Automated Scoliosis Diagnosis in Spinal Imaging: Laboratory Validation, Clinical Limitations, and Systematic Implementation Challenge Review

Ervin Gubin Moung[1], Xie Aishu[2], Ali Farzamnia[3]
Faculty of Computing and Informatics, Universiti Malaysia Sabah, Jalan UMS, Kota Kinabalu, 88400, Sabah, Malaysia [1, 2]
School of Computing and Engineering, University of Huddersfeld, West Yorkshire, HD1 3DH, United Kingdom [3]

***Abstract*—Technological advances in automated medical imaging diagnosis have created translation gaps between laboratory achievements and clinical implementation, with traditional manual Cobb angle measurement requiring considerable time with inevitable measurement errors. This review analyzes translation challenges in automated diagnosis systems using scoliosis assessment as a case study, examining 55 articles from 1948-2025 across three domains: Cobb angle measurement, classification, and segmentation. Despite research investment, fully automated approaches have not surpassed semi-automated performance in comparable validation studies. Within the 23 Cobb angle measurement studies, traditional methods outperform sophisticated deep learning systems with average error rates of 1.8° ± 0.4° MAD versus 4.2° ± 1.8° MAE, while validation degradation occurs with performance dropping from 95.28% to 85.9% when transitioning to real-world datasets. Non-standard classification achieves high accuracy but lacks clinical utility, while standard systems struggle with automation, revealing a translation paradox where technical sophistication does not correlate with clinical adoptability. Main problems include testing method gaps, performance drops, different automation approaches, and cost issues. This review recommends standard testing methods and step-by-step clinical implementation to help these systems work in real clinics.**

Keywords—Automated diagnosis; medical imaging; scoliosis; Cobb angle; clinical implementation; artificial intelligence

I. Introduction

Medical imaging has undergone major advances over the past seven decades, with automated systems achieving high performance in laboratory settings. Despite these achievements, translation into clinical practice remains limited. Scoliosis assessment illustrates this gap clearly. While AI systems for diabetic retinopathy and breast cancer detection have achieved clinical success with standardized protocols and FDA approval [IDx-DR, 100,000 patients annually; breast cancer AI 94.5% sensitivity], scoliosis measurement systems remain confined to research laboratories.

The field has experienced continuous development of diagnostic approaches since 1948, yet clinical challenges remain unresolved. Researchers have proposed multiple measurement principles for scoliosis assessment, including the Ferguson method [1], Greenspan method [2], Diab method [3], and Centroid method [4]. These developments demonstrate ongoing efforts to enhance diagnostic accuracy and reliability. The Cobb angle measurement method was developed by Cobb in 1948 and adopted by the Scoliosis Research Society as the standard method for quantifying scoliotic deformities in 1966. This method remains the clinical gold standard despite well-documented limitations. The traditional manual measurement process has problems with errors between different doctors and even the same doctor measuring twice [5], with studies showing significant differences in measurements [6][7][8][9]. The Ferguson method exhibits considerable intra-observer and inter-observer variation [1]. Measurement inconsistencies continue to exist across established methodologies [10][11][12]. Since 2017, focus has transferred to estimating Cobb angles using Artificial Intelligence (AI) models, with researchers [13] implementing computer-aided measurement based on automatic detection of vertebral slopes using deep neural network, and [14] proposing a Convolutional Neural Network (CNN)-based spine and Cobb angle estimator using moire images. Modern approaches like [15] achieved impressive results using convolutional neural networks for Cobb angle measurement, while [16] achieved high performance using U-Net webserver for spine segmentation, and [17] demonstrated effective lumbar vertebrae segmentation. This technological sophistication has not translated proportionally into clinical implementation rates, revealing limitations that distinguish scoliosis AI from successful medical AI implementations in radiology and pathology.

A. Scope and Objectives

To address this problem, this review asks three main questions: (1) What factors prevent laboratory-validated systems from achieving comparable performance in clinical settings? (2) What barriers hinder adoption of these systems in actual workflows? and (3) What strategies can bridge the gap between laboratory development and clinical implementation?

B. Paper Organization

The remainder of this paper is organized as follows: the Methodology section outlines the literature search and study selection. Background and Context review historical development and contrasts scoliosis AI with successful medical AI domains. Current State Analysis summarizes recent performance trends, while Critical Analysis investigates the causes of translation failures. Future Directions provide recommendations, and Conclusions synthesize key findings for stakeholders.

II. Methodology

A. Search Strategy and Study Selection

A comprehensive literature search was conducted across multiple databases including PubMed/MEDLINE, IEEE Xplore, ACM Digital Library, and Scopus from January 1948 to February 2025. The search strategy combined controlled vocabulary terms and keywords related to scoliosis, automated measurement, artificial intelligence, and clinical implementation. Primary search terms included: ("scoliosis" OR "spinal curvature") AND ("automated" OR "computer-assisted" OR "artificial intelligence" OR "deep learning" OR "machine learning") AND ("Cobb angle" OR "measurement" OR "classification" OR "segmentation"). Additional studies were identified through citation tracking of seminal papers and review of reference lists.

B. Inclusion and Exclusion Criteria

Inclusion criteria were: (1) English-language, peer-reviewed articles or conference proceedings; (2) studies on automated or computer-assisted scoliosis assessment; (3) work covering Cobb angle measurement, scoliosis classification, or spinal segmentation; (4) studies reporting quantitative performance metrics; and (5) articles providing enough methodological detail for analysis. Exclusion criteria were: (1) research focused only on treatment or surgical interventions; (2) articles lacking performance metrics or validation data; (3) duplicate publications, including conference papers later published as journal articles; (4) studies on non-human subjects; and (5) purely theoretical work without experimental validation.

C. Study Categorization and Data Extraction

Selected studies were categorized into three primary domains: (1) Cobb angle measurement approaches, (2) scoliosis classification systems, and (3) spinal structure segmentation methods. This categorization was based on the primary research objective and methodology reported in each study. Data extraction followed a standardized protocol that captured: study characteristics (authors, year, study design, sample size), technical approaches (algorithms employed, imaging modalities), performance metrics (accuracy measures, reliability statistics, validation methods), and clinical implementation details (automation level, workflow integration, clinical testing). Performance metrics were standardized where possible, with accuracy measures converted to comparable units (degrees for Cobb angle measurement, percentages for classification and segmentation). When studies reported multiple metrics, priority was given to the most clinically relevant measure (e.g., mean absolute difference for Cobb angles, overall accuracy for classification). The selection process for this review followed established guidelines for literature identification and screening. This review analyzed 55 studies spanning from 1948 to 2025 across the three primary domains of automated scoliosis assessment. Fig. 1 illustrates the comprehensive search and selection methodology employed to identify the final 55 studies included in this review, demonstrating the distribution across the three primary domains of automated scoliosis assessment.

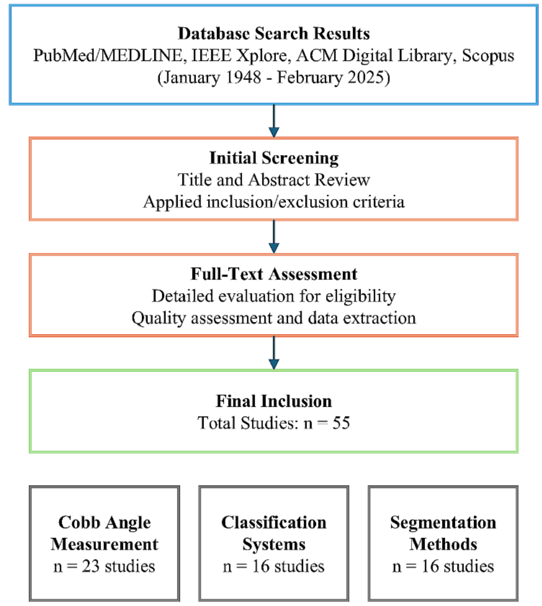

Fig. 1. Study selection flow diagram for review of automated scoliosis.

The flowchart shown in Fig. 1 demonstrates the systematic selection process from initial database searches through final inclusion, resulting in 55 studies categorized across three primary domains: Cobb angle measurement (n=23), classification systems (n=16), and segmentation methods (n=16). The distribution of studies across the three domains reflects the evolution of research focus over the 77-year timeframe examined. Cobb angle measurement studies represent the largest portion (42% of total), reflecting the central importance of angular measurement in scoliosis assessment. Classification and segmentation approaches each represent 29% of studies, indicating that researchers recognize effective automation requires progress across multiple complementary technologies. This balanced distribution provides comprehensive coverage of translation challenges across all aspects of automated scoliosis assessment.

D. Limitations

This review has several limitations: (1) differences in performance reporting across studies limited quantitative analysis; (2) publication bias may favor positive results, potentially overestimating automated system performance; (3) limited availability of real-world clinical validation data restricted translation analysis; (4) rapid technological changes may make some analyses outdated; (5) focus on English-language publications may have excluded relevant international research.

III. Background and Context

A. Evolution of Automated Scoliosis Assessment

The field has evolved through four technological phases, each with different strengths, weaknesses, and adoption rates. As technological sophistication increased, clinical implementation declined. Table I presents the development phases and their duration, showing rapid technological development but declining clinical implementation success.

TABLE I. TECHNOLOGY DEVELOPMENT PHASES

Phase	Period	Duration
Manual Era	1948-1990	42 years
Early Automation	1990-2010	20 years
Classical ML Peak	2010-2017	7 years
Deep Learning Era	2017-Present	7+ years

The rapid development in Table I differs from the gradual, steady progress seen in successful medical AI domains like mammography, where evolution occurred over similar timeframes but achieved progressive clinical integration rather than decreased adoption. The performance and clinical implementation patterns across these phases show the main problem. Table II shows that as technical performance improved, clinical implementation decreased.

TABLE II. PERFORMANCE VS. CLINICAL IMPLEMENTATION

Phase	Best Performance	Clinical Implementation
Manual Era	Intra: 5.6°, Inter: 6.6° [1]	100%
Early Automation	Variable improvement [18][11][12]	<5%
Classical ML Peak	MAD: 1.5°, ICC: >0.95 [10][19]	<5%
Deep Learning Era	MAE: 4.9°, DC: 95.8% [20][21]	<1%

Table II shows a pattern that differs from successful medical AI domains. In successful domains, better technology leads to more clinical use. Diabetic retinopathy screening moved from research to FDA approval to widespread use over a similar timeframe. It was found that specific factors in scoliosis context, rather than general AI problems, cause implementation difficulties, which suggest that targeted approaches may perform better and show improved progress within clinical practice settings.

B. Medical AI Translation Context: Lessons from Successful Domains

Scoliosis AI faces unique translation challenges. Comparison with successful medical AI implementations reveal differences in problem structure and validation approaches, which suggest that targeted solutions perform better within scoliosis treatment contexts.

1) Successful medical AI implementations: It was found that diabetic retinopathy screening achieved clinical success mainly because the task was simplified into binary classification of present or absent, with the added benefit of standardized imaging protocols that allowed for consistent practice within clinical workflows. Clear clinical decision points together with regulatory pathways appear to support adoption. Breast cancer detection AI perform better by supporting radiologists, which mean it show improved progress in accuracy without full automation. These cases suggest clinical acceptance is more likely when AI act as enhancement with professionals, not as substitutes.

2) Scoliosis AI differences: Scoliosis assessment was found to require continuous measurement instead of binary classification, and this make the task more complex than other screening-based AI applications. It lack standardized imaging protocols across hospitals, which suggest that consistent practice remain difficult to achieve. Scoliosis assessment also need to integrate with complex treatment decision algorithms, and this requirement only perform better when AI is used as enhancement rather than full replacement. In addition to that, the current regulatory frameworks lack provisions for automated measurement systems, so the approval pathway remain unclear. These fundamental differences, repeated again to emphasize, clarify why proven strategies from other medical AI applications prove ineffective when directly applied to scoliosis assessment, and this suggest improvement will only be achieved by adopting different design approaches. Comparing successful and failed medical AI implementations shows major differences in problem structure, implementation strategies, and clinical outcomes. Fig. 2 illustrates the key distinguishing factors between successful medical AI domains and the failed scoliosis AI domain, highlighting the strategic and technical decisions that determine translation success.

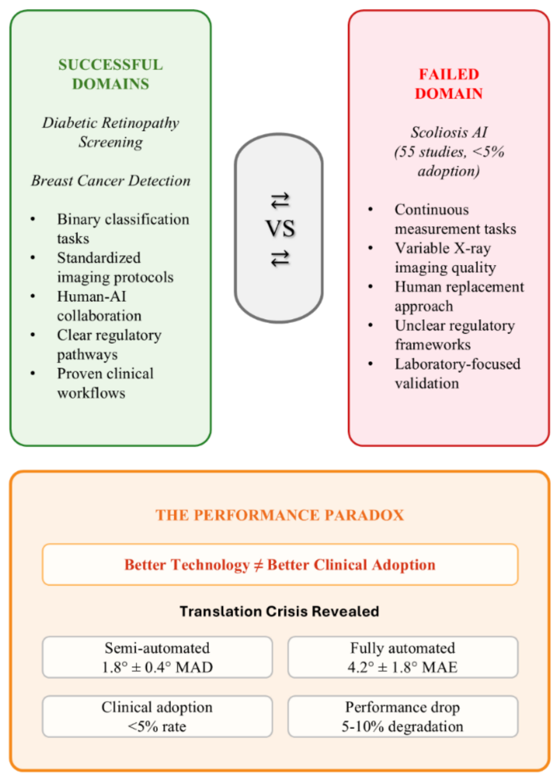

Fig. 2. Medical AI success vs failure analysis.

Fig. 2 shows how important it is to match the right technology with the right problem for medical AI success. Successful domains benefit from binary classification tasks that align with deep learning strengths, while scoliosis AI requires continuous geometric measurement that favors traditional mathematical approaches. The performance paradox contradicts conventional expectations. Semi-automated methods achieve better results (1.8° ± 0.4° MAD) compared to fully automated systems (4.2° ± 1.8° MAE). This finding shows that current AI approaches do not match the geometric precision needed for scoliosis assessment. This analysis helps explain why helping doctors work better than replacing doctors for successful clinical use. The foundation era established measurement principles through the Ferguson method [1], Greenspan method [2], and Diab method [3], but these were designed for manual implementation. Early computer-assisted approaches [18][11][12] showed promise but remained time-consuming due to manual endpoint identification requirements. In [18], authors compared radiographic and computer-assisted measurements, while in [11] and [12], the authors evaluated manual versus computer-assisted approaches. Traditional image processing and classical machine learning approaches achieved peak performance during 2010-2017. In [19], the authors achieved ICC>0.95 using Fuzzy Hough Transform [22] and [23] achieved ICC>0.9 using Active Shape Models. However, these successes occurred during a period when other medical AI domains were transitioning from research to clinical implementation. This timing highlights scoliosis AI's unique translation challenges.

C. Deep Learning Era and Divergence from Medical AI Success Patterns

Deep learning approaches led to a shift toward complete automation instead of human-AI collaboration that worked well in other medical domains. This divergence explains much of the current translation crisis.

Several researchers developed deep learning solutions for scoliosis assessment. In research by [13], authors developed computer-aided Cobb measurement using deep neural networks, while researchers in [14] proposed CNN-based estimators and those in [15] achieved significant results using convolutional neural networks.

Advanced segmentation approaches also emerged. In studies by [16], researchers deployed U-Net achieving 90.4% dice coefficient, and authors in [24] proposed comprehensive spine detection frameworks. In research by [25], authors developed automatic vertebrae localization for CT scans, researchers in [28] demonstrated CNN-based lumbar spine segmentation, and those in [29] contributed deep learning methods for spine analysis in MR images. Other technical approaches included [30] implementing BoostNet for automatic landmark estimation, [31] developing structured multi-output regression for direct Cobb angle estimation, and [32] proposing vertebra detection and corner regression techniques. However, this period showed a different pattern from successful medical AI domains. While domains like radiology succeeded through augmentation approaches that enhanced radiologist capability, scoliosis AI pursued replacement approaches that attempted complete automation. Classification approaches showed similar patterns, with [33] achieving results through local centroids evaluation, [34] implementing support vector classifiers for progression risk prediction, and [35] developing machine learning approaches for clinical classification using 3D surface data. Feature-based approaches like [36] demonstrated vertebrae localization and segmentation for curvature classification, while [37] applied support vector machines for severity assessment from surface topography. Recent developments include [38] analyzing scoliosis from spinal X-ray images, [39] implementing artificial neural networks for spinal deformity identification, and various classification frameworks [40][41][42][43] addressing different aspects of scoliosis assessment with varying degrees of automation success. Clinical reality involves complex decision-making beyond just measurement accuracy. The Lenke classification system [44] serves as the popular standard, but it requires automation of the entire Cobb method instead of providing decision support for specific components. Studies evaluating classification reliability [45] demonstrate the complexity of implementing automated systems that match clinical standards.

D. Summary

The analysis shows that scoliosis AI fails not because AI itself is bad, but because researchers chose the wrong approach compared to successful AI in other medical areas. Understanding these basic differences is important for analyzing current performance patterns and finding effective solutions. The next section examines how these different approaches affect current system performance and clinical implementation. Building on this analysis, the following section examines current performance patterns across the three primary domains.

IV. Current State Analysis

A. Cobb Angle Measurement Approaches

Automated Cobb angle measurement demonstrates a concerning trend where increased technological sophistication correlates with decreased clinical adoption. This paradox becomes more pronounced when compared to successful medical AI domains where sophistication enhances rather than hinders adoption. Analysis of 23 studies identified three distinct methodological approaches: traditional methods (3 studies), semi-automated approaches (8 studies), and fully automated systems (12 studies). This contradicts patterns in successful medical AI domains where deep learning typically outperforms traditional approaches, suggesting differences in problem structure. Table III presents an analysis of performance across different methodological approaches, including Mean Absolute Difference (MAD) and Mean Absolute Error (MAE) where available from the literature. Performance metrics vary across studies due to different datasets, validation methods, and measurement protocols. Direct numerical comparison should be interpreted cautiously due to methodological differences in error calculation across studies.

TABLE III. AVERAGE PERFORMANCE ANALYSIS ACROSS STUDY TYPES

Method Type	Performance	Studies (n)	Translation Rate
Manual	3.9° ± 0.3° MAD	Baseline	100% (standard)
Semi-Automated	1.8° ± 0.4° MAD	8	<5% adoption
Fully Automated	4.2° ± 1.8° MAE	12	<1% adoption

Table III shows performance metrics as mean ± standard deviation across multiple studies on semi-automated and fully automated approaches, revealing statistically significant performance differences that contradict expectations based on successful medical AI domains.

Semi-automated methods achieve better performance than deep learning systems (1.8° ± 0.4° MAD vs 4.2° ± 1.8° MAE) for three primary reasons. First, continuous measurement tasks favor geometric approaches over pattern recognition (Task-Method Mismatch). Second, deep learning models optimize for controlled datasets that poorly represent clinical variability (Overfitting to Laboratory Conditions). Third, end-to-end learning discards domain-specific geometric constraints that are essential for reliable measurement (Feature Engineering Loss). Traditional approaches achieved superior performance, with researchers in [19] reaching ICC>0.95 using Fuzzy Hough Transform [22] because geometric constraints built into spine anatomy provide more reliable guidance than learned features. Semi-automated approaches like [46] achieved significant accuracy using generalized Hough transform for cervical vertebrae detection. Statistical modeling approaches [47] demonstrated effective lumbar spine segmentation using multi-vertebrae anatomical shape models. Modern approaches struggle: [20] achieved MAE of 4.9º and [32] achieved SMAPE of 25.69% because they try to learn geometric relationships that work better when expressed through direct mathematical formulations. Advanced segmentation approaches [48] using iterative fully convolutional neural networks showed promise but faced translation challenges.

B. Classification Systems Analysis

Scoliosis classification analysis shows a key difference that separates this domain from successful medical AI implementations. While successful domains like pathology AI achieve clinical utility through high-performing automated classification, scoliosis classification faces a utility-automation trade-off that prevents clinical implementation. The analysis shows that automation success decreases as clinical utility increases, creating an implementation challenge unique to complex measurement-dependent medical domains. Table IV presents the classification performance characteristics across different system types.

TABLE IV. CLASSIFICATION TRADE-OFF ANALYSIS

Classification Type	Automation Success	Clinical Utility
Non-Standard Binary	High (88.3%) [34]	Limited
Non-Standard Ternary	High (94.69%) [37]	Limited
Standard (Lenke)	Low (72%) [43]	High

This utility-automation trade-off occurs because clinically relevant classification systems like Lenke [44] require multiple geometric and morphological features that are difficult to automate. Automated classification only succeeds when simplified to features that reduce clinical relevance. This differs fundamentally from successful medical AI domains where clinical relevance and automation potential align.

Non-standard approaches show better automation performance through simplified feature sets. Authors in [33] achieved 88.3% accuracy using local centroids evaluation because geometric simplification enables reliable automation, while researchers in [36] achieved 94.69% accuracy through feature-based approaches that sacrifice comprehensive clinical assessment for automation reliability. Support vector approaches [34][37] achieved different levels of success based on feature complexity and clinical requirements. Recent classification developments include [38] analyzing scoliosis from spinal X-ray images, [39] implementing total curvature analysis with artificial neural networks, and [43] developing segmentation network-based Lenke classification systems. However, standard classification systems face dependency on reliable Cobb angle automation, creating cascading failures where measurement errors spread through classification algorithms. Researchers in [40] achieved 92.9% accuracy using Decision trees, and those in [41] achieved kappa values of 0.94 through rule-based approaches, but these successes depend on manual measurement inputs rather than automated systems. Authors in [42] demonstrated novel classification methods using 3D ultrasound imaging, while reliability studies [45] highlight the challenges of automated implementation of established classification systems.

C. Segmentation Technologies and Validation Crisis

Spinal structure segmentation represents the most critical component for understanding validation failures that distinguish scoliosis AI from successful medical AI domains. Analysis of 55 articles demonstrates impressive laboratory achievements but shows real-world performance degradation patterns not observed in successful medical AI implementations. The consistent performance degradation stems from differences between scoliosis imaging characteristics and successful medical AI domains. Scoliosis assessment relies on X-ray imaging that carry variability in positioning, exposure, and anatomical presentation. This differs from the standardized imaging protocols that exist in diabetic retinopathy screening or mammography. Laboratory validation often fails to capture such clinical variability, which suggests that results observed in controlled settings do not reflect actual practice. Table V presents the validation crisis evidence across different segmentation approaches, and it demonstrates degradation patterns that appear unique to this domain.

TABLE V. VALIDATION CRISIS EVIDENCE

Approach	Lab Performance	Real-World Drop
Traditional (ASM)	93.6% [17]	5-8% typical
U-Net Variants	95.8% [21]	10% [50]
Instance Segmentation	96% [51]	7% [52]

Table V shows that performance drop mainly come from three causes. First is Dataset Bias Amplification, where laboratory datasets usually do not include the wide imaging variation that is often found in real clinical settings. Second is Anatomical Anomaly Underrepresentation, where the training data has too few cases of complex deformity and this makes generalization weaker. Third is Workflow Integration Failure, since experimental testing often ignores positioning and the image quality limitation that exist in daily practice. It was found that traditional approaches perform better and maintain more stability. For example, landmark-based segmentation [17] reports 93.6% accuracy, while the generalized Hough transform [46] also shows strong outcome. These methods stay more consistent because geometric constraints can handle imaging variation better. In comparison, statistical [47] and iterative methods [48] produce mixed stability. At some time, they perform better, but other times they become weaker. Advanced deep learning approaches suffer sharper drops. RAR-U-Net [21] with reported 95.8% accuracy degraded strongly under clinical-like testing. LPAQR [49] at 95.28% also degrade since the learned features do not transfer well when imaging condition changes. Newer studies bring in region-based CNNs [50], attention gate dual-pathway networks [51], detection-guided mixed-supervised segmentation [52], and automated vertebrae recognition [53]. Yet the validation issue still remains. This situation stands in clear contrast with other medical AI fields, where real-world results usually match or even surpass laboratory results. It was found that the main difference comes from dataset representation and workflow integration during system development. Whole spine segmentation [54] tries to address this gap, but translation barriers still not fully resolved.

This validation gap stands in sharp contrast with other medical AI fields that usually observe better real-world outcomes, where performance in clinical settings often match or even surpass laboratory benchmarks. It was found that the key difference lies in dataset representation and workflow integration during system development. Whole spine segmentation methods [54] attempt to respond to this challenge, but translation barriers still remain unresolved. Early computer-assisted systems [55] already establish the idea that digital methods can improve measurement consistency, though they still require much manual intervention.

D. Comparative Performance Analysis

Table VI presents comparative advantages of different approaches across scoliosis assessment and successful medical AI domains.

TABLE VI. IMPLEMENTATION PRIORITY MATRIX

Approach	Performance	Clinical Stability	Translation Rate	Key Advantage
Scoliosis - Semi-automated	1.8° ± 0.4° MAD	High	<5%	Geometric precision
Scoliosis - Fully automated	4.2° ± 1.8° MAE	5-10% degradation	<1%	Pattern learning
Manual (Baseline)	5.6°/6.6° variation	100% stable	100%	Human expertise
Diabetic retinopathy AI	94.5% sensitivity	Stable	>80%	Binary classification
Breast cancer AI	High accuracy	Stable	>60%	Augmentation strategy

Semi-automated methods offer clear precision benefit while also achieving clinical stability that looks similar to other successful medical AI applications. Mathematical approaches with geometric constraints [19][22] work better for scoliosis assessment compared to deep learning methods, which mainly excel in pattern recognition tasks. Measurement standardization provides unique advantage that is not available in diagnostic areas requiring subjective clinical judgment. Hybrid augmentation frameworks follow proven medical AI strategies and at the same time meet the specific demand of geometric measurement tasks. This combined approach delivers better reliability by merging accuracy with quality improvement, and it creates benefit that neither automation nor manual methods achieve alone.

E. Summary

These performance differences highlight gap between scoliosis AI and proven medical AI systems. The next section explores why standard methods struggle in this field. It also looks at what this tells us about requirement essential for clinical deployment in complex measurement tasks.

V. CRITICAL ANALYSIS: THE TRANSLATION CHALLENGE

A. Clinical Implementation Barriers for Scoliosis AI

Scoliosis AI struggles in clinical practice for reasons different from other medical AI applications. Understanding these barriers requires examining why standard methods break down in measurement tasks that demand geometric precision instead of only pattern recognition.

1) Primary barrier: *Task-Technology Mismatch:* The core implementation failure stem from mismatch between deep learning optimization goals and the geometric measurement need. This suggest improvement require redirecting effort from optimization aim toward accuracy of measurement that perform better for clinical use. Successful medical AI fields like image classification leverage pattern recognition strength of deep learning. Cobb angle measurement demand geometric accuracy that mathematical methods deliver more consistent.

2) Secondary barrier: *Validation Approach Differences:* Scoliosis AI validation method diverge from proven medical AI practice by focusing on laboratory performance score rather than workflow compatibility. Systems tested only in controlled setting fail to show improved progress for clinical use, while implementation in real hospital environment perform better and observe better alignment with workflow.

3) Tertiary barrier: *Regulatory Framework Gaps:* Other medical AI field perform better because they benefit from clear regulatory pathway for screening and diagnostic support. Scoliosis measurement automation operate without such structure, so approval obstacle limit adoption and prevent system from showing improved progress toward commercial development. Table VII analyze the factor that explain why scoliosis AI fail to observe better clinical integration across system aspect.

TABLE VII. TRANSLATION FAILURE MECHANISMS

Failure Dimension	Primary Cause	Success Comparison
Technical Performance	Task-Technology Mismatch	Radiology (pattern recognition)
Validation Crisis	Laboratory-Clinical Gap	Pathology (workflow integration)
Clinical Utility	Automation-Utility Trade-off	Diagnostic AI (augmentation)

Examining why implementations fail reveals that applying proven medical AI approach in unsuitable areas leads to consistent breakdown. Technical performance issue appears because deep learning performs better in pattern recognition task like diabetic retinopathy detection but struggle in geometric measurement task that require mathematical precision. The validation crisis emerges because controlled laboratory conditions fail to represent imaging variability that exists in clinical scoliosis assessment. This contrasts with standardized imaging protocol used in successful domain, where practice observes better consistency. Successful medical AI implementation shares common characteristics that scoliosis AI lack standardized imaging protocol, binary or categorical output, augmentation instead of replacement strategy, and clear regulatory pathway. The root cause of translation failure comes from pursuing full automation, while clinical adoption would show improved progress if human-AI collaboration performs better in practice.

B. Main Problems and Cross-Domain Comparison

The challenges preventing scoliosis AI from achieving clinical implementation differ fundamentally from the enabling factors in successful medical AI domains, indicating that scoliosis requires domain-specific solutions for clinical adoption.

1) Validation crisis analysis: The validation crisis results from structural differences between scoliosis AI and successful medical AI domains [26] [27]. Performance dropped in multiple studies, with [49] reporting a decline from 95.28% to 85.9%, [51] from 95.4% to 89%, [52] from 94.39% to 87.21%, and [53] from 95.19% to 93.89%. This pattern contrasts with successful medical AI domains where real-world performance typically meets or exceeds laboratory performance. Recent validation studies continue to confirm this pattern, with [56] reporting similar reliability challenges in mobile AI applications, and [57] demonstrating that even advanced CNN approaches maintain this performance gap between laboratory and clinical settings.

2) Understanding validation failures: Laboratory validation employs cross-validation techniques on limited datasets that fail to capture clinical variability. Successful medical AI domains use standardized imaging protocols. Scoliosis assessment involves variable positioning, exposure settings, and anatomical presentations. Laboratory datasets systematically exclude this clinical variability. This creates optimistic performance estimates that fail to predict clinical utility.

3) Cross-domain economic analysis: Economic challenge in scoliosis AI contrast strongly with successful application. Diabetic retinopathy screening show improved progress by delivering clear cost benefit through early detection and fewer specialist visits. In comparison, scoliosis measurement automation face setup cost that outweigh workflow gain. Manual measurement take about 20 minutes [10], yet automated system require heavy infrastructure investment that fail to perform better in workflow advantage.

4) Regulatory framework differences: Regulatory obstacle appear because no suitable framework exist for measurement automation. Other medical AI field perform better because they use existing approval pathway for diagnostic support and screening tool. Scoliosis automation, in contrast, need new regulatory structure that balance accuracy requirement with clinical safety.

C. Economic and Implementation Reality

Economic analysis observes better clarity when comparing scoliosis AI with successful field. The difference lies in cost structure, setup need, and value clarity.

1) Cost-benefit structure analysis: Real-world implementation show automation expense often outweigh saving once development, validation, training, maintenance, and regulatory cost are included. Successful medical AI field show improved progress because they provide clear value benefit. Diabetic retinopathy screening perform better by reducing specialist visits and allowing earlier treatment, while scoliosis automation show vague benefit compared to setup cost.

2) Infrastructure requirements: X-ray images suffer from poor quality and weak contrast, and this limit automated system performance. Alternative imaging such as CT and MR create high expense and radiation risk. Because of these constraint, X-ray remain the preferred choice due to cost, radiation, dataset availability, and maturity. Yet this preference not show improved progress for automation, since it create obstacle that prevent system from performing better.

3) Clinical integration complexity: Integration barriers go beyond technical performance to include workflow compatibility challenges, regulatory compliance demands, and organizational adaptation requirements. Major integration obstacles include workflow disruption needing staff retraining, system reliability standards exceeding current abilities, complicated error management for system breakdowns, and new quality control procedures for combined workflows.

4) Regulatory framework inadequacy: Existing classification systems like Lenke [44] lack development in fully automated applications, creating regulatory challenges where standard validation methods cannot properly evaluate AI systems. Regulatory obstacles include unclear FDA approval routes for AI measurement systems, responsibility assignment issues for automated errors, undefined quality benchmarks for AI medical measurements, and unspecified monitoring requirements for developing automated systems.

D. Summary

It was found that recognizing these barriers can create groundwork for building effective solutions that draw from proven medical AI approaches, and in terms of measurement this becomes very important for geometric tasks. Moreover, the following section will explain some strategic approaches that follow established success model in medical AI, while also trying to answer the particular challenge faced in automated scoliosis assessment.

VI. Discussion

This review of 70 years of scoliosis AI research shows that common belief in medical AI, where better technology directly creates clinical use, does not hold here. In fact, deployment depends more on strategic fit than on complexity alone. It was found that increasing complexity often reduces adoption rate, opposite to what seen in other successful medical AI fields, and this suggests scoliosis AI face structural issues rather than only simple implementation barrier. The difficulty is less about accuracy itself, but more about how the system can fit into clinical workflow.

The core mismatch between geometric measurement requirement and pattern recognition approach highlights a wider habit in medical AI: applying proven technique everywhere without enough task-specific adjustment. This study indicates that problem features, and not only algorithm performance, is what truly shape clinical success. Moreover, the consistent result that semi-automated methods perform better than fully automated systems suggest that human and AI partnership may actually be more effective path forward. In terms of accuracy this is very important, because certain areas need precise measurement rather than just pattern recognition, and this again confirms that the task requirement is more critical than algorithm complexity.

The validation crisis observed across multiple studies indicates systematic methodological inadequacies in how scoliosis AI research approaches real-world applicability. Unlike successful medical AI domains that prioritize clinical workflow integration during development, scoliosis AI research appears to treat clinical implementation as a secondary concern addressed after algorithmic optimization. This finding aligns with broader medical AI translation challenges identified by [58], who demonstrate that healthcare AI failures stem primarily from inadequate attention to clinical implementation factors during development, rather than technical limitations. This sequence reversal may explain why laboratory performance fails to predict clinical utility consistently.

Several limitations affect these findings. The focus on English-language publications may have excluded relevant international developments, particularly given the global nature of scoliosis research. The rapid pace of technological change means some recent advances may not yet demonstrate their clinical translation potential, and the 77-year timeframe may obscure recent improvement trends. Furthermore, viewing outcomes as simply success or failure may oversimplify the complex factors affecting clinical adoption, where partial implementation or hybrid workflows could represent significant progress that traditional metrics miss.

These insights extend beyond scoliosis to other measurement-based medical fields. Areas needing geometric accuracy, continuous measurement, or integration with established mathematical methods may work better with hybrid approaches than complete automation. This indicates the medical AI field could benefit from developing task classifications that systematically match technological methods to problem features instead of assuming successful techniques work universally.

Economic review shows implementation expenses frequently outweigh potential benefits when including system development, validation, training, maintenance, and regulatory costs. This challenges the belief that better technical performance automatically creates strong value propositions. Healthcare organizations seem to make logical economic choices based on realistic cost-benefit evaluations, suggesting that apparent resistance to implementation may reflect sound economic assessment. Recent comprehensive analysis of AI implementation obstacles [59] confirms economic factors as the main barrier to healthcare AI adoption, supporting the logical decision-making approach seen in scoliosis AI.

This review suggests medical AI research should consider clinical implementation factors earlier in algorithm development, rather than treating real-world application as an afterthought. The scoliosis AI experience offers important lessons for emerging medical AI fields about the value of strategic alignment, appropriate technology choices, and realistic validation approaches in achieving genuine clinical impact.

VII. Future Directions and Recommendations

A. Strategic Research Priorities Based on Current Findings

Future research needs to tackle the core challenges found in this study while applying lessons from successful medical AI applications in other fields. The top priority involves moving from complete automation strategies to enhancement approaches that have worked well in radiology and pathology AI systems.

1) Mechanism-Based research priorities: Our analysis shows research priorities must solve the task-technology mismatch by creating hybrid approaches that blend geometric mathematical methods with AI enhancement instead of full replacement. This matches successful medical AI patterns where technology supports rather than replaces human expertise. Recent progress in automated Cobb angle measurement [60], [61] shows that hardware-independent approaches can reach clinical accuracy when designed as enhancement rather than replacement systems.

2) Learning from successful medical AI domains: Diabetic retinopathy screening succeeded through standardized procedures, binary decision outputs, and clear regulatory routes. Breast cancer detection AI gained adoption through enhancement strategies that improved rather than replaced radiologist skills. Scoliosis AI must follow similar enhancement approaches, emphasizing measurement quality improvement and error reduction instead of complete automation. Table VIII shows the strategic implementation plan based on our analysis and successful medical AI patterns.

TABLE VIII. Implementation Priority Matrix

Priority Level	Strategy	Success Model
Immediate	Hybrid measurement systems	Radiology AI augmentation
	Standardized validation	Pathology AI validation
Medium-term	Regulatory pathways	FDA-approved frameworks
	Cost-effectiveness demo	Screening AI ROI models
Long-term	Complete spine analysis	Multi-modal integration

3) Immediate actions (1-2 years)

a) Hybrid system development: Research efforts should focus on hybrid system creation that combines traditional geometric methods for reliable measurement with AI enhancement for quality improvement and error detection. This strategy follows successful radiology AI applications where AI supports rather than replaces radiologist skills. Current spine segmentation approaches [62], [63] offer technical foundations for hybrid systems that blend deep learning capabilities with geometric constraints, enabling the enhancement strategy proven successful in other medical AI fields.

b) Validation protocol standardization: Validation protocol standardization represents another key priority. Organizations should create multi-site validation procedures that include real-world imaging variation, following successful pathology AI validation approaches that test systems within actual clinical workflows instead of controlled laboratory settings.

c) Workflow integration research: Implementation research must tackle clinical workflow integration from the start, following the diabetic retinopathy screening model where clinical workflow needs guided technical development rather than being handled as an afterthought.

4) Medium-term objectives (3-5 years)

a) Regulatory pathway development: Work with regulatory agencies to create suitable frameworks for measurement enhancement systems, learning from established FDA routes for diagnostic support AI while addressing the specific needs of measurement improvement applications.

b) Economic value demonstration: Conduct thorough cost-effectiveness studies that show clear ROI for hybrid systems, following successful screening AI models that offer clear value benefits through improved accuracy and workflow efficiency.

c) Clinical evidence generation: Create clinical evidence through pilot implementations that show improved measurement reliability and workflow efficiency, following the evidence development strategies used in successful medical AI fields.

B. Technology Selection Framework Based on Medical AI Success Patterns

Healthcare institutions considering automated system implementation can use our analysis for evidence-based guidance drawn from successful medical AI patterns rather than theoretical performance data.

1) Technology recommendation framework: Based on comprehensive review of 55 studies and comparison with successful medical AI fields, current evidence strongly favors hybrid enhancement approaches over complete automation systems for clinical use.

2) Learning from successful implementation patterns: Successful medical AI fields achieve clinical value through enhancement strategies that improve human capabilities while keeping human oversight and decision control. Scoliosis AI should follow similar approaches, emphasizing measurement quality improvement, error detection, and workflow support instead of pursuing complete measurement automation.

3) Clinical decision support framework: Healthcare institutions should consider enhancement systems for high-volume screening programs needing standardized procedures, quality assurance applications for manual measurement validation, educational environments for training standardization, and research studies requiring measurement standardization and error reduction.

4) Implementation readiness assessment: Healthcare institutions should assess implementation readiness using criteria from successful medical AI deployments. Key evaluation areas include technical infrastructure capable of supporting hybrid workflows, staff training capacity for new enhancement tools, quality assurance procedures adapted for human-AI collaboration, and organizational commitment to workflow adaptation and continuous improvement.

C. Policy and Regulatory Development Based on Medical AI Success

Policy and regulatory development for measurement enhancement system should take lessons from successful medical AI but still adapt to its own technical requirement.

1) Regulatory framework development: Diagnostic AI validation method don't fully work here. New framework should cover training data diversity, algorithm transparency, and performance monitoring. Post-market evaluation must be continuous. Liability keep human in decision loop but also address system error. Benchmarks set minimum measurable accuracy improvement.

2) Funding strategy realignment: Funding should shift to implementation-focused research that show clinical value. Priority include: (i) clinical integration studies from AI success stories, (ii) validation methods for real-world prediction, (iii) hybrid human–AI systems that combine skill and enhancement, and (iv) cross-disciplinary collaboration connecting technical, clinical, and implementation expertise.

3) Technology integration: Integration should follow proven models: multimodal imaging (X-ray, MRI, surface topography) from radiology AI, explainable decision support from pathology AI, and federated learning for multi-site training with privacy protection.

4) Economic model development: Economic structure should follow medical AI demonstration: ROI based on workflow improvement and error reduction, cost-effectiveness beyond setup, and reimbursement aligned with medical AI reimbursement examples.

D. Summary

This analysis, drawn from successful medical AI, points a way forward. By shifting to enhancement, aligning funding, and following proven patterns, scoliosis AI adoption can show improved progress. The recommendation is simple but important: success will come when technology helps the human expert, not when it tries to replace.

VIII. Conclusions

It was found from 55 studies that a paradox appears. Performance has kept improving for 70 years, but clinical usage keeps going down. Semi-automated method performs better than fully automated, with accuracy 1.8° ± 0.4° MAD against 4.2° ± 1.8° MAE. This suggests the opposite of what medical AI normally shows, where more complexity usually performs better. Moreover, moving from lab to clinic cause 5–10% accuracy drop, and clinical use still below 5%. This shows that technical gain does not equal real usage. The main cause is task-technology mismatch. Cobb angles need geometric accuracy, but AI chase full automation instead of enhancement. Other fields already show improved progress with enhancement, but scoliosis AI push too far. This reduces trust, reduces confidence, and reduces adoption. Thus, it can be concluded that success needs shift to enhancement. Hybrid system with mathematical accuracy plus AI support perform better and give clearer value. Institutions should focus on measurement quality, error detection, and workflow support with human control. The message is simple but repeat again: success will not come from replacing the human, it will come from helping the human. In terms of practical steps, implementation needs coordinated action across three phases. The immediate priority is to create hybrid enhancement system and to establish validation procedures that include real-world workflow integration. The medium-term goal is to set regulatory route for measurement enhancement system and to provide cost-effectiveness evidence through pilot implementation. The long-term aim is to integrate multi-modal approach and comprehensive spine analysis capability. Finally, success depends on testing system inside actual clinical workflow rather than only laboratory setting, following the validation pattern already established in successful medical AI. This analysis has several important limitations that should be recognized. The focus on English-language publications may have missed relevant international developments, particularly given the global nature of scoliosis research. Publication bias may favor positive results, potentially overestimating automated system performance compared to clinical reality. The rapid pace of technology change means some recent advances may not yet show their clinical implementation potential, while the 77-year study timeframe may hide recent improvement trends. Additionally, limited availability of real-world clinical validation data restricted the scope of implementation analysis, and the binary view of implementation success versus failure may oversimplify the complex factors affecting clinical adoption. The field must abandon complete automation approaches in favor of augmentation strategies that have proven successful across medical AI domains. Researchers should focus on hybrid system development, clinicians should advocate for workflow-integrated solutions, industry should invest in augmentation rather than replacement technologies, and regulators should develop appropriate frameworks for measurement enhancement systems. Only through coordinated stakeholder efforts that incorporate lessons from successful medical AI implementations while addressing the unique challenges of geometric measurement tasks can the field overcome this translation crisis and deliver meaningful benefits for patient care.

Acknowledgment

The authors gratefully acknowledge all researchers whose work was reviewed in this analysis and thank Penerbit Universiti Malaysia Sabah (UMS Press) for funding.

References

[1] A. B. Ferguson, "The study and treatment of scoliosis," *South Med J*, vol. 23, no. 2, pp. 116-120, 1930.

[2] A. Greenspan, J. W. Pugh, A. Norman, and R. S. Norman, "Scoliotic index: a comparative evaluation of methods for the measurement of scoliosis," *Bulletin of the Hospital for Joint Diseases*, vol. 39, no. 2, pp. 117-125, 1978.

[3] K. M. Diab, J. A. Sevastik, R. Hedlund, and I. A. Suliman, "Accuracy and applicability of measurement of the scoliotic angle at the frontal plane by Cobb's method, by Ferguson's method and by a new method," *European Spine Journal*, vol. 4, pp. 291-295, 1995.

[4] Y. Chen, W. Chen, and W. Chiou, "An alternative method for measuring scoliosis curvature," *Orthopedics-New Jersey*, vol. 30, no. 10, p. 828, 2007.

[5] B. Drerup and E. Hierholzer, "Evaluation of frontal radiographs of scoliotic spines---Part I measurement of position and orientation of vertebrae and assessment of clinical shape parameters," *Journal of Biomechanics*, vol. 25, no. 11, pp. 1357-1362, 1992.

[6] R. T. Morrissy, G. S. Goldsmith, E. C. Hall, D. Kehl, and G. H. Cowie, "Measurement of the Cobb angle on radiographs of patients who have scoliosis. Evaluation of intrinsic error," *JBJS*, vol. 72, no. 3, pp. 320-327, 1990.

[7] C. E. Beekman and V. Hall, "Variability of scoliosis measurement from spinal roentgenograms," *Physical Therapy*, vol. 59, no. 6, pp. 764-765, 1979.

[8] J. E. H. Pruijs, M. A. P. E. Hageman, W. Keessen, R. Van Der Meer, and J. C. Van Wieringen, "Variation in Cobb angle measurements in scoliosis," *Skeletal Radiology*, vol. 23, pp. 517-520, 1994.

[9] D. L. Carman, R. H. Browne, and J. G. Birch, "Measurement of scoliosis and kyphosis radiographs. Intraobserver and interobserver variation," *JBJS*, vol. 72, no. 3, pp. 328-333, 1990.

[10] R. Kundu, A. Chakrabarti, and P. K. Lenka, "Cobb angle measurement of scoliosis with reduced variability," *arXiv preprint arXiv:1211.5355*, 2012.

[11] K. G. Shea, P. M. Stevens, M. Nelson, J. T. Smith, K. S. Masters, and S. Yandow, "A comparison of manual versus computer-assisted radiographic measurement: intraobserver measurement variability for Cobb angles," *Spine*, vol. 23, no. 5, pp. 551-555, 1998.

[12] M. C. Tanure, A. P. Pinheiro, and A. S. Oliveira, "Reliability assessment of Cobb angle measurements using manual and digital methods," *The Spine Journal*, vol. 10, no. 9, pp. 769-774, 2010.

[13] J. Zhang, H. Li, L. Lv, and Y. Zhang, "Computer-aided cobb measurement based on automatic detection of vertebral slopes using deep neural network," *International Journal of Biomedical Imaging*, vol. 2017, no. 1, p. 9083916, 2017.

[14] R. Choi, K. Watanabe, H. Jinguji, N. Fujita, Y. Ogura, S. Demura, T. Kotani, K. Wada, M. Miyazaki, H. Shigematsu, and Y. Aoi, "CNN-based spine and Cobb angle estimator using moire images," *IIEEJ Transactions on Image Electronics and Visual Computing*, vol. 5, no. 2, pp. 135-144, 2017.

[15] M. H. Horng, C. P. Kuok, M. J. Fu, C. J. Lin, and Y. N. Sun, "Cobb angle measurement of spine from X-ray images using convolutional neural network," *Computational and Mathematical Methods in Medicine*, vol. 2019, p. 6357171, 2019.

[16] Y. J. Kim, B. Ganbold, and K. G. Kim, "Web-based spine segmentation using deep learning in computed tomography images," *Healthcare Informatics Research*, vol. 26, no. 1, pp. 61-67, 2020.

[17] B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "Shape representation for efficient landmark-based segmentation in 3-D," *IEEE Transactions on Medical Imaging*, vol. 33, no. 4, pp. 861-874, 2014.

[18] K. Singer, T. J. Jones, and P. D. Breidahl, "A comparison of radiographic and computer-assisted measurements of thoracic and

thoracolumbar sagittal curvature," *Skeletal Radiology*, vol. 19, pp. 21-26, 1990.

[19] J. Zhang, E. Lou, L. H. Le, D. L. Hill, J. V. Raso, and Y. Wang, "Automatic Cobb measurement of scoliosis based on fuzzy Hough Transform with vertebral shape prior," *Journal of Digital Imaging*, vol. 22, pp. 463-472, 2009.

[20] B. Chen, Q. Xu, L. Wang, S. Leung, J. Chung, and S. Li, "An automated and accurate spine curve analysis system," *IEEE Access*, vol. 7, pp. 124596-124605, 2019.

[21] Z. Wang, Z. Zhang, and I. Voiculescu, "RAR-U-Net: a residual encoder to attention decoder by residual connections framework for spine segmentation under noisy labels," in *2021 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2021, pp. 21-25.

[22] J. H. Han, L. Kóczy, and T. Poston, "Fuzzy hough transform," *Pattern Recognition Letters*, vol. 15, no. 7, pp. 649-658, 1994.

[23] S. Allen, E. Parent, M. Khorasani, D. L. Hill, E. Lou, and J. V. Raso, "Validity and reliability of active shape models for the estimation of Cobb angle in patients with adolescent idiopathic scoliosis," *Journal of Digital Imaging*, vol. 21, pp. 208-218, 2008.

[24] R. Korez, B. Ibragimov, B. Likar, F. Pernuš, and T. Vrtovec, "A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation," *IEEE Transactions on Medical Imaging*, vol. 34, no. 8, pp. 1649-1662, 2015.

[25] B. Glocker, J. Feulner, A. Criminisi, D. R. Haynor, and E. Konukoglu, "Automatic localization and identification of vertebrae in arbitrary field-of-view CT scans," in *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part III 15*, Springer Berlin Heidelberg, 2012, pp. 590-598.

[26] S. R. Eddy, "Hidden markov models," *Current Opinion in Structural Biology*, vol. 6, no. 3, pp. 361-365, 1996.

[27] B. Schuster-Böckler and A. Bateman, "An introduction to hidden Markov models," *Current Protocols in Bioinformatics*, vol. 18, no. 1, pp. A-3A, 2007.

[28] S. Kónya, T. S. Natarajan, H. Allouch, K. A. Nahleh, O. Y. Dogheim, and H. Boehm, "Convolutional neural network-based automated segmentation and labeling of the lumbar spine X-ray," *Journal of Craniovertebral Junction and Spine*, vol. 12, no. 2, pp. 136-143, 2021.

[29] B. Schlager, M. Krump, J. Boettinger, J. Koenigschulte, S. Doyle, J. Mayr, and R. Wegenkittl, "Deep learning methods for automatic vertebrae identification and morphometric analysis of the spine in MR images," *Computer Methods and Programs in Biomedicine*, vol. 208, p. 106230, 2021.

[30] H. Wu, C. Bailey, P. Rasoulinejad, and S. Li, "Automatic landmark estimation for adolescent idiopathic scoliosis assessment using BoostNet," in *Medical Image Computing and Computer Assisted Intervention− MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20*, Springer International Publishing, 2017, pp. 127-135.

[31] H. Sun, X. Zhen, C. Bailey, P. Rasoulinejad, Y. Yin, and S. Li, "Direct estimation of spinal cobb angles by structured multi-output regression," in *International Conference on Information Processing in Medical Imaging*, Springer International Publishing, Cham, 2017, pp. 529-540.

[32] B. Khanal, L. Dahal, P. Adhikari, and B. Khanal, "Automatic cobb angle detection using vertebra detector and vertebra corners regression," in *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, Springer International Publishing, Cham, 2019, pp. 81-87.

[33] H. S. Kim, S. Ishikawa, Y. Ohtsuka, H. Shimizu, T. Shinomiya, and M. A. Viergever, "Automatic scoliosis detection based on local centroids evaluation on moire topographic images of human backs," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1314-1320, 2001.

[34] P. O. Ajemba, L. Ramirez, N. G. Durdle, D. L. Hill, and V. J. Raso, "A support vectors classifier approach to predicting the risk of progression of adolescent idiopathic scoliosis," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, pp. 276-282, 2005.

[35] S. Rothstock, H. R. Weiss, D. Krueger, and L. Paul, "Clinical classification of scoliosis patients using machine learning and markerless 3D surface trunk data," *Medical & Biological Engineering & Computing*, vol. 58, no. 12, pp. 2953-2962, 2020.

[36] J. Fatima, M. Mohsan, A. Jameel, M. U. Akram, and A. Muzaffar Syed, "Vertebrae localization and spine segmentation on radiographic images for feature-based curvature classification for scoliosis," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 26, p. e7300, 2022.

[37] L. Ramirez, N. G. Durdle, V. J. Raso, and D. L. Hill, "A support vector machines classifier to assess the severity of idiopathic scoliosis from surface topography," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 84-91, 2006.

[38] A. A. Z. Imran, C. Huang, H. Tang, W. Fan, K. Cheung, M. To, Z. Qian, and D. Terzopoulos, "Analysis of scoliosis from spinal x-ray images," *arXiv preprint arXiv:2004.06887*, 2020.

[39] H. Lin, "Identification of spinal deformity classification with total curvature analysis and artificial neural network," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 376-382, 2007.

[40] P. Phan, N. Mezghani, M. L. Nault, C. É. Aubin, S. Parent, J. de Guise, and H. Labelle, "A decision tree can increase accuracy when assessing curve types according to Lenke classification of adolescent idiopathic scoliosis," *Spine*, vol. 35, no. 10, pp. 1054-1059, 2010.

[41] J. Zhang, H. Li, L. Lv, X. Shi, F. Guo, and Y. Zhang, "A Computer-aided Method for Improving the Reliability of Lenke Classification for Scoliosis," *Journal of Healthcare Engineering*, vol. 6, no. 2, pp. 145-158, 2015.

[42] D. Yang, T. T. Y. Lee, K. K. L. Lai, Y. S. Wong, L. N. Wong, J. L. Yang, T. P. Lam, R. M. Castelein, J. C. Y. Cheng, and Y. P. Zheng, "A novel classification method for mild adolescent idiopathic scoliosis using 3D ultrasound imaging," *Medicine in Novel Technology and Devices*, vol. 11, p. 100075, 2021.

[43] D. Liu, L. Zhang, J. Yang, and A. Lin, "Lenke classification of scoliosis based on segmentation network and adaptive shape descriptor," *Applied Sciences*, vol. 13, no. 6, p. 3905, 2023.

[44] L. G. Lenke, "A Decision Tree Can Increase Accuracy When Assessing Curve Types According to Lenke Classification of Adolescent Idiopathic Scoliosis Point of View," *Spine*, vol. 35, no. 10, pp. 1060-1060, 2010.

[45] M. Ogon, K. Giesinger, H. Behensky, C. Wimmer, M. Nogler, C. M. Bach, and M. Krismer, "Interobserver and intraobserver reliability of Lenke's new scoliosis classification system," *Spine*, vol. 27, no. 8, pp. 858-862, 2002.

[46] M. A. Larhmam, S. Mahmoudi, and M. Benjelloun, "Semi-automatic detection of cervical vertebrae in X-ray images using generalized Hough transform," in *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2012, pp. 396-401.

[47] A. Rasoulian, R. Rohling, and P. Abolmaesumi, "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+ pose model," *IEEE Transactions on Medical Imaging*, vol. 32, no. 10, pp. 1890-1900, 2013.

[48] N. Lessmann, B. Van Ginneken, P. A. De Jong, and I. Išgum, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification," *Medical Image Analysis*, vol. 53, pp. 142-155, 2019.

[49] L. Zhang, L. Shi, J. C. Y. Cheng, W. C. W. Chu, and S. C. H. Yu, "LPAQR-Net: efficient vertebra segmentation from biplanar whole-spine radiographs," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2710-2721, 2021.

[50] L. Zhang, J. Zhang, and S. Gao, "Region-Based Convolutional Neural Network-Based Spine Model Positioning of X-Ray Images," *BioMed Research International*, vol. 2022, no. 1, p. 7512445, 2022.

[51] W. Shi, T. Xu, H. Yang, Y. Xi, Y. Du, J. Li, and J. Li, "Attention gate based dual-pathway network for vertebra segmentation of X-ray spine images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 8, pp. 3976-3987, 2022.

[52] S. Pang, C. Pang, Z. Su, L. Lin, L. Zhao, Y. Chen, Y. Zhou, H. Lu, and Q. Feng, "DGMSNet: Spine segmentation for MR image by a detection-guided mixed-supervised segmentation network," *Medical Image Analysis*, vol. 75, p. 102261, 2022.

[53] M. U. Saeed, N. Dikaios, A. Dastgir, G. Ali, M. Hamid, and F. Hajjej, "An automated deep learning approach for spine segmentation and

vertebrae recognition using computed tomography images," *Diagnostics*, vol. 13, no. 16, p. 2658, 2023.

[54] R. Da Mutten, O. Zanier, S. Theiler, S. J. Ryu, L. Regli, C. Serra, and V. E. Staartjes, "Whole Spine Segmentation Using Object Detection and Semantic Segmentation," *Neurospine*, vol. 21, no. 1, p. 57, 2024.

[55] N. Chockalingam, P. H. Dangerfield, G. Giakas, T. Cochrane, and J. C. Dorgan, "Computer-assisted Cobb measurement of scoliosis," *European Spine Journal*, vol. 11, pp. 353-357, 2002.

[56] H. Li, C. Qian, W. Yan, D. Fu, Y. Zheng, Z. Zhang, J. Meng, and D. Wang, "Use of Artificial Intelligence in Cobb Angle Measurement for Scoliosis: Retrospective Reliability and Accuracy Study of a Mobile App," Journal of Medical Internet Research, vol. 26, p. e50631, 2024.

[57] Y. Maeda, T. Nagura, M. Nakamura, et al., "Automatic measurement of the Cobb angle for adolescent idiopathic scoliosis using convolutional neural network," Scientific Reports, vol. 13, p. 14576, 2023.

[58] V. K. Bürger, J. Amann, C. K. T. Bui, J. Fehr, and V. I. Madai, "The unmet promise of trustworthy AI in healthcare: why we fail at clinical translation," Frontiers in Digital Health, vol. 6, p. 1279629, 2024.

[59] E. Ahmed, R. Ward, S. Delves, B. Spooner, J. Isherwood, A. Dhami, S. Poyser, and J. Ghosh, "A Systematic Review of the Barriers to the Implementation of Artificial Intelligence in Healthcare," Cureus, vol. 15, no. 10, p. e46454, 2023.

[60] A. Suri, S. Tang, D. Kargilis, E. Taratuta, B. J. Kneeland, G. Choi, A. Agarwal, N. Anabaraonye, W. Xu, J. B. Parente, A. Terry, A. Kalluri, K. Song, and C. S. Rajapakse, "Conquering the Cobb Angle: A Deep Learning Algorithm for Automated, Hardware-Invariant Measurement of Cobb Angle on Radiographs in Patients with Scoliosis," Radiology: Artificial Intelligence, vol. 5, no. 4, p. e220158, 2023.

[61] W. Caesarendra, W. Rahmaniar, J. Mathew, and A. Thien, "Automated Cobb Angle Measurement for Adolescent Idiopathic Scoliosis Using Convolutional Neural Network," Diagnostics, vol. 12, no. 2, p. 396, 2022.

[62] M. U. Saeed, N. Dikaios, A. Dastgir, G. Ali, M. Hamid, and F. Hajjej, "An automated deep learning approach for spine segmentation and vertebrae recognition using computed tomography images," Diagnostics, vol. 13, no. 16, p. 2658, 2023.

[63] H. Lu, M. Li, K. Yu, Y. Zhang, and L. Yu, "Lumbar spine segmentation method based on deep learning," Journal of Applied Clinical Medical Physics, vol. 24, p. e13996, 2023.