

Research Paper

Automated measurement of lumbar pedicle screw parameters using deep learning algorithm on preoperative CT scans

Qian Zhang^{a,b,c,1}, Fanfan Zhao^{c,1}, Yu Zhang^{a,1}, Man Huang^a, Xiangyang Gong^{b,c,*}, Xuefei Deng^{d,**}

^a Department of Radiology, The 901st Hospital of the Joint Logistics Support Force of PLA, Hefei 230031, China

^b Soochow University, Soochow 215000, China

^c Department of Radiology, Zhejiang Provincial People's Hospital (Affiliated People's Hospital), Hangzhou Medical College, Hangzhou 310014, China

^d Department of Anatomy, Anhui Medical University, Hefei 230032, China

HIGHLIGHTS

- Developed an automated framework for lumbar pedicle screw measurement using VGG16 on CT scans.
- The model identifies screw placement landmarks with a PCK value of 93.5% to 99.0%.
- Showed high agreement with radiologists and a spinal surgeon, with ICCs from 0.82 to 0.98.
- Showcased the model's potential for efficient, precise lumbar pedicle screw measurements.

ARTICLE INFO

Keywords:

Deep learning-based algorithm
Pedicle screw
Computed tomography
Lumbar spine

ABSTRACT

Purpose: This study aims to devise and assess an automated measurement framework for lumbar pedicle screw parameters leveraging preoperative computed tomography (CT) scans and a deep learning algorithm.

Methods: A deep learning model was constructed employing a dataset comprising 1410 axial preoperative CT images of lumbar pedicles sourced from 282 patients. The model was trained to predict several screw parameters, including the axial angle and width of pedicles, the length of pedicle screw paths, and the interpedicular distance. The mean values of these parameters, as determined by two radiologists and one spinal surgeon, served as the reference standard.

Results: The deep learning model achieved high agreement with the reference standard for the axial angle of the left pedicle (ICC = 0.92) and right pedicle (ICC = 0.93), as well as for the length of the left pedicle screw path (ICC = 0.82) and right pedicle (ICC = 0.87). Similarly, high agreement was observed for pedicle width (left ICC = 0.97, right ICC = 0.98) and interpedicular distance (ICC = 0.91). Overall, the model's performance paralleled that of manual determination of lumbar pedicle screw parameters.

Conclusion: The developed deep learning-based model demonstrates proficiency in accurately identifying landmarks on preoperative CT scans and autonomously generating parameters relevant to lumbar pedicle screw placement. These findings suggest its potential to offer efficient and precise measurements for clinical applications.

1. Introduction

In recent years, the utilization of posterior pedicle screw placement has emerged as a predominant surgical approach for managing spinal

degenerative pathologies and traumatic injuries [1]. Significant variations exist in the precision of pedicle screw insertion, with inadequate screw positioning reported in 2 % to 31 % of cases. Additionally, post-operative monitoring has documented incidences of screw malposition

* Corresponding author at: Zhejiang Provincial People's Hospital, Hangzhou 310014, China (X. Gong).

** Corresponding author at: Anhui Medical University, Hefei 230032, China (X. Deng).

E-mail addresses: cjr.gxy@hotmail.com (X. Gong), dengxf@ahmu.edu.cn (X. Deng).

¹ These authors contributed equally to the study.

or loosening, ranging from 0 % to 42 % [2,3]. Proper screw placement was imperative in order to avoid serious complications such as nerve, vascular, or intervertebral disc injury, while also providing a strong stable construct [4–7]. Each of related parameters for measuring screw placement in preoperative computed tomography (CT) image is directly related to the accuracy of pedicle screw placement. Measurement is needed for selecting screw size and proper screw placement, including the axial angle, length of pedicle, the pedicle width and the distance between the two pedicles, to provide accurate preoperative planning in preoperative CT image for physicians. Traditional assessments of screw placement parameters have predominantly relied upon manual annotation and measurement techniques. Such methodologies are characterized by their labor-intensive nature and subjectivity, posing challenges for radiologists. Consequently, there exists considerable variability in the consistency and precision of parameter measurements among radiologists [8].

The integration of deep learning has witnessed a notable surge within the realm of medical imaging [9–11]. Particularly in spinal measurements, deep learning algorithms have found application in assessing a spectrum of parameters including the Cobb angle of the spine, vertebral body compression, and the evaluation of cervical spondylotic myelopathy [12–14]. However, in previous literature, measurements were mostly based on radiography rather than CT images, and there were few reports on using deep learning algorithms to automatically measure parameters related to lumbar screw placement. Drawing upon insights gleaned from clinical practice, the pivotal aspect in the assessment of parameters pertinent to lumbar pedicle screw placement lies in the precise identification and localization of anatomical landmarks [15,16]. Thus, the primary objective of this investigation entailed the development and assessment of a deep learning-driven model tailored for the automated quantification of parameters associated with lumbar pedicle screw placement on preoperative CT scans, with a specific focus on evaluating its efficacy and applicability across diverse clinical scenarios.

2. Materials and methods

Approval for this study was obtained from the Ethics Committee of Zhejiang Provincial People's Hospital (ZJPP Hospital). Utilizing anonymized patient data within a retrospective study framework obviated the requirement for explicit informed consent.

2.1. Subjects

We reviewed preoperative CT images at ZJPP Hospital through the PACS system from January 2017 to December 2022. This study collects patients who are clinically planned for pedicle screw fixation. Inclusion criteria comprised patients aged 18 years or older who had undergone lumbar CT (Siemens SOMATOM Definition 3.0) scans and were clinically slated for pedicle screw fixation. This retrospective analysis involved a cohort of 2589 patients. To uphold the reliability of manual labeling, cases presenting with indiscernible landmarks pertaining to screw placement parameters were omitted from the analysis. The exclusion criteria encompassed: (1) vertebral compression or burst fracture ($n = 989$); (2) history of lumbar spine surgery or significant calcification in vertebral bone ($n = 1245$); (3) bone defects in the lumbar spine due to surgery or congenital variations ($n = 52$); (4) there was a significant congenital variation or acquired deformity in the shape of the vertebral body ($n = 21$). After exclusion criteria, a total of 282 patients remained, and each patient could reconstruct five axial CT images parallel to the lumbar 1 to lumbar 5 pedicle, respectively. In the culmination of our efforts, we amassed a corpus of 1410 axial CT images depicting various lumbar pedicles. From this dataset, a subset of 200 axial CT images of lumbar pedicles was randomly allocated for use as the test set. The remaining 1210 images were further divided: 847 images (70 %) were allocated for the training set, and 363 images (30 %) were

used as the validation set to refine and adjust model hyperparameters (Fig. 1).

2.2. Definitions and landmark annotation

For model evaluation, seven pertinent parameters associated with screw placement were chosen based on preoperative lumbar CT images: the axial angle of the left and right pedicles, the length of the screw path for both sides, the width of the left and right pedicles, and the interpedicular distance [17,18] (Fig. 2). A total of 14,520 landmarks were meticulously annotated at the pixel-level from 1210 axial CT images depicting the lumbar pedicle. This annotation process was conducted collaboratively by two radiologists and one spinal surgeon utilizing a dedicated artificial intelligence data annotation platform (<https://warehouse.healthviewcn.com/>).

A total of 14,520 landmarks, meticulously labeled by an expert manual annotator certified by the board, who held the qualification of a radiologist (referred to as M1), served as the foundation for subsequent development of algorithm. Following model training, the algorithm demonstrated the capability to predict landmark locations without the necessity of explicit labeling. To ascertain the model's performance, a test set comprising 200 images was annotated independently by M1, alongside two additional manual annotators, denoted as M2 and M3, with diverse backgrounds—one being a radiologist and the other a surgeon. In the absence of an objective reference standard parameter, the mean of measurement parameters, as determined collectively by two radiologists and one spinal surgeon, was employed as the benchmark [19]. Furthermore, to evaluate intra-observer reliability, M1 reassessed the test set after a 4-week interval.

2.3. Dataset preparation

The dataset preparation phase involved comprehensive data pre-processing procedures conducted utilizing Python (version 3.11.1) and the ITK library (<https://www.simpleitk.org>). Notably, all CT images incorporated in the study underwent anonymization processes to ensure patient confidentiality. Augmentation of the training set was facilitated by incorporating random rotation within the range of -40° to $+40^\circ$, along with random scaling spanning from 0.7 to 1.3. Subsequently, standardization of the lumbar pedicle CT images was achieved through resizing to a uniform resolution of 512×512 pixels. The CT images parameters were as follows: window width:1500; window level:450; tube voltage:120 Kv; tube current: 20 mA ~ 100 mA; raw pixel size: 0.143×0.143 mm.

2.4. Measurement model development

Within the framework of the automated measurement model aimed at assessing parameters associated with lumbar pedicle screw placement, two pivotal components stand out: a landmark detection convolutional neural network (CNN) and an algorithm dedicated to mathematical calculation of screw parameters. The training phase revolves around the meticulous optimization of CNN parameters, a process necessitating the iterative exposure of training images to the network architecture. Through continuous evaluation of errors, adjustments to network connection weights are made, honing the model's predictive capabilities. Integral to the training regimen is the utilization of a validation set, tasked with scrutinizing the model's performance throughout the training iterations. Notably, this dataset assumes a dual role: not only does it serve as a litmus test for the model's efficacy, but it also plays a pivotal role in facilitating judicious model selection. By subjecting the model to rigorous evaluation against the validation set, researchers can discern the optimal configuration that strikes a balance between predictive accuracy and generalization prowess.

The landmark detection network employed in this study is a tailored variant of the U-net architecture, distinguished by the integration of

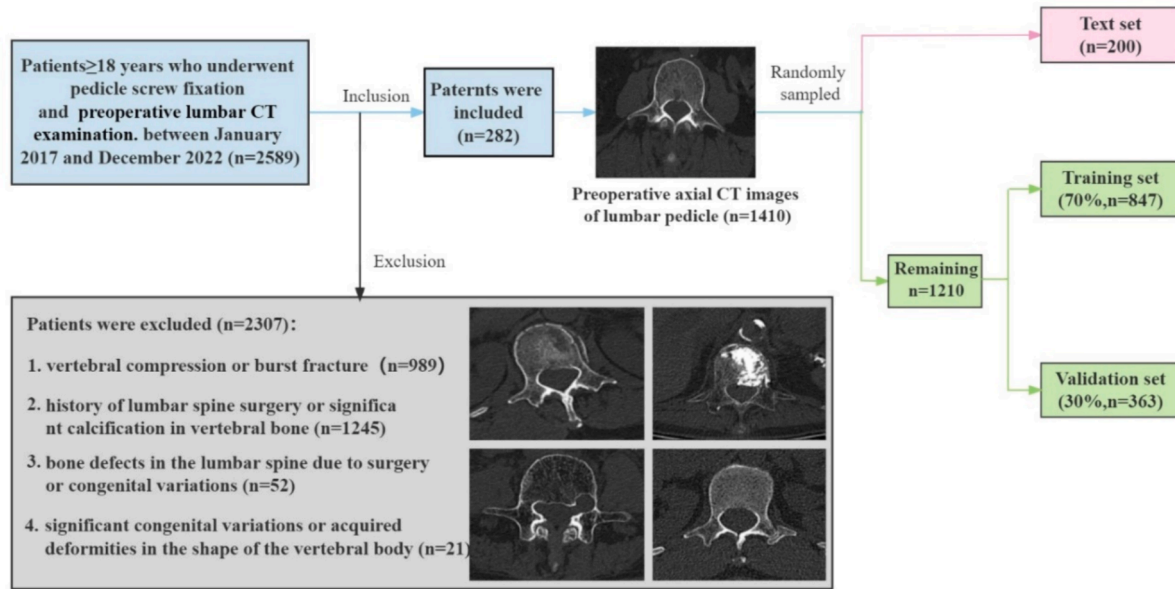


Fig. 1. Flowchart shows number of patients meeting inclusion and exclusion criteria.

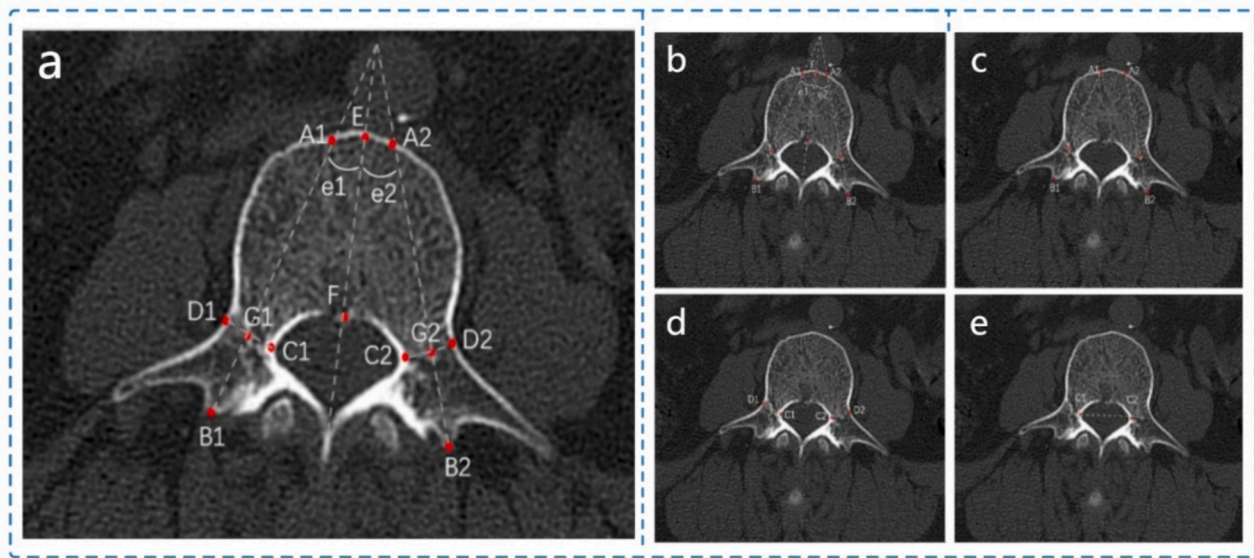


Fig. 2. Landmark annotations and parameter for assessing pedicle screw placement. (a) The landmarks C, D and G, denote the medial margin, outer margin, and midpoint of the narrowest width of right (1) and left (2) pedicles. Points A and B signify the intersection points between a straight line parallel to the pedicle and the anterior and posterior vertebral cortices passing through G. Point E marks the intersection between the midline and the anterior vertebral cortex, while point F designates the intersection between the midline and the posterior vertebral cortex. (b) Axial angle of the right and left pedicle (e1) (e2). (c) The length of bilateral pedicle screw path (A1B1, A2B2). (d) The bilateral pedicle width (C1D1, C2D2). (e) The distance between the two pedicles (C1C2).

ResNet34 as the encoder component. Meanwhile, the decoder segment retains the conventional U-net architecture, ensuring continuity and coherence within the network structure. At the core of this modified U-net architecture lies a pivotal 3×3 convolution layer, strategically positioned to yield a multi-channel heat map output [20]. Significantly, the pixel intensities within each channel of the heat map encode the probability distribution of key points essential for landmark localization. In addition to its architectural framework, the decoder network is distinguished by its incorporation of four consecutive bilinear upsampling layers, each subsequently paired with two convolutional layers [21]. These convolutional layers feature rectified linear activation (ReLU) [22] for enhanced gradient propagation and a batch normalization layer (BN) to stabilize and expedite training convergence.

Training of the CNN was executed utilizing TensorFlow (version 1.4), employing a mini-batch size of 8 to balance computational efficiency and model convergence. Furthermore, to augment the performance of the landmark detection network, pre-trained VGG16 weights derived from the ImageNet dataset (www.image-net.org/challenges/LSVRC/) were leveraged, enhancing the network's capacity for effective feature extraction and representation learning, particularly in scenarios characterized by limited data availability [23–26].

The landmark detection network based on VGG16 enables efficient transfer learning through the initialization of weights specifically tuned for the ImageNet dataset. This approach ensures the creation of a robust and generalizable model, particularly when dealing with limited training data. To guide the optimization process during training, a Mean

Squared Error loss function is employed in conjunction with the Adam optimizer. Notably, the learning rate undergoes a gradual linear decrease from $5e-4$ to $1e-6$, facilitating stable and consistent convergence of the model parameters. Moreover, the utilization of the Online Hard Key-points Mining technique serves to explicitly select crucial key points, enhancing the network's capacity for landmark detection by focusing on salient anatomical features.

Following rigorous training and model selection based on performance evaluation against the validation set, the CNN model is poised to demonstrate its efficacy on the test set. Subsequently, post-sufficient training, the CNN autonomously predicts landmarks within the test set, obviating the need for manual annotation by annotators. The predicted landmark coordinates are subsequently leveraged to compute the seven parameters pertinent to lumbar pedicle screw placement through automated algorithms implemented in Python and Pytorch (version 2.4.0). This streamlined process not only ensures efficiency but also minimizes potential errors associated with manual calculations. A schematic depiction of the study's experimental design and workflow is succinctly summarized in Fig. 3, elucidating the systematic progression from data acquisition to parameter calculation.

2.5. Statistical analysis

All statistical analyses were conducted using MedCalc version 20.2. Statistical significance was established at a two-sided p-value threshold of less than 0.05, indicative of significance. Categorical variables were compared utilizing chi-square tests, while the comparison among quantitative variables was facilitated through either t-tests or Mann-Whitney U-tests, depending on the nature of the data distribution and the equality of variances.

2.6. Reliability of landmark annotation

To comprehensively assess the intra- and inter-observer reliability of manual annotation, computations were undertaken to determine the percentage of landmark-to-landmark distances falling within specified thresholds of 1 mm, 2 mm, and 3 mm. This approach enabled a nuanced evaluation of the consistency and agreement levels between observers, shedding light on the precision of manual annotation across varying degrees of spatial proximity.

2.7. Landmark performance

To assess the efficacy of our measurement model in predicting landmarks, we utilized the Percentage of Correct Key Points (PCK) metric. PCK measures the proportion of predicted landmarks that fall within a normalized distance from a reference standard landmark, providing a quantitative evaluation of predictive accuracy. In the absence of definitive objective reference standards, the averages derived from assessments conducted by two radiologists and one spinal surgeon were utilized as the reference standard. This collective approach ensured a comprehensive and robust benchmark for evaluating the predictive performance of the model across diverse clinical contexts.

2.8. Model measurement performance

The measurements derived from assessments by two radiologists and one spinal surgeon were juxtaposed with the model's predictions across the seven parameters. In cases of differing opinions, consensus was reached through discussion. To comprehensively assess the overall performance of our measurement model, several statistical metrics were employed for comparison. These included the intraclass correlation coefficient (ICC), which serves as a metric of consistency, with ICC values equal to or exceeding 0.75 deemed indicative of satisfactory reliability [27]. Additionally, the Pearson or Spearman correlation coefficient (r) was utilized to gauge the strength and direction of the relationship between the reference standard and the model estimate. A correlation coefficient $|r| \geq 0.7$ signifies a robust correlation between the observed and predicted values.

Furthermore, the mean absolute error (MAE) and root mean square error (RMSE) were calculated to quantify the magnitude of the discrepancies between observed and predicted values. They provide a measure of the model's predictive precision. In addition to these metrics, violin and box plots were generated to visually depict the mean difference and distribution of parameter values. The statistical significance of the observed differences was determined through paired t-tests and Wilcoxon rank-sum tests, further elucidating the comparative performance of the model across the seven parameters of interest.

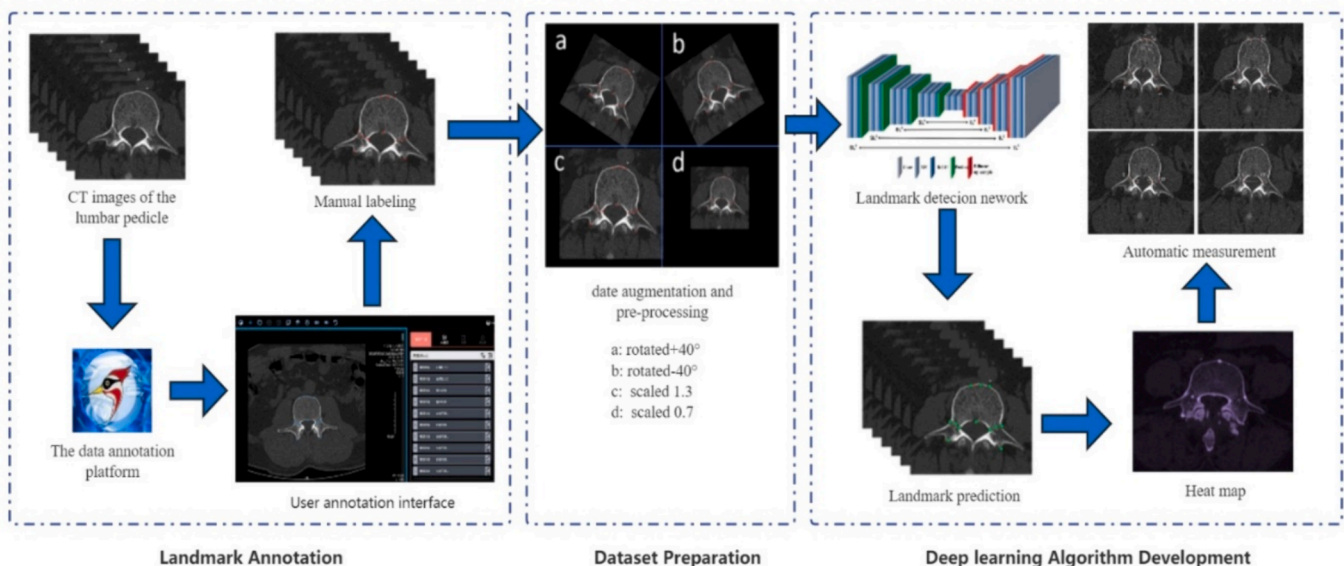


Fig. 3. Flowchart of the study design.

3. Results

3.1. General data distributions

Incorporated into the study were a total of 1410 preoperative axial CT images focusing on the lumbar pedicle as the subjects of analysis. Notably, thorough examination revealed no discernible disparities among the datasets concerning gender composition and age distribution, as illustrated comprehensively in Table 1. This congruence in demographic characteristics underscores the appropriateness of the dataset segmentation, ensuring robustness and consistency in subsequent analyses.

3.2. Reliability of landmark

The proportion of intra-observer landmark distances falling within the 3 mm threshold exhibited complete agreement at 100 %. Conversely, the inter-observer comparison revealed percentages of 98.54 % (M1 vs M2), 95.75 % (M2 vs M3), and 96.36 % (M1 vs M3), respectively, within the same threshold. Detailed breakdowns of intra- and inter-observer landmark distances within the 1 mm and 2 mm thresholds are delineated in Table 2 for comprehensive evaluation.

3.3. Landmark performance

At the 3 mm distance threshold, the PCK for landmark prediction consistently exceeded 93 %. Notably, the highest PCK value at this threshold was observed for the intersection point between the midline and the anterior vertebral cortex (E), while the lowest values were recorded for the intersection points between the parallel line to the pedicle and the anterior vertebral cortex (A1, A2) (Fig. 4, Table 3). Furthermore, Fig. 5 illustrates a representative example showcasing the predictive capabilities of our model in landmark measurement, offering a visual demonstration of its efficacy in action.

3.4. Model measurement performance

Regarding the axial angle of the pedicle, the reference standards were 15.59(12.14, 20.11)° (left) and 16.47(12.16, 20.03)° (right), while the models were 16.13(12.64, 19.83)° (left) and 16.13(12.56, 20.88)° (right). For the length of pedicle screw path, the reference standards were 53.91 ± 5.77 mm (left) and 53.67 ± 5.91 mm (right), while the models were 53.71 ± 5.76 mm (left) and 53.39 ± 6.23 mm (right). For

Table 1
Data distributions of various data sets.

Characteristics	Training data	Validation data	Text data	p-value (statistical value)
Number of images	847	363	200	
Sex (%) ^a				0.090(4.806 ^c)
Female	395(46.64)	193(53.17)	92 (46.00)	
Male	452(53.37)	170(46.83)	108 (54.00)	
Age (years) ^b	58(50,66)	59(51,66)	59 (50,67)	0.437(1.655 ^d)
Female	58(51,69)	59(53,68)	57 (50,67)	0.333(2.197 ^d)
Male	58(49,65)	60(49,66)	61 (52,67)	0.145(3.866 ^d)

^a Data are presented in terms of object counts, accompanied by corresponding percentages within parentheses.
^b Data are articulated in terms of medians, alongside the interquartile range denoted by the 25th and 75th percentiles in parentheses.
^c χ^2 .
^d H value.

Table 2
The intra-observer and inter-observer reliability of landmark annotation (%).

	Distance threshold		
	1 mm	2 mm	3 mm
Intra-observer reliability	72.58	98.98	100.00
Inter-observer reliability			
M1 vs M2	66.24	95.58	98.54
M1 vs M3	62.75	91.31	95.75
M2 vs M3	63.38	93.68	96.36

the pedicle width, the reference standards were 6.39(4.65, 7.87) mm (left) and 6.23(4.52, 7.82) mm (right) and the models were 6.23(4.74, 7.99) mm (left) and 6.30(4.51, 7.78) mm (right). For the distance between the two pedicles, the reference standards were 25.96(24.55, 29.86) mm and 26.02(24.36, 28.96) mm (Table 4).

Upon comprehensive evaluation of the model's overall performance against the reference standard, meticulous measurements were obtained for the seven parameters. The calculated metrics revealed robust consistency, with intraclass correlation coefficients (ICC) ranging from 0.81 to 0.98, correlation coefficients (r) ranging from 0.82 to 0.98, root mean square error (RMSE) values spanning from 0.48 to 3.46, and mean absolute error (MAE) values ranging from 0.33 to 2.23, as illustrated in Table 5. Additionally, to provide a visual depiction of the parameter distributions, violin and box plots analyses were conducted, showcasing the variability and central tendencies of the seven parameters, as depicted in Fig. 6.

4. Discussion

Accurate and dependable assessment of parameters associated with lumbar pedicle screw placement holds significant relevance in procedures such as posterior pedicle screw placement, which encompass widely utilized the minimally invasive spinal surgery [28]. As is well known, manual measurement is tedious, inefficient, and relying on the personal experience of spinal surgeons and radiologist can easily lead to significant variability [29]. This study employed preoperative lumbar CT scans alongside a deep learning algorithm to construct an automated measurement model for parameters concerning screw placement. Key findings from our analysis on the test set revealed the following: 1) the model successfully identified landmarks automatically, achieving a PCK exceeding 93 % at the 3 mm distance threshold; 2) precise measurements of various parameters, including the axial angle of pedicle, length of pedicle screw path, width of pedicle, and the distance between the two pedicles, were attained with ICC values ranging from 0.82 to 0.98; and 3) the performance of the developed automatic measurement model was comparable to the reference standards obtained from clinicians' measurements ($p > 0.05$).

Assuring the precision of manual annotations stands as a significant hurdle in assessing the efficacy of this model, as it directly impacts the fidelity of the model's landmark predictions. The main professionals engaged in the measurement of parameters related to screw placement are radiologists and spinal surgeons. Based on this, the model annotator selected radiologists and spinal surgeon to ensure the generalization of labeling. As per the findings presented, the discrepancy observed between observers averaged approximately 2.98 mm, a measurement deemed suitable for both biomechanical and clinical analyses [30]. Within the scope of our investigation, a substantial proportion of inter-observer distances, ranging from 93.5 % to 98.5 %, were contained within the 3 mm threshold, underscoring the consistency of performance exhibited by our manual annotator.

The performance of our model exhibited excellence, achieving PCK values ranging from 93.5 % to 98.5 % within a 3 mm distance threshold range for predicting landmarks. Notably, the predictive accuracy for landmarks A1, A2, B1, and B2 was comparatively lower when juxtaposed with other landmarks. In some cases, this may be due to the fact

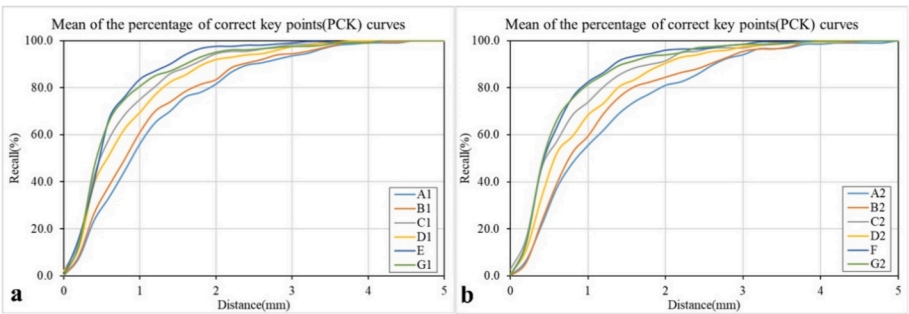


Fig. 4. Performance evaluation of the model in landmark detection. (a) PCK diagram of key points A1 ~ D1, E, and G1; (b) PCK diagram of key points A2 ~ D2, F, and G2.

Table 3
The PCK values of landmarks (%).

Distance threshold (mm)	A1	B1	C1	D1	E	G1	A2	B2	C2	D2	F	G2
1	56.00	61.00	75.00	69.50	83.50	80.50	55.00	59.50	74.00	68.50	82.50	81.00
2	81.50	83.50	94.50	92.00	97.50	95.50	81.00	84.50	91.50	90.00	96.00	94.00
3	93.50	94.50	98.50	97.50	99.00	97.50	94.00	95.50	97.00	97.50	98.50	98.50
4	99.00	99.00	100.00	100.00	99.00	99.00	98.50	99.50	100.00	99.50	100.00	100.00

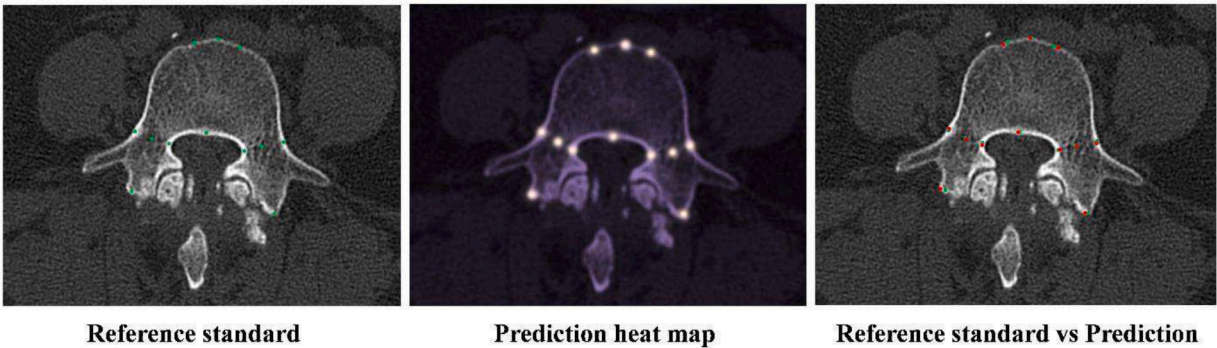


Fig. 5. Representative illustrations of landmark detection by the proposed model. On the left side, the mean values derived from R1, R2, and R3 (reference standard) are depicted. In the middle, the heat map representing the model's predictions is showcased. Additionally, the right side provides a visual comparison illustrating the disparities between the reference standard and the model's predictions, thereby offering insights into the model's performance and accuracy in landmark detection.

Table 4
The measured values of manual annotators are compared to the model estimates of the seven parameters.

Screw parameters	Reference standard	Model	statistical value	p-value
Axial angle of left pedicle(°)	15.59(12.14, 20.11)	16.70 (12.64, 19.83)	-1.432 ^b	0.151
Axial angle of right pedicle(°)	16.47(12.16, 20.03)	16.13 (12.56, 20.88)	-0.694 ^b	0.493
Length of left pedicle screw path [#] (mm)	53.91 ± 5.77	53.71 ± 5.76	0.827 ^a	0.412
Length of right pedicle screw path [#] (mm)	53.67 ± 5.91	53.39 ± 6.23	1.292 ^a	0.204
Left pedicle width (mm)	6.39(4.67, 7.84)	6.23(4.75, 7.99)	-1.953 ^b	0.507
Right pedicle width (mm)	6.23(4.74, 7.99)	6.30(4.51, 7.78)	-1.154 ^b	0.247
Distance between the two pedicles(mm)	25.96(24.55, 29.86)	26.02 (24.36, 28.96)	-1.919 ^b	0.061

^a t value.
^b z value.
[#] Data conforms to the Shapiro-Wilk normal distribution test.

Table 5
Comparison of model estimates and reference standards for parameters related to screw placement.

Screw parameters	ICC (95 %CI)	r	MAE	RMSE
Axial angle of left pedicle(°)	0.92 (0.89,0.94)	0.901 ^{b*}	1.614	2.393
Axial angle of right pedicle(°)	0.93 (0.90,0.94)	0.913 ^{b*}	1.592	2.332
Length of left pedicle screw path (mm)	0.82 (0.77,0.86)	0.824 ^{a*}	2.242	3.471
Length of right pedicle screw path (mm)	0.87 (0.83,0.90)	0.868 ^{a*}	1.774	3.092
Left pedicle width(mm)	0.97 (0.95,0.97)	0.973 ^{b*}	0.368	0.573
Right pedicle width(mm)	0.98 (0.97,0.98)	0.972 ^{b*}	0.332	0.483
Distance between the two pedicles (mm)	0.91 (0.89,0.94)	0.879 ^{b*}	1.254	1.431

^a Pearson correlation coefficient; ^b Spearman correlation coefficient; * $p < 0.01$.

that making a straight line parallel to the pedicle of the pedicle is more subjective compared to other landmarks with obvious anatomical structures. Moreover, our deep learning-driven automatic measurement model demonstrates favorable concordance and precision when

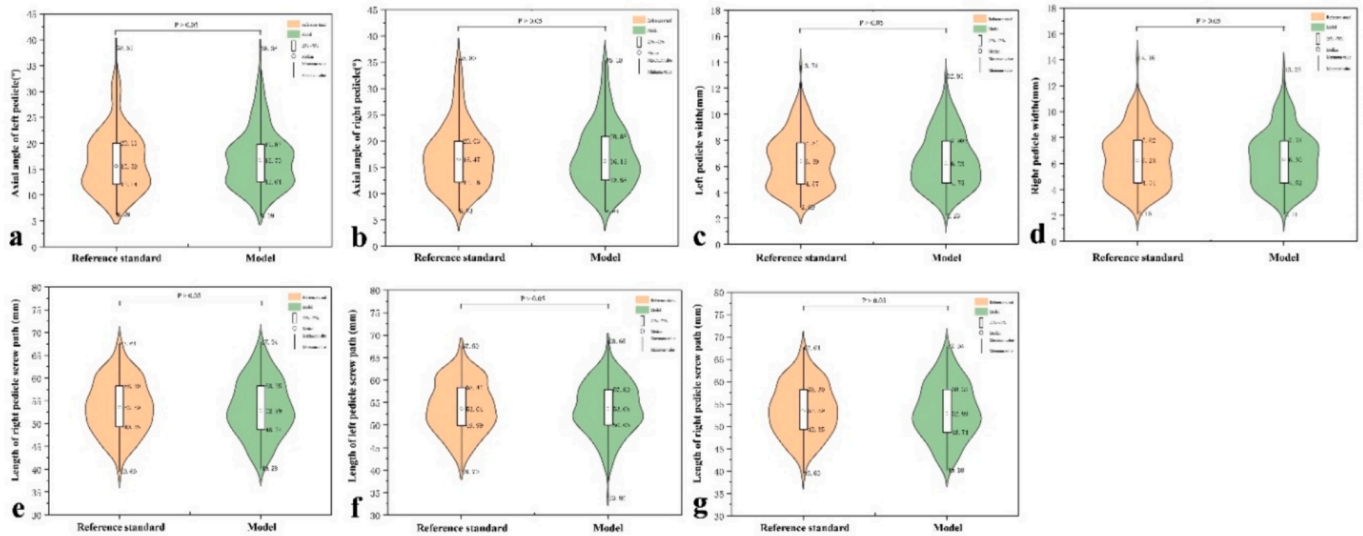


Fig. 6. Visualization of discrepancies in pedicle screw placement parameters. (a) Axial angle of left pedicle. (b) Axial angle of right pedicle. (c) Left pedicle width. (d) Right pedicle width. (e) Distance between the two pedicles. (f) Length of left pedicle screw path. (g) Right of left pedicle screw path.

compared to the reference standard for parameters associated with screw placement derived from preoperative lumbar CT images. The calculated ICC and r values, spanning from 0.82 to 0.98, along with MAE values ranging from 0.33 to 2.24, attest to the model's ability to accurately anticipate the seven frequently employed parameters linked to screw placement, thus aligning closely with the proficiency level of clinicians.

In the previous study, the intra-operative 3D CT or image guidance utilizing O-arm or Stealth navigation was employed to furnish screw-related parameters [1,31]. Notably, there existed a diversity in surgeon expertise across these studies, a fundamental aspect contributing to potential variations in outcomes. These require the purchase of additional medical equipment or software, which increases the difficulty of promoting in grassroots hospitals. Our deep learning-based automatic measurement model has better generalizability. Certain studies employed bibliometric and visual methodologies to retrieve articles related to spinal conditions using deep learning approaches. They found that most articles focused on algorithms for diagnosing spinal diseases, while there were not many articles based on automatic measurement [32]. In previous literature, the development of automatic measurement algorithms was mostly based on radiography [33,34]. Naik et al [35] formulated a hybrid registration framework that integrates 3D-2D Iterative Control Point (ICP) based on anatomical landmarks. This method enhances the alignment of pedicle screw trajectories between intra-operative X-ray images and preoperative CT scans. Nevertheless, this trajectory registration technique primarily relied on anatomical landmark alignment and generated projection images, limiting its applicability to scenarios primarily within surgical interventions where surgeons manipulate surgical tools. In contrast, our model achieves automatic landmark localization directly through landmark detection, thereby reducing annotation time and computational overhead while furnishing clinicians with essential measurement parameters for diverse lumbar interbody fusion procedures. To enhance the clinical applicability of the algorithm, all preoperative CT scans were acquired prior to actual surgical interventions.

Additionally, except for the left and right length of pedicle screw path, all other five parameters do not follow a normal distribution and instead exhibit two peaks. The main reason for the non-normal distribution observed is that the anatomical morphology of the 5th lumbar vertebral body differs significantly from other lumbar vertebral bodies, resulting in significant differences in most parameter measurements compared to other vertebral bodies. From this, it can be seen that even if

the dataset contains vertebral bodies with significant morphological differences, the model still achieves satisfactory performance. Perhaps this model is also worth further research and testing optimization in the measurement of screw parameters in the thoracic and cervical vertebrae.

The current investigation encountered several constraints. Initially, to ensure the accuracy and completeness of landmark annotation, a substantial number of patients with fractures, bone defects, and other conditions were excluded from the study cohort. Second, the lack of a definitive reference standard and the inherent variability in manual annotations posed significant challenges in directly comparing the performance of the measurement model with that of human annotators. Thirdly, the implementation of diverse neural networks [9,20,36] that leverage sophisticated feature extraction techniques [37] and optimization algorithms [38,39], which we can couple with computational analysis techniques [40], can significantly enhance the performance of automated frameworks for measuring lumbar pedicle screw parameters. Lastly, our study dataset was derived from cases necessitating screw insertion surgery, potentially influencing the overall robustness of the final system. Thus, in forthcoming studies, we intend to integrate multi-center, multi-disease patient datasets into model training protocols progressively.

5. Conclusions

A novel deep learning model was successfully constructed to automate the measurement of parameters associated with lumbar pedicle screw placement using preoperative CT scans, demonstrating performance on par with that of radiologists and spinal surgeons. This advancement holds promise in delivering efficient, precise, and comprehensive screw-related parameters necessary for diverse lumbar interbody fusion procedures performed by spinal surgeons. Moreover, the versatility of our model extends beyond lumbar regions, as it is anticipated to streamline the automatic measurement of screw placement in cervical and thoracic regions following transfer learning techniques.

CRedit authorship contribution statement

Qian Zhang: Writing – original draft, Methodology, Data curation. **Fanfan Zhao:** Validation, Methodology. **Yu Zhang:** Writing – original draft, Methodology, Data curation. **Man Huang:** Validation, Software. **Xiangyang Gong:** Writing – review & editing, Investigation. **Xuefei**

Deng: Writing – review & editing.

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.

Acknowledgment

This work was supported by the Natural Science Research Project of Anhui Colleges and Universities (2022AH050691).

References

- [1] C.A. Miller, C.G. Ledonio, M.A. Hunt, F. Siddiq, D.W. Polly Jr., Reliability of the planned pedicle screw trajectory versus the actual pedicle screw trajectory using intra-operative 3D CT and image guidance, *Int. J. Spine Surg.* 10 (2016) 38.
- [2] L. La Barbera, C. Ottardi, T. Villa, Comparative analysis of international standards for the fatigue testing of posterior spinal fixation systems: the importance of preload in ISO 12189, *Spine J.* 15 (10) (2015) 2290–2296.
- [3] O.P. Gautschi, B. Schatlo, K. Schaller, E. Tessitore, Clinically relevant complications related to pedicle screw placement in thoracolumbar surgery and their management: a literature review of 35,630 pedicle screws, *Neurosurg. Focus* 31 (4) (2011) E8.
- [4] J.E. Lonstein, F. Denis, J.H. Perra, M.R. Pinto, M.D. Smith, R.B. Winter, Complications associated with pedicle screws, *J. Bone Joint Surg. Am.* 81 (11) (1999) 1519–1528.
- [5] J. Tang, Z. Zhu, T. Sui, D. Kong, X. Cao, Position and complications of pedicle screw insertion with or without image-navigation techniques in the thoracolumbar spine: a meta-analysis of comparative studies, *J. Biomed. Res.* 28 (3) (2014) 228–239.
- [6] I.D. Gelalis, N.K. Paschos, E.E. Pakos, A.N. Politis, C.M. Arnaoutoglou, A. C. Karageorgos, A. Ploumis, T.A. Xenakis, Accuracy of pedicle screw placement: a systematic review of prospective in vivo studies comparing free hand, fluoroscopy guidance and navigation techniques, *Eur. Spine J.* 21 (2) (2012) 247–255.
- [7] V. Kosmopoulos, C. Schizas, Pedicle screw placement accuracy: a meta-analysis, *Spine* 32 (3) (2007) E111–E120.
- [8] J. Chen, C. Chen, M. Nour, D. Liu, Y. Zhu, W. Zhang, K. Polat, X. Deng, Morphology properties of scapular spine relative to reverse shoulder arthroplasty: a biomechanical study, 85 (2023) 104827.
- [9] P. Hamet, J. Tremblay, Artificial intelligence in medicine, *Metabolism* 69s (2017) S36–S40.
- [10] K.K.L. Wong, A. Zhang, K. Yang, S. Wu, D.N. Ghista, GCW-UNet segmentation of cardiac magnetic resonance images for evaluation of left atrial enlargement, *Comput. Methods Programs Biomed.* 221 (2022) 106915.
- [11] X. Zhu, Y. Wei, Y. Lu, M. Zhao, K. Yang, S. Wu, H. Zhang, K.K.L. Wong, Comparative analysis of active contour and convolutional neural network in rapid left-ventricle volume quantification using echocardiographic imaging, *Comput. Methods Programs Biomed.* 199 (2021) 105914.
- [12] D. Alukaev, S. Kiselev, T. Mustafaev, A. Ainur, B. Ibragimov, T. Vrtovec, A deep learning framework for vertebral morphometry and Cobb angle measurement with external validation, *Eur. Spine J.* 31 (8) (2022) 2115–2124.
- [13] J.W. Seo, S.H. Lim, J.G. Jeong, Y.J. Kim, K.G. Kim, J.Y. Jeon, A deep learning algorithm for automated measurement of vertebral body compression from X-ray images, *Sci. Rep.* 11 (1) (2021) 13732.
- [14] G.W. Lee, H. Shin, M.C. Chang, Deep learning algorithm to evaluate cervical spondylotic myelopathy using lateral cervical spine radiograph, *BMC Neurol.* 22 (1) (2022) 147.
- [15] S. Nakdhamabhorn, B.M. Pillai, A. Chotivichit, J. Suthakorn, Sensorless based haptic feedback integration in robot-assisted pedicle screw insertion for lumbar spine surgery: a preliminary cadaveric study, *Comput. Struct. Biotechnol. J.* 24 (2024) 420–433.
- [16] G. Kato, S. Baba, K. Kawaguchi, T. Watanabe, T. Mae, S. Tomari, Perpendicular probing and screwing technique: a simple method for accurate pedicle screw placement based on the human internal reference frame for angle estimation, *PLoS One* 17 (11) (2022) e0277229.
- [17] S.P. Mohanty, M. Pai Kanhangad, S.N. Bhat, S. Chawla, Morphometry of the lower thoracic and lumbar pedicles and its relevance in pedicle fixation, *Musculoskelet. Surg.* 102 (3) (2018) 299–305.
- [18] G. Wai, W. Rusli, S. Ghouse, D.C. Kieser, A. Kedgley, N. Newell, Statistical shape modelling of the thoracic spine for the development of pedicle screw insertion guides, *Biomech. Model. Mechanobiol.* 22 (1) (2023) 123–132.
- [19] D.B. Larson, M.C. Chen, M.P. Lungren, S.S. Halabi, N.V. Stence, C.P. Langlotz, Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs, *Radiology* 287 (1) (2018) 313–322.
- [20] K.K.L. Wong, *Cybernetical Intelligence: Engineering Cybernetics with Machine Intelligence*, Wiley-IEEE Press, Hoboken, New Jersey, 2023.
- [21] C. Payer, D. Stern, H. Bischof, M. Urschler, Integrating spatial configuration into heatmap regression based CNNs for landmark localization, *Med. Image Anal.* 54 (2019) 207–219.
- [22] Z. Han, S. Yu, S.B. Lin, D.X. Zhou, Depth selection for deep ReLU nets in feature extraction and generalization, *IEEE Trans. Pattern Anal. Mach. Intell.* 44 (4) (2022) 1853–1868.
- [23] N. Tajbakhsh, J.Y. Shin, S.R. Gurudu, R.T. Hurst, C.B. Kendall, M.B. Gotway, L. Jianming, Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* 35 (5) (2016) 1299–1312.
- [24] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, ImageNet large scale visual recognition, *Challenge* 115 (3) (2015) 211–252.
- [25] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer, *Learning* 35 (5) (2016) 1285–1298.
- [26] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, (2015).
- [27] J. Lee, D. Koh, C.N. Ong, Statistical evaluation of agreement between two methods for measuring a quantitative variable, *Comput. Biol. Med.* 19 (1) (1989) 61–70.
- [28] X. Deng, Y. Zhu, S. Wang, Y. Zhang, H. Han, D. Zheng, Z. Ding, K.K. Wong, CT and MRI determination of intermuscular space within lumbar paraspinal muscles at different intervertebral disc levels, *PLoS One* 10 (10) (2015) e0140315.
- [29] C.T. Lee, T. Kabir, J. Nelson, S. Sheng, H.W. Meng, T.E. Van Dyke, M.F. Walji, X. Jiang, S. Shams, Use of the deep learning approach to measure alveolar bone level, *J. Clin. Periodontol.* 49 (3) (2022) 260–269.
- [30] H.C. Chen, C.J. Lin, C.H. Wu, C.K. Wang, Y.N. Sun, Automatic Insall-Salvati ratio measurement on lateral knee x-ray images using model-guided landmark localization, *Phys. Med. Biol.* 55 (22) (2010) 6785–6800.
- [31] W. Feng, W. Wang, S. Chen, K. Wu, H. Wang, O-arm navigation versus C-arm guidance for pedicle screw placement in spine surgery: a systematic review and meta-analysis, *Int. Orthop.* 44 (5) (2020) 919–926.
- [32] B. Qu, J. Cao, C. Qian, J. Wu, J. Lin, L. Wang, L. Ou-Yang, Y. Chen, L. Yan, Q. Hong, G. Zheng, X. Qu, Current development and prospects of deep learning in spine image analysis: a literature review, *Quant. Imaging Med. Surg.* 12 (6) (2022) 3454–3479.
- [33] Y. Ishikawa, T. Kokabu, K. Yamada, Y. Abe, H. Tachi, H. Suzuki, T. Ohnishi, T. Endo, D. Ukeba, K. Ura, M. Takahata, N. Iwasaki, H. Sudo, Prediction of Cobb angle using deep learning algorithm with three-dimensional depth sensor considering the influence of garment in idiopathic scoliosis, *J. Clin. Med.* 12 (2) (2023).
- [34] Y. Zhao, J. Zhang, H. Li, X. Gu, Z. Li, S. Zhang, Automatic Cobb angle measurement method based on vertebra segmentation by deep learning, *Med. Biol. Eng. Comput.* 60 (8) (2022) 2257–2269.
- [35] R.R. Naik, A. Hoblidar, S.N. Bhat, N. Ampar, R. Kundangar, A hybrid 3D–2D image registration framework for pedicle screw trajectory registration between intraoperative X-ray image and preoperative CT image, *J. Imaging* 8 (7) (2022).
- [36] Z. Tang, M. Sui, X. Wang, W. Xue, Y. Yang, Z. Wang, T. Ouyang, Theory-guided deep neural network for boiler 3-D NOx concentration distribution prediction, *Energy* 299 (2024) 131500.
- [37] Z. Tang, S. Wang, Y. Li, Dynamic NOx emission concentration prediction based on the combined feature selection algorithm and deep neural network, *Energy* 292 (2024) 130608.
- [38] M. Liu, J. Lv, S. Du, Y. Deng, X. Shen, Y. Zhou, Multi-resource constrained flexible job shop scheduling problem with fixture-pallet combinatorial optimisation, *Comput. Ind. Eng.* 188 (2024) 109903.
- [39] Y. Zhou, S. Du, M. Liu, X. Shen, Machine-fixture-pallet resources constrained flexible job shop scheduling considering loading and unloading times under pallet automation system, *J. Manuf. Syst.* 73 (2024) 143–158.
- [40] S.C.P. Cheung, K.K.L. Wong, W. Yang, G.H. Yeoh, J.Y. Tu, R. Beare, T. Phan, Experimental and numerical study on the hemodynamics of stenosed carotid bifurcation, *Australasian Physical & Engineering Sciences in Medicine* 33 (4) (2010) 319–328.