Deep learning pipeline for automated detection and classification of central canal, lateral recess and neural foraminal stenosis on lumbar spine MRI

Presented by: Tan Jiong Hao Jonathan

James Thomas Patrick Decourcy Hallinan, Lei Zhu, Kaiyuan Yang, Andrew Makmur, Diyaa Abdul Rauf Algazwi, Yee Liang Thian, Samuel Lau, Yun Song Choo, Eide Sterling, Qai Ven Yap, Yiong Huak Chan, Naresh Kumar, Beng Chin Ooi, Hiroshi Yoshioka, Swee Tian Quek

National University Hospital Singapore, Yong Loo Lin School of Medicine Singapore, NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Department of Computer Science, School of Computing, National University of Singapore

National University Hospital

Introduction

- Lumbar spine MRIs are essential in the assessment of Lumbar spinal stenosis for accurate evaluation of the central canal, lateral recesses, and neural foramina.

- Detailing the degree of stenosis at each level can be time consuming and repetitive.

National University Hospital

Aim of study

- To develop a deep learning pipeline (DLP) to automatically detect and classify central canal, lateral recess, and neural foraminal stenosis in the lumbar spine using axial and sagittal MR sequences.

National University Hospital

Study Design

- 446 MRI Lumbar spine studies were analyzed, encompassing 12403 axial T2-weighted and 6161 sagittal T1-weighted images.
  - Studies with instrumentation or severe scoliosis were excluded

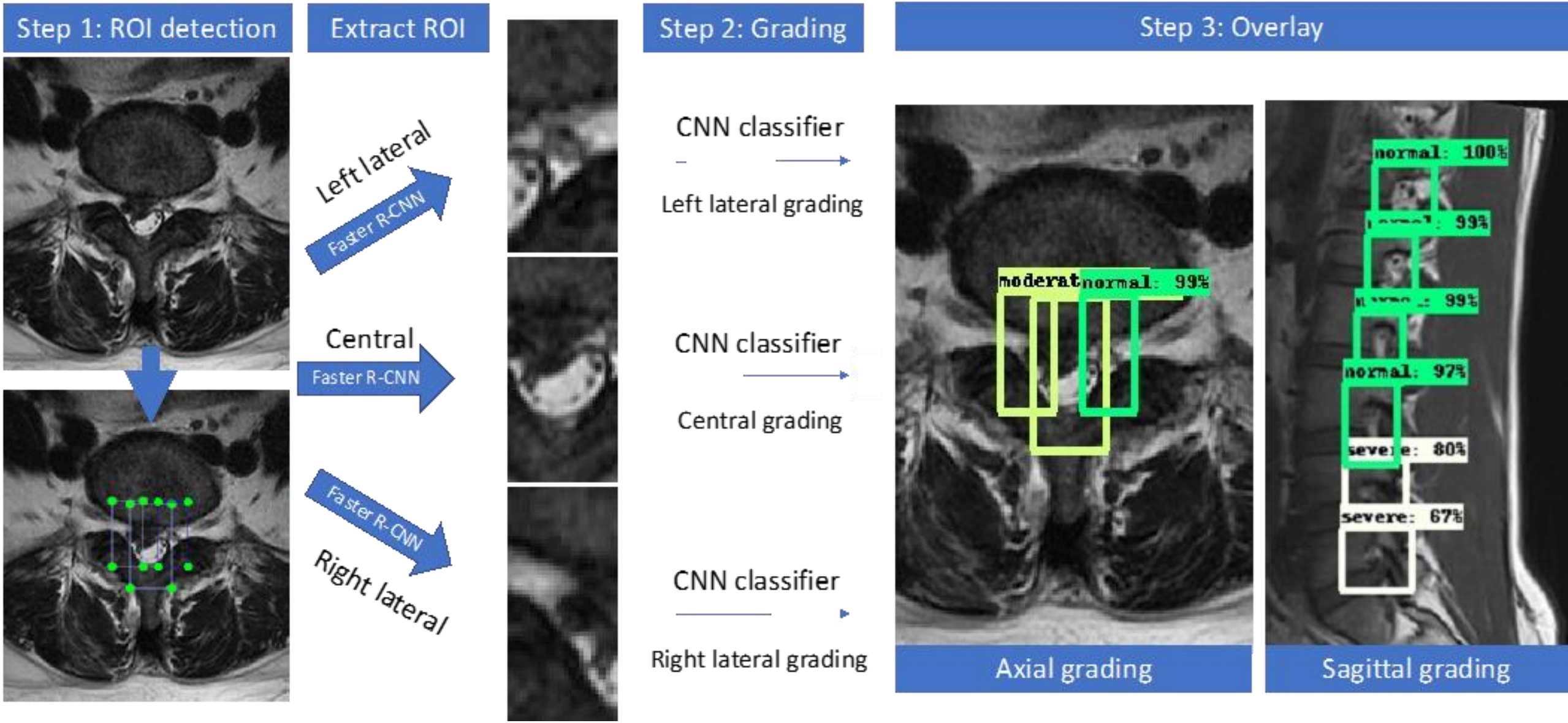
- Training/testing split was 89/11%
- Training data were labelled by 4 board-certified radiologists using pre-defined gradings (normal/mild/moderate/severe).

National University Hospital

Study Design

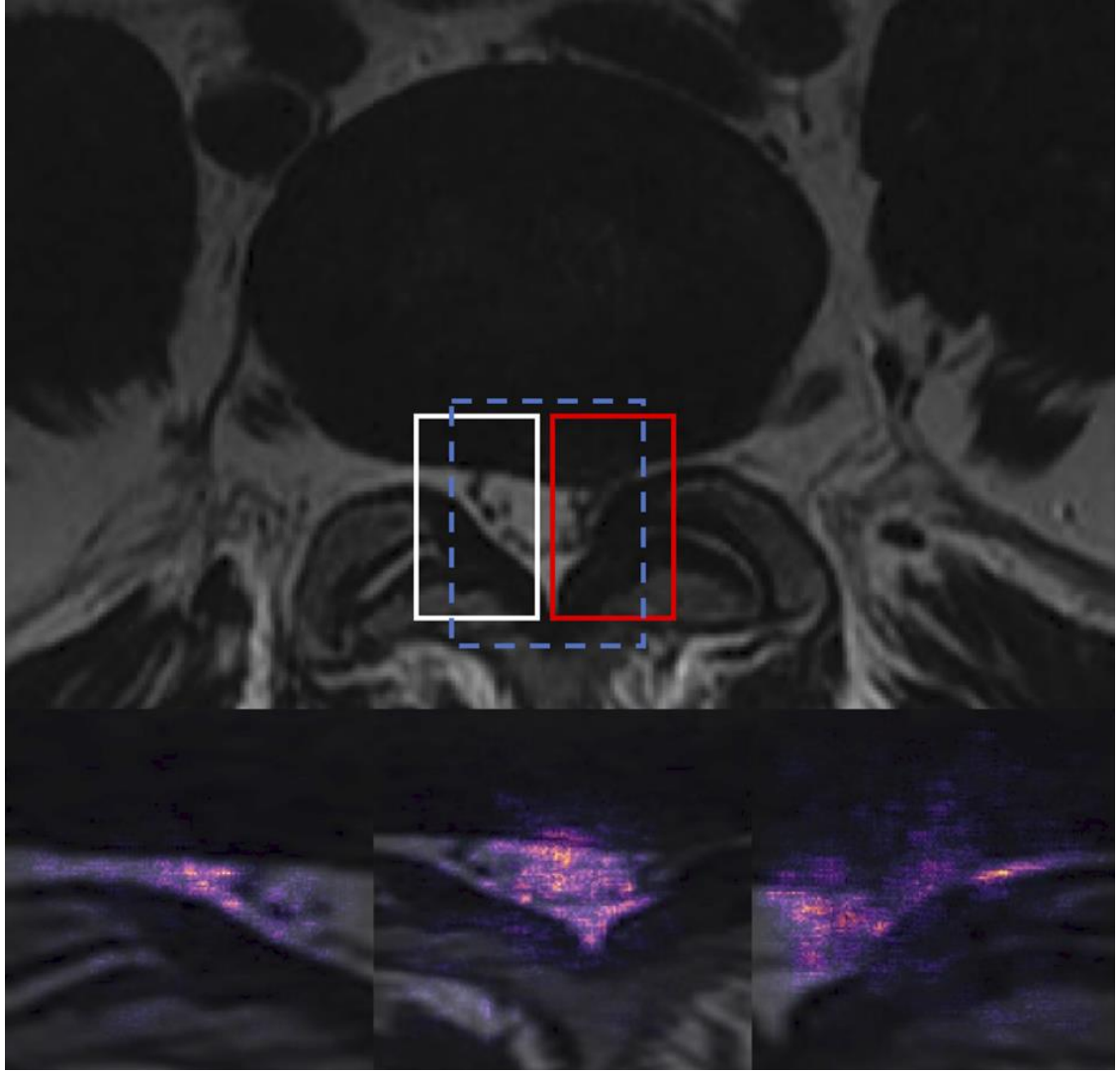
- A convolutional neural network (CNN) was trained to detect the region of interest (ROI)
- A second CNN was then trained for ROI classification.

National University Hospital



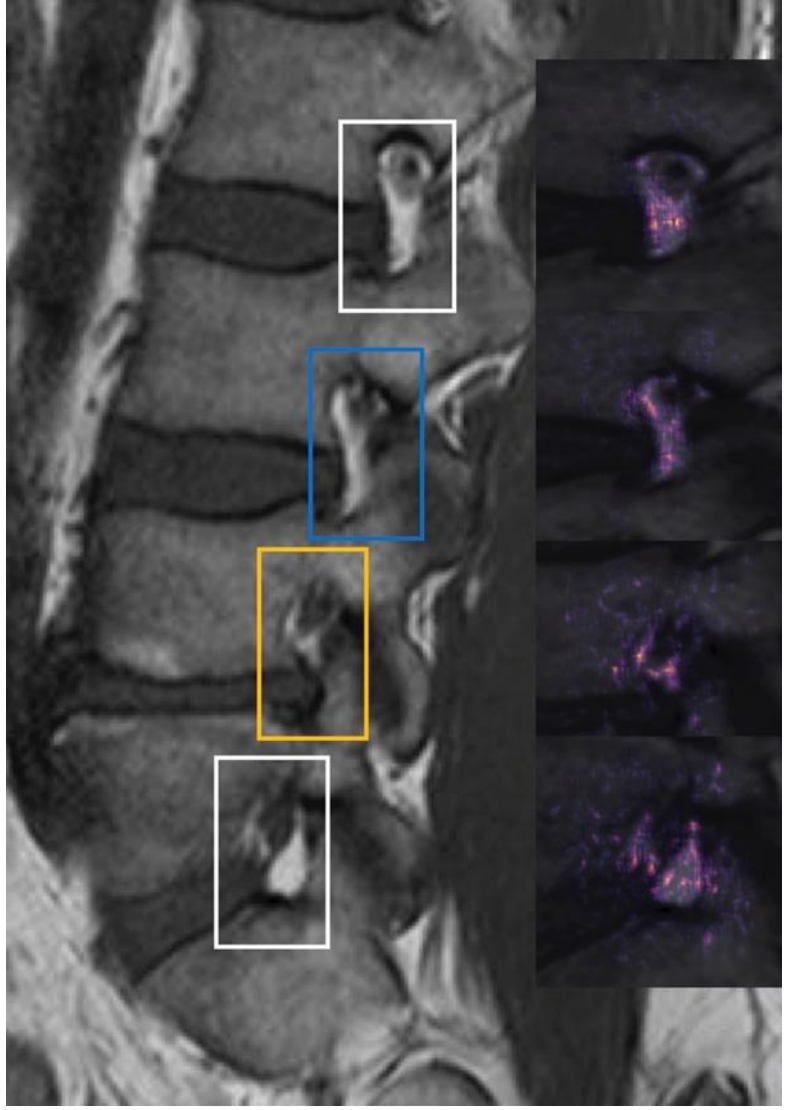
Two-step deep learning pipeline development. Region of interest (ROI) detection (central canal and lateral recesses shown) on the axial T2-weighted images, followed by classification of the ROI into normal, mild, moderate or severe stenosis. A two-step deep learning pipeline was also used for neural foraminal stenosis detection and classification on sagittal T1-weighted images. The percentage scores next to the grading represent the probability of the predicted class.

National University Hospital



Integrated Gradient was applied to the deep learning pipeline and highlights the axial T2-weighted MRI regions that predict the classification result. There is a normal (white box) right lateral recess, mild (blue dashed) central canal stenosis and severe (red) left lateral recess stenosis. Corresponding ‘heatmaps’ show that the activation of the deep learning pipeline is mainly at the high signal regions; Cerebrospinal fluid in the central canal and surrounding epidural fat.

National University Hospital



Integrated Gradient was applied to the deep learning pipeline and highlights the sagittal T1-weighted MRI regions that predict the classification result. Examples of normal (white box), mild (blue) and moderate (orange) neural foraminal stenoses are provided. Corresponding ‘heatmaps’ show that the activation of the deep learning pipeline is mainly at the high signal epidural fat surrounding the exiting nerve root (keyhole appearance).

Study Design

- A held-out test set of 50 spines were labelled by a musculoskeletal radiologist with 31-years-experience(reference standard), and subspecialist neuro(Rad1) and musculoskeletal(Rad2) radiologists with 5 and 9-years-experience, and compared to the performance of the DLP

- DLP performance on an external dataset of 100 spines was also evaluated.

- Detection recall(%), inter-rater agreement(Gwet's kappa) and sensitivity/specificity/PPV/NPV were calculated.

National University Hospital

Results

- DLP ROI detection recall ranged from 99.7-99.9%/95.2-99.3%/84.5-96.2% for central canal/lateral recesses/neural foramina respectively, which were comparable with subspecialist radiologists (range:83.9-99.9%).

- Internal dataset dichotomous DLP classification (normal/mild versus moderate/severe)
  - Almost-perfect agreement for rad1, rad2 and the average DLP
  - Kappas of 0.98/0.98/0.96 for central canal, 0.92/0.95/0.92 for lateral recesses, and 0.94/0.95/0.89 for neural foramina, respectively ($p<0.001$).

National University Hospital

Results

- External testing
  - Almost-perfect agreement for the DLP was seen for dichotomous classification of all ROIs (range:0.95-0.96).

- The DLP demonstrated high specificity (91.9-97.9%) and NPV (96.3-99.7%) for dichotomous classification of all ROIs.

National University Hospital

Strengths

- Generalizable results
  - Heterogeneous training dataset of lumbar spine MRI studies across a range of MRI scanners and vendors with variable imaging parameters.
  - Performance of our DLP on an external dataset showed similar performance to our internal dataset.

National University Hospital

Limitations

- Use of radiological grading scales as reference standard
  - Controversial

- Highly supervised labelling of images for model development
  - Time consuming and labor intensive
  - Limitation of number of studies that can be studied

National University Hospital

Conclusion

- The DLP showed comparable agreement to subspecialist radiologists for detection and classification of central canal and lateral recess stenosis, with slightly reduced agreement for neural foraminal stenosis on lumbar spine MRI.

- Further applications
  - Semi-automated reporting under radiologist supervision
  - Screening of MRI scans for prioritisation of early review
  - Detection and classification of other causes of cord compression e.g. spinal metastases

No personal/institutional conflicts of interest to disclose.

National University Hospital