CONESCAPANHONDURAS2025paper9.pdf

Institute of Electrical and Electronics Engineers (IEEE)

Document Details

Submission ID

trn:oid:::14348:477787965

Submission Date

Jul 31, 2025, 11:57 PM CST

Download Date

Aug 1, 2025, 1:08 PM CST

File Name

CONESCAPANHONDURAS2025paper9.pdf

File Size

355.3 KB

6 Pages

5,026 Words

30,303 Characters

43% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **89** Not Cited or Quoted 33%
  Matches with neither in-text citation nor quotation marks
- **1** Missing Quotations 0%
  Matches that are still very similar to source material
- **15** Missing Citation 10%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

39%	Internet sources
39%	Publications
0%	Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- **89** Not Cited or Quoted 33%
  Matches with neither in-text citation nor quotation marks
- **1** Missing Quotations 0%
  Matches that are still very similar to source material
- **15** Missing Citation 10%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

39% Internet sources
39% Publications
0% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	arxiv.org	13%
2	Internet	export.arxiv.org	2%
3	Publication	Kottu, Jyothirmai. "A Vision Transformer-Driven Method for Generating Medical R...	2%
4	Internet	www.arxiv-vanity.com	1%
5	Internet	cbr.robocup.org.br	1%
6	Internet	dare.uva.nl	1%
7	Internet	openaccess.dpu.edu.tr	<1%
8	Internet	studenttheses.uu.nl	<1%
9	Internet	deepai.org	<1%
10	Internet	ouci.dntb.gov.ua	<1%

11 Internet
www.frontiersin.org <1%

12 Publication
Qian Tang, Yongbin Yu, Xiao Feng, Chenhui Peng. "Semantic and Visual Enrichme... <1%

13 Publication
Amaan Izhar, Norisma Idris, Nurul Japar. "Medical radiology report generation: A ... <1%

14 Publication
Qinghua Huang, Guanghui Li. "Knowledge graph based reasoning in medical ima... <1%

15 Publication
Xiaodan Zhang, Sisi Yang, Yanzhao Shi, Junzhong Ji, Ying Liu, Zheng Wang, Huimi... <1%

16 Internet
open.metu.edu.tr <1%

17 Internet
orca.cardiff.ac.uk <1%

18 Internet
oa.upm.es <1%

19 Internet
proceedings.mlr.press <1%

20 Internet
www.dei.unipd.it <1%

21 Internet
www.nowpublishers.com <1%

22 Publication
Ting Pang, Peigao Li, Lijie Zhao. "A survey on automatic generation of medical im... <1%

23 Publication
Yiyao Sun, Xinran Wen, Yan Zhang, Lijun Jin et al. "Visual-Language Foundation M... <1%

24 Internet
nlp.cs.aueb.gr <1%

25 Publication

Abdullah Al Siam, Md Maruf Hassan, Md Atikur Rahaman, Masuk Abdullah. "Diegi... <1%

26 Publication

Angela Ayobi, Peter D. Chang, Daniel S. Chow, Brent D. Weinberg et al. "Performa... <1%

27 Internet

books.openedition.org <1%

28 Internet

www.researchgate.net <1%

29 Publication

Shuzhe Chen, Li Li, Zhichao Lin, Ke Zhang, Ying Gong, Lu Wang, Xu Wu, Maokun Li... <1%

30 Internet

birmingham.elsevierpure.com <1%

31 Internet

assets-eu.researchsquare.com <1%

32 Internet

www.medrxiv.org <1%

33 Internet

opus.lib.uts.edu.au <1%

34 Publication

K. Sinha, Z. Uddin, H.I. Kawsar, S. Islam, M.J. Deen, M.M.R. Howlader. "Analyzing c... <1%

35 Publication

Assia Boukhamla, Nabiha Azizi, Samir Brahim Belhaouari, Nilanjan Dey. "Explorin... <1%

36 Publication

Piotr Jarosik, Marcin Lewandowski, Ziemowit Klimonda, Michal Byra. "Pixel-Wise ... <1%

37 Publication

Héctor Cañas, Josefa Mula, Francisco Campuzano-Bolarín. "A General Outline of a ... <1%

38 Publication

Junsan Zhang, Xiuxuan Shen, Shaohua Wan, Sotirios K. Goudos, Jie Wu, Ming Che... <1%

39 Publication

Fang Li, Pengfei Wang, Kuan Lin, Jiangtao Wang. "Chapter 6 MeFormer: Generatin... <1%

40 Internet

www.ncbi.nlm.nih.gov <1%

41 Publication

Gomes, Nuno Teixeira. "Previsão de Séries Temporais com R^2 Negativo: Uma Aplic... <1%

42 Internet

cdn.iiit.ac.in <1%

43 Publication

Dehai Zhang, Anquan Ren, Jiashu Liang, Qing Liu, Haoxing Wang, Yu Ma. "Improvi... <1%

44 Publication

Kun Xia, Jinzhuo Wang. "Recent advances of Transformers in medical image analy... <1%

45 Publication

Xie, Huidong. "Deep Learning Strategies for PET and SPECT Image Enhancement",... <1%

46 Internet

orbilu.uni.lu <1%

47 Internet

www.ece.nus.edu.sg <1%

48 Internet

www.utupub.fi <1%

49 Publication

"Artificial Intelligence in Medicine", Springer Science and Business Media LLC, 2025 <1%

50 Publication

Faysal Petouo, Yaya Arafat, Mueeze Mushabbir, Kamrul Hasan, Hasan Mahmud. "... <1%

51 Publication

Shenshen Bu, Taiji Li, Zhiming Dai. "Enhancing Medical Report Generation in Mult... <1%

52 Publication

Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical ... <1%

53 Publication

Yuxue Liu, Junsan Zhang, Kai Liu, Lizhuang Tan. "ADCNet: Anomaly-Driven Cross-... <1%

54 Internet

frcrce.ac.in <1%

55 Internet

ieeecai.org <1%

56 Internet

link.springer.com <1%

57 Internet

mediatum.ub.tum.de <1%

58 Internet

raw.githubusercontent.com <1%

59 Internet

web.fe.up.pt <1%

60 Internet

www-emerald-com-443.webvpn.sxu.edu.cn <1%

61 Publication

Xinyao Liu, Junchang Xin, Qi Shen, Zhihong Huang, Zhiqiong Wang. "Automatic m... <1%

62 Publication

"AI for Health Equity and Fairness", Springer Science and Business Media LLC, 2024 <1%

63 Publication

Jian Zhang, Kaihao He, Zunlei Feng, Shuifa Sun, Xiaoyan Sun, Zhenming Yuan, Jun ... <1%

1 State of the Art in Automatic Generation of Medical Reports: Models, Applications and Challenges

5 1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

3rd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

22 ***Abstract*—The automatic generation of medical reports through artificial intelligence has become a fundamental tool for improving diagnostic efficiency and reducing the workload in the clinical field. This review examines the most current computational approaches used for this task, such as encoder-decoder architectures, transformer-based models, as well as multimodal and multitask models. An exhaustive search was conducted in relevant academic databases, covering articles published between 2020 and 2025. The results show notable advances in semantic 45 accuracy and narrative coherence, especially due to the use of transformers and multimodal clinical data. In addition, this study identifies the medical specialties most frequently addressed 46 in this field, the continents with the highest concentration of 41 related research, and the evaluation metrics commonly used to assess the performance of the models. However, challenges remain in terms of interpretability, generalization, and availability of representative public datasets. This review highlights the main research gaps and demonstrates the key challenges for clinical implementation in current medical settings.**

61 ***Keywords*—Automatic medical report generation, Artificial intelligence, Encoder-decoder architecture, Transformer models, Multimodal learning, Evaluation metrics.**

51 I. Introduction

1 The automatic generation of medical reports through artifi- 35 cial intelligence (AI) techniques has emerged as an innovative 21 tool in the healthcare field. This technology has the potential to streamline medical workflows, reduce the administrative 49 burden on healthcare professionals, and enhance the accuracy and clarity of clinical documentation. Most authors develop such tools using Deep Learning (DL) [1], which has become a key instrument for processing large volumes of complex data and replicating, with greater precision and efficiency, the visual interpretation process carried out by specialists.

1 Medical images play a fundamental role in this process, as they provide detailed visual information that, when integrated 53 with other clinical data sources, allows for the generation of more complete and accurate reports. Among the most 31 commonly used image types are X-rays, ultrasounds, magnetic resonance imaging (MRI), and computed tomography (CT) scans. These images not only offer a detailed view of internal body structures but also serve as the foundation upon which medical professionals base their diagnoses. Currently, visual analysis of these images is essential for interpreting patterns, identifying anomalies, and comparing them with previous studies. However, this process faces limitations such as variability in professional experience, workload, and fatigue. Moreover, the growing volume of imaging data, especially in complex studies like CT scans, increases the risk of errors or overlooking critical details [2]. In this context, AI models have been designed to assist in pattern detection and automatic report generation [3], [4], reducing the cognitive load on professionals and improving diagnostic accuracy.

It is in this context that Deep Learning and other computational techniques come into play. The most widely used approaches include Convolutional Neural Networks (CNNs) [5], Generative Adversarial Networks (GANs) [3], and Recurrent Neural Networks (RNNs) [6]. The automation of medical reports, already established in the field of radiology [7], has also begun to be implemented in other specialties such as cardiology, for the interpretation of electrocardiograms (ECG) and echocardiograms [8]; in pulmonology, for the automated analysis of spirometry [9]; and in digital pathology, for describing findings in biopsies and histopathological images [10]. Despite these advancements, the automatic generation of medical reports still faces multiple challenges. One of the main issues is the need to ensure the accuracy and consistency of the language used in the reports, avoiding errors that could compromise diagnostic quality. Furthermore, the interpretability and validation capabilities of AI models in clinical settings remain an active area of research, with concerns related to the transparency and understandability of algorithmic decisions.

This review analyzes the most recent approaches to automatic medical report generation applied to various pathologies. It presents the models and techniques used, available datasets, evaluation methods, and key challenges for clinical implementation. The objective of this research is to provide an updated and comprehensive overview of the topic, covering multiple specialties and highlighting both the advancements and current challenges.

II. Fundamental Concepts

A. Application of Deep Learning and NLP in Medicine.

In medicine, Deep Learning (DL) is fundamental for analyzing images and clinical data, identifying relevant patterns for the automatic generation of reports. After this analysis, Natural Language Processing (NLP) comes into play, allowing machines to understand and generate human text. NLP is divided into Natural Language Understanding (NLU), which extracts concepts, entities, and relationships from data, and Natural Language Generation (NLG) [11], which transforms structured information into coherent text through techniques such as word embeddings, which represent terms as high-dimensional vectors to facilitate semantic comparison.

The complete process of report generation includes the acquisition and preprocessing of medical images and electronic health records (EHR), the analysis using DL and NLP models, the integration of results into a coherent report, and expert validation, which adjusts the models to improve their accuracy and relevance. Moreover, standardization with regulations such as HL7-FHIR and DICOM [12] ensures interoperability and security in data management. Figure 1 illustrates this workflow, from clinical data to the final report.

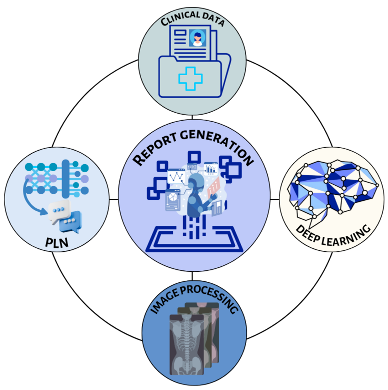

Fig. 1. Medical Report Generation.

Among the most relevant architectures for automatic report generation are encoder-decoder models, which use CNNs as encoders to extract visual features from medical images and RNNs as decoders, responsible for transforming these features into coherent textual descriptions [4], [13]. This architecture is especially useful when generating text from complex clinical images.On the other hand, transformer-based models, initially proposed by Google in 2017, have become a standard in NLP tasks thanks to the self-attention mechanism, which allows for more accurate contextual representations within a sequence. Models such as BERT and GPT-2 have demonstrated outstanding performance in medical tasks like classification, automatic summarization, and clinical question answering [14]. Additionally, some of these models integrate encoder-decoder structures, expanding their applicability in generating specialized natural language [15].

Furthermore, multimodal approaches integrate data from different sources (e.g., text, images, and signals) within the same model, offering a more holistic view of the clinical case [14]. This capability is especially valuable in medical environments where information is presented in multiple complementary formats. In addition, multitask models address several tasks simultaneously within a shared architecture, which enhances system generalization and optimizes resources by sharing parameters and representations during training [11]. This approach promotes greater efficiency in generating personalized and clinically relevant reports.

B. Computational Models

During the literature review, several relevant approaches in automatic medical report generation were identified. Fenglin Liu et al. proposed the KGAE (Knowledge Graph Auto-Encoder) model for radiographic reports, demonstrating that its supervised version (KGAE-S) significantly improves performance (BLEU-4 of 0.181 vs. 0.126) [13] by incorporating structured knowledge, with potential applications in other areas such as magnetic resonance imaging and dermoscopy.

Mingjie Li et al. presented ASGK [16], a model that fuses auxiliary visual and linguistic signals to mimic medical interpretation. It achieved high scores (BLEU-4 of 0.524, ROUGE-L of 0.674, CIDEr of 0.3124, and a Hit@% of 20%), highlighting its coherence and clinical applicability.

Finally, An Yan et al. (2021) [17] proposed a weakly supervised contrastive learning (WCL) approach to enhance the accuracy of radiological reports, overcoming the limitations of cross-entropy training. They used a pretrained BERT to identify erroneous yet semantically similar reports, thereby reinforcing the generation of clinically relevant content.

In the generation of medical reports, transformer-based models have been applied due to their high accuracy and adaptability. The CGT (Clinical Graph Transformer) model, proposed by Mingjie Li et al. (2022) [18], was integrated with CNNs to generate ophthalmological reports, incorporating clinical triplets as prior knowledge into the visual features. Despite the challenge of knowledge noise, it achieved promising results: a BLEU-4 of 0.243, ROUGE-L of 0.345, and a Hit@% of 44.7% in validation with ophthalmologists.

Similarly, Q. Gu et al. (2020) proposed PT-GEN [19], which utilized Transformers and GPT-2 to generate clinical sections from medical tables by applying table linearization. With fine-tuned GPT-2, they reached high metrics: a BLEU-4 of 0.695 and ROUGE of 0.861.

On the other hand, Di You et al. (2022) [20] proposed AlignTransformer to handle long sequences and biases. Their architecture combines the AHA and MGT modules to process medical images and generate extended reports. It achieved a BLEU-4 of 0.173, ROUGE-L of 0.379, and a Hit@% of 63%, demonstrating its effectiveness in clinical detection.

Representative examples illustrating the efficacy of multimodal and multitask approaches in the medical field include the CMCL model proposed by Liu et al. (2022) [21], which emulates the cognitive process of radiologists by jointly integrating radiographic images and clinical descriptions in natural language. This model not only enhances the contextual interpretation of visual findings but also reinforces clinical reasoning by merging complementary modalities of information.

Additionally, the Self-Boosting Framework model by Wang Z. et al. (2021) [22] falls within the multitask model category. This approach synergistically transforms the automatic generation of medical reports by incorporating a secondary image-

text correlation task, which feeds back into the training process and optimizes diagnostic accuracy. Both studies reflect how the combination and strategy of multiple modalities or tasks can significantly enhance the performance of AI-assisted clinical
22 diagnostic support systems.

III. Research Methodology

The review was conducted in February 2025 using the Scopus and IEEE Xplore databases, employing keywords related to the automatic generation of medical reports. Initially, 301 documents were retrieved. Forty were removed due to duplication, and 151 were excluded after analyzing titles, abstracts, and keywords due to lack of relevance, leaving 110
40 articles for detailed review. Subsequently, 88 articles were excluded for not meeting the research objectives, resulting in a final total of 22 documents for in-depth analysis.

This process followed a predefined protocol (see Table I), which includes the search strategy:

TABLE I
Summary of the Document Selection Process

DataBase	Scopus and IEEE Xplore
37 Keywords	automatic report generation, medical report generation, Automatic medical reports
Document type	Articles and conference papers
Date	2020 - 2024
Language	English
Review questions	• RQ1: What are the current AI-based tools and techniques used for the automatic generation of medical reports? (56) • RQ2: What are the main challenges and limitations in the clinical implementation of these tools in today's medical environment?
Search query	"automatic report generation" OR "medical report generation" OR "Automatic medical reports"
Initial search	301
Provisional collection	110
Selected articles	22

IV. Results and discussion

A. Datasets

43 Based on the conducted literature review, several commonly used datasets were identified for training and evaluating models for automatic medical report generation. Among them, various public datasets stand out, with radiology being the most represented specialty. Notable examples include MIMIC-CXR [35], IU-X-Ray [36], CheX-ray14 [2], and COV-CTR [4], all of which contain chest radiographic images paired with corresponding clinical reports.

In addition, datasets from other areas of medicine were identified. DeepEyeNet (DEN) [37] and FFA-IR [38] focus on ophthalmic images, while Gastrointestinal Endoscope (GE) [39] is centered on gastrointestinal tract endoscopy. Other datasets such as the Genotype-Tissue Expression (GTEx) [40] focus on genomic data, and biomedical signal datasets like MIME [41] and PTB-XL [42] include electromyography and electrocardiogram recordings, respectively.

This diversity of datasets reflects the wide range of potential applications for artificial intelligence in the clinical domain, enabling the development of automatic report generation tools across various medical specialties (see Table II).

B. Applications in medical specialties

The automatic generation of medical reports has emerged as a promising tool across multiple clinical specialties [8], largely facilitated by access to public databases containing medical images, physiological signals, and their corresponding annotated reports. This review has shown that one of the most extensively explored fields is radiology, accounting for 68.18% of the studies analyzed [21]. This predominance is attributed to the availability of multiple databases with large volumes of medical images, which has enabled various applications such as the automatic classification of thoracic pathologies, including pneumonia, pulmonary edema, among others.

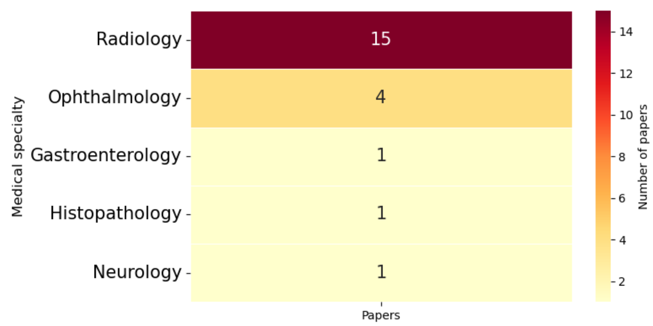

Fig. 2. Studies by Medical Specialty.

The second most recurrent specialty is ophthalmology, accounting for 18.18% of the studies [29]. In this field, medical reports are generated to describe visual anomalies, probable diagnoses, and follow-up recommendations, thereby assisting the work of specialists. In contrast, specialties such as histopathology, gastroenterology, and neurology each represent only 4.55% of the total studies [32], [31], [19]. Although these fields have clinically relevant applications for report generation, the lack of large-scale datasets hinders the implementation of deep learning-based computational models.

In absolute terms, 15 studies were identified in radiology, 4 in ophthalmology, and 1 in each of the remaining three specialties, as shown in Figure 2. These results highlight a concentration of technological development in areas with greater data availability and point to a significant opportunity to extend these solutions to other currently underserved medical specialties.

Regarding the computational models used, encoder-decoder architectures were the most prevalent, appearing in 40.9% of the papers [4], [13], followed by transformer-based models, present in 31.8% [14], [15]. Only 18.2% of the studies utilized multitask or multimodal architectures [21], [22], while 9.1% employed retrieval-based approaches [27]. This reflects a tendency towards traditional sequential approaches and the relatively unexplored potential of more complex or hybrid models.

TABLE II
SUMMARY OF DATABASES USED BY THE AUTHORS.

Database	Ref.	Description	Type	No. Data	Links
Mimic-CXR	[21], [20], [13], [17], [15], [23], [24], [25], [26], [27]	Hospital chest radiographs with clinical reports in text	X-ray	370,000	https://physionet.org/content/mimic-cxr/2.1.0/
IU-X-RAY	[21], [28], [20], [13], [22], [15], [24]	Chest X-rays with reports from Indiana University Hospital	X-ray	7,500	https://www.kaggle.com/datasets/raddar/chest-xrays-indiana-university
COV-CTR	[16], [22]	COVID-19 images (CT, X-ray) with medical reports.	CT,X-ray	500	https://github.com/mlii0117/COV-CTR
CX-CHR	[28], [16], [14]	Database with chest X-rays and Chinese medical reports.	X-ray	33,200	Private
DeepEyeNet	[29], [30]	Fundus imaging for automatic ophthalmologic diagnosis.	Equipo oftálmico	5,000	https://github.com/Jhhuangkay/DeepOpht-Medical-Report-Generation-for-Retinal-Images-via-Deep-Models-and-Visual-Explanation
FFA-IR	[18]	Fluorescein and infrared retinal imaging for medical analysis.	X-ray	300	https://physionet.org/content/ffa-ir-medical-report/1.1.0/
MIME	[19]	Diagnostic electromyography signals, represented as visual graphs.	Biomedical signals	5,000	Private
GTEx	[31]	Gene expression data and human tissues, including histological images.	Gene expression data	17,000	http://gtexportal.org/home/aboutAdultGtex
GE	[32]	Gastrointestinal endoscopic imaging for analysis and detection of pathologies.	Endoscopia	1,000	Private

TABLE III
SELECTED ARTICLES FOR THE LITERARY REVIEW. NUMBER OF CITATIONS TAKEN FROM GOOGLE SCHOLAR (ACCESSED MAY 2025).

Reference	Computational Model	Medical Specialty	Main Objective	Año	NC
[21]	CMCL	Radiology	Improve medical reporting with progressive competency-based learning.	2022	200
[28]	CMAS	Radiology	Generation of CXR reports using structural information between sections.	2020	193
[20]	AlignTransformer	Radiology	Generate consistent medical reports by overcoming visual data bias.	2021	167
[13]	KGAE	Radiology	Generate peerless medical reports with knowledge network.	2021	133
[16]	ASKG	Radiology	Improve the generation of medical reports using auxiliary signals.	2023	125
[17]	MDT + WCL	Radiology	Improve medical report generation using weakly supervised contrastive learning.	2021	86
[22]	Self-Boosting	Radiology	Improve radiographic report generation with cooperative self-improvement tasks.	2021	78
[15]	Baseline + ITM + MLC +TW	Radiology	Generate accurate radiographic reports using transformer and supervised criteria.	2022	75
[14]	Medical-VLBERT	Radiology	Create a system that automatically generates COVID-19 reports.	2021	69
[29]	DeepOpht	Ophthalmology	Generate automatic and explainable clinical reports for retinal imaging.	2021	67
[23]	MCGN	Radiology	Create a system that automatically generates CT COVID-19 reports.	2022	66
[24]	PromptMRG	Radiology	Improve the automatic generation of medical reports through diagnostic prompts.	2024	63
[25]	SV + MV + T + I	Radiology	Automate the generation of medical reports through a differentiable model of three complementary modules.	2021	61
[18]	CGT	Ophthalmology	Generate ophthalmic reports with structured clinical knowledge.	2022	55
[33]	NSL	Radiology	Automating medical report generation using neural-symbolic learning for spinal radiology.	2021	54
[26]	x-rem	Radiology	Generate accurate radiology reports using enhanced image-text similarity.	2024	49
[27]	MedWriter	Radiology	Generate accurate medical reports using hierarchical template retrieval.	2021	42
[30]	Keyword-Attention Model	Ophthalmology	Improve interpretability of retinal reports with keywords.	2023	31
[34]	ICGA-GPT	Ophthalmology	Develop a bilingual ICGA report generation and QA system.	2024	20
[32]	KdNet	Gastroenterology	Develop a knowledge-driven transformer for consistent medical reporting.	2022	15
[31]	HIPT-BERT	Histopathology	Develop automatic histopathology reporting system combining ViT and BERT.	2024	12
[19]	PT-GEN	Neurology	Automatically generate EMG and NCV reports using neural NLG.	2020	6

Table III presents a summary of the 22 articles selected in this review. The research shows that radiology is the most studied specialty, representing 68.18% of the total articles. After radiology, ophthalmology accounts for 18.18% of the studies, while other medical specialties such as gastroenterology, histopathology, and neurology each account for only 4.55%.

This indicates a need to encourage further research in these less-explored fields. Additionally, most of the reviewed papers originate from Asia, with 63.6% of the studies, highlighting a lack of research on automatic medical report generation in other continents such as Europe and the Americas, which contribute 13.6% and 22.7% respectively.

In terms of evaluation metrics, BLEU was the most frequently reported, appearing in 68.2% of the articles [13], [16], [18], followed by ROUGE (54.5%) [19], [16], METEOR (45.5%), and CIDEr (36.4%) [16]. However, a limited accuracy of these metrics was also observed. This does not necessarily indicate poor model performance but rather reflects the inherent limitations of the task. Medical language is highly variable, and many clinically valid expressions may not exactly match reference reports, which penalizes traditional metrics that rely heavily on lexical similarity.

All these findings demonstrate that although current models face significant challenges in terms of evaluation and performance, their development continues to progress toward more accurate and clinically relevant solutions, marking an important step in the integration of artificial intelligence into medical practice.

V. Conclusions

The automatic generation of medical reports through AI techniques is becoming established as a significant support tool in clinical practice. As evidenced in this review, the use of architectures such as encoder–decoder models, transformers, and multimodal and multitask approaches has enabled progress in the automatic interpretation of medical images and the creation of structured clinical reports, thus addressing RQ1. However, as outlined in RQ2, significant challenges persist, including the high variability of medical language, the scarcity of data in less-explored specialties such as neurology and gastroenterology, and limited interoperability. Despite these limitations, recent advances lay the groundwork for the adoption of AI systems, achieving a digital transformation that provides medicine with increasingly precise, interpretable, and human-centered technological tools.

References

[1] P. P. Shinde and S. Shah, "A review of machine learning and deep learning applications," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, August 2018, pp. 1–6.

[2] X. Wang, Y. Peng, L. Lu *et al.*, "Chestx-ray14: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," *IEEE Transactions on Medical Imaging*, vol. 36, no. 8, pp. 1945–1957, 2023.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, 06 2014.

[4] M. Li, R. Liu, F. Wang, X. Chang, and X. Liang, "Auxiliary signal-guided knowledge encoder-decoder for medical report generation," *World Wide Web*, pp. 1–18, 2022.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[6] C. Yin, B. Qian, J. Wei, X. Li, X. Zhang, Y. Li, and Q. Zheng, "Automatic generation of medical imaging diagnostic report with hierarchical recurrent neural network," in *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE, November 2019, pp. 728–737.

[7] A. Vallée, R. Quint, A. L. Brun, F. Mellot, and P. A. Grenier, "A deep learning-based algorithm improves radiology residents' diagnoses of acute pulmonary embolism on ct pulmonary angiograms," *European Journal of Radiology*, vol. 171, p. 111324, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0720048X24000408

[8] Z. Attia, S. Kapa, F. Lopez-Jimenez, P. McKie, D. Ladewig, G. Satam, P. Pellikka, M. Enriquez-Sarano, P. Noseworthy, T. Munger, S. Asirvatham, C. Scott, R. Carter, and P. Friedman, "Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram," *Nature Medicine*, vol. 25, no. 1, pp. 70–74, Jan 2019. [Online]. Available: https://doi.org/10.1038/s41591-018-0240-2

[9] M. Topalovic, S. Laval, J.-M. Aerts, T. Troosters, M. Decramer, W. Janssens, and on behalf of the Belgian Pulmonary Function Study investigators, "Automated interpretation of pulmonary function tests in adults with respiratory complaints," *Respiration*, vol. 93, no. 3, pp. 170–178, 2017. [Online]. Available: https://doi.org/10.1159/000454956

[10] G. Campanella, M. Hanna, L. Geneslaw *et al.*, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, vol. 25, pp. 1301–1309, 2019. [Online]. Available: https://doi.org/10.1038/s41591-019-0508-1

[11] A. Gatt and E. Krahmer, "Survey of the state of the art in natural language generation: Core tasks, applications and evaluation," in *Journal of Artificial Intelligence Research*. AI Access Foundation, 2018, vol. 61, pp. 65–170.

[12] S. Shivshankar, N. Makhija, and P. Mathusudhanan, "Digital imaging and communication in medicine (dicom): Biomedical and health informatics: Imaging and interoperability using hl7 and dicom," in *Smart Healthcare and Machine Learning*. Singapore: Springer Nature Singapore, 2024, pp. 299–317.

[13] F. Liu, C. You, X. Wu, S. Ge, S. Wang, and X. Sun, "Auto-encoding knowledge graph for unsupervised medical report generation," in *Proceedings of the 35th International Conference on Neural Information Processing Systems*, ser. NIPS '21. Red Hook, NY, USA: Curran Associates Inc., 2021.

[14] G. Liu, Y. Liao, F. Wang, B. Zhang, L. Zhang, X. Liang, X. Wan, S. Li, Z. Li, S. Zhang, and S. Cui, "Medical-vlbert: Medical visual language bert for covid-19 ct report generation with alternate learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 3786–3797, 2021.

[15] Z. Wang, H. Han, L. Wang, X. Li, and L. Zhou, "Automated radiographic report generation purely on transformer: A multicriteria supervised approach," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2803–2813, 2022.

[16] M. Li, R. Liu, F. Wang *et al.*, "Auxiliary signal-guided knowledge encoder-decoder for medical report generation," *World Wide Web*, vol. 26, pp. 253–270, 2023. [Online]. Available: https://doi.org/10.1007/s11280-022-01013-6

[17] A. Yan, Z. He, X. Lu, J. Du, E. Chang, A. Gentili, J. McAuley, and C.-N. Hsu, "Weakly supervised contrastive learning for chest X-ray report generation," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 4009–4015. [Online]. Available: https://aclanthology.org/2021.findings-emnlp.336/

[18] M. Li, W. Cai, K. Verspoor, S. Pan, X. Liang, and X. Chang, "Cross-modal clinical graph transformer for ophthalmic report generation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 20 624–20 633.

[19] Q. Gu, C. Nie, R. Zou, W. Chen, C. Zheng, D. Zhu, X. Mao, Z. Wei, and D. Tian, "Automatic generation of electromyogram diagnosis report," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 1645–1650.

[20] D. You, F. Liu, S. Ge, X. Xie, J. Zhang, and X. Wu, "Aligntransformer: Hierarchical alignment of visual regions and disease tags for medical report generation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III*. Springer International Publishing, 2021, pp. 72–82.

[21] F. Liu, S. Ge, Y. Zou, and X. Wu, "Competence-based multimodal curriculum learning for medical report generation," *arXiv preprint arXiv:2206.14579*, 2022.

[22] Z. Wang, L. Zhou, L. Wang, and X. Li, "A self-boosting framework for automated radiographic report generation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2433–2442.

[23] Z. Wang, M. Tang, L. Wang, X. Li, and L. Zhou, "A medical semantic-assisted transformer for radiographic report generation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, ser. Lecture Notes in Computer Science, L. Wang, Q. Dou, P. Fletcher, S. Speidel, and S. Li, Eds. Springer, Cham, 2022, vol. 13433, pp. 757–766. [Online]. Available: https://doi.org/10.1007/978-3-031-16437-8_63

[24] H. Jin, H. Che, Y. Lin, and H. Chen, "Promptmrg: diagnosis-driven prompts for medical report generation," in *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. [Online]. Available: https://doi.org/10.1609/aaai.v38i3.28038

[25] H. Nguyen, D. Nie, T. Badamdorj, Y. Liu, Y. Zhu, J. Truong, and L. Cheng, "Automated generation of accurate & fluent medical X-ray reports," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 3552–3569. [Online]. Available: https://aclanthology.org/2021.emnlp-main.288/

[26] J. Jeong, K. Tian, A. Li, S. Hartung, S. Adithan, F. Behzadi, and P. Rajpurkar, "Multimodal image-text matching improves retrieval-based chest x-ray report generation," in *Medical Imaging with Deep Learning*. PMLR, 2024, pp. 978–990.

[27] X. Yang, M. Ye, Q. You, and F. Ma, "Writing by memorizing: Hierarchical retrieval-based medical report generation," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, C. Zong, F. Xia, W. Li, and R. Navigli, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 5000–5009. [Online]. Available: https://aclanthology.org/2021.acl-long.387/

[28] B. Jing, Z. Wang, and E. Xing, "Show, describe and conclude: On exploiting the structure information of chest X-ray reports," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 6570–6580. [Online]. Available: https://aclanthology.org/P19-1657/

[29] J.-H. Huang, C.-H. Huck Yang, F. Liu, M. Tian, Y.-C. Liu, T.-W. Wu, I.-H. Lin, K. Wang, H. Morikawa, H. Chang, J. Tegner, and M. Worring, "Deepopht: Medical report generation for retinal images via deep models and visual explanation," in *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 2441–2451.

[30] T.-W. Wu, J.-H. Huang, J. Lin, and M. Worring, "Expert-defined keywords improve interpretability of retinal image captioning," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023, pp. 1859–1868.

[31] S. Sengupta and D. E. Brown, "Automatic report generation for histopathology images using pre-trained vision transformers and bert," in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 2024, pp. 1–5.

[32] Y. Cao, L. Cui, F. Yu, L. Zhang, Z. Li, N. Liu, and Y. Xu, "Kdtnet: Medical image report generation via knowledge-driven transformer," in *Database Systems for Advanced Applications*, A. Bhattacharya, J. Lee Mong Li, D. Agrawal, P. K. Reddy, M. Mohania, A. Mondal, V. Goyal, and R. Uday Kiran, Eds. Cham: Springer International Publishing, 2022, pp. 117–132.

[33] Z. Han, B. Wei, X. Xi, B. Chen, Y. Yin, and S. Li, "Unifying neural learning and symbolic reasoning for spinal medical report generation," *Medical Image Analysis*, vol. 67, p. 101872, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136184152030236X

[34] X. Chen, W. Zhang, Z. Zhao, P. Xu, Y. Zheng, D. Shi, and M. He, "Icga-gpt: report generation and question answering for indocyanine green angiography images," *The British Journal of Ophthalmology*, vol. 108, no. 10, pp. 1450–1456, 2024.

[35] A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, and S. Horng, "Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports," *Scientific Data*, vol. 6, p. 317, 2019. [Online]. Available: https://doi.org/10.1038/s41597-019-0322-0

[36] D. Demner-Fushman, M. D. Kohli, M. B. Rosenman, S. E. Shooshan, L. Rodriguez, S. Antani, G. R. Thoma, and C. J. McDonald, "Preparing a collection of radiology examinations for distribution and retrieval," *Journal of the American Medical Informatics Association*, vol. 23, no. 2, pp. 304–310, 2016. [Online]. Available: https://doi.org/10.1093/jamia/ocv080

[37] J.-H. Huang, C.-H. H. Yang, F. Liu, M. Tian, Y.-C. Liu, T.-W. Wu, I.-H. Lin, K. Wang, H. Morikawa, H. Chang, J. Tegner, and M. Worring, "Deepopht: medical report generation for retinal images via deep models and visual explanation," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2021, pp. 2442–2452.

[38] M. Li, W. Cai, R. Liu, Y. Weng, T. Liu, C. Wang, X. Chen, Z. Liu, C. Pan, M. Li, Y. Zheng, Y. Liu, F. Salim, K. Verspoor, X. Liang, and X. Chang, "FFA-IR: Towards an explainable and reliable medical report generation benchmark (version 1.1.0)," 2025. [Online]. Available: https://physionet.org/content/ffa-ir-medical-report/1.1.0/

[39] S. Zhu, J. Gao, L. Liu, M. Yin, J. Lin, C. Xu, C. Xu, and J. Zhu, "Public imaging datasets of gastrointestinal endoscopy for artificial intelligence: a review," *Journal of Digital Imaging*, vol. 36, no. 6, pp. 2578–2601, 2023. [Online]. Available: https://doi.org/10.1007/s10278-023-00844-7

[40] J. Lonsdale *et al.*, "The genotype-tissue expression (gtex) project," *Nature Genetics*, vol. 45, no. 6, pp. 580–585, 2013. [Online]. Available: https://doi.org/10.1038/ng.2653

[41] E. Choi, C. Xiao, W. Stewart, and J. Sun, "Mime: Multilevel medical embedding of electronic health records for predictive healthcare," in *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31. Curran Associates, Inc., 2018. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2018/file/934b535800b1cba8f96a5d72f72f1611-Paper.pdf

[42] P. Wagner, N. Strodthoff, R. D. Bousseljot *et al.*, "Ptb-xl, a large publicly available electrocardiography dataset," *Scientific Data*, vol. 7, p. 154, 2020. [Online]. Available: https://doi.org/10.1038/s41597-020-0495-6