Name : Tazmeen Afroz **Roll No:22P-9252** **Class & Section:BAI-7A**

Assignment - 5

Dataset Description:

The C-NMC 2019 dataset, also known as the ALL Challenge dataset of ISBI 2019, focuses on classifying microscopic images of cells to identify acute lymphoblastic leukemia (ALL), the most common childhood cancer affecting about 25% of pediatric cases. It contains 15,135 segmented cell images from 118 patients, labeled as either normal cells or leukemia blasts by an expert oncologist, including real-world artifacts like staining noise despite corrections. This binary classification dataset is designed for developing models to distinguish immature leukemic blasts from normal cells.

Dataset Visualization:

Code:

```
all_paths =
glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_0/all/*.bmp')
+ \

glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_1/all/*.bmp')
+ \

glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_2/all/*.bmp')

hem_paths =
glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_0/hem/*.bmp')
+ \

glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_1/hem/*.bmp')
+ \

glob.glob('../input/leukemia-classification/C-NMC_Leukemia/training_data/fold_2/hem/*.bmp')

print("Found Diseased:", len(all_paths))
print("Found Healthy:", len(hem_paths))
```

Output:

```
Found Diseased: 7272
Found Healthy: 3389
```

```
sample_all = random.sample(all_paths, 4)
sample_hem = random.sample(hem_paths, 4)

samples = sample_all + sample_hem
labels  = ["Diseased"]*4 + ["Healthy"]*4

fig, ax = plt.subplots(2, 4, figsize=(16, 8))
fig.suptitle("Leukemia Cell Samples ", fontsize=20)

idx = 0
for i in range(2):
   for j in range(4):
```

```
        img = imread(samples[idx])

        ax[i, j].imshow(img)
        ax[i, j].axis("off")
        ax[i, j].set_title(labels[idx], fontsize=13, fontweight="bold")

        idx += 1

plt.tight_layout()
plt.show()
```

Visualization: