

# Myotube Segregation

Grace Aviance<sup>[A01285158]</sup>, Carlos Sánchez<sup>[A01640495]</sup>, Fabian Lioner<sup>[A01633776]</sup>,  
and Samuel Padilla Esqueda<sup>[A01641383]</sup>

Tec de Monterrey, Av. Gral Ramón Corona No. 2514 Zapopan, Jal.  
<https://github.com/Elcasvi/MyotubeSegregation>

**Abstract.** This project merges biological research with machine learning to examine the effects of alcohol on human muscle tissue using in vitro models. The primary goal is to develop an AI-driven image segmentation tool that can identify and analyze both mature and immature myotubes (muscle fibers) in tissue samples exposed to varying concentrations of alcohol. By applying deep learning techniques, the project aims to automate the recognition and classification of tissue damage, shedding light on how alcohol affects cellular structures at different exposure levels. To achieve this, two pre-trained models are employed for image analysis, with their performance compared to identify the most accurate and effective method for classifying myotubes.

**Keywords:** Machine Learning models · Classification · Convolutional Neural Network (CNN) · Myotubes

## 1 Introduction

The study of cells is essential for understanding the structure and function of living organisms. This knowledge can help explain severe diseases, such as cancer, and other conditions, such as birth defects.

Traditionally, cellular analysis has been conducted using microscopes to observe and study cellular changes. However, advancements in technologies like artificial intelligence now enable the optimization of this manual process, delivering precise results more efficiently. These tools allow for the handling of large volumes of images and reduce human error in the analysis.

In this context, the project aims to develop a Convolutional Neural Network (CNN) to investigate the effect of alcohol on women's muscle tissue. The primary objective is for the model to identify structures called myotubes and analyze how their development is impacted by different alcohol concentrations. By automating the recognition and classification of tissue damage, this project seeks to provide valuable insights into how alcohol affects cellular structures at varying levels of exposure.

To address this challenge, the following objectives are proposed, aiming to determine whether the machine learning techniques used are effective in solving this problem.

## 2 Key Objectives:

- Biological Investigation:

Study the impact of alcohol on muscle tissue by exposing in vitro samples to 0 mM, 25 mM, and 100 mM alcohol concentrations. Analyze the difference in cellular response between mature and immature miotubes. Investigate potential markers of tissue damage, such as cellular morphology changes, and differences in structural integrity between various alcohol treatments.

- AI/ML Development:

Develop and train a machine learning model for image segmentation that can identify and differentiate between mature and immature myotubes in tissue samples. Create a model recognition system that classifies different regions (tiles) of tissue based on alcohol exposure, using these segmented images for further analysis. Manage and process large datasets (TBs of data), ensuring that the model is scalable and capable of handling high volumes of high-resolution imagery.

To achieve these objectives, it is necessary to approach each stage of the project with care. The development approach is as follows.

## 3 Approach:

- Data Preparation:

- Large datasets of tissue images will be gathered, annotated, and preprocessed, with distinct tiles representing different alcohol concentrations.
- The AI will be trained to segment miotubes and to recognize different degrees of maturity in the muscle fibers.

- Machine Learning:

- Image segmentation algorithms will be developed to differentiate between mature and immature miotubes, essential for analyzing the effect of alcohol at the cellular level.
- The model will also analyze tissue damage based on the different alcohol concentrations, enabling automatic classification.

- Validation:

- The model's performance will be validated by comparing its output with manually annotated data, ensuring accurate segmentation and classification.

Since we have completed the development and the defined process, it is important to establish metrics to determine whether the results obtained were satisfactory or not. Therefore, the expected outcomes are outlined here, specifying what the model should be able to achieve.

#### 4 Expected Outcomes:

- An advanced AI system that can accurately identify and segment muscle fibers in tissue samples, providing automated insights into how alcohol affects different types of myotubes.
- A better understanding of how alcohol concentrations impact tissue structure, contributing valuable data to the field of alcohol-related tissue damage studies.

At this point, we have a broad understanding of what we want to achieve and how we aim to achieve it. However, to fully comprehend the process, it is necessary to have a basic understanding of certain concepts that will be used throughout the project's development. Therefore, a brief explanation of the key concepts utilized in this research is presented below.

#### 5 Conceptual Framework

Machine learning is a set of algorithms that can learn and recognize the patterns or objects from the data provided. Therefore, machine learning can make accurate predictions for newly inserted data. Machine learning can be used as a very important tool to overcome challenges in computer vision such as object recognition and medical imaging. In recent decades, machine learning has been used in many applications around the world and successfully solved many AI problems (Lee 2010).

Deep learning is an extension algorithm from machine learning that uses complex architecture and the structure of deep learning consists of multiple layers used for hierarchical features extraction (Schmidhuber 2015). A deep neural network is a network that can be extended by adding new layers consisting of multiple units and the parameters of each layer can be trained (Bengio & Lecun 2007).

A convolutional neural network (CNN) is a class of deep neural networks and has become an efficient tool for solving pattern recognition problems. CNN architecture typically consists of convolutional layers that are fully connected with pooling layers to extract essential features from the image and the fully connected layer is used as a classifier. (Ozsert Yigit & Ozyildirim 2017).

Transfer Learning is a machine learning technique where a model trained on one task is reused as the starting point for a model on a second, related task. Instead of training a model from scratch, you leverage knowledge (parameters, weights, or features) learned by a pre-trained model on a large dataset and fine-tune it for a specific task with a smaller dataset. An application is Computer Vision with object detection, image classification, and segmentation.

Since the models will be trained for myotube detection and segmentation, it is important to demonstrate how these processes are carried out. Detection involves locating an object in the image and selecting that position with a bounding box, while segmentation is more precise and allows for selecting all the pixels

that belong to the object we are searching for. We can observe the differences in Figure 1.

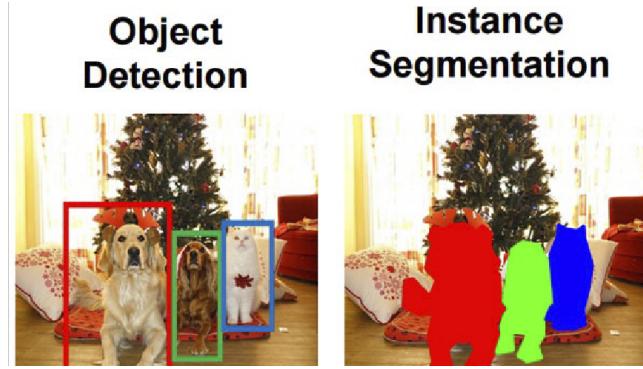


Fig. 1: Object Detection vs Object Segmentation

This conceptual framework outlines the key components and relationships that guide this study, serving as a blueprint for addressing the research objectives. To deepen this understanding, the following theoretical references provide the foundational knowledge and established principles that support the framework.

## 6 Theoretical references

### 6.1 "Deep Learning for an Automated Image-Based Stem Cell Classification"

Hematopoiesis is a process in which hematopoietic stem cells produce other mature blood cells in the bone marrow through cell proliferation and differentiation. The hematopoietic cells are cultured on a petri dish to form a different colony-forming unit (CFU). The idea is to identify the type of CFU produced by the stem cell by several selected convolutional neural network (CNN) pre-trained models to overcome these constraints for automated CFU classification. An example of the process followed is shown in Figure 2.

The images are acquired from the CODTIS. A total of 728 stem cell images were collected as a dataset for this research. The CFU databases have been divided into 70%, 15%, and 15% of the total dataset that represents training data, validation data and classification data, respectively, and categorized into three types which are CFU-erythroid (E), CFU-granulocyte/macrophage (GM) and CFU-PreB.

These images are then pre-processed before being fed into CNN pre-trained models. The image pre-processing process is important to improve and enhance

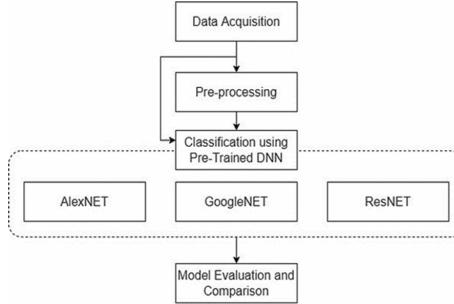


Fig. 2: Workflow diagram

image quality. Through *data augmentation* in pre-processing, the number of images can be increased by generating additional synthetic images with slight modifications to the existing dataset, such as reflection, rotation, and translation; which helps in overcoming the overfitting problem in the training process. Also, *Image cropping* is performed to eliminate unwanted objects or backgrounds from the images, in this case isolate the colonies from the background. In this research, the contrast of the image is enhanced using the contrast-limited adaptive histogram equalization (CLAHE) technique. CLAHE is an improved type of adaptive histogram equalization (AHE) technique which to improve the image pixels transformation from general histogram equalization

The models adopt a deep learning neural network approach to extract informative features from the CFU image. Those three different models are AlexNet, GoogleNet and ResNet-18. The performance of each one in identifying CFU types is evaluated based on their precision and accuracy.

## 6.2 "A deep learning-based segmentation pipeline for profiling cellular morphodynamics using multiple types of live cell microscopy"

Quantitative analysis of cellular morphodynamics requires precise cell edge detection from live cell imaging. However, this is challenging due to issues like noisy, low-contrast images in fluorescence microscopy caused by phototoxicity and photobleaching, and artifacts such as halo and shade-off in phase contrast microscopy, which conventional segmentation methods cannot handle. To address these challenges, the study introduces MARS-Net (Multiple microscopy-type-based Accurate and Robust Segmentation Network), a deep learning-based pipeline that leverages transfer learning and data from various microscopy types to accurately localize cell edges, enabling detailed quantitative profiling of cellular morphodynamics.

Among deep learning models, convolutional neural network (CNN) excels in pattern recognition in images by learning complex features directly from the input images using its hierarchical structure. CNN has achieved great success in

image classification and segmentation. In particular, U-Net is the most widely adopted CNN-based structure for image segmentation and has demonstrated promising segmentation results in static and live cell images.

For instance, U-Net-based models such as StarDist and CellPose have additional structures or outputs to segment images of crowded cells and nuclei effectively.

In this pipeline, they used the U-Net based structure and incorporated the transfer learning technique that initializes the weights of the network with those of the same network trained on ImageNet for the image recognition task. Transfer learning has been applied to many deep learning segmentation models DeepEdge, TernausNetV2 and classification tasks to achieve high performance with a limited dataset. In addition, transfer learning allows the model to achieve higher edge-localization accuracy on multiple types of microscopy datasets. They replaced the U-Net encoder with one of the image classification networks, such as VGG16/VGG19, ResNet50V2 and EfficientNetB7, and used the initial weights from the ImageNet training. Among them, the pretrained VGG19 encoder coupled with U-Net decoder (VGG19-U-Net) segmented the boundary of the cell with the highest accuracy.

Having explored the concepts that support this project, we begin the development process with an analysis of the dataset available for this work.

## 7 Dataset

The dataset contains images of cellular samples exposed to different alcohol concentrations, collected over a period of 5 days at 3-hour intervals.

### 7.1 Description of the Dataset

These images, with dimensions of 2592x1944 pixels and saved in PNG format, collectively exceed 20 GB in size, providing a substantial collection of high-resolution images for detailed analysis.

The folder structure in which the images were provided is as follows:

#### 1. Plates (Alcohol Levels):

- **Plate 1:** Contains images with alcohol concentrations ranging from 0  $\mu\text{M}$  to 10  $\mu\text{M}$ .
- **Plate 2:** Contains images with alcohol concentrations ranging from 25  $\mu\text{M}$  to 100  $\mu\text{M}$ .

#### 2. Date and Time:

- Each plate includes multiple folders organized by the date and time of capture, within the 5-day period, with images taken every 3 hours.

#### 3. Image Types:

The images from each time segment are divided into two folders.

- **Raw:** Contains the original, unprocessed images.
- **Derived:** Contains processed images with filters applied to facilitate visualization.

#### 4. Groups by Alcohol Level:

Each set of images, whether in raw or derived, is classified into six folders based on the alcohol level.

- **Plate 1:**

- A1: 0  $\mu\text{M}$
- A2: 0  $\mu\text{M}$
- A3: 0  $\mu\text{M}$
- B1: 10  $\mu\text{M}$
- B2: 10  $\mu\text{M}$
- B3: 10  $\mu\text{M}$

- **Plate 2:**

- A1: 25  $\mu\text{M}$
- A2: 25  $\mu\text{M}$
- A3: 25  $\mu\text{M}$
- B1: 100  $\mu\text{M}$
- B2: 100  $\mu\text{M}$
- B3: 100  $\mu\text{M}$

#### 5. Sample Sections:

- Within each alcohol level, images are further organized into folders labeled from 00\_01 to 06\_07. These folders represent different sections of each sample.

## 7.2 Understanding the Dataset

It is important to note that the dataset contains a large amount of images, which must be manually processed before being used in the model. This processing involves identifying in each image which structures correspond to myotubes. Given the time-consuming nature of this task, it is crucial to prioritize the most relevant images for our objective.

For model training, only a specific type of processed image will be used. As such, images in the raw folder will not be utilized. Instead, we will work with images from the derived group, specifically those labeled as "BestImages". These images use a filter that enhances contrast between image elements, making it easier to identify myotubes.

It is worth mentioning that images from the first and last days of the experiment do not show Myotubes, either because they had not yet formed or because they had already decayed. Therefore, the model will be trained exclusively with samples collected on August 15 and 16, corresponding to the third and fourth days of the experiment, where mature myotubes in formation are clearly visible.

For model training, images from Plate 1 corresponding to the following sections were selected:

- A1:** 02\_04, 03\_04 (0  $\mu M$ )
- 02\_03, 03\_03 (0  $\mu M$ )
- A2:** 02\_06, 03\_06 (0  $\mu M$ )
- 02\_05, 03\_05 (0  $\mu M$ )
- A3:** 03\_06, 04\_06 (0  $\mu M$ )
- 03\_05, 04\_05 (0  $\mu M$ )

These sections, combined with the specified dates, provide the highest-quality images where the myotubes of interest can be clearly observed. After this image filtering, we are left with a total of 135 images that will be used for training the model.

In Figure 3, we can observe a comparison between a raw image and a best image of one of the samples in the selected group. Although there doesn't seem to be much difference at first glance, the increased contrast at the edges makes a significant difference for the model.

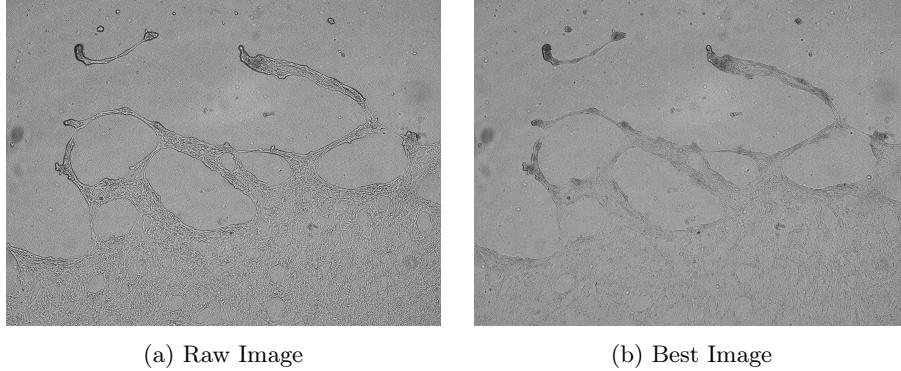


Fig. 3: Comparison Raw Image vs BestImage.

Now that we understand the dataset and know which images we will use for training the model, we can begin with the methodology.

## 8 Methodology:

### 8.1 Pretrained Model and Transfer Learning

To address this challenge, two pretrained object detection models were utilized: `retinanet_resnetfpn_coco` and `Mask R-CNN-ResNet50-FPN`. Both models use

Convolutional Neural Networks (CNNs) to extract features and identify patterns in input images.

Both models are pretrained on large datasets with over 80 categories, meaning that its parameters, or *backbone*, are optimized to extract useful features from images.

The retinanet\_resnetfpn\_coco model was pretrained using bounding boxes for annotations, while the Mask R-CNN-ResNet50-FPN model was trained using segmentation masks. The primary goal of using both models is to compare their performance and determine which approach yields higher precision in identifying myotubes.

To apply these models to our problem, we utilize *Transfer Learning*, a deep learning technique where a pretrained model “transfers” its knowledge to solve a specific problem. In other words, the model already knows how to identify features in images, such as edges or patterns, and does not need to relearn this. Instead, it uses the transferred knowledge to learn how to identify *myotubes*. This approach saves both time and computational resources.

## 8.2 Data Labeling

Before training the models, manual data labeling is required for the selected images from the dataset. For retinanet\_resnetfpn\_coco, bounding boxes were annotated using the Label Studio tool. For Mask R-CNN-ResNet50-FPN, segmentation masks were created using CVAT (Computer Vision Annotation Tool), a specialized software for precise labeling.

These annotated datasets were then used to train the respective models.

An example of the labeling process using Label Studio and CVAT is provided in Figure 4.

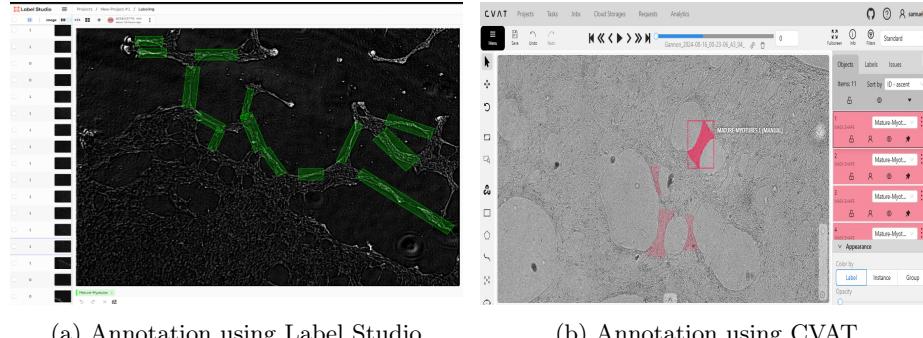


Fig. 4: Comparison of annotation tools: Label Studio (bounding boxes) and CVAT (segmentation masks).

### 8.3 Model Training, Validation, and Testing

After labeling the images, the training process begins. The model training process is divided into three stages. It is important to highlight that for each part of the process, the images selected for model training are divided into the three respective groups. Therefore, there is a training set, a validation set, and a testing set, with percentages of 70%, 20%, and 10%, respectively.

1. **Training:** This phase involves the model learning from a labeled dataset. During training, the model adjusts its parameters (weights) to minimize the error (or loss function) and improve its ability to make accurate predictions.
2. **Validation:** This phase occurs during training and is used to adjust the model's hyperparameters (such as the learning rate). A separate dataset, known as the validation set, is used to evaluate the model's performance on unseen data. This helps prevent overfitting.
3. **Testing:** After training and validation, the model is evaluated on a test dataset that has never been used during training or validation. This provides a final assessment of the model's performance and its ability to generalize to new, unseen data.

## 9 Results

### 9.1 Results for the RetinaNet\_ResNetFPN\_COCO Model:

The model successfully identifies myotubes in the images provided in the test set. In Figure 5, we observe an improvement in the model's precision. In earlier versions, not all myotubes in the images were fully identified. However, in the latest version, the model is now able to detect all the myotubes present in the images.

We can evaluate the model's performance based on the AP (Average Precision) and box\_loss graphs, which provide information about the model's ability to learn and its accuracy.

In Figure 6 we can observe the box\_loss graph, the box\_loss measures the accuracy of the bounding box predictions compared to the true bounding boxes. It combines both the localization of the objects and the presence of the object in the detection box.

The fact that the box\_loss curve is moving downward indicates that the model is improving in object localization. This suggests that the model is learning to predict the locations of the myotubes more accurately.

In Figure 7 we can observe the AP graph, the AP measures how accurately the model identifies objects in the images, considering both precision (the percentage of correct predictions) and recall (how well the model detects all true objects).

An increase in AP reflects better performance in object identification. This indicates that the model is becoming more precise and efficient at classifying objects correctly, suggesting that the training phase is progressing well.

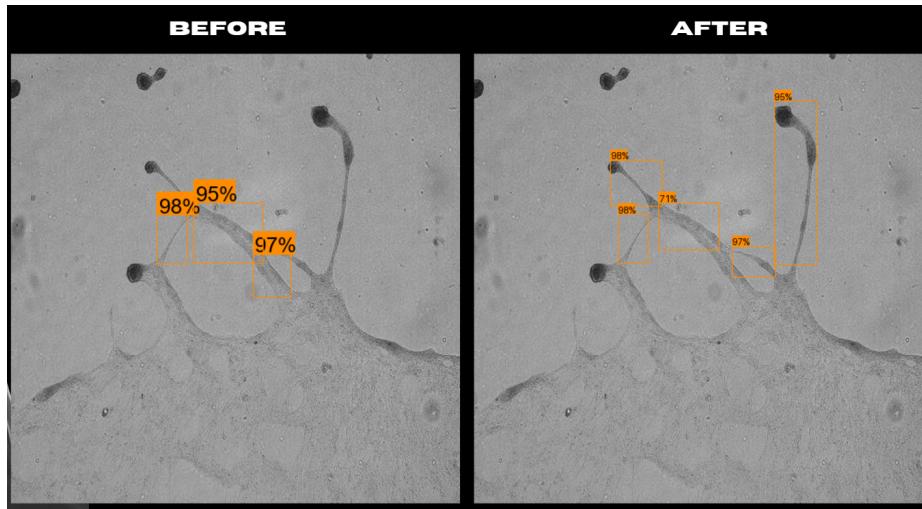


Fig. 5: Myotube identification using the RetinaNet\_ResNetFPN\_COCO model

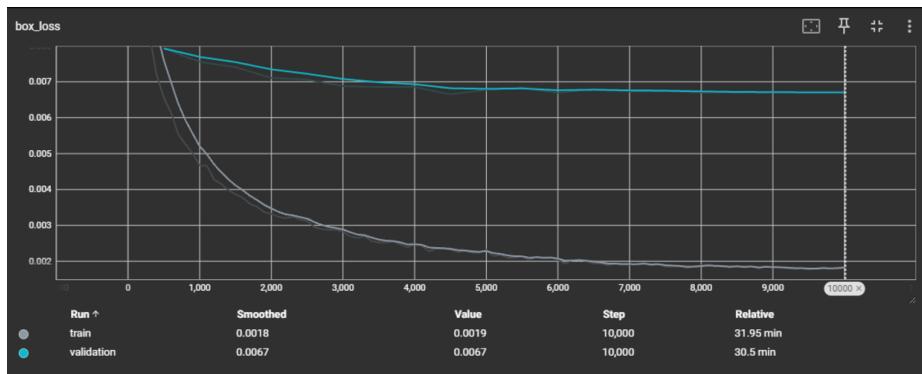


Fig. 6: Box Loss Graph

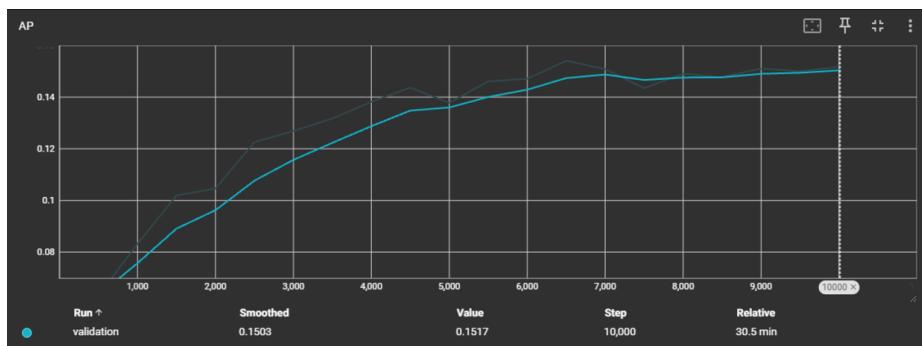


Fig. 7: AP (Average Precision) graph

## 9.2 Results for the Mask R-CNN-ResNet50-FPN Model:

The results of this model proved to be more accurate, identifying more myotubes than the previous model and successfully marking their start and end points. Additionally, by using masks, we are also able to gather relevant information about the myotubes, such as their length and width.

An example of the results of the model is provided in Figure 8.

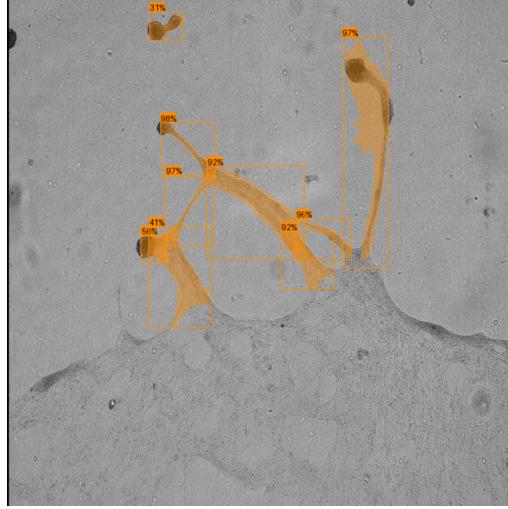


Fig. 8: Myotube identification using the R-CNN-ResNet50-FPN Model

We can evaluate the model's performance based on the Mask\_AP (Average Precision) and model\_loss graphs, which provide information about the model's ability to learn and its accuracy.

In Figure 9 we can observe the model\_loss graph, The model\_loss reflects the overall error of the model, considering both classification and segmentation (mask) losses.

The downward trend in model\_loss indicates that the model is improving in both classification accuracy and segmentation quality. This means the model is learning to reduce errors over time.

In Figure 10 we can observe the mask\_AP graph, The mask\_AP evaluates the model's ability to correctly segment objects (in this case, myotubes) by considering both precision and recall at different confidence levels.

The rising curve in the mask\_AP indicates improved performance in segmenting myotubes accurately. This suggests that the model is not only detecting the objects but also capturing their shape and boundaries more effectively.

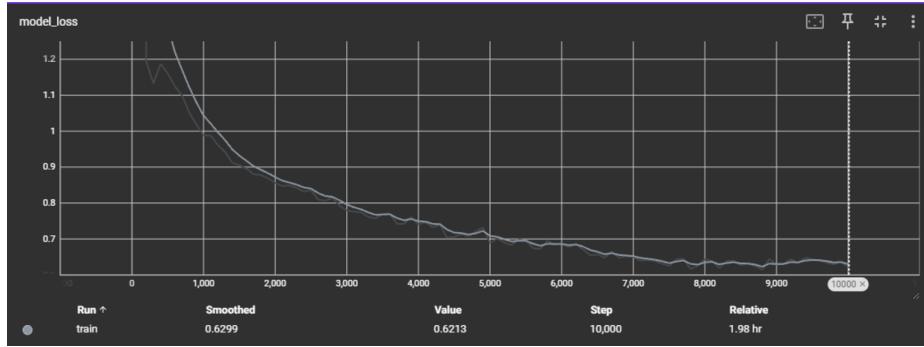


Fig. 9: Model Loss Graph

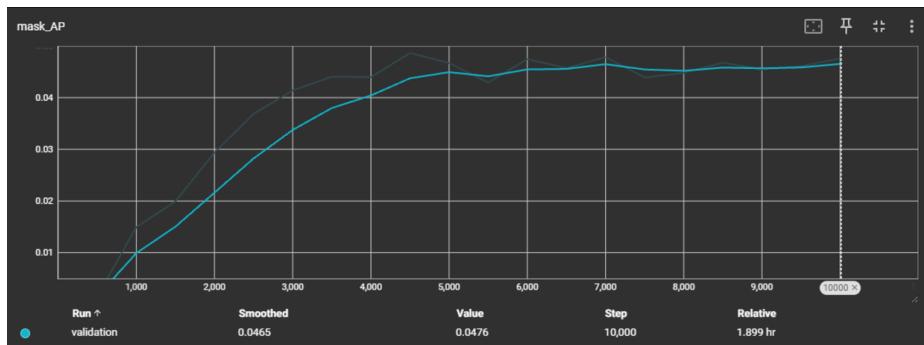


Fig. 10: Mask AP Graph

### 9.3 Web Application

Finally, to make this model more useful and accessible to people without knowledge of machine learning, a web application was created where you can upload images for segmentation, and the model will return the results.

You can access the application using the following QR code or by using the link below.

<https://cellforceai.sanchezapps.net>



Fig. 11: QR code for web application

## 10 Conclusions

The tests show that both models are effective in locating myotubes in the images. However, the Mask R-CNN-ResNet50-FPN model demonstrates higher precision and allows us to gather more data about the identified myotubes. While the results are satisfactory, there is still the possibility of errors, such as incorrectly classifying a structure as a myotube or failing to classify a structure that is indeed a myotube. However, these errors are less frequent than human mistakes and can be corrected with further training. We believe this model is capable of automating the myotube detection process in images, requiring only a quick review by experts to ensure there are no errors, instead of spending hours manually classifying myotubes.

## REFERENCES

Olveres J, González G, Torres F, Moreno-Tagle JC, Carballo-Degante E, Valencia-Rodríguez A, Méndez-Sánchez N, Escalante-Ramírez B. What is new in computer vision and artificial intelligence in medical image analysis applications. Quant Imaging Med Surg 2021;11(8):3830 3853. doi:10.21037/qims-20-1151

## Authors' Contributions

- **Grace Aviance:**
  - Abstract
  - Introduction
  - Conceptual Framework
  - Theoretical references
- **Carlos Sánchez:**
  - Results for the RetinaNet\_ResNetFPN\_COCO Model
  - Results for the Mask R-CNN-ResNet50-FPN Model
  - Pretrained Model and Transfer Learning
- **Fabian Lioner:**
  - key Objectives
  - Approach
  - Expected Outcome
  - Data Labeling
- **Samuel Padilla:**
  - Description of the Dataset
  - Understanding the Dataset
  - Model Training, Validation, and Testing
  - Conclusions