

Article

Not peer-reviewed version

Development of End-to-End Artificial Intelligence Models for Surgical Planning in Transforaminal Lumbar Interbody Fusion

[Anh Tuan Bui](#) , [Hieu Le](#) , Tung Thanh Hoang , [Giam Minh Trinh](#) , Hao-Chiang Shao , [Pei-I Tsai](#) , [Kuan-Jen Chen](#) , Kevin Li-Chun Hsieh , [E-Wen Huang](#) , [Ching-Chi Hsu](#) , [Mathew Mathew](#) , [Ching-Yu Lee](#) , Po-Yao Wang , [Tsung-Jen Huang](#) , [Meng-Huang Wu](#) *

Posted Date: 22 December 2023

doi: 10.20944/preprints202312.1699.v1

Keywords: Spinal fusion; Interbody cage; Sagittal balance; Artificial intelligence; Machine learning; Spinal parameters



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Development of End-to-End Artificial Intelligence Models for Surgical Planning in Transforaminal Lumbar Interbody Fusion

Anh Tuan Bui ^{1,2}, Hieu Le ³, Tung Thanh Hoang ², Giam Minh Trinh ^{4,5}, Hao-Chiang Shao ⁶,
Pei-I Tsai ⁷, Kuan-Jen Chen ⁷, Kevin Li-Chun Hsieh ^{8,9,10}, E-Wen Huang ¹¹, Ching-Chi Hsu ¹²,
Mathew Mathew ¹³, Ching-Yu Lee ^{14,15}, Po-Yao Wang ^{14,15}, Tsung-Jen Huang ^{14,15}
and Meng-Huang Wu ^{14,15,16,*}.

¹ International Ph.D. Program in Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

² Department of Spine Surgery, Military Hospital 103, Vietnam Military Medical University, Hanoi, Vietnam

³ School of Computer and Communication Sciences, Swiss Federal Institute of Technology in Lausanne, Lausanne, Switzerland

⁴ Department of Trauma-Orthopedics, College of Medicine, Pham Ngoc Thach Medical University, Ho Chi Minh City, Vietnam

⁵ Department of Pediatric Orthopedics, Hospital for Traumatology and Orthopedics, Ho Chi Minh City, Vietnam

⁶ Institute of Data Science and Information Computing, National Chung Hsing University, Taichung City, Taiwan

⁷ Biomedical Technology and Device Research Laboratories, Industrial Technology Research Institute, Hsinchu, Taiwan

⁸ Department of Radiology, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

⁹ Department of Medical Imaging, Taipei Medical University Hospital, Taipei, Taiwan

¹⁰ Research Center of Translational Imaging, Taipei Medical University Hospital, Taipei, Taiwan

¹¹ Department of Materials Science and Engineering, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

¹² Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

¹³ Department of Biomedical Engineering, Colleges of Engineering and Medicine, University of Illinois Chicago, Chicago, USA

¹⁴ Department of Orthopedics, Taipei Medical University Hospital, Taipei, Taiwan

¹⁵ Department of Orthopedics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

¹⁶ TMU Biodesign Center, Taipei Medical University, Taipei, Taiwan

* Correspondence: maxwutmu@gmail.com

Abstract: Transforaminal lumbar interbody fusion (TLIF) is a commonly used technique for treating lumbar degenerative diseases. In this study, we developed a fully computer-supported pipeline to predict both the cage height and the degree of lumbar lordosis subtraction from the pelvic incidence (PI-LL) after TLIF surgery, utilizing preoperative X-ray images. The automated pipeline comprised two primary stages. First, a deep learning model was employed to extract essential features from X-ray images. Subsequently, five machine learning algorithms were trained to identify the optimal models to predict interbody cage height and postoperative PI-LL. LASSO regression and support vector regression demonstrated superior performance in predicting interbody cage height and postoperative PI-LL, respectively. For cage height prediction, the root mean square error (RMSE) was calculated as 1.01, and the model achieved the highest accuracy at a height of 12 mm, with exact prediction achieved in 54.43% (43/79) of cases. In most of the remaining cases, the prediction error of the model was within 1 mm. Additionally, the model demonstrated satisfactory performance in predicting PI-LL, with an RMSE of 5.19 and an accuracy of 0.81 for PI-LL stratification. In conclusion, our results indicate that machine learning models can reliably predict interbody cage height and postoperative PI-LL.

Keywords: spinal fusion; interbody cage; sagittal balance; artificial intelligence; machine learning; spinal parameters

1. Introduction

Over the past few decades, transforaminal lumbar interbody fusion (TLIF) has been commonly used to treat lumbar degenerative diseases, demonstrating the benefits of achieving satisfactory arthrodesis through a unilateral approach with minimal impingement on neural components [1,2]. In addition to relieving spinal nerve compression, the primary objective of TLIF is to restore sagittal balance and the intervertebral body height [3–5].

In terms of sagittal alignment, several studies have reported a close relationship between postoperative sagittal malalignment and postoperative residual symptoms in patients with lumbar fusion [5,6]. Among the parameters of spinal alignment, subtraction of lumbar lordosis (LL) from the pelvic incidence (PI) is a crucial indicator of postoperative outcomes after short-segment lumbar interbody fusion for lumbar pathologies. Patients with PI-LL (PI minus LL) mismatch have increased risks of adjacent segment disease (ASD), late surgical complications, and revision surgery [7–9]. Therefore, postoperative alignment prognosis, especially for critical parameters such as PI-LL, is required for optimal preoperative planning for lumbar fusion. However, predicting postoperative alignment in patients is challenging. Ailon et al. [10] reported that only 42% of cases were accurately predicted by 17 experienced surgeons specializing in treating spinal deformity. Although various methods exist for predicting postoperative parameters in patients with adult spinal deformity [11,12], a method for predicting the value of PI-LL in TLIF procedures still needs to be developed.

Selecting an interbody cage with the correct height is a crucial aspect of lumbar interbody fusion. Utilizing an undersized cage may result in the inability to restore the intervertebral height and segmental lordosis, as well as in complications such as pseudarthrosis and cage migration [13–15]. By contrast, utilizing an oversized cage may increase the likelihood of nerve root compression, ASD, or cage subsidence [15]. In clinical practice, the cage height has long been selected subjectively by surgeons depending on their operational experience. Few studies have predicted the height of fusion cages on the basis of the intervertebral height of the pathological segment [16] or the anterior and posterior disc height on a preoperative computed tomography (CT) image [17]. However, in severe degenerative diseases, such as spondylolisthesis and spinal deformity, when the disc height is greatly reduced, these methods are often inaccurate. Thus, estimating the height of interbody cages remains a challenge.

The choice of the cage height affects sagittal balance (and vice versa), and preoperative spinal parameters play a key role in determining the appropriate size of the implanted device for achieving favorable parameters after surgery [16,18]. Therefore, it is imperative to develop regression models for predicting interbody cage height and postoperative parameters based on preoperative data. However, manual measurements are time-consuming for obtaining all parameters and are prone to rater-dependent errors. Presently, automated tools involving artificial intelligence (AI) are employed to enhance the accuracy and efficiency of measuring spinal alignment parameters from radiographic images [12,19]. Despite these advancements, there is a notable gap in the literature as, to the best of our knowledge, the integration of AI-derived parameters into regression models for surgical planning remains underdeveloped. This study aims to develop a dedicated pipeline utilizing AI and machine learning (ML) to reliably predict interbody cage height and postoperative PI-LL in TLIF surgery based on preoperative X-ray images.

2. Materials and Methods

2.1. Patient selection

A total of 311 patients who underwent L4-L5 TLIF surgery between January 2019 and December 2021 at our institution were included in this retrospective study. The following patients were

included: (1) patients with lumbar degenerative diseases, such as lumbar disc herniation, lumbar spinal stenosis, and spondylolisthesis; (2) patients who underwent TLIF surgery to implant a single interbody cage; and (3) patients who did not experience any complications, such as cage migration, pseudarthrosis, or fusion failure, and did not require revision surgery because of cage problems or ASD during the follow-up period (at least 6 months). The following patients were excluded: (1) patients with a history of lumbar fractures or patients who received a diagnosis of one-segment lumbar degenerative disease at other levels, multiple lumbar degenerative diseases, lumbar scoliosis, spinal tumors, or severe osteoporosis; (2) patients who received two interbody cage implants; (3) patients with unstandardized sagittal radiographs with low image quality for segmentation or radiographs lacking a femoral head; and (4) patients who experienced neurological or neuromuscular episodes during the follow-up period.

In addition to preoperative and postoperative X-ray images and the size of the surgically implanted interbody fusion cage, the demographics of each patient were obtained. Standing lateral X-ray images were used because they offer higher quality and standardization than intraoperative X-ray images, thereby minimizing segmentation bias and error in parameter measurements. Imaging data were obtained using a Radnext 50 X-ray machine from Hitachi Global (Tokyo, Japan).

2.2. X-ray segmentation and feature extraction

A pretrained BiLuNet model was employed to segment each input X-ray image into various semantic regions, including L1, L2, L3, L4, and L5 regions, a sacrum region, and two femoral head regions (Figure 1) [20]. After resizing the original image to 512×512 pixels, the model generated an output image with four labels: background, lumbar vertebral regions, sacrum, and two femoral heads. Nearest-neighbor interpolation was then used to resize the segmented image to its original size. Based on the contours of the segmented areas, a computer vision algorithm obtained multiple corner points to measure the spinal parameters on preoperative X-ray images. Subsequently, these features were combined with four demographic features – namely age, gender, body mass index (BMI), and fusion indication – to derive input features for ML algorithms. Finally, the PI-LL value was measured from the postoperative X-ray image by two experienced surgeons (C.-Y.L. and M.-H.W.) and served as a validation standard for ML models.

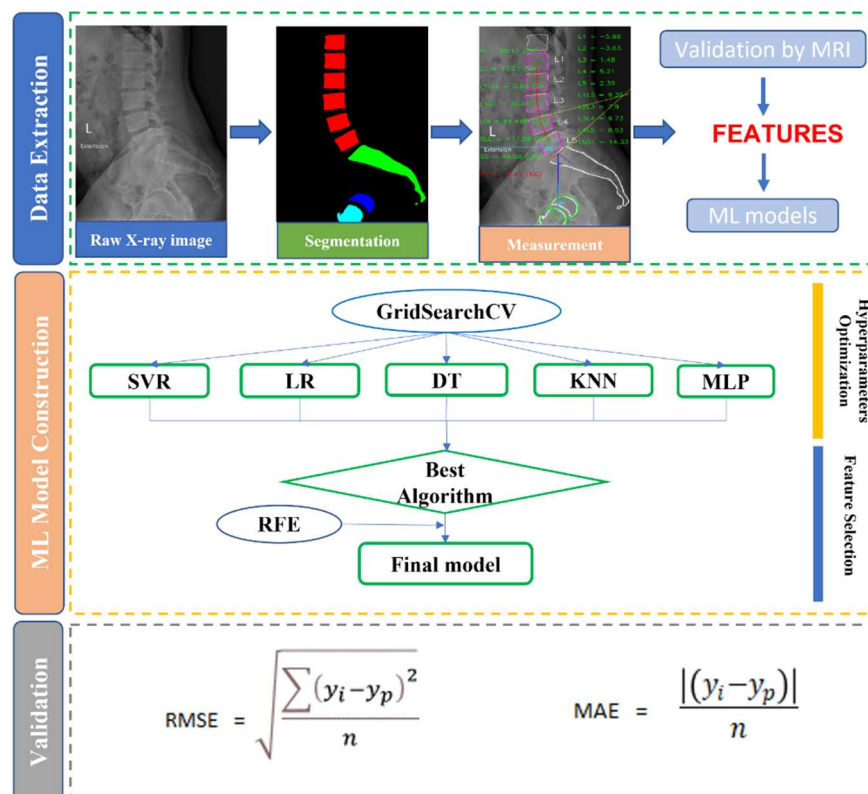


Figure 1. Study flowchart depicting four subprocesses: data cohort collection, feature extraction, feature validation, and ML model construction and validation. ML: machine learning; SVR: support vector regression; LR: LASSO regression; DT: decision tree; KNN: K-nearest neighbor; MLP: multilayer perceptron; RFE: recursive feature elimination; RMSE: root mean square error; MAE: mean absolute error.

To assess the measurement precision of the BiLuNet model, two authors (A.T.B. and G.M.T.) independently measured the aforementioned parameters using magnetic resonance imaging (MRI) and compared their results with those of the model. Since the MRI angle parameters in the supine position differ from those obtained from standing X-ray images, only bone distance features were selected to evaluate interobserver reliability.

2.3. ML implementation

We divided our ML pipeline into three steps: data extraction, model building, and validation (Figure 1). All steps were performed using Python 3.7 and scikit-learn 1.1.2 package [21].

2.3.1. Data preprocessing

Each missing value in the dataset of all aforementioned features was examined and replaced by the mean value of each parameter. Due to distinct units and large differences between feature ranges, the z-score was employed in the data normalization step [22].

2.3.2. Regression models

Various ML models were evaluated to determine their performance for the aforementioned features. These models included five regression algorithms: decision tree (DT), LASSO regression (LR), support vector regression (SVR), K-nearest neighbor (KNN), and multilayer perceptron (MLP). Hyperparameter optimization was conducted for each ML algorithm through the GridSearchCV method to achieve improved results. The algorithm with the highest performance was selected as the baseline model to construct the final ML model. After baseline ML models were obtained for either cage height or postoperative PI-LL prediction, feature selection with Recursive Feature Elimination (RFE) was used to remove the least crucial features and rebuild models with the remaining features.

To determine the optimal number of features, an RFE loop was performed with cross-validation (RFECV function). The mean absolute error (MAE) of the model was then calculated across all repetitions and folds of the RFECV function. Generally, the scikit-learn library represents the MAE as a negative value to maximize it. Therefore, a model with a large negative MAE value is regarded as superior for RFE visualization. After the RFE process, the final model was built using the optimal subset of features, with the SHapley Additive exPlanations (SHAP) value indicating the importance of each feature in model prediction [23].

2.4. Statistical analysis and measurement metrics

A five-fold cross-validation ($k=5$) was performed to assess the efficacy of the ML regression algorithms. The model was then trained on $k - 1$ data splits, and the trained model was tested on the remaining held-out split. Subsequently, the performance of each model was averaged across all data splits for comparison. This cross-validation scheme provided a more reliable test result than that derived using a single fixed testing data split, especially when training data were limited. It also guaranteed that each data point was tested exactly once.

To compare the performance of all ML algorithms, both the root mean square error (RMSE) and the MAE of each model were calculated. The testing error in each case was then visualized to evaluate the accuracy of prediction. To examine the reliability of features in the deep learning model, the intraclass correlation coefficient (ICC) was calculated using SPSS version 18.0 (SPSS, Chicago, IL, USA). The 95% confidence interval of the ICC estimate suggests poor reliability for values below 0.5, moderate reliability for values between 0.5 and 0.75, adequate reliability for values between 0.75 and

0.9, and excellent reliability for values greater than 0.9 [24]. Schwab classification was then performed with three levels of PI-LL, and the final model was evaluated in terms of its ability to stratify postoperative PI-LL based on the accuracy index and F1-score. Generally, a PI-LL value below 10° yields a modifier of 0, a value between 10° and 20° yields a modifier of 1, and a value greater than 20° yields a modifier of 2 [25].

3. Results

3.1. Patient characteristics

This study included 126 men and 185 women, with a mean age of 64.08 years (standard deviation: 11.19) and a mean BMI of 25.4 kg/m² (standard deviation: 3.62). In total, 88 patients had lumbar disc herniation, 154 patients had lumbar spinal stenosis, and 69 patients had lumbar spondylolisthesis. Figure 2 depicts the ground truth distribution of two predictable parameters. Most of the cases (149/311 cases) had cage heights of 12–13 mm, with only few cases having fusion cage heights of 8, 9, and 15 mm. Similar uneven distribution was observed in PI-LL values after surgery, with the majority of patients having PI-LL values ranging from 0 to 20. These unbalanced proportions posed a challenge for the optimization of the ML algorithms.

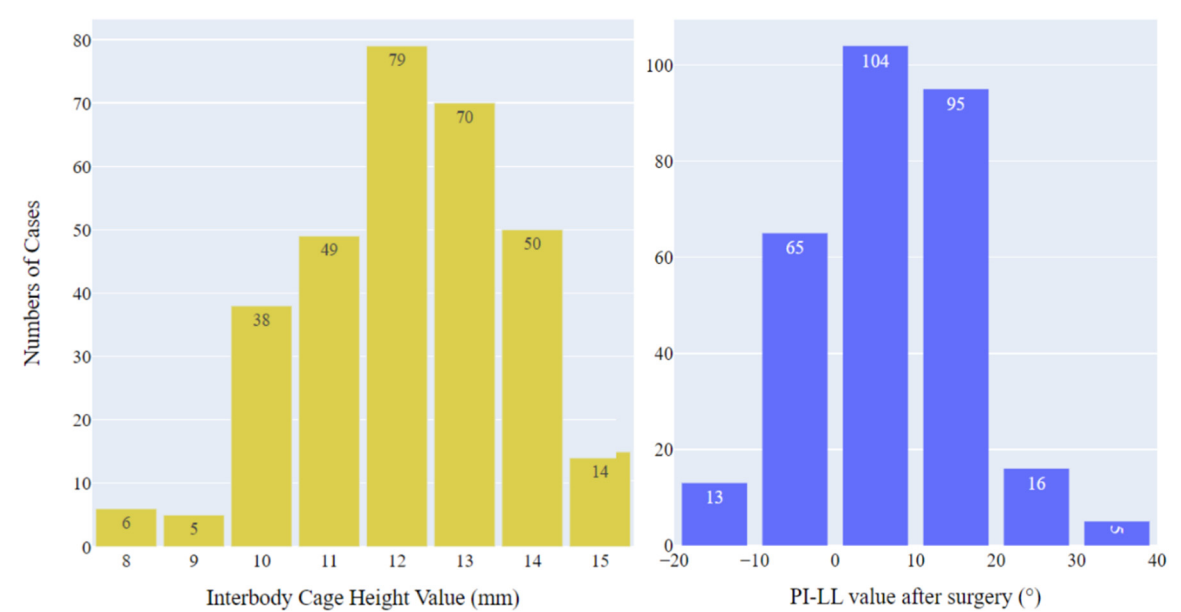


Figure 2. Distribution of actual interbody cage heights and postoperative PI-LL values.

3.2. Performance of ML algorithms

A total of 53 features were extracted from preoperative X-ray images using a deep learning model (Supplementary Table S1). These features demonstrated highly reliability, as evidenced by interobserver reliability within an ICC range of 0.78–0.947 (Supplementary Table S2). These results affirm the robust performance of the deep learning model in accurately measuring spinal parameters. Following the inclusion of four clinical features, a total of 57 features were input into the regression models.

Subsequent experiments were conducted to determine the optimal parameters of each ML algorithm in predicting both cage height and postoperative PI-LL. Table 1 enumerates the ranges of all scrutinized hyperparameters and their corresponding optimal values. Upon comparison of the five algorithms with optimal parameters, LR exhibited superior performance in predicting the cage height, with an RMSE of 1.06 and an MAE of 0.76. Notably, SVR emerged as the optimal model for predicting postoperative PI-LL, displaying the lowest RMSE (5.4) and MAE (4.15) among the algorithms considered, followed by LR, MLP, KNN, and DT (Table 2). Consequently, LR was selected as the baseline model for predicting the cage height, while SVR was selected for predicting PI-LL.

Table 1. Hyperparameter optimization for ML algorithms for the prediction of interbody cage height and postoperative PI-LL.

ML algorithm	Hyperparameter ranges	Optimal values for cage height prediction	Optimal values for PI-LL prediction
LR	Alpha = [0, 1], interval = 0.001	0.001	0.01
DT	Criterion = [squared_error, friedman_mse, absolute_error, poisson]	poisson	squared_error
	min_samples_split = [10, 20, 30, 40, 50]	30	50
	min_samples_leaf = [5, 10, 20, 30, 40]	20	5
SVR	kernels = [poly, linear, rbf, sigmoid]	sigmoid	linear
	C = [0.1, 1, 10, 100]	10	0.1
	gamma = [0.001, 0.01, 0.1, 1]	0.001	1
MLP	hidden_layer_sizes = [(50, 50, 50), (100, 100, 100), (200, 200, 200)]	(200, 200, 200)	(200, 200, 200)
	activation = [tanh, relu]	relu	tanh
	solver = [sgd, adam, lbfgs]	lbfgs	sgd
	alpha = [0.0001, 0.001, 0.05]	0.05	0.0001
KNN	n_neighbors = [5, 10, 20, 30, 40, 50]	20	5
	metric = [euclidean, manhattan, minkowski]	euclidean	euclidean
	weights = [uniform, distance]	uniform	distance

Table 2. Performance of ML algorithms in the prediction of interbody cage height and postoperative PI-LL. RMSE: root mean square error; MAE: mean absolute error.

Algorithm	Cage height		Postoperative PI-LL	
	RMSE	MAE	RMSE	MAE
DT	1.12	0.85	7.05	5.39
LR	1.06	0.76	5.42	4.2
SVR	1.09	0.77	5.4	4.15
MLP	1.16	0.87	6.36	4.84
KNN	1.25	0.498	6.98	5.21

3.3. Final model

3.3.1. Feature selection

Figure 3 depicts the RFECV results for two baseline modes. In the LR model for predicting interbody cage height, the RFE curve identified 23 features as the optimal input for achieving peak performance, with a negative optimum cut-off MAE of -0.693. Likewise, the SVR model for predicting postoperative PI-LL identified 24 features as the optimal number, with a negative cut-off MAE of -4.096. The two subsets of features were subsequently employed to retrain the models (Supplementary Table S3), and the final models underwent validation using the testing set.

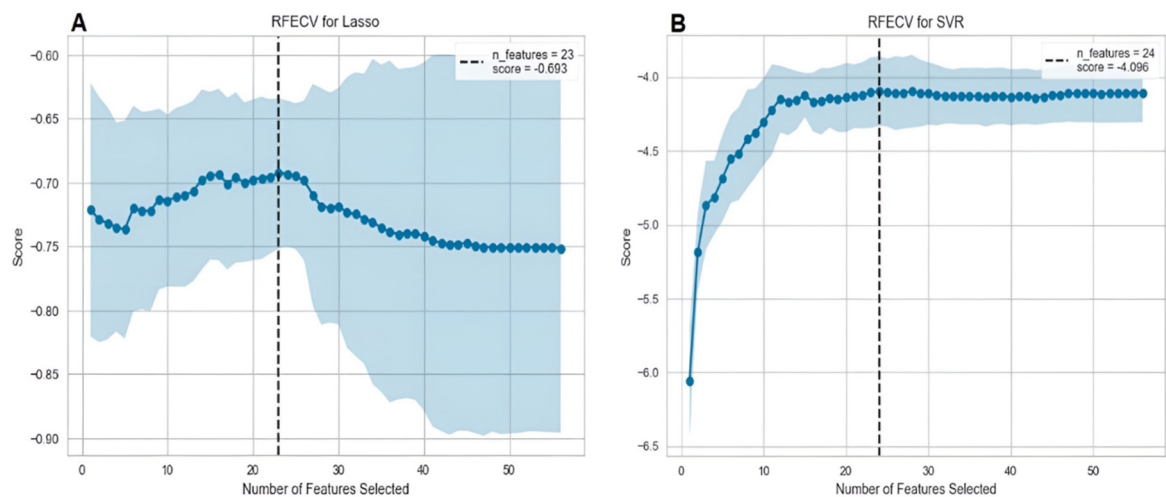


Figure 3. RFECV curves of two baseline models with negative MAEs for different numbers of features: (A) a LR model for interbody cage height prediction and (B) an SVR model for postoperative PI-LL prediction.

3.3.2. Optimal model performance

As shown in Table 3, the finalized LASSO algorithm for cage height prediction demonstrated an RMSE of 1.01 and an MAE of 0.7. These values reflect an enhancement over the metrics obtained prior to feature reduction (i.e., 1.06 and 0.76, respectively). Figure 4 depicts the accuracy of cage height prediction using the testing set, with 42.12% (131/311) of cases achieving exact values. Our model demonstrated commendable accuracy for interbody cage heights with the range of 10–13 mm. Notably, the most accurate prediction was obtained for a height of 12 mm, with 54.43% (43/79) of cases accurately predicted. Simultaneously, the accuracy ratios for sizes 10, 11, and 13 mm were 52.63% (20 of 38 cases), 51.02% (25 of 49 cases), and 42.86% (30 of 70 cases), respectively. In the majority of the remaining cases, the model exhibited a 1 mm prediction error, resulting in an overall accuracy rate of 88.75% (276 out of 311 cases) within the acceptable margin of 1 mm.

Table 3. Performance of two final models. RMSE: root mean square error; MAE: mean absolute error.

	Baseline model performance		Optimal model performance	
	RMSE	MAE	RMSE	MAE
Cage height prediction	1.06	0.76	1.01	0.7
Postoperative PI-LL prediction	5.5	4.15	5.19	3.86

Due to the limited sample sizes in the 8, 9, and 15 mm fusion cage groups, the model encountered elevated prediction errors. Specifically, four of six cases with an actual cage height of 8 mm were erroneously predicted to have a height of 9 mm. Within the 9 mm group, predicted values were 8 mm in three cases and 9 mm in two cases. Notably, for the 15 mm group, the model tended to predict interbody cage heights within the range of 13 to 14 mm in 10 out of 14 cases.

Moving on to postoperative PI-LL prediction, the final SVR model achieved lower RMSE and MAE values on the testing set compared to the baseline model (5.19 and 3.86 versus 5.4 and 4.15; Table 3). In Figure 5A, the model’s performance on both the training and testing data is depicted, indicating a well-calibrated model where most points cluster around the regression line. This observation suggests close alignment between predicted PI-LL values and actual values. However, in cases with PI-LL values exceeding 20, more considerable errors were observed. Furthermore, the model exhibited high precision in stratifying postoperative PI-LL, achieving an accuracy of 0.81 and a high F1-score for the 0 group (Figure 5B).

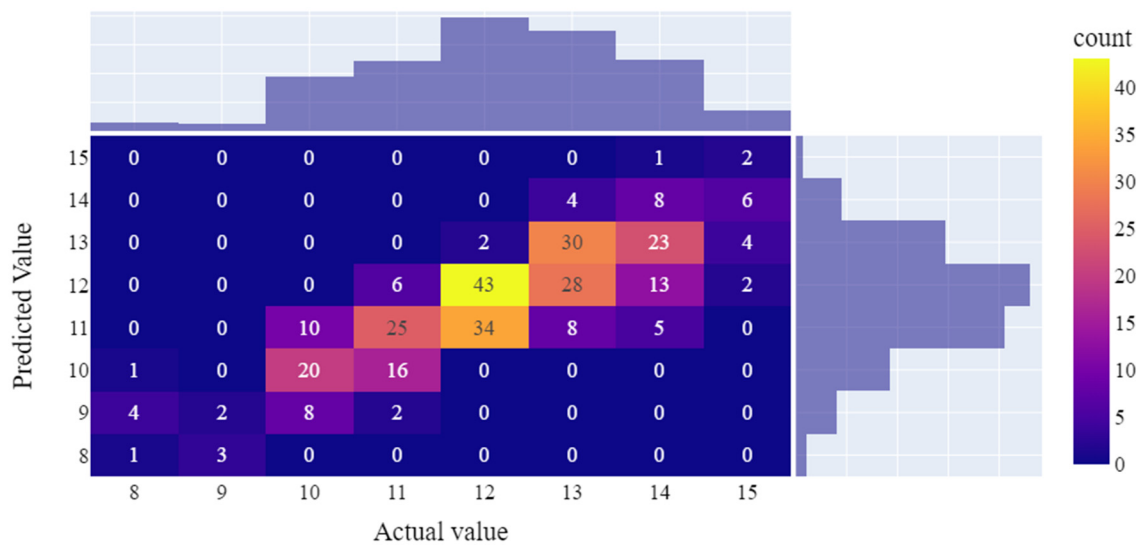


Figure 4. Confusion matrix for final model performance in the prediction of interbody cage height.

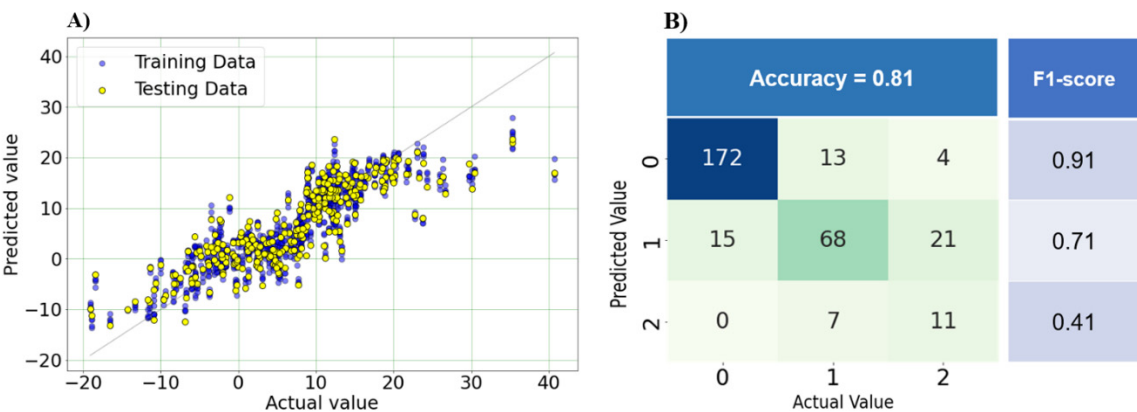


Figure 5. Performance of SVR. (A) Calibration plot (actual and predicted values) for predicting postoperative PI-LL on both training and testing data. (B) Confusion matrix for stratifying postoperative PI-LL on the testing set into three groups: 0 (<10), 1 (10–20), and 2 (>20).

3.3.3. Feature importance.

Figure 6 visualizes the ten most influential features in the two final models. In predicting interbody cage height, the intervertebral height at the midpoint of L4-L5 (L4L5_mid) emerged as the most crucial factor. This prediction was notably influenced by three angles: LL, PI, and the L4-L5 intervertebral disc angle (L4L5_angle). Additionally, crucial parameters included the intervertebral heights of lumbar segments from L3 to S1, encompassing the intervertebral height at the midpoint of L3L4 and L5S1 (L3L4_mid and L5S1_mid), the posterior intervertebral height of L3L4 (L3L4_post), and the anterior intervertebral height of L3L4 and L4L5 (L3L4_ant and L4L5_ant). Among the factors related to vertebral body size, only the upper vertebral width of L3 (L3Width_up), was included in this influential list.

Preoperative LL, relative LL (RLL), and PI played crucial roles in predicting postoperative PI-LL. Essential features associated with PI-LL after surgery predominantly involved angles related to preoperative sagittal alignment, such as sacrum slope (SS), pelvic tilt (PT), and L5S1 intervertebral disc angle (L5S1_angle). Additionally, factors influencing PI-LL prediction were linked to the height of the vertebral body, including the anterior height of the L5 vertebra and the posterior height of the L2 and L3 vertebrae (L2Height_Post and L3Height_Post).

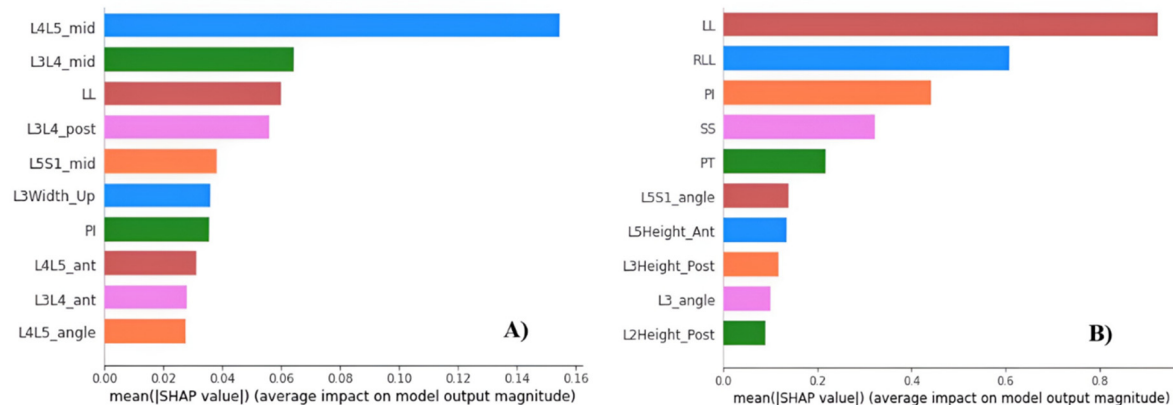


Figure 6. (A) Most crucial features for the model of interbody cage height prediction. (B) Most crucial features for the model of postoperative PI-LL prediction. *Note: The explanations of feature abbreviations are provided in Supplementary Table S1.

4. Discussion

Spinopelvic alignment restoration is essential for both adult spinal deformity surgery and short-segment lumbar interbody fusion [8,26,27]. However, determining the influence of each factor on sagittal alignment is difficult because the normal standing posture is jointly determined by multiple lumbosacral factors [28,29]. As shown in Figure 6 in the present study, the postoperative value of PI-LL is substantially influenced by the preoperative values of LL, RLL, and PI. However, because the PI value is regarded as *a constant anatomic feature with slight variation in pathologic disorders or lumbar spine interventions* [30], *determining the postoperative LL is typically necessary for predicting the optimal PI-LL*. According to previous research, LL restoration after surgery is closely linked to preoperative LL and PI [31–34]. Therefore, LL and PI can be used to predict the LL and PI-LL values after surgery, as in our model.

Appropriate parameters must be obtained for enhancing surgical quality, and surgeons must develop effective strategies to achieve harmonious sagittal alignment. Our model demonstrated a strong capacity to generate a satisfactory PI-LL value while being able to forecast the potential range of this value. By selecting patients without ASD for the dataset, the algorithm trained on these data was able to generate a favorable PI-LL value, which can be used to reduce the incidence of ASD in patients [7]. Our PI-LL prediction model was also able to provide predictions for surgical planning in selecting the appropriate surgical technique and instruments. Actually, the optimal PI-LL has been the subject of debate. Satoshi et al. [35] reported that this value is inconsistent. Meanwhile, multiple studies have suggested that surgeons must strive to reduce PI-LL to 10° or less whenever possible [8,36,37]. According to our model, if unsatisfactory PI-LL prediction values are obtained before surgery, surgeons could consider implementing additional intraoperative techniques. To achieve an adequate LL value, strong fixation with a curved rod system can be implemented. In some cases of severe hypolordosis, osteotomy techniques such as pedicle subtraction osteotomy are also a viable option [38]. Furthermore, the predictive results of postoperative PI-LL from our algorithm may aid in rod bending or in the determination of the number of spinal levels requiring fixation when a surgeon receives intraoperative fluoroscopic images. However, previous studies have revealed substantial discrepancies between standing and prone angle measurements [39,40]. Therefore, these models must be further developed to ensure their seamless integration from preoperative planning to actual surgery.

Size, shape, and position play a crucial role in the insertion of an intervertebral cage. However, findings regarding the importance of the implant shape and placement have been inconsistent. Cage lordosis and final LL after surgery are strongly correlated, with more anterior placement resulting in greater intervertebral lordosis [18]. Conversely, some *in vitro* biomechanical and clinical studies have reported that the cage position and geometry do not affect sagittal alignment after lumbar interbody fusion [41–43]. The cage height typically serves as a key factor applied by surgeons for improving

lordosis [44,45], and our research has primarily focused on predicting this index. Most of our cage height values were between 12 and 13 mm, which are consistent with the recommended cage heights of 11, 12, or 13 mm for the L3-4 and L4-5 levels in a previous study conducted in the Chinese population [16]. In addition, our model performed well for cases within this range, indicating its potential clinical applicability for the Asian population. Overall, predicting the appropriate interbody cage size can assist surgeons in decision-making and improve postoperative outcomes, particularly for inexperienced surgeons. Prediction using our model can also provide the cage height with an error of approximately 1 mm only (Figure 4). Consequently, fewer cages need to be sterilized, thus reducing the costs of surgery. In addition, the costs of treatment decrease due to the reduced operation duration and complication rates. Therefore, patients evidently benefit from the development of these models.

Our results indicated that the disc height of the pathological segment and the two adjacent levels plays a crucial role in predicting the height of the interbody cage (Figure 6). To predict this value, Wang et al. [16] developed a regression model that emphasizes the importance of the intervertebral height at the midpoint of the pathological segment (MIVH): $\text{interbody cage height} = 11.123 - 0.563 * \text{gender} + 0.149 * \text{MIVH}$. In our study, gender was one of the final 23 features used to build the optimal model, but its influence was not as evident as that of the other parameters. With the exception of the parameters associated with the intervertebral disc height, PI and LL contributed to the prediction of the interbody cage height. These two parameters also contributed to the aforementioned prediction of postoperative PI-LL. Lafage et al. [11] discovered that pelvic retroversion and global sagittal balance in adult patients with spinal deformities were primarily influenced by the PI and LL values. Here, we emphasized that PI and LL are among the most crucial parameters for both long- and short-segment fusion surgeries.

Multiple researchers have attempted to develop algorithms for predicting postoperative sagittal parameters and the interbody cage height, aiming to enhance accuracy and applicability in clinical practice. Traditionally, these formulas featured a limited set of variables to simplify computations. Lafage et al. [11,46] developed one of the most accurate formulas for predicting the sagittal vertical axis (SVA). They used only four variables in their formula: PI, LL, thoracic kyphosis, and age. Legaye and Duval-Beaupère [30,47] proposed multilinear regression models for calculating LL by using only basic parameters, such as thoracic kyphosis, SS, PI, PT, and T9 spinopelvic inclination. In contrast to prior approaches, our goal was to incorporate all significant lumbar parameters into algorithm development. Because our prediction models (LR for the interbody cage height and SVR for postoperative PI-LL) and previous models share the same characteristic of utilizing multiple linear algorithms, we took advantage of the current technological advancements to incorporate as many variables as possible. However, certain factors, such as the width and length of the vertebral body, were found to be crucial features in our model, a novel finding absent from existing medical literature. While this discovery might be serendipitous during model training with our dataset, it necessitates further verification in subsequent research. Previously, employing multiple parameters may have been impractical for routine clinical use. However, leveraging computational power, contemporary methods now facilitate improved predictive accuracy, rendering these predictions applicable in clinical scenarios. According to Langella et al. [48], computer-assisted methods are associated with a failure rate below 20% for predicting PI and SVA. To the best of our knowledge, this study is pioneering in presenting a pipeline and diverse models for predicting PI-LL and cage height from preoperative X-ray images through AI.

This study has several limitations. Firstly, it is imperative to acknowledge the retrospective nature and single-center design, which inherently presents constraints due to a modest sample size. As a result, optimal interbody cage height or postoperative PI-LL may be subject to variability influenced by subjective factors such as the surgeon's technique and patient demographics. Despite these constraints, the introduction of multiple algorithms in this study introduced a pioneering concept, setting the groundwork for enhanced predictive accuracy in future multicenter studies. Secondly, this study was limited to patients with monosegmental TLIF at the L4L5 level, and only one sagittal parameter, PI-LL, was predicted. However, using our algorithms, a large number of

postoperative parameters can be predicted not only for single-level fusion surgery but also for surgeries involving multiple levels. Thirdly, sagittal balance is associated with factors such as SVA, T1 spinopelvic inclination, and C7 plumb line, which are evaluated using full-length spine radiographs [37,49,50]. Because we focused only on short-segment fusion, we examined only the lumbar region. Therefore, global sagittal balance factors must be examined for TLIF surgery in the future. Lastly, the complexity of our model, involving multiple steps, increases the probability of errors. To increase predictive accuracy, a synthetic model must be developed, integrating radiographic parameters from X-ray, MRI, and CT scans. This comprehensive approach will contribute to refining and validating the predictive capabilities of our model in diverse clinical scenarios.

5. Conclusion

This study marks a significant stride in the development of an end-to-end AI models tailored for predicting interbody cage height and postoperative PI-LL in TLIF surgery. Our findings underscore the efficacy of sophisticated computer-assisted models in spinal morphometry, showcasing the remarkable accuracy of ML algorithms. These models emerge as valuable tools for surgeons, offering substantial support in both preoperative planning and postoperative assessment. Our results highlights the significance of integrating multiple crucial parameters, particularly preoperative PI and LL, into multilinear regression equations. This innovative approach demonstrates promise in predicting outcomes for spinal fusion surgery, emphasizing the potential for improved precision in patient-specific treatment strategies. However, to ensure model reliability and generalizability, further validation and refinement with larger datasets and multicenter studies are required.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org; Table S1: Spinal parameter features extracted using a deep learning model; Table S2: ICCs validating the reliability of the deep learning model in measuring bone distance parameters compared with the MRI results; Table S3: Two subsets of crucial features for two baseline ML models.

Author Contributions: Conceptualization, A.T.B., T.-J.H. and M.-H.W.; Methodology, A.T.B. and M.-H.W.; Software, H.L., P.-I.T., and K.-J.C.; Validation, H.-C.S., E.-W.H. and C.-C.H.; Formal Analysis, H.L. and K.L.-C.H.; Data Curation, K.L.-C.H., C.-Y.L. and P.-Y.W. ; Writing – Original Draft Preparation, A.T.B., H.L., and G.M.T.; Writing – Review & Editing, T.T.H., H.-C.S., M.M., C.-Y.L., T.-J.H., and M.-H.W.; Visualization, A.T.B. and H.L.; Supervision, T.T.H., H.-C.S., T.-J.H., and M.-H.W.; Project Administration, M.-H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Higher Education Sprout Project of the Ministry of Education of Taiwan.

Institutional Review Board Statement: This study was approved by the Joint Institutional Review Board of Taipei Medical University (N201807084).

Informed Consent Statement: The patient's consent was waived for this retrospective study using a clinical database, in accordance with the IRB's statement and regulations.

Data Availability Statement: Access to dataset shall be provided by the corresponding authors upon reasonable request and in accordance with the policies of the relevant institution.

Acknowledgments: The author(s) express their gratitude to Mr. Po-Yu Hsieh and Mr. Chen-Wei Lai from the Industrial Technology Research Institute, Taiwan, for their invaluable assistance in constructing the AI models utilized in this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mummaneni PV, Dhall SS, Eck JC, et al. Guideline update for the performance of fusion procedures for degenerative disease of the lumbar spine. Part 11: interbody techniques for lumbar fusion. *J Neurosurg Spine*. Jul 2014;21(1):67-74. doi:10.3171/2014.4.Spine14276
2. Noshchenko A, Hoffecker L, Lindley EM, Burger EL, Cain CM, Patel VV. Perioperative and long-term clinical outcomes for bone morphogenetic protein versus iliac crest bone graft for lumbar fusion in

- degenerative disk disease: systematic review with meta-analysis. *J Spinal Disord Tech.* May 2014;27(3):117-35. doi:10.1097/01.bsd.0000446752.34233.ca
3. Xiao Y, Li F, Chen Q. Transforaminal lumbar interbody fusion with one cage and excised local bone. *Arch Orthop Trauma Surg.* May 2010;130(5):591-7. doi:10.1007/s00402-009-0917-6
 4. Ould-Slimane M, Lenoir T, Dauzac C, et al. Influence of transforaminal lumbar interbody fusion procedures on spinal and pelvic parameters of sagittal balance. *Eur Spine J.* Jun 2012;21(6):1200-6. doi:10.1007/s00586-011-2124-8
 5. Watkins RGt, Hanna R, Chang D, Watkins RG, 3rd. Sagittal alignment after lumbar interbody fusion: comparing anterior, lateral, and transforaminal approaches. *J Spinal Disord Tech.* Jul 2014;27(5):253-6. doi:10.1097/BSD.0b013e31828a8447
 6. Yamasaki K, Hoshino M, Omori K, et al. Risk Factors of Adjacent Segment Disease After Transforaminal Inter-Body Fusion for Degenerative Lumbar Disease. *Spine (Phila Pa 1976).* Jan 15 2017;42(2):E86-e92. doi:10.1097/brs.0000000000001728
 7. Rothenfluh DA, Mueller DA, Rothenfluh E, Min K. Pelvic incidence-lumbar lordosis mismatch predisposes to adjacent segment disease after lumbar spinal fusion. *Eur Spine J.* Jun 2015;24(6):1251-8. doi:10.1007/s00586-014-3454-0
 8. Aoki Y, Nakajima A, Takahashi H, et al. Influence of pelvic incidence-lumbar lordosis mismatch on surgical outcomes of short-segment transforaminal lumbar interbody fusion. *BMC Musculoskelet Disord.* Aug 20 2015;16:213. doi:10.1186/s12891-015-0676-1
 9. Senteler M, Weisse B, Snedeker JG, Rothenfluh DA. Pelvic incidence-lumbar lordosis mismatch results in increased segmental joint loads in the unfused and fused lumbar spine. *Eur Spine J.* Jul 2014;23(7):1384-93. doi:10.1007/s00586-013-3132-7
 10. Ailon T, Scheer JK, Lafage V, et al. Adult Spinal Deformity Surgeons Are Unable to Accurately Predict Postoperative Spinal Alignment Using Clinical Judgment Alone. *Spine Deform.* Jul 2016;4(4):323-329. doi:10.1016/j.jspd.2016.02.003
 11. Lafage V, Schwab F, Vira S, Patel A, Ungar B, Farcy JP. Spino-pelvic parameters after surgery can be predicted: a preliminary formula and validation of standing alignment. *Spine (Phila Pa 1976).* Jun 2011;36(13):1037-45. doi:10.1097/BRS.0b013e3181eb9469
 12. Lafage R, Pesenti S, Lafage V, Schwab FJ. Self-learning computers for surgical planning and prediction of postoperative alignment. *Eur Spine J.* Feb 2018;27(Suppl 1):123-128. doi:10.1007/s00586-018-5497-0
 13. Abbushi A, Cabraja M, Thomale UW, Woiciechowsky C, Kroppenstedt SN. The influence of cage positioning and cage type on cage migration and fusion rates in patients with monosegmental posterior lumbar interbody fusion and posterior fixation. *Eur Spine J.* Nov 2009;18(11):1621-8. doi:10.1007/s00586-009-1036-3
 14. Li H, Wang H, Zhu Y, Ding W, Wang Q. Incidence and risk factors of posterior cage migration following decompression and instrumented fusion for degenerative lumbar disorders. *Medicine (Baltimore).* Aug 2017;96(33):e7804. doi:10.1097/md.00000000000007804
 15. Aoki Y, Yamagata M, Nakajima F, et al. Examining risk factors for posterior migration of fusion cages following transforaminal lumbar interbody fusion: a possible limitation of unilateral pedicle screw fixation. *J Neurosurg Spine.* Sep 2010;13(3):381-7. doi:10.3171/2010.3.Spine09590
 16. Wang H, Chen W, Jiang J, Lu F, Ma X, Xia X. Analysis of the correlative factors in the selection of interbody fusion cage height in transforaminal lumbar interbody fusion. *BMC Musculoskelet Disord.* Jan 12 2016;17:9. doi:10.1186/s12891-016-0866-5
 17. Makino T, Honda H, Fujiwara H, Yoshikawa H, Yonenobu K, Kaito T. Low incidence of adjacent segment disease after posterior lumbar interbody fusion with minimum disc distraction: A preliminary report. *Medicine (Baltimore).* Jan 2018;97(2):e9631. doi:10.1097/md.00000000000009631
 18. Landham PR, Don AS, Robertson PA. Do position and size matter? An analysis of cage and placement variables for optimum lordosis in PLIF reconstruction. *Eur Spine J.* Nov 2017;26(11):2843-2850. doi:10.1007/s00586-017-5170-z
 19. Cho BH, Kaji D, Cheung ZB, et al. Automated Measurement of Lumbar Lordosis on Radiographs Using Machine Learning and Computer Vision. *Global Spine J.* Aug 2020;10(5):611-618. doi:10.1177/2192568219868190
 20. Tran V, Lin H-Y, Liu H-W, Jang F-J, Tseng C-H. *BiLuNet: A Multi-path Network for Semantic Segmentation on X-ray Images.* 2021:10034-10041.
 21. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research.* 2011;12:2825-2830.
 22. Shalabi LA, Shaaban Z, Kasasbeh B. Data Mining: A Preprocessing Engine. *Journal of Computer Science.* 2006;2(9)doi:10.3844/jcssp.2006.735.739
 23. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Advances in neural information processing systems.* 2017;30

24. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*. Jun 2016;15(2):155-63. doi:10.1016/j.jcm.2016.02.012
25. Schwab F, Ungar B, Blondel B, et al. Scoliosis Research Society-Schwab adult spinal deformity classification: a validation study. *Spine (Phila Pa 1976)*. May 20 2012;37(12):1077-82. doi:10.1097/BRS.0b013e31823e15e2
26. Kong LD, Zhang YZ, Wang F, Kong FL, Ding WY, Shen Y. Radiographic Restoration of Sagittal Spinopelvic Alignment After Posterior Lumbar Interbody Fusion in Degenerative Spondylolisthesis. *Clin Spine Surg*. Mar 2016;29(2):E87-92. doi:10.1097/bsd.000000000000104
27. Glassman SD, Berven S, Bridwell K, Horton W, Dimar JR. Correlation of radiographic parameters and clinical symptoms in adult scoliosis. *Spine (Phila Pa 1976)*. Mar 15 2005;30(6):682-8. doi:10.1097/01.brs.0000155425.04536.f7
28. Weisz G, Houang M. Classification of the normal variation in the sagittal alignment of the human lumbar spine and pelvis in the standing position. *Spine (Phila Pa 1976)*. Jul 1 2005;30(13):1558-9; author reply 1559. doi:10.1097/01.brs.0000167527.42783.76
29. Lafage V, Schwab F, Skalli W, et al. Standing balance and sagittal plane spinal deformity: analysis of spinopelvic and gravity line parameters. *Spine (Phila Pa 1976)*. Jun 15 2008;33(14):1572-8. doi:10.1097/BRS.0b013e31817886a2
30. Legaye J, Duval-Beaupère G, Hecquet J, Marty C. Pelvic incidence: a fundamental pelvic parameter for three-dimensional regulation of spinal sagittal curves. *Eur Spine J*. 1998;7(2):99-103. doi:10.1007/s005860050038
31. Chou D. Commentary: Retrospective Review of Immediate Restoration of Lordosis in Single-Level Minimally Invasive Transforaminal Lumbar Interbody Fusion: A Comparison of Static and Expandable Interbody Cages. *Operative Neurosurgery*. 2020;18(5):E153-E154. doi:10.1093/ons/opz272
32. McMordie JH, Schmidt KP, Gard AP, Gillis CC. Clinical and Short-Term Radiographic Outcomes of Minimally Invasive Transforaminal Lumbar Interbody Fusion With Expandable Lordotic Devices. *Neurosurgery*. Feb 1 2020;86(2):E147-e155. doi:10.1093/neuros/nyz402
33. Porche K, Dru A, Moor R, Kubilis P, Vaziri S, Hoh DJ. Preoperative Radiographic Prediction Tool for Early Postoperative Segmental and Lumbar Lordosis Alignment After Transforaminal Lumbar Interbody Fusion. *Cureus*. Sep 2021;13(9):e18175. doi:10.7759/cureus.18175
34. Schwab F, Lafage V, Patel A, Farcy JP. Sagittal plane considerations and the pelvis in the adult patient. *Spine (Phila Pa 1976)*. Aug 1 2009;34(17):1828-33. doi:10.1097/BRS.0b013e3181a13c08
35. Inami S, Moridaira H, Takeuchi D, Shiba Y, Nohara Y, Taneichi H. Optimum pelvic incidence minus lumbar lordosis value can be determined by individual pelvic incidence. *Eur Spine J*. Nov 2016;25(11):3638-3643. doi:10.1007/s00586-016-4563-8
36. Schwab F, Patel A, Ungar B, Farcy JP, Lafage V. Adult spinal deformity-postoperative standing imbalance: how much can you tolerate? An overview of key parameters in assessing alignment and planning corrective surgery. *Spine (Phila Pa 1976)*. Dec 1 2010;35(25):2224-31. doi:10.1097/BRS.0b013e3181ee6bd4
37. Schwab FJ, Blondel B, Bess S, et al. Radiographical spinopelvic parameters and disability in the setting of adult spinal deformity: a prospective multicenter analysis. *Spine (Phila Pa 1976)*. Jun 1 2013;38(13):E803-12. doi:10.1097/BRS.0b013e318292b7b9
38. Berjano P, Aebi M. Pedicle subtraction osteotomies (PSO) in the lumbar spine for sagittal deformities. *Eur Spine J*. Jan 2015;24 Suppl 1:S49-57. doi:10.1007/s00586-014-3670-7
39. Brink RC, Colo D, Schlösser TPC, et al. Upright, prone, and supine spinal morphology and alignment in adolescent idiopathic scoliosis. *Scoliosis Spinal Disord*. 2017;12:6. doi:10.1186/s13013-017-0111-5
40. Salem W, Coomans Y, Brismée JM, Klein P, Sobczak S, Dugailly PM. Sagittal Thoracic and Lumbar Spine Profiles in Upright Standing and Lying Prone Positions Among Healthy Subjects: Influence of Various Biometric Features. *Spine (Phila Pa 1976)*. Aug 1 2015;40(15):E900-8. doi:10.1097/brs.0000000000000918
41. Takahashi H, Suguro T, Yokoyama Y, Iida Y, Terashima F, Wada A. Effect of cage geometry on sagittal alignment after posterior lumbar interbody fusion for degenerative disc disease. *J Orthop Surg (Hong Kong)*. Aug 2010;18(2):139-42. doi:10.1177/230949901001800202
42. Kepler CK, Rihn JA, Radcliff KE, et al. Restoration of lordosis and disk height after single-level transforaminal lumbar interbody fusion. *Orthop Surg*. Feb 2012;4(1):15-20. doi:10.1111/j.1757-7861.2011.00165.x
43. Faundez AA, Mehbod AA, Wu C, Wu W, Ploumis A, Transfeldt EE. Position of interbody spacer in transforaminal lumbar interbody fusion: effect on 3-dimensional stability and sagittal lumbar contour. *J Spinal Disord Tech*. May 2008;21(3):175-80. doi:10.1097/BSD.0b013e318074bb7d
44. Gambhir S, Wang T, Pelletier MH, Walsh WR, Ball JR. How Does Cage Lordosis Influence Postoperative Segmental Lordosis in Lumbar Interbody Fusion. *World Neurosurg*. Jun 2019;126:e606-e611. doi:10.1016/j.wneu.2019.02.100
45. Uribe JS, Harris JE, Beckman JM, Turner AW, Mundis GM, Akbarnia BA. Finite element analysis of lordosis restoration with anterior longitudinal ligament release and lateral hyperlordotic cage placement. *Eur Spine J*. Apr 2015;24 Suppl 3:420-6. doi:10.1007/s00586-015-3872-7

46. Smith JS, Bess S, Shaffrey CI, et al. Dynamic changes of the pelvis and spine are key to predicting postoperative sagittal alignment after pedicle subtraction osteotomy: a critical analysis of preoperative planning techniques. *Spine (Phila Pa 1976)*. May 1 2012;37(10):845-53. doi:10.1097/BRS.0b013e31823b0892
47. Legaye J, Duval-Beaupère G. Sagittal plane alignment of the spine and gravity: a radiological and clinical evaluation. *Acta Orthop Belg*. Apr 2005;71(2):213-20.
48. Langella F, Villafañe JH, Damilano M, et al. Predictive Accuracy of Surgimap Surgical Planning for Sagittal Imbalance: A Cohort Study. *Spine (Phila Pa 1976)*. Nov 15 2017;42(22):E1297-e1304. doi:10.1097/brs.0000000000002230
49. Glassman SD, Bridwell K, Dimar JR, Horton W, Berven S, Schwab F. The impact of positive sagittal balance in adult spinal deformity. *Spine (Phila Pa 1976)*. Sep 15 2005;30(18):2024-9. doi:10.1097/01.brs.0000179086.30449.96
50. Lafage V, Schwab F, Patel A, Hawkinson N, Farcy JP. Pelvic tilt and truncal inclination: two key radiographic parameters in the setting of adults with spinal deformity. *Spine (Phila Pa 1976)*. Aug 1 2009;34(17):E599-606. doi:10.1097/BRS.0b013e3181aad219

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.