**Data Analysis:**

**English Data:**

* Modern English:
  + Images in .png format.
  + Preprocess needed:
    - Crop the images to extract handwriting.
* Historical English (Do not have, do we need?)

**Hebrew Data:**

* Modern Hebrew:
  + Images in .tif and .png format.
  + .tif Preprocess needed:
    - Conversion to .png format
    - Rotation of 180deg.
    - Crop the images to extract handwriting.
    - Remove yellow lines.
  + .pdf Preprocess needed:
    - Conversion to .png format
    - Rotation of 180deg.
    - Crop the images to extract handwriting.
    - Remove yellow lines.
* Cairo Geniza:
  + Images in .tif and .png format
  + Preprocess needed:
    - Crop the images to extract handwriting.
* Middle Ages:
  + Images in .jpg format.
  + Preprocess needed:
    - Crop the images to extract handwriting.

**Arabic Data:**

* Modern Arab:
  + Images in .tif format.
  + Preprocess needed:
    - Conversion to .png format.
* Historical Arab:
  + Images in .jpg format.

**Latin Data:**

* Historical Latin:
  + Images in .jpg format.