The train repo contains QA pairs and images extracted from two textbooks and crawled from the PEIR digital library website. The val and test repo contain QA pairs and images for validation, all images are crawled from the PEIR digital library website.

The QA pairs are in the json format with the first column corresponds to image names, the second column contains questions, and the third column contains answers.

The copyright of images used in this dataset belongs to the ‘Textbook of Pathology’ by Harsh Mohan, the ‘Basic Pathology’ by Purchase Robbins, and the PEIR digital library website.