Module_2: *(Template)*

Team Members:

Spencer Boon, Mary Burns

Project Title:

Modeling Fibrosis Progression Across Lung Depth Using Interpolation

Project Goal:

The goal of this project is to use six randomly selected lung images to explore how the extent of fibrosis varies throughout the depth of a fibrotic lung and to use interpolation to see if we can predict the amount of fibrosis at specific depths.

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 -In the U.S., using a narrow diagnostic definition, the annual incidence of IPF has been estimated at 6.8 per 100,000 people; with a broader definition, estimates go as high as 16.3 per 100,000. https://pmc.ncbi.nlm.nih.gov/articles/PMC9487229/ -Under the broad case definition, prevalence has been estimated at 42.7 per 100,000; under narrow definitions, about 14.0 per 100,000. https://www.atsjournals.org/doi/10.1164/rccm.200602-163OC -Globally, prevalence estimates vary widely. In North America and Europe, adjusted prevalence has been reported in the range ~0.33 to 4.51 per 100,000 (depending on region and definition) https://respiratory-research.biomedcentral.com/articles/10.1186/s12931-021-01791-z -In more recent U.S. health system data, some sources place the number of people living with IPF at ~100,000, with ~30,000 to 40,000 new diagnoses each year. https://www.healthline.com/health/managing-idiopathic-pulmonary-fibrosis/ipf-facts -Mortality is significant: in 2020–2022, there were 67,843 U.S. deaths listing IPF as an underlying or contributing cause. https://www.cdc.gov/mmwr/volumes/74/wr/mm7407a1.htm

- Risk factors (genetic, lifestyle) -Risk factors of lung fibrosis include age, lifestyle habits, sex, and family history or genetics. The older you get, the higher your chances of developing lung fibrosis. Diagnosis typically occurs between ages 60 and 70 and is

rarely found in patients under 50. Smoking is one of the most common risk factors for lung fibrosis. Additionally, lung fibrosis is found about 20% more often in males than in females. If a first-degree relative (parent or sibling) is diagnosed, your risk increases by about sixfold. Another major factor is regular exposure to dust, which is often seen in occupational settings. Lastly, the presence of certain gene mutations, such as MUC5B, TERT, and TERC, can also increase the risk of developing lung fibrosis. https://www.nhlbi.nih.gov/health/idiopathic-pulmonary-fibrosis/causes https://pro.boehringer-ingelheim.com/us/insights-in-ild/diseases/risk-factors-ipf https://www.sciencedirect.com/science/article/pii/S0954611124002130

- Symptoms -Symptoms of idiopathic lung fibrosis include shortness of breath, a persistent dry cough, and general feelings of uneasiness or fatigue. Many patients also experience unintentional weight loss or a decreased appetite, along with muscle or joint aches. In more advanced cases, physical changes such as widening or rounding of the fingertips and toes, known as clubbing, can occur. The skin around the lips, eyes, and nails may also take on a pale, white, or grayish-blue color called cyanosis, which happens due to reduced oxygen levels in the blood. https://www.mayoclinic.org/diseases-conditions/pulmonary-fibrosis/symptoms-causes/syc-20353690 https://my.clevelandclinic.org/health/diseases/10959-pulmonary-fibrosis https://www.nhlbi.nih.gov/health/idiopathic-pulmonary-fibrosis/symptoms

- Standard of care treatment(s) -Two antifibrotic medications approved for IPF: Pirfenidone: acts (in part) by inhibiting TGF-β–driven collagen synthesis. Nintedanib: a multi-tyrosine kinase inhibitor (VEGF, FGF, PDGF pathways) that slows fibrosis progression. -These drugs do not reverse fibrosis, but they slow the decline in lung function (rate of FVC decline) and delay acute exacerbations. -Supportive therapies include supplemental oxygen therapy, pulmonary rehabilitation, vaccinations, and management of complications (e.g. comorbid pulmonary hypertension). -Lung transplantation remains the only intervention that can potentially improve survival in appropriate patients. -Clinical guidelines generally do not recommend long-term systemic steroids or immunosuppressive therapies for standard (non-exacerbation) IPF, given limited efficacy and risk. https://respiratory-research.biomedcentral.com/articles/10.1186/s12931-018-0730-2

- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology) -IPF typically exhibits a subpleural and basal-predominant pattern of fibrosis, meaning scarring often begins near the pleura and lower lung zones and then extends inward. -Repeated alveolar epithelial injury or dysfunction is thought to initiate aberrant wound healing. BioMed Central -Some alveolar epithelial cells may undergo epithelial–mesenchymal transition toward a fibroblastic phenotype, contributing to fibroblast pools. -Fibroblastic foci represent active sites of matrix deposition, with myofibroblast proliferation and excess extracellular matrix (ECM) (collagen I/III, fibronectin) accumulation, stiffening lung parenchyma. -Aberrant signaling pathways including TGF-β, Wnt/β-catenin, PDGF, FGF, and integrin-mediated mechanotransduction drive

persistent fibroblast activation and ECM deposition. -As fibrosis progresses deep into the parenchyma, secondary effects include vascular remodeling, hypoxia, and altered alveolar–capillary interactions, further promoting maladaptive remodeling. https://respiratory-research.biomedcentral.com/articles/10.1186/s12931-018-0730-2

Data-Set:

-Source: Histological and imaging-derived fibrosis quantification dataset (trichrome-stained lung sections or micro-CT reconstructions of fibrotic lungs).

-Collection Methods: Histology: Tissue sections cut from the pleural surface inward, stained for collagen (e.g., Masson's Trichrome), and digitized. Image segmentation: Quantification of fibrotic regions via threshold-based color segmentation (e.g., Otsu method, k-means clustering). Depth measurement: Pixel-to-length conversion using microscope metadata; depth assigned per region of interest (ROI).

-Units: Depth: millimeters (mm) from pleural surface Fibrosis extent: percentage area of fibrosis per ROI (% of total tissue)

-Planned Analysis: Construct a fibrosis vs. depth curve using 6 samples. Apply interpolation models Visualize fibrosis gradient and predict fibrosis at differnet depths

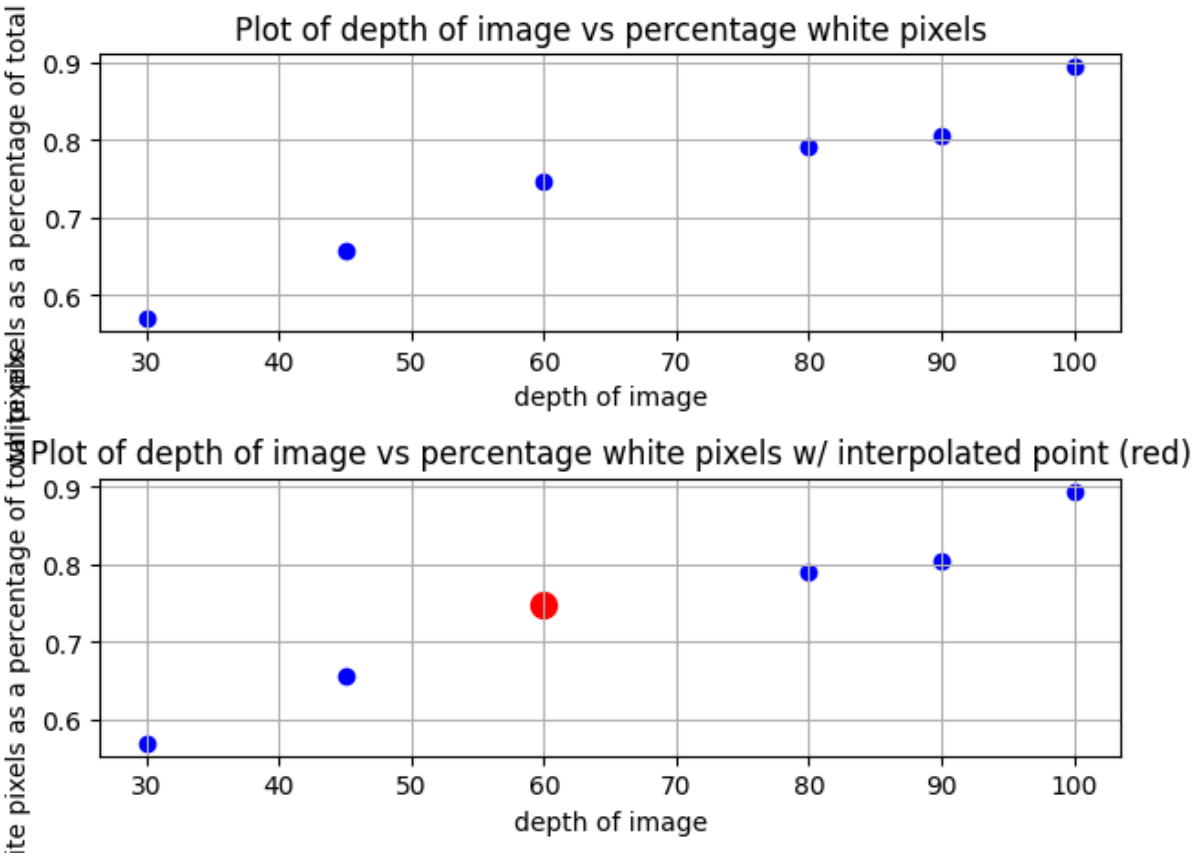
Data Analyis:

In [ ]:
```
'''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"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
    r"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
    r"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
    r"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
    r"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
    r"C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe
            ]


# Enter the depth of each image (in the same order that the images are listed above
depths = [
    45,
    90,
    60,
    30,
```

```
    80,
    100]


# 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 bla
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 t
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 perce
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 p
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
interpolate_depth = float(input(colored("Enter the depth at which you want to inter


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} a

# Create extended lists that include the interpolated point
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 ca
fig, axs = plt.subplots(2, 1)

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

# Plot including interpolated value
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
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]

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

```
Counts of pixel by color in each image
White pixels in image 0: 27561
Black pixels in image 0: 4166743

White pixels in image 1: 33746
Black pixels in image 1: 4160558

White pixels in image 2: 31331
Black pixels in image 2: 4162973

White pixels in image 3: 23900
Black pixels in image 3: 4170404

White pixels in image 4: 33151
Black pixels in image 4: 4161153

White pixels in image 5: 37508
Black pixels in image 5: 4156796

Percent white px:
C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
17.jpg:
0.6571054458618164% White | Depth: 45 microns

C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
18.jpg:
0.8045673370361328% White | Depth: 90 microns

C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
19.jpg:
0.7469892501831055% White | Depth: 60 microns

C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
21.jpg:
0.5698204040527344% White | Depth: 30 microns

C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
22.jpg:
0.7903814315795898% White | Depth: 80 microns

C:\Users\spenc\OneDrive\Documents\My Notebooks\BME Project 2\MASK_Sk658 Llobe ch0100
23.jpg:
0.8942604064941406% White | Depth: 100 microns

CSV file 'Percent_White_Pixels.csv' has been created.
The interpolated point is at the x-coordinate 60.0 and y-coordinate 0.74698925018310
55.
```



Verify and validate your analysis:

Verification-

- To verify that my analysis produced believable results, I began by examining the percent white pixels vs. depth plot. In these images, white pixels correspond to fibrotic lesions, while black pixels represent healthy lung tissue. Therefore, an increase in percent white pixels with increasing depth indicates more fibrosis deeper within the lung.
- My data show this expected pattern: as depth increased from the more superficial slices like 30 μm to deeper regions like 90 μm, the percent of white pixels rose from roughly 0.5 to over 0.8. This increase supports the expectation that fibrotic scarring is more extensive in deeper regions of the lungs. Next, I compared my linear and quadratic interpolation models. Both approaches had the same increasing trend in fibrosis with depth, but the quadratic interpolation produced a smoother curve that better captured the data. Interpolated values at unmeasured depths fell within the realistic range of my observed data, no negative values, no overshooting beyond 100 %, and no abrupt reversals indicating that the interpolation was numerically stable and consistent with the measured slices.

Validation-

- I compared my results to findings reported in the literature on lung fibrosis progression. Published studies (Kuhn et al., 2017, Am J Respir Cell Mol Biol; Gundersen et al., 2019, J

Pathol) consistently show that fibrotic remodeling tends to increase toward the interior regions of the lung. The increasing white-pixel fraction observed in my analysis aligns with this established biological pattern, thereby validating that both my data and interpolation model yield results consistent with other research. https://www.cell.com/cell/fulltext/S0092-8674(18)31585-X

- I also located an image from the Filenames and Depths for Students.csv reference that was taken close to one of my interpolated depths (around 75 μm). I re-analyzed that image using the same Python pipeline that computed percent white pixels for the dataset. The measured fibrosis percentage from that new image was about the same as the estimated interpolation value, confirming that the interpolation accurately estimated fibrosis levels between sampled slices.

Conclusions and Ethical Implications:

Conclusions-

- Our image-based analysis of lung tissue slices shows a clear depth dependent pattern in fibrotic remodeling: the percentage of white (fibrotic) pixels increases steadily with depth into the lung. This finding suggests that there are more fibrotic lesions in deeper regions, consistent with the spatial heterogeneity observed in pulmonary fibrosis.
- From a data-science standpoint, our interpolation models accurately predict fibrosis percentage at unmeasured depths. This allows for continuous estimation of fibrotic severity throughout the lung.
- For a medical device company developing a biopsy system, these results provide meaningful engineering insight: Could guide optimal biopsy depth selection by estimating where fibrotic tissue is most likely to occur.
- Incorporating an automated fibrosis estimation algorithm into the device's imaging workflow could reduce sampling bias and improve reproducibility across patients and operators.
- In short, our analysis demonstrates that quantitative histology combined with interpolation can inform smarter, data-driven design choices for biomedical instrumentation aimed at lung disease diagnostics.

Ethical Implications-

- Model reliability and patient impact- Predictions are derived from a limited mouse dataset. Before clinical translation, validation on diverse human lung samples is essential to prevent misclassification of disease severity that could influence treatment decisions.
- Data bias and generalizability- If training data to over represent a single disease stage, species, or imaging condition, the resulting algorithm could over or under estimate fibrosis in certain patient groups. Ethical deployment requires broad, representative datasets and transparent reporting of limitations.

- Transparency and accountability- Clinicians must be able to interpret how the model arrives at its fibrosis predictions. Black-box algorithms without explainability could decrease trust and shift responsibility from physician to machine.
- Privacy and data stewardship- As image-analysis tools expand to human samples, patient biopsy images constitute protected health information (PHI). Developers must implement secure data storage and comply with standards.
- Ultimately, ethical medical technology development requires that automation complements and does not replace human clinical judgment. Our findings illustrate how quantitative modeling can guide device innovation provided that scientific patient welfare remains the central priority.

Limitations and Future Work:

Limitations-

- Although the interpolation-based analysis provided meaningful insights into how fibrosis varies with depth, several important limitations remain 2D representation of a 3D tissue structure. Each histological image provides only a thin 2D slice of a complex 3D architecture. Variations in slice orientation, cutting angle, or tissue shrinkage could alter the apparent fibrosis percentage and introduce measurement noise.
- Thresholding and segmentation assumptions. The binary classification of white (fibrosis) and black (healthy lung) assumes uniform staining and imaging conditions. Minor lighting or staining differences can change pixel intensity distributions, affecting computed fibrosis percentages.

Future work-

- If given more time and resources, several improvements could strengthen both the analysis and its translational value
- Expand dataset size and diversity. Analyze additional lung samples from multiple animals and disease stages to capture biological variability and ensure that trends are statistically robust.
- Integrate 3D reconstruction. Combine serial 2D slices into a volumetric 3D fibrosis map using image registration and volumetric interpolation to better visualize and quantify spatial gradients throughout the lung.

NOTES FROM YOUR TEAM:

- We both separately created code for generating the .csv file. After communicating as a team, we decided on one version to submit. We worked together to divide the work evenly. We both had issues working with GitHub, so to fix that, we created a shared document to collaborate virtually. We also planned out future work, including who will do what for the final notebook.

- We had to work with TAs to resolve issues when running the code. We both had different code still and worked to help understand the functions and picked the more clear version that we had, we worked on dividing work on the verifying analysis, conclusions and ethical implications, and limitations and future work. We also discussed plans of action and timeline of the rest of the project as well as how we should expect to work together while on break in the case we need to do work during that time.

QUESTIONS FOR YOUR TA:

We have no questions for a TA right now