Multimodal Deep Learning Framework for PFT Estimation in Scoliosis Patients

Utilizing chest radiographs, radiographic measurements, and clinical information

Beom Su Kim, Wounsuk Rhee, Jeuk Lee, Ihn Seok Chae,
Hyoungmin Kim,
Bong-Soon Chang, and Sam Yeol Chang
Department of Orthopaedic Surgery

SNUH SEOUL NATIONAL UNIVERSITY HOSPITAL

Background, Clinical Need, and Study Objectives

Clinical problem

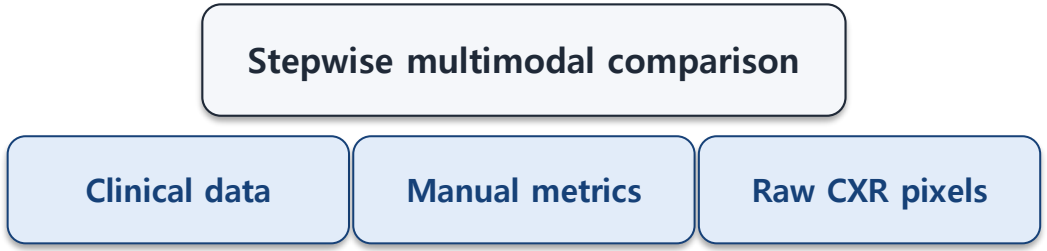
- Scoliosis-related thoracic deformity can restrict lung expansion and reduce chest wall compliance.
- PFT remains the reference standard, but may be difficult in pediatric, neuromuscular, cognitive, or physically impaired patients.
- Routine standing PA chest radiographs may contain latent structure-function information.

Study objectives

- Estimate continuous spirometric indices: FVC and FEV1.
- Identify restrictive ventilatory pattern using predicted FVC% (<80%).
- Quantify the incremental contribution of clinical variables, manual metrics, and raw CXR pixels.

Key novelty: structure-driven pulmonary dysfunction in scoliosis

Unlike prior CXR-based PFT estimation (e.g., Ueda et al., Lancet Digit Health 2024), this study isolates **structure-driven pulmonary dysfunction in scoliosis**, a defined model in which thoracic deformity is the primary driver of restrictive physiology.



Cohort Selection & Patient Characteristics

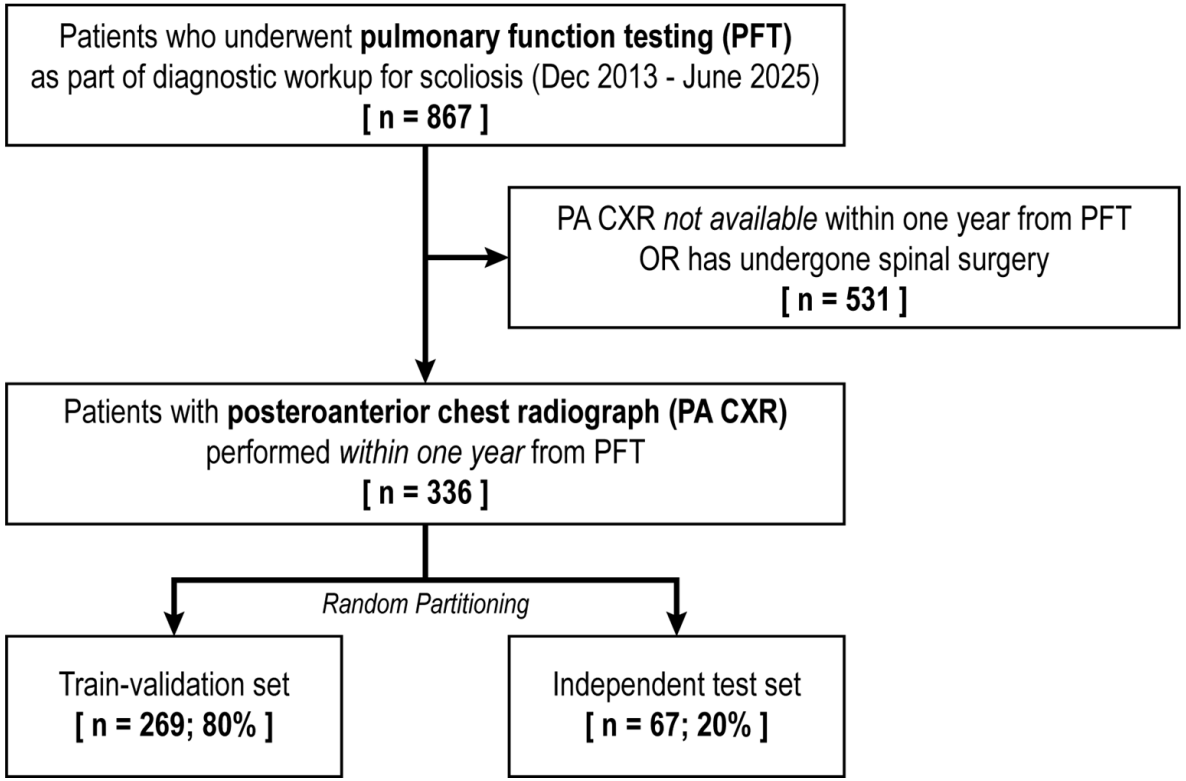


Figure 1. Cohort selection flowchart

Study cohort

867 screened → 336 included

Single tertiary care hospital

December 2013 – June 2025

Standing PA CXR within 1 year of PFT

Prior spinal surgery excluded

Final split

Train-validation: n=269 (80%)

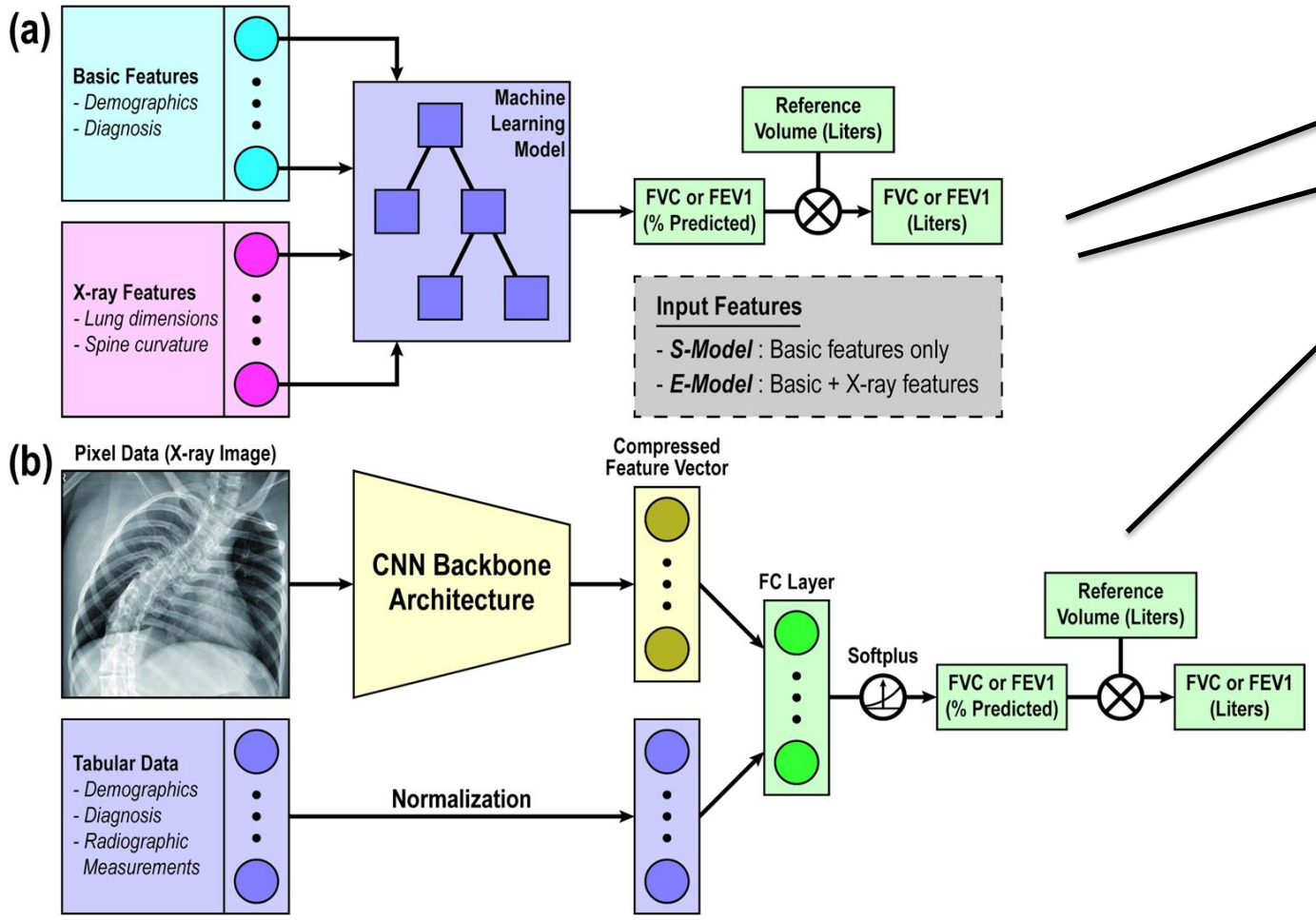
Independent test set: n=67 (20%)

Key Demographics (Train-Val / Test, p-value)

- Age: 31.1±24.1 / 30.9±24.3 yr (p=0.96) | Male: 34.6% / 47.8% (p=0.06)
- Height: 154.7±13.6 / 157.7±9.5 cm (p=0.09)
- **FVC / FEV1 distributions did not differ between sets (p=0.86 / 0.96)**
- **Right lung area & height: significant group diff. (p<0.005)**

Model Architecture Framework

Figure 2



Model	Input data	Algorithms
S-Model	Clinical variables	LR, SVM, RF, XGB, LGBM
E-Model	Clinical + manual imaging metrics	Same ML algorithms
M-Model	Clinical + manual + raw CXR pixels	CNN backbones + FC layers

Training strategy

- 5-fold cross-validation within train-validation set
- Independent test set held out throughout development
- Final predictions averaged across folds

*** Softplus activation ensured physiologically positive outputs**

Figure 2. Schematic illustration of S-, E-, and M-model designs

Regression Performance

Figure 3

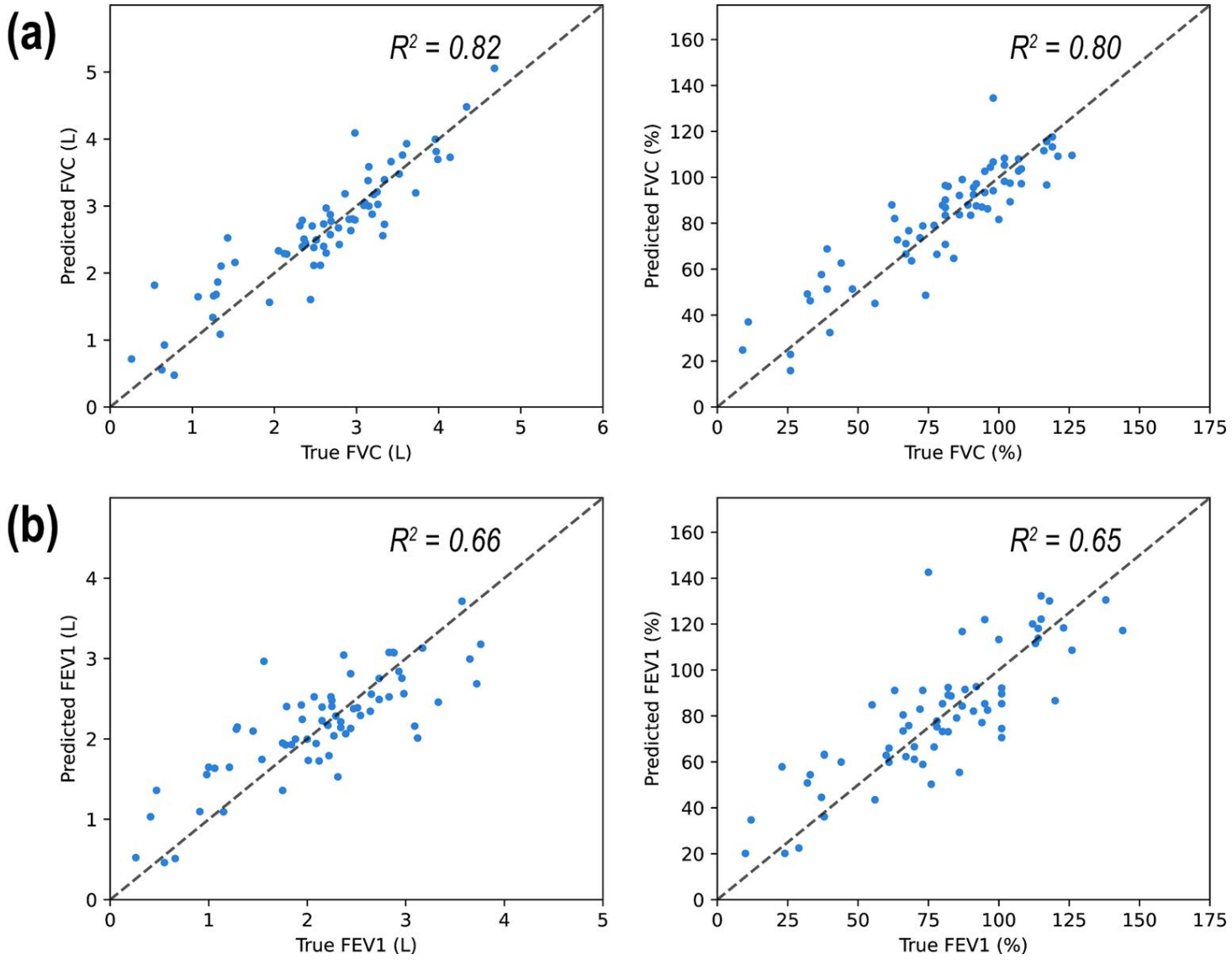


Figure 3. Scatter plots of FVC and FEV1 predictions on the independent test set

Key test-set results

FVC: MAE 0.31 L
RMSE 0.41 L, R^2 0.82

FEV1: MAE 0.37 L
RMSE 0.48 L, R^2 0.66

Selected M-Model Performance

Selected by smallest validation RMSE (§2.4)

- **FVC (EfficientNetV2-B3)**
  MAE 0.31L | RMSE 0.41L | R^2 0.82
  ≈ volume of one tidal breath
- **FEV1 (EfficientNetV2-B0)**
  MAE 0.37L | RMSE 0.48L | R^2 0.66

- FVC: statistically significant gain S vs E (p=0.04); E vs M (p=0.002)
- FEV1: E vs M not significant — reflects airway resistance & effort beyond static structure

Restrictive Ventilatory Pattern Classification

Table 5

Classification based on predicted FVC% threshold <80%

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	F1
S-Model (LGBM)	0.76	0.69	0.81	0.69	0.81	0.69
E-Model (LR)	0.85	0.81	0.88	0.81	0.88	0.81
M-Model (ResNet-34)	0.90	0.85	0.93	0.88	0.90	0.86

Accuracy
0.90

Specificity
0.93

F1-score
0.86

- Restrictive label: FVC% < 80% predicted (ATS/ERS criteria). Optimal backbone differs by task — ResNet-34 for classification vs. EfficientNetV2-B3 for regression.
- High specificity supports use as a conservative complementary screening adjunct. ResNet-34 selected by primary metric (Accuracy / F1).

Clinical emphasis: screening and risk stratification, not replacement of formal pulmonary function testing.

Model Interpretability: Feature Importance

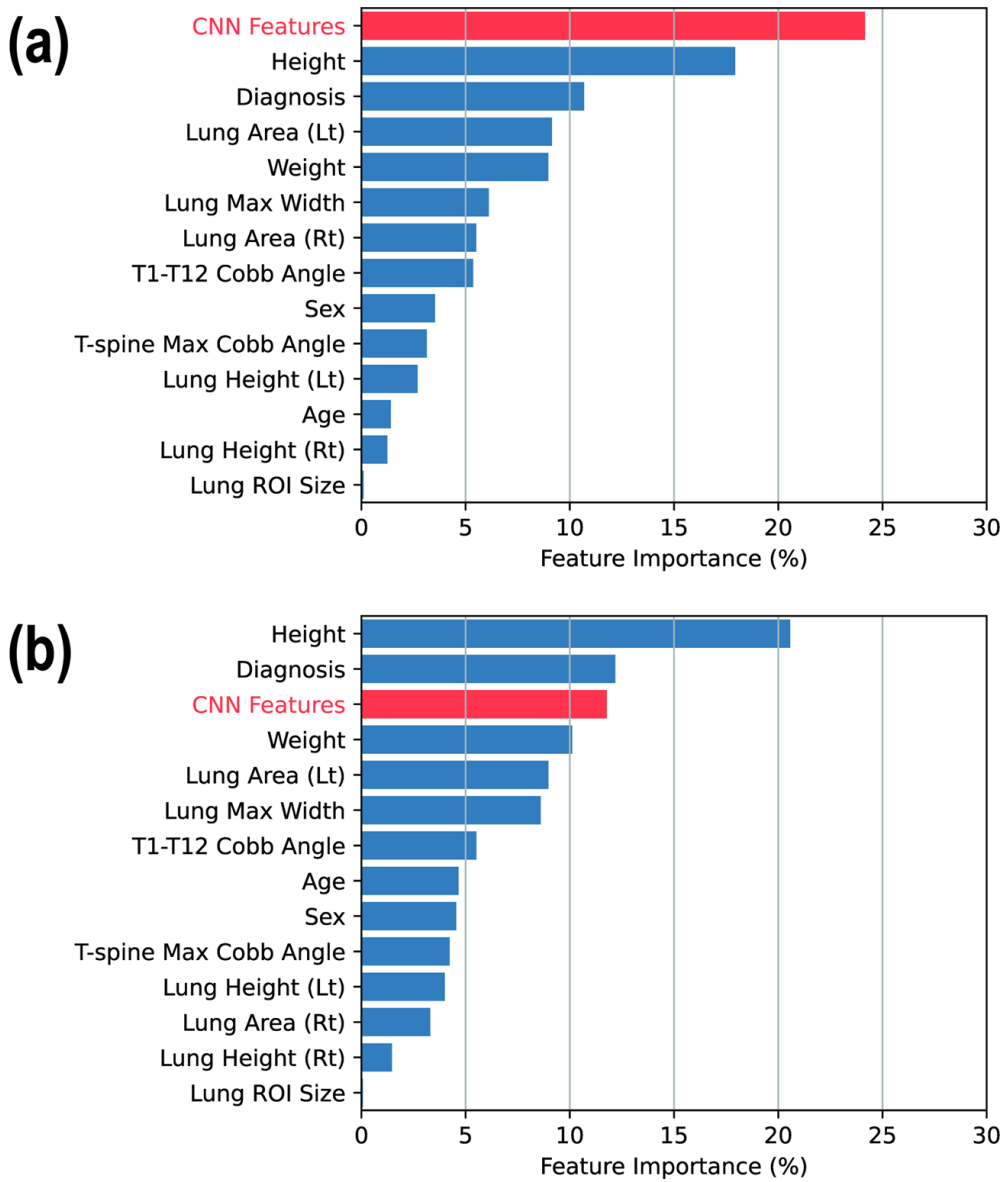


Figure 4. SHAP-based feature importance for FVC and FEV1 prediction

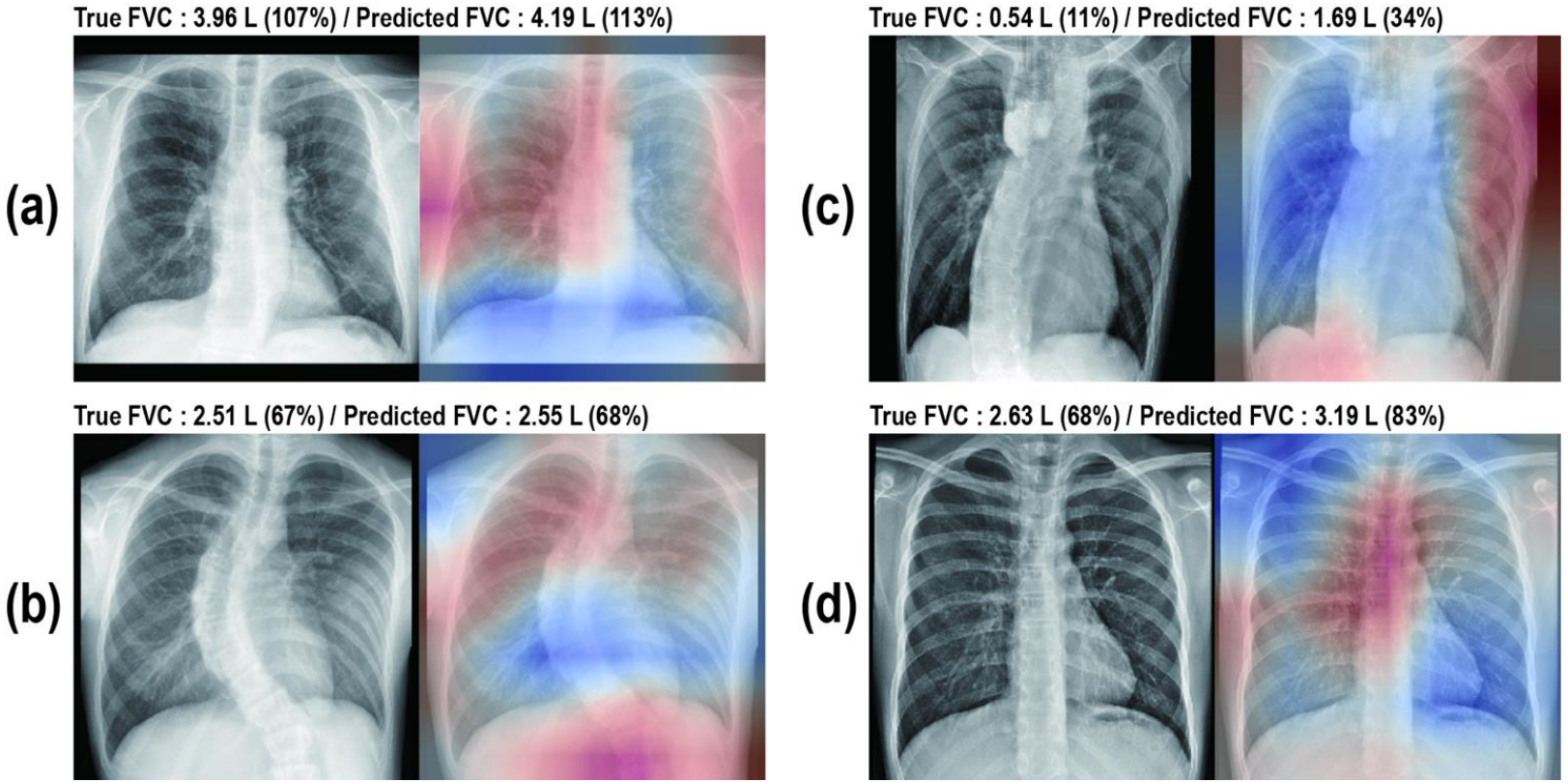
SHAP interpretation

- For FVC prediction, CNN-derived image features accounted for approximately 25% of total contribution.
- Height, diagnosis, and lung area were also major contributors.
- For FEV1, height showed the largest contribution; CNN features contributed less than in FVC.

Interpretation: raw pixels add structural information beyond manual measurements.

The feature profile aligns with the clinical expectation that FVC is more directly tied to thoracic geometry, whereas FEV1 also reflects airway resistance, muscle strength, and patient effort.

Grad-CAM Visualization: Where the Model Looks



Visual activation

Success modes

(a) Normal CXR: positive signal along the straight thoracic spine.

(b) Severe scoliosis, accurate prediction: focal negative activation at the curve apex.

Failure modes

(c, d) Restrictive cases with diffuse lung-field negative signals — errors arise when dysfunction is disproportionate to radiographic structure.

Error patterns highlight non-structural limits of radiograph-based estimation.

Figure 5. Representative Grad-CAM visualizations for FVC prediction on selected test samples

Activation patterns reflect a global structure–function association rather than a focal landmark, while highlighting an inherent ceiling when pulmonary dysfunction is driven by non-structural factors.

Limitations and Future Work

Current limitations

- Single-institution retrospective design: potential selection bias and temporal discordance between imaging and PFT.
- Standing PA CXR only: sagittal deformity and dynamic thoracic mechanics could not be assessed.
- Non-structural determinants of respiratory performance were not captured by static morphology.
- External validation has not yet been performed.

This model is a complementary screening adjunct, not a replacement for confirmatory PFT.

Future work

- Multi-institutional external validation across diverse scoliosis populations.
- Longitudinal modeling to track pulmonary function change over time.
- Incorporation of lateral radiographs or 3D imaging to capture sagittal and rotational deformity.
- Evaluation as a preoperative risk stratification and follow-up monitoring tool.

Conclusion and Clinical Implications

Pulmonary function in scoliosis patients can be estimated from routinely acquired CXR, manual radiographic measurements, and basic clinical information.

Clinical significance

- Non-invasive complementary screening when formal PFT is unavailable or impractical.
- Risk stratification for patients with structurally driven restrictive dysfunction.
- Potential support for preoperative assessment and longitudinal monitoring.

Scientific implication

- Raw image features contributed meaningfully to FVC prediction.
- FEV1 showed weaker image-driven gains, reflecting non-structural physiological determinants.
- Grad-CAM supports global thoracic structure-function learning.

Key metrics

FVC MAE 0.31 L

Accuracy 0.90

Specificity 0.93

Thank you.