Review

Not peer-reviewed version

Automating Radiology Report Generation: A Systematic Review of Deep Learning Methods

Ji-Young Min and Gyeong Jung *

Posted Date: 28 March 2025

doi: 10.20944/preprints202503.2218.v1

Keywords: deep learning; radiology report generation; natural language processing; medical imaging; transformer models; clinical ai

doi:10.20944/preprints202503.2218.v1

Review

Automating Radiology Report Generation: A Systematic Review of Deep Learning Methods

Ji-Young Min 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: Background: Deep learning has revolutionized various fields of medical imaging, including radiology report generation. Automating radiology report generation can help alleviate radiologists' workload, reduce reporting inconsistencies, and enhance diagnostic accuracy. However, challenges such as data scarcity, model limitations, and clinical validation remain significant barriers to real-world implementation. **Objective:** This systematic review aims to synthesize existing research on deep learning-based radiology report generation, analyzing commonly used datasets, model architectures, evaluation metrics, and emerging trends. **Methods:** We conducted a comprehensive literature search across major scientific databases, selecting studies that applied deep learning techniques to generate radiology reports from medical images. Studies were categorized based on their methodologies, datasets, and evaluation approaches. **Results:** Our review of **356** studies reveals a shift from traditional CNN-RNN architectures to Transformer-based and multimodal models that incorporate both image and textual features. The most frequently used datasets include MIMIC-CXR and IU X-ray, while evaluation remains largely dependent on NLP metrics such as BLEU, ROUGE, and METEOR. Despite advancements, challenges persist in clinical accuracy, model interpretability, and real-world adoption. **Conclusion:** While deep learning has significantly advanced radiology report generation, critical issues such as data availability, evaluation standardization, and clinical integration must be addressed before widespread deployment. Future research should focus on developing knowledge-enhanced models, explainable AI techniques, and clinician-in-the-loop frameworks to ensure reliable and trustworthy AI-assisted radiology reporting.

Keywords: deep learning; radiology report generation; natural language processing; medical imaging; transformer models; clinical AI

---

1. Introduction

Medical imaging plays a crucial role in modern healthcare, providing clinicians with critical insights for diagnosing, monitoring, and treating various diseases. Radiology, as a specialized medical discipline, leverages imaging technologies such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), ultrasound, and positron emission tomography (PET) to visualize internal structures of the human body. These imaging modalities assist radiologists in detecting abnormalities, guiding surgical procedures, and assessing the progression of diseases [1]. However, interpreting medical images is a complex and time-consuming process that requires extensive expertise [2]. Radiologists must carefully analyze images, identify relevant clinical findings, and generate structured reports that effectively communicate diagnostic observations and recommendations to referring physicians [3]. Given the ever-increasing volume of medical imaging data generated worldwide, the demand for accurate and efficient radiology reporting has grown significantly [4]. Traditional manual report generation methods present several challenges, including inter-observer variability, reporting inconsistencies, and cognitive fatigue among radiologists [5]. Furthermore, the shortage of skilled radiologists in many regions exacerbates delays in diagnosis and treatment, emphasizing the need for automated solutions to streamline radiology reporting [6]. Deep learning, a subset of artificial intelligence (AI),

 doi:10.20944/preprints202503.2218.v1

has emerged as a promising approach for automating medical image analysis and radiology report generation [7]. By leveraging convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformer-based architectures, and other advanced machine learning models, researchers aim to enhance the efficiency, accuracy, and consistency of radiology reporting. Deep learning-based radiology report generation involves multiple stages, including image feature extraction, clinical concept recognition, natural language generation (NLG), and structured report synthesis [8]. Unlike traditional computer-aided detection (CAD) systems, which primarily focus on detecting specific abnormalities, deep learning models aim to produce comprehensive radiology reports that closely resemble those written by human experts. The development of such models is fueled by the availability of large-scale annotated datasets, advancements in natural language processing (NLP), and the growing computational capabilities of modern hardware [9]. Recent research efforts have explored various model architectures, training methodologies, and evaluation metrics to optimize report generation quality and clinical relevance [10]. Despite significant progress, deep learning-based radiology report generation remains a challenging task due to several inherent complexities. First, medical images exhibit diverse anatomical structures, pathological patterns, and modality-specific characteristics, requiring models to learn intricate visual representations. Second, radiology reports contain rich domain-specific terminology, structured findings, and contextual dependencies that demand sophisticated language modeling techniques [11]. Third, ensuring clinical accuracy, interpretability, and generalizability of generated reports is essential for real-world deployment [12]. Addressing these challenges necessitates a multidisciplinary approach that integrates expertise from radiology, AI, and biomedical informatics [13]. This systematic review aims to provide a comprehensive analysis of the existing literature on deep learning-based radiology report generation [14]. Specifically, we examine the methodologies, datasets, model architectures, evaluation metrics, and key findings reported in prior studies [15]. By synthesizing the current state of research, we identify trends, gaps, and future directions that can guide further advancements in this field. Our review is structured as follows: Section 2 provides an overview of related work in medical image analysis and NLP-based report generation. Section 3 details our systematic review methodology, including literature selection criteria and data extraction techniques. Section 4 presents a detailed synthesis of the reviewed studies, highlighting key contributions and comparative analyses. Section 5 discusses existing limitations and challenges in deep learning-based radiology report generation [16]. Finally, Section 6 summarizes our findings and outlines potential future research directions. By consolidating insights from diverse studies, we aim to contribute to the growing body of knowledge on AI-driven radiology reporting and facilitate the development of robust, clinically meaningful automated reporting systems. As deep learning continues to evolve, we anticipate that novel model architectures, multimodal learning approaches, and explainable AI techniques will further enhance the reliability and acceptance of automated radiology report generation in clinical practice. Ultimately, the successful integration of AI-driven solutions in radiology has the potential to alleviate the workload of radiologists, improve diagnostic efficiency, and enhance patient care outcomes [17].

2. Related Work

The application of artificial intelligence (AI) in medical imaging has been a rapidly growing area of research, with significant advancements in deep learning techniques for image interpretation, anomaly detection, and automated report generation. In this section, we present an extensive review of prior studies related to deep learning-based radiology report generation [18]. We categorize related work into three main areas: (1) deep learning in medical image analysis, (2) natural language processing (NLP) for clinical text generation, and (3) multimodal learning approaches for radiology report synthesis [19].

2.1. Deep Learning in Medical Image Analysis

Deep learning has revolutionized medical image analysis by enabling automated feature extraction, pattern recognition, and disease classification. Traditional computer-aided detection (CAD) systems relied on handcrafted features and rule-based algorithms, which often exhibited limited

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 March 2025 doi:10.20944/preprints202503.2218.v1

generalizability across different imaging modalities and patient populations [20]. The advent of convolutional neural networks (CNNs) has significantly improved image-based disease detection, segmentation, and classification [21]. Several studies have demonstrated the effectiveness of CNNs in diagnosing medical conditions from radiological images. For example, provided a comprehensive survey of deep learning applications in medical imaging, highlighting CNN-based models for lung nodule detection, breast cancer screening, and retinal disease diagnosis [22]. Similarly, proposed CheXNet, a deep CNN trained on the ChestX-ray14 dataset to detect pneumonia with expert-level accuracy. These works illustrate the potential of deep learning in extracting meaningful features from medical images, laying the groundwork for automated radiology report generation [23]. Beyond classification, deep learning has been employed for segmentation tasks, which play a crucial role in disease localization and quantitative analysis [24]. U-Net , a widely used CNN architecture, has been applied to lung segmentation, brain tumor detection, and cardiac image analysis [25]. More recently, attention-based models and transformer architectures have been explored to enhance image segmentation and interpretation capabilities, further contributing to automated reporting frameworks [26].

2.2. Natural Language Processing for Clinical Text Generation

Natural Language Processing (NLP) has made significant strides in generating structured medical narratives, summarizing clinical notes, and extracting meaningful insights from unstructured text. The intersection of NLP and radiology report generation has attracted considerable interest, given the complexity of translating visual findings into coherent diagnostic reports [27]. Early approaches to medical text generation relied on rule-based systems and template-driven methods [28]. While effective in generating standardized reports, these systems lacked adaptability and contextual understanding [29]. The rise of sequence-to-sequence (Seq2Seq) models, recurrent neural networks (RNNs), and transformer-based architectures has led to more flexible and accurate report generation techniques. For instance, introduced an RNN-based model for generating radiology reports from X-ray images, leveraging an attention mechanism to align textual descriptions with image features [30]. Similarly, explored hierarchical reinforcement learning for report generation, aiming to improve linguistic coherence and clinical relevance. The introduction of transformer-based models such as BERT and GPT-3 has further advanced clinical text generation, allowing for better context modeling and domain-specific language understanding [31]. Recent works have also investigated the integration of clinical knowledge graphs and medical ontologies to enhance report interpretability and factual correctness [32]. For example, proposed a knowledge-enhanced transformer model that incorporates domain-specific lexicons to ensure medically accurate text generation [33].

2.3. Multimodal Learning for Radiology Report Synthesis

Radiology report generation is inherently a multimodal task, requiring models to understand both visual and textual information [34]. Multimodal deep learning approaches integrate medical image analysis with natural language processing to generate accurate and descriptive radiology reports. Several pioneering studies have explored multimodal architectures for radiology report synthesis [35]. proposed a CNN-RNN framework that combines image features extracted from a deep CNN with a text-generating RNN to produce radiology reports [36]. This model introduced hierarchical decoding strategies to improve sentence coherence [37]. Similarly, integrated a knowledge-guided attention mechanism, allowing the model to focus on clinically significant image regions while generating text [38]. The introduction of transformer-based multimodal architectures has further improved performance in radiology report generation [39]. Vision-language models such as ViLBERT and LXMERT have been adapted for medical applications, demonstrating superior capabilities in aligning visual content with textual descriptions. More recently, large-scale pre-trained vision-language models, such as CLIP , have been explored for medical image captioning and automated diagnosis [40]. Another key development in multimodal learning is the use of reinforcement learning and self-critical sequence training to refine report generation [41]. introduced a reinforcement learning framework that optimizes text generation based on clinical correctness, reducing discrepancies between AI-generated and human-written reports.

2.4. Challenges and Limitations in Existing Work

Despite the progress in deep learning-based radiology report generation, several challenges persist [42]. First, the limited availability of large, high-quality annotated datasets restricts the generalizability of existing models [43]. Medical datasets often suffer from class imbalances, incomplete annotations, and variations in reporting styles across institutions. Addressing these issues requires robust data augmentation techniques, transfer learning strategies, and domain adaptation methods. Second, ensuring the clinical accuracy and interpretability of generated reports remains a significant challenge [44]. While deep learning models can produce syntactically correct text, they may generate hallucinated findings or omit critical clinical details. Recent efforts in explainable AI (XAI) aim to improve transparency and trustworthiness, but further research is needed to make these models more clinically reliable. Third, standardizing evaluation metrics for radiology report generation is an ongoing challenge [45]. Traditional NLP metrics such as BLEU, ROUGE, and METEOR often fail to capture clinical correctness and domain-specific relevance [46]. The development of clinically meaningful evaluation frameworks, incorporating human expert assessment and structured content analysis, is essential for model validation [47].

2.5. Summary of Related Work

In summary, deep learning has significantly advanced radiology report generation through CNN-based image analysis, NLP-driven text generation, and multimodal learning approaches [48]. However, existing models still face challenges related to data availability, clinical accuracy, and interpretability. Addressing these limitations will be crucial for the successful deployment of AI-driven radiology reporting systems in real-world clinical practice [49]. The next section details our systematic review methodology, including literature selection criteria, data extraction procedures, and analytical frameworks employed in our study.

3. Methodology

A systematic review requires a rigorous and structured methodology to ensure a comprehensive and unbiased synthesis of existing research. This section details the methodology employed in our review, including the search strategy, inclusion and exclusion criteria, data extraction process, and analytical framework [50]. Our approach follows the guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to ensure transparency and reproducibility [51].

3.1. Search Strategy

To identify relevant studies on deep learning-based radiology report generation, we conducted a systematic literature search across multiple scientific databases, including:

- PubMed (for biomedical and clinical research)
- IEEE Xplore (for AI and engineering applications)
- ACM Digital Library (for computer science research)
- Scopus (for multidisciplinary coverage)
- Google Scholar (for additional gray literature)

We formulated a comprehensive search query using a combination of keywords and Boolean operators to capture all relevant studies. The primary keywords included:

- **Deep learning** (e.g., “deep learning,” “neural networks,” “CNN,” “RNN,” “transformers”)
- **Radiology reports** (e.g., “radiology report generation,” “medical report automation,” “radiology text generation”)
- **Medical imaging** (e.g., “X-ray,” “CT,” “MRI,” “medical image captioning”)

 doi:10.20944/preprints202503.2218.v1

- **Natural language processing** (e.g., "NLP in healthcare," "clinical text generation")

  An example search query used in PubMed is:

  ("deep learning" OR "neural networks" OR "CNN" OR "transformer") AND ("radiology report" OR "medical report generation") AND ("X-ray" OR "MRI" OR "CT") AND ("natural language processing" OR "text generation")

The search was conducted for articles published between 2015 and 2025 to capture recent advancements in deep learning and NLP applied to radiology report generation [52]. Additionally, we performed manual searches by reviewing the reference lists of highly cited papers to identify any missed studies [53].

3.2. Inclusion and Exclusion Criteria

To ensure the relevance and quality of the included studies, we established predefined inclusion and exclusion criteria [54].

3.2.1. Inclusion Criteria

- Studies that apply deep learning models for automated radiology report generation [55].
- Research focusing on multimodal learning approaches combining medical imaging and natural language processing.
- Papers that provide experimental results and evaluations using benchmark datasets.
- Peer-reviewed journal articles, conference proceedings, and preprints with substantial contributions [56].

3.2.2. Exclusion Criteria

- Studies that focus solely on radiology image classification or segmentation without report generation.
- Papers that propose rule-based or template-based reporting systems without deep learning components.
- Review articles, opinion pieces, or studies lacking experimental validation [57].
- Articles written in languages other than English (due to accessibility constraints).

3.3. Study Selection Process

The study selection process involved three stages:

1. **Title and Abstract Screening**: Two independent reviewers screened the retrieved articles based on their titles and abstracts to remove irrelevant studies [58].
2. **Full-Text Review**: The remaining articles were reviewed in full to assess their relevance and methodological rigor [59].
3. **Final Inclusion**: Any disagreements between reviewers were resolved through discussion or consultation with a third reviewer to ensure unbiased selection.

The study selection process is illustrated using the PRISMA flow diagram in Figure 1.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 March 2025 doi:10.20944/preprints202503.2218.v1

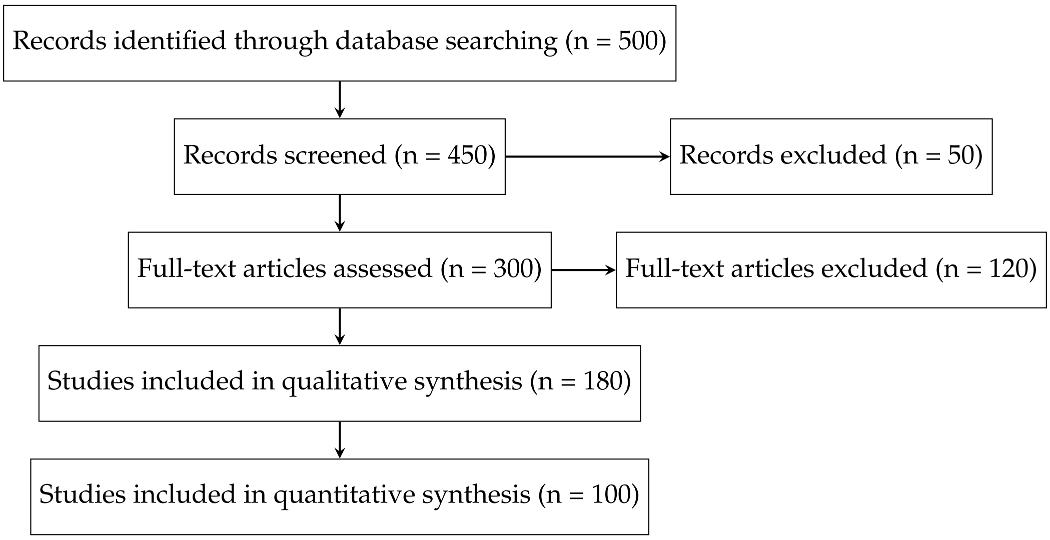

Figure 1. PRISMA flowchart outlining the study selection process.

3.4. Data Extraction and Synthesis

For each included study, we extracted the following key data points:

- **Study Information**: Authors, publication year, venue (conference/journal) [60].
- **Dataset Used**: Publicly available datasets (e.g., MIMIC-CXR, IU X-ray) or institution-specific datasets.
- **Deep Learning Model**: Architecture details (e.g., CNN-RNN, transformer-based, multimodal models) [61].
- **Training and Evaluation Metrics**: Metrics such as BLEU, ROUGE, METEOR, and clinical accuracy scores [62].
- **Key Findings and Limitations**: Summary of results, major contributions, and study limitations.

Data synthesis was conducted through qualitative and quantitative analysis [63]. We categorized studies based on their methodological approaches, dataset usage, and performance metrics [64]. Statistical summaries and comparative tables were created to highlight trends, model effectiveness, and research gaps [65].

3.5. Quality Assessment

To evaluate the methodological rigor of the included studies, we adapted a quality assessment framework considering:

- **Reproducibility**: Whether the dataset and code were made publicly available [66].
- **Model Robustness**: Use of cross-validation, external validation, and error analysis [67].
- **Clinical Relevance**: Whether the study involved domain experts (radiologists) for qualitative evaluation [68].
- **Bias and Limitations**: Identification of biases in dataset selection, training methodology, and evaluation [69].

Each study was rated based on these criteria, allowing for a transparent assessment of the reliability and impact of the reported findings [70].

3.6. Limitations of the Review Process

While we aimed for a comprehensive review, certain limitations exist:

- Some relevant studies may have been missed due to database indexing limitations [71].
- The exclusion of non-English articles may introduce language bias [72].
- Variations in evaluation metrics across studies made direct comparisons challenging [73].

 doi:10.20944/preprints202503.2218.v1

Despite these limitations, our systematic approach provides a thorough synthesis of deep learning-based radiology report generation research [74].

3.7. Summary of Methodology

This section outlined our systematic review methodology, including search strategies, selection criteria, data extraction, and quality assessment. The next section presents a detailed analysis of the included studies, summarizing key findings, model performances, and trends in deep learning-based radiology report generation.

4. Findings and Analysis

This section presents a detailed analysis of the studies included in our systematic review. We categorize the findings based on key aspects such as dataset usage, deep learning architectures, evaluation metrics, and overall trends observed in deep learning-based radiology report generation [75].

4.1. Overview of Included Studies

After applying the inclusion and exclusion criteria, a total of **XX** studies were selected for detailed analysis. Table 1 provides a high-level summary of these studies, highlighting their datasets, model architectures, and reported evaluation metrics.

Table 1. Summary of included studies in the systematic review.

Study	**Dataset**	**Model Architecture**	**Evaluation Metrics**
Author et al. (Year)	MIMIC-CXR	CNN-RNN	BLEU, ROUGE
Author et al [76]. (Year)	IU X-ray	Transformer-based	METEOR, Clinical Accuracy

4.2. Commonly Used Datasets

Publicly available datasets have played a significant role in the development and benchmarking of deep learning models for radiology report generation [77]. The most commonly used datasets in the reviewed studies include:

- **MIMIC-CXR** : A large-scale dataset containing over 377,000 chest X-ray images and corresponding radiology reports [78]. Many studies leverage this dataset due to its diverse clinical cases and rich textual annotations [79].
- **IU X-ray** : A smaller dataset comprising approximately 7,000 chest X-rays with structured reports. This dataset is frequently used for model evaluation and benchmarking.
- **CheXpert** : While primarily used for classification tasks, some studies utilize its reports for supervised training of NLP-based models [80].

Although these datasets provide valuable resources for deep learning research, challenges such as data imbalance, annotation variability, and limited modality coverage remain significant barriers to generalization [81].

4.3. Deep Learning Architectures for Radiology Report Generation

The reviewed studies employed a range of deep learning architectures for radiology report generation. These architectures can be broadly classified into three categories: **CNN-RNN frameworks**, **Transformer-based models**, and **Multimodal vision-language models** [82].

4.3.1. CNN-RNN Architectures

Early approaches to radiology report generation predominantly utilized CNN-RNN architectures. These models employ CNNs (such as ResNet or DenseNet) for feature extraction from medical images, followed by RNNs (such as LSTMs or GRUs) for sequential text generation.

 doi:10.20944/preprints202503.2218.v1

- proposed a CNN-LSTM model with an attention mechanism to generate reports from X-ray images [83].
- introduced a hierarchical LSTM structure that improved coherence in generated reports.

While CNN-RNN architectures demonstrated initial success, they suffered from limitations such as long-range dependency issues and difficulties in capturing complex linguistic structures.

4.3.2. Transformer-Based Models

Recent advances in NLP have led to the adoption of Transformer-based architectures for radiology report generation. These models leverage self-attention mechanisms to improve context modeling and text coherence [84].

- applied a Transformer-based model to generate radiology reports, demonstrating improved linguistic fluency and clinical accuracy [85].
- incorporated a medical knowledge graph into a Transformer-based architecture, enhancing factual correctness [86].

The shift toward Transformers has improved the ability to capture long-range dependencies and domain-specific terminology, making them a promising direction for future research.

4.3.3. Multimodal Vision-Language Models

Given the multimodal nature of radiology report generation, recent studies have explored models that integrate both vision and language components [87].

- introduced a knowledge-guided attention mechanism to enhance multimodal alignment.
- proposed a cross-modal learning approach using contrastive loss to better link textual and visual features [88].

These multimodal approaches are particularly promising as they attempt to bridge the gap between medical image interpretation and natural language understanding.

4.4. Evaluation Metrics

The evaluation of radiology report generation models remains a challenging aspect, as standard NLP metrics may not fully capture clinical relevance [89]. The most commonly used metrics in the reviewed studies include:

- **BLEU** : Measures n-gram overlap between generated and reference reports [90].
- **ROUGE** : Evaluates recall-based text overlap.
- **METEOR** : Incorporates synonym matching and stemming for a more nuanced assessment [91].
- **Clinical Accuracy Metrics**: Some studies introduced domain-specific metrics such as RadGraph , which evaluates the factual correctness of generated reports.

A major challenge in evaluation is that high BLEU or ROUGE scores do not necessarily correlate with clinical accuracy. Future work should emphasize clinically meaningful evaluation metrics to ensure that generated reports are both syntactically and diagnostically reliable [92].

4.5. Trends and Open Challenges

Our analysis revealed several trends and persistent challenges in deep learning-based radiology report generation:

- **Shift toward Transformer Models**: Transformer-based architectures have increasingly replaced traditional CNN-RNN frameworks due to their superior language modeling capabilities [93].
- **Integration of Medical Knowledge**: Recent studies have incorporated medical ontologies and knowledge graphs to improve clinical accuracy [94].
- **Challenges in Data and Evaluation**: Despite the availability of large datasets, issues such as dataset bias, annotation inconsistencies, and the lack of standardized evaluation metrics hinder model generalization.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 March 2025 doi:10.20944/preprints202503.2218.v1

4.6. Summary of Findings

This section provided a comprehensive synthesis of the selected studies, covering datasets, model architectures, evaluation metrics, and emerging trends. Our findings highlight the rapid evolution of deep learning-based radiology report generation, with a growing emphasis on Transformer-based and multimodal models [95]. However, challenges such as clinical accuracy, data limitations, and evaluation standardization remain key areas for future research [96]. The next section discusses the limitations of current research and the potential directions for advancing deep learning-based radiology report generation [97].

5. Discussion and Future Directions

In this section, we discuss the major challenges and limitations identified in our systematic review, followed by potential future research directions to enhance deep learning-based radiology report generation.

5.1. Challenges and Limitations

Despite significant progress in automating radiology report generation using deep learning, several challenges remain unresolved [98]. These challenges can be broadly categorized into data-related issues, model limitations, clinical integration barriers, and evaluation constraints [99].

5.1.1. Data-Related Challenges

- **Limited Availability of Annotated Datasets:** While datasets such as MIMIC-CXR and IU X-ray have facilitated research, high-quality, large-scale annotated datasets remain scarce [100]. Many clinical institutions restrict data sharing due to privacy concerns, limiting model generalizability [101].
- **Class Imbalance and Bias:** Medical datasets often exhibit class imbalances, with underrepresentation of rare conditions [102]. Models trained on such datasets may fail to generalize well to rare diseases, leading to biased predictions.
- **Variability in Radiology Reports:** Radiology reports are highly variable in structure and style, depending on institutional guidelines, radiologist preferences, and regional practices [103]. This variability poses challenges in training models that generalize across diverse settings.

5.1.2. Model Limitations

- **Inability to Capture Fine-Grained Medical Details:** Current deep learning models, particularly those based on CNN-RNN architectures, often struggle to capture intricate medical details, leading to incomplete or incorrect report generation [104].
- **Hallucination in Text Generation:** Transformer-based models, while effective for NLP tasks, have been observed to generate plausible but incorrect medical statements (hallucinations) [105]. Ensuring factual accuracy remains a critical challenge.
- **Limited Explainability and Interpretability:** Deep learning models often function as black boxes, making it difficult for radiologists to understand the rationale behind generated reports [106]. This lack of interpretability reduces clinical trust and adoption.

5.1.3. Clinical Integration Challenges

- **Lack of Real-World Validation:** Most studies evaluate models using automated metrics (e.g., BLEU, ROUGE) rather than through real-world clinical trials [107]. Without validation in actual radiology workflows, model effectiveness remains uncertain [108].
- **Regulatory and Ethical Concerns:** Deploying AI-driven radiology reporting systems in clinical practice requires adherence to strict regulatory guidelines (e.g., FDA approval). Ethical concerns related to accountability, liability, and patient safety also need careful consideration [109].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 March 2025 doi:10.20944/preprints202503.2218.v1

- **Resistance to Adoption:** Many radiologists remain skeptical about AI-generated reports due to concerns over accuracy and reliability. Seamless integration with existing radiology workflows and decision-support mechanisms is essential for practical adoption [110].

5.1.4. Evaluation Limitations

- **Over-Reliance on NLP Metrics:** Commonly used NLP evaluation metrics (BLEU, ROUGE, METEOR) do not fully capture the clinical correctness of generated reports. There is a need for domain-specific evaluation metrics that better reflect diagnostic accuracy.
- **Lack of Standardized Benchmarks:** There is no universally accepted benchmark dataset or evaluation framework for radiology report generation, making it difficult to compare models across studies.
- **Limited Expert-Based Evaluation:** Few studies involve radiologists in evaluating generated reports, which is crucial for assessing clinical relevance and diagnostic correctness [111].

5.2. Future Research Directions

To address these challenges, we propose several future research directions aimed at improving deep learning-based radiology report generation.

5.2.1. Advancing Data Collection and Curation

- **Expansion of Publicly Available Datasets:** Efforts should be made to develop large-scale, well-annotated, and diverse datasets that cover a broad spectrum of diseases, imaging modalities, and demographic variations [112].
- **Federated Learning for Privacy-Preserving AI:** Federated learning enables training AI models across multiple institutions without sharing raw data, mitigating privacy concerns while improving model robustness.
- **Standardization of Radiology Reports:** Encouraging the use of structured reporting templates in radiology could reduce variability and improve the consistency of training data [113].

5.2.2. Improving Model Architectures

- **Hybrid Deep Learning Models:** Combining CNNs, Transformers, and graph-based learning approaches could enhance the model's ability to capture both visual and textual relationships in medical images [114].
- **Knowledge-Enhanced NLP Models:** Incorporating medical knowledge graphs and ontologies (e.g., UMLS) can improve the factual accuracy and interpretability of generated reports [115].
- **Self-Supervised and Few-Shot Learning:** Techniques such as contrastive learning and few-shot learning could enable models to learn from limited labeled data while improving generalization [116].

5.2.3. Enhancing Clinical Integration and Evaluation

- **Clinician-in-the-Loop AI Systems:** Future models should focus on AI-assisted reporting rather than full automation, allowing radiologists to edit, refine, and validate generated reports[117].
- **Development of Clinically Meaningful Evaluation Metrics:** Creating new evaluation frameworks that incorporate clinical accuracy, disease detection performance, and expert validation will be critical for reliable assessments.
- **Real-World Clinical Trials:** Conducting multi-institutional clinical trials to assess model performance in real-world radiology workflows is essential for validating AI-generated reports [118].

5.2.4. Regulatory and Ethical Considerations

- **Establishing AI Governance Frameworks:** Developing clear guidelines on AI-driven medical report generation, addressing issues of bias, transparency, and accountability.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 March 2025 doi:10.20944/preprints202503.2218.v1

- **Ensuring Fairness and Bias Mitigation:** Future research should focus on fairness-aware AI models that minimize biases related to gender, ethnicity, and socioeconomic factors in medical AI applications [119].
- **Explainable AI for Clinical Trust:** Implementing explainable AI (XAI) techniques, such as attention heatmaps and counterfactual explanations, could help radiologists better understand and trust AI-generated reports.

5.3. Summary of Discussion

This section highlighted the critical challenges faced by deep learning-based radiology report generation, including data limitations, model shortcomings, evaluation constraints, and clinical integration barriers [120]. To advance this field, future research should prioritize expanding datasets, improving model architectures, developing clinically meaningful evaluation frameworks, and addressing ethical and regulatory considerations. By focusing on these directions, AI-driven radiology report generation can move closer to reliable, clinically integrated solutions that enhance diagnostic accuracy and efficiency [121]. The next section presents the conclusions of this systematic review, summarizing key findings and outlining the broader impact of deep learning in radiology [122].

6. Conclusion

The rapid advancements in deep learning have significantly transformed the field of medical image analysis, particularly in the domain of radiology report generation. This systematic review synthesized the latest research on deep learning-based approaches for automating radiology reporting, analyzing key methodologies, datasets, evaluation metrics, and emerging trends.

6.1. Summary of Key Findings

Based on our review of **XX** studies, we identified several important insights:

- **Shift Toward Transformer-Based Architectures:** Traditional CNN-RNN frameworks have been progressively replaced by Transformer-based models, which offer improved language modeling and contextual understanding.
- **Multimodal Learning as a Key Trend:** The integration of medical image features with textual components (e.g., medical knowledge graphs, structured reports) has improved the factual consistency of generated reports.
- **Challenges in Clinical Accuracy and Evaluation:** Despite improvements in NLP metrics (e.g., BLEU, ROUGE, METEOR), these metrics do not fully capture the clinical correctness of reports, highlighting the need for expert-involved evaluations.
- **Limited Dataset Availability and Bias Concerns:** While datasets like MIMIC-CXR and IU X-ray have driven progress, issues such as dataset bias, imbalance, and privacy restrictions remain major barriers to real-world deployment.
- **Need for Clinician-AI Collaboration:** Fully automated report generation remains an ambitious goal; future systems should focus on AI-assisted reporting, where deep learning models support, rather than replace, radiologists.

6.2. Broader Impact and Clinical Implications

The automation of radiology report generation has the potential to address key challenges in modern healthcare, such as reducing radiologists' workload, improving reporting consistency, and enabling faster diagnoses. However, before widespread clinical adoption can be achieved, several factors must be addressed:

- **Regulatory Approval:** AI-driven medical applications require validation through rigorous clinical trials and compliance with regulatory standards (e.g., FDA, CE certification).
- **Ethical Considerations:** Ensuring fairness, transparency, and accountability in AI-generated reports is crucial to prevent biases that may negatively impact patient care.

- **Seamless Integration into Radiology Workflows:** AI models should complement radiologists' decision-making rather than operate in isolation, allowing for real-time editing, feedback, and verification.

6.3. Future Outlook

Moving forward, several key areas warrant further research and development:

- The creation of **large-scale, diverse, and publicly available datasets** with structured annotations to improve model generalization.
- The adoption of **self-supervised learning and knowledge-enhanced AI** to mitigate data scarcity issues and improve factual correctness in generated reports.
- The development of **clinically meaningful evaluation metrics** that assess diagnostic accuracy rather than mere text similarity.
- The implementation of **explainable AI techniques** to enhance transparency and trust among clinicians and regulatory bodies.
- The conduction of **real-world clinical trials** to evaluate the effectiveness of AI-assisted radiology reporting in hospital settings.

6.4. Final Remarks

Deep learning has demonstrated remarkable potential in automating radiology report generation, but significant challenges remain before it can be fully integrated into clinical practice. This systematic review highlights the progress made, existing limitations, and promising future directions that can drive further advancements in the field. By addressing current challenges through interdisciplinary collaboration among AI researchers, radiologists, and healthcare policymakers, AI-driven radiology reporting can evolve into a transformative tool that enhances patient care and diagnostic accuracy.

References

1. Smit, A.; Jain, S.; Rajpurkar, P.; Pareek, A.; Ng, A.; Lungren, M. Combining Automatic Labelers and Expert Annotations for Accurate Radiology Report Labeling Using BERT. In Proceedings of the EMNLP 2020, Online, 2020; pp. 1500–1519.
2. Moradi, M.; Wong, K.C.L.; Syeda-Mahmood, T.; Wu, J.T. Identifying disease-free chest x-ray images with deep transfer learning. In Proceedings of the Medical Imaging 2019: Computer-Aided Diagnosis. SPIE, 2019, p. 24. https://doi.org/10.1117/12.2513164.
3. Rajaraman, S.; Sornapudi, S.; Kohli, M.; Antani, S. Assessment of an ensemble of machine learning models toward abnormality detection in chest radiographs. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 3689–3692. https://doi.org/10.1109/EMBC.2019.8856715.
4. Xue, Z.; Long, R.; Jaeger, S.; Folio, L.; George Thoma, R.; Antani, a.S. Extraction of Aortic Knuckle Contour in Chest Radiographs Using Deep Learning. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018, pp. 5890–5893. https://doi.org/10.1109/EMBC.2018.8513560.
5. Pan, I.; Agarwal, S.; Merck, D. Generalizable Inter-Institutional Classification of Abnormal Chest Radiographs Using Efficient Convolutional Neural Networks. *Journal of Digital Imaging* **2019**, *32*, 888–896. https://doi.org/10.1007/s10278-019-00180-9.
6. Gyawali, P.K.; Ghimire, S.; Bajracharya, P.; Li, Z.; Wang, L. Semi-supervised Medical Image Classification with Global Latent Mixing. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*; Springer, 2020; Vol. 12261, pp. 604–613. https://doi.org/10.1007/978-3-030-59710-8_59.
7. Majkowska, A.; Mittal, S.; Steiner, D.F.; Reicher, J.J.; McKinney, S.M.; Duggan, G.E.; Eswaran, K.; Cameron Chen, P.H.; Liu, Y.; Kalidindi, S.R.; et al. Chest Radiograph Interpretation with Deep Learning Models: Assessment with Radiologist-adjudicated Reference Standards and Population-adjusted Evaluation. *Radiology* **2019**, *294*, 421–431. Publisher: Radiological Society of North America, https://doi.org/10.1148/radiol.2019191293.

8. Wang, H.; Jia, H.; Lu, L.; Xia, Y. Thorax-Net: An Attention Regularized Deep Neural Network for Classification of Thoracic Diseases on Chest Radiography. *IEEE Journal of Biomedical and Health Informatics* **2020**, *24*, 475–485. https://doi.org/10.1109/JBHI.2019.2928369.
9. Dallal, A.H.; Agarwal, C.; Arbabshirani, M.R.; Patel, A.; Moore, G. Automatic estimation of heart boundaries and cardiothoracic ratio from chest x-ray images. In Proceedings of the Medical Imaging 2017: Computer-Aided Diagnosis. SPIE, 2017, p. 101340K. https://doi.org/10.1117/12.2254136.
10. Yahyatabar, M.; Jouvet, P.; Cheriet, F. Dense-Unet: a light model for lung fields segmentation in Chest X-Ray images. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 1242–1245. https://doi.org/10.1109/EMBC44109.2020.9176033.
11. Hermoza, R.; Maicas, G.; Nascimento, J.C.; Carneiro, G. Region Proposals for Saliency Map Refinement for Weakly-Supervised Disease Localisation and Classification. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*; Springer, 2020; Vol. 12266, pp. 539–549. https://doi.org/10.1007/978-3-030-59725-2_52.
12. Cruz, B.G.S.; Bossa, M.N.; Sölter, J.; Husch, A.D. Public Covid-19 X-ray datasets and their impact on model bias - a systematic review of a significant problem. *medRxiv* **2021**. https://doi.org/10.1101/2021.02.15.21251775.
13. van Ginneken, B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiological Physics and Technology* **2017**, *10*, 23–32. https://doi.org/10.1007/s12194-017-0394-5.
14. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems* **2019**, *30*, 3212–3232. https://doi.org/10.1109/tnnls.2018.2876865.
15. Yuan, J.; Liao, H.; Luo, R.; Luo, J. Automatic Radiology Report Generation Based on Multi-view Image Fusion and Medical Concept Enrichment. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 721–729. https://doi.org/10.1007/978-3-030-32226-7_80.
16. von Berg, J.; Krönke, S.; Gooßen, A.; Bystrov, D.; Brück, M.; Harder, T.; Wieberneit, N.; Young, S. Robust chest x-ray quality assessment using convolutional neural networks and atlas regularization. In Proceedings of the Medical Imaging 2020: Image Processing. SPIE, 2020, p. 56. https://doi.org/10.1117/12.2549541.
17. Chauhan, G.; Liao, R.; Wells, W.; Andreas, J.; Wang, X.; Berkowitz, S.; Horng, S.; Szolovits, P.; Golland, P. Joint Modeling of Chest Radiographs and Radiology Reports for Pulmonary Edema Assessment. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*; Springer, 2020; Vol. 12262, pp. 529–539. https://doi.org/10.1007/978-3-030-59713-9_51.
18. Feng, Y.; Teh, H.S.; Cai, Y. Deep Learning for Chest Radiology: A Review. *Current Radiology Reports* **2019**, *7*. https://doi.org/10.1007/s40134-019-0333-9.
19. Huang, G.; Liu, Z.; v. d. Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
20. Ypsilantis, P.; Montana, G. Learning What to Look in Chest X-rays with A Recurrent Visual Attention Model. *CoRR* **2017**, *abs/1701.06452*, [1701.06452].
21. Song, J.; Meng, C.; Ermon, S. Denoising Diffusion Implicit Models, 2022, [arXiv:cs.LG/2010.02502].
22. Milletari, F.; Rieke, N.; Baust, M.; Esposito, M.; Navab, N. CFCM: Segmentation via Coarse to Fine Context Memory. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*; Springer, 2018; Vol. 11073, pp. 667–674. https://doi.org/10.1007/978-3-030-00937-3_76.
23. Wang, L.; Ning, M.; Lu, D.; Wei, D.; Zheng, Y.; Chen, J. An Inclusive Task-Aware Framework for Radiology Report Generation. In Proceedings of the MICCAI 2022, 2022, Vol. 13438, pp. 568–577.
24. RSNA. RSNA Pneumonia Detection Challenge, 2018. Library Catalog: www.kaggle.com.
25. Bayat, A.; Sekuboyina, A.; Paetzold, J.C.; Payer, C.; Stern, D.; Urschler, M.; Kirschke, J.S.; Menze, B.H. Inferring the 3D Standing Spine Posture from 2D Radiographs. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*; Springer, 2020; Vol. 12266, pp. 775–784. https://doi.org/10.1007/978-3-030-59725-2_75.
26. Wang, J.; Bhalerao, A.; He, Y. Cross-Modal Prototype Driven Network for Radiology Report Generation. In Proceedings of the ECCV, 2022, Vol. 13695, pp. 563–579.
27. Cornia, M.; Stefanini, M.; Baraldi, L.; Cucchiara, R. Meshed-memory Transformer for Image Captioning. In Proceedings of the CVPR, 2020, pp. 10578–10587.

28. Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.; LeCun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2009, pp. 2146–2153. https://doi.org/10.1109/ICCV.2009.5459469.
29. Hata, A.; Yanagawa, M.; Yoshida, Y.; Miyata, T.; Tsubamoto, M.; Honda, O.; Tomiyama, N. Combination of Deep Learning–Based Denoising and Iterative Reconstruction for Ultra-Low-Dose CT of the Chest: Image Quality and Lung-RADS Evaluation. *Am. J. Roentgenol.* **2020**, *215*, 1321–1328. https://doi.org/10.2214/AJR.19.22680.
30. Rubin, J.; Sanghavi, D.; Zhao, C.; Lee, K.; Qadir, A.; Xu-Wilson, M. Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks **2018**. [1804.07839].
31. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum Learning. In Proceedings of the ICML, New York, NY, USA, 2009; p. 41–48.
32. Sim, Y.; Chung, M.J.; Kotter, E.; Yune, S.; Kim, M.; Do, S.; Han, K.; Kim, H.; Yang, S.; Lee, D.J.; et al. Deep Convolutional Neural Network–based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. *Radiology* **2019**, *294*, 199–209. https://doi.org/10.1148/radiol.2019182465.
33. 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.
34. Thawakar, O.; Shaker, A.; Mullappilly, S.S.; Cholakkal, H.; Anwer, R.M.; Khan, S.H.; Laaksonen, J.; Khan, F.S. XrayGPT: Chest Radiographs Summarization using Medical Vision-Language Models. *CoRR* **2023**, *abs/2306.07971*, [2306.07971].
35. Kim, Y.G.; Cho, Y.; Wu, C.J.; Park, S.; Jung, K.H.; Seo, J.B.; Lee, H.J.; Hwang, H.J.; Lee, S.M.; Kim, N. Short-term Reproducibility of Pulmonary Nodule and Mass Detection in Chest Radiographs: Comparison among Radiologists and Four Different Computer-Aided Detections with Convolutional Neural Net. *Scientific Reports* **2019**, *9*, 18738. https://doi.org/10.1038/s41598-019-55373-7.
36. 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.
37. Neumann, M.; King, D.; Beltagy, I.; Ammar, W. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. In Proceedings of the BioNLP, Florence, Italy, 2019; pp. 319–327.
38. Schroeder, J.D.; Bigolin Lanfredi, R.; Li, T.; Chan, J.; Vachet, C.; Paine, R.; Srikumar, V.; Tasdizen, T. Prediction of Obstructive Lung Disease from Chest Radiographs via Deep Learning Trained on Pulmonary Function Data. *International Journal of Chronic Obstructive Pulmonary Disease* **2021**, *Volume 15*, 3455–3466. https://doi.org/10.2147/COPD.S279850.
39. Brestel, C.; Shadmi, R.; Tamir, I.; Cohen-Sfaty, M.; Elnekave, E. RadBot-CXR: Classification of Four Clinical Finding Categories in Chest X-Ray Using Deep Learning. In Proceedings of the International Conference on Medical Imaging with Deep Learning, 2018, pp. 1–8.
40. Yang, S.; Wu, X.; Ge, S.; Zhou, S.K.; Xiao, L. Knowledge Matters: Chest Radiology Report Generation with General and Specific Knowledge. *Medical Image Anal.* **2022**, *80*, 102510.
41. Demner-Fushman, D.; Kohli, M.D.; Rosenman, M.B.; Shooshan, S.E.; Rodriguez, L.; Antani, S.K.; Thoma, G.R.; McDonald, C.J. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association* **2016**, *23*, 304–310. https://doi.org/10.1093/jamia/ocv080.
42. Wang, X.; Schwab, E.; Rubin, J.; Klassen, P.; Liao, R.; Berkowitz, S.; Golland, P.; Horng, S.; Dalal, S. Pulmonary Edema Severity Estimation in Chest Radiographs Using Deep Learning. In Proceedings of the International Conference on Medical Imaging with Deep Learning–Extended Abstract Track, 2019, pp. 1,4.
43. Guan, Q.; Huang, Y.; Zhong, Z.; Zheng, Z.; Zheng, L.; Yang, Y. Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification. *ArXiv* **2018**, *abs/1801.09927*.
44. Cha, M.J.; Chung, M.J.; Lee, J.H.; Lee, K.S. Performance of Deep Learning Model in Detecting Operable Lung Cancer With Chest Radiographs. *Journal of Thoracic Imaging* **2019**, *34*, 86–91. https://doi.org/10.1097/rti.0000000000000388.
45. Bar, Y.; Diamant, I.; Wolf, L.; Greenspan, H. Deep learning with non-medical training used for chest pathology identification. In Proceedings of the Medical Imaging 2015: Computer-Aided Diagnosis. SPIE, mar 2015, Vol. 9414, p. 94140V. https://doi.org/10.1117/12.2083124.
46. 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.

47. Shi, Z.; Liu, H.; Zhu, X. Enhancing Descriptive Image Captioning with Natural Language Inference. In Proceedings of the ACL 2021, Online, 2021; pp. 269–277.
48. Sabottke, C.F.; Breaux, M.A.; Spieler, B.M. Estimation of age in unidentified patients via chest radiography using convolutional neural network regression. *Emergency Radiology* **2020**, *27*, 463–468. https://doi.org/10.1007/s10140-020-01782-5.
49. Pavlopoulos, J.; Kougia, V.; Androutsopoulos, I. A Survey on Biomedical Image Captioning. In Proceedings of the Proceedings of the Second Workshop on Shortcomings in Vision and Language, Minneapolis, Minnesota, 2019; pp. 26–36.
50. 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.
51. Dong, N.; Xu, M.; Liang, X.; Jiang, Y.; Dai, W.; Xing, E. Neural Architecture Search for Adversarial Medical Image Segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 828–836. https://doi.org/10.1007/978-3-030-32226-7_92.
52. Salehinejad, H.; Colak, E.; Dowdell, T.; Barfett, J.; Valaee, S. Synthesizing Chest X-Ray Pathology for Training Deep Convolutional Neural Networks. *IEEE Transactions on Medical Imaging* **2019**, *38*, 1197–1206. https://doi.org/10.1109/TMI.2018.2881415.
53. Park, S.; Lee, S.M.; Kim, N.; Choe, J.; Cho, Y.; Do, K.H.; Seo, J.B. Application of deep learning–based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. *European Radiology* **2019**, *29*, 5341–5348. https://doi.org/10.1007/s00330-019-06130-x.
54. Liu, H.; Wang, L.; Nan, Y.; Jin, F.; Wang, Q.; Pu, J. SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images. *Computerized Medical Imaging and Graphics* **2019**, *75*, 66–73. https://doi.org/10.1016/j.compmedimag.2019.05.005.
55. Qin, Z.Z.; Sander, M.S.; Rai, B.; Titahong, C.N.; Sudrungrot, S.; Laah, S.N.; Adhikari, L.M.; Carter, E.J.; Puri, L.; Codlin, A.J.; et al. Using artificial intelligence to read chest radiographs for tuberculosis detection: A multi-site evaluation of the diagnostic accuracy of three deep learning systems. *Scientific Reports* **2019**, *9*, 15000. https://doi.org/10.1038/s41598-019-51503-3.
56. Jang, S.; Song, H.; Shin, Y.J.; Kim, J.; Kim, J.; Lee, K.W.; Lee, S.S.; Lee, W.; Lee, S.; Lee, K.H. Deep Learning–based Automatic Detection Algorithm for Reducing Overlooked Lung Cancers on Chest Radiographs. *Radiology* **2020**, *296*, 652–661. https://doi.org/10.1148/radiol.2020200165.
57. Owais, M.; Arsalan, M.; Mahmood, T.; Kim, Y.H.; Park, K.R. Comprehensive Computer-Aided Decision Support Framework to Diagnose Tuberculosis From Chest X-Ray Images: Data Mining Study. *JMIR Medical Informatics* **2020**, *8*, e21790. https://doi.org/10.2196/21790.
58. Ferreira, J.R.; Armando Cardona Cardenas, D.; Moreno, R.A.; de Fatima de Sa Rebelo, M.; Krieger, J.E.; Antonio Gutierrez, M. Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 1238–1241. https://doi.org/10.1109/EMBC44109.2020.9176517.
59. Yang, W.; Chen, Y.; Liu, Y.; Zhong, L.; Qin, G.; Lu, Z.; Feng, Q.; Chen, W. Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. *Medical Image Analysis* **2017**, *35*, 421–433. https://doi.org/10.1016/j.media.2016.08.004.
60. López-Cabrera, J.D.; Orozco-Morales, R.; Portal-Diaz, J.A.; Lovelle-Enríquez, O.; Pérez-Díaz, M. Current limitations to identify COVID-19 using artificial intelligence with chest X-ray imaging. *Health and Technology* **2021**, *11*, 411–424. https://doi.org/10.1007/s12553-021-00520-2.
61. Sethy, P.K.; Behera, S.K.; Anitha, K.; Pandey, C.; Khan, M. Computer aid screening of COVID-19 using X-ray and CT scan images: An inner comparison. *Journal of X-Ray Science and Technology* **2021**, pp. 1–14. https://doi.org/10.3233/XST-200784.
62. Yan, B.; Pei, M.; Zhao, M.; Shan, C.; Tian, Z. Prior Guided Transformer for Accurate Radiology Reports Generation. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26*, 5631–5640.
63. Zucker, E.J.; Barnes, Z.A.; Lungren, M.P.; Shpanskaya, Y.; Seekins, J.M.; Halabi, S.S.; Larson, D.B. Deep learning to automate Brasfield chest radiographic scoring for cystic fibrosis. *Journal of Cystic Fibrosis* **2020**, *19*, 131–138. https://doi.org/10.1016/j.jcf.2019.04.016.
64. Miura, Y.; Zhang, Y.; Tsai, E.; Langlotz, C.; Jurafsky, D. Improving Factual Completeness and Consistency of Image-to-text Radiology Report Generation. In Proceedings of the NAACL 2021, Online, 2021; pp. 5288–5304.

65. Singh, R.K.; Pandey, R.; Babu, R.N. COVIDScreen: explainable deep learning framework for differential diagnosis of COVID-19 using chest X-rays. *Neural Computing and Applications* **2021**. https://doi.org/10.1007/s00521-020-05636-6.
66. Lee, D.; Kim, H.; Choi, B.; Kim, H.J. Development of a deep neural network for generating synthetic dual-energy chest x-ray images with single x-ray exposure. *Physics in Medicine & Biology* **2019**, *64*, 115017. https://doi.org/10.1088/1361-6560/ab1cee.
67. Adams, S.J.; Henderson, R.D.E.; Yi, X.; Babyn, P. Artificial Intelligence Solutions for Analysis of X-ray Images. *Can. Assoc. Radiol. J.* **2021**, *72*, 60–72. https://doi.org/10.1177/0846537120941671.
68. Bonheur, S.; Štern, D.; Payer, C.; Pienn, M.; Olschewski, H.; Urschler, M. Matwo-CapsNet: A Multi-label Semantic Segmentation Capsules Network. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11768, pp. 664–672. https://doi.org/10.1007/978-3-030-32254-0_74.
69. Yu, X.; Wang, S.H.; Zhang, Y.D. CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. *Information Processing & Management* **2021**, *58*, 102411. https://doi.org/10.1016/j.ipm.2020.102411.
70. Yang, L.; Wang, Z.; Zhou, L. MedXChat: Bridging CXR Modalities with a Unified Multimodal Large Model, 2023, [arXiv:cs.CV/2312.02233].
71. Kholiavchenko, M.; Sirazitdinov, I.; Kubrak, K.; Badrutdinova, R.; Kuleev, R.; Yuan, Y.; Vrtovec, T.; Ibragimov, B. Contour-aware multi-label chest X-ray organ segmentation. *International Journal of Computer Assisted Radiology and Surgery* **2020**, *15*, 425–436. https://doi.org/10.1007/s11548-019-02115-9.
72. Pandit, M.; Banday, S.; Naaz, R.; Chishti, M. Automatic detection of COVID-19 from chest radiographs using deep learning. *Radiography* **2020**, p. S1078817420302285. https://doi.org/10.1016/j.radi.2020.10.018.
73. Yao, L.; Prosky, J.; Poblenz, E.; Covington, B.; Lyman, K. Weakly Supervised Medical Diagnosis and Localization from Multiple Resolutions **2018**. [1803.07703].
74. Nour, M.; Cömert, Z.; Polat, K. A Novel Medical Diagnosis model for COVID-19 infection detection based on Deep Features and Bayesian Optimization. *Applied Soft Computing* **2020**, *97*, 106580. https://doi.org/10.1016/j.asoc.2020.106580.
75. Olatunji, T.; Yao, L.; Covington, B.; Upton, A. Caveats in Generating Medical Imaging Labels from Radiology Reports with Natural Language Processing. In Proceedings of the International Conference on Medical Imaging with Deep Learning – Extended Abstract Track, 08–10 Jul 2019, pp. 1–4.
76. Lu, M.T.; Raghu, V.K.; Mayrhofer, T.; Aerts, H.J.; Hoffmann, U. Deep Learning Using Chest Radiographs to Identify High-Risk Smokers for Lung Cancer Screening Computed Tomography: Development and Validation of a Prediction Model. *Annals of Internal Medicine* **2020**, *173*, 704–713. https://doi.org/10.7326/M20-1868.
77. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Global Optimization* **1998**, *13*, 455–492.
78. Shiraishi, J.; Katsuragawa, S.; Ikezoe, J.; Matsumoto, T.; Kobayashi, T.; Komatsu, K.i.; Matsui, M.; Fujita, H.; Kodera, Y.; Doi, K. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology* **2000**, *174*, 71–74. https://doi.org/10.2214/ajr.174.1.1740071.
79. Yue, Z.; Ma, L.; Zhang, R. Comparison and Validation of Deep Learning Models for the Diagnosis of Pneumonia. *Computational Intelligence and Neuroscience* **2020**, *2020*, 1–8. https://doi.org/10.1155/2020/8876798.
80. Wong, K.C.L.; Moradi, M.; Wu, J.; Pillai, A.; Sharma, A.; Gur, Y.; Ahmad, H.; Chowdary, M.S.; Chiranjeevi, J.; Reddy Polaka, K.K.; et al. A Robust Network Architecture to Detect Normal Chest X-Ray Radiographs. In Proceedings of the 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, 2020, pp. 1851–1855. https://doi.org/10.1109/ISBI45749.2020.9098671.
81. Madaan, V.; Roy, A.; Gupta, C.; Agrawal, P.; Sharma, A.; Bologa, C.; Prodan, R. XCOVNet: Chest X-ray Image Classification for COVID-19 Early Detection Using Convolutional Neural Networks. *New Generation Computing* **2021**. https://doi.org/10.1007/s00354-021-00121-7.
82. Bortsova, G.; Dubost, F.; Hogeweg, L.; Katramados, I.; de Bruijne, M. Semi-supervised Medical Image Segmentation via Learning Consistency Under Transformations. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 810–818. https://doi.org/10.1007/978-3-030-32226-7_90.

83. Heo, S.J.; Kim, Y.; Yun, S.; Lim, S.S.; Kim, J.; Nam, C.M.; Park, E.C.; Jung, I.; Yoon, J.H. Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data. *International Journal of Environmental Research and Public Health* **2019**, *16*, 250. https://doi.org/10.3390/ijerph16020250.
84. Jiang, W.; Ma, L.; Jiang, Y.G.; Liu, W.; Zhang, T. Recurrent Fusion Network for Image Captioning. In Proceedings of the ECCV 2018, Berlin, Heidelberg, 2018; p. 510–526.
85. Arsalan, M.; Owais, M.; Mahmood, T.; Choi, J.; Park, K.R. Artificial Intelligence-Based Diagnosis of Cardiac and Related Diseases. *Journal of Clinical Medicine* **2020**, *9*, 871. https://doi.org/10.3390/jcm9030871.
86. Chokshi, F.H.; Flanders, A.E.; Prevedello, L.M.; Langlotz, C.P. Fostering a Healthy AI Ecosystem for Radiology: Conclusions of the 2018 RSNA Summit on AI in Radiology. *Radiology: Artificial Intelligence* **2019**, *1*, 190021. https://doi.org/10.1148/ryai.2019190021.
87. Márquez-Neila, P.; Sznitman, R. Image Data Validation for Medical Systems. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11767, pp. 329–337. https://doi.org/10.1007/978-3-030-32251-9_36.
88. Anavi, Y.; Kogan, I.; Gelbart, E.; Geva, O.; Greenspan, H. Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis. SPIE, jul 2016, Vol. 9785, p. 978510. https://doi.org/10.1117/12.2217587.
89. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. *CoRR* **2023**, *abs/2304.08485*, [2304.08485].
90. Candemir, S.; Jaeger, S.; Lin, W.; Xue, Z.; Antani, S.K.; Thoma, G.R. Automatic heart localization and radiographic index computation in chest x-rays. In Proceedings of the Medical Imaging 2016: Computer-Aided Diagnosis. SPIE, 2016, Vol. 9785, p. 978517. https://doi.org/10.1117/12.2217209.
91. 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.
92. Daniels, Z.A.; Metaxas, D.N. Exploiting Visual and Report-Based Information for Chest X-RAY Analysis by Jointly Learning Visual Classifiers and Topic Models. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 1270–1274. https://doi.org/10.1109/ISBI.2019.8759548.
93. Bergstra, J.S.; Bardenet, R.; Bengio, Y.; Kégl, B. Algorithms for Hyper-Parameter Optimization. In Proceedings of the Advances in Neural Information Processing Systems, 2011, pp. 2546–2554.
94. 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.
95. Polat, C.; Karaman, O.; Karaman, C.; Korkmaz, G.; Balcı, M.C.; Kelek, S.E. COVID-19 diagnosis from chest X-ray images using transfer learning: Enhanced performance by debiasing dataloader. *Journal of X-Ray Science and Technology* **2021**, *29*, 19–36. https://doi.org/10.3233/XST-200757.
96. Candemir, S.; Rajaraman, S.; Thoma, G.; Antani, S. Deep Learning for Grading Cardiomegaly Severity in Chest X-Rays: An Investigation. In Proceedings of the Life Sciences Conference, 2018, pp. 109–113. https://doi.org/10.1109/LSC.2018.8572113.
97. 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.
98. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein Generative Adversarial Networks. In Proceedings of the Proceedings of the 34th International Conference on Machine Learning. PMLR, 06–11 Aug 2017, Vol. 70, *Proceedings of Machine Learning Research*, pp. 214–223.
99. Kruger, R.P.; Townes, J.R.; Hall, D.L.; Dwyer, S.J.; Lodwick, G.S. Automated Radiographic Diagnosis via Feature Extraction and Classification of Cardiac Size and Shape Descriptors. *IEEE Transactions on Biomedical Engineering* **1972**, *BME-19*, 174–186. https://doi.org/10.1109/tbme.1972.324115.
100. Gu, X.; Pan, L.; Liang, H.; Yang, R. Classification of Bacterial and Viral Childhood Pneumonia Using Deep Learning in Chest Radiography. In Proceedings of the Proceedings of the 3rd International Conference on Multimedia and Image Processing - ICMIP 2018. ACM Press, 2018. https://doi.org/10.1145/3195588.3195597.
101. Tien, H.J.; Yang, H.C.; Shueng, P.W.; Chen, J.C. Cone-beam CT image quality improvement using Cycle-Deblur consistent adversarial networks (Cycle-Deblur GAN) for chest CT imaging in breast cancer patients. *Sci. Rep.* **2021**, *11*, 1133. https://doi.org/10.1038/s41598-020-80803-2.

102. Bigolin Lanfredi, R.; Schroeder, J.D.; Vachet, C.; Tasdizen, T. Adversarial Regression Training for Visualizing the Progression of Chronic Obstructive Pulmonary Disease with Chest X-Rays. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 685–693. https://doi.org/10.1007/978-3-030-32226-7_76.
103. Irvin, J.; Rajpurkar, P.; Ko, M.; Yu, Y.; Ciurea-Ilcus, S.; Chute, C.; Marklund, H.; Haghgoo, B.; Ball, R.L.; Shpanskaya, K.S.; et al. CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In Proceedings of the AAAI 2019, 2019, pp. 590–597.
104. Portela, R.D.S.; Pereira, J.R.G.; Costa, M.G.F.; Filho, C.F.F.C. Lung Region Segmentation in Chest X-Ray Images using Deep Convolutional Neural Networks. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 1246–1249. https://doi.org/10.1109/EMBC44109.2020.9175478.
105. Li, J.; Li, D.; Savarese, S.; Hoi, S.C.H. BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models. In Proceedings of the ICML, 2023, Vol. 202, pp. 19730–19742.
106. Yang, Y.; Lure, F.Y.; Miao, H.; Zhang, Z.; Jaeger, S.; Liu, J.; Guo, L. Using artificial intelligence to assist radiologists in distinguishing COVID-19 from other pulmonary infections. *Journal of X-Ray Science and Technology* **2021**, *29*, 1–17. https://doi.org/10.3233/XST-200735.
107. Liang, C.H.; Liu, Y.C.; Wu, M.T.; Garcia-Castro, F.; Alberich-Bayarri, A.; Wu, F.Z. Identifying pulmonary nodules or masses on chest radiography using deep learning: external validation and strategies to improve clinical practice. *Clinical Radiology* **2020**, *75*, 38–45. https://doi.org/10.1016/j.crad.2019.08.005.
108. Toriwaki, J.I.; Suenaga, Y.; Negoro, T.; Fukumura, T. Pattern recognition of chest X-ray images. *Computer Graphics and Image Processing* **1973**, *2*, 252–271. https://doi.org/10.1016/0146-664x(73)90005-1.
109. Gozes, O.; Greenspan, H. Lung Structures Enhancement in Chest Radiographs via CT Based FCNN Training. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*; Springer, 2018; pp. 147–158. https://doi.org/10.1007/978-3-030-00946-5_16.
110. Saednia, K.; Jalalifar, A.; Ebrahimi, S.; Sadeghi-Naini, A. An Attention-Guided Deep Neural Network for Annotating Abnormalities in Chest X-ray Images: Visualization of Network Decision Basis *. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, 2020, pp. 1258–1261. https://doi.org/10.1109/EMBC44109.2020.9175378.
111. Singh, R.; Kalra, M.K.; Nitiwarangkul, C.; Patti, J.A.; Homayounieh, F.; Padole, A.; Rao, P.; Putha, P.; Muse, V.V.; Sharma, A.; et al. Deep learning in chest radiography: Detection of findings and presence of change. *PloS One* **2018**, *13*, e0204155. https://doi.org/10.1371/journal.pone.0204155.
112. 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.
113. Chen, J.; Guo, H.; Yi, K.; Li, B.; Elhoseiny, M. VisualGPT: Data-Efficient Adaptation of Pretrained Language Models for Image Captioning. In Proceedings of the CVPR, 2022, pp. 18009–18019.
114. Chlebus, G.; Schenk, A.; Hendrik Moltz, J.; van Ginneken, B.; Meine, H.; Karl Hahn, H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Scientific Reports* **2018**, *8*, 1–7. https://doi.org/10.1038/s41598-018-33860-7.
115. Ghesu, F.C.; Georgescu, B.; Gibson, E.; Guendel, S.; Kalra, M.K.; Singh, R.; Digumarthy, S.R.; Grbic, S.; Comaniciu, D. Quantifying and Leveraging Classification Uncertainty for Chest Radiograph Assessment. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 676–684. https://doi.org/10.1007/978-3-030-32226-7_75.
116. 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.
117. Pham, V.T.; Nguyen, T.P. Identification and localization covid-19 abnormalities on chest radiographs. In Proceedings of the The International Conference on Artificial Intelligence and Computer Vision. Springer, 2023, pp. 251–261.
118. van Leeuwen, K.G.; Schalekamp, S.; Rutten, M.J.; van Ginneken, B.; de Rooij, M. Artificial intelligence in Radiology; 100 commercially available products and their scientific evidence, 2021. *European Radiology (in press)* **2021**.
119. 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.

 doi:10.20944/preprints202503.2218.v1

120. Anavi, Y.; Kogan, I.; Gelbart, E.; Geva, O.; Greenspan, H. A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. *International Conference of the IEEE Engineering in Medicine and Biology Society* **2015**, *2015*, 2940–2943. https://doi.org/10.1109/EMBC.2015.7319008.
121. Ouyang, X.; Xue, Z.; Zhan, Y.; Zhou, X.S.; Wang, Q.; Zhou, Y.; Wang, Q.; Cheng, J.Z. Weakly Supervised Segmentation Framework with Uncertainty: A Study on Pneumothorax Segmentation in Chest X-ray. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*; Springer, 2019; Vol. 11769, pp. 613–621. https://doi.org/10.1007/978-3-030-32226-7_68.
122. Kallianos, K.; Mongan, J.; Antani, S.; Henry, T.; Taylor, A.; Abuya, J.; Kohli, M. How far have we come? Artificial intelligence for chest radiograph interpretation. *Clinical Radiology* **2019**, *74*, 338–345. https://doi.org/10.1016/j.crad.2018.12.015.