

Editorial

Deep Learning and Vision Transformer for Medical Image Analysis

Yudong Zhang ^{1,*} , Jiaji Wang ¹ , Juan Manuel Gorriz ²  and Shuihua Wang ¹

¹ School of Computing and Mathematical Sciences, University of Leicester, Leicester LE1 7RH, UK; jw933@le.ac.uk (J.W.); shuihuawang@ieee.org (S.W.)

² Department of Signal Theory, Networking, and Communications, University of Granada, 52005 Granada, Spain; gorriz@ugr.es

* Correspondence: yudongzhang@ieee.org; Tel.: +44-754-870-0453

Artificial intelligence (AI) refers to the field of computer science theory and technology [1] that is focused on creating intelligent machines capable of simulating human intelligence [2]. AI systems [3] are designed to perform tasks that typically require human intelligence [4], such as perception, learning, reasoning [5], problem-solving [6], decision-making [7], etc.

Machine learning (ML) [8] is a subfield of AI that encompasses algorithms and statistical models, enabling computer systems to automatically learn from data, identify patterns, and make predictions or decisions without being explicitly programmed [9]. It involves the development of mathematical models and algorithms [10] that allow machines to iteratively process and analyze large datasets, learn from examples or experiences, and improve their performance over time. By leveraging ML theories and techniques [11], computers can discover complex patterns, extract meaningful insights, and generate reliable predictions, making ML a powerful tool for various applications in fields such as finance, smart healthcare [12], the Internet of Things [13], natural language processing (NLP) [14], recommendation systems, etc.

Deep learning (DL) is a specialized branch of ML that focuses on the development and training of artificial neural networks with multiple layers of interconnected nodes [15], which are known as deep neural networks. It enables computers to automatically learn hierarchical representations of data, allowing for the extraction of intricate patterns and features from complex datasets [16]. DL leverages the power of large-scale computing and vast amounts of data [17] to enable neural networks to perform sophisticated tasks, such as image and speech recognition, NLP, and even autonomous decision-making. By emulating the structure and functionality of the human brain, DL has revolutionized AI by significantly enhancing the accuracy and performance of various applications [18] including medical image analysis (MIA) [19], while also demanding substantial computational resources.

Transformers are a revolutionary DL method that have greatly impacted the field of NLP. They are an example of a neural network model designed to process sequential data, such as sentences or paragraphs, by leveraging attention mechanisms. Unlike traditional recurrent neural networks (RNNs) [20] that process input sequentially, transformers [21] employ a parallelized approach, allowing for more efficient and scalable computation. By focusing on the relationships and dependencies between different words or tokens within a sequence, the transformer model excels at tasks like machine translation, text generation, sentiment analysis, and language understanding [22]. Transformers' self-attention mechanisms enable them to capture contextual information effectively, resulting in state-of-the-art performance on a wide range of NLP benchmarks and applications. Transformers have become the foundation for many advanced language models, such as BERT, ChatGPT [23], and T5, and have significantly advanced the capabilities of language understanding and generation systems. Vision transformers (ViTs) [24] are an adaptation



Citation: Zhang, Y.; Wang, J.; Gorriz, J.M.; Wang, S. Deep Learning and Vision Transformer for Medical Image Analysis. *J. Imaging* **2023**, *9*, 147. <https://doi.org/10.3390/jimaging9070147>

Received: 15 June 2023

Accepted: 18 July 2023

Published: 21 July 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

of the classical transformer architecture that apply self-attention mechanisms to process image data [25], making them an exemplary powerful model for tasks in computer vision, showcasing the extension of transformers' effectiveness beyond NLP. Figure 1 shows the relationship between AI, ML, DL, and Transformers.

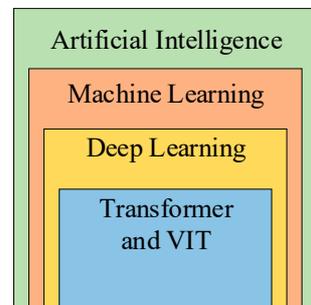


Figure 1. Relationship between AI, ML, DL, and Transformers.

Medical image analysis (MIA) [26] is an important field of application for AI. MIA involves a series of common procedures [27], starting with image acquisition, wherein medical imaging modalities capture anatomical or functional information. The acquired images then undergo preprocessing techniques [28] to correct artifacts, enhance quality, and standardize the data. Next, segmentation methods [29] are employed to separate and identify specific structures or regions of interest within the images. Registration techniques [30] are applied to align multiple images or different modalities for spatial correspondence.

Feature extraction algorithms [31] extract relevant quantitative or qualitative information from the segmented regions for subsequent analysis. Classification methods [32] are then utilized to classify the extracted features, enabling the identification of diseases or conditions. Visualization techniques [33] help in the interpretation and display of the analysis results for clinicians and researchers. Localization methods [34] precisely determine the spatial location of abnormalities or structures within the images, aiding in diagnosis and treatment planning. These procedures, shown in Figure 2, collectively contribute to the comprehensive analysis and interpretation of medical images, ultimately facilitating improved patient care and medical research [35].

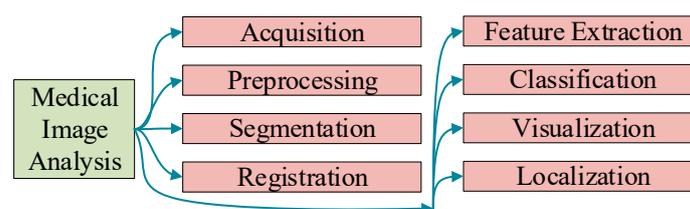


Figure 2. Eight common procedures in medical image analysis.

DL for MIA faces several challenges. Acquiring a sufficient quantity of high-quality annotated medical images can be challenging due to privacy concerns, limited availability, and the time-consuming process of manual annotation [36]. DL and ViT models often require a large amount of labeled data to achieve optimal performance, and this data may be limited for rare diseases [37] or specific subpopulations. Further, DL and ViT models typically have a large number of parameters, making them demanding and in need of substantial computational resources [38] for training and inference.

Author Contributions: Conceptualization, Y.Z. and J.W.; methodology, J.M.G. and S.W.; validation, Y.Z. and J.W.; formal analysis, J.M.G. and S.W.; investigation, Y.Z.; resources, J.W.; data curation, J.M.G. and S.W.; writing—original draft preparation, Y.Z. and J.W.; writing—review and editing, J.M.G. and S.W.; supervision, J.M.G. and S.W.; project administration, Y.Z. and J.W.; funding acquisition, Y.Z., J.M.G. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was partially supported by MRC, UK (MC_PC_17171); Royal Society, UK (RP202G0230); Hope Foundation for Cancer Research, UK (RM60G0680); GCRF, UK (P202PF11); Sino-UK Industrial Fund, UK (RP202G0289); LIAS, UK (P202ED10, P202RE969); Data Science Enhancement Fund, UK (P202RE237); Fight for Sight, UK (24NN201); Sino-UK Education Fund, UK (OP202006); BBSRC, UK (RM32G0178B8); MCIN/AEI (10.13039/501100011033); FEDER ‘Una manera de hacer Europa’ (RTI2018-098913-B100) by the Consejería de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía); FEDER (CV20-45250, A-TIC-080-UGR18, B-TIC-586-UGR20, and P20-00525).

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

References

- Ghouri, A.M.; Khan, H.R.; Mani, V.; ul Haq, M.A.; Jabbour, A. An artificial-intelligence-based omnichannel blood supply chain: A pathway for sustainable development. *J. Bus. Res.* **2023**, *164*, 113980. [\[CrossRef\]](#)
- Cundall, P. Human intelligence seems capable of anything to me. *New Sci.* **2023**, *246*, 30.
- Lee, M.C.M.; Scheepers, H.; Lui, A.K.H.; Ngai, E.W.T. The implementation of artificial intelligence in organizations: A systematic literature review. *Inf. Manag.* **2023**, *60*, 103816. [\[CrossRef\]](#)
- Raspanti, M.A.; Palazzani, L. Artificial intelligence and human intelligence: Contributions of christian theology and philosophy of the person. *Biola J.-Riv. Biodiritto* **2022**, 457–471. [\[CrossRef\]](#)
- Saleem, K.; Saleem, M.; Ahmad, R.Z.; Javed, A.R.; Alazab, M.; Gadekallu, T.R.; Suleman, A. Situation-aware bdi reasoning to detect early symptoms of covid 19 using smartwatch. *IEEE Sens. J.* **2023**, *23*, 898–905. [\[CrossRef\]](#)
- Goudar, V.; Peysakhovich, B.; Freedman, D.J.; Buffalo, E.A.; Wang, X.J. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nat. Neurosci.* **2023**, *26*, 879–890. [\[CrossRef\]](#)
- Gomez, C.; Unberath, M.; Huang, C.M. Mitigating knowledge imbalance in ai-advised decision-making through collaborative user involvement. *Int. J. Hum.-Comput. Stud.* **2023**, *172*, 102977. [\[CrossRef\]](#)
- Shakibi, H.; Faal, M.Y.; Assareh, E.; Agarwal, N.; Yari, M.; Latifi, S.A.; Ghodrat, M.; Lee, M. Design and multi-objective optimization of a multi-generation system based on pem electrolyzer, ro unit, absorption cooling system, and orc utilizing machine learning approaches; a case study of australia. *Energy* **2023**, *278*, 127796. [\[CrossRef\]](#)
- Bhowmik, R.T.; Jung, Y.S.; Aguilera, J.A.; Prunicki, M.; Nadeau, K. A multi-modal wildfire prediction and early-warning system based on a novel machine learning framework. *J. Environ. Manag.* **2023**, *341*, 117908. [\[CrossRef\]](#)
- Kozikowski, P. Machine learning for grouping nano-objects based on their morphological parameters obtained from sem analysis. *Micron* **2023**, *171*, 103473. [\[CrossRef\]](#)
- Vinod, D.N.; Prabaharan, S.R.S. Elucidation of infection asperity of ct scan images of COVID-19 positive cases: A machine learning perspective. *Sci. Afr.* **2023**, *20*, e01681. [\[CrossRef\]](#)
- Abd Rahman, N.H.; Zaki, M.H.M.; Hasikin, K.; Abd Razak, N.A.; Ibrahim, A.K.; Lai, K.W. Predicting medical device failure: A promise to reduce healthcare facilities cost through smart healthcare management. *PeerJ Comput. Sci.* **2023**, *9*, e1279. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yazdanpanah, S.; Chaeikar, S.S.; Jolfaei, A. Monitoring the security of audio biomedical signals communications in wearable iot healthcare. *Digit. Commun. Netw.* **2023**, *9*, 393–399. [\[CrossRef\]](#)
- Pyne, Y.; Wong, Y.M.; Fang, H.S.; Simpson, E. Analysis of ‘one in a million’ primary care consultation conversations using natural language processing. *BMJ Health Care Inform.* **2023**, *30*, e100659. [\[CrossRef\]](#)
- Ahmed, S.; Raza, B.; Hussain, L.; Aldweesh, A.; Omar, A.; Khan, M.S.; Eldin, E.T.; Nadim, M.A. The deep learning resnet101 and ensemble xgboost algorithm with hyperparameters optimization accurately predict the lung cancer. *Appl. Artif. Intell.* **2023**, *37*, 2166222. [\[CrossRef\]](#)
- Tyson, R.; Gavalian, G.; Ireland, D.G.; McKinnon, B. Deep learning level-3 electron trigger for clas12. *Comput. Phys. Commun.* **2023**, *290*, 108783. [\[CrossRef\]](#)
- Almutairy, F.; Scekcic, L.; Matar, M.; Elmoudi, R.; Wshah, S. Detection and mitigation of gps spoofing attacks on phasor measurement units using deep learning. *Int. J. Electr. Power Energy Syst.* **2023**, *151*, 109160. [\[CrossRef\]](#)
- Alizadehsani, Z.; Ghaemi, H.; Shahraki, A.; Gonzalez-Briones, A.; Corchado, J.M. Dcservcg: A data-centric service code generation using deep learning. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106304. [\[CrossRef\]](#)
- Zhang, Y.; Dong, Z. Medical imaging and image processing. *Technologies* **2023**, *11*, 54. [\[CrossRef\]](#)
- Kessler, S.; Schroeder, D.; Korlakov, S.; Hettlich, V.; Kalkhoff, S.; Moazemi, S.; Lichtenberg, A.; Schmid, F.; Aubin, H. Predicting readmission to the cardiovascular intensive care unit using recurrent neural networks. *Digit. Health* **2023**, *9*, 20552076221149529. [\[CrossRef\]](#)
- Alam, F.; Ananbeh, O.; Malik, K.M.; Odayani, A.A.; Hussain, I.B.; Kaabia, N.; Aidaroos, A.A.; Saudagar, A.K.J. Towards predicting length of stay and identification of cohort risk factors using self-attention-based transformers and association mining: COVID-19 as a phenotype. *Diagnostics* **2023**, *13*, 1760. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fuad, K.A.A.; Chen, L.Z. A survey on sparsity exploration in transformer-based accelerators. *Electronics* **2023**, *12*, 2299. [\[CrossRef\]](#)
- Gradonm, K.T. Electric sheep on the pastures of disinformation and targeted phishing campaigns: The security implications of chatgpt. *IEEE Secur. Priv.* **2023**, *21*, 58–61. [\[CrossRef\]](#)

24. Hoshi, T.; Shibayama, S.; Jiang, X.A. Employing a hybrid model based on texture-biased convolutional neural networks and edge-biased vision transformers for anomaly detection of signal bonds. *J. Electron. Imaging* **2023**, *32*, 023039. [[CrossRef](#)]
25. Chen, S.; Lu, S.; Wang, S.; Ni, Y.; Zhang, Y. Shifted window vision transformer for blood cell classification. *Electronics* **2023**, *12*, 2442. [[CrossRef](#)]
26. Apostolidis, K.D.; Papakostas, G.A. Digital watermarking as an adversarial attack on medical image analysis with deep learning. *J. Imaging* **2022**, *8*, 155. [[CrossRef](#)]
27. Kiryati, N.; Landau, Y. Dataset growth in medical image analysis research. *J. Imaging* **2021**, *7*, 155. [[CrossRef](#)]
28. Wang, S. Advances in data preprocessing for biomedical data fusion: An overview of the methods, challenges, and prospects. *Inf. Fusion* **2021**, *76*, 376–421. [[CrossRef](#)]
29. Shan, C.X.; Li, Q.; Guan, X. Lightweight brain tumor segmentation algorithm based on multi-view convolution. *Laser Optoelectron. Prog.* **2023**, *60*, 1010018. [[CrossRef](#)]
30. Baum, Z.M.C.; Hu, Y.P.; Barratt, D.C. Meta-learning initializations for interactive medical image registration. *IEEE Trans. Med. Imaging* **2023**, *42*, 823–833. [[CrossRef](#)]
31. Shamna, N.V.; Musthafa, B.A. Feature extraction method using hog with ltp for content-based medical image retrieval. *Int. J. Electr. Comput. Eng. Syst.* **2023**, *14*, 267–275. [[CrossRef](#)]
32. Hida, M.; Eto, S.; Wada, C.; Kitagawa, K.; Imaoka, M.; Nakamura, M.; Imai, R.; Kubo, T.; Inoue, T.; Sakai, K.; et al. Development of hallux valgus classification using digital foot images with machine learning. *Life* **2023**, *13*, 1146. [[CrossRef](#)] [[PubMed](#)]
33. Niemitz, L.; van der Stel, S.D.; Sorensen, S.; Messina, W.; Sekar, S.K.V.; Sterenborg, H.; Andersson-Engels, S.; Ruers, T.J.M.; Burke, R. Microcamera visualisation system to overcome specular reflections for tissue imaging. *Micromachines* **2023**, *14*, 1062. [[CrossRef](#)] [[PubMed](#)]
34. Bodard, S.; Denis, L.; Hingot, V.; Chavignon, A.; Helenon, O.; Anglicheau, D.; Couture, O.; Correas, J.M. Ultrasound localization microscopy of the human kidney allograft on a clinical ultrasound scanner. *Kidney Int.* **2023**, *103*, 930–935. [[CrossRef](#)]
35. Zhang, Y.; Gorriz, J.M. Deep learning in medical image analysis. *J. Imaging* **2021**, *7*, 74. [[CrossRef](#)]
36. Sylolypavan, A.; Sleeman, D.; Wu, H.H.; Sim, M. The impact of inconsistent human annotations on ai driven clinical decision making. *NPJ Digit. Med.* **2023**, *6*, 26. [[CrossRef](#)]
37. Talesh, S.A.; Mahmoudi, S.; Mohebbali, M.; Mamishi, S. A rare presentation of visceral leishmaniasis and epididymo-orchitis in a patient with chronic granulomatous disease. *Clin. Case Rep.* **2023**, *11*, e7426. [[CrossRef](#)]
38. Court, L.E.; Fave, X.; Mackin, D.; Lee, J.; Yang, J.Z.; Zhang, L.F. Computational resources for radiomics. *Transl. Cancer Res.* **2016**, *5*, 340–348. [[CrossRef](#)]

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.