Medical Engineering and Physics 140 (2025) 104332

Contents lists available at ScienceDirect

Medical Engineering and Physics

journal homepage: www.elsevier.com/locate/medengphy

Paper

Ensemble learning of deep CNN models and two stage level prediction of Cobb angle on surface topography in adolescents with idiopathic scoliosis

Mostafa Hassan [a,c,*], Jose Maria Gonzalez Ruiz [a], Nada Mohamed [a], Thomaz Nogueira Burke [e], Qipei Mei [b], Lindsey Westover [a,d]

[a] Department of Mechanical Engineering, University of Alberta, 10th Floor, Donadeo Innovation Centre for Engineering, Edmonton, T6G 1H9, AB, Canada
[b] Department of Civil and Environmental Engineering, University of Alberta, 7th Floor, Donadeo Innovation Centre for Engineering, Edmonton, T6G 1H9, AB, Canada
[c] Mechanical Design & Production Department, Faculty of Engineering, Cairo University, Cairo University Rd, Oula, Giza District, Cairo, 12613, Egypt
[d] Department of Biomedical Engineering, University of Alberta, 13th Floor, Donadeo Innovation Centre for Engineering, Edmonton, T6G 1H9, AB, Canada
[e] Allied Health Institute, Federal University of Mato Grosso do Sul, Cidade Universitária, Av. Costa e Silva, Campo Grande, 79070-900, MS, Brazil

ARTICLE INFO

Keywords:
Adolescent idiopathic scoliosis
Convolutional neural network
Cobb angle
Surface topography
Medical imaging
Deep learning

ABSTRACT

This study employs Convolutional Neural Networks (CNNs) as feature extractors with appended regression layers for the non-invasive prediction of Cobb Angle (CA) from Surface Topography (ST) scans in adolescents with Idiopathic Scoliosis (AIS). The aim is to minimize radiation exposure during critical growth periods by offering a reliable, non-invasive assessment tool. The efficacy of various CNN-based feature extractors—DenseNet121, EfficientNetB0, ResNet18, SqueezeNet, and a modified U-Net—was evaluated on a dataset of 654 ST scans using a regression analysis framework for accurate CA prediction. The dataset comprised 590 training and 64 testing scans. Performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and accuracy in classifying scoliosis severity (mild, moderate, severe) based on CA measurements. The EfficientNetB0 feature extractor outperformed other models, demonstrating strong performance on the training set ($R = 0.96$, $R^2 = 0.93$) and achieving an MAE of 6.13° and RMSE of 7.5° on the test set. In terms of scoliosis severity classification, it achieved high precision (84.62%) and specificity (95.65% for mild cases and 82.98% for severe cases), highlighting its clinical applicability in AIS management. The regression-based approach using the EfficientNetB0 as a feature extractor presents a significant advancement for accurately determining CA from ST scans, offering a promising tool for improving scoliosis severity categorization and management in adolescents.

1. Introduction

AIS is a prevalent spinal disorder characterized by a three-dimensional curvature and poses significant challenges in clinical assessment and management. This condition, marked by abnormal lateral curvature, vertebral rotation, and sagittal malalignment of the spine, predominantly arises during the adolescent growth spurt. AIS, accounting for up to 80% of scoliosis cases, affects approximately 0.47% to 5.2% of the global population, as indicated by various epidemiological studies [1–5]. Diagnosing AIS during the initial phases of adolescent growth significantly improves the likelihood of effectively preventing the development of more pronounced spinal curvatures as the child matures [4].

CA, derived from Posterior-Anterior (PA) X-ray imaging, remains the gold standard for scoliosis evaluation, being determinant in therapeutic decisions and scoliosis severity classification. CA between 10° and 25° corresponds to mild, 25° to 40° are considered moderate, and above 40° scoliosis is classified as severe [6]. However, the two-dimensional CA measurement does not fully encapsulate the three-dimensional complexities inherent in scoliosis, such as vertebral rotation. Further, the planar X-ray measurement does not capture aesthetic features appearance, such as torso asymmetry which is often more important to patients. This limitation, compounded by considerable intra- and inter-observer variability in CA assessments [7], raises concerns about measurement accuracy and subsequent treatment decisions. Notably, a measurement error of just 5°, equivalent to the clinical threshold for condition progression, could lead to inappropriate treatment choices [8,9]. Additionally, the frequent radiographic monitoring required for CA tracking also poses potential long-term health risks. Particularly, females with AIS undergoing regu-

* Corresponding author at: Department of Mechanical Engineering, University of Alberta, Ab, Canada.
E-mail address: mkhassan@ualberta.ca (M. Hassan).

https://doi.org/10.1016/j.medengphy.2025.104332
Received 4 March 2024; Received in revised form 20 March 2025; Accepted 4 April 2025
Available online 9 April 2025

lar X-ray evaluations face a significantly higher risk of developing breast cancer in later life [10]. These concerns highlight the necessity for safer, radiation-free imaging alternatives.

These non-invasive techniques provide a safer environment for scoliosis monitoring, reducing the risks associated with repeated radiation exposure. Two prominent methods are ultrasound and ST. Ultrasound offers real-time imaging and can effectively detect vertebral rotation by measuring the rotation of the laminae [11]. It correlates well with radiographic CA measurements, particularly when prior radiographs are available for reference [12,13]. However, its effectiveness diminishes for severe curves exceeding 45° due to the obstruction of spinous processes [13–15], and it is highly operator-dependent, requiring significant expertise for reliable measurements [15]. In contrast, ST provides a three-dimensional representation of scoliosis on the torso, enabling non-invasive and radiation-free monitoring [16]. It demonstrates strong correlation and reliability compared to radiographic CA measurements [16–18] and is particularly advantageous for serial monitoring to track curve progression, potentially reducing the frequency of radiographs [18]. Despite its advantages, advancements are needed to optimize the use of external torso asymmetries for accurate scoliosis evaluation. Machine learning techniques, such as CNNs, offer a promising avenue to overcome these limitations by analyzing complex patterns in ST data and predicting spinal deformities.

Markerless 3D ST has emerged as a promising, non-invasive technology that offers three-dimensional imaging of the torso without radiation exposure risks [19]. This technique provides a more accurate torso representation compared to marker-based STs, which frequently suffer from marker placement errors [20,21]. Various scanning technologies, such as Laser Optical Scanning Systems [22,23] and Handheld 3D scanners [24], have been employed to capture the 3D ST of the torso. Our group introduced an innovative asymmetry analysis method, significantly enhancing ST's clinical utility in managing scoliosis [25,26]. This method employs a 'best plane' of sagittal symmetry, enabling the reflection of the original torso and its superimposition with the reflected image. This process yields a complete three-dimensional representation of torso asymmetries. Rosthstock et al. [27] applied a similar method, but they used the anatomical sagittal plane itself for reflecting the original torso scan. These aforementioned asymmetry techniques were based on scans of patients in a standing position. Kokabu et al. [28] Kokabu et al. (2021) applied the asymmetry analysis to scans of patients in the Adam's Forward Bend Test position, utilizing a median sagittal plane for torso reflection.

One significant challenge in using ST for AIS assessment is accurately predicting the CA from external torso maps, as these scans lack direct spinal visualization. This necessitates a sophisticated interpretation of surface asymmetry to accurately estimate spinal geometry, a task well-suited for advanced machine learning techniques, particularly CNNs [29,30]. CNNs have shown potential in scoliosis analysis from both 2D PA X-ray images and ST scans [28,31,32]. Unfortunately, limitations arising from a lack of external validation and overfitting of the deep learning algorithms employed mean that accurately predicting CA remains a challenge.

In this work, a comprehensive comparative analysis of several CNN models for CA prediction from AIS 3D ST scans is conducted. It is important to emphasize that the value predicted by our machine learning model is not a new or distinct parameter; rather, it represents an estimation of the actual CA for the maximum scoliosis curve in a specific patient scan. The model is designed to approximate the traditional CA measurement, which continues to serve as the standard for assessing scoliosis severity. Due to the limited dataset typically encountered in medical imaging, standard deep CNN architectures with extensive training parameters were not included, as they are prone to overfitting in such scenarios. Instead, the focus is on architectures that are either specifically tailored for small datasets or have demonstrated exceptional performance in medical imaging applications. This includes SqueezeNet [33], ResNet-18 [34], DenseNet-121 [35], U-Net [36], and EfficientNet [37]. These models were selected based on their proven effectiveness in medical contexts [36,38,39] and their suitability for limited data environments. The performance of these models will be assessed on an unseen testing set to identify the most appropriate architecture for CA prediction in this specialized field. A comprehensive list of acronyms used throughout this article is provided in Table 1.

Table 1
List of acronyms used in this study.

Acronym	Definition	Acronym	Definition
AIS	Adolescent Idiopathic Scoliosis	FP	False Positive
CNNs	Convolutional Neural Networks	TN	True Negative
RMSE	Root Mean Square Error	TP	True Positive
NPV	Negative Predictive Value	FN	False Negative
ANN	Artificial Neural Network	CA	Cobb Angle
ReLU	Rectified Linear Unit	PA	Posteroanterior
R^2	Coefficient of Determination	ST	Surface Topography
R	Pearson's Correlation Coefficient	MAE	Mean Absolute Error

2. Materials and methods

2.1. Clinical data

In prior research conducted by our group, ST scans were performed on 233 patients diagnosed with AIS using high-resolution Konica Minolta (vivid 910 3D laser scanners) [40]. Four 3D images (left, right, back, front) for each scan were merged to form a comprehensive 3D representation of each patient's torso [23,25]. Additional 195 follow-up scans at six and twelve-month intervals were captured, bringing the total to 428 data points. Parallel to this, a Structure Sensor Mark I (Occipital, Inc.) attached to an iPad (3rd generation, Apple Inc.) was used to capture ST scans of patients. The operator circled the patient to ensure complete coverage of the torso surface. The inclusion of 242 scans from this dataset increased the available data to 670 total scans.

The patient group was aged between 10 to 18 years (mean 14.0 ± 1.8). These patients were non-surgical cases, each with at least one spinal curve greater than 10°. All ST scans had corresponding PA X-ray radiographs. X-ray-based CA measurements, obtained from these radiographs, were utilized as the ground truth labels for the CNN models. These measurements were conducted by scoliosis specialists in dedicated scoliosis clinics. The mean maximum CA across our combined clinic dataset is 32.6° ± 13.1°, ranging from 10° to 97°. The testing set was randomly selected to comprise approximately 10% of the entire dataset, resulting in 64 data points. The mean maximum CA in the testing set was 31.73° ± 10.83°, with values ranging from 12° to 51°. The data split was conducted without stratification by scoliosis type, severity, or gender, ensuring a completely random selection process. Fig. 1 presents a histogram plot of the entire dataset, showing the frequency of data counts across CA ranges.

In our study, data points corresponding to CA ranging from 60° to 90° were excluded. This exclusion, prompted by the sparse representation of these ranges, was essential to uphold the statistical reliability and efficacy of our model training. Such a measure is in line with established machine learning and data preprocessing guidelines. Influential literature, including Goodfellow et al., [41] and García et al. [42], underscores the imperative of judicious data distribution and representation in model training. Kohavi et al.'s [43] insights on cross-validation additionally emphasize the need for data balance, crucial in mitigating biases toward underrepresented data ranges. The resultant dataset, comprising 654 subjects, features a well-distributed range of CA, thus providing a statistically sound basis for our CNN models.

2.2. Preprocessing

Distinct preprocessing protocols were employed for the vivid 910 3D laser and the structure sensor scanning systems. For the 3D laser

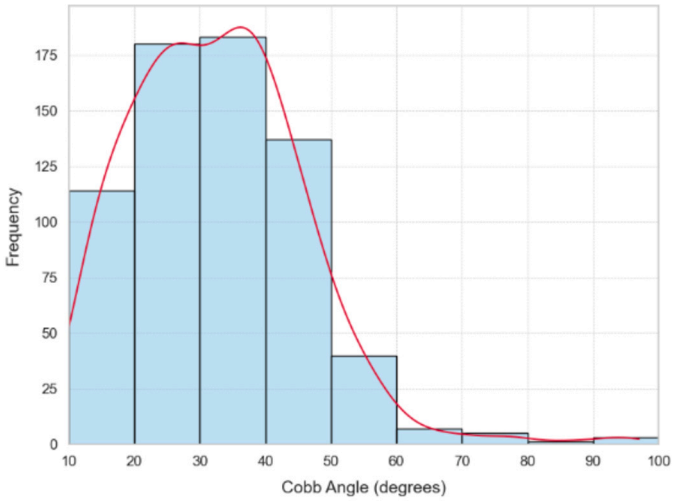


Fig. 1. Histogram plot of the frequency of data points across maximum CA ranges in the current dataset for each ST scan.

scanning, four views captured by vivid 910 scanners were integrated using our custom macro in Geomagic Studio 12 software (Morrisville, NC, USA) [23,25]. This integration process yielded a detailed 3D mesh representation of each patient's torso. Regarding the Structure Sensor scanning, a live mesh was directly generated using the scanning app of the Structure Sensor.

The subsequent stage involved refining these meshes using the Geomagic software. This critical step entailed the removal of extraneous elements such as the frame, arms, hips, head, and neck, focusing exclusively on the torso [40]. Additionally, a cleaning process was undertaken to rectify any scanning imperfections, such as gaps or spikes, which are common in 3D scanning procedures. The outcome of this process was a set of refined, high-quality 3D meshes [40].

These meshes were then converted into point clouds using a custom-developed Python script, specifically tailored for this task. This process employed a vertex sampling method, wherein the vertices of the mesh were extracted and utilized as the points of the point cloud. Then a mirrored version of each point cloud was created. This allowed for the alignment and comparison of the original and mirrored clouds to identify the optimal plane of symmetry. The process involved minimizing the distances between corresponding points across both meshes, with the discrepancies visualized in a detailed deviation map (refer to Fig. 2).

For the final stage of data preprocessing, The focus was placed solely on the posterior half of the torso ST scan. The anterior points were omitted, following the findings of our previous group work which established a correlation between the asymmetry patches on the back and the severity of AIS curves. Therefore, the anterior half of the scan was not utilized in our analysis. The deviation and depth information of the points on the posterior half of the ST was mapped into a 2D image array. The mapping process involved positioning these points within the array according to their respective x and y coordinates. Subsequently, each element in the array was assigned values corresponding to the deviation and depth of its respective point in the scan. This bi-channel image array, representing the posterior torso point cloud, constituted the input for our deep learning model. The entire preprocessing pipeline is comprehensively depicted in Fig. 2, providing a clear visual representation of each step.

2.3. Proposed technique for CA prediction using CNN

This study confronts the prevalent issue of overfitting, especially critical in datasets with limited size, through the implementation of a comprehensive validation strategy. Our approach, designed to ensure the robustness and generalizability of our models, is underpinned by a 10 repeat of stratified 10-fold cross-validation method. This approach culminates in the creation of 100 unique models, each subjected to statistical analysis to verify unbiased performance. A spectrum of CNN feature extractors is explored to evaluate their capability in accurately interpreting scoliotic ST. The process is outlined as follows:

a) **Data Partitioning**: The dataset is initially split into two subsets: training and testing. The training subset is used for model learning, while the testing subset, composed of external, unseen data, is earmarked for the conclusive evaluation of the CNN models.
b) **Repeated stratified k-Fold Cross-Validation**: In this phase, the training data is segmented into 10 distinct folds, facilitating the development of 10 individual CNN models. Each fold serves as a validation set in rotation, enabling exhaustive feature extraction through repeated training cycles. This cross-validation process is repeated 10 times to ensure comprehensive feature identification.
c) **Training Set Evaluation using mean predictions**: Utilizing the 100 models via repeated cross-validation, average prediction values are computed across the training set. This method, in line with previous literature such as Bishop et al. [44], ensures comprehensive data coverage and reduces bias, providing a clear measure of model fitting effectiveness.
d) **Application on unseen data using mean predictions**: The same mean prediction method is applied to an external, unseen testing dataset. This step is crucial for evaluating the model's predictive accuracy and generalizability, offering a distinct assessment of its performance in unfamiliar contexts.
e) **AIS severity classification Evaluation**: The resulting predictions from both sets will be evaluated against established clinical standards for AIS severity classes, including mild($10° \leq CA < 25°$), moderate($25° \leq CA < 40°$), and severe($CA \geq 40°$).

The whole methodology is illustrated in Fig. 3, with 590 training data points and 64 testing data points.

2.4. Convolutional neural network

CNN is composed of multiple types of interrelated layers that handle various tensor orders depending on the type of input, such as pseudo, grayscale, and RGB images [41]. This process involves dividing the input image into smaller sections called receptive fields. These fields, in turn, combine to form feature maps or filters. The optimization of these filters is made possible by a weight matrix that contains bias components. An essential part of this procedure is the updating of each epoch's weight matrix during the backpropagation phase of training to minimize the loss score [30]. Typically, this framework includes a Rectified Linear Unit (ReLU) following the convolutional layers, introducing non-linearity into the overall system. Subsequently, a pooling layer is integrated to downsize the dimensions of the filters. There are various criteria for pooling such as assimilation of the maximal pixel value, known as max pooling, or calculating the average value, known as mean pooling [45]. Fully connected layers replicate a standard Artificial Neural Network (ANN) where dense hidden layers can contribute to the output layer. Additionally, the inclusion of dropout layers aids in avoiding overfitting by excluding a calculated percentage of the overall training parameters [41]. For a more thorough understanding of the mathematical underpinnings, the potential advantages, the available alternatives, and inherent tradeoffs of this process, refer to the existing literature in this field [30,41,45,46].

In this study, benchmark CNN architectures were evaluated to determine their suitability for scoliosis detection using ST scans. Since this study represents an early effort in applying benchmark architectures for scoliosis detection from ST scans, no established guidelines exist in the literature regarding the optimal architecture for this task. Therefore, our selection was based on three key criteria: (1) proven efficiency in medical imaging, (2) capability to handle small datasets effectively, and (3) demonstrated performance in scoliosis detection using other medical imaging techniques, such as X-ray imaging.

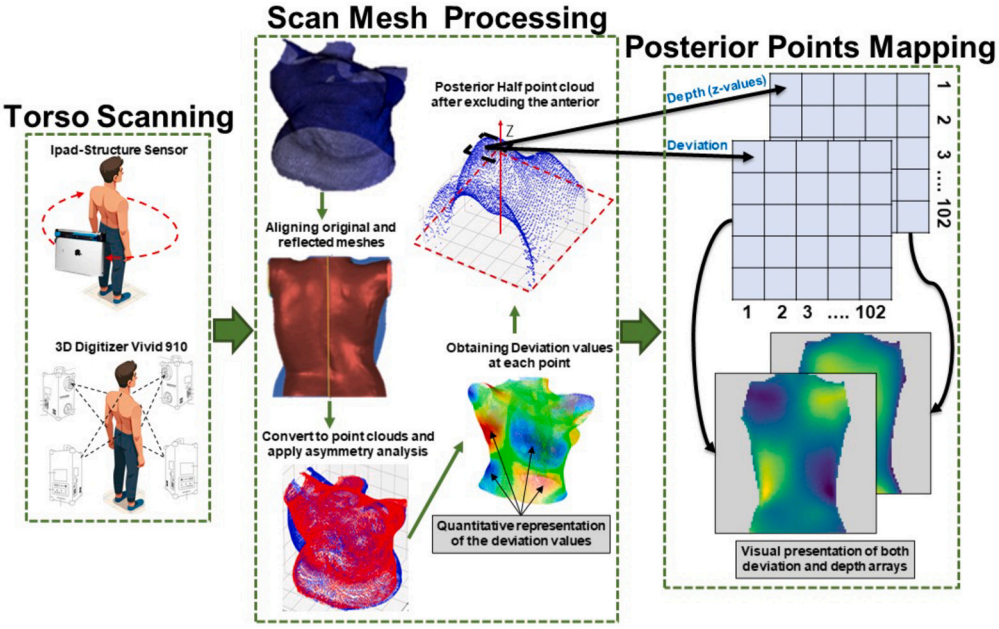


Fig. 2. Preprocessing steps. The 2D array was mapped from the posterior half of the torso point cloud of the scan which will be used as an input to the CNN model.

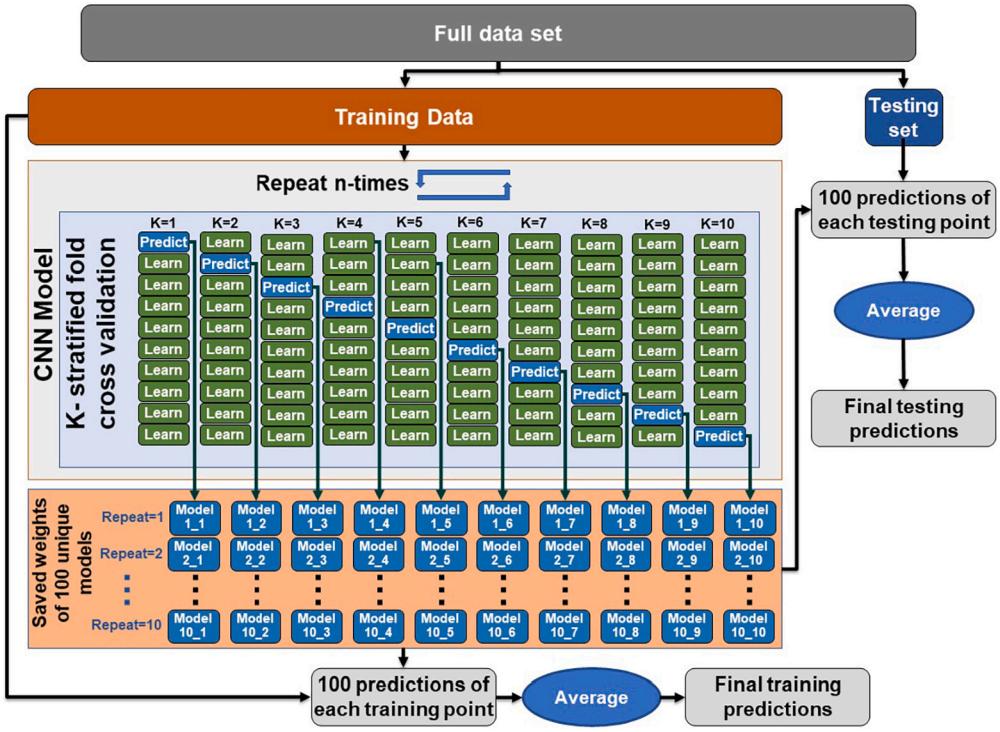


Fig. 3. CA prediction. The full data set will be split into training and testing sets, then applied a 10-repeated stratified 10-fold cross-validation. Finally, the resulting 100 model predictions will be averaged to get the final prediction.

The chosen architectures have been widely utilized in medical imaging, including disease detection from chest X-ray radiographs [47] and COVID-19 identification from CT images [48]. Additionally, they were selected for their ability to perform well on small datasets [36,48,49], a critical factor in medical applications where data collection is often limited. While these models have not been extensively explored for scoliosis detection using ST data, they have shown strong predictive performance in scoliosis assessment through other imaging modalities, such as X-ray and MRI [50,51]. Their established strengths in medical imaging, combined with their potential applicability to scoliosis detection, make them well-suited for this study.

The architectures used in this study are briefly described as follows:

- **DenseNet121** [35]: This network employs a dense concatenation of filters across its neural network layers, to focus on essential partial filters throughout an image.
- **EfficientNetB0** [37]: This network advances learning by scaling image parameters like width, depth, and resolution using a specialized compound coefficient. Unlike other architectures, it focuses on layer dimensions rather than just the training parameters. With limited data sets, scaling down these image parameters is essential to prevent overfitting.

Table 2
Dynamic Data Augmentation techniques used during training the CNNs.

Technique	Setting
rotation_range	10
width_shift_range	0.1
height_shift_range	0.1
shear_range	0.1
zoom_range	0.1
horizontal_flip	'True'
fill_mode	'nearest'

- **ResNet18** [34]: ResNet, an abbreviation for Residual Network, revolutionized deep neural networks by incorporating residual layers and skip connections. These additions combat the problem of vanishing gradients often encountered in deep learning. The key breakthrough of ResNet lies in its residual layers, enabling the network to focus on learning residual mappings instead of striving to directly learn the complete mapping from input to output.
- **SqueezeNet** [33]: SqueezeNet was developed with the innovative goal of minimizing the number of parameters required during the training process of neural networks. Central to its design is the "Fire Module", which incorporates a conv2D 1x1 layer, a strategic choice that substantially reduces the volume of parameters across the entire network.
- **Modified U-Net**: This network is mainly used for medical imaging segmentation [36], in which upsampling and downsampling paths are implemented. This happens by detecting specific features at the start of the network and then concatenating them at the end of the network. Modifications have been implemented to the original U-Net architecture for application to the data utilized for this study. Firstly, a ZeroPadding2D layer was introduced at the outset of the network. This adjustment ensures that the input dimensions (102x102x2) are consistently divisible by 2, thereby facilitating smoother downscaling and upscaling processes within the network. Additionally, a BatchNormalization layer has been incorporated within the 'conv block' to improve the model's learning efficiency. Furthermore, unlike the original U-Net, which does not explicitly employ cropping during the upsampling and downsampling phases, our revised architecture utilizes explicit cropping. This method is particularly beneficial for handling the two-dimensional nature of the input.

A regression head is added to these feature extractors, incorporating a 50% dropout layer and L2 regularization to reduce overfitting risks. Fig. 4 illustrates these CNNs. The Huber Loss function was utilized as the primary loss metric, chosen for its lower sensitivity to outliers compared to the squared error loss [52]. This function (denoted as α) operates on a dual mechanism: it behaves quadratically for small discrepancies between predicted and observed values, and linearly for larger discrepancies. An α value of 0.1 was found optimal in our experiments. A detailed account of the hyperparameters used is provided in Table 3. To bolster model generalization, Dynamic data augmentation techniques were employed, varying each image in every epoch. These variations included rotations up to 10 degrees, zoom ranges of 0.1, shifts in both width and height by up to 10%, and random horizontal and vertical flips, detailed description of the augmentation parameters used are listed in Table 2.

2.5. Statistical analysis and performance metrics

In this study, the fitting performance between the actual and predicted maximum CA for each ST scan was assessed across both the training and testing datasets. Four metrics were employed to evaluate

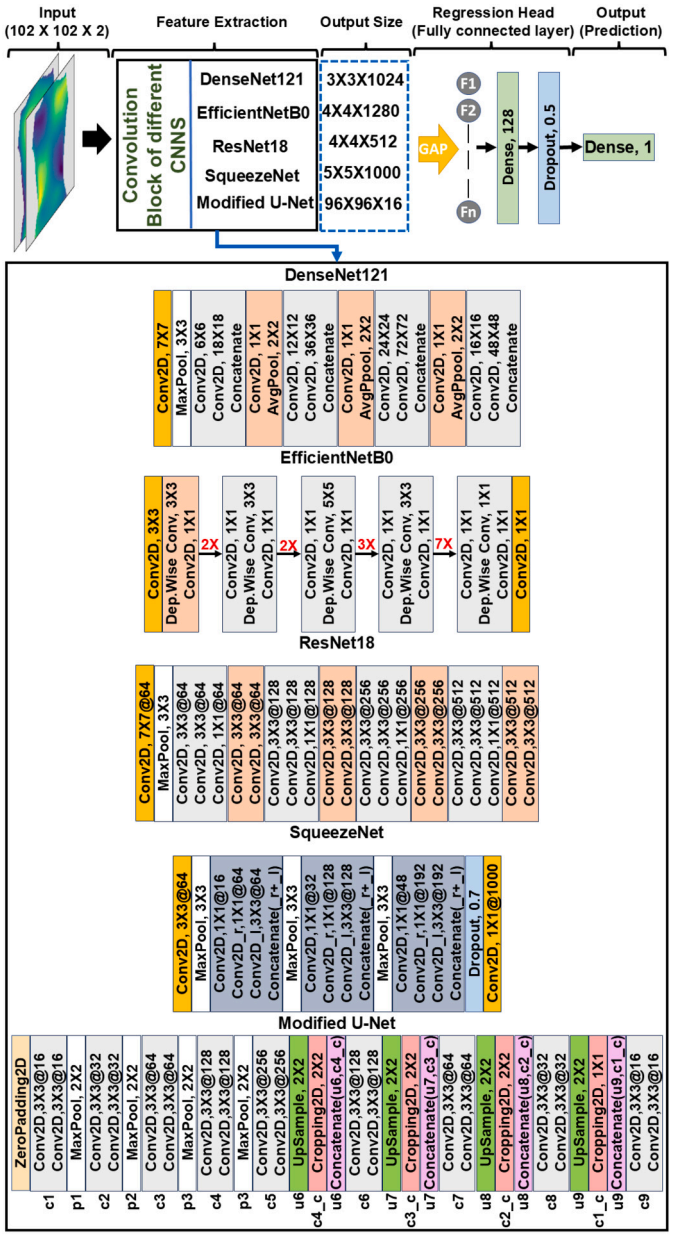


Fig. 4. Details of suggested CNN feature extractor models.

Table 3
Hyperparameters of the CNN model.

Parameters	Values
Batch size	32
Learning rate	0.001
Epochs	1500
Optimizer	Adam
L2 regularization strength	0.0001

model prediction accuracy: MAE, RMSE, Pearson's correlation coefficient (R), and the Coefficient of Determination (R^2).

For the practical assessment in a clinical setting, categorizing scoliosis severity as mild($10° \leq CA < 25°$), moderate($25° \leq CA < 40°$), or severe($CA \geq 40°$). Model performance was evaluated using an array of metrics. Beyond precision and recall (sensitivity), the F1-Score, specificity (true negative rate), and negative predictive value (NPV) were computed. These metrics furnish a comprehensive understanding of the model's efficacy at varying clinical decision thresholds.

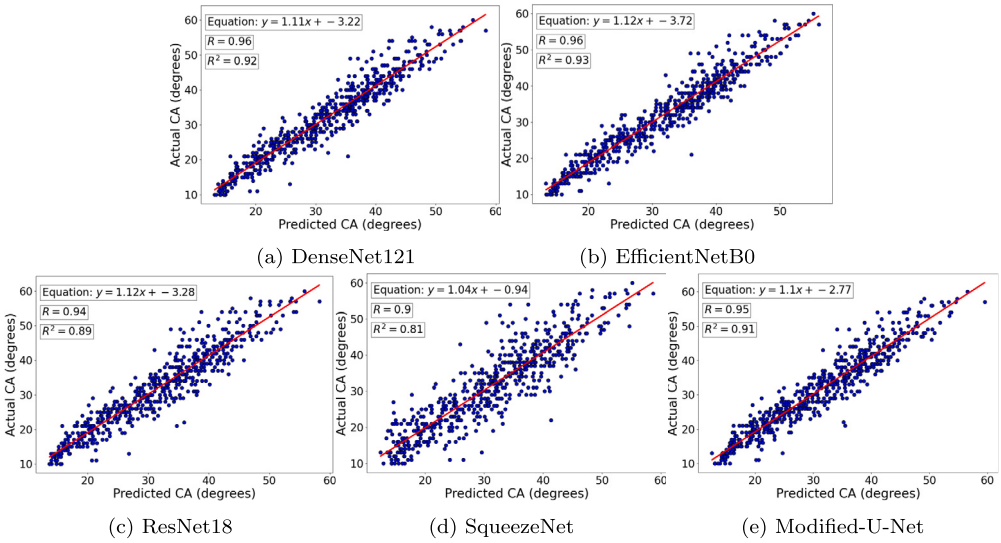


Fig. 5. Correlation between the actual and predicted values of the CA for the training set for the different architectures.

$$\text{Precision (Positive Predictive Rate)} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

$$\text{Specificity (True Negative Rate)} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{Negative Predictive Value (NPV)} = \frac{TN}{TN + FN} \tag{5}$$

where, TP (true positives) refers to instances where the predictive model accurately identifies a patient's scoliosis as falling within the correct severity category. TN (true negatives) corresponds to cases where the model appropriately recognizes patients not meeting the criteria for a specified severity category. Conversely, FP (false positives) occurs when the model erroneously assigns a scoliosis severity category to a patient, and FN (false negatives) designates situations in which the model fails to classify a patient's condition as belonging to a particular severity category when it does. These classifications are instrumental in constructing the confusion matrix—a critical tool for the derivation of the performance metrics previously delineated.

3. Results

3.1. Training performance

After 10 stratified cross-validations were conducted on the dataset, key regression metrics (MAE, RMSE, R, R^2) were extracted for each fold and architecture. Table 4 summarizes the average MAE, RMSE, R, and R^2 values, showing minimal variation across extractors, particularly with the EfficientNetB0 model exhibiting the narrowest MAE (4.84° to 4.98°) and RMSE (6.40° to 6.66°) range. In contrast, the modified U-Net had broader MAE (4.74° to 5.05°) and RMSE (6.36° to 6.65°) ranges. The performance of different extractors was closely matched.

R and R^2 values across models also showed tight ranges (0.82 to 0.84 for R and 0.66 to 0.70 for R^2), indicating consistent predictive accuracy. The 100 CNN models generated from cross-validation were used to predict within the training set, averaging predictions to maximize data use. Fig. 5 displays the high correlation between actual and predicted values across CNNs, with EfficientNetB0, U-Net, and DenseNet121 achieving R values of 0.96, 0.95, and 0.96, and R^2 values of 0.93, 0.91, and 0.92, respectively. These results highlight the models' effectiveness and training performance robustness.

3.2. External testing dataset performance

To assess the robustness and efficiency of our proposed models, previously unseen data were analyzed. Predictions for the testing set were generated by averaging the results from 100 model predictions for each architecture. Fig. 6 illustrates the correlation between the actual and predicted values CA for the testing set. Our models demonstrate a strong correlation in the testing set, with R values ranging from 0.73 to 0.75 and R^2 values ranging from 0.54 to 0.56. Notably, the EfficientNetB0 architecture exhibits the highest correlation on the testing set, achieving an R-value of 0.75 and a R^2 value of 0.56. Table 6 presents the MAE and the RMSE values for both the training and testing sets. Across the different extractors, MAE ranges from 5.99° to 6.34°, while RMSE ranges from 7.53° to 7.91°. As the performance ranges across parameters were relatively narrow, we conducted a statistical significance analysis of differences between CNN architectures. A one-way ANOVA test comparing results from the cross-validation procedure (100 models per architecture) revealed statistically significant differences in correlation metrics (R values) between architectures ($F(4, 495) = 2.52, p < 0.05$), while differences in MAE were not statistically significant ($F(4, 495) = 1.48, p > 0.05$). Table 5 summarizes the statistical significance analysis across all architectures. The cross-validation results demonstrate that all architectures achieved comparable R values (0.83-0.84) and R^2 values (0.67-0.68) during training. However, on the test set, EfficientNetB0 maintained superior correlation performance, achieving the highest R (0.75) and R^2 (0.56) values.

While the MAE differences between architectures were minimal and not statistically significant (ranging from 5.99° for U-Net to 6.29° for DenseNet121), the superior correlation metrics of EfficientNetB0 indicate it more effectively captures the relationship between predicted and actual CA values. This improvement in explained variance (R^2) is particularly valuable in the clinical context of CA prediction, as it enhances the model's ability to track changes in CA across the severity spectrum. The combination of competitive error metrics and superior correlation performance supports our selection of EfficientNetB0 as the optimal architecture for CA prediction from ST scans in patients with AIS.

3.3. Scoliosis severity classification

In many scoliosis clinics, scoliosis severity is classified as mild($10° \leq CA < 25°$), moderate($25° \leq CA < 40°$), or severe($CA \geq 40°$), based on the CA range. This classification scheme was employed in our analysis to assess the model's capability on the training set, specifically, its

Table 4
Values of MAE, RMSE, R, and R^2 Average results obtained from each repetition of the cross-validation folds for the proposed CNN architectures.

		Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9	Rep 10
DenseNet121	MAE (°)	4.81	4.82	4.73	4.80	4.84	4.91	4.98	4.83	4.91	4.75
	RMSE (°)	6.44	6.33	6.30	6.41	6.60	6.66	6.67	6.56	6.61	6.36
	R	0.83	0.84	0.84	0.83	0.82	0.82	0.82	0.83	0.83	0.84
	R^2	0.69	0.70	0.70	0.69	0.67	0.66	0.66	0.67	0.67	0.69
EfficientNetB0	MAE (°)	4.92	4.98	4.98	4.84	4.98	4.93	4.86	4.92	4.96	4.91
	RMSE (°)	6.66	6.62	6.65	6.48	6.65	6.47	6.40	6.66	6.59	6.51
	R	0.82	0.82	0.82	0.83	0.82	0.83	0.84	0.82	0.83	0.83
	R^2	0.66	0.67	0.66	0.68	0.66	0.68	0.69	0.67	0.67	0.68
ResNet18	MAE (°)	4.92	4.73	4.89	4.76	4.84	4.80	4.76	4.87	4.82	4.84
	RMSE (°)	6.61	6.34	6.50	6.41	6.40	6.41	6.35	6.47	6.56	6.52
	R	0.83	0.84	0.83	0.83	0.84	0.84	0.84	0.83	0.83	0.83
	R^2	0.67	0.70	0.68	0.69	0.69	0.69	0.69	0.68	0.68	0.68
SqueezeNet	MAE (°)	4.86	4.81	4.78	4.86	4.80	4.82	4.74	4.88	4.86	4.91
	RMSE (°)	6.49	6.45	6.39	6.44	6.40	6.45	6.43	6.48	6.35	6.51
	R	0.83	0.83	0.84	0.84	0.84	0.83	0.84	0.83	0.84	0.83
	R^2	0.68	0.68	0.69	0.69	0.69	0.68	0.69	0.68	0.69	0.68
Modified U-Net	MAE (°)	4.89	4.98	4.81	4.93	4.93	4.91	5.05	4.91	4.74	4.88
	RMSE (°)	6.62	6.65	6.36	6.55	6.54	6.60	6.62	6.61	6.49	6.61
	R	0.83	0.82	0.84	0.83	0.83	0.83	0.82	0.83	0.83	0.83
	R^2	0.67	0.66	0.69	0.68	0.68	0.67	0.67	0.67	0.68	0.67

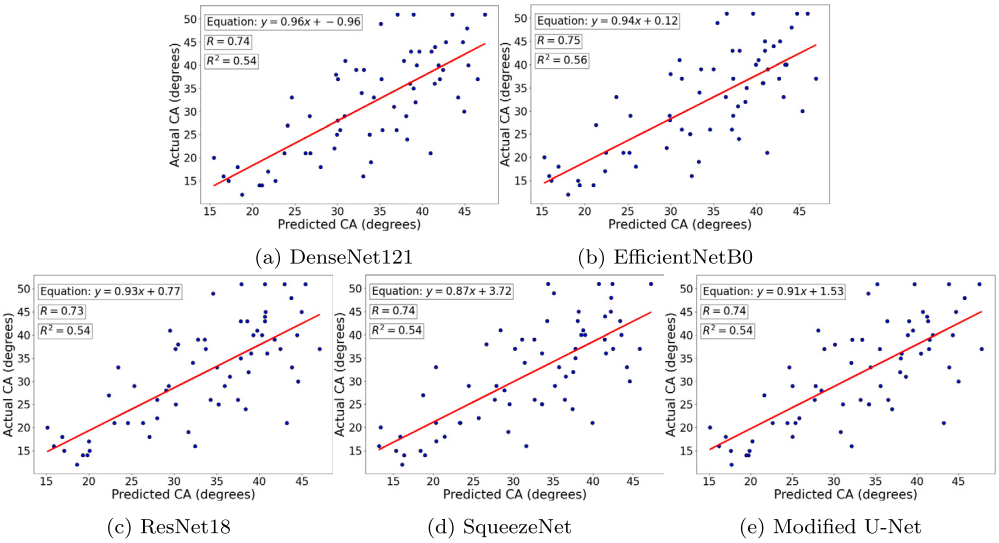


Fig. 6. Correlation between the actual and predicted value of CA of the testing set for the different architectures.

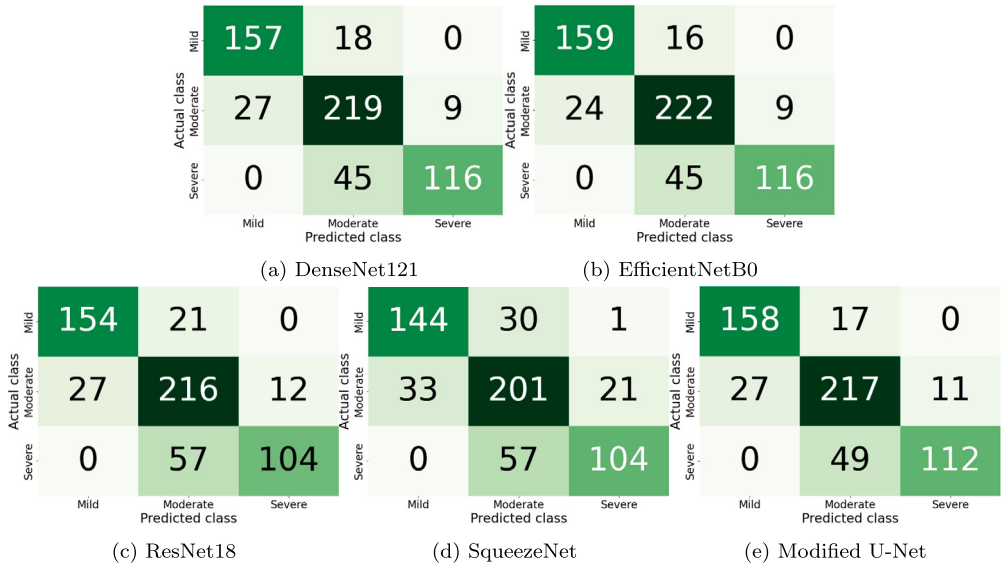
Table 5
Statistical significance analysis across all architectures.

Model	CV R (95% CI)[a]	Test R	CV R^2 (95% CI)[a]	Test R^2	Test MAE
DenseNet121	0.83 (0.82-0.84)	0.74	0.68 (0.67-0.69)	0.54	6.29°
EfficientNetB0	0.83 (0.82-0.84)	**0.75**	0.67 (0.66-0.68)	**0.56**	6.13°
ResNet18	0.83 (0.82-0.84)	0.73	0.68 (0.67-0.69)	0.54	6.14°
SqueezeNet	0.84 (0.83-0.85)	0.74	0.68 (0.67-0.69)	0.54	6.00°
U-Net	0.83 (0.82-0.84)	0.74	0.67 (0.66-0.68)	0.54	5.99°

[a] CV = Cross-validation results based on 100 models per architecture (10 repetitions × 10 folds).

ability to accurately categorize scoliosis severity based on ST. Additionally, the model's effectiveness was tested on an independent testing set to determine its reliability in classifying unseen torso scan data. Our analysis utilized a scoliosis severity classification based on CA ranges—mild($10° \leq CA < 25°$), moderate($25° \leq CA < 40°$), and severe($CA \geq 40°$)—to evaluate our model's accuracy on a training set and its reliability on an independent testing set. Figs. 7 and 8 present the confusion matrices for the training and testing datasets across the various CNNs. EfficientNetB0 demonstrated the highest classification accuracy on the training set However, in the testing set, the classification accuracy was less distinct, with DenseNet121, ResNet, and a modified U-Net exhibiting comparable performance levels.

To further elucidate the model's efficacy, performance classification metrics were applied to each clinical class, as detailed in Table 7. It was observed that different architectures excelled in specific classes; for example, SqueezeNet showed the highest NPV for the mild class in the testing set. Yet, when considering all classification metrics across both training and testing datasets, EfficientNetB0 emerged as the best CNN for AIS severity classification in torso scans.



(a) DenseNet121 (b) EfficientNetB0

(c) ResNet18 (d) SqueezeNet (e) Modified U-Net

Fig. 7. Confusion matrix of the training for the different architectures.

Table 6
Mean prediction analysis using resulting 100 CNN models of data points in the training set as a whole and the testing set.

		Training Set	Testing Set
DenseNet121	MAE (°)	2.61	6.29
	RMSE (°)	3.47	7.69
	R	0.96	0.74
	R^2	0.92	0.54
EfficientNetB0	MAE (°)	2.54	6.13
	RMSE (°)	3.36	7.50
	R	0.96	0.75
	R^2	0.93	0.56
ResNet18	MAE (°)	3.02	6.14
	RMSE (°)	4.02	7.59
	R	0.94	0.73
	R^2	0.89	0.54
SqueezeNet	MAE (°)	3.81	6.00
	RMSE (°)	5.05	7.46
	R	0.90	0.74
	R^2	0.81	0.54
Modified U-Net	MAE (°)	2.74	5.99
	RMSE (°)	3.64	7.53
	R	0.95	0.74
	R^2	0.91	0.54

4. Discussion

This study presents a comparative analysis of various CNN architectures, focusing on their efficacy in predicting the maximum CA from ST torso scans in patients with AIS. In previous studies, CNNs have utilized CA prediction from torso images or scans, as evidenced in the works of Watanabe et al. [32], Choi et al. [53,54], and Kokabu et al. [28]. Building upon Choi et al.'s foundational research, Watanabe et al. employed 1,996 pairs of moiré stereotopography ST images and 2D X-ray radiographs as input for a CNN to estimate spinal alignment. This approach involved using the CNN to align the spine to the moiré image using 17 vertebrae, followed by applying the B-spline method for CA measurement. Despite achieving high prediction accuracy, the study noted a considerable range of MAE across patients (2.6°–10.7°). However, Watanabe et al. [32], did not report additional regression performance metrics such as R and R^2 values. Moreover, their reliance on 2D X-ray images for alignment negates the advantage of using moiré scans for diagnosis, since X-ray data is already required.

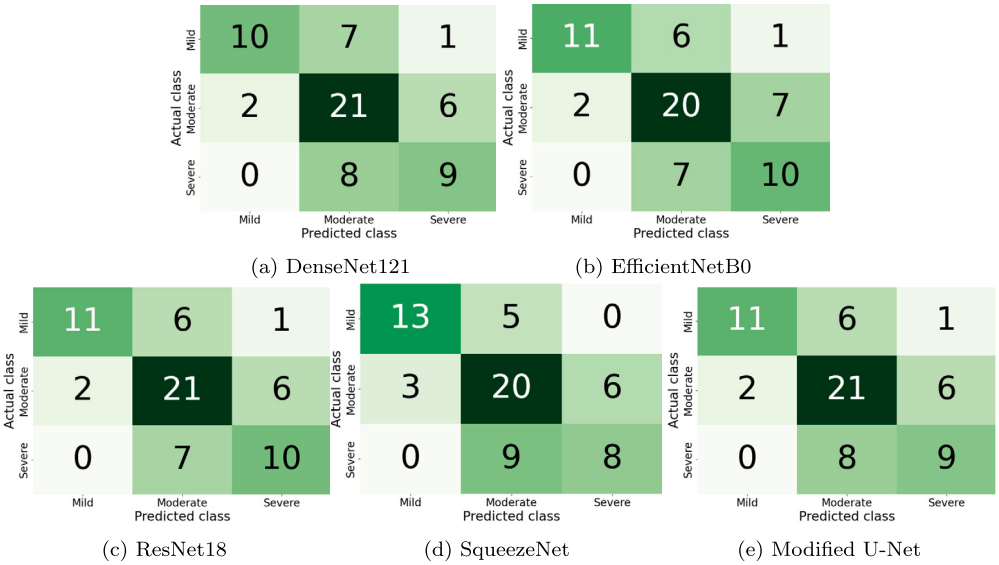
According to the literature, Kokabu et al. [28] appears to be a prominent study that has specifically applied a regression-based CNN approach to data derived from torso ST scans. They implemented a straightforward CNN for CA value prediction, utilizing cross-validation within their dataset and reporting an MAE range from 4.4° to 4.7° through repeated cross-validation. Additionally, when applying the CNN model to the entire dataset, they achieved an R-value of 0.91 and an R^2 value of 0.82. Nonetheless, this study has several limitations. The use of the entire dataset for training, without testing on new data, raises questions about the possibility of overfitting and how well their model might generalize to other datasets. Additionally, it would be beneficial to specify which of the 50 developed CNN models was employed when comparing actual and predicted Cobb Angle values. Finally, since the model was trained solely on data from the 3d depth sensor, its effectiveness with other scanning techniques remains uncertain.

In this study, a comprehensive comparison and evaluation of various CNN models were conducted for their application in CA prediction using ST torso scans. Among the models analyzed, EfficientNetB0 emerged as the most effective architecture for CA prediction, both statistically and clinically. In terms of statistical performance, EfficientNetB0 demonstrated a correlation coefficient (R) of 0.96 and a coefficient of determination (R^2) of 0.93 between the actual and predicted CA of the subjects in the training set. This performance surpasses that reported by Kokabu et al. ($R = 0.91$, $R^2 = 0.82$). Furthermore, the model showcased a Mean Absolute Error (MAE) range of 4.84° to 4.98° in repeated cross-validation runs. This range is slightly higher than Kokabu et al.'s reported range of 4.4° to 4.7°, a discrepancy that may be attributed to our study's use of the entire posterior half of the 3D scan, as opposed to the more focused area of interest employed by Kokabu et al.

The innovative nature of our approach lies in its expansion of the scope of analysis by utilizing the entire posterior half of the 3D scan, a methodology that, to the best of our knowledge, has not been previously implemented in the literature. By analyzing a more holistic anatomical region, the model captures subtle surface variations and asymmetries that may be overlooked in smaller, predefined areas of interest. This comprehensive analysis aligns closely with clinical practices, where a full assessment of the back surface is often critical for accurate diagnosis and monitoring of scoliosis. Furthermore, the inclusion of cross-validation to develop an ensemble prediction model enhances the robustness and generalizability of the results, mitigating overfitting and ensuring consistent performance across datasets. Together, these ad-

Table 7
Scoliosis clinical categories metrics of the training and the testing sets.

		Precision		Recall		F1-Score		Specificity		NPV	
		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
DenseNet121	Mild	85.33%	83.33%	89.71%	55.56%	87.47%	66.67%	93.51%	95.65%	95.58%	84.62%
	Moderate	77.66%	58.33%	85.88%	72.41%	81.56%	64.62%	81.25%	57.14%	88.35%	71.43%
	Severe	92.80%	56.25%	72.05%	52.94%	81.12%	54.55%	97.91%	85.11%	90.34%	83.33%
EfficientNetB0	Mild	86.89%	84.62%	90.86%	61.11%	88.83%	70.97%	94.23%	95.65%	96.08%	86.27%
	Moderate	78.45%	60.61%	87.06%	68.97%	82.53%	64.52%	81.85%	62.86%	89.29%	70.97%
	Severe	92.80%	55.56%	72.05%	58.82%	81.12%	57.14%	97.91%	82.98%	90.34%	84.78%
ResNet18	Mild	85.08%	84.62%	88.00%	61.11%	86.52%	70.97%	93.51%	95.65%	94.88%	86.27%
	Moderate	73.47%	61.76%	84.71%	72.41%	78.69%	66.67%	76.79%	62.86%	86.87%	73.33%
	Severe	89.66%	58.82%	64.60%	58.82%	75.09%	58.82%	97.21%	85.11%	88.00%	85.11%
SqueezeNet	Mild	81.36%	81.25%	82.29%	72.22%	81.82%	76.47%	92.07%	93.48%	92.51%	89.58%
	Moderate	69.79%	58.82%	78.82%	68.97%	74.033%	63.49%	74.11%	60.00%	82.18%	70.00%
	Severe	82.54%	57.14%	64.60%	47.06%	72.47%	51.61%	94.88%	87.23%	87.74%	82.00%
Modified U-Net	Mild	85.41%	84.62%	90.29%	61.11%	87.78%	70.97%	93.51%	95.65%	95.81%	86.27%
	Moderate	76.68%	60.00%	85.10%	72.41%	80.67%	65.62%	80.34%	60.00%	87.66%	72.41%
	Severe	91.06%	56.25%	69.57%	52.94%	78.87%	54.55%	97.44%	85.11%	89.53%	83.33%



(a) DenseNet121 (b) EfficientNetB0

(c) ResNet18 (d) SqueezeNet (e) Modified U-Net

Fig. 8. Confusion matrix of the testing set for the different architectures.

vancements establish a new benchmark for integrating machine learning into non-invasive scoliosis assessment, bridging the gap between clinical needs and advanced computational methodologies.

These findings underscore the necessity of employing advanced CNN architectures to achieve superior model fitting and information capture with reasonable complexity. A key feature of the EfficientNetB0 model is its utilization of compound coefficient scaling, which harmoniously scales model depth, width, and image resolution [37]. This approach effectively maintains the relative proportions of these parameters throughout the training process, unlike other architectures that scale these parameters randomly. This characteristic is particularly advantageous for ST torso scans, where the torso's shape and its deviations are intricately linked. By preserving this relationship, EfficientNetB0 enables more accurate quantification of image features, significantly enhancing the reliability of CA predictions in clinical settings. In this discussion, the predicted CA values from the CNN model are utilized in scoliosis severity classes (mild, moderate, and severe). This approach represents a shift from directly employing CNNs for classification purposes, as Our investigation reveals that utilizing CNNs directly for classification in such cases may not yield a fully representative understanding of the data. This is primarily due to the absence of discernible features, as these classes are not represented or connected to the ST torso scans of the patients.

Clinically, our model demonstrates notable proficiency in classifying scans into the mild severity class. It achieves a precision of 86.89% in training subjects and 84.62% in testing subjects, coupled with a high specificity of 94.23% in the training cohort and 95.65% in the testing set. These consistent results across training and testing groups suggest that the model is effectively trained on relevant features for the mild class, without apparent overfitting to specific local features within this class. In practical terms, for subjects with previously undetermined spinal parameters, our model accurately identifies mild cases 85% of the time. Moreover, it correctly recognizes non-mild conditions (moderate or severe) 95% of the time.

However, our study is limited by its relatively small dataset size of 654 data points, potentially leading to an under-representation of some critical features for a comprehensive analysis. This limitation is particularly evident in the performance gap observed between the training and testing sets, with the R^2 value on the testing set being approximately half of that on the training set. This substantial disparity highlights the model's susceptibility to overfitting, which poses challenges for its clinical implementation. Despite the use of data augmentation,

the current model exhibits significant overfitting, emphasizing the need for alternative approaches. To address this, a transfer learning framework is necessary, as further training of CNNs on the current dataset is unlikely to yield substantial improvements in performance. Future work should prioritize leveraging transfer learning techniques by utilizing pretrained CNN models to extract robust features. Subsequently, the training process can focus on regression models applied to the extracted feature vectors, potentially enhancing both generalizability and clinical applicability. Additionally, future studies could explore integrating machine learning-based measurement algorithms from ultrasound and X-ray radiographs. Combining these imaging modalities with surface topography could provide a more comprehensive framework for scoliosis assessment, improving both diagnostic precision and clinical utility.

Furthermore, the model demonstrated suboptimal performance in the severe category, which can be attributed to the limited amount of data available in the CA range of 50° to 60° compared to other categories. Generally, the severe category is less represented in our study, primarily because Many patients within this CA range typically undergo surgical procedures and were, therefore, excluded from our analysis. Typically, if this analysis were based on a classification machine learning technique, subsampling could be used to reduce the number of mild category data points and balance the dataset. However, since this is a regression problem, our focus is on representing the entire CA range to avoid introducing bias. Based on our analysis, training a machine learning algorithm to classify ST scans into mild, moderate, and severe categories is not advisable, as these classes are clinically derived and not directly correlated with features in the scans. Instead, the classification analysis was included to align with clinical categorization, facilitating its integration into routine clinical workflows.

Here, a significant divergence in model performance was observed, with a precision of 92.80% and specificity of 97.91% in the training set, compared to a notably lower precision of 55.56% and specificity of 82.98% in the testing set. This discrepancy underscores the need for additional data encompassing the severe class to adequately capture the full spectrum of features characteristic of this specific CA range.

A noteworthy advantage of our study is the diversity of scanning techniques employed in generating the scans used for analysis. This variety in the source of scans suggests that our model's effectiveness is not dependent on the specific setup of the scanning equipment. Instead, the model prioritizes focusing on the analysis of asymmetry. This aspect significantly broadens the potential utility of our model in diverse clinical environments, making it a versatile tool for spinal analysis.

5. Conclusions

In this paper, five CNN models were evaluated as feature extractors—DenseNet121, EfficientNetB0, ResNet18, SqueezeNet, and a modified U-Net—for predicting maximum CA in ST torso scans of AIS patients. The EfficientNetB0 emerged as the most promising, demonstrating high accuracy with an R of 0.96 and R^2 of 0.93, alongside an MAE of 6.13° and RMSE of 7.5° on unseen data. It also showed high precision (84.62%) and specificity (95.65% for mild and 82.98% for severe cases) in AIS severity classification. These results underscore EfficientNetB0's potential for precise CA estimation in AIS, offering significant advancements in diagnosing and managing scoliosis, particularly for mild cases. It suggests a pathway for integrating this model into clinical routines for more effective scoliosis severity classification, reducing reliance on invasive radiographic monitoring and aligning with non-invasive, patient-centered care practices.

Ethical approval statement

This research was conducted in strict adherence to all applicable national regulations and institutional guidelines, aligning with the principles outlined in the Helsinki Declaration. The study presents a secondary analysis of data that was originally gathered from participants across multiple centers, notably in Edmonton, Canada, and Campo Grande, Brazil. Ethical clearance for this analysis was obtained from the University of Alberta Health Research Ethics Board and the Universidade Santo Amaro - UNISA ethics committee, under approval number No. Pro00118197 and Pro00117065. Consent was secured from all participants in the original studies, ensuring the ethical integrity of the data utilized in this analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Canadian Institutes of Health Research (CIHR), WCHRI - Women and Children's Health Research Institute, and Alberta Innovates.

References

[1] Choudhry MN, Ahmad Z, Verma R. Adolescent idiopathic scoliosis. Open Orthop J 2016;10:143.
[2] Soucacos PN, Soucacos PK, Zacharis KC, Beris AE, Xenakis TA. School-screening for scoliosis. A prospective epidemiological study in northwestern and central Greece. J Bone Jt Surg 1997;79(10):1498–503.
[3] CĠLLĠ K, Tezeren G, Tas T, et al. School screening for scoliosis in sivas, Turkey. Acta Orthop Traumatol Turc 2009;43(5):426–30.
[4] Konieczny MR, Senyurt H, Krauspe R. Epidemiology of adolescent idiopathic scoliosis. J Child Orthop 2013;7:3–9.
[5] Wong H-K, Hui JH, Rajan U, Chia H-P. Idiopathic scoliosis in Singapore schoolchildren: a prevalence study 15 years into the screening program. Spine 2005;30(10):1188–96.
[6] Kittleson AC, Lim LW. Measurement of scoliosis. Am J Roentgenol 1970;108(4):775–7.
[7] Langensiepen S, Semler O, Sobottke R, Fricke O, Franklin J, Schönau E, et al. Measuring procedures to determine the Cobb angle in idiopathic scoliosis: a systematic review. Eur Spine J 2013;22:2360–71.
[8] Lonstein JE, Carlson J. The prediction of curve progression in untreated idiopathic scoliosis during growth. J Bone Jt Surg 1984;66(7):1061–71.
[9] Weinstein S, Ponseti I. Curve progression in idiopathic scoliosis. J Bone Jt Surg 1983;65(4):447–55.
[10] Ronckers CM, Doody MM, Lonstein JE, Stovall M, Land CE. Multiple diagnostic X-rays for spine deformities and risk of breast cancer. Cancer Epidemiol Biomark Prev 2008;17(3):605–13.
[11] Suzuki S, Yamamuro T, Shikata J, Shimizu K, Iida H. Ultrasound measurement of vertebral rotation in idiopathic scoliosis. J Bone Jt Surg, Br 1989;71(2):252–5.
[12] Chen W, Le LH, Lou EH. Ultrasound imaging of spinal vertebrae to study scoliosis. Open J Acoust 2012.
[13] Chen W, Lou EH, Zhang PQ, Le LH, Hill D. Reliability of assessing the coronal curvature of children with scoliosis by using ultrasound images. J Child Orthop 2013;7(6):521–9.
[14] Young M, Hill DL, Zheng R, Lou E. Reliability and accuracy of ultrasound measurements with and without the aid of previous radiographs in adolescent idiopathic scoliosis (ais). Eur Spine J 2015;24:1427–33.
[15] Zheng Y-P, Lee TT-Y, Lai KK-L, Yip BH-K, Zhou G-Q, Jiang W-W, et al. A reliability and validity study for scolioscan: a radiation-free scoliosis assessment system using 3d ultrasound imaging. Scoliosis Spinal Disord 2016;11:1–15.
[16] Girdler S, Cho B, Mikhail CM, Cheung ZB, Maza N, Cho SK-W. Emerging techniques in diagnostic imaging for idiopathic scoliosis in children and adolescents: a review of the literature. World Neurosurg 2020;136:128–35.
[17] Frerich JM, Hertzler K, Knott P, Mardjetko S. Comparison of radiographic and surface topography measurements in adolescents with idiopathic scoliosis. Open Orthop J 2012;6:261.
[18] Knott P, Sturm P, Lonner B, Cahill P, Betsch M, McCarthy R, et al. Multicenter comparison of 3d spinal measurements using surface topography with those from conventional radiography. Spine Deform 2016;4(2):98–103.
[19] Grant CA, Johnston M, Adam CJ, Little JP. Accuracy of 3d surface scanners for clinical torso and spinal deformity assessment. Med Eng Phys 2019;63:63–71.
[20] Goldberg CJ, Kaliszer M, Moore DP, Fogarty EE, Dowling FE. Surface topography, Cobb angles, and cosmetic change in scoliosis. Spine 2001;26.
[21] Ajemba PO, Durdle NG, Hill DL, Raso VJ. A torso-imaging system to quantify the deformity associated with scoliosis. IEEE Trans Instrum Meas 2007;56:1520–6.
[22] Poncet P, Delorme S, Ronsky JL, Dansereau J, Clynch G, Harder J, et al. Reconstruction of laser-scanned 3d torso topography and stereoradiographical spine and rib-cage geometry in scoliosis. Comput Methods Biomech Biomed Eng 2001;4(1):59–75.

[23] Komeili A, Westover LM, Parent EC, Moreau M, El-Rich M, Adeeb S. Surface topography asymmetry maps categorizing external deformity in scoliosis. Spine J 2014;14(6):973–83.
[24] Sanz-Pena I, Arachchi S, Curtis-Woodcock N, Silva P, McGregor AH, Newell N. Obtaining patient torso geometry for the design of scoliosis braces. A study of the accuracy and repeatability of handheld 3d scanners. Prosthet Orthot Int 2022;46(4):e374–82.
[25] Komeili A, Westover L, Parent EC, El-Rich M, Adeeb S. Correlation between a novel surface topography asymmetry analysis and radiographic data in scoliosis. Spine Deform 2015;3(4):303–11.
[26] Ghaneei M, Ekyalimpa R, Westover L, Parent EC, Adeeb S. Customized k-nearest neighbourhood analysis in the management of adolescent idiopathic scoliosis using 3d markerless asymmetry analysis. Comput Methods Biomech Biomed Eng 2019;22(7):696–705.
[27] Rothstock S, Weiss H-R, Krueger D, Paul L. Clinical classification of scoliosis patients using machine learning and markerless 3d surface trunk data. Med Biol Eng Comput 2020;58:2953–62.
[28] Kokabu T, Kanai S, Kawakami N, Uno K, Kotani T, Suzuki T, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. Spine J 2021;21(6):980–7.
[29] Shinde PP, Shah S. A review of machine learning and deep learning applications. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA). IEEE; 2018. p. 1–6.
[30] LeCun Y, Bengio Y, et al. Convolutional networks for images, speech, and time series. In: The handbook of brain theory and neural networks, vol. 3361. 1995. p. 1995.
[31] Vergari C, Skalli W, Gajny L. A convolutional neural network to detect scoliosis treatment in radiographs. Int J Comput Assisted Radiol Surg 2020;15:1069–74.
[32] Watanabe K, Aoki Y, Matsumoto M. An application of artificial intelligence to diagnostic imaging of spine disease: estimating spinal alignment from moiré images. Neurospine 2019;16(4):697.
[33] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. Squeezenet: alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. arXiv preprint. arXiv:1602.07360, 2016.
[34] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
[35] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4700–8.
[36] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference. Proceedings, part III 18. Springer; 2015. p. 234–41.
[37] Tan M, Le Q. Efficientnet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning; 2019. p. 6105–14.
[38] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017;42:60–88.
[39] Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, et al. Chexnet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint. arXiv:1711.05225, 2017.
[40] Komeili A. Adolescent idiopathic scoliosis: classifying, assessing, and monitoring using surface topography asymmetry analysis. Ph.D. thesis. University of Alberta; 2014.
[41] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.
[42] García S, Luengo J, Herrera F. Data preprocessing in data mining, vol. 72. Springer; 2015.
[43] Kohavi R, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Ijcai, vol. 14. 1995. p. 1137–45.
[44] Bishop C. Pattern recognition and machine learning. Springer Google Schola 2006;2:35–42.
[45] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[46] Mitchell TM. Machine learning; 1997.
[47] Chhabra M, Kumar R. A smart healthcare system based on classifier densenet 121 model to detect multiple diseases. In: Mobile radio communications and 5G networks: proceedings of second MRCN 2021. Springer; 2022. p. 297–312.
[48] Yang Y, Zhang L, Du M, Bo J, Liu H, Ren L, et al. A comparative analysis of eleven neural networks architectures for small datasets of lung images of covid-19 patients toward improved clinical decisions. Comput Biol Med 2021;139:104887.
[49] Ke A, Ellsworth W, Banerjee O, Ng AY, Rajpurkar P. Chextransfer: performance and parameter efficiency of imagenet models for chest X-ray interpretation. In: Proceedings of the conference on health, inference, and learning; 2021. p. 116–24.
[50] Yi W, Zhao J, Tang W, Yin H, Yu L, Wang Y, et al. Deep learning-based high-accuracy detection for lumbar and cervical degenerative disease on t2-weighted mr images. Eur Spine J 2023;32(11):3807–14.
[51] Singh A, Prakash S, Das A, Kushwaha N. Colonnet: a hybrid of densenet121 and u-net model for detection and segmentation of gi bleeding. arXiv preprint. arXiv:2412.05216, 2024.
[52] Huber PJ. Robust estimation of a location parameter. In: Breakthroughs in statistics: methodology and distribution. Springer; 1992. p. 492–518.
[53] Choi R, Watanabe K, Jinguji H, Fujita N, Ogura Y, Demura S, et al. Cnn-based spine and Cobb angle estimator using moire images. IIEEJ Trans Image Electron Vis Comput 2017;5(2):135–44.
[54] Choi R, Watanabe K, Fujita N, Ogura Y, Matsumoto M, Demura S, et al. Measurement of vertebral rotation from moire image for screening of adolescent idiopathic scoliosis. IIEEJ Trans Image Electron Vis Comput 2018;6(2):56–64.