References

39. OCEBM Levels of Evidence Working Group. OCEBM Levels of Evidence [cited 2023 Nov 11]. Available online: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
40. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.
41. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36.
42. Baratloo A, Hosseini M, Negida A, El Ashal G. Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. Emerg (Tehran) 2015;3:48-9.
43. Trevethan R. Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. Front Public Health 2017;5:307.
44. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22:276-82.
45. Tanaka M, Kanazawa A, Yonenobu K. Diagnosis of OPLL and OYL: Overview. In: Yonenobu K, Nakamura K, Toyama Y, editors. OPLL: Ossification of the Posterior Longitudinal Ligament. 2nd Edition. Tokyo: Springer; 2006. p. 111–3.
46. Tsuyama N. Ossification of the posterior longitudinal ligament of the spine. Clin Orthop Relat Res 1984;(184):71-84.
47. South A. rworldmap: Mapping Global Data [cited 2023 Nov 11]. Available online: https://cran.r-project.org/web/packages/rworldmap/rworldmap.pdf
48. McGuinness LA. Introduction to robvis, a visualization tool for risk-of-bias assessments; 2019 [cited 2023 Nov 11]. Available online: https://cran.r-project.org/web/packages/robvis/vignettes/Introduction_to_robvis.html
49. Gehlenborg, Jake Conway and Nils. Basic Usage; 2022 [cited 2023 Nov 11]. Available online: https://cran.r-project.org/web/packages/UpSetR/vignettes/basic.usage.html
50. Schwarzer G, Carpenter JR, Rücker G. Meta-Analysis with R. Cham: Springer; 2015. (EBL-Schweitzer) [cited 2023 Nov 11]. Available online: https://cran.r-project.org/web/packages/meta/meta.pdf

Table S1 Inclusion and exclusion criteria

Inclusion criteria
Published in the English language
Peer-reviewed journals
Specified radiological method (e.g., radiography, CT, MRI)
Adult human patients with cervical OPLL
Use of deep learning model (CNN)
Exclusion criteria
All non-English languages
Commentaries, case reports, narrative reviews, letters to editors, books
Animal studies or lab-based studies
Studies on children and adolescents (<18 years)
Patient with previous spinal surgery

CNN, convolutional neural network; CT, computed tomography; MRI, magnetic resonance imaging; OPLL, ossification of the posterior longitudinal ligament.

https://dx.doi.org/10.21037/qims-24-1485

Table S2 Search strategy conducted on October 1st, 2023, delineating the databases used, search terms employed, publication dates, and the corresponding results retrieved from each database

Database	Search terms	Publication dates	Results (n)
Google Scholar	"Deep Learning" AND "Ossification of the Posterior Longitudinal Ligament"	1998–2023	106
ScienceDirect	"Deep Learning" AND "Ossification of the Posterior Longitudinal Ligament"	2020–2023	10
PubMed	("deep learning"[MeSH Terms] OR ("deep"[All Fields] AND "learning"[All Fields]) OR "deep learning"[All Fields]) AND ("ossification of posterior longitudinal ligament"[MeSH Terms] OR ("ossification"[All Fields] AND "posterior"[All Fields] AND "longitudinal"[All Fields] AND "ligament"[All Fields]) OR "ossification of posterior longitudinal ligament"[All Fields])	2021–2023	6
BASE	"Deep Learning" AND "Ossification of the Posterior Longitudinal Ligament"	2021–2023	6
Cochrane Library	"Deep Learning" AND "Ossification of the Posterior Longitudinal Ligament"	2021	1

Table S3 Level of evidence of each of the included studies based on the Oxford Centre of Evidence-Based Medicine (OCEBM) Levels of Evidence (39)

Study number	Author, Year	Level of evidence
1	Ogawa *et al.* (2022)	3
2	Miura *et al.* (2021)	3
3	Murata *et al.* (2021)	3
4	Chae *et al.* (2022)	3
5	Tamai *et al.* (2022)	3
6	Ito *et al.* (2023)	2b
7	Shemesh *et al.* (2023)	3

Table S4 Grading of Recommendations, Assessment, Development and Evaluations (GRADE) scoring for all studies (40)

Study	Risk of Bias	Imprecision	Inconsistency	Indirectness	Publication bias	Overall
Ogawa *et al.* (2022)	Low	Low	Low	Moderate	Low	Moderate
Miura *et al.* (2021)	Moderate	Moderate	Low	Very low	Low	Moderate
Murata *et al.* (2021)	Low	Low	Low	Moderate	Low	Moderate
Chae *et al.* (2022)	Low	Low	Low	Moderate	Low	Moderate
Tamai *et al.* (2022)	Low	Moderate	Low	Low	Low	Moderate
Ito *et al.* (2023)	Low	Moderate	Low	Low	Low	Moderate
Shemesh *et al.* (2023)	Moderate	Low	Very low	Very low	Low	Moderate

https://dx.doi.org/10.21037/qims-24-1485

Table S5 Table of extracted items from both the qualitative synthesis (systematic review) and quantitative synthesis (meta-analysis)

Extracted items for qualitative synthesis
Study, sample size, study type and design, country, deep learning model, model construction, radiological technique, control, patient demographic, indication for radiology, accuracy, AUC, sensitivity, specificity, TP, FN, TN, FP, F1 value, k1 value, recall, precision, NPV, PPV, OPLL subtypes, main conclusion, risk of bias
Extracted items for quantitative synthesis
Accuracy, AUC, sensitivity, OPLL subtypes, human performance

AUC, area under the curve; FN, false negative; FP, false positive; NPV, negative predictive value; OPLL, ossification of the posterior longitudinal ligament; PPV, positive predictive value; TN, true negative; TP, true positive.

Table S6 Risk of bias analysis of all included studies (QUADAS-2 tool) (41)

Study	Risk of Bias				Applicability Concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ogawa *et al.* (2022)	Unclear	Low	Unclear	Low	Low	Unclear	Low
Miura *et al.* (2021)	Low	Low	Low	Low	Low	Unclear	Unclear
Murata *et al.* (2021)	Low	Unclear	Low	Low	Low	Low	Low
Chae *et al.* (2022)	Low	Low	Low	Low	Low	Unclear	Unclear
Tamai *et al.* (2022)	Low	Unclear	Unclear	Low	Low	Low	Low
Ito *et al.* (2023)	Low	Unclear	Low	Low	Low	Low	Low
Shemesh *et al.* (2023)	Low	Low	Low	Low	Unclear	Unclear	Unclear

https://dx.doi.org/10.21037/qims-24-1485

Table S7 Table of extracted outcome variables with an definition for each variable (42-44)

Variable	Definition
Accuracy	Accuracy is the ratio of correctly predicted positive and negative observations to the total observations. Formula: (TP + TN)/(TP + TN + FP + FN)
Sensitivity	Sensitivity (also known as true positive rate/recall) measures the percentage of actual positive cases that were correctly identified. Formula: TP/(TP + FN)
Specificity	Specificity (also known as true negative rate) indicates how well a test correctly identifies negative cases. Formula: TN/(TN + FP)
Precision	Ratio of true positive predictions to the total predicted positives. Formula: TP/(TP + FP)
AUC	The AUC measures the classifier's performance across all possible threshold values and represents the area under the ROC curve
ROC	ROC is a graphical representation of the trade-off between sensitivity and specificity at various thresholds.
NPV	The NPV is the percentage of actual negative cases that were correctly identified. Formula: TN/(TN + FN)
PPV	The PPV indicates the likelihood of actual positive results being correctly identified. Formula: TP/(TP + FP)
K1 score	The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall, taking both false positives and false negatives into account. Formula: 2 * (precision * recall)/(precision + recall), where precision = TP/(TP + FP) and recall = TP/(TP + FN)
Cohen's kappa score	Cohen's kappa score is a measure of inter-rater agreement for categorical items. It considers the agreement occurring by chance and adjusts the observed agreement accordingly. The formula involves observed and expected agreement between two raters or classifiers

TP, true positive; TN, true negative; FP, false positive; FN, false negative; AUC, area under the curve; ROC, receiver operating characteristic; NPV, negative predictive value; PPV, positive predictive value.

Table S8 OPLL subtypes definitions (based on Tanaka *et al.* and proposed by the Investigation Committee on OPLL of the Japanese Ministry of Public Health and Welfare by Tsuyama *et al.*). (45,46)

OPLL subtype	Definition
Segmental	Involves ossification behind each vertebral body
Continuous	Ossified mass that spans several vertebral bodies and the intervening disk spaces
Mixed	Mixture of both continuous and segmental types
Localized	Ossification is localized to the intervertebral disk space without involvement of the vertebral body

OPLL, ossification of the posterior longitudinal ligament.

https://dx.doi.org/10.21037/qims-24-1485

Table S9 The R code utilized in this review

Graph	Code

World Map (47)	install.packages("rworldmap")
	library(rworldmap)
	studien_daten <- data.frame(
	  Land = c("Japan", "S. Korea", "Israel"),
	  Studies = c(5, 1, 1)
	)
	worldprep <- getMap()
	worldprep$Studies <- NA
	for (i in 1:nrow(studien_daten)) {
	  country_name <- studien_daten$Land[i]
	  worldprep$Studies[worldprep$NAME == country_name] <- studien_daten$Studies[i]
	}
	mapCountryData(
	  mapToPlot = worldprep,
	  nameColumnToPlot = "Studies",
	  xlim = c(20, 180), # Begrenzung der Längengrade
	  ylim = c(-10, 100), # Begrenzung der Breitengrade
	  catMethod = c(0:5),
	  colourPalette = "heat",
	  addLegend = TRUE,
	  borderCol = "black",
	  mapTitle = "", # Leere Graph-Überschrift
	  aspect = 1,
	  missingCountryCol = "lightgrey",
	  add = FALSE,
	  nameColumnToHatch = TRUE,
	  lwd = 0.5,
	  oceanCol = NA
	)
	title("Number of studies per country", line = -7)

Robvis traffic light (48)	install.packages("robvis") library(robvis) summary_plot <- rob_traffic_light(data = Risk_of_Bias_assessment, tool = "QUADAS-2", colour = "cochrane", psize = 10) summary_plot
Robvis summary (48)	library(robvis) summary_plot <- rob_summary(data = Risk_of_Bias_assessment, tool = "QUADAS-2", weighted = FALSE, overall = TRUE) summary_plot
Upset (49)	install.packages("UpSetR") library(UpSetR) library(readxl) main_bar_color <- "choose color" matrix_color <- "choose color") excel_data <- readxl::read_excel study <- as.data.frame(excel_data) upset(fromList(study), sets = names(study), sets.bar.color = "choose color", main.bar.color = main_bar_color, matrix.color = matrix_color)

Table S9 (*continued*)

https://dx.doi.org/10.21037/qims-24-1485

Table S9 (*continued*)

Graph	Code
Funnel (50)	install.packages("ggplot2") install.packages("meta") library(ggplot2) library(meta) library(readxl) excel_data <- read_excel meta_analysis <- metaprop(event = excel_data$Sensitivity * excel_data$SampleSize, n = excel_ data$SampleSize) funnel(meta_analysis)
Scatter plot matrix	install.packages("readxl") library(readxl) excel_data <- read_excel plot <- pairs(~ Accuracy + AUC + Sensitivity + Specificity, data = excel_data, main = "Scatterplot Matrix", labels = c("Accuracy", "AUC", "Sensitivity", "Specificity"), cex.labels = 2, font.labels = 2)
Metaanalysis (50)	library(meta) library(readxl) data <- read_excel(excel_file) overall_meta <- metaprop(event = data$Events, n = data$Total, studlab = data$Study, byvar = data$OPLL, comb.random = TRUE) forest(overall_meta, fixed = FALSE, random = TRUE, overall = TRUE, col.random = "red", prediction = TRUE, text.fixed = "Test for subgroup differences", rowsize = 2, cex = 0.2)

https://dx.doi.org/10.21037/qims-24-1485