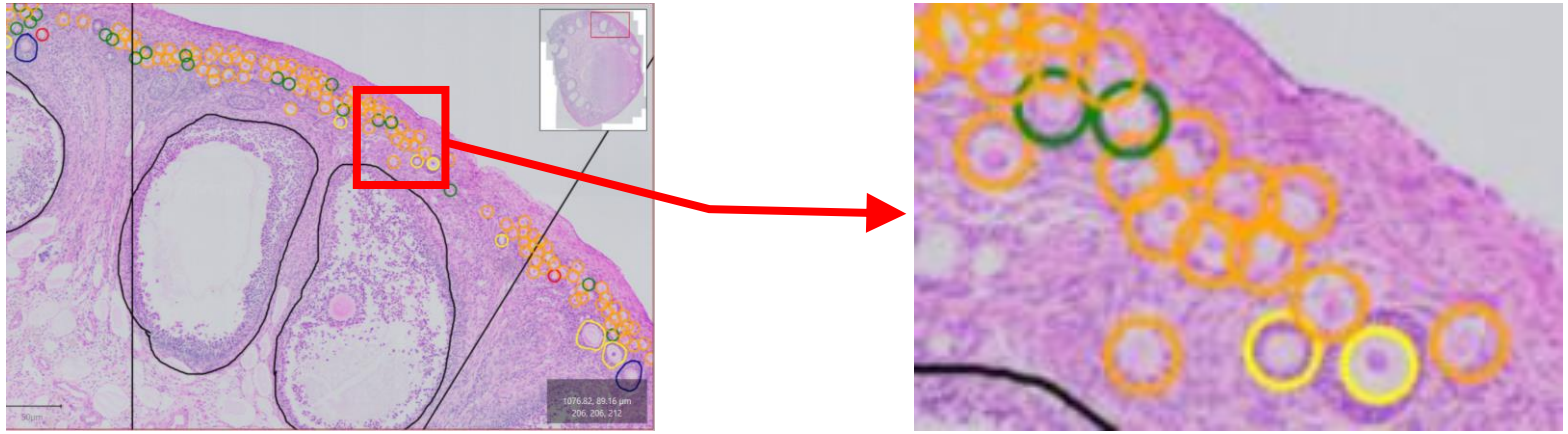
# *Machine learning for the classification of ovarian follicles from histology images*:

Dr. James Sluka, Akash Bhapkar, Param Nagda, Parth Rao

**Goals:** Part of the MOTHER DB* project is to use its ovary image repository to develop machine learning and modeling applications that can segment developing follicles in images into the appropriate developmental stage. This information is used in studies of normal ovarian function as well as in studies of reductive problems, disease and toxicity studies.
A follicle is a cluster of cell centered around a growing oocyte (egg) cell.

Primordial → Transitional Primordial → Primary → Transitional Primary → Secondary → Multilayer → AMF → Antral → Atretic Antral → Multi-oocytic → Corpus Luteum



**Data:** A set of manually annotated ovarian histology slides provided by the MOTHER team. The set is still growing but currently consists of 6 annotated histology slides with several thousand annotated follicles in various stages of follicular development.

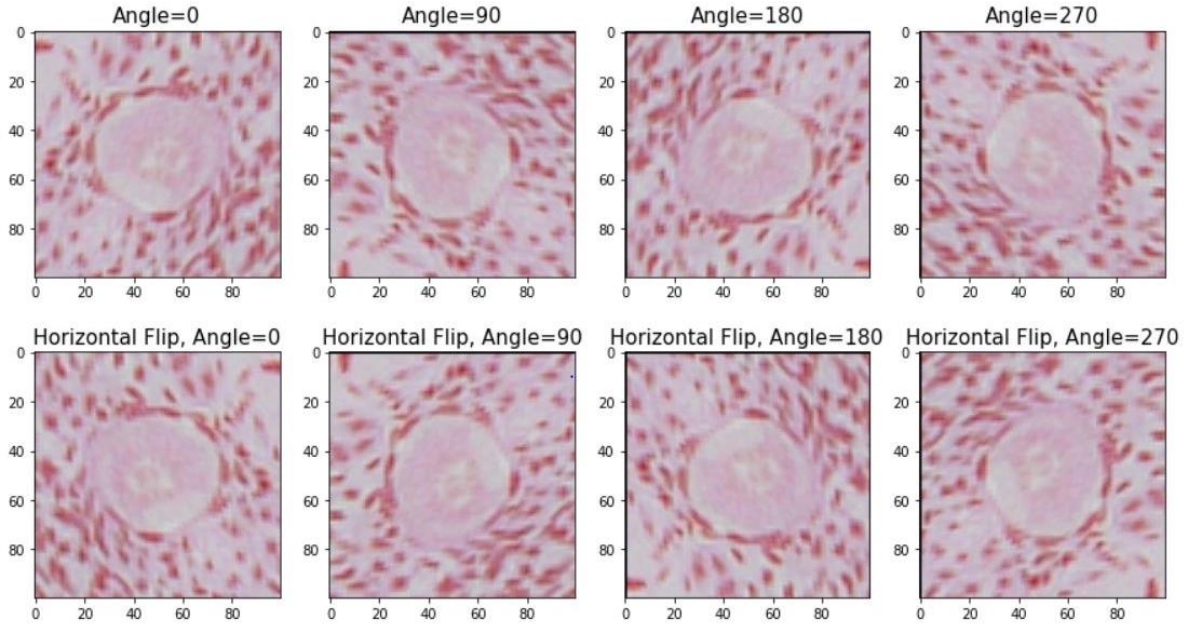
**Approach:** Machine learning, using Convolutional Neural Networks in Jupyter Notebooks using PyTorch.

* *Multispecies Ovary Tissue Histology Electronic Repository MOTHER DB is funded by the National Science Foundation DBI-2054061, 3/2021 – 2/2024. MOTHER-DB.org*

*Machine learning for the classification of ovarian follicles from histology images*:
Dr. James Sluka, Akash Bhapkar, Param Nagda, Parth Rao

## Data Prep Pipeline by Param Nagda:

1. Jupyter Notebook extracts individual 100x100 sub-images from the full histology slide image based on a text file containing follicle center coordinates and type information provided by the MOTHER annotators.
2. Image intensities are normalized.
3. Additional images created for each master image via rotations (90°, 180°, 270°) and inversions giving 8 unique images per input annotation.




4. Data partitioned into training and testing data sets.

# *Machine learning for the classification of ovarian follicles from histology images*:

Dr. James Sluka, Akash Bhapkar, Param Nagda, Parth Rao

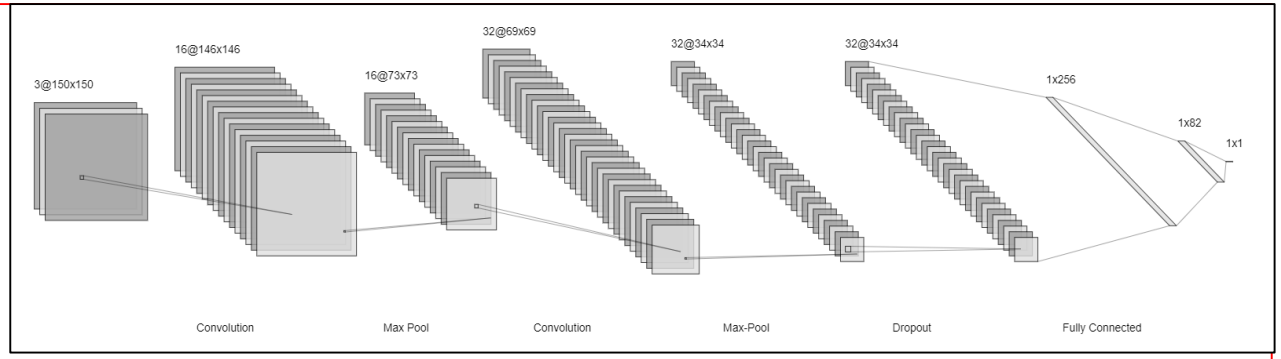
## Akash Bhapkar:

Testing ResNet model as an image classifier and was able to achieve 85% accuracy with the hyperparameter tuning.
Also tried same architecture with continuous numbers (representing the first 6 follicular stages) and performance was similar.

## Parth Rao and Param Nagda:

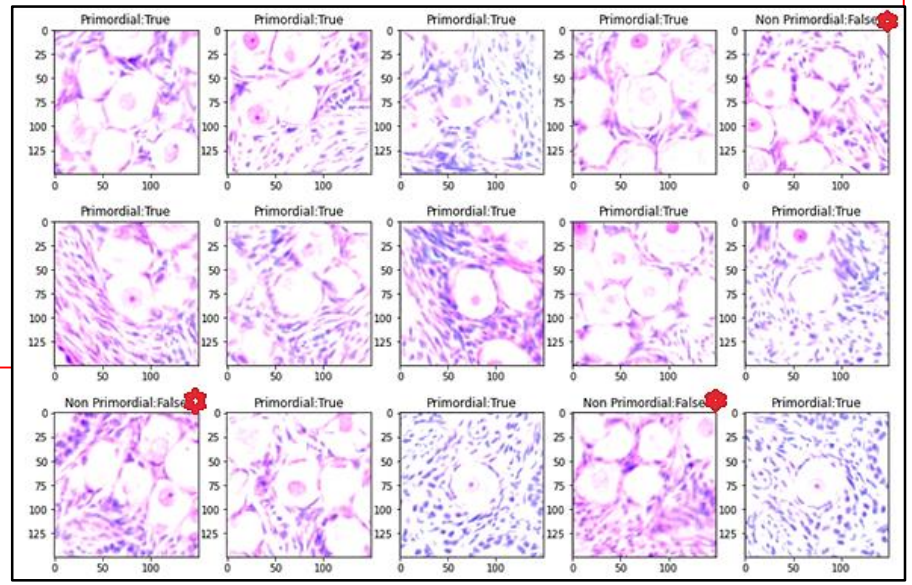
Initially developed CNN models to classify the first 6 follicle types into their classes, obtaining a test accuracy of 76%.

Based on that experience, the focus was shifted to creating binary classifiers that differentiates primordial follicles from non-primordial. Initially, due to the lack of data (1 cell C/S) accuracies were low.




But as the training dataset got larger (6 cell C/S) a significant improvement in results were observed. It was also observed that as we train for greater number of epochs the model quickly began to overtrain.

After testing various models, our best model has a test accuracy of 93%, though this has not been checked by our experts.




## Future Work:

The training data is still limited, refine the classifier as more data becomes available.
Create binary classifiers for the other 5 early follicle stages.