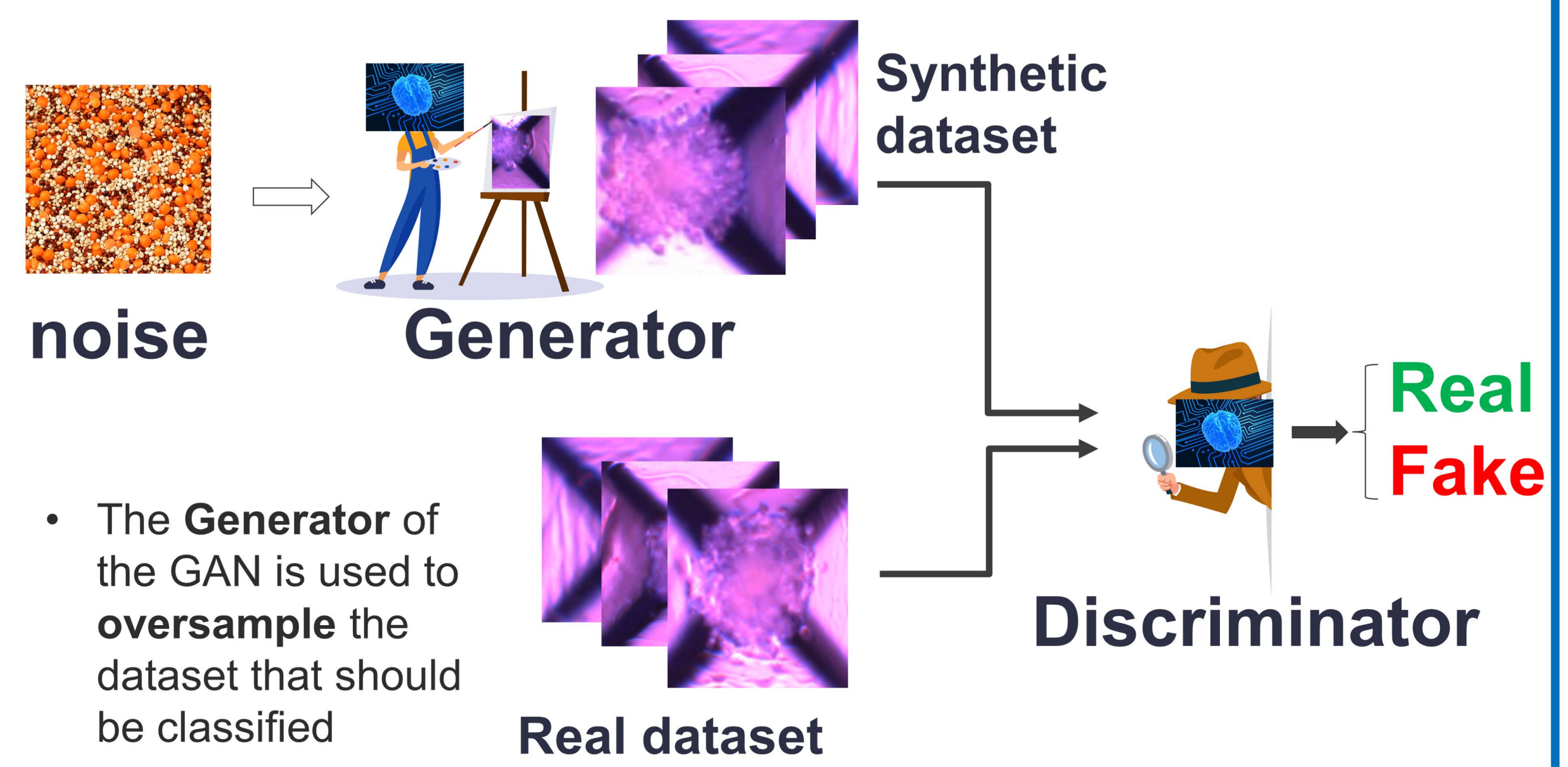
- **Biological datasets** are often **small and unbalanced** because of rarer phenotypes
- There are less occurrences of unhealthy spheroids, thus **the classification** can be **biased** towards the more represented class
- As nature is cruel, **variation** of healthy classes is tremendous

Deep Learning



- **Data is key** and major contributor to project's success
- However, **data acquisition and annotation** is very costly

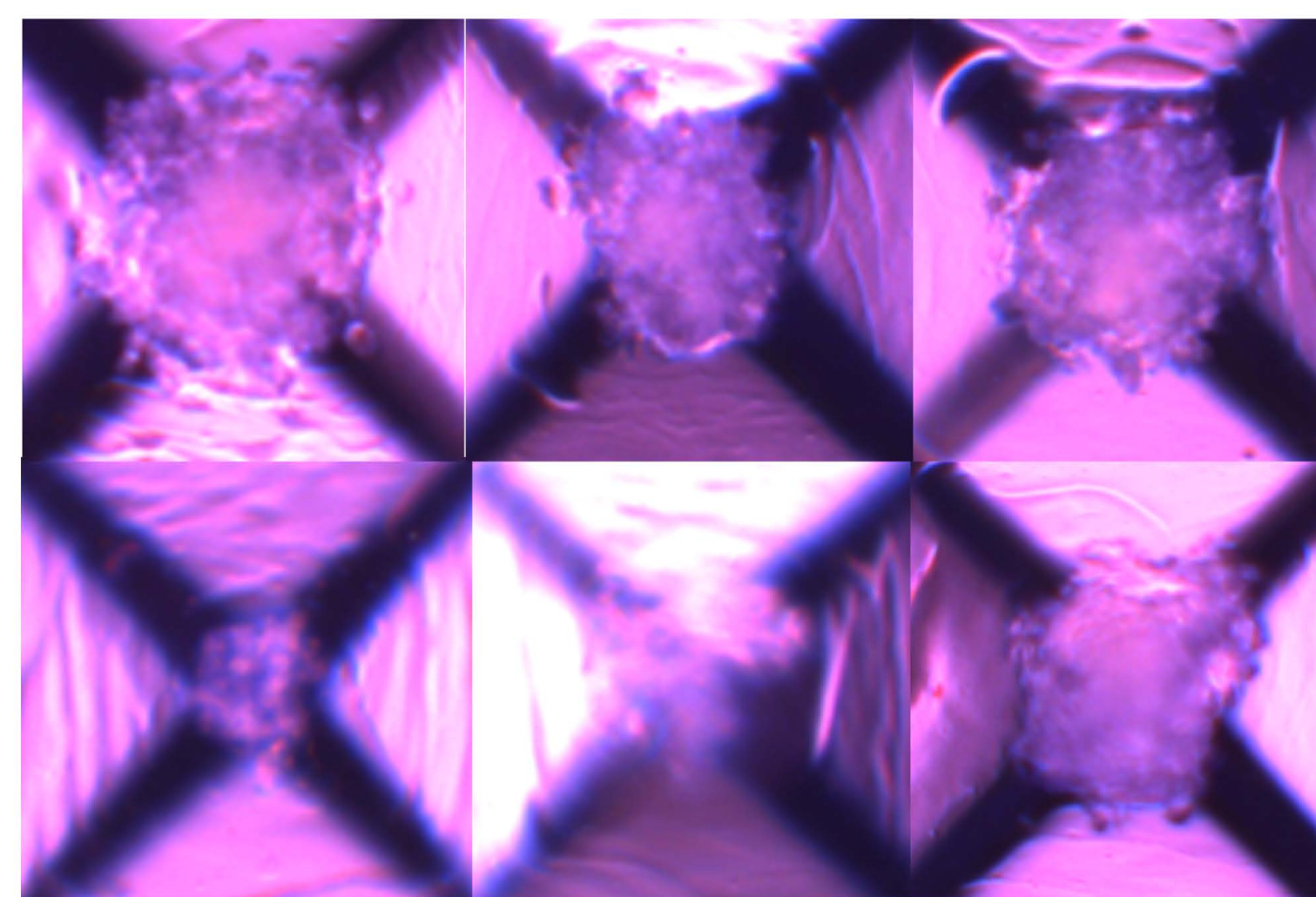
Generative Adversarial Networks



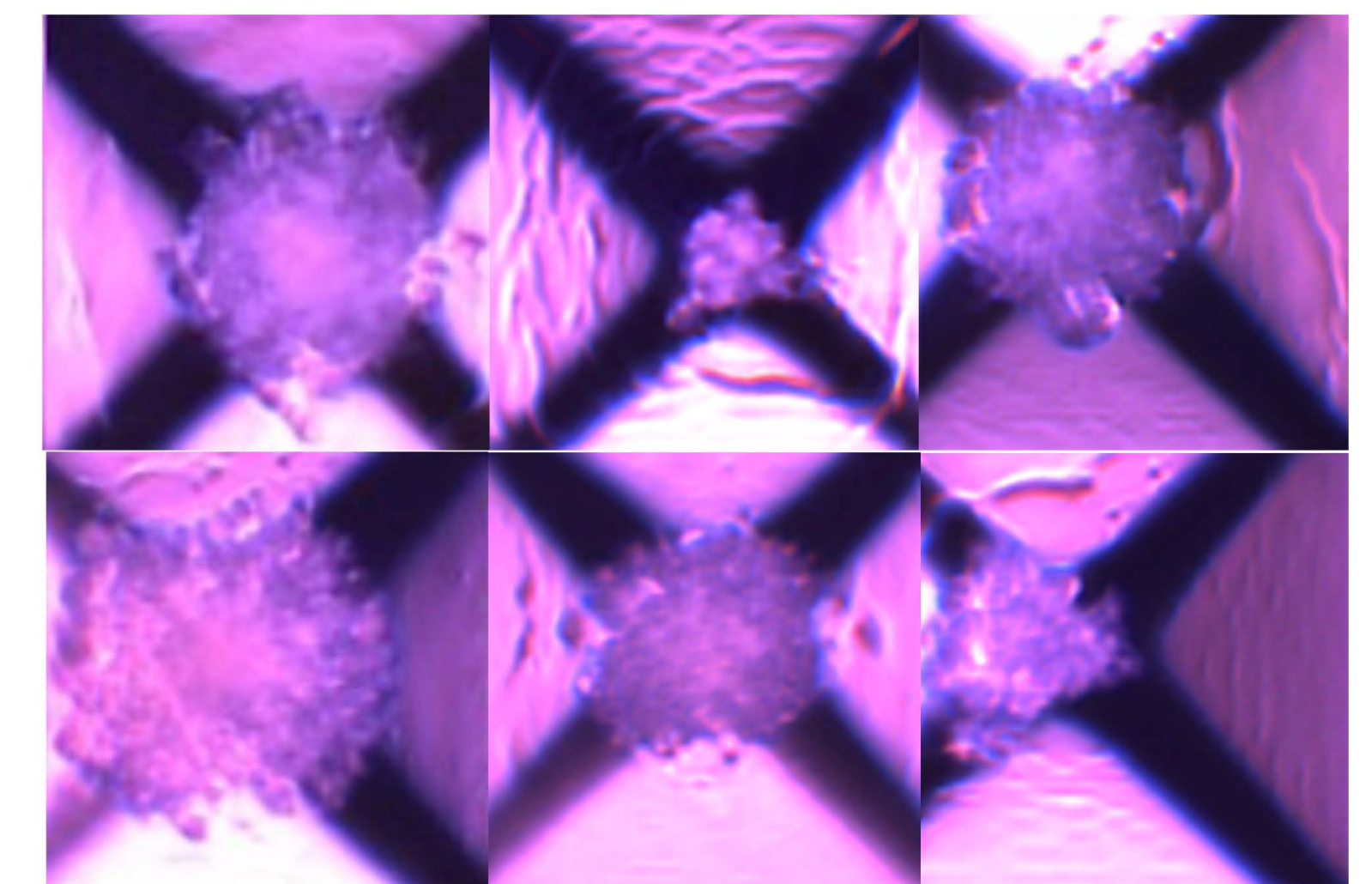
- The **Generator** of the GAN is used to **oversample** the dataset that should be classified

Generative Data Augmentation

- The **synthetic GANs generated images** resemble the real dataset distribution. Experts **couldn't distinguish** between **real/fake spheroids**
- The **classification accuracy** significantly **increased** with the input of **synthetic** unhealthy spheroids
- There are a lot of **limitations** to include **deep fakes** in the **medical field**, but they could be a potential **solution** to small and unbalance data problems



• Real Spheroids



• Synthetic GANs Spheroids