bioengineering

MDPI

Editorial

Application of Deep Learning in Medical Diagnosis

Zhanlin Ji [1,*,†] **and Li Zhao** [2,†]

[1] College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China
[2] Beijing National Research Center for Information Science and Technology, Institute for Precision Medicine, Tsinghua University, Beijing 100084, China
* Correspondence: zhanlin.ji@ncst.edu.cn
† These authors contributed equally to this work.

1. Introduction

This article belongs to the Special Issue titled "Application of Deep Learning in Medical Diagnosis".

Deep learning is reshaping medical diagnosis across image reconstruction, harmonization, segmentation, disease classification, and outcome prediction. However, the field has reached a point at which improvements in benchmark accuracy, although important, are no longer sufficient evidence of clinical value. Reliable medical artificial intelligence must also remain robust across institutions and devices, communicate uncertainty, fit within clinical workflows, and demonstrate that its use benefits patients and clinicians. The eight papers collected in this Special Issue of *Bioengineering*, "Application of Deep Learning in Medical Diagnosis", illuminate complementary parts of this transition from promising algorithms to dependable clinical intelligence.

2. Recent Advances Represented in This Special Issue

The first group of contributions addresses the quality and consistency of medical images before diagnostic inference. Lee et al. investigated deep learning-based harmonization of abdominal computed tomography images to improve the reproducibility of radiomic features [1]. Bertl et al. examined how deep learning-based image reconstruction influences tumor visibility and diagnostic confidence in computed tomography [2]. Together, these studies underscore a foundational point: diagnostic models cannot be separated from the acquisition and reconstruction pipelines that shape their inputs. Harmonization and reconstruction may improve downstream analysis, but their effects must be evaluated not only through image-quality metrics; the stability of quantitative features and clinically meaningful interpretation are also required.

A second theme is efficient and task-focused model design. Lin et al. combined denoised CT images, EfficientNet, and a voting strategy for the identification of locally advanced rectal cancer [3]. Hao et al. introduced DTONet as a lightweight model for melanoma segmentation [4], while Rai et al. presented LightweightUNet with multimodal and generative-adversarial-network-augmented imaging data for breast cancer detection [5]. These studies reflect the practical need to balance discrimination, segmentation quality, computational cost, and data availability. Smaller and more efficient architectures may support deployment in settings with limited hardware, but efficiency should be assessed together with calibration, robustness, inference latency, and the consequences of failure.

This Special Issue also broadens the discussion beyond conventional image classification. Oskouei et al. proposed OKEN, an optimizable dimensionality-reduction framework

Check for updates

Received: 2 November 2026
Accepted: 3 November 2026
Published: 18 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

for whole-slide image classification [6], in which an evolutionary algorithm evolves a transformation matrix using the silhouette score as the fitness function, yielding a latent space that is compact and explicitly optimized for class separability rather than variance preservation alone. Khosravi et al. externally validated deep learning models for determining the etiology of retinal hemorrhage across diverse fundus-photography datasets [7], directly addressing the persistent gap between internal performance and transportability. Finally, Razo et al. used hospital-level variables and novel physiological markers to predict neurological outcomes after out-of-hospital cardiac arrest [8]. Beyond conventional clinical variables, their model incorporated ocular ultrasound parameters, plasma protein biomarkers, and sex steroid hormones, and SHAP analysis identified interferon gamma and TARC as features of high importance, demonstrating that structured physiological information can extend the variable space of prognostic modeling when imaging alone is insufficient. This contribution illustrates the value of integrating imaging-oriented advances with structured clinical and physiological information when the diagnostic or prognostic question cannot be answered by a single modality.

Taken together, the eight papers span the diagnostic pipeline from input standardization and reconstruction to segmentation, classification, representation learning, external validation, and multimodal prognosis. Their diversity is a strength. It shows that progress in medical diagnosis depends not on a single architecture, but on coordinated improvements in data quality, model design, validation, and clinical interpretation.

3. Current Knowledge Gaps

Several gaps remain. First, data heterogeneity continues to limit reproducibility and external validity. Differences in scanners, acquisition protocols, reconstruction settings, populations, prevalence, and annotation practice can produce performance shifts that are obscured by random internal splits. External validation, as demonstrated in this Special Issue, should become routine rather than exceptional, and it should include transparent reporting across clinically relevant subgroups.

Second, limited and uneven annotations constrain representation learning. Synthetic augmentation, weak supervision, transfer learning, and self-supervised learning can reduce this burden, but each introduces its own assumptions. Synthetic data may amplify hidden biases or create unrealistic shortcuts; weak labels may encode institutional practice rather than biological truth. These methods therefore require validation against independent, representative, and prospectively collected data.

Third, predictive accuracy is not equivalent to trustworthy clinical utility. Many studies still provide incomplete evidence on calibration, uncertainty, failure detection, interpretability, and the handling of unfamiliar inputs. A model can achieve strong average performance while remaining unsafe for particular subgroups or rare presentations. The field also lacks sufficient evidence from prospective studies of clinician–AI interactions, where automation bias, alert fatigue, time pressure, and differences in expertise can alter the real-world effect of a system.

Finally, deployment raises governance questions that are inseparable from technical performance. Dataset provenance, privacy protection, model updating, auditability, and post-deployment monitoring should be planned from the outset. Reproducible code and clear reporting are essential, but they do not replace local validation or accountability for clinical decisions.

4. Priorities for Future Research

Future work should prioritize disease-relevant representations that remain stable across institutions, devices, and acquisition protocols. Multicenter and prospective vali-

dation should report calibration and subgroup performance, not only aggregate discrimination. Where domain adaptation or harmonization is used, investigators should show that clinically important signals are preserved rather than merely making datasets appear statistically similar.

Efficient learning should be pursued through lightweight architectures, self-supervised pretraining, weak supervision, and rigorously evaluated synthetic augmentation. Comparisons should account for memory use, inference time, energy consumption, and maintenance requirements alongside predictive performance. This broader view of efficiency is especially important for resource-constrained settings and point-of-care applications.

Multimodal integration should be clinically purposeful. Combining imaging, physiological signals, laboratory values, medical records, and patient-reported information may improve diagnosis and prognosis, but more data do not automatically produce better care. Future studies should address missingness, temporal alignment, acquisition cost, and whether each modality contributes information that changes a decision. Models should also remain useful when one or more inputs are unavailable.

Evaluation must become uncertainty-aware and clinically grounded. Interpretability methods should be tested for stability and relevance to the intended users. Calibration, out-of-distribution detection, abstention strategies, and human-readable confidence communication should be treated as core design elements. Ultimately, prospective multicenter studies should measure changes in clinician performance, diagnostic delay, downstream testing, patient outcomes, and health equity.

5. Concluding Perspective

The decisive test of medical AI is whether it improves care when clinicians use it. The contributions to this Special Issue demonstrate meaningful progress across data preparation, efficient modeling, compact representation, external validation, and multimodal prediction. They also make clear that no single performance metric can establish readiness for clinical use. The next phase of research should connect robust data processing, efficient and multimodal learning, uncertainty-aware prediction, and prospective clinical evaluation so that AI systems can communicate both the value and the limits of their recommendations. We sincerely thank all authors, reviewers, and members of the editorial team for advancing this interdisciplinary effort.

Author Contributions: Z.J. and L.Z. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Lee, S.B.; Hong, Y.; Cho, Y.J.; Jeong, D.; Lee, J.; Choi, J.W.; Hwang, J.Y.; Lee, S.; Choi, Y.H.; Cheon, J.-E. Enhancing Radiomics Reproducibility: Deep Learning-Based Harmonization of Abdominal Computed Tomography (CT) Images. *Bioengineering* **2024**, *11*, 1212. [CrossRef] [PubMed]
2. Bertl, M.; Hahne, F.-G.; Gräger, S.; Heinrich, A. Impact of Deep Learning-Based Image Reconstruction on Tumor Visibility and Diagnostic Confidence in Computed Tomography. *Bioengineering* **2024**, *11*, 1285. [CrossRef] [PubMed]

3. Lin, C.-Y.; Wu, J.C.-H.; Kuan, Y.-M.; Liu, Y.-C.; Chang, P.-Y.; Chen, J.-P.; Lu, H.H.-S.; Lee, O.K.-S. Precision Identification of Locally Advanced Rectal Cancer in Denoised CT Scans Using EfficientNet and Voting System Algorithms. *Bioengineering* **2024**, *11*, 399. [CrossRef] [PubMed]
4. Hao, S.; Wang, H.; Chen, R.; Liao, Q.; Ji, Z.; Lyu, T.; Zhao, L. DTONet a Lightweight Model for Melanoma Segmentation. *Bioengineering* **2024**, *11*, 390. [CrossRef] [PubMed]
5. Rai, H.M.; Yoo, J.; Agarwal, S.; Agarwal, N. LightweightUNet: Multimodal Deep Learning with GAN-Augmented Imaging Data for Efficient Breast Cancer Detection. *Bioengineering* **2025**, *12*, 73. [CrossRef] [PubMed]
6. Oskouei, S.; Pedersen, A.; Valla, M.; Dale, V.G.; Wahl, S.G.F.; Haugum, M.D.; Langø, T.; Ramnefjell, M.P.; Akslen, L.A.; Kiss, G.; et al. OKEN: A Supervised Evolutionary Optimizable Dimensionality Reduction Framework for Whole Slide Image Classification. *Bioengineering* **2025**, *12*, 733. [CrossRef] [PubMed]
7. Khosravi, P.; Huck, N.A.; Shahraki, K.; Ghafari, E.; Azimi, R.; Kim, S.Y.; Crouch, E.; Xie, X.; Suh, D.W. External Validation of Deep Learning Models for Classifying Etiology of Retinal Hemorrhage Using Diverse Fundus Photography Datasets. *Bioengineering* **2025**, *12*, 20. [CrossRef] [PubMed]
8. Razo, M.; Kotini, P.; Li, J.; Khosla, S.; Buhimschi, I.A.; Vanden Hoek, T.; Del Rios, M.; Darabi, H. Using a Neural Network Architecture for the Prediction of Neurologic Outcome for Out-of-Hospital Cardiac Arrests Using Hospital Level Variables and Novel Physiologic Markers. *Bioengineering* **2025**, *12*, 124. [CrossRef] [PubMed]

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