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

Rather than treating pulmonary function prediction as a generic imaging task, this study isolates **how thoracic structural information contributes to FVC/FEV1 estimation in scoliosis.**

Stepwise multimodal comparison

Clinical data	Manual metrics	Raw CXR pixels

Cohort Selection & Patient Characteristics

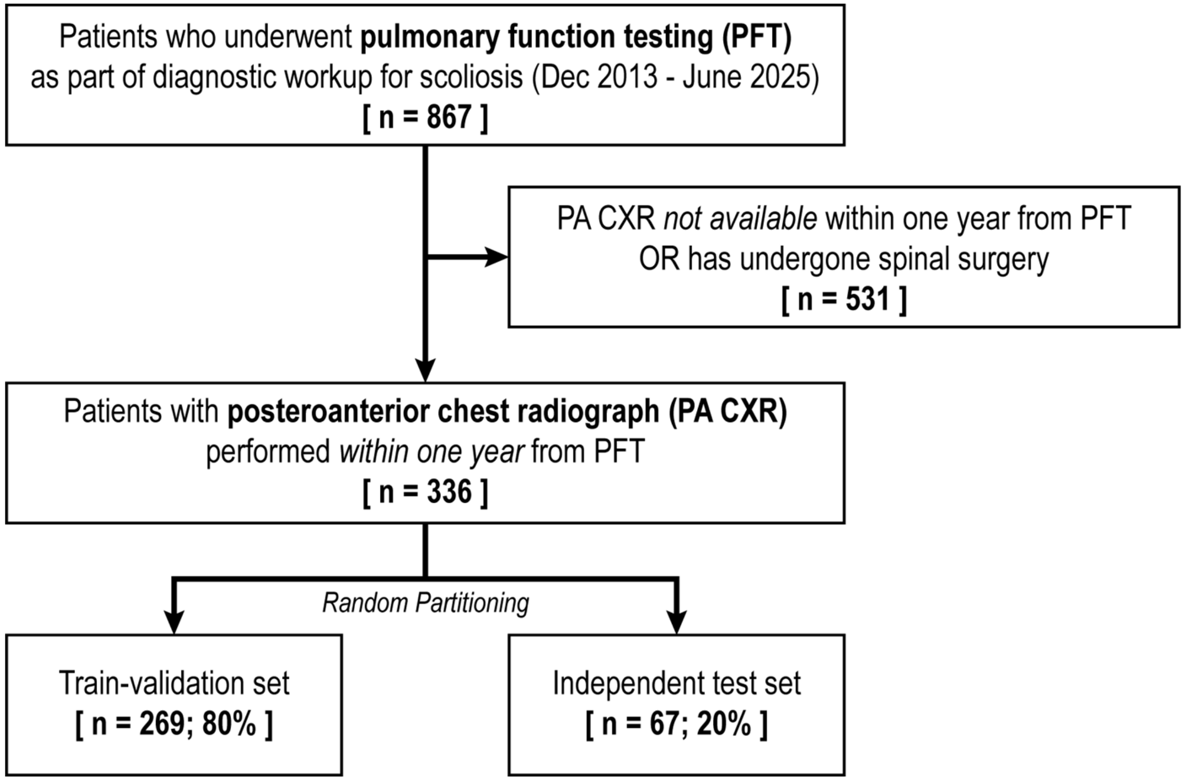


Figure 1. Cohort selection flowchart

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

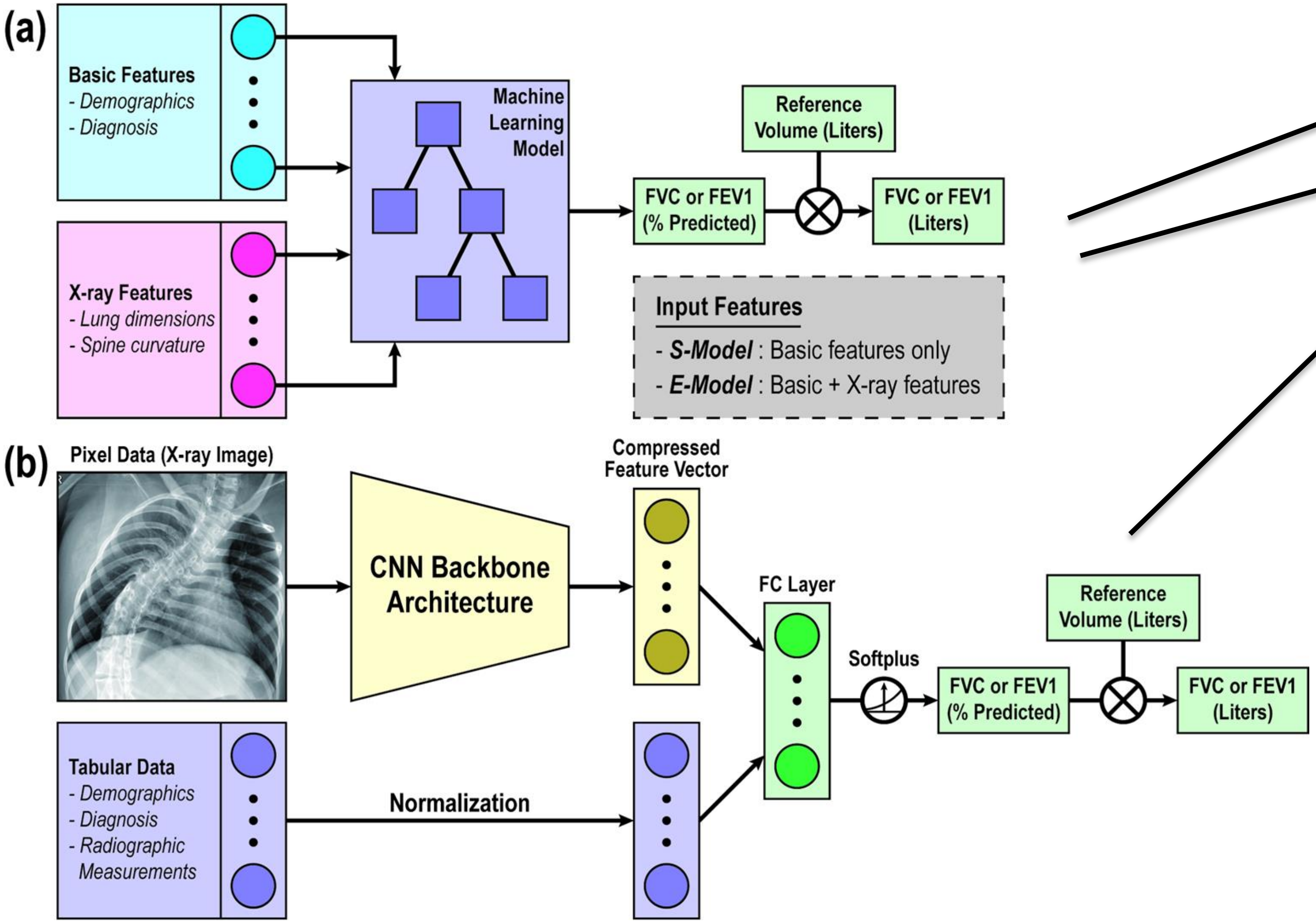
Train-validation: n=269 (80%)

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)
- Height: 154.7±13.6 / 157.7±9.5 cm (p=0.09)
- FVC: 2.6±1.0 / 2.6±1.0 L (p=0.86) | Restrictive: 39.4% / 38.8%
- **Right lung area & height: significant group diff. (p<0.005)**

Model Architecture Framework



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

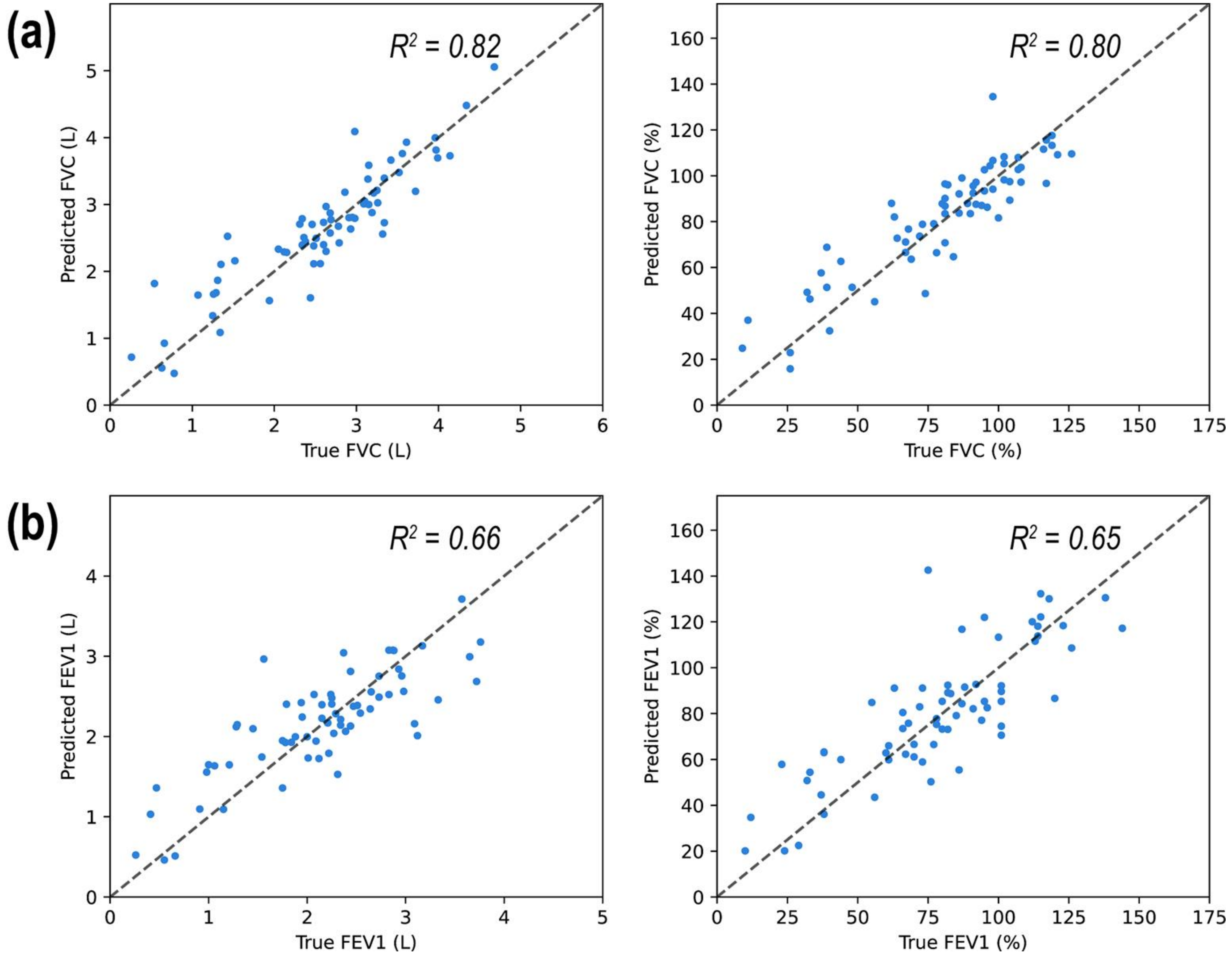


Figure 3. Scatter plots of FVC and FEV1 predictions on the independent test set

Key test-set results

FVC: MAE 0.31 L
RMSE 0.41 L, R² 0.82

FEV1: MAE 0.37 L
RMSE 0.48 L, R² 0.66

Selected M-Model Performance

- **FVC (EfficientNetV2-B3)**
  MAE 0.31L | RMSE 0.41L | R² 0.82
- **FEV1 (EfficientNetV2-B0)**
  MAE 0.37L | RMSE 0.48L | R² 0.66

- FVC: S→E→M Statically significant improvement(p=0.04, p=0.002)
- FEV1: E vs M Not statistically significant

Restrictive Ventilatory Pattern Classification

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

- M-Model improved specificity and F1-score compared with tabular-only model families.

- High specificity supports use as a conservative complementary screening adjunct to prioritize patients who may benefit from confirmatory PFT.

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

Model Interpretability: Feature Importance

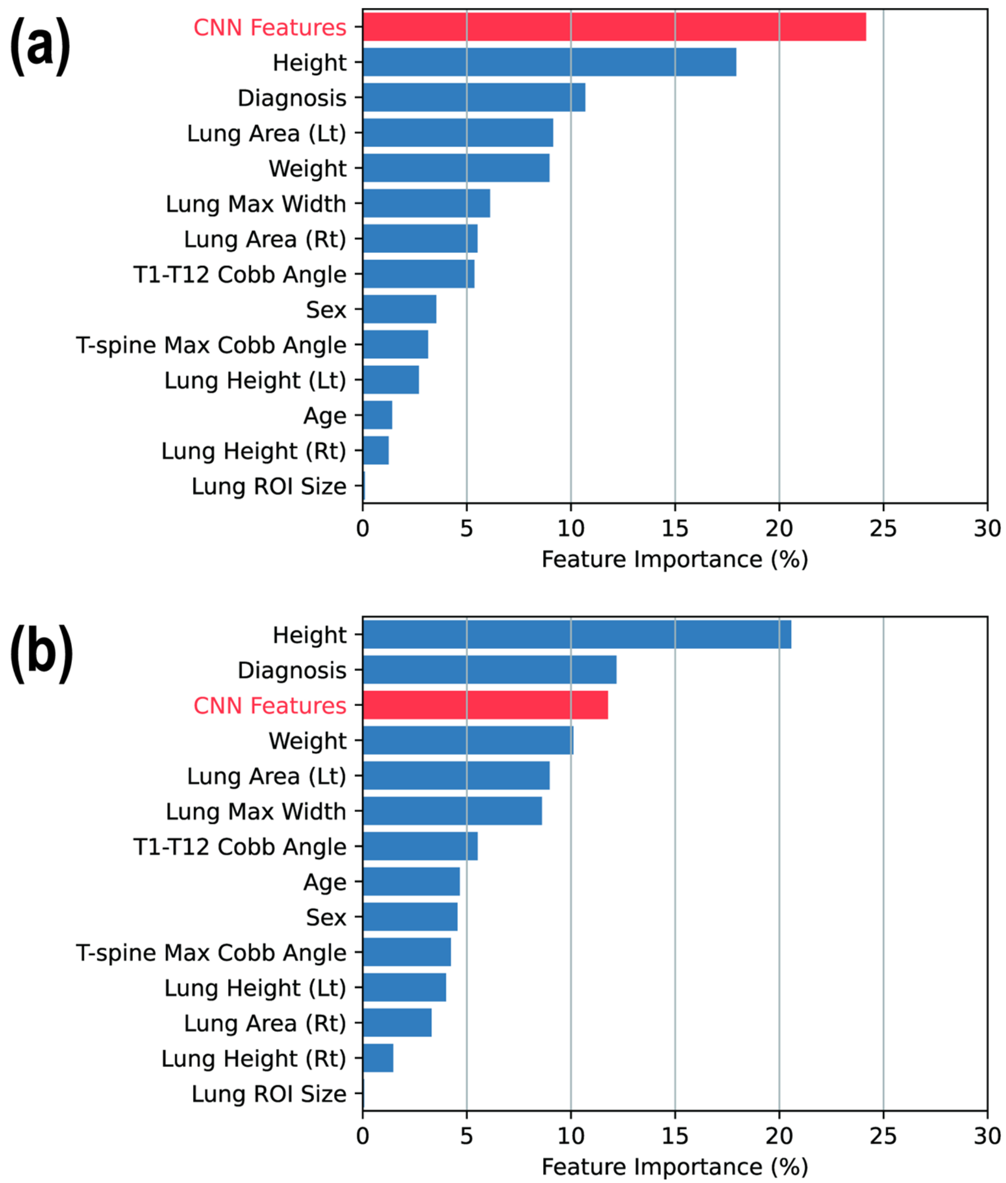


Figure 4. SHAP-based feature importance for FVC and FEV1 prediction

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

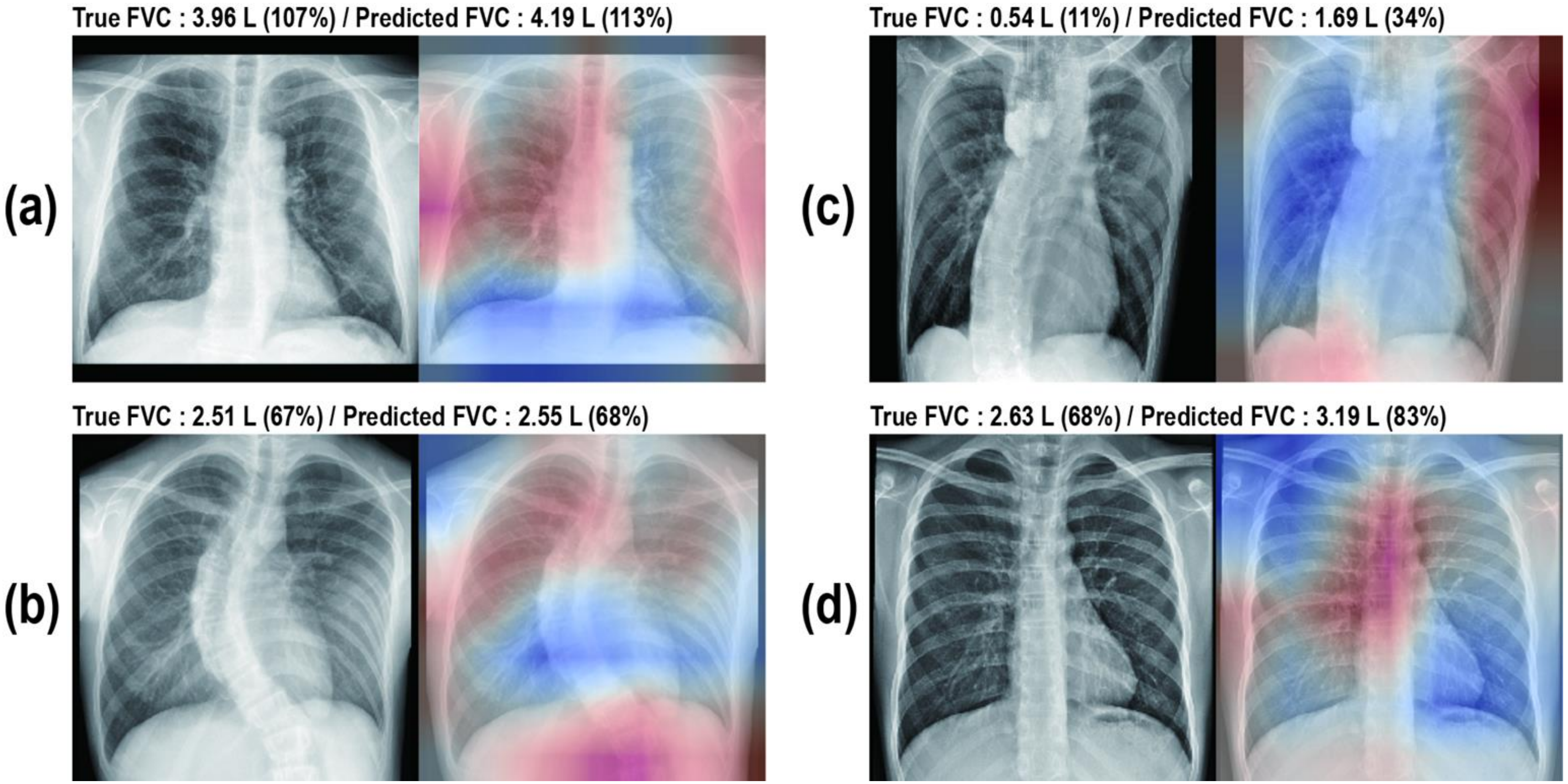


Figure 5. Representative Grad-CAM visualizations for FVC prediction on selected test samples

Visual activation

- Positive signals in preserved-function examples were observed along the thoracic spine and lung fields.
- Severe scoliosis showed strong negative activation near the curve apex.
- Diffuse lung-field signals suggest global thoracic structure-function learning.

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

These visualization patterns suggest that the model leverages thoracic curvature and lung-field information rather than a single focal anatomical landmark.

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.