

# Prompt2SegCXR: Prompt to Segment all Organs and Diseases in Chest X-rays.

---

**Prompt2SegCXR** is designed to generate segmentation masks for multiple organs and diseases in chest X-rays using doodle prompts. These doodles indicate the probable location of the target regions, enabling precise and efficient segmentation.

Our dataset combines chest X-rays and segmentation masks from three renowned sources:

1. **Japanese Society of Radiological Technology (JSRT) dataset**
2. **VinDr-RibCXR dataset**
3. **Shenzhen Hospital (China) dataset**

Below are the details of these datasets and how we curated them:

---

## Dataset Details

---

### 1. JSRT Dataset

- **Description:** The JSRT dataset provides chest X-ray images without segmentation masks ([link](#)).
  - **Segmentation Masks:** Generated using organ-wise annotations from Ginneken et al. ([link](#))
  - **Selection:** 150 chest X-ray images covering these organ classes:
    - Right Clavicle, Heart, Right Lung, Left Clavicle, and Left Lung.
- 

### 2. VinDr-RibCXR Dataset

- **Description:** This dataset includes both chest X-rays and segmentation masks for ribs ([link](#)).
  - **Selection:** 100 chest X-ray images with their corresponding rib masks.
- 

### 3. Shenzhen Hospital Dataset

- **Description:** Provides disease-specific segmentation masks for lung-related conditions ([link](#)).
  - **Selection:** Up to 50 samples per disease class, spanning a range of common conditions.
- 

## Combined Dataset Overview

---

Our final dataset contains **23 classes**, including **6 organ classes** and **17 disease classes**:

### Organ Classes

1. Right Clavicle
2. Right Lung
3. Left Clavicle
4. Left Lung
5. Rib

### Disease Classes

1. Apical Thickening
  2. Calcified Lymph Node
  3. Calcified Nodule
  4. Cavity
  5. Pleural Effusion
  6. Miliary
  7. Severe Infiltrate (Consolidation)
  8. Linear Density
  9. Small Infiltrate (Non-linear)
  10. Retraction
  11. Clustered Nodule (2mm–5mm)
  12. Adenopathy
  13. Pleural Thickening (Non-apical)
  14. Calcification (Other than Nodule or Lymph Node)
  15. Moderate Infiltrate (Non-linear)
  16. Thickening of Interlobar Fissure
  17. Single Nodule (Non-calcified)
-

# Doodle Prompts

---

- **Created by:** Medical experts from our team.
  - **Tools Used:** Microsoft Image Viewer for annotations.
  - **Purpose:** Highlight regions of interest related to organs and diseases.
  - **Format:**
    - Target pixels set to class-specific values.
    - All other pixels set to zero for clarity.
  - **Metadata:** Ensures accurate mapping to original X-rays and segmentation masks.
- 

## Public Availability and Copyright

---

- **Published by Us:** Only the created doodle prompts.
  - **Additional Requirements:** To train and evaluate the model, you must collect the corresponding chest X-ray images and segmentation masks which are also publicly available or will be available by request to their publisher.
- 

## File Naming Conventions

---

### 1. Shenzhen Dataset

- **Filenames:**
  - Prompt files: CHNCXR\_0330\_1\_Adenopathy\_9.png
  - Mask files: CHNCXR\_0330\_1\_Adenopathy\_9.png
  - Chest X-ray: CHNCXR\_0330\_1.png
- **Special Case:** For prompts containing `Calcification_(other_than_nodule_lymphnod)`, replace `_lymphnod` with `&lymphnod` in the corresponding mask filename.
  - Example:
    - Prompt: CHNCXR\_0367\_1\_Calcification\_(other\_than\_nodule\_lymphnod)\_2.png
    - Mask: CHNCXR\_0367\_1\_Calcification\_(other\_than\_nodule&lymphnod)\_2.png

### 2. VinDr-RibCXR Dataset

- **Filenames:** All filenames for prompts, masks, and X-rays are identical.
  - Example:
    - VinDr\_RibCXR\_train\_000.png

### 3. JSRT Dataset

- **Filenames:**
    - Prompt filenames indicate the organ using abbreviations:
      - RC (Right Clavicle), H (Heart), RL (Right Lung), LC (Left Clavicle), LL (Left Lung).
    - Masks for all organs are combined in a single file and need to be separated.
    - Example:
      - Prompt: JPCLN001\_H.png
      - Chest X-ray: JPCLN001.png
      - Mask: JPCLN001.png (Separate by organ, e.g., JPCLN001\_H.png).
- 

We hope this documentation provides clarity for using the **Prompt2SegCXR** dataset. Happy experimenting!

---