

Large Language Model–Assisted Estimation of Left Double-Lumen Tube Insertion Depth Using Preoperative Chest Radiography

Cansu Kılınç Berktaş

cansukilinc87@gmail.com

Sağlık Bilimleri Üniversitesi

Hilal Akça

Sağlık Bilimleri Üniversitesi

Furkan Tontu

Sağlık Bilimleri Üniversitesi

Emine Şeyda Teloğlu

Sağlık Bilimleri Üniversitesi

Funda Gümüş Özcan

Sağlık Bilimleri Üniversitesi

Research Article

Keywords: Artificial Intelligence, Large Language Model, Endobronchial Intubation, One-Lung Ventilation, Bronchoscopy, Radiography, Chest

Posted Date: June 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9727861/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective(s):

To evaluate the accuracy of a large language model (LLM)–based approach for predicting the optimal insertion depth of left-sided double-lumen endobronchial tubes (DLTs) using routinely available demographic data alone or combined with posteroanterior (PA) chest radiographs, using fiberoptic bronchoscopy (FOB)–confirmed measurements as the reference standard.

Design:

Retrospective observational validation study.

Setting:

Single-center tertiary academic hospital.

Participants:

Adult patients undergoing thoracic surgery requiring one-lung ventilation with a left-sided DLT.

Interventions:

Predicted insertion depths were generated using a pretrained large language model (ChatGPT, OpenAI) based on standardized prompts incorporating demographic data alone and in combination with PA chest radiograph–derived anatomical information.

Measurements & Main Results:

Predictions were compared with intraoperative FOB-confirmed insertion depths. Agreement was evaluated using Bland–Altman analysis and non-parametric comparisons (Friedman test with Dunn’s post-hoc analysis). A total of 154 patients were included. The median insertion depth was 29.00 cm for physician measurements and for both LLM-based approaches. Friedman testing demonstrated a statistically significant overall difference among the three methods ($\chi^2(2) = 12.14$, $p = 0.0023$). Post-hoc analysis revealed a significant difference between physician measurements and LLM predictions without radiographic input (adjusted $p = 0.0039$), whereas no significant difference was observed between physician measurements and LLM predictions with PA radiograph input (adjusted $p = 0.7634$). Bland–Altman analysis demonstrated minimal systematic bias for both LLM-based approaches, with mean differences of 0.20 cm (without PA) and 0.15 cm (with PA), and 95% limits of agreement within approximately ± 2.8 cm. More than 90% of predictions were within ± 2 cm of physician-confirmed depth.

Conclusions:

LLM-based prediction of DLT insertion depth demonstrated reasonable agreement with physician-confirmed measurements. Incorporation of PA chest radiographs reduced statistical differences between LLM predictions and physician measurements, although overall predictive performance remained comparable. These findings suggest that LLM-assisted estimation may serve as a supportive adjunct for initial DLT positioning, pending prospective validation.

INTRODUCTION

One-lung ventilation (OLV) is a standard technique in thoracic surgery to optimize surgical exposure and maintain adequate gas exchange (1, 2). Double-lumen endobronchial tubes (DLTs) remain the most commonly used devices for achieving lung isolation and are considered the primary method for OLV in adult patients (3). Accurate positioning of the DLT is critical, as malposition may lead to hypoxemia, inadequate lung isolation, increased airway pressures, or airway trauma (1, 2).

Fiberoptic bronchoscopy is regarded as the gold standard for confirming correct DLT placement; however, it requires specific equipment and operator expertise, and it may not always be immediately available in all clinical settings (2). Consequently, anesthesiologists frequently rely on estimated insertion depths before bronchoscopic confirmation.

Several studies have demonstrated a statistically significant correlation between patient height and optimal DLT insertion depth (4, 5). Height-based formulae have therefore been proposed to guide initial placement. Nevertheless, subsequent investigations have shown that although such correlations exist, they may not be sufficiently precise for individual patients, particularly in those with atypical body habitus or below-average height (6, 7). These findings highlight the limitations of single-parameter predictive approaches.

Anatomical variability and device-related differences further complicate optimal depth estimation. Dimensional variations among double-lumen tubes from different manufacturers have been documented and may influence final positioning (8, 9). In response to these limitations, imaging-based strategies have been explored. Computed tomography-guided measurements have demonstrated improved accuracy in predicting optimal DLT insertion depth (10, 11).

In recent years, artificial intelligence (AI) and machine learning techniques have shown substantial promise in medical image analysis and clinical decision support (12–14). Artificial intelligence-based approaches using radiographic parameters have also been applied to predict DLT insertion depth and tube size (15), suggesting that image-derived anatomical features may improve individualized airway management strategies. The present study aimed to evaluate the accuracy of an LLM-based approach in predicting the insertion depth of left-sided DLTs using demographic data, with and without preoperative posteroanterior chest radiographs. Model predictions were compared with physician-confirmed insertion depths to assess agreement under real-world clinical conditions.

METHODS

Study Approval and Study Design

This retrospective single-center study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of the University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital (Approval No: E-96317027-514.10-

301752463; Date: 20.01.2026). The requirement for informed consent was waived by the Clinical Research Ethics Committee due to the retrospective nature of the study. The study was designed as a retrospective observational validation study conducted at a tertiary academic hospital. A total of 154 patients who underwent thoracic surgery requiring one-lung ventilation with double-lumen endobronchial tubes (DLTs) were included.

Institutional anesthesia and surgical records were retrospectively reviewed to evaluate the agreement between LLM-based predictions and fiberoptic bronchoscopy (FOB)–confirmed insertion depths of DLTs. The primary aim of the study was not to influence or alter clinical decision-making, but to assess the accuracy of the LLM in predicting optimal DLT insertion depth under real-world clinical conditions.

Reference Standard

For all included patients, DLT positioning was confirmed intraoperatively using flexible fiberoptic bronchoscopy (FOB) as part of routine anesthesia practice. After optimal positioning was achieved, the insertion depth measured from the lip margin (in centimeters) was documented in the anesthesia record. FOB-confirmed insertion depth was accepted as the reference standard for all analyses.

Participants

Adult patients aged 18 years or older who had undergone thoracic surgery requiring one-lung ventilation with a DLT were eligible for inclusion. Patients were included if a preoperative digital PA chest radiograph was available. Exclusion criteria were age younger than 18 years, missing demographic data or poor-quality radiographs, and lung isolation achieved using devices other than a DLT.

Large Language Model–Based Prediction Strategy

All patient data were anonymized prior to analysis, and no identifiable personal information was included in the dataset. Predicted insertion depths were generated using a publicly available pretrained large language model (ChatGPT, OpenAI).

For each patient, demographic variables including age, sex, height, weight, and DLT size were entered into the model using a standardized prompt structure. In a second approach, descriptive anatomical information derived from the preoperative PA chest radiograph was incorporated into the same standardized prompt format.

The prompt wording and structure were kept identical for all patients to ensure methodological consistency. Default system settings were used, and the model was not fine-tuned, retrained, or otherwise modified for the purposes of this study. For each case, the first generated output was recorded without repeated querying.

All predictions were generated retrospectively and were not available to the anesthesia team during clinical care.

Rationale for Use of PA Chest Radiographs

Although thoracic surgery patients routinely undergo chest computed tomography (CT), PA chest radiographs were selected for LLM-based analysis to enhance clinical feasibility and generalizability. PA chest radiographs are universally available, standardized, and routinely obtained in thoracic surgery patients. In contrast, CT imaging is associated with greater data complexity, variability in acquisition protocols, and limited real-time accessibility for bedside decision support. Therefore, the use of PA radiographs was considered more suitable for a pragmatic, clinically applicable large language model–assisted approach.

Model Interaction Protocol

Model Interaction Protocol

Predictions were generated using a publicly available multimodal large language model (ChatGPT, OpenAI) via its web-based interface. Demographic data were entered using a standardized prompt structure. For radiograph-assisted predictions, preoperative PA chest radiographs were uploaded in JPEG format together with the same prompt.

Cases were entered sequentially within an ongoing model session. No feedback regarding the actual insertion depth was provided to the model at any time. Default system inference settings were used, and no manual adjustments (e.g., temperature modification) were performed by the investigators. For each case, the first generated output was recorded without repeated querying.

Statistical Analysis

Predicted insertion depths generated by the LLM were compared with FOB-confirmed reference measurements. Agreement between physician measurements and LLM predictions was assessed using Bland–Altman analysis, including calculation of mean bias, standard deviation (SD) of the bias, and 95% limits of agreement (LoA). Comparisons among physician measurements, LLM predictions without radiographic input, and LLM predictions with PA chest radiograph were performed using the Friedman test for related samples. When significant, post-hoc pairwise comparisons were conducted using Dunn’s multiple comparisons test with adjusted p-values.

RESULTS

A total of 154 patients were included in the final analysis. A left-sided double-lumen tube was used in all cases. The demographic characteristics of the study population are summarized in Table 1. The median insertion depth was 29.00 cm (IQR: 27.38–30.00) for physician-confirmed measurements, 29.00 cm

(IQR: 27.48–29.30) for LLM predictions without radiographic input, and 29.00 cm (IQR: 27.18–29.50) for LLM predictions incorporating PA chest radiographs.

Table 1
Baseline Characteristics of the Study Population (n = 154)

Variable	Value
Age (years)	53 (28–69)
Height (cm)	168 (160–175)
Weight (kg)	72 (60–83)
Double-lumen tube size, n (%)	
35 Fr	46 (29.9%)
37 Fr	67 (43.5%)
39 Fr	41 (26.6%)

Comparison of the three methods using the Friedman test demonstrated a statistically significant overall difference ($\chi^2(2) = 12.14$, $p = 0.0023$) (Table 2). Post-hoc pairwise comparisons with Dunn’s multiple comparisons test revealed a significant difference between physician measurements and LLM predictions without radiographic input (adjusted $p = 0.0039$). No statistically significant difference was observed between physician measurements and LLM predictions with PA radiograph input (adjusted $p = 0.7634$), nor between LLM predictions with and without radiographic input (adjusted $p = 0.1126$).

Table 2
Comparison of physician measurements and AI predictions

	Physician	AI (without PA chest radiograph)	AI (with PA chest radiograph)
Median (IQR), cm	29.00 (27.38–30.00)	29.00 (27.48–29.30)*	29.00 (27.18–29.50)
* Significantly different from physician measurements based on Dunn’s post-hoc test following Friedman analysis (adjusted $p = 0.0039$); no statistically significant differences were observed between the other group comparisons.			

Bland–Altman analysis comparing physician measurements and LLM predictions without radiographic input demonstrated a mean bias of 0.2019 cm (SD 1.327), with 95% limits of agreement ranging from – 2.398 to 2.802 cm (Fig. 1). For LLM predictions incorporating PA chest radiographs, the mean bias was 0.1506 cm (SD 1.337), with 95% limits of agreement ranging from – 2.469 to 2.771 cm (Fig. 2).

Bland–Altman analysis demonstrated minimal systematic bias for both approaches, with 95% limits of agreement within approximately ± 2.8 cm. In addition, 91.6% of demographic-only predictions and 90.3%

of radiograph-assisted predictions were within ± 2 cm of physician-confirmed depth. The mean absolute error was approximately 1 cm for both approaches.

Overall, both LLM-based approaches demonstrated minimal systematic bias and reasonable agreement with physician-confirmed insertion depths.

DISCUSSION

In this retrospective observational study, we demonstrated that a pretrained large language model generated insertion depth predictions for left-sided double-lumen endobronchial tubes with clinically acceptable agreement when compared with physician-confirmed measurements. Although a statistically significant difference was observed between physician measurements and LLM predictions without radiographic input, incorporation of PA chest radiographs eliminated this statistical difference, suggesting improved concordance with clinician-determined positioning. Importantly, LLM-generated predictions did not influence clinical tube placement decisions in this study and therefore represent retrospective agreement rather than real-time guidance.

The importance of accurate DLT positioning in thoracic anesthesia is well established, as improper placement may result in inadequate lung isolation and perioperative complications (1, 2). Meta-analytic data further support the central role of DLTs in OLV practice (3), reinforcing the clinical relevance of optimizing initial insertion depth.

Height-based formulae have historically been proposed to guide DLT placement (4, 5, 16). However, studies have consistently demonstrated that while statistically significant correlations exist, interindividual variability limits the clinical reliability of such approaches (6, 7). Our findings are consistent with this literature, as demographic-only LLM prediction demonstrated measurable deviation from physician-confirmed depth.

Imaging-based strategies have attempted to address anatomical variability. CT-guided approaches have shown improved initial positioning accuracy compared with blind insertion techniques (10, 11). However, routine CT-based planning may not be practical for real-time decision support. In contrast, the present LLM-based approach utilizes routinely obtained PA chest radiographs, thereby enhancing feasibility and potential clinical applicability.

Recent advances in artificial intelligence, including large language models, have expanded the potential for data-driven clinical decision support tools (12–14). Prior work has demonstrated the feasibility of using radiographic parameters to predict DLT depth and size (15). Our study extends this concept by directly evaluating agreement with physician-confirmed insertion depth in a real-world cohort.

Bland–Altman analysis demonstrated minimal systematic bias for both LLM-based approaches, with limits of agreement within clinically acceptable ranges. Although radiographic input did not substantially

alter the magnitude of bias, statistical comparison indicated improved agreement when PA chest radiographs were incorporated.

Although the 95% limits of agreement approached ± 2.8 cm, more than 90% of predictions were within ± 2 cm of clinician-determined depth, and the mean absolute error was approximately 1 cm for both approaches. This distribution indicates that the majority of deviations were small in magnitude. Furthermore, incorporation of image-derived anatomical information may help reduce discrepancies between LLM-generated depth estimates and clinician-confirmed positioning.

Several limitations should be acknowledged. First, the retrospective design may introduce selection bias, and external validation in independent cohorts is required before broader clinical implementation. Second, the study evaluated left-sided DLTs exclusively; therefore, the findings may not be directly generalizable to right-sided devices.

Finally, predictions were generated sequentially within an ongoing model session using default system-level inference settings. Although a standardized prompt and single-response strategy were applied to minimize variability and no outcome feedback was provided to the model, potential contextual influence between consecutive cases cannot be entirely excluded. In addition, because large language models may evolve over time and inference behavior may vary across sessions or system updates, full reproducibility across different environments cannot be guaranteed. These methodological considerations should be taken into account when interpreting the findings.

In conclusion, LLM-based prediction of DLT insertion depth demonstrated reasonable agreement with physician-confirmed measurements, with minimal systematic bias and low absolute error. Although the 95% limits of agreement approached ± 2.8 cm, the majority of predictions were within ± 2 cm of clinician-determined depth.

While bronchoscopic confirmation remains indispensable for optimal airway management, LLM-assisted depth estimation may serve as a supportive adjunct for guiding initial DLT placement, particularly in settings where fiberoptic bronchoscopy resources are limited or delayed. Prospective validation under controlled inference conditions is warranted before real-time clinical integration can be considered.

Abbreviations

AI
Artificial Intelligence
DLT
Double-Lumen Tube
FOB
Fiberoptic Bronchoscopy
LLM
Large Language Model

OLV
One-Lung Ventilation
PA
Posteroanterior
SD
Standard Deviation
LoA
Limits of Agreement

Declarations

Ethics approval and consent to participate

This retrospective, single-center observational study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Clinical Research Ethics Committee of the University of Health Sciences Turkey, Basaksehir Cam and Sakura City Hospital (Approval No: E-96317027-514.10-301752463; Date: 20.01.2026). The requirement for informed consent was waived by the ethics committee due to the retrospective nature of the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution

C.K.B. conceived and designed the study, collected and analyzed the data, performed the statistical analysis, and drafted the manuscript. H.A. contributed to study design, data interpretation, and critical revision of the manuscript. F.T. contributed to data collection and interpretation of the findings. E.Ş.T. contributed to methodological planning and manuscript revision. F.G.Ö. supervised the study, contributed

to interpretation of the results, and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Weiskopf RB, Campos JH. Current techniques for perioperative lung isolation in adults. *Anesthesiology*. 2002;97(5):1295–301.
2. Brodsky JB. Lung separation and the difficult airway. *Br J Anaesth*. 2009;103(Suppl 1):i66–75.
3. Palaczynski P, Misiolek H, Szarpak L, Smereka J, Pruc M, Rydel M, et al. Systematic review and meta-analysis of efficiency and safety of double-lumen tube and bronchial blocker for one-lung ventilation. *J Clin Med*. 2023;12(5):1877.
4. Brodsky JB, Lemmens HJM. Tracheal width and left double-lumen tube size: A formula to estimate left bronchial width. *J Clin Anesth*. 2005;17(4):267–70.
5. Lin WL, Cherng CH. Predicting optimal insertion depth of a left-sided double-lumen endobronchial tube. *J Cardiothorac Vasc Anesth*. 2016;30(4):942–6.
6. Yasumoto M, Higa K, Nitahara K, Shono S, Hamada T. Optimal depth of insertion of left-sided double-lumen endobronchial tubes cannot be predicted from body height in below average-sized adult patients. *Eur J Anaesthesiol*. 2006;23(1):42–4.
7. Woo JH, Cho S, Kim YJ, Kim DY, Choi Y, Lee JW. Depth of double-lumen endobronchial tube: A comparison between real practice and clinical recommendations using height-based formulae. *Anesth Pain Med*. 2023;18(1):37–45.
8. Hegland N, Schnitzler S, Ellensohn J, Steurer MP, Weiss M, Dullenkopf A. Dimensional variations of left-sided double-lumen endobronchial tubes. *Anesthesiol Res Pract*. 2019;2019:3634202.
9. Russell WJ, Strong TS. Dimensions of double-lumen tracheobronchial tubes. *Anaesth Intensive Care*. 2003;31(1):50–3.
10. Li J, Qian Y, Lei Y, Huo W, Xu M, Zhang Y, et al. Combination of computed tomography measurements and flexible video bronchoscope guidance for accurate placement of the right-sided double-lumen tube: A randomized controlled trial. *BMJ Open*. 2023;13(11):e066541.

11. Liu Z, Zhao L, Jia Q, Yang X, Liang SJ, He W. Chest computed tomography image for accurately predicting the optimal insertion depth of left-sided double-lumen tube. *J Cardiothorac Vasc Anesth*. 2018;32(2):855–9.
12. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
13. Rajpurkar P, Irvin J, Ball RL, Zhu K, Yang B, Mehta H, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15(11):e1002686.
14. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24–9.
15. Chang TR, Yuan MK, Pan SF, Chuang CC, So EC. Double-lumen endotracheal tube—predicting insertion depth and tube size based on patient’s chest X-ray image data and four other body parameters. *Diagnostics (Basel)*. 2022;12(12):3162.
16. Chow MYH, Goh MH, Ti LK. Predicting the depth of insertion of left-sided double-lumen endobronchial tubes. *J Cardiothorac Vasc Anesth*. 2002;16(4):456–8.

Figures

Bland-Altman of Physician vs AI

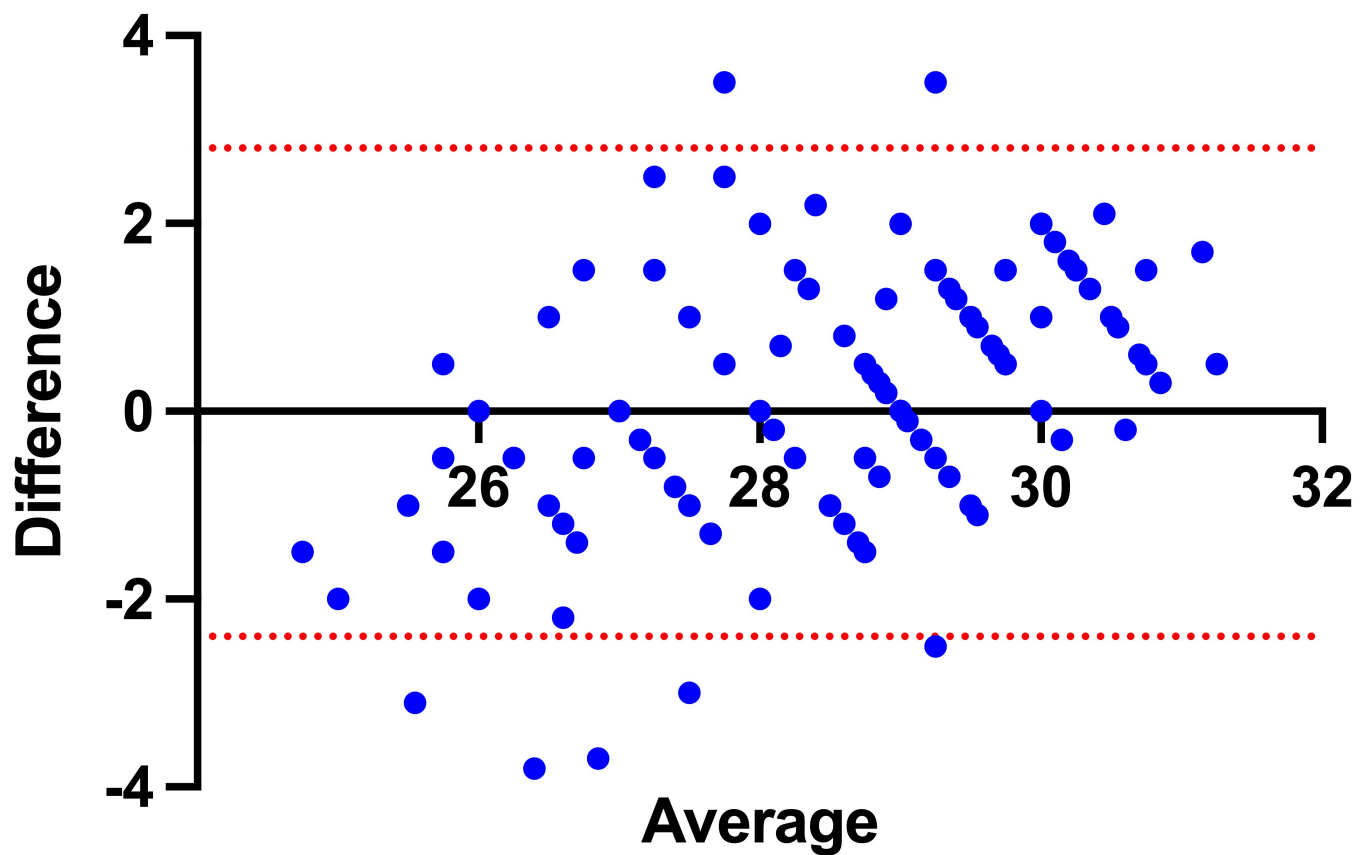


Figure 1

Bland-Altman of Physician vs AI

Bland-Altman of Physician vs AI + PA

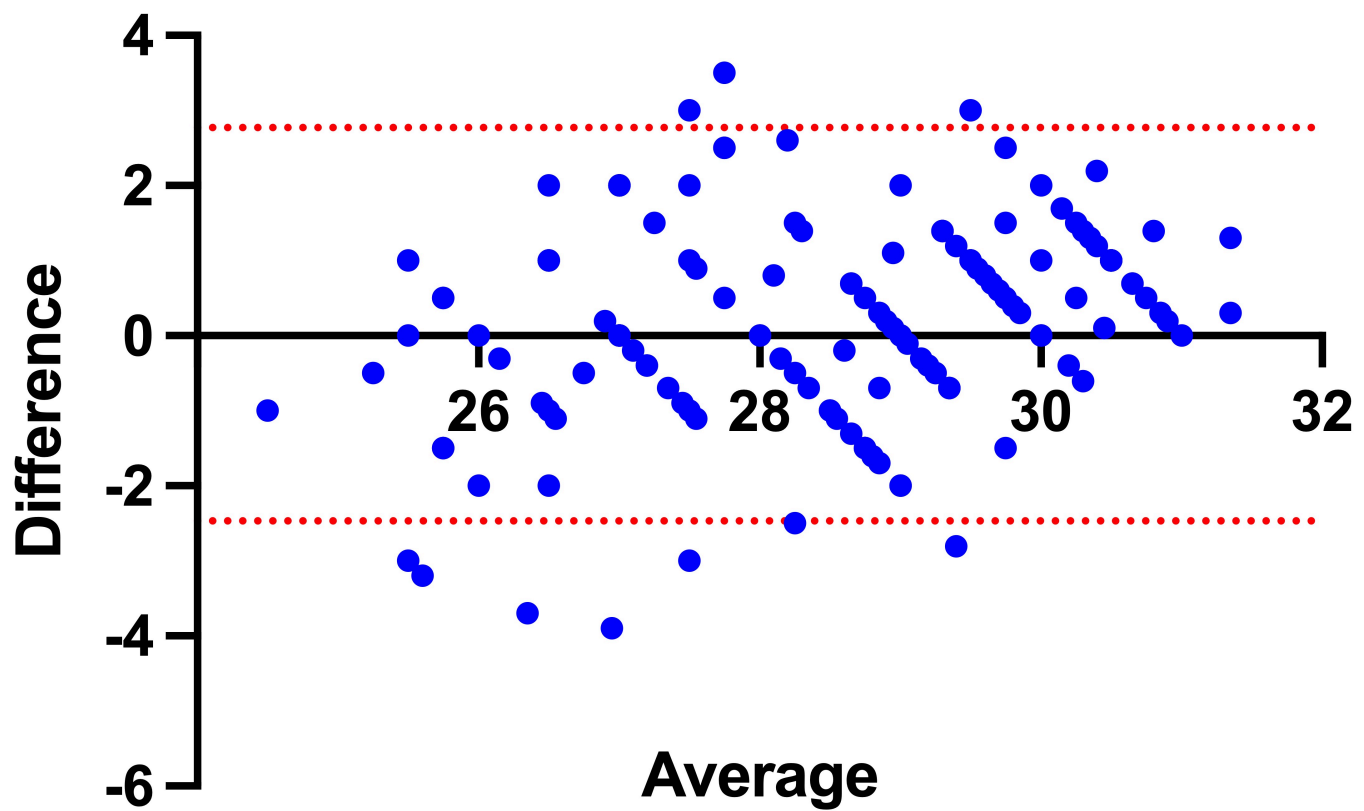


Figure 2

Bland-Altman of Physician vs AI + PA