Module_2

Team Members:

Ana, Marco

Project Title:

Fibrosis in the lung at different biopsy depths

Project Goal:

This project seeks to explore how the extent of lung fibrosis changes at different depths in the lung by analyzing black-and-white biopsy images and using interpolation to predict fibrosis levels, helping us learn how computational tools can be applied to real biomedical problems.

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).

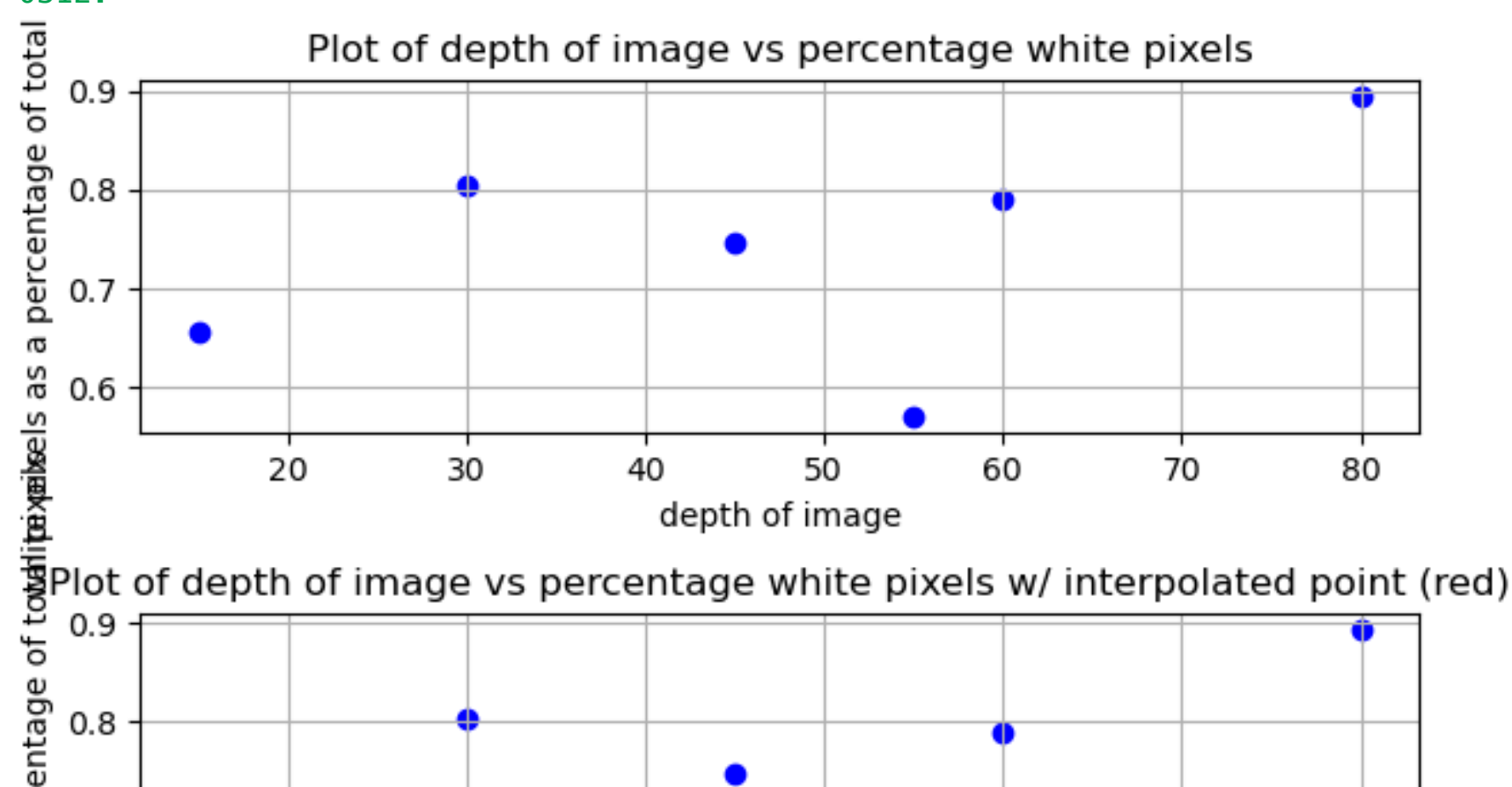
- Prevalence & incidence:
    - Incidence: Around 9 to 18 new cases per 100000 person-years are estimated in the US.
    - Prevalence: Around 10 to 50 per 100000 people have been diagnosed with Lung fibrosis in the US.
- Risk factors (genetic, lifestyle):
    - Genetic: - telomere-related Genes: shortened telomeres cause premature cell senescence and defective lung epithelial repair. - MUC5B gene variant: most common risk factor. Overexpression of this gene can lead to lack of mucus clearance, causing scarring. - Surfactant protein gene: mutations affect surfactant protein production and folding, which lead to damage in the alveolar epithelial cells in the lungs.
    - Lifestyle: - Smoking: retired and non retired smokers have a 2-3 times higher risk for developing IPF. Increases disease rate of mortality and worsens its progression. - Environmental Exposures: long term exposure to metal dusts, wood dusts, stone dusts and exposure to other types of irritants cause repeated alveolar injury leading to Fibrosis. - Viral infections: Some viral infections are suspected to trigger or accelerate Fibrosis, due to chronic inflammation. - Gatroesophacal Reflux disease: Micro-aspiration of stomach acid in the lungs repeatedly can injure the alveoli. - Aging: Due to cell senscence and decreased ability to regenrate alveolar epithelium.
- Symptoms
    - Shortness of breath: Most common symptom. At the beginning it only happens when excercising, but later it happens even at rest.
    - Dry cough: Long lasting, no mucus cough. Gets worse over time.
    - Fatigue: Fatigue and weakness is caused by lack of oxygen reaching tissues.
    - clubbing: Approximately 50% of IPF patients present rounded and enlarged fingertips due to low oxygen levels.
    - crackles: When a patient's lungs open up the stiffness and scarring cause a crackling sounds that doctors can hear with a stethoscope.
    - weight loss: low appetite and energy cause unintentional weight loss.
- Standard of care treatment(s):
    - Pulmonary rehabilitation: Supervised excercise, education and the correct information to help a patient eat correctly, quit smoking and rearrange their lifestyle to keep their symptoms under control.
    - Medication: Immunosuppressants and steroids are commonly used to treat IPF either through oral or IV form. Two anti-scarring drugs, pirfenidone and nintedanib, are now available to treat this disease. Pirfenidone slows down the scarring, while nintenadib blocks some enxymes that cause scarring.
    - supplemental oxygen: Receiving oxygen through a tube or mask can help patients breathe easier and sleep better. Used when blood oxygen is low.
    - lung transplant: When medications and other treatments do not work anymore. Lung transplant becomes an option and it can greatly improve survival and quality of life.
- Biological mechanisms (anatomy, organ physiology, cell & molecular physiology):
    - Anatomy: The lungs contain alveoli, tiny air sac, that are sorrounded by capillaries where gas exchange occurs. In fibrosis, alveolar walls become stiff and scarred, therefore gas exchange efficiency drops. This causes shortness of breath and low oxygen levels.
    - Organ Physiology: Fibrotic tissue is stiff, harder for alveoli to expand. Blood flow and air flow no longer match efficiently.
    - cell and molecular physiology: repeated alveolar injury causes abnormal wound healing. Persistent myofibroblasts and excess extracellular matrix, result in chronic scarring and loss of lung structure.

Data-Set:

The dataset we are analyzing includes black-and-white images of mouse lung tissue that show different levels of fibrosis at various depths in the lung. Each image comes from a thin lung section that was stained for the protein desmin, which helps highlight myofibroblasts—the cells that make scar tissue. After staining, the images were turned into black-and-white masks so the white pixels represent fibrotic (scarred) areas and the black pixels show healthy tissue. These images were collected from a Bleomycin-induced lung injury model, which is commonly used to study idiopathic pulmonary fibrosis (IPF) in mice. The “Filenames and Depths for Students.csv” file lists the depth (in micrometers) for each image, which gives us the depths we need for the code.

In [1]:
```python
## Data Analyis:
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/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010017.jpg",
    r"/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010018.jpg",
    r"/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010019.jpg",
    r"/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010021.jpg",
    r"/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010022.jpg",
    r"/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010023.jpg",
            ]

# Enter the depth of each image (in the same order that the images are listed abo

depths = [
    15,
    30,
    45,
    55,
    60,
    80
    ]

# 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 b

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

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 per

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
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'''

#interpolation

interpolate_depth = float(input(colored("Enter the depth at which you want to int

x = depths
y = white_percents

i = interp1d(x, y, kind='linear')
interpolate_point = i(interpolate_depth)
print(colored(f'The interpolated point is at the x-coordinate {interpolate_depth}

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

#Two plots: one with the interpolated point and the other with just the data we o

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')
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/ interpolat
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)-


# 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:
/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010017.jpg:
0.6571054458618164% White | Depth: 15 microns

/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010018.jpg:
0.8045673370361328% White | Depth: 30 microns

/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010019.jpg:
0.7469892501831055% White | Depth: 45 microns

/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010021.jpg:
0.5698204040527344% White | Depth: 55 microns

/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010022.jpg:
0.7903814315795898% White | Depth: 60 microns

/Users/anamorales/Desktop/coding 1/MASK_Sk658 Llobe ch010023.jpg:
0.8942604064941406% White | Depth: 80 microns

CSV file 'Percent_White_Pixels.csv' has been created.
The interpolated point is at the x-coordinate 47.0 and y-coordinate 0.711555480957
0312.
```



Verify and validate your analysis:

To verify our analysis, I checked that my code correctly counted the black and white pixels in each image and that the percent white values made sense based on what I saw visually. The graph showed a reasonable pattern where fibrosis levels changed with lung depth, and the interpolated point fit smoothly between the data. To validate my results, I compared them to research on Bleomycin-induced lung fibrosis, which shows fibrosis can increase with depth, supporting that my results were realistic.

Conclusions and Ethical Implications:

From this project, we learned that interpolation is a useful tool for predicting data points within the range of values we already have. When we tried to plot a point outside the dataset, we realized that it wasn’t possible because that would be extrapolation, which goes beyond the data and can lead to unreliable results. Understanding this difference helped me see why it’s important to stay within the data range when analyzing this type of information. Ethically, it matters because making predictions without real data could lead to misleading conclusions, especially in health-related research where accuracy is critical.

Limitations and Future Work:

One limitation of this project was that the data only covered a small range of depths, so we could only interpolate within that limited range. This made it harder to see the full pattern of how fibrosis changes throughout the lung. For future work, collecting images from a wider range of depths would help give a clearer picture and make our analysis more accurate and complete.

NOTES FROM YOUR TEAM:

- On the first day of Module 2, we introduced ourselves and divided the tasks. Ana started researching the background of the disease, while Marco worked on understanding the code. Once we finished the research, we began analyzing what we were asked to do.
- We ran into a problem with our code but were able to fix it by Thursday, October 9.
- After analyzing the dataset, we discussed our project’s limitations and what we would plan to do if we had the opportunity to expand our research.
- Finally, after sharing and understanding each other’s points of view, we completed the remaining sections together.

QUESTIONS FOR YOUR TA:

- Is there any way we tell if our interpolation results are accurate beyond just looking at the graph?