BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 14, 2026
Research Article

Limitations of Parameter-Based AI Models in the MRI Evaluation of Lumbar Spinal Stenosis: A Clinical Perspective

Bilal Ertugrul[1]

***Abstract*— Lumbar spinal stenosis is a common degenerative spinal disorder in neurosurgical practice, and magnetic resonance imaging plays a central role in diagnosis and surgical planning. In recent years, artificial intelligence, particularly deep learning–based models, has shown promising results in automated magnetic resonance imaging analysis, including stenosis grading and segmentation. However, most existing systems rely heavily on predefined quantitative parameters, raising concerns about their clinical applicability. This study critically evaluates the limitations of parameter-based artificial intelligence models in the magnetic resonance imaging assessment of lumbar spinal stenosis and proposes a conceptual framework for clinically aligned, explainable hybrid artificial intelligence approaches. A structured literature-based analysis was conducted to examine the major limitations of parameter-based artificial intelligence models and to explore future directions for clinically aligned, explainable hybrid systems. The limitations were assessed in terms of standardization, clinical decision-making complexity, clinicoradiological mismatch, multilevel disease evaluation, imposed parameterization, and labeling variability. Parameter-based artificial intelligence models show important limitations in representing real-world clinical decision-making. The absence of standardized thresholds, inability to integrate multidimensional clinical information, and difficulty identifying symptomatic levels in multilevel disease reduce their clinical usefulness. In addition, interobserver variability in labeling introduces noise that negatively affects model performance and generalizability. Although parameter-based artificial intelligence models contribute to the quantitative evaluation of lumbar spinal stenosis, they remain insufficient for comprehensive clinical decision support. Explainable hybrid artificial intelligence approaches integrating imaging findings with clinical information and structured knowledge may provide a more reliable and clinically meaningful framework for future applications.**

Index Terms— **Lumbar spinal stenosis, artificial intelligence, MRI, deep learning, hybrid models, explainable AI.**

I. Introduction

Lumbar spinal stenosis (LSS) is a common degenerative spinal disorder that is particularly prevalent in the elderly population. It is a significant aspect of neurosurgical practice (1, 2). Magnetic resonance imaging (MRI) plays a central role in diagnosis and surgical planning. Recently, artificial intelligence (AI), particularly deep learning–based approaches, has received considerable attention for its use in the automated analysis of spinal MRIs, including stenosis grading and segmentation tasks (3–5).

However, many current AI models are developed based on predefined quantitative parameters, and their decision-making processes rely heavily on these measurable variables. Parameters such as the anteroposterior canal diameter, cross-sectional area, and ligamentum flavum thickness are frequently used as model inputs. While these metrics provide objective information, they represent only a limited component of the clinical decision-making process. They may fail to adequately capture factors such as symptomatology, neurological examination findings, and the complex interplay of multilevel pathologies. The key MRI parameters commonly used in the evaluation of lumbar spinal stenosis are illustrated in Figure 1.

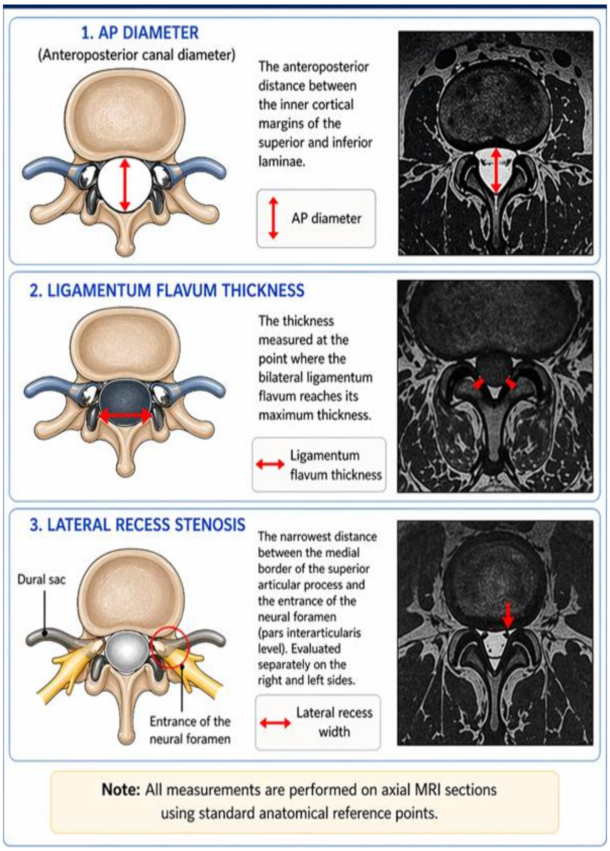

Fig. 1. Key MRI parameters used in the evaluation of lumbar spinal stenosis.

[1] **Bilal Ertugrul**, is with the Department of Neurosurgery, Faculty of Medicine, Fırat University, Elazig, Türkiye (e-mail: bilalertugruldr@gmail.com).
https://orcid.org/0000-0001-7812-3332

Manuscript received Apr 20, 2026; accepted Jul. 28, 2026. DOI: 10.17694/bajece.1933652

Ertugrul, B. (2026). Limitations of Parameter-Based AI Models in the MRI Evaluation of Lumbar Spinal Stenosis: A Clinical Perspective. Balkan Journal of Electrical and Computer Engineering, 14(1), 142-147.

Accordingly, significant advancements have been made in dataset development and modeling strategies. For example, Abdulmahmod et al.'s framework provides a standardized data infrastructure. However, a recent systematic review demonstrated that most existing models rely heavily on quantitative parameters (6, 7). Nevertheless, approaches that incorporate multimodal data integration and clinical knowledge representation are emerging. In this context, the multimodal computer-aided diagnosis (CAD) approach proposed by Salem et al. and the knowledge graph–based model developed by Raza et al. offer more comprehensive modeling strategies that extend beyond imaging data alone (8, 9).

Despite these advances, many existing systems remain insufficiently aligned with real-world clinical decision-making processes and continue to exhibit limitations in terms of explainability. Consequently, there is an increasing need for explainable hybrid AI approaches that integrate imaging findings, clinical information, and structured knowledge representations (10,11). Accordingly, the present study aims to systematically analyze the limitations of parameter-based AI models in the MRI evaluation of lumbar spinal stenosis and to provide a clinically oriented framework for the development of explainable hybrid artificial intelligence systems. By identifying the major sources of discrepancy between quantitative AI outputs and real-world clinical decision-making, this study seeks to highlight future directions for clinically applicable, interpretable, and trustworthy AI-assisted diagnostic models.

II. METHODOLOGY

A structured literature-based analysis was conducted to investigate the limitations of parameter-based artificial intelligence models in the MRI evaluation of lumbar spinal stenosis. The study focused on identifying recurring challenges and clinically relevant factors that may influence the applicability of AI systems in real-world decision-making. In addition, recent developments in explainable and multimodal artificial intelligence were reviewed to construct a clinically oriented framework for future AI-assisted diagnostic systems.

III. MATERIALS AND METHODS

3.1 Search Strategy

Relevant studies were identified through searches of PubMed, Web of Science, Scopus, and Google Scholar databases. The search focused on studies published in English between January 2015 and March 2026.

3.2 Search Terms

The search strategy was based on combinations of the following keywords: “lumbar spinal stenosis”, “artificial intelligence”, “deep learning”, “magnetic resonance imaging”, “MRI”, “segmentation”, “classification”, “explainable artificial intelligence”, “hybrid AI”, “multimodal AI”, “knowledge graph”, and “large language models”.

3.3 Literature Selection

Studies focusing on MRI-based artificial intelligence applications in lumbar spinal stenosis were preferentially considered, including those addressing segmentation, classification, prediction, explainable AI approaches, multimodal frameworks, and clinically oriented decision-support systems. In addition, studies discussing the limitations, interpretability, and clinical applicability of AI models were reviewed. Publications unrelated to lumbar spinal stenosis or MRI-based artificial intelligence were not considered.

3.4 Data Extraction and Thematic Analysis

Information obtained from the reviewed studies was qualitatively synthesized to evaluate the major challenges associated with parameter-based artificial intelligence models. Particular attention was given to issues related to standardization, clinicoradiological mismatch, multilevel disease assessment, parameter dependency, labeling variability, and explainability. These aspects were comparatively evaluated with respect to their potential impact on real-world clinical decision-making.

3.5 Framework Development

Based on current developments in multimodal and explainable artificial intelligence, a clinically oriented framework was developed to illustrate potential strategies for integrating imaging findings, clinical variables, and structured knowledge representations into future AI-assisted diagnostic systems.

This section focuses on the practical aspects of the research, such as experimental setups, hardware or software used, data acquisition, simulation parameters, and performance metrics. Describe all materials, tools, or datasets used so that the work can be replicated. Ensure figures and diagrams related to the experimental design are properly referenced.

IV. RESULTS

Theme 1. Lack of Standardized Parameters

Quantitative parameters such as the anteroposterior (AP) canal diameter, dural sac cross-sectional area, lateral recess depth, foraminal area, and ligamentum flavum thickness are commonly used in the evaluation of lumbar spinal stenosis and constitute the primary inputs for many artificial intelligence (AI) models (5). However, universally accepted and clinically validated threshold values for these parameters have not yet been established. Patients with similar quantitative measurements may exhibit markedly different clinical manifestations, reducing the reliability of purely measurement-based classifications. Consequently, parameter-based models may not fully reflect clinical reality and may provide limited support for real-world decision-making.

Theme 2. Multidimensional Nature of Clinical Decision-Making

Clinical decision-making in lumbar spinal stenosis is inherently complex and multidimensional. Surgical indications are determined not only by radiological findings but also by pain severity, neurogenic claudication distance, neurological deficits, quality of life, and associated comorbidities. Therefore, AI models relying exclusively on imaging-based parameters may not adequately represent the complexity of real-world clinical decision-making processes.

Theme 3. Clinicoradiological Mismatch

Clinicoradiological mismatch represents another important challenge in lumbar spinal stenosis. While some patients may demonstrate severe radiological stenosis with minimal symptoms, others may present with substantial clinical complaints despite only mild imaging abnormalities (12). Consequently, AI systems based solely on imaging findings may fail to generate clinically meaningful outputs.

Theme 4. Challenges in Multilevel Disease

Lumbar spinal stenosis frequently affects multiple spinal levels. Although current AI systems are capable of detecting stenosis at different levels independently, determining the clinically dominant symptomatic level remains difficult. This limitation is particularly important in surgical planning and represents a major obstacle to the clinical implementation of parameter-based AI systems.

Theme 5. Parameter Dependency

Another important limitation of parameter-based AI systems is their dependence on predefined quantitative variables. Some of these variables are not routinely used in everyday clinical practice and may not adequately reflect clinical reasoning. Consequently, forced parameterization may reduce the clinical relevance of these systems.

Theme 6. Labeling Variability

The performance of artificial intelligence models depends heavily on the quality and consistency of labeled datasets. In lumbar spinal stenosis, expert-dependent grading introduces label noise and interobserver variability, negatively affecting model performance and generalizability (13). This issue remains one of the major challenges limiting the reliability and clinical applicability of current AI-based approaches.

V. DISCUSSION

In recent years, artificial intelligence (AI) technologies have made significant advances in the evaluation of lumbar spinal stenosis (LSS), particularly in the automated analysis of magnetic resonance imaging (MRI) data and stenosis grading (4, 14). However, despite the promising performance metrics reported in the literature, current models remain limited in their integration into clinical decision-making processes. This discrepancy highlights a fundamental gap between model performance and clinical applicability (Figure 2).

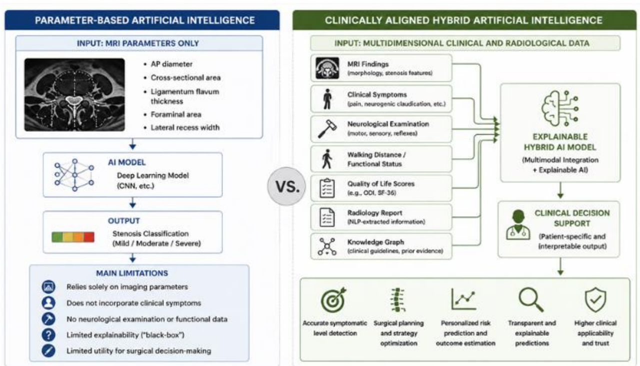

Fig. 2. Comparison of parameter-based artificial intelligence and clinically aligned hybrid artificial intelligence for the evaluation of lumbar spinal stenosis.

A closer examination of the literature reveals substantial progress in segmentation and stenosis detection through large datasets and deep learning–based approaches (4, 15). In particular, the development of standardised datasets has played a crucial role in improving model performance. Nevertheless, as emphasised in systematic reviews, the majority of these models rely heavily on quantitative parameters such as the anteroposterior canal diameter, dural sac area and ligamentum flavum thickness. While this parameter-driven approach may enhance model accuracy, it does not adequately capture the complexity of clinical decision-making processes.

Table I. Summarizes representative artificial intelligence approaches currently used for MRI evaluation of lumbar spinal stenosis

Study	AI approach	Input data	Main output	Principal limitation
Abdulmahmod et al.	Deep learning segmentation	Lumbar MRI	Segmentation of spinal structures	Imaging-only analysis
Al-Antari et al.	Predictive deep learning	MRI parameters	Stenosis classification	Dependence on predefined parameters
Salem et al.	Multimodal CAD	MRI + clinical information	Computer-aided diagnosis	Limited external validation
Raza et al.	Knowledge graph + LLM	MRI reports + structured knowledge	Clinical decision support	Early-stage clinical implementation
Conventional AI models	Parameter-based AI	Quantitative MRI measurements	Severity grading	Limited clinical integration
Proposed framework (Present study)	Explainable hybrid AI	MRI + clinical + functional + knowledge-based data	Clinically meaningful decision support	Requires prospective validation

The findings of this study suggest that parameter-based models are limited by their inability to adequately represent the multidimensional nature of clinical decision-making. In clinical practice, decisions are not based solely on radiological findings, but also on multiple factors such as patient-reported symptoms, functional capacity and neurological status (2). This inherently complex, multifactorial process cannot be fully captured by imaging-based analyses alone.

Furthermore, the discrepancy between radiological findings and clinical symptoms is one of the most significant challenges in evaluating LSS. As emphasised frequently in the literature, patients with severe radiological stenosis may be asymptomatic, whereas mild imaging findings may be associated with substantial clinical impairment (17). As imaging-based AI models cannot directly account for this discrepancy, they may fail to generate reliable and meaningful predictions.

The presence of multilevel disease is another important limitation of current models. While deep learning algorithms can accurately detect stenosis at multiple levels, they cannot identify the level that is clinically dominant and responsible for the patient’s symptoms. This is a particularly critical limitation in the context of surgical planning, significantly restricting the clinical applicability of existing models.

Another major issue is the 'parameter dependency' that arises during the development of the model. Artificial intelligence systems tend to learn from the measurable variables presented in the training data. However, clinical decision-making is not always confined to quantifiable parameters. Consequently, models may overlook clinically relevant factors that are difficult to express numerically. This creates a substantial gap between the representations learned by the model and real-world clinical reasoning.

Interobserver variability in the labelling process is another well-documented challenge. Variations in stenosis grading among radiologists introduce 'label noise' into datasets, thereby negatively affecting model performance (5). This issue is further exacerbated in multicentre datasets, where it can limit model generalisability even more.

When considered collectively, these findings clearly indicate that future artificial intelligence (AI) systems must move beyond purely parameter-based approaches. Hybrid AI models, which integrate imaging data with clinical information, have the potential to generate more realistic and clinically meaningful outputs. Furthermore, the incorporation of explainable AI techniques can enhance the transparency of model outputs and strengthen clinicians' trust in these systems (10).

Accordingly, explainable hybrid AI approaches based on multidimensional data integration are likely to represent one of the most important future research directions in the evaluation of lumbar spinal stenosis.

The findings of the present study are supported by recent developments in multimodal and hybrid AI approaches reported in the literature. In particular, systems that combine imaging data with clinical information and offer explainability appear to have the potential to overcome the inherent limitations of models relying solely on quantitative parameters. In this context, multimodal computer-aided diagnosis (CAD) systems and knowledge graph–based approaches can be considered important steps toward developing models that more closely align with real-world clinical decision-making processes. By demonstrating the discrepancy between parameter-based AI outputs and real-world clinical decision-making, the present study highlights the need for clinically oriented, explainable, and hybrid artificial intelligence frameworks that may facilitate the translation of AI technologies into routine practice.

Recent developments in multimodal artificial intelligence and knowledge-based systems further support the feasibility of clinically translatable approaches (8,9,16). The integration of imaging findings with clinical variables, structured knowledge representations, and explainable learning strategies may help overcome the limitations of conventional parameter-based models. Such hybrid systems have the potential to improve interpretability, enhance model reliability, and provide more personalized and clinically meaningful decision-support tools.

Study Limitations

Several limitations should be acknowledged. First, this study did not involve the development or external validation of a specific artificial intelligence model. Therefore, the findings should be interpreted in the context of currently available evidence and clinically relevant considerations regarding parameter-based AI systems.

Second, considerable heterogeneity exists among previously published studies in terms of datasets, labeling strategies, and model architectures. Such variability limits direct comparisons between studies and may affect the generalizability of reported findings.

Furthermore, the observer-dependent nature of lumbar spinal stenosis grading remains an important source of variability that may influence the performance and reliability of AI systems. These factors should be considered when interpreting the findings and designing future clinically applicable AI models. The key limitations of parameter-based artificial intelligence models in lumbar spinal stenosis are summarized in Table 2.

Table II. Major limitations of parameter-based artificial intelligence models in lumbar spinal stenosis and their clinical implications

Domain	Parameter-based AI Approach	Clinical Reality	Limitation	Clinical Impact
Measurement	Relies on predefined MRI parameters (e.g., AP canal diameter, dural sac area, ligamentum flavum thickness)	No universally accepted or standardized threshold values exist	Lack of standardization	Reduced reproducibility and inconsistent interpretation
Decision-making	Based primarily on imaging findings	Integrates symptoms, neurological examination, and functional status	Inability to capture multidimensional clinical factors	Limited clinical applicability
Clinicoradiological correlation	Assumes a direct relationship between imaging severity and clinical symptoms	A mismatch between imaging findings and symptoms is common	Failure to account for clinicoradiological discrepancy	Potentially misleading outputs
Multilevel disease	Evaluates each spinal level independently	Identification of the dominant symptomatic level is essential	Inability to determine the symptomatic level	Limited value in surgical planning
Model structure	Requires predefined quantitative inputs	Clinical reasoning is not always fully quantifiable	Forced parameterization	Oversimplified decision-making processes
Dataset labeling	Relies on expert-based annotations	High interobserver variability exists	Label noise	Decreased model performance and generalizability

Future Directions

Future studies should move beyond approaches that focus solely on imaging-based parameters. In particular, the development of hybrid artificial intelligence (AI) models incorporating multimodal data integration may enable a more accurate representation of clinical decision-making processes (10,16).

In this context, the integration of clinical variables such as patient symptoms, functional status, and neurological findings into AI models is of critical importance. Additionally, the incorporation of radiological reports through knowledge graph–based approaches and natural language processing techniques may further enhance the model's ability to understand clinical context.

Moreover, the use of explainable AI methods can facilitate the interpretation of model outputs by clinicians, thereby improving trust and promoting the integration of these systems into clinical practice. A clinically aligned and explainable hybrid artificial intelligence framework integrating imaging and clinical data is presented in Figure 3.

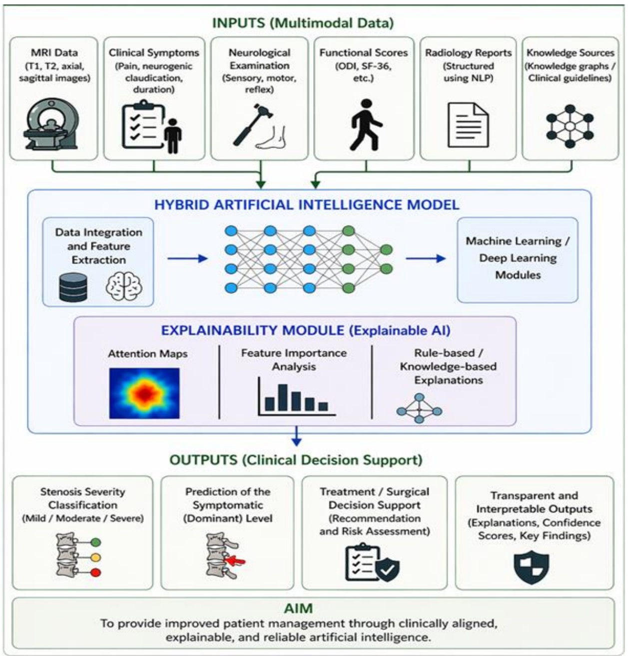

Fig. 3. Proposed framework for a clinically aligned explainable hybrid artificial intelligence model.

Finally, the establishment of large-scale, multicenter, and standardized datasets is essential for improving model generalizability and accelerating translation into real-world clinical applications.

Recent advances in multimodal and explainable AI systems further support the feasibility of clinically translatable approaches and highlight the potential of integrating imaging findings with structured clinical information.

Overall, although parameter-based AI models have substantially contributed to the quantitative evaluation of lumbar spinal stenosis, their limited ability to capture the multidimensional nature of clinical decision-making underscores the need for clinically aligned, explainable, and hybrid AI frameworks.

VI. CONCLUSION

The findings of this study suggest that explainable and hybrid artificial intelligence (AI) approaches, which integrate imaging data with clinical information, have the potential to play a pivotal role in the development of more reliable and clinically meaningful decision-support systems.

ACKNOWLEDGMENT (OPTIONAL)

There are no funding organizations, institutions or individuals supporting the research.

REFERENCES

[1] Katz JN, Zimmerman ZE, Mass H, Makhni MC. Diagnosis and management of lumbar spinal stenosis: a review. JAMA. 2022;327(17):1688–1699. https://doi.org/10.1001/jama.2022.5921

[2] Genevay S, Atlas SJ. Lumbar spinal stenosis. Best Pract Res Clin Rheumatol. 2010;24(2):253–265. https://doi.org/10.1016/j.berh.2009.11.001

[3] Alpert JN. Review of the diagnosis and management of lumbar spinal stenosis. JAMA. 2022;328(8):779–780. https://doi.org/10.1001/jama.2022.11384

[4] 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. https://doi.org/10.1016/j.media.2017.07.005

[5] Schizas C, Theumann N, Burn A, Tansey R, Wardlaw D, Smith FW, et al. Qualitative grading of severity of lumbar spinal stenosis based on the morphology of the dural sac on MRI. Spine (Phila Pa 1976). 2010;35(21):1919–1924. https://doi.org/10.1097/BRS.0b013e3181d359bd

[6] Abdulmahmod OF, Al-Antari MA, Kwon H, Habib A. Medical spine sagittal MRI dataset for segmentation and foraminal stenosis detection. Sci Data. In press.

[7] Al-Antari MA, et al. Evaluating AI-powered predictive solutions for MRI in lumbar spinal stenosis: a systematic review. Artif Intell Rev. In press.

[8] Salem S, Habib A, Raza M, Al-Huda Z, Al-Maqtari O, Ertuğrul B, Yıldırım Ö. AutoSpineAI: lightweight multimodal CAD framework for lumbar spine MRI assessments. Proc IEEE EMBS Int Conf Biomed Health Inform (BHI). 2025.

[9] Raza M, Hussain J, Gu YH, Ertuğrul B, Han I, Al-Antari MA. SR-MKG: self-refining medical knowledge graph via LLM-based topological confidence scoring for radiological reports of lumbar spinal stenosis. Int Artif Intell Data Process Symp. 2025.

[10] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9(4):e1312. https://doi.org/10.1002/widm.1312

[11] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. https://doi.org/10.1038/s41591-018-0300-7

[12] Lee S, Lee JW, Yeom JS, Kim KJ, Kim HJ, Chung SK, et al. A practical MRI grading system for

lumbar foraminal stenosis. AJR Am J Roentgenol. 2010;194(4):1095–1098. https://doi.org/10.2214/AJR.09.2772

[13] Karimi D, Dou H, Warfield SK, Gholipour A. Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. Med Image Anal. 2020;65:101759. https://doi.org/10.1016/j.media.2020.101759

[14] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

[15] Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus nonsurgical treatment for lumbar spinal stenosis. Spine (Phila Pa 1976). 2016;41(14):E857–E868. https://doi.org/10.1097/BRS.0000000000001635

[16] Huang SC, Pareek A, Seyyedi S, Banerjee I, Lungren MP. Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med. 2020;3:136. https://doi.org/10.1038/s41746-020-00341-z

[17] Jensen MC, Brant-Zawadzki MN, Obuchowski N, Modic MT, Malkasian D, Ross JS. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med. 1994;331(2):69–73. https://doi.org/10.1056/NEJM199407143310201

BIOGRAPHIES

Bilal Ertuğrul was born in Batman, Türkiye, in 1979. He received the M.D. degree from the Faculty of Medicine, Dicle University, Diyarbakır, Türkiye, in 2004. He is currently an Assistant Professor with the Department of Neurosurgery, Faculty of Medicine, Fırat University, Elazığ, Türkiye. His clinical and research interests include spinal surgery, neuro-oncology, stereotactic and functional neurosurgery, medical image analysis, and artificial intelligence applications in neurosurgery.