How Artificial Intelligence is Shaping Medical Imaging Technology: A Survey of Innovations and Applications

Luís Pinto-Coelho *

Review

How Artificial Intelligence Is Shaping Medical Imaging Technology: A Survey of Innovations and Applications

Luís Pinto-Coelho [1,2,*]

[1] ISEP—School of Engineering, Polytechnic Institute of Porto, 4200-465 Porto, Portugal
[2] INESCTEC, Campus da Faculdade de Engenharia da Universidade do Porto, 4200-465 Porto, Portugal
[*] Correspondence: lfc@isep.ipp.pt

Abstract: The integration of artificial intelligence (AI) into medical imaging has guided in an era of transformation in healthcare. This literature review explores the latest innovations and applications of AI in the field, highlighting its profound impact on medical diagnosis and patient care. The innovation segment explores into cutting-edge developments in AI, such as deep learning algorithms, convolutional neural networks, and generative adversarial networks, which have significantly improved the accuracy and efficiency of medical image analysis. These innovations have enabled rapid and accurate detection of abnormalities, from identifying tumors during radiological examinations to detecting early signs of eye disease in retinal images. The article also highlights various applications of AI in medical imaging, including radiology, pathology, cardiology, and more. AI-based diagnostic tools not only speed up the interpretation of complex images but also improve early detection of disease, ultimately delivering better outcomes for patients. Additionally, AI-based image processing facilitates personalized treatment plans, thereby optimizing healthcare delivery. This literature review highlights the paradigm shift that AI has brought to medical imaging, highlighting its role in revolutionizing diagnosis and patient care. By combining cutting-edge AI techniques and their practical applications, it is clear that AI will continue shaping the future of healthcare in profound and positive ways.

Keywords: transformers; artificial intelligence; GAN; medical images

---

Introduction

Advancements in medical imaging and artificial intelligence (AI) have ushered in a new era of possibilities in the field of healthcare. The fusion of these two domains has revolutionized various aspects of medical practice, ranging from early disease detection and accurate diagnosis to personalized treatment planning and improved patient outcomes.

Medical imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) play a pivotal role in providing clinicians with detailed and comprehensive visual information about the human body. These imaging modalities generate vast amounts of data that require efficient analysis and interpretation, and this is where AI steps in.

AI, particularly deep learning algorithms, has demonstrated remarkable capabilities in extracting valuable insights from medical images. Deep learning models, trained on large datasets, are capable of recognizing complex patterns and features that may not be readily discernible to the human eye. These algorithms can even provide a new perspective about what image features should be valued to support decisions. One of the key advantages of AI in medical imaging is its ability to enhance the accuracy and efficiency of disease diagnosis. Through this process, AI can assist healthcare professionals in detecting abnormalities, identifying specific structures, and predicting disease outcomes.

By leveraging machine learning algorithms, AI systems can analyze medical images with speed and precision, aiding in the identification of early-stage diseases that may be difficult to detect

through traditional methods. This early detection is crucial as it can lead to timely interventions, potentially saving lives and improving treatment outcomes.

Furthermore, AI has opened up new possibilities in image segmentation and quantification. By employing sophisticated algorithms, AI can accurately delineate structures of interest within medical images, such as tumors, blood vessels, or organs. This segmentation capability is invaluable in treatment planning, as it enables clinicians to precisely target areas for intervention, optimize surgical procedures, and deliver targeted therapies.

The integration of AI and medical imaging has also facilitated the development of personalized medicine. Through the analysis of medical images and patient data, AI algorithms can generate patient-specific insights, enabling tailored treatment plans that consider individual variations in anatomy, physiology, and disease characteristics. This personalized approach to healthcare enhances treatment efficacy and minimizes the risk of adverse effects, leading to improved patient outcomes and quality of life.

Additionally, AI has paved the way for advancements in image-guided interventions and surgical procedures. By combining preoperative imaging data with real-time imaging during surgery, AI algorithms can provide surgeons with augmented visualization, navigation assistance, and decision support. These tools enhance surgical precision, reduce procedural risks, and enable minimally invasive techniques, ultimately improving patient safety and surgical outcomes.

Recently several cutting-edge articles have been published covering a wide variety of topics within the scope of medical imaging and AI. Many of these outstanding advancements are directed to cancer, a major cause of severe disease and mortality. The main contributions and fields will be addressed in the next sections.

Technological Innovations

Mathematical models and algorithms stand at the forefront of scientific exploration, serving as powerful tools that enable us to unravel complex phenomena, make predictions, and uncover hidden patterns in vast datasets. These essential components of modern research have not only revolutionized our understanding of the natural world but have also played a pivotal role in driving technological breakthroughs that open up numerous application possibilities across various domains. The synergy between mathematical models and algorithms has not only enhanced our understanding of the world but has also been a driving force behind technological advancements that have transformed our daily lives.

The earliest multilayer perceptron networks, while representing a crucial step in the evolution of neural networks, had notable limitations. One of the primary constraints was their shallow architecture, which consisted of only a few layers, limiting their ability to model complex patterns. Besides the model expansion restrictions imposed by the limited computing power, training these networks with multiple layers was also challenging. In particular, the earliest activation functions used in neural networks, including the sigmoid and hyperbolic tangent (tanh), leaded to the vanishing gradient problem [1], as their gradients became exceedingly small as inputs moved away from zero. This issue impeded the efficient propagation of gradients during training, resulting in slow convergence or training failures. Furthermore, the limited output range of these functions and their symmetric nature constrained the network's ability to represent complex, high-dimensional data. Additionally, the computational complexity of these functions, particularly the exponential calculations, hindered training and inference in large networks. These shortcomings led to the development and widespread adoption of more suitable activation functions, such as the Rectified Linear Unit (ReLU) [2] and its variants, which successfully addressed these issues and became integral components of modern deep learning architectures [3]. For these reasons, early multilayer perceptron networks struggled to capture complex patterns in data, making them unsuitable for tasks requiring the modeling of intricate relationships, ultimately leading to the necessity of exploration of more advanced architectures and training techniques.

Improvements in the artificial neurons functionality, more advanced architectures, and improved training algorithms supported by graphical computational units (GPU) came to open

promising possibilities. The LeNet-5 architecture, developed for the recognition of handwritten digits [4], is a fundamental milestone for Convolutional Neural Networks (CNNs) [5,6].

CNNs, inspired by the biological operation of animals' vision system, assume that the input is the representation of image data. Current architectures follow a structured sequence of layers, each with specific functions to process and extract features from the input data [7]. The journey begins with the input layer, which receives raw image data, typically represented as a grid of pixel values, often with three color channels (Red, Green, Blue) for color images. Following the input layer, the network employs Convolutional Layers, which are responsible for feature extraction. These layers use convolutional operations (of several types [6]) to detect local patterns and features in the input data. Early convolutional layers focus on detecting basic features like edges, corners, and textures. After each convolution operation, Activation Layers with Rectified Linear Unit (ReLU) activation functions are applied to introduce non-linearity. ReLU units help the network learn more complex patterns and enhance its ability to model the data effectively. Pooling (Subsampling) Layers come next, reducing the spatial dimensions of the feature maps while preserving important information. Max-pooling and average-pooling are common operations that help make the network more robust to variations in scale and position. The sequence of Convolutional Layers continues, with additional layers stacked to capture increasingly complex and abstract features. These deeper layers are adept at detecting higher-level patterns, shapes, and objects in the data. Similar to the earlier Convolutional Layers, Activation Layers with ReLU functions are applied after each convolution operation, maintaining non-linearity and enhancing feature learning. Pooling (Subsampling) Layers may be used again, further decreasing the spatial dimensions of the feature maps and retaining essential information. At the end of this sequence, after the network has extracted the most relevant information from the input data, a special set of vectors are obtained, designated by deep features [8]. These, located deep in the network, distill data into compact, meaningful forms that are highly discriminative. They often have lower dimensionality than the raw input data, which not only conserves computational resources but also simplifies subsequent processing, making it especially beneficial in the analysis of high-dimensional data, such as images. This process also eliminates the tedious and error prone process of handcrafted feature selection, leading to optimized feature sets and to the possibility of building the so-called "end-to-end" systems. Deep features can also help mitigate overfitting, a common challenge in machine learning, since, by learning relevant representations, they prevent models from memorizing the training data and encourage more robust generalization.

Another great advantage of deep feature extraction pipelines is the possibility of using transfer learning techniques. In this case, a deep feature extraction network, previously successfully developed on one task or dataset can be transferred and fine-tuned to another related task, significantly reducing the need for large labeled datasets and speeding up model training. This versatility is a game changer in many applications.

After this extraction front-end, continuing with the processing pipeline and moving towards the end of the network, Fully Connected Layers are introduced. These layers come after the convolutional and pooling layers and play a pivotal role in feature aggregation and classification. The deep features extracted by the previous layers are flattened and processed through one or more fully connected layers.

Finally, the Output Layer emerges as the last layer of the network. The number of neurons in this layer corresponds to the number of classes in a classification task or the number of output units in a regression task. For classification tasks, a softmax activation function is typically used to calculate class probabilities, providing the final output of the CNN.

The described structured sequence of layers, from the input layer to the output layer, captures the hierarchical feature learning process in a CNN, allowing it to excel in image classification tasks (among others). Specific CNN architectures may introduce variations, additional components, or specialized layers based on the network's design goals and requirements.

Transformers

CNNs are well-suited for grid-like data, such as images, where local patterns can be captured efficiently. However, they struggle with sequential data because they lack a mechanism for modeling dependencies between distant elements (for example in distinct time instants, or far in the image). Also, CNNs do not inherently model the position or order of elements within the data. They rely on shared weight filters, which makes them translation-invariant but can be problematic when absolute spatial relationships are important [9]. To overcome these limitations (handling sequential data, modeling long-range dependencies, incorporating positional information, and addressing tasks involving multimodal data, among others), transformers were introduced [10]. In the context of machine learning applied to images, transformers are a type of neural network architecture that extends the transformer model originally designed for natural language processing [10] to handle computer vision tasks. These models are often referred to as Vision Transformers (ViTs) or Image Transformers [11].

In the context of Vision Transformers (ViT), the initial image undergoes a transformation process, wherein it is divided into a sequence of patches. Each of these patches is associated with a positional encoding technique, which captures and encodes the spatial positions of the patches, thus preserving spatial information. These patches, together with a class token, are then input into a Transformer model to perform Multi-Head Self-Attention (MHSA) and generate embeddings that represent the learned characteristics of the patches. The class token's state in the ViT's output functions as the representation of the entire image. Finally, a multi-layer perceptron (MLP) is employed for the purpose of classifying the learned image representation. Notably, in addition to using raw images, it is also possible to supply feature maps generated by Convolutional Neural Networks (CNNs) as input to a Vision Transformer for the purpose of establishing relational mappings [12].

The attention mechanism is a fundamental component in transformers. It plays a pivotal role in enabling the model to selectively focus on different parts of the input data with varying degrees of attention. At its core, the attention mechanism allows the model to assign varying levels of importance to different elements within the input data. This means the model can "pay attention" to specific elements while processing the data, prioritizing those that are most relevant to the task at hand. This selective attention enhances the model's ability to capture essential information and relationships within the input. The mechanism operates as follows: First, the input data is organized into a sequence of elements, such as tokens in a sentence for NLP or patches in an image for computer vision. Then, the mechanism introduces three sets of learnable parameters: query (Q), key (K), and value (V). The query represents the element of interest, while the key and value pairs are associated with each element in the input sequence. For each element in the input sequence, the attention mechanism calculates an attention score, reflecting the similarity between the query and the key for that element. The method used to measure this similarity can vary, with techniques like dot product and scaled dot product being common choices. These attention scores represent how relevant each element is to the query. The next step involves applying the softmax function to the attention scores. This converts them into weights that sum to 1, effectively determining the importance of each input element concerning the query. The higher the weight, the more attention the model allocates to that specific element in the input data. Finally, the attention mechanism computes a weighted sum of the values, using the attention weights. The resulting output is a combination of information from all input elements, with elements more relevant to the query receiving higher weight in the final representation [13,14].

The attention mechanism can be used in various ways (attention-gate [15], mixed-attention [16], among others in the medical field), with one prominent variant being self-attention. In self-attention, the query, key, and value all originate from the same input sequence. This allows the model to model relationships and dependencies between elements within the same sequence, making it particularly useful for tasks that involve capturing long-range dependencies and context [17–19].

These innovations have been crucial in advancing the state of the art in medical image, covering machine learning tasks such as classification, segmentation, synthesis (image or video), detection,

and captioning. By enhancing the model's ability to focus on relevant information and understand complex relationships within the data, the attention mechanism represents a significant step in the improvement of the quality and effectiveness of various deep learning applications in the medical field.

Generative Networks

Generative Adversarial Networks, or GANs, are a class of machine learning models introduced in 2014 [20] that excel at generating data, often in the form of images, but applicable to other data types like text or audio as well. GANs consist of two neural networks: a generator and a discriminator. The generator creates synthetic data from random noise and aims to produce data that is indistinguishable from real data, while the discriminator tries to distinguish between real and fake data. Through an adversarial training process, these networks compete, with the generator continually improving its ability to create realistic data and the discriminator enhancing its capacity to identify real from fake data.

GANs have revolutionized the field of data generation. They offer a highly effective way to create synthetic data that closely resembles real data. This is highly valuable, especially when dealing with limited datasets, as GANs can help augment training data for various machine learning tasks. For instance, in medical imaging, where obtaining large, diverse datasets can be challenging, GANs enable researchers to generate additional, realistic medical images for training diagnostic models, ultimately improving the accuracy of disease detection [21]. A recent study by Armanious et al. proposed a new framework called MedGAN [22] for medical image-to-image translation that operates on the image level in an end-to-end manner. MedGAN builds upon recent advances in the field of GANs by merging the adversarial framework with a new combination of non-adversarial losses. The framework utilizes a discriminator network as a trainable feature extractor which penalizes the discrepancy between the translated medical images and the desired modalities. Style-transfer losses are also utilized to match the textures and fine-structures of the desired target images to the translated images. Additionally, a new generator architecture, titled CasNet, enhances the sharpness of the translated medical outputs through progressive refinement via encoder-decoder pairs. MedGAN was applied on three different tasks: PET-CT translation, correction of MR motion artefacts, and PET image denoising. Perceptual analysis by radiologists and quantitative evaluations illustrate that MedGAN outperforms other existing translation approaches.

Generative Adversarial Networks (GANs) have been a promising tool in the field of medical image analysis [23], particularly in image-to-image translation. In a study be Skandarani et al. [24] conducted an empirical study on GANs for medical image synthesis. The results revealed that GANs are far from being equal as some are ill-suited for medical imaging applications while others are much better off. The top-performing GANs are capable of generating realistic-looking medical images by FID standards that can fool trained experts in a visual Turing test and comply to some metrics [25]. The introduction of these models into clinical practice has been cautious [26], but the advantages and performance that have been successively achieved with their development have allowed GANs to become a successful technology .

Along with GANs Variational Autoencoders (VAEs) are a popular technique for image generation. While both models are capable of generating images, they differ in their approach and training methodology. VAEs are a type of generative model that learns to encode the fundamental information of the input data into a latent space. The encoder network maps the input data to a latent space, which is then decoded by the decoder network to generate the output image. VAEs are trained using a probabilistic approach that maximizes the likelihood of the input data given the latent space. VAEs are better suited for applications that require probabilistic modeling, such as image reconstruction and denoising. This approach is capable of generating high-quality images but may suffer from blurry outputs [27–29].

Generative approaches are vital in machine learning for medical images due to their capacity to generate realistic data, drive innovation in image generation and manipulation, facilitate image-to-image translation, and open up creative opportunities for content generation across various domains.

Deep Learning Techniques and Performance Optimization

Medical imaging techniques are based on different physical principles, each with their benefits and limitations. The ability to deal with such diverse modalities is also an important aspect to be addressed by AI. In [30] a set of "tricks" are presented to improve the performance of deep learning models for multimodal image classification tasks. The authors start by emphasizing the increasing importance of multimodal image classification, which involves utilizing information from multiple modalities such as images, text, and other data sources. For this they also address the challenges specific to multimodal datasets, including data preprocessing, feature extraction, data imbalance, heterogeneity of modalities, data fusion, and model optimization. As defined by the authors, a "bag of tricks" or techniques can enhance the effectiveness of these models in handling multimodal data. These tricks can focus on the data, covering feature alignment, modality-specific preprocessing, and class balancing techniques, and also on the processing, using architectural modifications, training strategies, and regularization techniques. For the evaluation of such systems, benchmarking approaches are also presented and explored. These are valuable insights for researchers and practitioners working in the field of multimodal image classification.

Applications

Medical Image Analysis for Disease Detection and Diagnosis

Medical image analysis for disease detection and diagnosis is a rapidly evolving field that holds immense potential for improving healthcare outcomes. By harnessing advanced computational techniques and machine learning algorithms, medical professionals are now able to extract invaluable insights from various medical imaging modalities.

Artificial intelligence is an area where great progress has been observed and the number of techniques applicable to medical image processing has been increasing significantly. In this context of diversity, review articles where different techniques are presented and compared are useful. For example, in the area of Automated Retinal Disease Assessment (ARDA), AI can be used to help healthcare workers in the early detection, screening, diagnosis and grading of retinal diseases such as Diabetic Retinopathy (DR), Retinopathy of Prematurity (RoP), and Age-related Macular Degeneration (AMD), as shown in the comprehensive survey presented in [31]. The authors highlight the significance of medical image modalities, such as optical coherence tomography (OCT), fundus photography, and fluorescein angiography, in capturing detailed retinal images for diagnostic purposes and explain how AI can cope with these distinct information sources, either isolated or combined. The limitations and subjectivity of traditional manual examination and interpretation methods are emphasized, leading to the exploration of AI-based solutions. For this, an overview of the utilization of deep learning models is presented, and the most promising results in the detection and classification of retinal diseases, including age-related macular degeneration (AMD), diabetic retinopathy, and glaucoma are thoroughly covered. The role of AI in facilitating the analysis of large-scale retinal datasets and the development of computer-aided diagnostic systems is also highlighted. However, AI is not always a perfect solution, and the challenges and limitations of AI-based approaches are also covered, addressing issues related to data availability, model interpretability, and regulatory considerations. Given the significant interest in this field and the promising results that AI has yielded, other studies have also emerged to cover various topics related to eye image analysis [32,33].

Another area of great interest is brain imaging, whose techniques play a crucial role in understanding the intricate workings of the human brain and in the diagnose neurological disorders. Methods such as magnetic resonance imaging (MRI), functional MRI (fMRI), positron emission tomography (PET), or electroencephalography signals (EEG) provide valuable insights into brain structure, function, and connectivity. However, the analysis of these complex data, be it images or signals, requires sophisticated tools and expertise. Again, artificial intelligence (AI) comes into play. The synergy between brain imaging and AI has the potential to revolutionize neuroscience and improve patient care by unlocking deeper insights into the intricacies of the human brain. In [34], a

powerful combination of deep learning techniques and the Sine-Cosine Fitness Grey Wolf Optimization (SCFGWO) algorithm is used on the detection and classification of brain tumors. It addresses the importance of accurate tumor detection and classification as well as the associated challenges. Complexity and variability are tackled by convolutional neural networks (CNNs) that can automatically learn and extract relevant features for tumor analysis. In this case the SCFGWO algorithm is used to fine-tune the parameters of the CNN leading to an optimized performance. Metrics, such as accuracy, sensitivity, specificity, and F1-score are compared with other existing approaches to showcase the effectiveness and benefits of the proposed method in brain tumor detection and classification. The advantages and limitations of the proposed approach and the potential impact of the research in clinical practice are also mentioned.

Lung imaging has been a subject of extensive research interest [35,36], primarily due to the aggressive nature of lung cancer and its tendency to be detected at an advanced stage, leading to high mortality rates among cancer patients. In this context, accurate segmentation of lung fields in medical imaging plays a crucial role in the detection and analysis of lung diseases. In a recent study [37], the authors focused on segmenting lung fields in chest X-ray images using a combination of superpixel resizing and encoder-decoder segmentation networks. The study effectively addresses the challenges associated with lung field segmentation, including anatomical variations, image artifacts, and overlapping structures. It emphasizes the potential of deep learning techniques and the utilization of encoder-decoder architectures for semantic segmentation tasks. The proposed method, which combines superpixel resizing with an encoder-decoder segmentation network, demonstrates a high level of effectiveness compared to other approaches, as assessed using evaluation metrics such as the Dice similarity coefficient, Jaccard index, sensitivity, specificity, and accuracy.

More recently, the interest in lung imaging has been reinforced due to its importance in the diagnosis and monitoring of COVID-19 disease. In a notable study [38], the authors delve into the data-driven nature of AI and its need for high-quality data. They specifically focus on the generation of synthetic data, which involves creating artificial instances that closely mimic real data. In fact, using the proposed approach the synthetic images are nearly indistinguishable from read images, when compared using structural similarity index (SSIM), peak signal-to-noise ratio (PSNR), and Fréchet Inception Distance (FID). In this case, lung CT for COVID-19 diagnosis is used as an application example where this proposed approach has shown to be successful. The problem is tackled by means of a new regularization strategy, which refers to a technique used to prevent overfitting in ML models. This strategy does not require making significant changes to the underlying neural network architecture, making it easier to implement. Furthermore, the proposed method's efficacy extends beyond lung CT for COVID-19 diagnosis and can be easily adapted to other image types or imaging modalities. Consequently, future research endeavors can explore its applicability to diverse diseases and investigate its relevance to emerging AI topics, such as zero-shot or few-shot learning.

Breast cancer, the second most reported cancer worldwide, must be diagnosed as early as possible for a good prognostic. In this case, medical imaging is paramount for disease prevention and diagnosis. The effectiveness of an AI based approach is evaluated in [39]. The authors present a novel investigation that constructs and evaluates two computer-aided detection (CAD) systems for digital mammograms. The objective was to differentiate between malignant and benign breast lesions by employing two state-of-the-art approaches based on radiomics (with features such as intensity, shape, and texture) and deep transfer learning concepts and technologies (with deep features). Two CAD systems were trained and assessed using a sizable and diverse dataset of 3000 images. The findings of this study indicate that deep transfer learning can effectively extract meaningful features from medical images, even with limited training data, offering more discriminatory information than traditional handcrafted radiomics features. However, explainability, a desired characteristic in artificial intelligence and in medical decision systems in particular, must be further explored to fully unravel the mysteries of these "black-box" models.

Still concerning breast imaging, and addressing the typical high data needs of machine learning systems, a study was made to compare and optimize models using small datasets [40]. The article

discusses the challenges associated with limited data, such as overfitting and model generalization. Distinct CNN architectures, such as AlexNet, VGGNet, and ResNet, are trained using small datasets. The authors discuss strategies to mitigate these limitations, such as data augmentation techniques, transfer learning, and model regularization. With these premises, a multiclass classifier, based on the BI-RADS lexicon on the INBreast dataset [41], was developed. Compared with the literature, the model was able to improve the state-of-the-art results. This comes to reinforce that discriminative fine-tuning works well with state-of-the-art CNN models and it is possible to achieve excellent performance even on small datasets.

Liver cancer is the third most common cause of death from cancer worldwide [42] and its incidence has been growing. Again, the development of the disease is often asymptomatic, making screening and early detection crucial for a good prognosis. In [43], the authors focus on the segmentation of liver lesion in CT images of the LiTS dataset [44]. As a novelty, the paper proposes an intelligent decision system for segmenting liver and hepatic tumors by integrating four efficient neural networks (ResNet152, ResNeXt101, DenseNet201, and InceptionV3). These classifiers are independently operated and a final result is obtained by postprocess to eliminate artifacts. The obtained results were better than those obtained by the individual networks.

Imaging and Modeling Techniques for Surgical Planning and Intervention

Imaging and 3D modeling techniques, coupled with the power of artificial intelligence (AI), have revolutionized the field of surgical planning and intervention, offering numerous advantages to both patients and healthcare professionals. By leveraging the capabilities of AI, medical imaging data, such as CT scans and MRI images, can be transformed into detailed three-dimensional models that provide an enhanced understanding of a patient's anatomy. This newfound precision and depth of information allow surgeons to plan complex procedures with greater accuracy, improving patient outcomes and minimizing risks. Furthermore, AI-powered algorithms can analyze vast amounts of medical data, assisting surgeons in real-time during procedures, guiding them with valuable insights, and enabling personalized surgical interventions. Additionally, the integration of 3D printing technology with imaging and 3D modeling techniques further amplifies the advantages of surgical planning and intervention. With 3D printing, these intricate anatomical models can be translated into physical objects, allowing surgeons to hold and examine patient-specific replicas before the actual procedure. This tangible representation aids in comprehending complex anatomical structures, identifying potential challenges, and refining surgical strategies. Surgeons can also utilize 3D-printed surgical guides and implants, customized to fit each patient's unique anatomy, thereby enhancing precision and reducing operative time.

These benefits are described and explored in [45]. The authors explore the operative workflow involved in the process of creating 3D-printed models of the heart using computed tomography (CT) scans. The authors begin by emphasizing the importance of accurate anatomical models in surgical planning, particularly in complex cardiac cases. They also discuss how 3D printing technology has gained prominence in the medical field, allowing for the creation of patient-specific anatomical models. In their developments they thoroughly describe the operative workflow for generating 3D-printed heart models. Throughout the process the challenges and limitations of the operative workflow from CT to 3D printing of the heart are covered. It is also discussed factors such as cost, time, expertise required, and the need for validation studies to ensure the accuracy and reliability of the printed models.

A similar topic is presented in [46]. Here the authors focus specifically on coronary artery bypass graft (CABG) procedures and describe the feasibility of using a 3D modelling and printing process to create surgical guides, contributing to the success of the surgery, and enhancing patient outcomes. In this paper the authors also discuss the choice of materials for the 3D-printed guide, considering biocompatibility and sterility requirements. In addition, a case study that demonstrates the successful application of the workflow in a real clinical scenario is presented.

The combination of AI-driven imaging, 3D modeling, and 3D printing technologies revolutionizes surgical planning and intervention, empowering healthcare professionals with

unparalleled tools to improve patient outcomes, create personalized solutions, and redefine the future of surgical practice. These advancements in imaging and 3D modeling techniques, driven by AI, are driving a new era of surgical precision and innovation in healthcare.

Image and Model Enhancement for Improved Analysis

Decision making and diagnosis are important purposes for clinical applications, but AI can also play an important role in other applications of the clinical process. For example, in [47] the authors focus on the application of colorization techniques to medical images, with the goal of enhancing the visual interpretation and analysis by adding chromatic information. The authors highlight the importance of color in medical imaging as it can provide additional information for diagnosis, treatment planning, and educational purposes. They also address the challenges associated with medical image colorization, including the large variability in image characteristics and the need for robust and accurate colorization methods. The proposed method utilizes a spatial mask-guided colorization with generative adversarial network (SMCGAN) technique to focus on relevant regions of the medical image while preserving important structural information during the process. The evaluation was based on a dataset from the Visible Human Project [48] and from the prostate dataset NCI-ISBI 2013 [49]. With the presented experimental setup and evaluation metrics used for performance assessment, the proposed technique was able to outperform the state-of-the-art GAN-based image colorization approaches with an average improvement of 8.48% in the peak signal-to-noise ratio (PSNR) metric.

In complex healthcare scenarios, it is crucial for clinicians and practitioners to understand the reasoning behind AI models' predictions and recommendations. Explainable AI (XAI) plays a pivotal role in the domain of medical imaging techniques for decision support, where transparency and interpretability are paramount. In [50], the authors address the problem of nuclei detection in histopathology images, which is a crucial task in digital pathology for diagnosing and studying diseases. They specifically propose a technique called NDG-CAM (Nuclei Detection in Histopathology Images with Semantic Segmentation Networks and Grad-CAM). Grad-CAM (Gradient-weighted Class Activation Mapping) [51] is a technique used in computer vision and deep learning to visualize and interpret the regions of an image that are most influential in the prediction made by a convolutional neural network. Hence, in the proposed methodology, the semantic segmentation network aims to accurately segment the nuclei regions in histopathology images, while Grad-CAM helps visualize the important regions that contribute to the model's predictions, helping to improve the accuracy and interpretability of nuclei detection. The authors compare the performance of their method with other existing nuclei detection methods, demonstrating that NDG-CAM achieves improved accuracy while providing interpretable results.

Still with the purpose of making AI providing human understandable results, the authors in [52] focus on the development of an open-source COVID-19 CT dataset that includes automatic lung tissue classification for radiomics analysis. The challenges associated with COVID-19 research, including the importance of large-scale datasets and efficient analysis methods are covered. The potential of radiomics, which involves extracting quantitative features from medical images, in aiding COVID-19 diagnosis, prognosis, and treatment planning, are also mentioned. The proposed dataset consists of CT scans from COVID-19 patients, which are annotated with labels indicating different lung tissue regions, such as ground-glass opacities, consolidations, and normal lung tissue.

Novel machine learning techniques are also being used to enhance the resolution and quality of medical images. These techniques aim to recover fine details and structures that are lost or blurred in low-resolution images, which can improve the diagnosis and treatment of various diseases. One of the novel machine learning techniques is based on GANs. For example, in Bing at al. [53] propose the use of an improved squeeze and excitation block that selectively amplifies the important features and suppresses the non-important ones in the feature maps. A simplified EDSR (enhanced deep super-resolution) model to generate high-resolution images from low-resolution inputs is also proposed, along with a new fusion loss function. The proposed method was evaluated on public medical image datasets and compared with state-of-the-art deep learning-based methods such as

SRGAN, EDSR, VDSR, and D-DBPN. The results show that the proposed method achieves better visual quality and preserves more details, especially for high upscaling factors.

Vision transformers, with their ability to treat images as sequences of tokens and learn global dependencies among them, can capture long-range and complex patterns in images, which can benefit super-resolution tasks. Zhu et al. [54] propose the use of vision transformers with residual dense connections and local feature fusion. This method proposes an efficient vision transformer architecture that can achieve high-quality single-image super-resolution for various medical modalities, such as MRI, CT, and X-ray. The key idea is to use residual dense blocks to enhance the feature extraction and representation capabilities of the vision transformer, and to use local feature fusion to combine the low-level and high-level features for better reconstruction. Moreover, this method also introduces a novel perceptual loss function that incorporates prior knowledge of medical image segmentation to improve the image quality of desired aspects, such as edges, textures, and organs. In another work, Wei et al. [55] propose to adapt the Swin transformer, which is a hierarchical vision transformer that uses shifted windows to capture local and global information, to the task of automatic medical image segmentation. The high-resolution Swin transformer uses a U-Net-like architecture that consists of an encoder and a decoder. The encoder converts the high-resolution input image into low-resolution feature maps using a sequence of Swin transformer blocks, and the decoder gradually generates high-resolution representations from low-resolution feature maps using upsampling and skip connections. The high-resolution Swin transformer can achieve state-of-the-art results on several medical image segmentation datasets, such as BraTS, LiTS, and KiTS.

In addition, perceptual loss functions, can be used to further enhance generative techniques. These are designed to measure the similarity between images in terms of their semantic content and visual quality, rather than their pixel-wise differences. Perceptual loss functions can be derived from pre-trained models, such as image classifiers or segmenters, that capture high-level features of images. By optimizing the perceptual loss functions, the super-resolution models can generate images that preserve the important structures and details of the original images, while avoiding artifacts and distortions [53,56].

Medical images often suffer from noise, artifacts, and limited resolution due to the physical constraints of the imaging devices. Therefore, developing effective and efficient methods for medical image super-resolution is a challenging and promising research topic, searching to obtain previously unachievable details and resolution [57,58].

Conclusion

Cutting-edge techniques that come to push the limits of current knowledge have been covered in this editorial. For those focused on the AI aspects of technology, evolutions have been reported in all stages of the medical imaging machine learning pipeline. As mentioned, the data-driven nature of these techniques requires that a special attention is given to it. Beyond high quality dataset [52], attention can be given to the generation of more data [38] and better data [37]. The training process can be optimized to deal with small datasets [40] or techniques can be used to improve the parameter optimization process [34]. To better understand the models' operating we can use explainable AI techniques [50]. We can also focus on generating a better output by combining several classifiers [43] or by adding useful information such as colors [47]. Many of the involved challenges throughout the process can address using a "bag-of-tricks" [30]. These advantages of using AI in medical imaging applications is explored in [31] and its ability to perform better that feature-based approach is covered in [39]. Finally, applications of AI to 3D modeling and physical object generation are covered in [45,46].

The field of medical imaging and AI is evolving rapidly, driven by ongoing research and technological advancements. Researchers are continuously exploring novel algorithms, architectures, and methodologies to further enhance the capabilities of AI in medical imaging. Additionally, collaborations between clinicians, computer scientists, and industry professionals are vital in translating research findings into practical applications that can benefit patients worldwide.

In conclusion, the fusion of medical imaging and AI has brought about significant advancements in healthcare. From early disease detection to personalized diagnosis and therapy, AI has demonstrated its potential to revolutionize medical practice. By harnessing the power of AI, medical professionals can leverage the wealth of information contained within medical images to provide accurate diagnoses, tailor treatment plans, and improve patient outcomes. As technology continues to advance, we can expect even more groundbreaking innovations that will further transform the landscape of medical imaging and AI in the years to come.

References

1. Roodschild, M.; Gotay Sardiñas, J.; Will, A. A New Approach for the Vanishing Gradient Problem on Sigmoid Activation. *Prog. in Artif. Intell.* **2020**, *9*, 351–360, doi:10.1007/s13748-020-00218-y.

2. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the Proceedings of the 27th International Conference on International Conference on Machine Learning; Omnipress: Madison, WI, USA, June 21 2010; pp. 807–814.

3. Agarap, A.F. Deep Learning Using Rectified Linear Units (ReLU). *arXiv e-prints* 2018.

4. Deng, L. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine* **2012**, *29*, 141–142, doi:10.1109/MSP.2012.2211477.

5. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444, doi:10.1038/nature14539.

6. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent Advances in Convolutional Neural Networks. *Pattern Recognition* **2018**, *77*, 354–377, doi:10.1016/j.patcog.2017.10.013.

7. O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks 2015.

8. Yandex, A.B.; Lempitsky, V. Aggregating Local Deep Features for Image Retrieval. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); December 2015; pp. 1269–1277.

9. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90, doi:10.1145/3065386.

10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems; Curran Associates Inc.: Red Hook, NY, USA, dezembro 2017; pp. 6000–6010.

11. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale Available online: https://arxiv.org/abs/2010.11929v2.

12. He, K.; Gan, C.; Li, Z.; Rekik, I.; Yin, Z.; Ji, W.; Gao, Y.; Wang, Q.; Zhang, J.; Shen, D. Transformers in Medical Image Analysis. *Intelligent Medicine* **2023**, *3*, 59–78, doi:10.1016/j.imed.2022.07.002.

13. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks 2019.

14. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going Deeper with Image Transformers 2021.

15. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas 2018.

16. Nwoye, C.I.; Yu, T.; Gonzalez, C.; Seeliger, B.; Mascagni, P.; Mutter, D.; Marescaux, J.; Padoy, N. Rendezvous: Attention Mechanisms for the Recognition of Surgical Action Triplets in Endoscopic Videos. *Medical Image Analysis* **2022**, *78*, 102433, doi:10.1016/j.media.2022.102433.

17. Sinha, A.; Dolz, J. Multi-Scale Self-Guided Attention for Medical Image Segmentation 2020.

18. Rao, A.; Park, J.; Woo, S.; Lee, J.-Y.; Aalami, O. Studying the Effects of Self-Attention for Medical Image Analysis 2021.

19. You, H.; Wang, J.; Ma, R.; Chen, Y.; Li, L.; Song, C.; Dong, Z.; Feng, S.; Zhou, X. Clinical Interpretability of Deep Learning for Predicting Microvascular Invasion in Hepatocellular Carcinoma by Using Attention Mechanism. *Bioengineering* **2023**, *10*, 948, doi:10.3390/bioengineering10080948.

20. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks 2014.

21. Platscher, M.; Zopes, J.; Federau, C. Image Translation for Medical Image Generation: Ischemic Stroke Lesion Segmentation. *Biomedical Signal Processing and Control* **2022**, *72*, 103283, doi:10.1016/j.bspc.2021.103283.

22. Armanious, K.; Jiang, C.; Fischer, M.; Küstner, T.; Hepp, T.; Nikolaou, K.; Gatidis, S.; Yang, B. MedGAN: Medical Image Translation Using GANs. *Computerized Medical Imaging and Graphics* **2020**, *79*, 101684, doi:10.1016/j.compmedimag.2019.101684.

23. Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for Medical Image Analysis. *Artificial Intelligence in Medicine* **2020**, *109*, 101938, doi:10.1016/j.artmed.2020.101938.

24. Skandarani, Y.; Jodoin, P.-M.; Lalande, A. GANs for Medical Image Synthesis: An Empirical Study. *Journal of Imaging* **2023**, *9*, 69, doi:10.3390/jimaging9030069.

25. Skandarani, Y.; Jodoin, P.-M.; Lalande, A. GANs for Medical Image Synthesis: An Empirical Study 2021.

26. Wang, T.; Lei, Y.; Fu, Y.; Wynne, J.F.; Curran, W.J.; Liu, T.; Yang, X. A Review on Medical Imaging Synthesis Using Deep Learning and Its Clinical Applications. *Journal of Applied Clinical Medical Physics* **2021**, *22*, 11–36, doi:10.1002/acm2.13121.

27. Ehrhardt, J.; Wilms, M. Chapter 8 - Autoencoders and Variational Autoencoders in Medical Image Analysis. In *Biomedical Image Synthesis and Simulation*; Burgos, N., Svoboda, D., Eds.; The MICCAI Society book Series; Academic Press, 2022; pp. 129–162 ISBN 978-0-12-824349-7.

28. Kebaili, A.; Lapuyade-Lahorgue, J.; Ruan, S. Deep Learning Approaches for Data Augmentation in Medical Imaging: A Review. *Journal of Imaging* **2023**, *9*, 81, doi:10.3390/jimaging9040081.

29. Elbattah, M.; Loughnane, C.; Guérin, J.-L.; Carette, R.; Cilia, F.; Dequen, G. Variational Autoencoder for Image-Based Augmentation of Eye-Tracking Data. *Journal of Imaging* **2021**, *7*, 83, doi:10.3390/jimaging7050083.

30. Adeshina, S.A.; Adedigba, A.P. Bag of Tricks for Improving Deep Learning Performance on Multimodal Image Classification. *Bioengineering* **2022**, *9*, 312, doi:10.3390/bioengineering9070312.

31. Saleh, G.A.; Batouty, N.M.; Haggag, S.; Elnakib, A.; Khalifa, F.; Taher, F.; Mohamed, M.A.; Farag, R.; Sandhu, H.; Sewelam, A.; et al. The Role of Medical Image Modalities and AI in the Early Detection, Diagnosis and Grading of Retinal Diseases: A Survey. *Bioengineering* **2022**, *9*, 366, doi:10.3390/bioengineering9080366.

32. Han, J.-H. Artificial Intelligence in Eye Disease: Recent Developments, Applications, and Surveys. *Diagnostics* **2022**, *12*, 1927, doi:10.3390/diagnostics12081927.

33. Daich Varela, M.; Sen, S.; De Guimaraes, T.A.C.; Kabiri, N.; Pontikos, N.; Balaskas, K.; Michaelides, M. Artificial Intelligence in Retinal Disease: Clinical Application, Challenges, and Future Directions. *Graefes Arch Clin Exp Ophthalmol* **2023**, doi:10.1007/s00417-023-06052-x.

34. Zain Eldin, H.; Gamel, S.A.; El-Kenawy, E.-S.M.; Alharbi, A.H.; Khafaga, D.S.; Ibrahim, A.; Talaat, F.M. Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. *Bioengineering* **2023**, *10*, 18, doi:10.3390/bioengineering10010018.

35. Forte, G.C.; Altmayer, S.; Silva, R.F.; Stefani, M.T.; Libermann, L.L.; Cavion, C.C.; Youssef, A.; Forghani, R.; King, J.; Mohamed, T.-L.; et al. Deep Learning Algorithms for Diagnosis of Lung Cancer: A Systematic Review and Meta-Analysis. *Cancers* **2022**, *14*, 3856, doi:10.3390/cancers14163856.

36. Hunger, T.; Wanka-Pail, E.; Brix, G.; Griebel, J. Lung Cancer Screening with Low-Dose CT in Smokers: A Systematic Review and Meta-Analysis. *Diagnostics* **2021**, *11*, 1040, doi:10.3390/diagnostics11061040.

37. Lee, C.-C.; So, E.C.; Saidy, L.; Wang, M.-J. Lung Field Segmentation in Chest X-Ray Images Using Superpixel Resizing and Encoder–Decoder Segmentation Networks. *Bioengineering* **2022**, *9*, 351, doi:10.3390/bioengineering9080351.

38. Lee, K.W.; Chin, R.K.Y. Diverse COVID-19 CT Image-to-Image Translation with Stacked Residual Dropout. *Bioengineering* **2022**, *9*, 698, doi:10.3390/bioengineering9110698.

39. Danala, G.; Maryada, S.K.; Islam, W.; Faiz, R.; Jones, M.; Qiu, Y.; Zheng, B. A Comparison of Computer-Aided Diagnosis Schemes Optimized Using Radiomics and Deep Transfer Learning Methods. *Bioengineering* **2022**, *9*, 256, doi:10.3390/bioengineering9060256.

40. Adedigba, A.P.; Adeshina, S.A.; Aibinu, A.M. Performance Evaluation of Deep Learning Models on Mammogram Classification Using Small Dataset. *Bioengineering* **2022**, *9*, 161, doi:10.3390/bioengineering9040161.

41. Zebari, D.A.; Ibrahim, D.A.; Zeebaree, D.Q.; Mohammed, M.A.; Haron, H.; Zebari, N.A.; Damaševičius, R.; Maskeliūnas, R. Breast Cancer Detection Using Mammogram Images with Improved Multi-Fractal Dimension Approach and Feature Fusion. *Applied Sciences* **2021**, *11*, 12122, doi:10.3390/app112412122.

13

42.   World Health Organization Cancer Key Facts Available online: https://www.who.int/news-room/fact-sheets/detail/cancer (accessed on 16 June 2023).

43.   Popescu, D.; Stanciulescu, A.; Pomohaci, M.D.; Ichim, L. Decision Support System for Liver Lesion Segmentation Based on Advanced Convolutional Neural Network Architectures. *Bioengineering* **2022**, *9*, 467, doi:10.3390/bioengineering9090467.

44.   Bilic, P.; Christ, P.; Li, H.B.; Vorontsov, E.; Ben-Cohen, A.; Kaissis, G.; Szeskin, A.; Jacobs, C.; Mamani, G.E.H.; Chartrand, G.; et al. The Liver Tumor Segmentation Benchmark (LiTS). *Medical Image Analysis* **2023**, *84*, 102680, doi:10.1016/j.media.2022.102680.

45.   Bertolini, M.; Rossoni, M.; Colombo, G. Operative Workflow from CT to 3D Printing of the Heart: Opportunities and Challenges. *Bioengineering* **2021**, *8*, 130, doi:10.3390/bioengineering8100130.

46.   Cappello, I.A.; Candelari, M.; Pannone, L.; Monaco, C.; Bori, E.; Talevi, G.; Ramak, R.; La Meir, M.; Gharaviri, A.; Chierchia, G.B.; et al. 3D Printed Surgical Guide for Coronary Artery Bypass Graft: Workflow from Computed Tomography to Prototype. *Bioengineering* **2022**, *9*, 179, doi:10.3390/bioengineering9050179.

47.   Zhang, Z.; Li, Y.; Shin, B.-S. Robust Medical Image Colorization with Spatial Mask-Guided Generative Adversarial Network. *Bioengineering* **2022**, *9*, 721, doi:10.3390/bioengineering9120721.

48.   National Library of Medicine Visible Human Project Available online: https://www.nlm.nih.gov/research/visible/visible_human.html (accessed on 16 June 2023).

49.   Bloch, B.N.; Madabhushi, A.; Huisman, H.; Freymann, J.; Kirby, J.; Grauer, M.; Enquobahrie, A.; Jaffe, C.; Clarke, L.; Farahani, K. NCI-ISBI 2013 Challenge: Automated Segmentation of Prostate Structures (ISBI-MR-Prostate-2013) 2015.

50.   Altini, N.; Brunetti, A.; Puro, E.; Taccogna, M.G.; Saponaro, C.; Zito, F.A.; De Summa, S.; Bevilacqua, V. NDG-CAM: Nuclei Detection in Histopathology Images with Semantic Segmentation Networks and Grad-CAM. *Bioengineering* **2022**, *9*, 475, doi:10.3390/bioengineering9090475.

51.   Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Int J Comput Vis* **2020**, *128*, 336–359, doi:10.1007/s11263-019-01228-7.

52.   Zaffino, P.; Marzullo, A.; Moccia, S.; Calimeri, F.; De Momi, E.; Bertucci, B.; Arcuri, P.P.; Spadea, M.F. An Open-Source COVID-19 CT Dataset with Automatic Lung Tissue Classification for Radiomics. *Bioengineering* **2021**, *8*, 26, doi:10.3390/bioengineering8020026.

53.   Bing, X.; Zhang, W.; Zheng, L.; Zhang, Y. Medical Image Super Resolution Using Improved Generative Adversarial Networks. *IEEE Access* **2019**, *7*, 145030–145038, doi:10.1109/ACCESS.2019.2944862.

54.   Zhu, J.; Yang, G.; Lio, P. A Residual Dense Vision Transformer for Medical Image Super-Resolution with Segmentation-Based Perceptual Loss Fine-Tuning 2023.

55.   Wei, C.; Ren, S.; Guo, K.; Hu, H.; Liang, J. High-Resolution Swin Transformer for Automatic Medical Image Segmentation. *Sensors* **2023**, *23*, 3420, doi:10.3390/s23073420.

56.   Zhang, K.; Hu, H.; Philbrick, K.; Conte, G.M.; Sobek, J.D.; Rouzrokh, P.; Erickson, B.J. SOUP-GAN: Super-Resolution MRI Using Generative Adversarial Networks. *Tomography* **2022**, *8*, 905–919, doi:10.3390/tomography8020073.

57.   Yang, H.; Wang, Z.; Liu, X.; Li, C.; Xin, J.; Wang, Z. Deep Learning in Medical Image Super Resolution: A Review. *Appl Intell* **2023**, *53*, 20891–20916, doi:10.1007/s10489-023-04566-9.

58.   Chen, C.; Wang, Y.; Zhang, N.; Zhang, Y.; Zhao, Z. A Review of Hyperspectral Image Super-Resolution Based on Deep Learning. *Remote Sensing* **2023**, *15*, 2853, doi:10.3390/rs15112853.