**Title: Lung segmentation ?**

**Introduction**

Image identification tasks in machine learning can broadly be divided into three main categories; classification, segmentation and detection. In image classification, the model identifies what is in the image for a given object (such as a cat or a dog), in image detection the model identifies *where* in the image the object is with a bounding box around the object, while in a segmentation task each individual pixel belonging to the object is identified. Hence image segmentation, where automatic detection of boundaries within images is performed, is a more advanced task compared to image classification or detection. In this report, we train a U-Net model to segment human lungs based on data obtained from two open source datasets of chest X-rays.

The chest X-ray is the most common radiological test performed worldwide and accounts for 25% of all diagnostic imaging techniques (1, 2). It allows for an evaluation of the airways, mediastinum, the heart and pleura and chest wall. The indications for performing a chest X-ray are broad and include for asymptomatic people undergoing health screening, as part of investigations for sepsis even in the absence of respiratory symptoms, detecting respiratory disease and for patients who are critically unwell or medically unstable (3). Hence the chest X-ray is a frequently performed diagnostic investigation which can provide useful information for many clinical conditions, at a relatively low cost and with minimal risk to the patient per examination (4). As a result, two datasets, known as the Montgomery County and Shenzhen chest X-ray sets (5), were made available by the United States Library of Medicine in an effort to provide data for the scientific community for training on chest X-rays.

The Montgomery County chest X-ray set was obtained in collaboration with the department of health and human services in Montgomery county in the United States (5). All chest X-rays were obtained as part of the counties tuberculosis screening programme and consists of 138 frontal chest X-rays. Of these, 80 are normal studies and 58 have manifestations of tuberculosis. A clinical reading of the chest-X ray is included for each case. The second dataset, the Shenzhen X-ray set, was obtained through collaboration with Shenzhen People’s hospital in Shenzhen, China (5). Chest X-rays were performed over a one month period as part of routine outpatient clinics. The dataset consists of 662 frontal chest X-rays, of which 326 have no pathology (normal) and 336 have features of tuberculosis. As for the Montgomery dataset, ground truth clinical labelling is available for all images.

Given the volume of chest X-rays performed on a daily basis in both community and hospital care, clinicians are faced with a high volume of images to review (6). Furthermore, conditions such as tuberculosis, which is a significant global health burden, are screened for using chest X-rays, hence efforts to utilise machine learning and artificial intelligence (AI) in the triage and diagnose of chest X-rays have emerged (7, 8) . Results to date have been promising with AI algorithms outperforming radiologists in the detection of conditions such as tuberculosis (7). For many of these studies, an initial task is segmenting the lungs. A collaboration between the Radiological Society of North America (RSNA) and Kaggle, as well as the US National Institute of Heath and MD.ai developed the RSNA Pneumonia detection challenge. The challenge was to build an algorithm for the detection of pneumonia from the above mentioned open source datasets. In this paper, we use a Kaggle notebook developed as part of the RSNA pneumonia challenge to train a U-net lung segmentation task.

**Team assignment author contributions:**

Study design: All authors

Introduction: Helena Bartels

Methods: Ian McBride

Results: Sidharth Mehra, Sicheng Zhoa

Discussion and future improvements: Tianxia Song, Pengcheng Ren

Manuscript review: Helena Bartels, Sidharth Mehra

Video presentation: All authors

**References**

1. de Stoppelaar SF, Pereverzeva L, Hafkamp B, Lips N, Tielbeke F, Rustenburg L, et al. Diagnostic Value of Chest X-Ray in Patients With Suspected Infection and No Respiratory Signs or Symptoms. Open Forum Infectious Diseases. 2020;7(6).

2. United Nations Scientific Committee on the Effects of Atomic Radiation. Medical radiation exposures. United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR 2000 Report to the General Assembly.

3. Speets AM, van der Graaf Y, Hoes AW, Kalmijn S, Sachs AP, Rutten MJ, et al. Chest radiography in general practice: indications, diagnostic yield and consequences for patient management. Br J Gen Pract. 2006;56(529):574-8.

4. Bradley SH, Hatton NLF, Aslam R, Bhartia B, Callister ME, Kennedy MP, et al. Estimating lung cancer risk from chest X-ray and symptoms: a prospective cohort study. Br J Gen Pract. 2021;71(705):e280-e6.

5. Jaeger S, Candemir S, Antani S, Wáng YX, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg. 2014;4(6):475-7.

6. approaches. CritdsocWragop, World Health Organization G.

7. Qin ZZ, Ahmed S, Sarker MS, Paul K, Adel ASS, Naheyan T, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. Lancet Digit Health. 2021;3(9):e543-e54.

8. Çallı E, Sogancioglu E, van Ginneken B, van Leeuwen KG, Murphy K. Deep learning for chest X-ray analysis: A survey. Med Image Anal. 2021;72:102125.