

## Module\_2: (*Template*)

### Team Members:

Alex and Diya

### Project Title:

Does fibrosis increase as you look deeper in the lung?

### Project Goal:

This project sees to find out if fibrosis increase as you look deeper into the lung using image analysis and interpolation.

### Disease Background:

*Fill in information and please note that this module is truncated and only has 5 bullets (instead of the 11 that you did for Module #1).*

- Prevalence & incidence: Close to 200,000 people live with IPF in the U.S. The disease typically occurs in people older than 50, and it affects more men than women (Genentech). Prevalence for patients ages 18 to 34 is 4.0 per 100,000 persons. For patients ages 75 and older, prevalence is 227.2 per 100,000 persons. The annual incidence is estimated to range from 1.2 to 76.4 per 100,000 in 2006 (American Journal of Respiratory and Critical Care)
- Risk factors (genetic, lifestyle): Diseases that can lead to IPF include obstructive sleep apnea, GERD, pulmonary hypertension, COPD, diabetes, depression, pulmonary embolism, coronary artery disease, and lung cancer (Genentech). In addition, there are genetic factors and viral exposures that can lead to the development of IPF (Pulmonary Fibrosis Foundation).
- Symptoms: Symptoms include difficulty breathing, persistent dry cough, fatigue, and rales. Symptoms of IPF are similar to those of other diseases, thus requiring additional testing to identify the disease (Genentech). Chest scans can display the honeycombing pattern of damaged alveoli, a hallmark sign of IPF. Patients will also display cyanosis, digital clubbing, and cyanosis of the nailbeds as their condition continues to worsen (Pham). Patients may experience an acute exacerbation, which is an episode of rapid worsening of shortness of breath and hypoxia often requiring hospitalization (Pulmonary Fibrosis Foundation)
- Standard of care treatment(s): Patients live 3-5 years after diagnosis on average. Other than a lung transplant, there is no cure for IPF. Patients will receive supplemental oxygen, pulmonary rehabilitation, and drug therapies to help manage the disease's effects (Genentech). Some antifibrotic medications can help slow the

progression of the disease, but not eliminate it (Pham). The two most commonly used antifibrotics are nintedanib and pirfenidone. According to the Pulmonary Fibrosis Foundation, the prognosis for patients is very unpredictable, and the statistic of 3-5 years of life expectancy is inaccurate. Different patients have different levels of aggression of IPF, thus making it even more difficult to quantify life expectancy.

- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology): In IPF, the soft tissue of the alveoli forms permanent scarring, affecting the exchange of oxygen and carbon dioxide. It is unknown what causes IPF, but there are underlying factors linked to the disease. Breathing becomes more difficult as the scarring spreads (Genentech). The fibrosis, or scarring, is caused by excess collagen in the connective and interstitial tissue. In a healthy patient, gas exchange happens between the alveoli and capillaries. Alveoli are lined by type I and type II pneumocytes, also known as type I and type II epithelial cells. The alveoli are predominantly lined with type I pneumocytes, which are made of squamous cells and form a nearly continuous barrier between the air and connective tissue. On the other hand, type II pneumocytes are embedded throughout the type I layer. They have a cubic shape, have microvilli for protection, and secrete surfactant. Surfactant is used to prevent the collapse of the alveoli during exhalation. Between the pneumocytes and capillaries lies the interstitial tissue of the lung, as mentioned earlier. When the alveolar lining is damaged, type I pneumocytes release TGF $\beta$ 1, stimulating the type II pneumocytes to trigger the formation of myofibroblasts. These myofibroblasts secrete reticular fibers, a type of collagen used for structural strength, and elastic fibers that give the lungs their elasticity. At the end of this process, the myofibroblasts undergo apoptosis and are removed. In IPF, the type II pneumocytes overproliferate during the repair process and form a dangerous excess of myofibroblasts and collagen. In addition, apoptosis of the myofibroblasts does not occur, leading to more excess collagen production. These factors thicken the interstitial tissue layer, causing issues with ventilation and oxygenation. Furthermore, excess collagen leads to stiffening of the lungs, reduction in lung capacity, honeycombing, and loss of alveoli. (Pham)

## Data-Set:

(Describe the data set(s) you will analyze. Cite the source(s) of the data. Describe how the data was collected -- What techniques were used? What units are the data measured in? Etc.)

- Describe the data set: The data set contained 78 black and white jpeg images at different lengths of a mouses fibrotic lung. In the image the white indicates fibrotic lesions and the black indicate healthy lung.
- Cite the sources of the data set: The data set and images were collected in the Pierce-Cottler Lab at the University of Virginia.
- How was the data collected: Bleomycin, antibiotic from Streptomyces verticillus that induces lung fibrosis, was injected into the mice. Three weeks later the mice were sacrificed and their lungs were harvested. The lung was than fixed onto paraformaldehyde, mounted in gel and than sliced with a microtome in the transverse

plane. After the slices were placed on a microscope slide and a stain was put on for the protein of interest. The slide was than put under a microscope and the images from the data set were taken. The mouse lungs were stained for three proteins: desmin(myofibroblasts), smooth muscle alpha actin ( large smooth blood muscle cells) and CD-31 (endothelial cells in all blood vessels, including capillaries). From there using desmin signaling they were converted into black and white images.

- What techniques were used: Induction of antibiotic into mice, lung harvest, slicing, tissue fixing, tissue embedding, immunofluorescence, microscopy.
- What units is the data measured in: The black and white images had pixels associated to the amount of black versus white which were than converted into percentages.

## Data Analyis:

(Describe how you analyzed the data. This is where you should intersperse your Python code so that anyone reading this can run your code to perform the analysis that you did, generate your figures, etc.)

```
'''Module 2: count black and white pixels in a .jpg and extrapolate points'''

from termcolor import colored
import cv2
import numpy as np
import matplotlib.pyplot as plt
from scipy.interpolate import interp1d
import pandas as pd

# Load the images you want to analyze

filenames = [
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_Sk658 Llobe ch010026.jpg",
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_Sk658 Llobe ch010168.jpg",
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_SK658 Slobe ch010115.jpg",
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_SK658 Slobe ch010147.jpg",
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_SK658 Slobe ch010158.jpg",
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_SK658 Slobe ch010159.jpg",
]

# Enter the depth of each image (in the same order that the images are listed above; you can find these in the .csv file provided to you which is tilted: "Filenames and Depths for Students")

depths = [
```

```

2800,
6000,
8800,
3000,
920,
860
]

# Make the lists that will be used

images = []
white_counts = []
black_counts = []
white_percents = []

# Build the list of all the images you are analyzing

for filename in filenames:
    img = cv2.imread(filename, 0)
    images.append(img)

# For each image (until the end of the list of images), calculate the number of black and white pixels and make a list that contains this information for each filename.

for x in range(len(filenames)):
    _, binary = cv2.threshold(images[x], 127, 255, cv2.THRESH_BINARY)

    white = np.sum(binary == 255)
    black = np.sum(binary == 0)

    white_counts.append(white)
    black_counts.append(black)

# Print the number of white and black pixels in each image.

print(colored("Counts of pixel by color in each image", "yellow"))
for x in range(len(filenames)):
    print(colored(f"White pixels in image {x}: {white_counts[x]}", "white"))
    print(colored(f"Black pixels in image {x}: {black_counts[x]}", "black"))
    print()

# Calculate the percentage of pixels in each image that are white and make a list that contains these percentages for each filename

for x in range(len(filenames)):
    white_percent = (100 * (white_counts[x] / (black_counts[x] + white_counts[x])))

```

```

white_percents.append(white_percent)

# Print the filename (on one line in red font), and below that line
print the percent white pixels and depth

print(colored("Percent white px:", "yellow"))
for x in range(len(filenames)):
    print(colored(f'{filenames[x]}:', "red"))
    print(f'{white_percents[x]}% White | Depth: {depths[x]} microns')
    print()

'''Write your data to a .csv file'''

# Create a DataFrame that includes the filenames, depths, and
percentage of white pixels
df = pd.DataFrame({
    'Filenames': filenames,
    'Depths': depths,
    'White percents': white_percents
})

# Write that DataFrame to a .csv file

df.to_csv('Percent_White_Pixels.csv', index=False)

print("CSV file 'Percent_White_Pixels.csv' has been created.")

'''the .csv writing subroutine ends here'''
# Interpolate a point: given a depth, find the corresponding white
pixel percentage -- for linear

interpolate_depth = float(input(colored("Enter the depth at which you", "yellow")))
want to interpolate a point: ", "yellow"))

x = depths
y = white_percents

i = interp1d(x, y, kind='linear') # You can also use 'quadratic',
'cubic', etc.
interpolate_point = i(interpolate_depth)
print(colored(f'The interpolated point is at the x-coordinate {interpolate_depth} and y-coordinate {interpolate_point}.', "green"))

depths_i = depths[:]
depths_i.append(interpolate_depth)
white_percents_i = white_percents[:]
white_percents_i.append(interpolate_point)

# make two plots: one that doesn't contain the interpolated point,
just the data calculated from your images, and one that also contains

```

```

the interpolated point (shown in red)
fig, axs = plt.subplots(2, 1)

axs[0].scatter(depths, white_percents, marker='o', linestyle='--',
color='blue')
axs[0].set_title('Plot of depth of image vs percentage white pixels -- linear')
axs[0].set_xlabel('depth of image')
axs[0].set_ylabel('white pixels as a percentage of total pixels')
axs[0].grid(True)

axs[1].scatter(depths_i, white_percents_i, marker='o', linestyle='--',
color='blue')
axs[1].set_title('Plot of depth of image vs percentage white pixels w/ interpolated point (red)--linear')
axs[1].set_xlabel('depth of image')
axs[1].set_ylabel('white pixels as a percentage of total pixels')
axs[1].grid(True)
axs[1].scatter(depths_i[len(depths_i)-1],
white_percents_i[len(white_percents_i)-1], color='red', s=100,
label='Highlighted point')

# Adjust layout to prevent overlap
plt.tight_layout()
plt.show()

# Interpolate a point: given a depth, find the corresponding white pixel percentage -- for quadratic

interpolate_depth = float(input(colored("Enter the depth at which you want to interpolate a point: ", "yellow")))

x = depths
y = white_percents

i = interp1d(x, y, kind='quadratic') # You can also use 'quadratic', 'cubic', etc.
interpolate_point = i(interpolate_depth)
print(colored(f'The interpolated point is at the x-coordinate {interpolate_depth} and y-coordinate {interpolate_point}.', "green"))

depths_i = depths[:]
depths_i.append(interpolate_depth)
white_percents_i = white_percents[:]
white_percents_i.append(interpolate_point)

# make two plots: one that doesn't contain the interpolated point, just the data calculated from your images, and one that also contains

```

```

the interpolated point (shown in red)
fig, axs = plt.subplots(2, 1)

axs[0].scatter(depths, white_percents, marker='o', linestyle='--',
color='blue')
axs[0].set_title('Plot of depth of image vs percentage white pixels -- quadratic')
axs[0].set_xlabel('depth of image')
axs[0].set_ylabel('white pixels as a percentage of total pixels')
axs[0].grid(True)

axs[1].scatter(depths_i, white_percents_i, marker='o', linestyle='--',
color='blue')
axs[1].set_title('Plot of depth of image vs percentage white pixels w/ interpolated point (red) -- quadratic')
axs[1].set_xlabel('depth of image')
axs[1].set_ylabel('white pixels as a percentage of total pixels')
axs[1].grid(True)
axs[1].scatter(depths_i[len(depths_i)-1],
white_percents_i[len(white_percents_i)-1], color='red', s=100,
label='Highlighted point')

# Adjust layout to prevent overlap
plt.tight_layout()
plt.show()

```

```

Counts of pixel by color in each image
White pixels in image 0: 73224
Black pixels in image 0: 4121080

White pixels in image 1: 99350
Black pixels in image 1: 4094954

White pixels in image 2: 151132
Black pixels in image 2: 4043172

White pixels in image 3: 80534
Black pixels in image 3: 4113770

White pixels in image 4: 59788
Black pixels in image 4: 4134516

White pixels in image 5: 59426

```

```
Black pixels in image 5: 4134878
```

```
Percent white px:
```

```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_Sk658 Llobe ch010026.jpg:  
1.7457962036132812% White | Depth: 2800 microns
```

```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_Sk658 Llobe ch010168.jpg:  
2.3686885833740234% White | Depth: 6000 microns
```

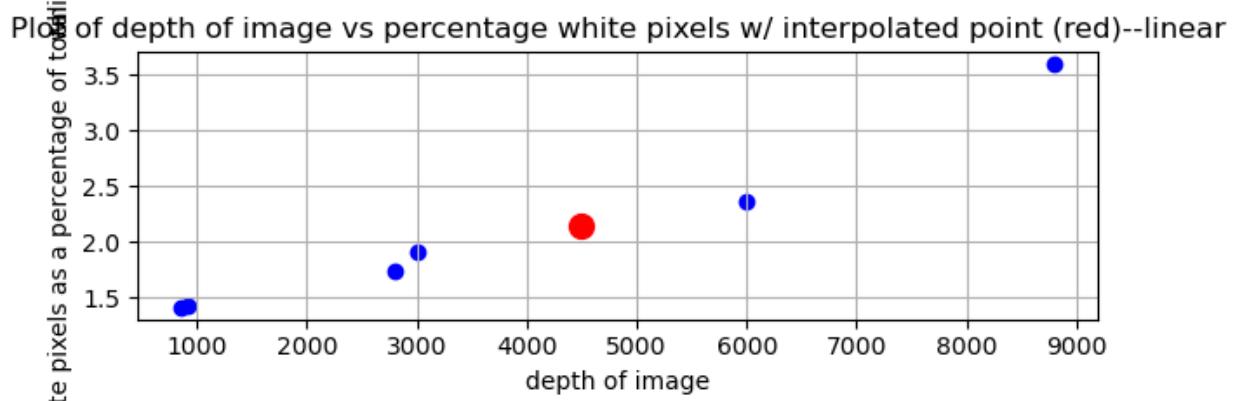
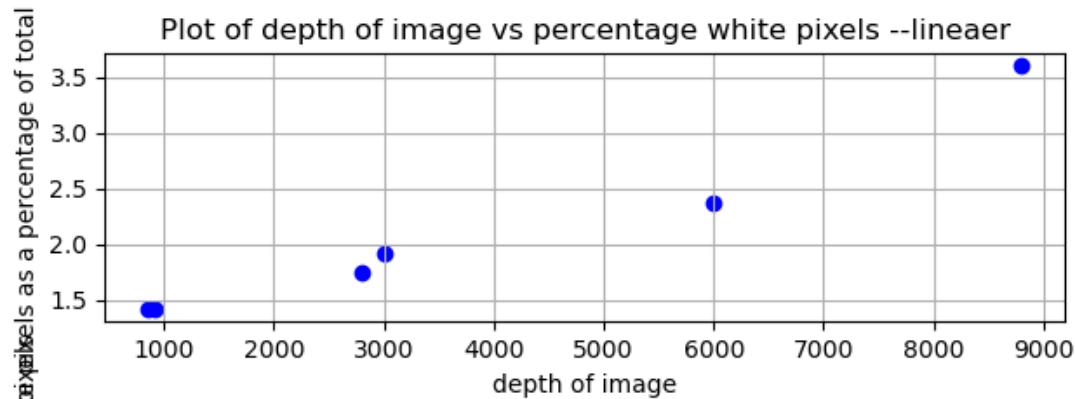
```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_SK658 Slobe ch010115.jpg:  
3.6032676696777344% White | Depth: 8800 microns
```

```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_SK658 Slobe ch010147.jpg:  
1.9200801849365234% White | Depth: 3000 microns
```

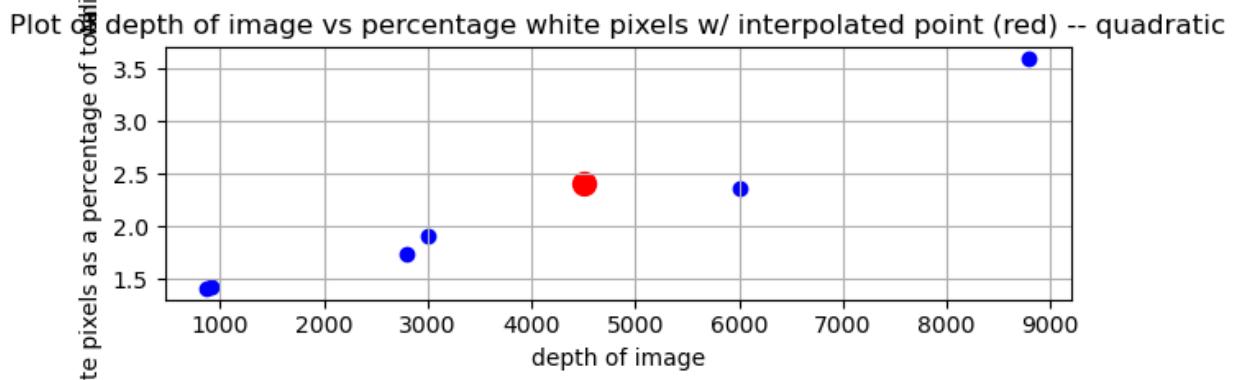
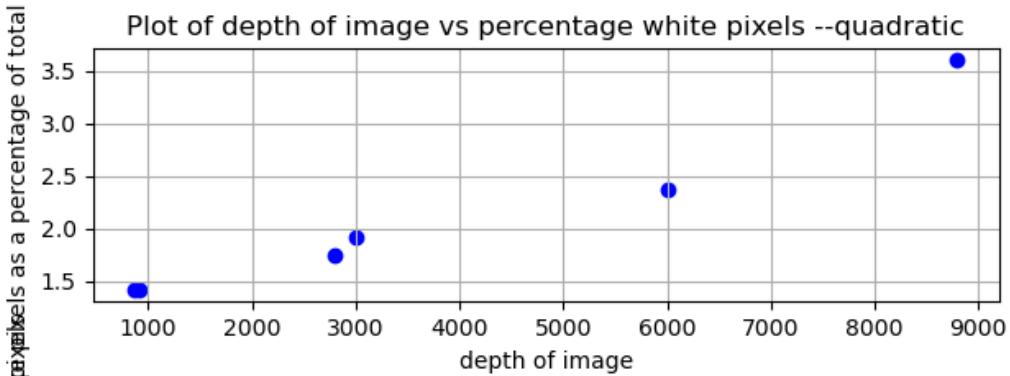
```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_SK658 Slobe ch010158.jpg:  
1.4254570007324219% White | Depth: 920 microns
```

```
/Users/diyamahendravadi/Downloads/Computational BME/Module  
2/MASK_SK658 Slobe ch010159.jpg:  
1.4168262481689453% White | Depth: 860 microns
```

```
CSV file 'Percent_White_Pixels.csv' has been created.  
The interpolated point is at the x-coordinate 4500.0 and y-coordinate  
2.1443843841552734.
```



The interpolated point is at the x-coordinate 4500.0 and y-coordinate 2.410273611756219.



#For this we are doing the same thing as above for just the one at 2000 microns in depth

#This is because we are going to use this to see if our interpolated values are similar to the values we find from the actual data set

```
from termcolor import colored
import cv2
import numpy as np
import matplotlib.pyplot as plt
from scipy.interpolate import interp1d
import pandas as pd

# Load the images you want to analyze

filenames = [
    r"/Users/diyamahendravadi/Downloads/Computational BME/Module 2/MASK_SK658 Slobe ch010134.jpg"
]
```

# Enter the depth of each image (in the same order that the images are listed above; you can find these in the .csv file provided to you which is titled: "Filenames and Depths for Students")

```
depths = [
    4500
```

```

]

# Make the lists that will be used

images = []
white_counts = []
black_counts = []
white_percents = []

# Build the list of all the images you are analyzing

for filename in filenames:
    img = cv2.imread(filename, 0)
    images.append(img)

# For each image (until the end of the list of images), calculate the
# number of black and white pixels and make a list that contains this
# information for each filename.

for x in range(len(filenames)):
    _, binary = cv2.threshold(images[x], 127, 255, cv2.THRESH_BINARY)

    white = np.sum(binary == 255)
    black = np.sum(binary == 0)

    white_counts.append(white)
    black_counts.append(black)

# Print the number of white and black pixels in each image.

print(colored("Counts of pixel by color in each image", "yellow"))
for x in range(len(filenames)):
    print(colored(f"White pixels in image {x}: {white_counts[x]}",
    "white"))
    print(colored(f"Black pixels in image {x}: {black_counts[x]}",
    "black"))
    print()

# Calculate the percentage of pixels in each image that are white and
# make a list that contains these percentages for each filenamae

for x in range(len(filenames)):
    white_percent = (100 * (white_counts[x] / (black_counts[x] +
white_counts[x])))
    white_percents.append(white_percent)

# Print the filename (on one line in red font), and below that line
# print the percent white pixels and depth

print(colored("Percent white px:", "yellow"))
for x in range(len(filenames)):
```

```

print(colored(f'{filenames[x]}:', "red"))
print(f'{white_percents[x]}% White | Depth: {depths[x]} microns')
print()

'''Write your data to a .csv file'''

# Create a DataFrame that includes the filenames, depths, and
# percentage of white pixels
df = pd.DataFrame({
    'Filenames': filenames,
    'Depths': depths,
    'White percents': white_percents
})

Counts of pixel by color in each image
White pixels in image 0: 91653
Black pixels in image 0: 4102651

Percent white px:
/Users/diyamahendravadi/Downloads/Computational BME/Module
2/MASK_SK658 Slobe ch010134.jpg:
2.185177803039551% White | Depth: 4500 microns

```

## Verify and validate your analysis:

- Looking at the graph of percent white pixels versus depth into the lung, our graph shows that as the depth into the lung increases, the percentage of white pixels also increases. For example when the depth was 860 microns the percentage was 1.417%, when the depth was 2800 microns the percentage was 1.746%, and when the depth was 6000 microns the percent was 2.369. This shows an increase as depth increases which goes along with our hypothesis as you get deeper into the lung there is a higher extent of lung fibrosis and that lung fibrosis does vary throughout the lung.
- For both the linear and quadratic interpolations, the y-values made sense. For both of them we used a x-coordinate of 4500 microns meaning the depth was 4500. For the linear it gave a y-value of 2.1444% and for the quadratic it gave a y-value of 2.4103% which are pretty similar. Compared to the other data points measured the do make sense, however, the linear one make a bit more sense than the quadratic one. For the quadratic it showed 2.4103% at 4500 microns where the 6000 was a 2.3686885833740234% which is lower, which does not make much sense as our trend shows increasing. This could just mean that the linear fit was a better approximation because the 2.1444% given by the linear approximation makes sense in terms of the trend of increasing shown in our graph.
- However when comparing our interpolated values to the value from the dataset at 4500 microns there was a clear difference in which interpolated model is better. The percentage white pixels for the one at 4500 microns in depth was 2.1851% which closer to linear than quadratic. This helps support the conclusion that the linear model is more

fitting of the data than the quadratic model. However this needs to be taken with caution as only 6 images were used to generate these interpolation which is too small.

- <https://www.ncbi.nlm.nih.gov/books/NBK448162/> : This article by Abdulghani Sankari; Kyle Chapman; Saad Ullah published in StatPearls supports our conclusion that deeper in the lung has worse fibrosis. In this article it states, imaging techniques display a distinctive pattern with fibrosis which is typically concentrated at the base of the lung. This agrees with our results showing that at the deepest part of the lung should have the most fibrosis which goes with both interpolation models the quadratic and the linear ones.
- <https://pmc.ncbi.nlm.nih.gov/articles/PMC4518345/> : Additionally this article also supports our conclusion about how lung fibrosis increases as you go deeper into the lung. This article is from A Thoracic Medicine article. This article states "IPF usually predominates at the base of the lung." This also validates our conclusion.

## Conclusions and Ethical Implications:

*(Think about the answer your analysis generated, draw conclusions related to your overarching question, and discuss the ethical implications of your conclusions.)*

- Our data analysis indicates that fibrotic tissue generally increases with depth into the lung, as shown by the trend of higher white pixel percentages at greater depths and therefore varies throughout the lung. This supports the hypothesis that fibrosis is not uniform throughout the lung, but rather concentrated in certain regions.
- For the medical device company developing a depth-specific biopsy tool, this data shows the need for a device that can collect multiple samples at specific and accurate depths that are deeper which is where there is higher areas of fibrosis. This company currently should go with a linear model as that was the best for our interpolation however, this was only with six data points which is clearly not enough to know anything. I would advise this company to get more mice and repeat this process on several as this device will be sold to millions of hospitals and used so it needs to be more accurate rather than off six points.
- For ethical implications, this testing for this device does use mice and making sure all procedures are followed is key. Additionally knowing that the results from this assignment are based on six data points from various regions in the lung and not at a specific depth away from each other needs to be considered. Additionally since this is done on mice and will be translated to work on humans this also needs to be highlighted before the medical device company takes this public. This could also include doing a clinical trial or something to even see if this data translates to the human lung. Additionally since it is a biopsy device it must be safe for patients and all the imaging needs to have doctor patient confidentiality.

## Limitations and Future Work:

*(Think about the answer your analysis generated, draw conclusions related to your overarching question, and discuss the ethical implications of your conclusions.)*

- There were two big limitations regarding the data set. One was the small amount of images created and an even smaller amount of images used to create the interpolations. This data set had a total of 78 images which is not enough to base a medical device off of. Additionally for the interpolations only 6 images were used to create those graphs. Since

this data was so small the conclusions need to be taken with caution. Additionally another big limitation was with the red blood cells causing additional noise in the images. Red blood cells do not show lung fibrosis however with the way the imaging was done the red blood cells in the lung do show as white pixels which changes the data a bit.

- For future work I would figure a way to get rid of the red blood cells and than do the interpolation on all 78 images instead of 6. This will allow more data points to interpolate over which gives more accurate results and removing the red blood cells will only allow the white pixels to be from the fibrosis not allowing outside factors to factor into our results. Additionally get more data sets with different lungs to compare results as well would be needed for a medical device company. Since this company is releasing a device they do not want it to fail due to lack of testing so the more data points and images they get the more likely their company will succeed. Finally if possible and safe I would reccommed and clinincal trial to test to see if these results transfer to humans as this device will be used on humans not mice.

## NOTES FROM YOUR TEAM:

*This is where our team is taking notes and recording activity.*

- Checkpoint 1:
  - We finished the CSV file with the files, depth, and percentage
  - We did the interpolation graphs as well
  - Did all of the bullets for the background information
- Checkpoint 2
  - We did the quadratic interpolation
  - Finished the data set, verify and validate results, conclusions, and limiatation sections.

## QUESTIONS FOR YOUR TA:

*These are questions we have for our TA.*

- We do not have any questions for TA

## References

Genentech. (2019, June 19). Behind Each Breath: Shedding Light on Idiopathic Pulmonary Fibrosis [Video]. YouTube. <https://www.youtube.com/watch?v=s2owdw!5Vac>

Hung Long Pham. (2018, June 14). Idiopathic pulmonary fibrosis: causes, symptoms, diagnosis, treatment, pathology [Video]. YouTube. [https://www.youtube.com/watch?v=Z6\\_0WkceY88](https://www.youtube.com/watch?v=Z6_0WkceY88)

Pulmonary Fibrosis Foundation. (n.d.). Idiopathic pulmonary fibrosis [Video]. Facebook. <https://www.facebook.com/watch/?v=1042168686536691>

Raghu, G., Weycker, D., Edelsberg, J., Bradford, W. Z., & Oster, G. (2006). Incidence and prevalence of idiopathic pulmonary fibrosis. American Journal of Respiratory and Critical Care Medicine, 174(7), 810–816. <https://doi.org/10.1164/rccm.200602-163OC>

Sankari, A., Chapman, K., & Ullah, S. (2024, April 23). Idiopathic pulmonary fibrosis. In StatPearls. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK448162/>

Sarkar, M., Madabhavi, I., Niranjan, N., & Dogra, M. (2015). Auscultation of the respiratory system. Annals of Thoracic Medicine, 10(3), 158–168. <https://doi.org/10.4103/1817-1737.160831>