Penn Medicine
Neurosurgery

Development of a Novel Artificial Intelligence Model to Predict Post-Operative Correction in Adolescent Idiopathic Scoliosis Surgery

Mert Marcel Dagli, M.D.
Sharpe Postdoctoral Research Fellow & Incoming Neurosurgery Resident

International College of Surgeons
UNITED STATES SECTION

AANOS

86th Annual Surgical Update

International College of Surgeons - United States Section

April 10-12, 2025

48th Annual Scientific Meeting

American Academy of Neurological and Orthopaedic Surgeons

April 11-12, 2025

Author Disclosures

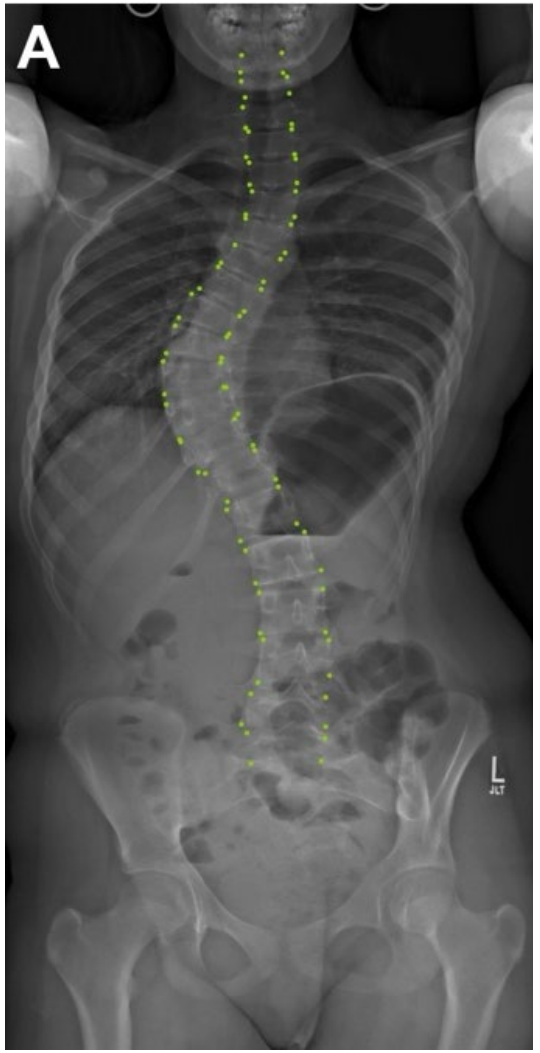
- Presenter:
- No financial disclosures

- Can AI tools predict radiographic correction success in AIS patients using common predictors?

- Development and Internal Validation of a Feasibility Model

Approach

- Development of Artificial Intelligence prediction model utilizing R21 AIS population
- Surgical outcomes can differ by patient anatomy and clinical factors
- Accurate models can aid in personalized treatment planning and patient education



Methods

- **Design & Setting:**
- Retrospective study using an institutional AIS registry (83 patients, age <18, single institution).
- **Inclusion Criteria:**
- AIS patients undergoing multilevel thoracolumbosacral PSF.
- Available pre- and post-operative upright whole-spine X-rays.
- **Exclusion Criteria:**
- Age ≥18 at time of surgery.
- Missing radiographic data.
- Surgery for tumor, trauma, or infection.
- **Outcome Definition:**
- Binary label: “Successful Correction” (≥75% TCA reduction) vs. “Less Successful Correction” (<75%).

Methods

- **Design & Setting:**
  - Retrospective study using an institutional AIS registry (83 patients, age <18, single institution).
- **Inclusion Criteria:**
  - AIS patients undergoing multilevel thoracolumbosacral PSF.
  - Available pre- and post-operative upright whole-spine X-rays.
- **Exclusion Criteria:**
  - Age ≥18 at time of surgery.
  - Missing radiographic data.
  - Surgery for tumor, trauma, or infection.
- **Outcome Definition:**
  - Binary label: “Successful Correction” (≥75% TCA reduction) vs. “Less Successful Correction” (<75%).

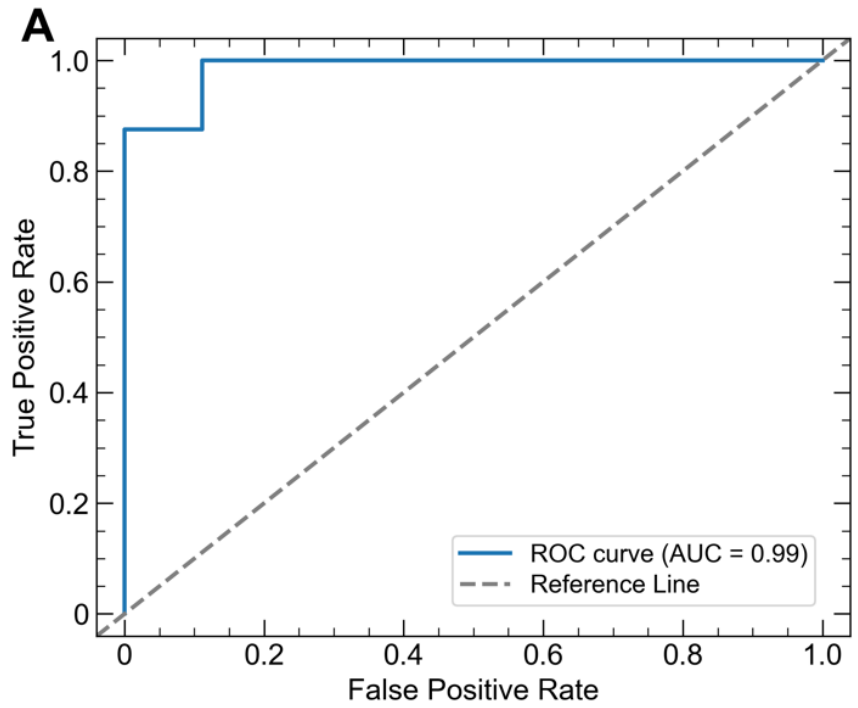
Methods

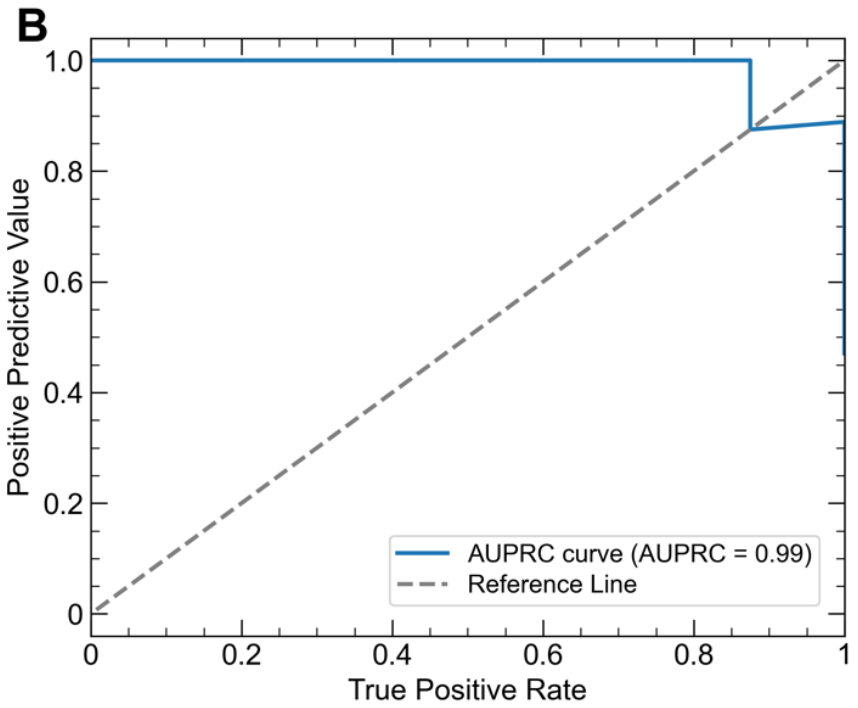
- **Machine Learning Approach:**
  - Compared multiple algorithms; XGBoost (extreme gradient boosting) emerged as the best performer.
- **Hyperparameter Tuning:**
  - Exhaustive grid search with 24 parameters → 3,840 model variants.
  - Optimized settings → considerations
- **Performance Metrics:**
  - Accuracy, Sensitivity (Recall), Precision, F1-score, AUC-ROC, AUPRC, Youden's Index.
- **Train-Test Split:**
  - 80:20 ratio from the registry (80% training; 20% validation) → considerations

Results

Variables	Overall (n=83)	Correction ≤75% (n=43)	Correction >75% (n=40)
Age	13 (11-14)	13 (11-14)	13 (11-14)
Female sex	61 (73.5%)	29 (72.5%)	32 (74.4%)
Race			
White	56 (67.5%)	27 (67.5%)	29 (67.4%)
Black	18 (21.7%)	10 (25%)	8 (18.6%)
Other	9 (10.8%)	3 (7.5%)	6 (14%)
Hispanic ethnicity	6 (7.2%)	2 (5%)	4 (9.3%)
BMI (kg/m2)			
Prior non-surgical intervention			
Radiographic measurements			
Right TC			
TCA (°)			
Left LC			
LCA (°)			

Parameter	Values
n_estimators	[50, 100, 200, 400]
learning_rate	[0.01, 0.05, 0.1, 0.2]
max_depth	[2, 3, 4, 5]
subsample	[0.5, 0.7, 0.8, 0.9, 1.0]





Variables
OR time (min)
PSF (levels)
PSO (levels)
EBL (mL)
Cell Saver Return (mL
pRBCs (mL)
Total RBCs (mL)
Intraoperative complic

Variables				
Radiographic measurements				
TC Direction				
TCA (°)				
TCA correction (°)				
Percentage T correction (%)	75.7 (63.9,82.7)	63.5 (59.3-70)	82.3 (79.2-86.5)	<.001
LC direction	82 (98.8)	-	1 (2.3)	.999
LCA (°)	13.0 (7.5,18.0)	16.5 (11.5-22.2)	9 (6-14)	<.001
Time until first physical activity (days)	0 (0-1)	0 (0-1)	0 (0-1)	.979
LOS (days)	3 (2-3)	3 (2-3)	3 (2-3)	.360
Postoperative complications	1 (1.2)	1 (2.5)	-	.482
SSI	1 (1.2)	1 (2.5)	-	.482

, 'binary:hinge']

	XGBoost
	0.941
	1.000
AUC-ROC	0.986
AUPRC	0.986
F1-score	0.941
Precision	0.888
Youden's index	0.663

Penn Medicine Neurosurgery

Take Aways

- **Feasibility Established** – An AI-driven model can predict post-operative correction in AIS.
- **Clinical Utility** – High sensitivity and specificity hold promise for improved patient counseling and surgical planning → considerations
- **Path Forward** – Larger, prospective, and multi-center studies needed to validate and generalize these findings
- **Considerations**
  - Cross-validation
  - Bayesian optimization vs grid search
  - Sample size -> ASD-AIS
  - Comparison vs. simpler models (f.e. LR)

Acknowledgements

- Mert Marcel Dagli
- Jang Yoon
- Ali Ozturk
- John Shin
- William C. Welch
- Yohannes Ghenbot
- Hasan Ahmad
- Daksh Chauhan
- Ryan Turlip
- Marie Kerr
- Patrick Cahill
- Jason Anari
- Kevin Bryan
- Jonathan Sussman
- Connor Wathen
- Beth Winkelstein
- Josh Golubovsky
- Daniel Yoshor
- Isaac Chen
- Bhargavi R. Budihal

Patients and Families