

AI based automatic processing of full spine radiographs for clinical parameters quantification

Augustin Daridon (Palaiseau, France), Ayman Assi (Bayrut, Liban), Le Folgoc Loïc (Palaiseau, France), Cédric Barrey (Lyon, France), Mourad Ould-Slimane (Rouen, France), Marc Khalifé (Paris, France), Nolwen Lemonnier (Rouen, France), Emmanuelle Ferrero (Paris, France), Clément Jacquemin (Périgueux, France), Rami El Rachkidi (Bayrut, Liban), Soufiane Ghailane (Périgueux, France), Vincent Challier (Périgueux, France), Mohamad Karam (Bayrut, Liban), Wafa Skalli (Paris, France), Elsa Angelini (Palaiseau, France, France)

Introduction

Spinal radiograph annotation is widely used to compute spinal and pelvic parameters for diagnosis and surgical planning. However, manual or semi-automatic annotation remains time consuming and limits large scale clinical implementation. Although several AI-based methods have been published, few have undergone rigorous multicenter evaluation suitable for routine clinical use. The aim of this study was to evaluate the performance of an original AI-based pipeline for automatic landmark detection and computation of clinically relevant radiographic parameters.

Materials and Methods

A dataset of frontal and lateral full-spine radiographs (from the tip of odontoid to femoral heads) was collected, including asymptomatic spines (AS), adolescent idiopathic scoliosis (AIS), and adult spinal deformity (ASD) cases. Exclusion criteria included prior surgical instrumentation, L5 sacralization, and missing visible landmarks. expert-annotated radiographs were used to train a multi-step AI algorithm. An additional independent validation set (VS) of 127 radiograph pairs (36 AS, 59 AIS, 32 ASD) from three centers was used for evaluation. All X-rays were manually annotated by an experienced operator using SpineView software whose measurement uncertainty was previously reported in Champain S et al, 2006. Clinical parameters were computed from both AI-derived and manual landmarks using the same software. The analysis was focused on: T4–T12 kyphosis (TK), L1–S1 lordosis (LL), sacral slope (SS), pelvic tilt (PT), pelvic incidence (PI), odontoid-femoral head angle (OD-HA), and sagittal vertical axis (SVA). Agreement was assessed using Bland–Altman analysis and Intraclass correlation coefficient (ICC), using the classification from Koo TK et al, 2016 (0.5-0.75 moderate, 0.75-0.9 good, >0.9 excellent)

Results

Bland–Altman analysis revealed near-zero systematic bias for all parameters (range: -0.6° to $+0.6^{\circ}$). SD of differences was 0.6° for OD-HA, 2.5 mm for SVA, 2.9° for PT, and 4.9° for TK, all within reported reproducibility of manual measurements. ICC values ranged from 0.94 to 0.98 (Excellent agreement). Parameters dependent on S1 endplate detection showed higher variability: SD = 5.5° , 6.9° , and 7.5° for SS, LL, and PI with ICC values of 0.81, 0.86, and 0.79 (good agreement).

Conclusion

The proposed AI pipeline demonstrated good or excellent agreement across key sagittal parameters. Variability was higher for sacral endplate-dependent parameters (SS, LL, PI), particularly with marked deformity, although manual correction was feasible within a short time (approximately 1 mn) These findings support the potential of the system to streamline quantitative radiographic analysis in clinical practice. Further evaluation is ongoing for coronal X-Rays.