SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

# Imaging Management of Lumbar Spine MRI Annotation in Machine Learning Model Validation with an Emphasis on Subcohort Variation

Veysel Kocaman, PhD, Gesund.ai; Sumir S. Patel, MD, MBA; Enes Hosgor, PhD

## Introduction/Background

Lumbar spine disorders, prevalent across various demographics, often require precise imaging analysis for accurate diagnosis and treatment planning. Magnetic Resonance Imaging (MRI) stands as the gold standard for visualizing spinal pathologies owing to its detailed soft tissue contrast. Although radiologists are predominantly trained in the clinical interpretation of lumbar spine MRIs, there exists a need to for radiologists to be adept in the annotation of these exams for machine learning training, validation, and monitoring. We aim to assess radiologist annotation to determine if automatic subcohort analysis of a dataset can improve performance.

## Methods/Intervention

For this study, three subspecialty trained attending radiologists each annotated 100 lumbar spine magnetic resonance imaging (MRI) examinations of patients across various clinical settings. The annotations included vertebral body height, vertebral body area, intervertebral disc area, and neuroforaminal area. Automatic analysis on a software platform (Gesund.ai) was used to identify relative underperformance of the radiologists with respect to clinical subcohorts. We analyzed multiple subcohorts and data annotation metrics to include Cohort Time Consumption, Gender and Institution Influence, and Most Time-Consuming Measurement.

## Results/Outcome

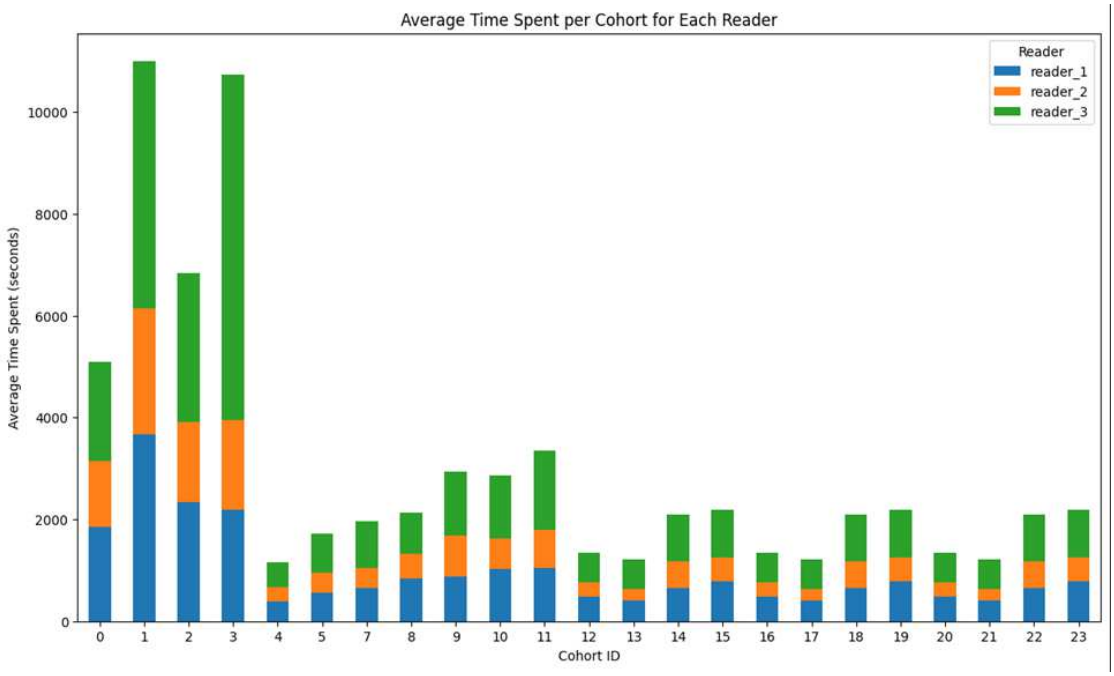
Cohorts 1 and 3 have significantly high average times, both from a Missouri institution with 'M' and 'M/F' genders, suggesting case nature or protocols affect processing times (Figure 1). Cohorts from two clinical sites, especially male and Missouri cohorts, indicate challenging diagnostic criteria. Patterns show certain geographical cohorts, particularly males, require more time, potentially reflecting complex anatomy or pathology. L1 and L4 vertebral bodies and area measurements are consistently time-consuming, suggesting a need for better training or tools (Figure 2). Older age groups (60-74) have higher times than younger ones due to age-related anatomical changes.

## Conclusion

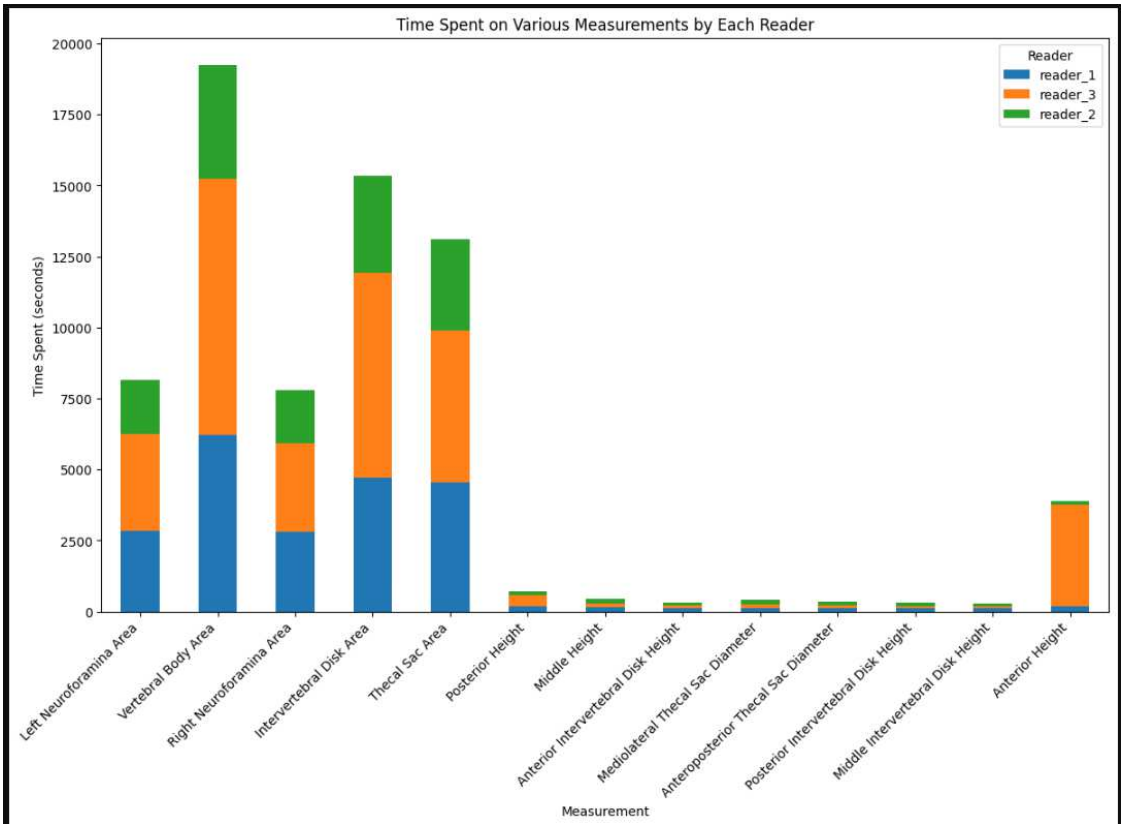
This study highlights the effectiveness of a sophisticated software platform designed to accustom radiologists to the type of annotation tasks required in modern machine learning. Our approach utilizes this software platform to automate the process of identifying deficiencies in the analysis of patient subcohorts. Once these deficiencies are identified, the system can tailor the training regimen by assigning additional cases from the same subcohort to the respective radiologists.

## Statement of Impact

Automated performance analysis with respect to clinical subcohorts has the potential to improve radiologist




Average time spent per cohort for each reader. Cohorts 1 and 3 have higher annotation times.




Time spent on various measurements per reader. Area measurements have higher annotation times.


## Keywords

Machine Learning; Validation; Radiologist; Subcohort; Training; MRI