Preprints.org

Article

Not peer-reviewed version

Integrating Multimodal Data with Large Foundation Models in Healthcare

Hyunwoo Choi , Eunji Kang , Daehyun Song , Gyeong Jung *

Posted Date: 7 August 2025

doi: 10.20944/preprints202508.0498.v1

Keywords: large foundation models; medical analysis; multimodal learning; clinical decision support; interpretability; fairness in AI; privacy-preserving machine learning; electronic health records; medical imaging; genomics; ethical AI in healthcare

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

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.

Article

Integrating Multimodal Data with Large Foundation Models in Healthcare

Hyunwoo Choi, Eunji Kang, Daehyun Song and Gyeong Jung *

Ulsan National Institute of Science and Technology (UNIST), 50 UNIST-gil, Ulsan 44919, Ulsan, South Korea
* Correspondence: gyeong.jung@unist.ac.kr

Abstract

Large Foundation Models (LFMs) have rapidly become a transformative force in the field of medical analysis, enabling unprecedented advancements in the interpretation, integration, and reasoning over complex and heterogeneous healthcare data. By leveraging massive datasets encompassing medical images, electronic health records, clinical notes, and genomic sequences, LFMs facilitate a unified framework that transcends the limitations of traditional, narrowly scoped machine learning models. This survey comprehensively reviews the state-of-the-art developments in LFMs tailored for medical applications, elucidating their architectural paradigms, training methodologies, and deployment strategies. We begin by examining the foundational building blocks that enable LFMs to handle multimodal data through modality-specific encoders and sophisticated fusion mechanisms, emphasizing the critical role of cross-attention and contrastive learning techniques in producing semantically aligned latent representations. The survey further explores practical applications, including diagnosis prediction, automated report generation, treatment recommendation, and personalized medicine, highlighting how LFMs enhance clinical decision-making by providing richer contextual understanding and reasoning capabilities. Despite their promise, the integration of LFMs into clinical practice faces significant challenges related to interpretability, data privacy, fairness, and scalability. We delve into these issues in depth, discussing the implications of model opacity, bias amplification, regulatory constraints, and the scarcity of labeled medical data. Cutting-edge solutions such as federated learning, self-supervised pretraining, and fairness-aware algorithms are examined as potential mitigations. Ethical considerations are addressed to ensure responsible AI deployment that safeguards patient rights, promotes equitable healthcare, and fosters trust among medical professionals and patients alike. Finally, the survey outlines future research opportunities, including advances in efficient training paradigms, improved model transparency, robust multimodal integration, and privacy-preserving technologies. The discussion underscores the necessity of interdisciplinary collaboration and human–AI partnership to realize the full potential of LFMs in improving health outcomes globally. Through this extensive analysis, we aim to provide researchers, clinicians, and policymakers with a holistic understanding of LFMs' capabilities, challenges, and prospects in the rapidly evolving landscape of medical artificial intelligence.

Keywords: large foundation models; medical analysis; multimodal learning; clinical decision support; interpretability; fairness in AI; privacy-preserving machine learning; electronic health records; medical imaging; genomics; ethical AI in healthcare

1. Introduction

The intersection of artificial intelligence (AI) and healthcare has witnessed transformative advancements over the past decade, driven in large part by the development of increasingly sophisticated machine learning models. Among these, Large Foundation Models (LFMs)—a class of deep learning architectures characterized by their scale, general-purpose nature, and pretraining on massive datasets—have emerged as a groundbreaking paradigm with the potential to revolutionize medical

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

analysis. These models, which include but are not limited to large language models (LLMs) such as GPT, BERT, PaLM, and multimodal architectures like CLIP, Flamingo, and Med-PaLM, have demonstrated exceptional capabilities across a variety of downstream tasks [1]. Originally designed for general-purpose natural language understanding and generation, LFMs have since been adapted and fine-tuned for an increasingly diverse set of clinical applications, ranging from medical image interpretation to patient record summarization, clinical decision support, biomedical literature mining, and even drug discovery [2]. The core principle behind LFMs is the use of transfer learning at an unprecedented scale. By training on broad and often heterogeneous corpora—spanning general web data, scientific literature, and structured clinical databases—these models acquire a robust and flexible representation of language, vision, or multimodal inputs [3]. Once pretrained, LFMs can be fine-tuned on smaller, task-specific datasets to achieve state-of-the-art performance on specialized medical tasks, often with significantly fewer labeled examples than traditional supervised approaches require [4]. This shift from task-specific, data-hungry models to generalist, data-efficient architectures has profound implications for medical AI, where high-quality labeled data is often scarce, expensive, or ethically challenging to obtain [5]. Moreover, the architecture of LFMs facilitates capabilities such as few-shot learning, zero-shot inference, and in-context learning, which allow models to generalize beyond their training distribution and perform novel tasks without explicit retraining [6]. In clinical settings, where variability in disease presentation, patient demographics, and diagnostic modalities is the norm rather than the exception, such generalization is not merely advantageous—it is essential [7]. LFMs also enable continual learning and domain adaptation, allowing them to remain current with evolving medical knowledge and guidelines, a critical feature in fast-moving domains such as oncology or infectious disease [8]. The promise of LFMs in medicine is further enhanced by the development of domain-specific variants trained on curated medical corpora [9]. Models such as BioBERT, ClinicalBERT, PubMedGPT, and Med-PaLM represent a growing ecosystem of biomedical LFMs that are designed to meet the stringent accuracy, interpretability, and safety requirements of clinical environments [10]. These models are increasingly being integrated into medical workflows, powering tools that assist radiologists in detecting anomalies, help clinicians interpret unstructured electronic health records (EHRs), and support researchers in synthesizing vast quantities of biomedical literature [11]. Despite these promising advances, the deployment of LFMs in real-world medical settings raises numerous technical, ethical, and regulatory challenges. Issues of bias, fairness, explainability, data privacy, and model robustness are particularly salient in healthcare, where decisions can have life-altering consequences. Furthermore, the black-box nature of many LFMs complicates their integration into clinical workflows that demand transparency, accountability, and rigorous validation. The computational cost and environmental footprint associated with training and maintaining these large-scale models also pose sustainability concerns, especially for resource-constrained healthcare systems [12]. In this survey, we aim to provide a comprehensive overview of the landscape of medical analysis using Large Foundation Models. We begin by categorizing the types of LFMs used in the biomedical domain and describing their architectural underpinnings [13]. We then examine the breadth of applications enabled by these models across various medical subfields, including radiology, pathology, genomics, pharmacology, and clinical informatics [14]. Special attention is given to the challenges and limitations associated with their deployment, with a discussion of ongoing research efforts aimed at improving interpretability, mitigating bias, and ensuring regulatory compliance. Finally, we explore emerging trends and future directions, including the integration of multimodal data sources, the use of federated learning and privacy-preserving techniques, and the development of next-generation foundation models tailored specifically for healthcare [15]. Through this survey, we seek to elucidate the transformative potential of LFMs in medicine, while also highlighting the critical considerations that must guide their responsible and equitable adoption [16]. As the field of medical AI moves toward increasingly general and powerful models, a clear understanding of the capabilities, limitations, and ethical implications of LFMs is essential for clinicians, researchers, developers, and policymakers alike [17]. We hope this work will serve as a foundational reference for the growing community at the

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

intersection of AI and medicine, and as a catalyst for interdisciplinary collaboration in the design of safe, effective, and human-centered medical AI systems.

2. Theoretical Foundations of Large Foundation Models in Medical Analysis

Large Foundation Models (LFMs), particularly those rooted in transformer architectures, are fundamentally built upon the framework of representation learning, statistical inference, and optimization over high-dimensional function spaces [18]. In the context of medical analysis, these models serve as approximators of complex, often nonlinear mappings between input modalities (e.g., clinical text, radiological images, genetic sequences) and outputs that may include diagnostic labels, treatment recommendations, or patient outcome probabilities. The goal is to learn a parameterized function $f_{\boldsymbol{\theta}} : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ denotes a potentially multimodal input space and $\mathcal{Y}$ the output task space, with parameters $\boldsymbol{\theta} \in \mathbb{R}^d$ optimized to minimize a task-specific loss function [19]. The training of LFMs typically begins with a self-supervised pretraining phase on a large corpus of unlabeled data [20]. Formally, let $\mathcal{D}_{\text{pre}} = \{x_i\}_{i=1}^{N}$ denote the pretraining dataset, where each $x_i \in \mathcal{X}$ may represent a tokenized medical sentence, a patch of a medical image, or a segment of structured EHR data [21]. A common pretraining objective is the masked language modeling (MLM) loss, defined as:

$$\mathcal{L}_{\text{MLM}}(\boldsymbol{\theta}) = -\mathbb{E}_{x \sim \mathcal{D}_{\text{pre}}} \sum_{t \in \mathcal{M}(x)} \log p_{\boldsymbol{\theta}}(x_t \mid x_{\setminus t}),$$

where $\mathcal{M}(x)$ denotes the set of masked positions in x, and $x_{\setminus t}$ is the input with the t-th token removed [22]. This objective encourages the model to learn bidirectional contextual representations, which can be transferred to downstream clinical tasks [23]. In multimodal settings, LFMs are trained to jointly model heterogeneous medical data modalities. Let $x = (x^{(1)}, x^{(2)}, \ldots, x^{(m)})$ be an input consisting of m modalities, such as $x^{(1)}$ for radiology images, $x^{(2)}$ for free-text reports, and $x^{(3)}$ for structured lab values [24]. The multimodal encoder seeks to learn a joint embedding space $\mathcal{H} \subset \mathbb{R}^k$ such that:

$$\mathbf{h} = \phi(x^{(1)}, x^{(2)}, \ldots, x^{(m)}; \boldsymbol{\theta}),$$

where ϕ denotes the multimodal fusion function, typically instantiated as a cross-attention transformer or a contrastive encoder-decoder architecture. A popular objective in this context is the contrastive loss:

$$\mathcal{L}_{\text{contrast}}(\boldsymbol{\theta}) = -\log \frac{\exp(\text{sim}(\phi(x^{(i)}), \phi(x^{(j)}))/\tau)}{\sum_{k=1}^{K} \exp(\text{sim}(\phi(x^{(i)}), \phi(x^{(k)}))/\tau)},$$

where $\text{sim}(\cdot, \cdot)$ denotes cosine similarity, τ is a temperature parameter, and K includes positive and negative pairs. In clinical settings, this may correspond to aligning a chest X-ray with its corresponding radiology report while distinguishing it from non-matching report-image pairs [25]. Once pretrained, the model is adapted to downstream medical tasks via fine-tuning [26]. Let $\mathcal{D}_{\text{fine}} = \{(x_i, y_i)\}_{i=1}^{M}$ be a task-specific labeled dataset, such as diagnosis codes for EHR entries. The fine-tuning objective is typically:

$$\mathcal{L}_{\text{task}}(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \ell(f_{\boldsymbol{\theta}}(x_i), y_i),$$

where ℓ is a task-dependent loss function, such as cross-entropy for classification or mean squared error for regression. Notably, the parameter vector $\boldsymbol{\theta}$ is often initialized from the pretrained weights, enabling improved generalization in low-data regimes—especially critical in domains like rare disease diagnosis, where labeled samples are limited. Moreover, the probabilistic interpretation of LFMs allows for uncertainty quantification, which is indispensable in high-stakes medical applications [27].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

Let $f_{\boldsymbol{\theta}}(x)$ denote the predictive distribution over outcomes [28]. The model's epistemic uncertainty can be estimated via Bayesian approximations, such as Monte Carlo dropout:

$$\mathbb{V}[f_{\boldsymbol{\theta}}(x)] \approx \frac{1}{T}\sum_{t=1}^{T} f_{\boldsymbol{\theta}_t}(x)^2 - \left(\frac{1}{T}\sum_{t=1}^{T} f_{\boldsymbol{\theta}_t}(x)\right)^2,$$

where $\boldsymbol{\theta}_t$ are sampled from a variational posterior over model weights. This estimation guides clinical users by signaling when a model is uncertain and may need human review or additional testing [29]. A further theoretical perspective comes from information theory and mutual information. The objective of pretraining can be viewed as maximizing the mutual information between inputs and learned representations:

$$\max_{\boldsymbol{\theta}} I(X; Z_{\boldsymbol{\theta}}),$$

where $Z_{\boldsymbol{\theta}}$ denotes the latent representation induced by the model [30]. In a medical context, maximizing this quantity ensures that critical diagnostic information (e.g., lesion presence, temporal changes in symptoms, biomarker anomalies) is preserved in the learned representations, thereby enhancing interpretability and utility in clinical decision-making [31]. Collectively, these mathematical formulations provide a rigorous foundation for understanding and advancing LFMs in medicine [32]. By leveraging self-supervised learning, multimodal fusion, probabilistic modeling, and information-theoretic principles, LFMs offer a powerful and generalizable framework for tackling a wide array of clinical challenges [33]. The next section explores how these models are operationalized across different medical domains and the empirical results that support their adoption.

3. Applications of Large Foundation Models in Medical Domains

The deployment of Large Foundation Models (LFMs) across diverse medical domains has resulted in substantial improvements in both predictive accuracy and workflow efficiency. These models, when fine-tuned or prompted appropriately, can be applied to a wide spectrum of healthcare challenges, including medical imaging interpretation, clinical text summarization, patient risk stratification, genomics, and personalized treatment planning [34]. Their generalizability enables transfer across tasks and modalities, reducing the need for developing bespoke solutions for each individual application.

Table 1. Representative Applications of Large Foundation Models in Medical Domains

Domain	Example LFM Models	Application Task	Performance
Radiology	CLIP, BioViL, GLoRIA, MedCLIP	Zero-shot or few-shot classification of pathologies in chest X-rays and CT scans [35]	Outperforms supervised baselines in low-label regimes; aligns image and report semantics effectively
Clinical NLP	BioBERT, ClinicalBERT, PubMedGPT, GatorTron	Named entity recognition (NER), relation extraction, clinical note summarization	Achieves state-of-the-art results on multiple clinical NLP benchmarks such as i2b2 and MIMIC-III
Health Records	RETAIN, Med-PaLM, BEHRT, TransformerEHR	Disease progression modeling, risk prediction, medication recommendation	Improves AUC/ROC and calibration in longitudinal patient modeling; captures temporal dependencies
Pathology	PaLM-E, HEAL, Vision Transformers	Whole slide image (WSI) classification, cancer subtype prediction	Enables WSI analysis without patch-level supervision

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

In the domain of radiology, LFMs have been particularly effective in bridging the gap between imaging data and associated reports [36]. Models such as CLIP and GLoRIA use contrastive objectives to align visual and textual modalities, allowing for semantic understanding of image features without the need for dense annotations [37]. In practice, this enables radiologists to perform zero-shot labeling of findings such as pneumothorax or cardiomegaly from chest X-rays, a task previously reliant on supervised CNNs trained on labeled corpora like CheXpert or MIMIC-CXR. Furthermore, vision-language models can generate natural language explanations that mirror the style and structure of radiology reports, enhancing interpretability and clinician trust. Clinical natural language processing (NLP) has also seen transformative progress due to foundation models trained on biomedical text corpora [38]. Models such as BioBERT and ClinicalBERT, pretrained on PubMed abstracts and EHR notes, outperform previous baselines on tasks like named entity recognition (NER), relation extraction, and coreference resolution [39]. With the emergence of GPT-based models, zero-shot and few-shot learning have become practical in clinical settings, enabling question answering over unstructured clinical narratives or summarization of discharge notes with minimal task-specific data [40]. These capabilities are vital in environments with high documentation burdens and limited annotation resources. Electronic Health Record (EHR) modeling presents unique challenges due to its heterogeneous, temporal, and sparse nature. LFMs tailored for EHRs, such as BEHRT and TransformerEHR, are capable of modeling longitudinal sequences of coded medical events. These models treat visits as tokenized events and leverage self-attention mechanisms to capture temporal relationships and interdependencies between diagnoses, procedures, and medications [41]. They are used for risk prediction (e.g., predicting sepsis onset), disease progression modeling, and personalized treatment recommendation, often yielding improved metrics such as AUROC and precision-recall compared to traditional models like RNNs or logistic regression [42]. In pathology, the processing of gigapixel whole slide images (WSIs) has traditionally required tile-based training pipelines. LFMs based on vision transformers (ViTs), pretrained on general or medical image datasets, can now process these slides in a more holistic manner [43]. Moreover, multimodal models that integrate image and textual report data have been developed for cancer subtype classification, enabling explainable predictions and report generation. These models reduce the annotation burden and provide better alignment with clinical narratives [44]. Genomics is another frontier where LFMs show immense promise [45]. Models like DNABERT and Enformer adapt transformer architectures to sequence-based inputs, treating DNA as a structured language [46]. These models can predict gene regulatory effects, model enhancer-promoter interactions, and perform variant effect prediction by learning patterns in raw nucleotide sequences [47]. The ability to model long-range dependencies (e.g., >10kb) using self-attention mechanisms has led to breakthroughs in understanding gene expression regulation and the effects of non-coding variants [48]. The domain of drug discovery has been enhanced through molecular LFMs that embed chemical representations such as SMILES strings or molecular graphs [49]. Using transformer-based encoders and pretrained on millions of molecules, these models predict physical and biochemical properties, facilitate drug-target interaction modeling, and even generate novel drug-like compounds in silico. The incorporation of graph-based encoders such as GraphGPT allows models to better capture molecular topology, leading to higher accuracy in virtual screening tasks [50]. Lastly, LFMs are being deployed as generalist models for clinical decision support by integrating multiple modalities. Med-PaLM 2 and Flamingo are examples of such systems that process medical questions alongside image inputs and structured records [51]. These models achieve performance close to or surpassing that of domain experts in benchmark medical exams such as the USMLE [52]. By reasoning over images, labs, and text jointly, they provide holistic assessments that reflect real-world diagnostic processes. However, their deployment still requires careful validation due to the high stakes of clinical decision-making. These diverse applications collectively highlight the versatility and strength of LFMs in tackling some of the most pressing challenges in medical AI. In the following sections, we examine in greater detail the current limitations and risks associated with their use, particularly in relation to fairness, interpretability, and trustworthiness [53].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

4. Architectural Paradigms of Large Foundation Models in Medical Analysis

The architecture of Large Foundation Models (LFMs) applied to medical analysis is both modular and hierarchical, designed to process, align, and reason across various data modalities [54]. These include clinical free-text, radiological images, structured tabular data from electronic health records (EHRs), and genomic sequences [55]. At a high level, a typical LFM pipeline consists of modality-specific encoders, a multimodal fusion module, a shared representation space, and task-specific heads [56]. This structure enables cross-modal information exchange while preserving domain-specific priors essential for accurate clinical reasoning [57]. At the input stage, raw data from diverse sources is first processed through domain-specific tokenization or embedding layers [58]. For instance, radiology images are divided into non-overlapping patches, each of which is flattened and projected into a latent space via a linear embedding layer, similar to the ViT paradigm [59]. Clinical text, on the other hand, is tokenized using subword or byte-pair encoding (BPE) schemes, which allow the model to generalize across varying terminologies and abbreviations. Structured EHRs may be converted into event sequences, where each token represents a timestamped medical code, lab test, or procedure [60]. The modality-specific encoders (e.g., CNNs for images, transformers for text and EHRs) transform these inputs into intermediate embeddings in a latent space $\mathbb{R}^d$ [61]. These embeddings are then passed into a fusion module—often a cross-attention transformer or a contrastive learning block—which aligns features from different modalities and learns joint representations. The resulting unified embedding captures semantic relationships across modalities, enabling the model to perform complex inference tasks such as diagnosis, prognosis, or triage with a holistic understanding of patient data.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

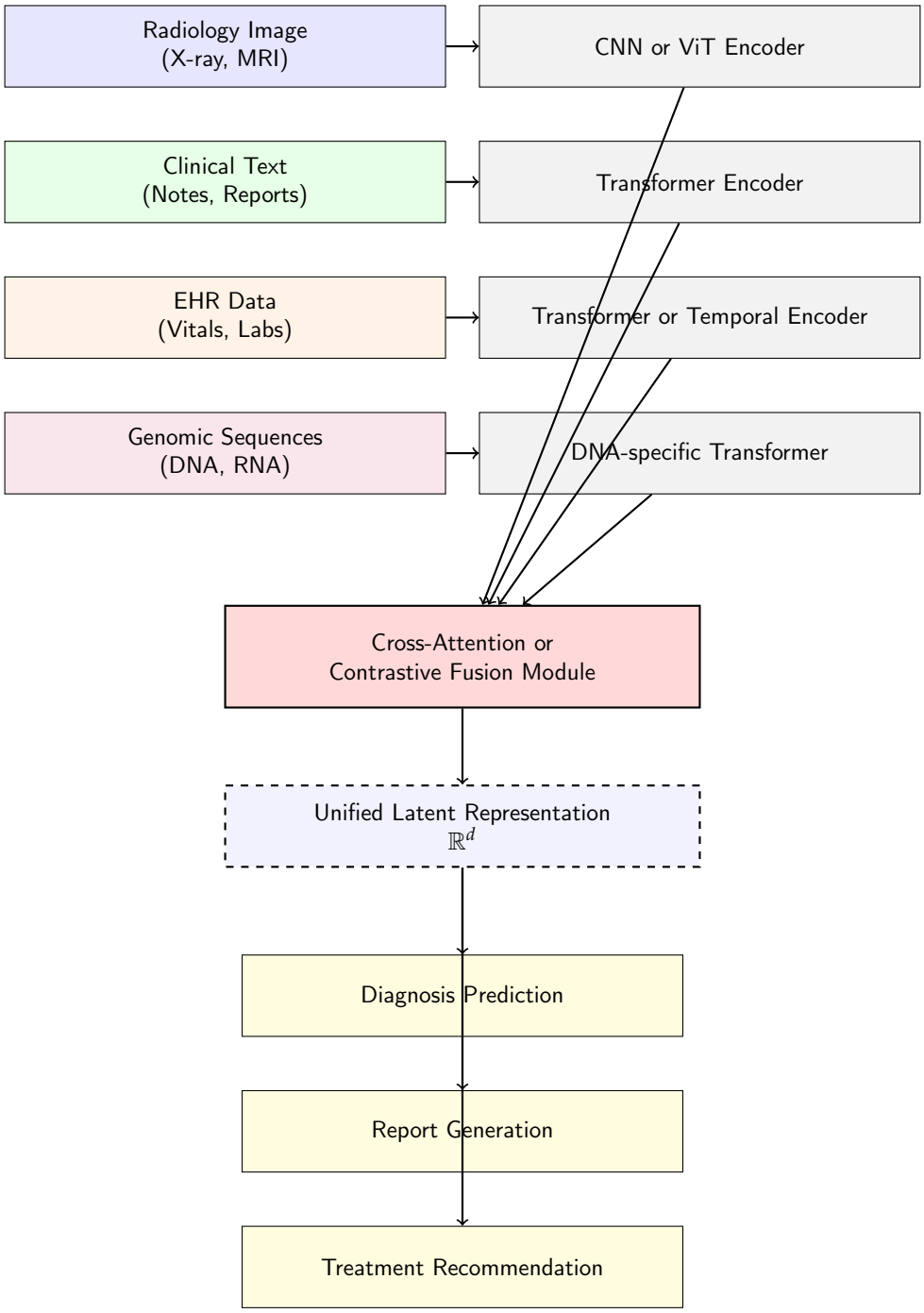

Figure 1. Schematic architecture of a multimodal Large Foundation Model in medical analysis. Inputs from distinct medical data sources are encoded by modality-specific encoders, fused via cross-modal attention mechanisms, and projected into a unified latent representation.

The design of this architecture accommodates the inherent variability and granularity of healthcare data. Radiology images vary in resolution and modality (e.g., DICOM-based CT scans vs [62]. low-resolution X-rays), clinical notes are often unstructured and include domain-specific abbreviations, and EHRs contain sparse and asynchronous measurements [63]. LFMs address these issues by aligning modalities in a common embedding space, where semantic consistency is preserved across vastly different input types [64]. Another crucial feature of these architectures is their ability to support both pretraining and fine-tuning in modular ways [65]. Modality-specific encoders can be pretrained independently on domain-relevant tasks (e.g., image captioning or masked token prediction) and then jointly fine-tuned with fusion and output layers [66]. This decoupled training process not only facilitates scalable development but also promotes model reuse across institutions and tasks [67]. In

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

clinical deployment scenarios, frozen encoders can be used in conjunction with lightweight adapters, enabling rapid domain adaptation without full retraining—a vital requirement in privacy-sensitive healthcare environments. Furthermore, these architectures increasingly incorporate interpretable attention mechanisms and saliency-based explainability tools. By visualizing attention weights across modalities, clinicians can inspect which parts of a radiology scan or which terms in a clinical note contributed most to a model's prediction. This is essential for building trust and supporting human–AI collaboration in decision-making [68]. Taken together, this architectural paradigm provides the necessary flexibility, scalability, and transparency for applying LFMs to real-world medical tasks [69]. It enables both horizontal scaling (across institutions and populations) and vertical specialization (for rare diseases or personalized care), positioning LFMs as a central pillar of the next generation of medical AI systems.

5. Challenges and Ethical Considerations in Deploying Large Foundation Models in Medical Analysis

Despite the transformative potential of Large Foundation Models (LFMs) in advancing medical analysis, their deployment in clinical settings is fraught with multifaceted challenges and ethical considerations [70]. These range from technical obstacles inherent to the models themselves, to broader societal implications such as fairness, privacy, and trustworthiness. A thorough understanding of these issues is essential to responsibly harness LFMs in healthcare, ensuring that their benefits are equitably distributed and aligned with medical ethics [71]. One of the most pressing technical challenges stems from the complexity and opacity of LFMs [72]. Their sheer scale—often encompassing billions of parameters—and their training on heterogeneous, noisy data complicate interpretability [73]. While attention mechanisms provide some insight into the model's decision-making process, these are often insufficient to fully explain predictions in a clinical context where stakes are high. The black-box nature of these models raises concerns about reliability, especially when deployed for diagnoses or treatment recommendations [74]. Moreover, LFMs can inadvertently learn and amplify biases present in training data. For instance, underrepresentation of minority populations or specific disease phenotypes in large biomedical corpora can lead to systematic disparities in model performance, thereby exacerbating healthcare inequities [75]. Addressing these issues requires rigorous evaluation protocols, development of robust uncertainty quantification methods, and integration of fairness constraints during training. Another significant challenge lies in the availability, quality, and privacy of medical data used to train LFMs. Medical data is inherently sensitive, regulated by frameworks such as HIPAA and GDPR, which restrict data sharing and complicate large-scale data aggregation. This scarcity and fragmentation of data can limit the diversity and representativeness of training sets, impacting generalizability [76]. Federated learning and differential privacy techniques offer promising avenues to mitigate these issues, allowing models to learn from distributed datasets without compromising patient confidentiality. However, these approaches introduce additional computational overhead and complexity, and their efficacy in large-scale clinical deployments remains an active area of research [77]. From an ethical standpoint, transparency and accountability are paramount when integrating LFMs into clinical workflows [78]. The delegation of decision-making to AI systems must not undermine the clinician's authority or responsibility [79]. Instead, LFMs should function as assistive tools that augment human judgment, providing interpretable outputs and clear indications of confidence or uncertainty. Furthermore, there is an ethical imperative to ensure equitable access to these advanced technologies across healthcare settings, including under-resourced and rural areas. This involves not only technological deployment but also education and training for medical professionals to effectively utilize and critically evaluate AI-generated insights. Regulatory frameworks must evolve in tandem with the rapid development of LFMs to safeguard patient safety and promote ethical standards [80]. Current guidelines for medical devices and software are being adapted to accommodate AI's continuous learning and data dependency. Transparency in model provenance, validation on diverse populations, and post-deployment monitoring are critical components of regulatory compliance [81].

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

Additionally, mechanisms for reporting and addressing adverse outcomes related to AI use need to be established, fostering an environment of continual improvement and accountability. Finally, the societal implications of deploying LFMs in medicine extend beyond clinical efficacy [82]. Issues such as data ownership, informed consent for AI-assisted care, and the potential for automation to disrupt clinical jobs warrant careful consideration [83]. Engaging diverse stakeholders—including patients, clinicians, ethicists, and policymakers—in the development and governance of medical LFMs is essential to align technological advances with societal values and patient-centered care principles [84]. In summary, while LFMs hold immense promise for revolutionizing medical analysis, realizing this potential requires addressing substantial technical, ethical, and regulatory challenges. Responsible development, transparent deployment, and inclusive governance are crucial to ensure that these powerful models enhance healthcare delivery without compromising fairness, privacy, or human dignity.

6. Future Directions and Research Opportunities in Large Foundation Models for Medical Analysis

The field of Large Foundation Models (LFMs) in medical analysis is rapidly evolving, with numerous promising avenues for future research that can further unlock their transformative potential while addressing existing limitations. As LFMs mature, interdisciplinary collaboration between machine learning researchers, clinicians, and ethicists will be critical to steer innovation towards impactful and responsible applications in healthcare [85]. This section outlines key future directions and research opportunities that will likely shape the next generation of medical LFMs. One major direction is the development of more efficient and scalable training methods tailored to the unique challenges of medical data [86]. Current LFMs require enormous computational resources and vast amounts of labeled data, which are often scarce in the medical domain [87]. Research into data-efficient learning paradigms such as self-supervised learning, continual learning, and few-shot adaptation holds promise to reduce these demands [88]. For instance, leveraging large volumes of unlabeled medical imaging and textual data through masked token prediction or contrastive learning can provide rich pretraining signals that improve downstream task performance with limited annotation. Additionally, exploring modular architectures that allow selective updating of model components can enable lifelong learning from incremental clinical data without catastrophic forgetting, an important capability for adapting to evolving medical knowledge and practices [89]. Another fruitful area of investigation is enhancing the interpretability and explainability of LFMs in clinical settings [90]. Bridging the gap between high model complexity and human-understandable insights remains a fundamental challenge [91]. Future research could focus on developing novel attention visualization tools, causal inference frameworks, and counterfactual explanations that provide clinicians with actionable rationale behind model predictions [92]. Embedding interpretability objectives directly into the training process may also yield models that are inherently more transparent. Furthermore, integrating domain knowledge, such as clinical guidelines or biological pathways, into model architectures could produce explanations grounded in medical reasoning rather than purely data-driven correlations. Improving fairness and mitigating bias in medical LFMs represent another critical frontier [93]. Future work should focus on developing robust methods for detecting and correcting disparities across demographic groups, disease subtypes, and healthcare systems [94]. Techniques such as adversarial debiasing, fairness-aware loss functions, and synthetic data augmentation can be adapted and extended for the complexities of medical data. Moreover, establishing standardized benchmarks and datasets that reflect diverse patient populations is essential to fairly evaluate and compare model performance. Ethical frameworks and community-driven governance structures will be needed to ensure equitable model development and deployment, particularly for marginalized and underserved groups [95]. Multimodal learning will continue to be a pivotal area of growth, as integrating heterogeneous data sources unlocks richer clinical insights [96]. Future research may explore advanced fusion strategies that dynamically weight modalities based on data quality, clinical context, and uncertainty [97]. Incorporating real-time

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

streaming data such as continuous vital sign monitoring alongside static modalities like imaging and genomics can enable more precise and timely clinical interventions. Additionally, the expansion of LFMs to incorporate emerging data types—such as wearable sensor data, digital pathology, and patient-reported outcomes—will broaden their applicability and relevance in personalized medicine [98]. The intersection of LFMs with privacy-preserving technologies also offers fertile ground for innovation [99]. Techniques like federated learning, homomorphic encryption, and secure multiparty computation could be refined to accommodate the scale and complexity of foundation models, allowing collaborative model training across institutions without compromising patient confidentiality [100]. Developing practical frameworks and tools to balance privacy, model utility, and regulatory compliance will be essential for widespread adoption [101]. Finally, integrating LFMs into clinical workflows demands robust evaluation protocols and human–AI interaction research. Future studies should rigorously assess how clinicians interpret and act on model outputs, the impact on diagnostic accuracy and patient outcomes, and potential unintended consequences [102]. Designing user interfaces that seamlessly incorporate AI assistance and provide meaningful feedback loops can foster trust and effective collaboration [103]. Moreover, developing adaptive models that learn from clinician corrections and feedback in real time could continuously improve system performance and user satisfaction [104]. In conclusion, the future of LFMs in medical analysis is rich with opportunities for innovation that can enhance healthcare delivery, diagnostics, and patient outcomes [105]. Addressing current challenges through targeted research and multidisciplinary cooperation will be pivotal in realizing the full potential of these powerful models while upholding ethical standards and clinical relevance [106].

7. Conclusion

Large Foundation Models (LFMs) have emerged as a groundbreaking paradigm in medical analysis, offering unprecedented capabilities in processing and interpreting complex, multimodal healthcare data. Their ability to learn rich, generalizable representations from vast amounts of unstructured and structured data has unlocked new possibilities across a diverse array of clinical tasks, ranging from imaging interpretation and natural language processing to genomics and drug discovery. This survey has explored the foundational architectures, applications, challenges, and future directions associated with LFMs, highlighting their potential to revolutionize modern medicine by augmenting clinical decision-making and improving patient outcomes.

Despite their impressive performance and versatility, LFMs also bring significant challenges that cannot be overlooked. The technical hurdles of model interpretability, data scarcity, bias, and computational expense must be addressed to ensure safe and effective deployment in clinical environments. Moreover, the ethical and regulatory dimensions surrounding privacy, fairness, transparency, and accountability are critical to gaining trust from both clinicians and patients. Careful attention to these considerations will be essential in guiding the responsible integration of LFMs into healthcare systems worldwide.

Looking forward, the continued evolution of LFMs will depend on collaborative efforts that bridge the fields of artificial intelligence, medicine, ethics, and policy. Advances in scalable training techniques, interpretable model design, and privacy-preserving learning will drive the development of more robust and equitable medical AI systems. Additionally, embedding human-centered design principles and fostering effective human–AI collaboration will ensure that these technologies serve as reliable partners to healthcare professionals rather than opaque replacements. The dynamic nature of medical knowledge and patient populations calls for adaptive LFMs that can learn continuously and respond to emerging health challenges in real time.

In essence, LFMs represent both a technological leap and a profound opportunity to transform healthcare delivery. Their ability to synthesize heterogeneous data and provide nuanced clinical insights promises to enhance diagnostic accuracy, streamline workflows, and enable personalized treatment strategies. By confronting the attendant challenges with rigorous research, ethical foresight,

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

and stakeholder engagement, the medical community can harness the full power of LFMs to improve health outcomes and equity on a global scale. This ongoing journey underscores a new era where artificial intelligence and human expertise converge to redefine the future of medicine.

References

1. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. *CoRR* **2023**, *abs/2302.13971*, [2302.13971].
2. Li, M.D.; Arun, N.T.; Gidwani, M.; Chang, K.; Deng, F.; Little, B.P.; Mendoza, D.P.; Lang, M.; Lee, S.I.; O'Shea, A.; et al. Automated assessment of COVID-19 pulmonary disease severity on chest radiographs using convolutional Siamese neural networks. *medRxiv* **2020**. https://doi.org/10.1101/2020.05.20.20108159.
3. Rajaraman, S.; Antani, S.K. Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs. *IEEE Access* **2020**, *8*, 27318–27326. https://doi.org/10.1109/ACCESS.2020.2971257.
4. Faghri, F.; Fleet, D.J.; Kiros, J.R.; Fidler, S. VSE++: Improving Visual-Semantic Embeddings with Hard Negatives. In Proceedings of the BMVC 2018, 2018, p. 12.
5. RSNA. RSNA Pneumonia Detection Challenge, 2018. Library Catalog: www.kaggle.com.
6. Arbabshirani, M.R.; Dallal, A.H.; Agarwal, C.; Patel, A.; Moore, G. Accurate segmentation of lung fields on chest radiographs using deep convolutional networks. In Proceedings of the Medical Imaging 2017: Image Processing. SPIE, 2017, p. 1013305. https://doi.org/10.1117/12.2254526.
7. Omkar Thawkar, Abdelrahman Shaker, S.S.M.H.C.R.M.A.S.K.J.L.; Khan, F.S. XrayGPT: Chest Radiographs Summarization using Large Medical Vision-language Models. *arXiv: 2306.07971* **2023**.
8. Varela-Santos, S.; Melin, P. A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks. *Information Sciences* **2021**, *545*, 403–414. https://doi.org/10.1016/j.ins.2020.09.041.
9. Dalla Serra, F.; Clackett, W.; MacKinnon, H.; Wang, C.; Deligianni, F.; Dalton, J.; O'Neil, A.Q. Multimodal Generation of Radiology Reports using Knowledge-Grounded Extraction of Entities and Relations. In Proceedings of the AACL 2022, Online only, 2022; pp. 615–624.
10. Xue, Z.; Antani, S.; Long, R.; Thoma, G.R. Using deep learning for detecting gender in adult chest radiographs. In Proceedings of the Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications. SPIE, 2018, p. 10. https://doi.org/10.1117/12.2293027.
11. Gilanie, G.; Bajwa, U.I.; Waraich, M.M.; Asghar, M.; Kousar, R.; Kashif, A.; Aslam, R.S.; Qasim, M.M.; Rafique, H. Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks. *Biomed. Signal Process. Control* **2021**, *66*, 102490. https://doi.org/10.1016/j.bspc.2021.102490.
12. Zhang, Y.; Miao, S.; Mansi, T.; Liao, R. Task Driven Generative Modeling for Unsupervised Domain Adaptation: Application to X-ray Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*; Springer, 2018; Vol. 11071, pp. 599–607. https://doi.org/10.1007/978-3-030-00934-2_67.
13. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-resolution Image Synthesis with Latent Diffusion Models. In Proceedings of the CVPR, 2022, pp. 10684–10695.
14. Ausawalaithong, W.; Marukatat, S.; Thirach, A.; Wilaiprasitporn, T. Automatic Lung Cancer Prediction from Chest X-ray Images Using Deep Learning Approach **2018**. [arxiv:cs,eess/1808.10858].
15. Crosby, J.; Chen, S.; Li, F.; MacMahon, H.; Giger, M. Network output visualization to uncover limitations of deep learning detection of pneumothorax. In Proceedings of the Medical Imaging 2020: Image Perception, Observer Performance, and Technology Assessment. SPIE, 2020, p. 22. https://doi.org/10.1117/12.2550066.
16. Zhang, T.; Fu, H.; Zhao, Y.; Cheng, J.; Guo, M.; Gu, Z.; Yang, B.; Xiao, Y.; Gao, S.; Liu, J. SkrGAN: Sketching-Rendering Unconditional Generative Adversarial Networks for Medical Image Synthesis. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11767, pp. 777–785. https://doi.org/10.1007/978-3-030-32251-9_85.
17. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **2021**, *18*, 203–211. https://doi.org/10.1038/s41592-020-01008-z.
18. Zhang, B.; Jia, C.; Wu, R.; Lv, B.; Li, B.; Li, F.; Du, G.; Sun, Z.; Li, X. Improving rib fracture detection accuracy and reading efficiency with deep learning-based detection software: a clinical evaluation. *The British Journal of Radiology* **2021**, *94*, 20200870. https://doi.org/10.1259/bjr.20200870.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

19. Oliveira, H.N.; Ferreira, E.; Santos, J.A.D. Truly Generalizable Radiograph Segmentation With Conditional Domain Adaptation. *IEEE Access* **2020**, *8*, 84037–84062. https://doi.org/10.1109/access.2020.2991688.
20. Zhang, L.; Rong, R.; Li, Q.; Yang, D.M.; Yao, B.; Luo, D.; Zhang, X.; Zhu, X.; Luo, J.; Liu, Y.; et al. A deep learning-based model for screening and staging pneumoconiosis. *Sci. Rep.* **2021**, *11*, 2201. https://doi.org/10.1038/s41598-020-77924-z.
21. Wu, J.T.; Agu, N.; Lourentzou, I.; Sharma, A.; Paguio, J.A.; Yao, J.S.; Dee, E.C.; Mitchell, W.; Kashyap, S.; Giovannini, A.; et al. Chest ImaGenome Dataset for Clinical Reasoning. In Proceedings of the NeurIPS Datasets and Benchmarks 2021, 2021, pp. 1–14.
22. Oh, Y.; Park, S.; Ye, J.C. Deep Learning COVID-19 Features on CXR Using Limited Training Data Sets. *IEEE Transactions on Medical Imaging* **2020**, *39*, 2688–2700. https://doi.org/10.1109/TMI.2020.2993291.
23. Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical Text-conditional Image Generation with CLIP Latents. *arXiv preprint arXiv:2204.06125* **2022**.
24. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017, NIPS'17, p. 6629–6640.
25. Nicolson, A.; Dowling, J.; Koopman, B. Improving Chest X-ray Report Generation by Leveraging Warm Starting. *Artificial Intelligence in Medicine* **2023**, *144*, 102633.
26. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.
27. Ouyang, X.; Karanam, S.; Wu, Z.; Chen, T.; Huo, J.; Zhou, X.S.; Wang, Q.; Cheng, J.Z. Learning Hierarchical Attention for Weakly-supervised Chest X-Ray Abnormality Localization and Diagnosis. *IEEE Transactions on Medical Imaging* **2020**, pp. 1–1. https://doi.org/10.1109/TMI.2020.3042773.
28. Ezzat, D.; Hassanien, A.E.; Ella, H.A. An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. *Appl. Soft Comput.* **2021**, *98*, 106742. https://doi.org/10.1016/j.asoc.2020.106742.
29. Chen, Z.; Shen, Y.; Song, Y.; Wan, X. Cross-modal Memory Networks for Radiology Report Generation. In Proceedings of the ACL-IJCNLP 2021, Online, 2021; pp. 5904–5914.
30. Philipsen, R.H.H.M.; Sánchez, C.I.; Melendez, J.; Lew, W.J.; van Ginneken, B. Automated chest X-ray reading for tuberculosis in the Philippines to improve case detection: a cohort study. *The International Journal of Tuberculosis and Lung Disease: The Official Journal of the International Union Against Tuberculosis and Lung Disease* **2019**, *23*, 805–810. https://doi.org/10.5588/ijtld.18.0004.
31. Anis, S.; Lai, K.W.; Chuah, J.H.; Shoaib, M.A.; Mohafez, H.; Hadizadeh, M.; Ding, Y.; Ong, Z.C. An Overview of Deep Learning Approaches in Chest Radiograph. *IEEE Access* **2020**. https://doi.org/10.1109/access.2020.3028390.
32. Xing, Y.; Ge, Z.; Zeng, R.; Mahapatra, D.; Seah, J.; Law, M.; Drummond, T. Adversarial Pulmonary Pathology Translation for Pairwise Chest X-Ray Data Augmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 757–765. https://doi.org/10.1007/978-3-030-32226-7_84.
33. El Asnaoui, K. Design ensemble deep learning model for pneumonia disease classification. *International Journal of Multimedia Information Retrieval* **2021**. https://doi.org/10.1007/s13735-021-00204-7.
34. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. AAAI Press, 2017, AAAI'17, p. 4278–4284.
35. Pham, V.T.; Tran, C.M.; Zheng, S.; Vu, T.M.; Nath, S. Chest x-ray abnormalities localization via ensemble of deep convolutional neural networks. In Proceedings of the 2021 International Conference on Advanced Technologies for Communications (ATC). IEEE, 2021, pp. 125–130.
36. Matsubara, N.; Teramoto, A.; Saito, K.; Fujita, H. Bone suppression for chest X-ray image using a convolutional neural filter. *Physical and Engineering Sciences in Medicine* **2020**, *43*, 97–108. https://doi.org/10.1007/s13246-019-00822-w.
37. Zhang, L.; Rao, A.; Agrawala, M. Adding Conditional Control to Text-to-Image Diffusion Models. In Proceedings of the ICCV, October 2023, pp. 3836–3847.
38. Wang, L.; Lin, Z.Q.; Wong, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci. Rep.* **2020**, *10*, 19549. https://doi.org/10.1038/s41598-020-76550-z.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

39. Bustos, A.; Pertusa, A.; Salinas, J.M.; de la Iglesia-Vayá, M. PadChest: A Large Chest X-Ray Image Dataset with Multi-Label Annotated Reports. *Medical Image Analysis* **2020**, *66*, 101797.
40. Jang, S.B.; Lee, S.H.; Lee, D.E.; Park, S.Y.; Kim, J.K.; Cho, J.W.; Cho, J.; Kim, K.B.; Park, B.; Park, J.; et al. Deep-learning algorithms for the interpretation of chest radiographs to aid in the triage of COVID-19 patients: A multicenter retrospective study. *PLoS One* **2020**, *15*, e0242759. https://doi.org/10.1371/journal.pone.0242759.
41. Carlile, M.; Hurt, B.; Hsiao, A.; Hogarth, M.; Longhurst, C.A.; Dameff, C. Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department. *Journal of the American College of Emergency Physicians Open* **2020**, *1*, 1459–1464. https://doi.org/10.1002/emp2.12297.
42. You, D.; Liu, F.; Ge, S.; Xie, X.; Zhang, J.; Wu, X. AlignTransformer: Hierarchical Alignment of Visual Regions and Disease Tags for Medical Report Generation. In Proceedings of the MICCAI 2021, 2021, Vol. 12903, pp. 72–82.
43. Haghighi, F.; Hosseinzadeh Taher, M.R.; Zhou, Z.; Gotway, M.B.; Liang, J. Learning Semantics-Enriched Representation via Self-discovery, Self-classification, and Self-restoration. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*; Springer, 2020; Vol. 12261, pp. 137–147. https://doi.org/10.1007/978-3-030-59710-8_14.
44. Li, Z.; Hou, Z.; Chen, C.; Hao, Z.; An, Y.; Liang, S.; Lu, B. Automatic Cardiothoracic Ratio Calculation With Deep Learning. *IEEE Access* **2019**, *7*, 37749–37756. https://doi.org/10.1109/ACCESS.2019.2900053.
45. Kikkisetti, S.; Zhu, J.; Shen, B.; Li, H.; Duong, T.Q. Deep-learning convolutional neural networks with transfer learning accurately classify COVID-19 lung infection on portable chest radiographs. *PeerJ* **2020**, *8*, e10309. https://doi.org/10.7717/peerj.10309.
46. An, J.Y.; Seo, H.; Kim, Y.G.; Lee, K.E.; Kim, S.; Kong, H.J. Codeless Deep Learning of COVID-19 Chest X-Ray Image Dataset with KNIME Analytics Platform. *Healthcare Informatics Research* **2021**, *27*, 82–91. https://doi.org/10.4258/hir.2021.27.1.82.
47. Conjeti, S.; Roy, A.G.; Katouzian, A.; Navab, N. Hashing with Residual Networks for Image Retrieval. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2017*; Springer, 2017; Vol. 10435, pp. 541–549. https://doi.org/10.1007/978-3-319-66179-7_62.
48. Thammarach, P.; Khaengthanyakan, S.; Vongsurakrai, S.; Phienphanich, P.; Pooprasert, P.; Yaemsuk, A.; Vanichvarodom, P.; Munpolsri, N.; Khwayotha, S.; Lertkowit, M.; et al. AI Chest 4 All. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC), 2020, pp. 1229–1233. ISSN: 2694-0604, https://doi.org/10.1109/EMBC44109.2020.9175862.
49. Gooßen, A.; Deshpande, H.; Harder, T.; Schwab, E.; Baltruschat, I.; Mabotuwana, T.; Cross, N.; Saalbach, A. Deep Learning for Pneumothorax Detection and Localization in Chest Radiographs, 2019, [arXiv:eess.IV/1907.07324].
50. Clevert, D.; Unterthiner, T.; Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) **2015**. [1511.07289].
51. Cornia, M.; Stefanini, M.; Baraldi, L.; Cucchiara, R. Meshed-memory Transformer for Image Captioning. In Proceedings of the CVPR, 2020, pp. 10578–10587.
52. Rueckel, J.; Kunz, W.G.; Hoppe, B.F.; Patzig, M.; Notohamiprodjo, M.; Meinel, F.G.; Cyran, C.C.; Ingrisch, M.; Ricke, J.; Sabel, B.O. Artificial Intelligence Algorithm Detecting Lung Infection in Supine Chest Radiographs of Critically Ill Patients With a Diagnostic Accuracy Similar to Board-Certified Radiologists. *Critical Care Medicine* **2020**, *Publish Ahead of Print*. https://doi.org/10.1097/CCM.0000000000004397.
53. Bougias, H.; Georgiadou, E.; Malamateniou, C.; Stogiannos, N. Identifying cardiomegaly in chest X-rays: a cross-sectional study of evaluation and comparison between different transfer learning methods. *Acta Radiol.* **2020**, p. 028418512097363. https://doi.org/10.1177/0284185120973630.
54. Li, C.; Yang, Y.; Liang, H.; Wu, B. Transfer learning for establishment of recognition of COVID-19 on CT imaging using small-sized training datasets. *Knowledge-Based Systems* **2021**, *218*, 106849. https://doi.org/10.1016/j.knosys.2021.106849.
55. Walsh, S.L.F.; Humphries, S.M.; Wells, A.U.; Brown, K.K. Imaging research in fibrotic lung disease; applying deep learning to unsolved problems. *The Lancet Respiratory Medicine* **2020**, *8*, 1144–1153. https://doi.org/10.1016/S2213-2600(20)30003-5.
56. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, jul 2017, pp. 1125–1134. https://doi.org/10.1109/cvpr.2017.632.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

57. Cai, Q.; Du, S.Y.; Gao, S.; Huang, G.L.; Zhang, Z.; Li, S.; Wang, X.; Li, P.L.; Lv, P.; Hou, G.; et al. A model based on CT radiomic features for predicting RT-PCR becoming negative in coronavirus disease 2019 (COVID-19) patients. *BMC Med. Imaging* **2020**, *20*, 118. https://doi.org/10.1186/s12880-020-00521-z.
58. Shiri, I.; Akhavanallaf, A.; Sanaat, A.; Salimi, Y.; Askari, D.; Mansouri, Z.; Shayesteh, S.P.; Hasanian, M.; Rezaei-Kalantari, K.; Salahshour, A.; et al. Ultra-low-dose chest CT imaging of COVID-19 patients using a deep residual neural network. *Eur. Radiol.* **2021**, *31*, 1420–1431. https://doi.org/10.1007/s00330-020-07225-6.
59. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, 2017, pp. 1–14.
60. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE, dec 2015, pp. 1440–1448. https://doi.org/10.1109/iccv.2015.169.
61. Ranem, A.; Babendererde, N.; Fuchs, M.; Mukhopadhyay, A. Exploring SAM Ablations for Enhancing Medical Segmentation in Radiology and Pathology, 2023, [arXiv:cs.CV/2310.00504].
62. Huang, Z.; Zhou, Q.; Zhu, X.; Zhang, X. Batch Similarity Based Triplet Loss Assembled into Light-Weighted Convolutional Neural Networks for Medical Image Classification. *Sensors* **2021**, *21*, 764. https://doi.org/10.3390/s21030764.
63. Wang, X.; Yu, J.; Zhu, Q.; Li, S.; Zhao, Z.; Yang, B.; Pu, J. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography. *Occupational and Environmental Medicine* **2020**, *77*, 597–602. https://doi.org/10.1136/oemed-2019-106386.
64. Klambauer, G.; Unterthiner, T.; Mayr, A.; Hochreiter, S. Self-Normalizing Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, 2017, pp. 972–981.
65. Patel, B.N.; Langlotz, C.P. Beyond the *AJR* : "Deep Learning Using Chest Radiographs to Identify High-Risk Smokers for Lung Cancer Screening Computed Tomography: Development and Validation of a Prediction Model". *Am. J. Roentgenol.* **2020**, p. AJR.20.25334. https://doi.org/10.2214/AJR.20.25334.
66. Kim, M.; Lee, B.D. Automatic Lung Segmentation on Chest X-rays Using Self-Attention Deep Neural Network. *Sensors* **2021**, *21*, 369. https://doi.org/10.3390/s21020369.
67. Lodwick, G.S.; Keats, T.E.; Dorst, J.P. The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. *Radiology* **1963**, *81*, 185–200. https://doi.org/10.1148/81.2.185.
68. Yi, X.; Adams, S.; Babyn, P.; Elnajmi, A. Automatic Catheter and Tube Detection in Pediatric X-ray Images Using a Scale-Recurrent Network and Synthetic Data. *Journal of Digital Imaging* **2019**, *33*, 181–190. https://doi.org/10.1007/s10278-019-00201-7.
69. object-CXR - Automatic detection of foreign objects on chest X-rays.
70. Oliveira, H.; Mota, V.; Machado, A.M.; dos Santos, J.A. From 3D to 2D: Transferring knowledge for rib segmentation in chest X-rays. *Pattern Recognition Letters* **2020**, *140*, 10–17. https://doi.org/10.1016/j.patrec.2020.09.021.
71. Li, M.; Liu, R.; Wang, F.; Chang, X.; Liang, X. Auxiliary Signal-guided Knowledge Encoder-decoder for Medical Report Generation. *World Wide Web* **2022**, pp. 1–18.
72. Woźniak, M.; Połap, D.; Capizzi, G.; Sciuto, G.L.; Kośmider, L.; Frankiewicz, K. Small lung nodules detection based on local variance analysis and probabilistic neural network. *Computer Methods and Programs in Biomedicine* **2018**, *161*, 173–180. https://doi.org/10.1016/j.cmpb.2018.04.025.
73. Fricks, R.B.; Abadi, E.; Ria, F.; Samei, E. Classification of COVID-19 in chest radiographs: assessing the impact of imaging parameters using clinical and simulated images. In Proceedings of the Medical Imaging 2021: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2021, Vol. 11597, p. 115970A. https://doi.org/10.1117/12.2582223.
74. Wang, Y.; Sun, L.L.; Jin, Q. Enhanced Diagnosis of Pneumothorax with an Improved Real-time Augmentation for Imbalanced Chest X-rays Data Based on DCNN. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2019**, pp. 1–1. https://doi.org/10.1109/TCBB.2019.2911947.
75. von Berg, J.; Young, S.; Carolus, H.; Wolz, R.; Saalbach, A.; Hidalgo, A.; Giménez, A.; Franquet, T. A novel bone suppression method that improves lung nodule detection. *International Journal of Computer Assisted Radiology and Surgery* **2015**, *11*, 641–655. https://doi.org/10.1007/s11548-015-1278-y.
76. Murphy, K.; Habib, S.S.; Zaidi, S.M.A.; Khowaja, S.; Khan, A.; Melendez, J.; Scholten, E.T.; Amad, F.; Schalekamp, S.; Verhagen, M.; et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. *Scientific Reports* **2020**, *10*, 5492. https://doi.org/10.1038/s41598-020-62148-y.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

77. Blain, M.; T Kassin, M.; Varble, N.; Wang, X.; Xu, Z.; Xu, D.; Carrafiello, G.; Vespro, V.; Stellato, E.; Ierardi, A.M.; et al. Determination of disease severity in COVID-19 patients using deep learning in chest X-ray images. *Diagnostic and Interventional Radiology (Ankara, Turkey)* **2020**. https://doi.org/10.5152/dir.2020.20205.
78. Liu, F.; Wu, X.; Ge, S.; Fan, W.; Zou, Y. Exploring and Distilling Posterior and Prior Knowledge for Radiology Report Generation. In Proceedings of the CVPR 2021, 2021, pp. 13753–13762.
79. Wu, M.; Zhang, X.; Sun, X.; Zhou, Y.; Chen, C.; Gu, J.; Sun, X.; Ji, R. DIFNet: Boosting Visual Information Flow for Image Captioning. In Proceedings of the CVPR 2022, 2022, pp. 17999–18008.
80. Williams, R.J. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.* **1992**, *8*, 229–256.
81. Uzunova, H.; Ehrhardt, J.; Jacob, F.; Frydrychowicz, A.; Handels, H. Multi-scale GANs for Memory-efficient Generation of High Resolution Medical Images. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 112–120. https://doi.org/10.1007/978-3-030-32226-7_13.
82. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the NAACL 2019), Minneapolis, Minnesota, 2019; pp. 4171–4186.
83. Ginneken, B.V.; Romeny, B.T.H.; Viergever, M. Computer-aided diagnosis in chest radiography: a survey. *IEEE Transactions on Medical Imaging* **2001**, *20*, 1228–1241. https://doi.org/10.1109/42.974918.
84. Wang, Z.; Wu, Z.; Agarwal, D.; Sun, J. MedCLIP: Contrastive Learning from Unpaired Medical Images and Text. In Proceedings of the EMNLP 2022, Abu Dhabi, United Arab Emirates, 2022; pp. 3876–3887.
85. Kim, D.W.; Jang, H.Y.; Kim, K.W.; Shin, Y.; Park, S.H. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean Journal of Radiology* **2019**, *20*, 405. https://doi.org/10.3348/kjr.2019.0025.
86. Frid-Adar, M.; Amer, R.; Greenspan, H. Endotracheal Tube Detection and Segmentation in Chest Radiographs Using Synthetic Data. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 784–792. https://doi.org/10.1007/978-3-030-32226-7_87.
87. Shah, U.; Abd-Alrazeq, A.; Alam, T.; Househ, M.; Shah, Z. An Efficient Method to Predict Pneumonia from Chest X-Rays Using Deep Learning Approach. *Studies in Health Technology and Informatics* **2020**, *272*, 457–460. https://doi.org/10.3233/SHTI200594.
88. Zarshenas, A.; Liu, J.; Forti, P.; Suzuki, K. Separation of bones from soft tissue in chest radiographs: Anatomy-specific orientation-frequency-specific deep neural network convolution. *Medical Physics* **2019**, *46*, 2232–2242. https://doi.org/10.1002/mp.13468.
89. Murugan, R.; Goel, T. E-DiCoNet: Extreme learning machine based classifier for diagnosis of COVID-19 using deep convolutional network. *Journal of Ambient Intelligence and Humanized Computing* **2021**. https://doi.org/10.1007/s12652-020-02688-3.
90. Wang, Z.; Han, H.; Wang, L.; Li, X.; Zhou, L. Automated Radiographic Report Generation Purely on Transformer: A Multicriteria Supervised Approach. *IEEE Transactions on Medical Imaging* **2022**, *41*, 2803–2813.
91. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the ICCV, October 2023, pp. 4015–4026.
92. ACR. SIIM-ACR Pneumothorax Segmentation, 2019.
93. Mao, J.; Xu, W.; Yang, Y.; Wang, J.; Yuille, A.L. Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN). In Proceedings of the ICLR 2015, 2015, pp. 1–17.
94. Umer, M.; Ashraf, I.; Ullah, S.; Mehmood, A.; Choi, G.S. COVINet: a convolutional neural network approach for predicting COVID-19 from chest X-ray images. *Journal of Ambient Intelligence and Humanized Computing* **2021**. https://doi.org/10.1007/s12652-021-02917-3.
95. Al-Waisy, A.S.; Al-Fahdawi, S.; Mohammed, M.A.; Abdulkareem, K.H.; Mostafa, S.A.; Maashi, M.S.; Arif, M.; Garcia-Zapirain, B. COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. *Soft Computing* **2020**. https://doi.org/10.1007/s00500-020-05424-3.
96. Lassau, N.; Ammari, S.; Chouzenoux, E.; Gortais, H.; Herent, P.; Devilder, M.; Soliman, S.; Meyrignac, O.; Talabard, M.P.; Lamarque, J.P.; et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. *Nat. Commun.* **2021**, *12*, 634. https://doi.org/10.1038/s41467-020-20657-4.
97. Rezaei, M.; Shahidi, M. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. *Intelligence-Based Medicine* **2020**, *3-4*, 100005. https://doi.org/10.1016/j.ibmed.2020.100005.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 7 August 2025 doi:10.20944/preprints202508.0498.v1

98. Nazarov, O.; Yaqub, M.; Nandakumar, K. On the Importance of Image Encoding in Automated Chest X-Ray Report Generation. In Proceedings of the 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022, 2022, p. 475.
99. Li, M.; Cai, W.; Liu, R.; Weng, Y.; Zhao, X.; Wang, C.; Chen, X.; Liu, Z.; Pan, C.; Li, M.; et al. FFA-IR: Towards an Explainable and Reliable Medical Report Generation Benchmark. In Proceedings of the NeurIPS Datasets and Benchmarks 2021, 2021.
100. Michalopoulos, G.; Williams, K.; Singh, G.; Lin, T. MedicalSum: A Guided Clinical Abstractive Summarization Model for Generating Medical Reports from Patient-Doctor Conversations. In Proceedings of the Findings of EMNLP 2022, Abu Dhabi, United Arab Emirates, 2022; pp. 4741–4749.
101. Pan, Y.; Chen, Q.; Chen, T.; Wang, H.; Zhu, X.; Fang, Z.; Lu, Y. Evaluation of a computer-aided method for measuring the Cobb angle on chest X-rays. *European Spine Journal* **2019**, *28*, 3035–3043. https://doi.org/10.1007/s00586-019-06115-w.
102. Banerjee, S.; Lavie, A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In Proceedings of the Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization, Ann Arbor, Michigan, 2005; pp. 65–72.
103. Sogancioglu, E.; Murphy, K.; Calli, E.; Scholten, E.T.; Schalekamp, S.; Ginneken, B.V. Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification. *IEEE Access* **2020**, *8*, 94631–94642. https://doi.org/10.1109/access.2020.2995567.
104. Schalekamp, S.; van Ginneken, B.; Koedam, E.; Snoeren, M.M.; Tiehuis, A.M.; Wittenberg, R.; Karssemeijer, N.; Schaefer-Prokop, C.M. Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images. *Radiology* **2014**, *272*, 252–261. https://doi.org/10.1148/radiol.14131315.
105. E, L.; Zhao, B.; Guo, Y.; Zheng, C.; Zhang, M.; Lin, J.; Luo, Y.; Cai, Y.; Song, X.; Liang, H. Using deep-learning techniques for pulmonary-thoracic segmentations and improvement of pneumonia diagnosis in pediatric chest radiographs. *Pediatric Pulmonology* **2019**, *54*, 1617–1626. https://doi.org/10.1002/ppul.24431.
106. Browne, R.F.J.; O'Reilly, G.; McInerney, D. Extraction of the Two-Dimensional Cardiothoracic Ratio from Digital PA Chest Radiographs: Correlation with Cardiac Function and the Traditional Cardiothoracic Ratio. *Journal of Digital Imaging* **2004**, *17*, 120–123. https://doi.org/10.1007/s10278-003-1900-3.

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.

© 2025 by the author(s). Distributed under a Creative Commons CC BY license.