AstraZeneca AI Challenge:

Description / Background-

Despite breaching into Web and collecting information for analysis; it's always a bottleneck to process the data wrapped in images of all sorts because underline it's nothing but just pixel values. So, we rather decide to dump it in unstructured space. Alert - That pile is building up

Let's talk about data analysis, we generally build a story by visualizing data through scatterplot, line graph, histogram, Kaplan-Meier charts etc. According to recent trend, in scientific literature this analysis is performing with the rate of approx. 110k/yr. isn't it interesting if we could somehow reverse engineer the plots images automatically to extract the raw data values from it!!

Challenge:-

Kaplan Meier Chart is the visual representation of depreciating or appreciate probability of an event with respect to time interval. It is generally used in survivorship analysis. In medical research, it is often used to measure the quantifiable amount of patient living for a certain period after treatment.

We, AstraZeneca Team challenges you to reverse engineer Kaplan-Meier (KM) plots from scientific literature and digitize them into probabilistic raw data points. We are providing you a set of sample KM curves to kick start.

After exploring through KM plots from multiple scientific literature, you can take assumptions such as background of plots will always be "white" in images.

Evaluation Criteria: -

Handling of below variances will decide calculation of result: -

- Undefined no. of colors shades is possible in Kaplan-Meier curve.
- Undefined no. of groups/Lines(K) could be possible in Kaplan-Meier curve (where K<10)
- Model/Algorithm should produce multiple csv/excel sheet for each group/line respectively.
- No user-input should be given while after running the model/algorithm i.e. solution should be completely automatic.

- Preferred: - Dashed Lines in Kaplan-Meier curve.

Hint: - Don't google it. We already did that! 😉

Result Submission:

- Input – Images (.png,.jpg,.jpeg,.tiff)

  Sample images link for which code can be tested:-

    1. Image 1
    2. Image 2
    3. Image 3
    4. Image 4
    5. Image 5

- Output –
    - Excel/CSV files for each group/line present in Kaplan-Meier curve. Each output file will consist of two column- "X-axis-values" and their respective "Y-axis-values".
    - A draft (word doc/ppt) briefly explaining the approach and milestone points in your algorithm.