

An Integrated Deep Learning Approach for Chest X-Ray Abnormality Recognition and Automated Radiological Report Synthesis

Haragopal V.P.¹, Shaik Munna², Penujuvi Sandeep³,
Thummaluru Soumith Reddy⁴

¹Assistant professor, ^{2,3,4}Student, Department of Computer Science and Engineering and Business Systems, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, India.

E-mail: ¹v.p.haragopal@gmail.com, ²shaikmunna04808@gmail.com, ³sandeepenujuvi857@gmail.com,

⁴soumithreddy192@gmail.com

Abstract

Chest X-ray (CXR) interpretation and radiology report creation depend on efficient feature extraction and report creation in medical terms. This research work proposes a multimodal AI framework for the automatic analysis of CXRs, which includes deep learning techniques to represent visual data, vision-language models, generate reports, and interpretability via Grad-CAM. For chest abnormality classification, this study uses visual deep neural network architectures as Swin Transformer, ConvNeXt, EfficientNet, and ResNet. This research utilizes ViT+BART, ViT+GPT-2, Swin+BART, and Swin+GPT-2 for the input image to generate report. The performance of visual classifiers is measured in Accuracy, Precision, Recall, and Macro F1-score. The generated reports are evaluated by BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and BERT-F1. Swin Transformer achieved an accuracy of 0.77, while ViT+BART achieved BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and BERT-F1 scores of 0.13, 0.42, 0.17, 0.33, and 0.89 respectively

Keywords: Chest Radiography, Thoracic Abnormality Classification, Vision-Language Modeling, Multimodal Report Generation, Hierarchical Visual Representation, Explainable Artificial Intelligence, Radiological Text Synthesis.

1. Introduction

Chest X-ray (CXR) imaging is one of the most commonly used diagnostic modalities for detecting chest pathologies such as pneumonia, tuberculosis, lung nodules, and cardiopulmonary diseases [1]. The advantages associated with its non-invasiveness, reduced cost, and wide availability make it a crucial tool for the clinical examination process in hospitals and diagnostic centres [2]. Nonetheless, the growing number of patients being examined using imaging techniques results in heavy workloads for the radiologists, a factor that can result in delayed and varied recording of diagnosis [3]. Manually interpreting these images involves skills and attentiveness, which increases the likelihood of identifying subtle radiographic findings [4].

Recent progress in Artificial Intelligence (AI) and Deep Learning (DL) has allowed for the automated analysis of medical imaging based on the learning of visual discriminative representations corresponding to certain pathological appearances [5]. Hierarchical and contextual information can be extracted from radiographic images using the convolutional neural network architecture and other transformer-based architectures, allowing clinicians to detect abnormalities of interest [6]. In addition, the combination of visual analysis with natural language generation allows for the automated conversion of analysed radiographic information into the structured report, minimizing manual report writing and increasing reporting consistency [7]. Techniques of explainable artificial intelligence, such as the gradient-weighted class activation map (Grad-CAM) can highlight the image regions that contribute to specific model predictions [8]. Web-based implementation can increase the availability of this technology by providing the possibility of real-time image analysis, visualization, and report generation via the integrated platform [9]. These capabilities will help to expedite clinical work processes while simultaneously ensuring standardized and understandable radiology reports [10].

This research will investigate the development of an integrated AI system for performing analysis and generating reports of CXRs, which utilizes image classification using deep learning models, vision-language approaches, Grad-CAM explainability technique, and Flask web application. Different visual backbones will be analyzed for its performance in CXR image classification, whereas different vision-language combinations will be analysed for the development of automated medical report generation.

2. Related Works

Fetiler et al. [11] researched transformers for generating CXR reports, revealing their ability to learn the visual–textual relationship for structured reporting in radiology. Akbar et al. [12] developed TraHeaLRG employing LSTM and GRU architectures, concentrating on sequence learning for better CXR report generation. Yoon et al. [13] researched the use of artificial intelligence-enhanced dual-energy X-ray imaging for better image quality and improved diagnostic interpretation.

The research of Sharma et al. [14] explored the performance discrepancy between artificial intelligence models and radiologists, showing how trained AI can be used for the task of classifying and reporting CXRs. Cxpmrg-bench was presented by Wang et al. [15], highlighting the importance of large-scale pre-training and benchmarking in X-ray medical report generation. Ali et al. [16] presented a semantic evaluation framework to evaluate the consistency of reports produced by LLMs. Zou et al. [17] presented CorBenchX, which is focused on error detection and correction in automated CXR reports.

X-Ray-CoT by Ng et al. was presented, stressing interpretable vision-language reasoning in CXR interpretation [18]. Yadav et al. [19] considered the use of AI-based tactile and 3D-printable X-ray visualization techniques in order to enhance access to medical image data. Yang et al. [20] offered spatio-temporal and retrieval-augmented modeling for incorporating contextual data into automated CXR reporting.

Overall, previous researches have shown great progress in representation learning, sequence modeling, generation of vision-language, semantic evaluation, error correction, explanation of reasoning, and retrieval augmentation using transformers. These studies provide the methodological foundation for integrated CXR systems combining image classification, automated report generation, and explainability.

3. Proposed Methodology

The proposed framework integrates deep learning models for CXR abnormality classification and automated radiological report generation. This system will employ multiple deep learning models such as SWIN Transformer, ConvNeXt, EfficientNet, and ResNet to categorize the images using the right model. Radiography images will be utilized to extract

relevant visual information from these models [21]. Models like Vision Transformer with BART, Vision Transformer with GPT2, Swin Transformer with BART, Swin Transformer with GPT2 will be used to transform visual representations into clinical explanations that make sense [22]. This framework has some preprocessing tools to help in preprocessing raw images to improve their consistency and the quality of information [23]. The grad-CAM explainability model is developed in, where the parts of the image making the biggest impact on prediction are visualized. This provides visual interpretability of the predictions and facilitates model-output verification [24]. The proposed system works as a web-based application using Flask with authentication methods for the login procedure, image uploading, automated analysis, visualization of the identified regions, and downloading the report [25]. This comprehensive approach helps radiologists receive the full assistance by automatic detection of anomalies with the visualization of comprehensible data and report generation. This speed up the process of diagnosis and improves the efficiency of the decision-making process in a medical institution.

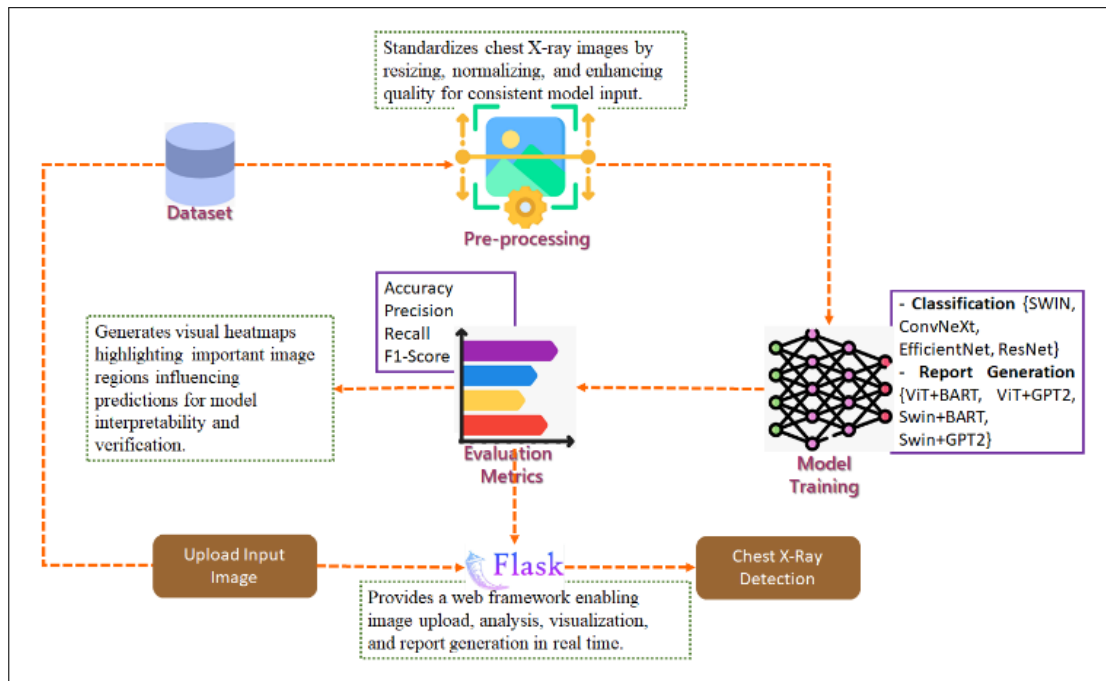


Figure 1. System architecture of the CXR analysis framework

The proposed workflow comprises the major stages involved in automated CXR analysis. It starts with collecting and pre-processing the data where the inputs are resized and normalized. As shown in Figure 1, different DL architectures, SWIN, ConvNeXt, EfficientNet, and ResNet are trained to classify the data while multimodal DL architectures, such as ViT+BART are

trained to generate the reports. Proper performance analysis is carried out with the help of metrics such as Accuracy, Precision and F1-score. Finally, Grad-CAM heatmaps are generated to highlight regions contributing to model predictions.

3.1 Dataset Collection

The dataset contains 7,470 chest x-ray images and 3,955 reports on radiology. The reports contain clinically significant details, and the dataset includes frontal and lateral projection information associated with the radiographic images. The key features of the dataset are listed in Table 1 [26]. The pairing of the images and reports was done with the help of their respective identifiers, making it possible for the dataset to perform the tasks of CXR classification and multimodal report generation [26].

Table 1. Dataset description of the IU X-ray dataset

Dataset Attribute	Description
Dataset	Indiana University Chest X-ray (IU X-ray)
Source	Open-I / Indiana University
Distribution Source	Kaggle
Chest X-ray Images	7,470
Radiological Reports	3,955
Projection Types	Frontal and lateral
Image Format	PNG
Text Information	Radiological reports
Training Set	70%
Validation Set	15%
Testing Set	15%
Primary Tasks	CXR classification and report generation
Data Organization	Paired image-report samples

Figure 2 shows the database structure of the experiment, depicting the connection between the image IDs, projection details, and their radiology reports. The processed samples were then fed into the classification model and report generator after image and text pre-processing.

	image_id	projection	org_caption
0	CXR3468_IM-1684-0001-0001.png	Frontal	heart size is at the upper limits of normal. t...
1	CXR3468_IM-1684-0001-0002.png	Lateral	heart size is at the upper limits of normal. t...
2	CXR1853_IM-0555-1001.png	Frontal	the lungs are clear bilaterally, specifically,...
3	CXR1853_IM-0555-2001.png	Lateral	the lungs are clear bilaterally, specifically,...
4	CXR1317_IM-0205-1001.png	Frontal	lungs are clear. there is minimal atelectasis ...

Figure 2. IU X-ray dataset structure

3.2 Data Pre-processing

Preprocessing is the initial stage that helps prepare the chest X-ray images and radiological reports to be classified and generated in the next stages. All images are resized to a consistent spatial resolution and normalized to maintain uniform pixel intensity distributions. Enhancement and augmentation techniques are used on the images to ensure that the model is robust enough to adapt to any patient position and image capturing process conditions.

Regarding the textual modality, the radiological reports are broken down into tokens, which can either be individual words or subword pieces, after which all the irrelevant symbols are stripped out and the text is standardized. The preprocessed radiological reports are then paired with their respective X-ray images to ensure the desired image-report pairing for multimodal learning. The standardized image and text data are then saved in an organized format.

3.3 Algorithms

3.3.1 Swin Transformer

Swin Transformer is a hierarchical vision transformer that uses shifted window self-attention to capture local and cross-window contextual information. The image is segmented into disjoint local windows on which the self-attention operation is performed. At times, it shifts the windows in order for them to communicate with each other. This strategy simplifies the coding process as compared to the global attention strategy. The swin transformer employs multi-scale feature representation on CXR and this helps it capture small thoracic abnormalities. It is structured hierarchically, and hence one can reduce the features gradually, and thus suitable for medical image classification tasks.

3.3.2 ConvNeXt

CNN architecture ConvNeXt is based on the transformers; however, convolutional layers are also used. Large kernels, inverted bottlenecks, layer normalization, and simpler stages in the network allow feature learning to become better. Representation capacity and training stability of ConvNeXt make it possible to use both classic CNNs and transformers. For the analysis of chest X-ray images, it becomes able to distinguish between different areas of lungs by finding abnormal patterns. ConvNeXt contains highly optimized convolutional layers and they help it acquire high generalization ability and efficient feature extraction from high-resolution CXR images.

3.3.3 EfficientNet

EfficientNet of CNN is made up of compound scaling in such a way that the depth, width, and resolution of the network are scaled sensibly. The network does not simply enlarge itself but rather alters these measures in a sensible way to get better performance with fewer parameters. EfficientNet uses mobile inverted bottleneck convolution blocks and squeeze-and-excitation to improve feature representation. It depends on a balanced scaling technique to provide the best means of detecting abnormalities and maximize resource utilization.

3.3.4 ResNet

The deep CNN models having skip links involving residual learning. With these shortcut connections, gradients can easily diffuse through many layers. This also helps solve the problem of vanishing gradient and makes extremely deep architectures feasible. ResNet blocks may also be capable of learning identity mapping that can efficiently map features without slowing the process. In the interpretation of CXR images, ResNet recognizes the image features that are hierarchically arranged, i.e., from simple edges to complex pathological features. It is deeply layered, making it easy to learn strong representations; therefore, problems in x-rays can be detected effectively.

3.3.5 ViT and BART

The ViT and BART architecture for vision and language is the merging of bidirectional and autoregressive transformer decoders and ViT encoders. The ViT transforms patches of an image into a global visual representation which then becomes subject to the operations of the self-attention mechanism. This representation then passes through BART to generate sensible

textual patterns based on encoder-decoder attention. CXR images become narrative clinical reports in the process of creating medical reports. In such a manner, radiological results and texts get semantically linked together through this multimodal correspondence.

3.3.6 ViT and GPT2

Vision Transformer and GPT2 are two modules that are used in the system for linking the feature extraction of images and autoregressive language models. Vision Transformer uses self-attention on picture patches to transform CXRs pictures into visual tokens representing image-level features. With the help of embeddings, GPT2 generates text in the form of a given sequence by predicting words based on tokens that were already written. With the help of this system, one can create visual patterns that can be described in a medical way. The autoregressive decoding ensures the syntactic fluency of the generated sentences and the clinical observations related to the image.

3.3.7 Swin and BART

This hybrid architecture of top-down visual encoding using Swin Transformer and sequence-to-sequence text model using BART. They create multiscale context-aware embeddings by giving Swin the window attention for CXR. Once submitted to BART, they are sent for autoregressive decoding and bidirectional encoding to enhance and decode them into structured medical descriptions. The combination of features from image and textual tokens makes the architecture highly suitable for cross-modal attention. This will allow localization of radiographic patterns to be converted into diagnostic outcomes understandable as part of the generated clinical reports.

3.3.8 Swin and GPT2

The Swin Transformer and GPT-2 architecture combines hierarchical visual feature extraction with autoregressive text generation. Swin Transformer extracts multi-scale visual representations from CXR images by modeling both local and cross-window contextual information. These visual representations are subsequently mapped to GPT-2, which generates the radiological report sequentially through autoregressive token prediction. This architecture establishes a visual-to-text in which image-derived features are transformed into clinically relevant textual descriptions. The integration of hierarchical image representations with sequential language modeling enables automated generation of radiological reports from CXR images.

4. Experimental Results and Discussion

The proposed CXR analysis framework was evaluated in two stages: CXR abnormality classification and automated radiological report generation. Performance of classification was assessed using Accuracy, Precision, Recall, and Macro F1-score, while that of report generation using BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and BERT-F1 scores.

The definitions of the evaluation metrics used for the purposes of classification are presented below. Accuracy indicates the percentage of correctly classified instances:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Precision is the fraction of the correctly predicted positive instances among all predicted positives:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall is the ratio of correctly identified positive instances to the total number of positive instances:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

F1 score refers to the harmonic mean of precision and recall:

$$F1\text{-Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \times 100 \quad (4)$$

For the purpose of multiclass classification, Macro F1-Score can be computed as the average of individual F1-scores calculated per each class.

4.1 Classification Performance

Comparative classification accuracy of SWIN, ConvNeXt, EfficientNet, and ResNet is displayed in Figure 3. SWIN was able to obtain the maximum classification accuracy of 0.77,

followed by ConvNeXt (0.73), ResNet (0.72), and EfficientNet (0.64). This indicates that Swin Transformer had the highest classification accuracy among the reviewed visual models.

However, the rest of the metrics reveal different performance levels. ConvNeXt had the best Macro F1 score (0.27) and Precision (0.30), but EfficientNet had the best Recall (0.27). ResNet also offered competitive Macro F1 score, Recall, and Precision results. In summary, even though SWIN has the best overall accuracy level, other frameworks have shown competitive levels in class-specific metrics.

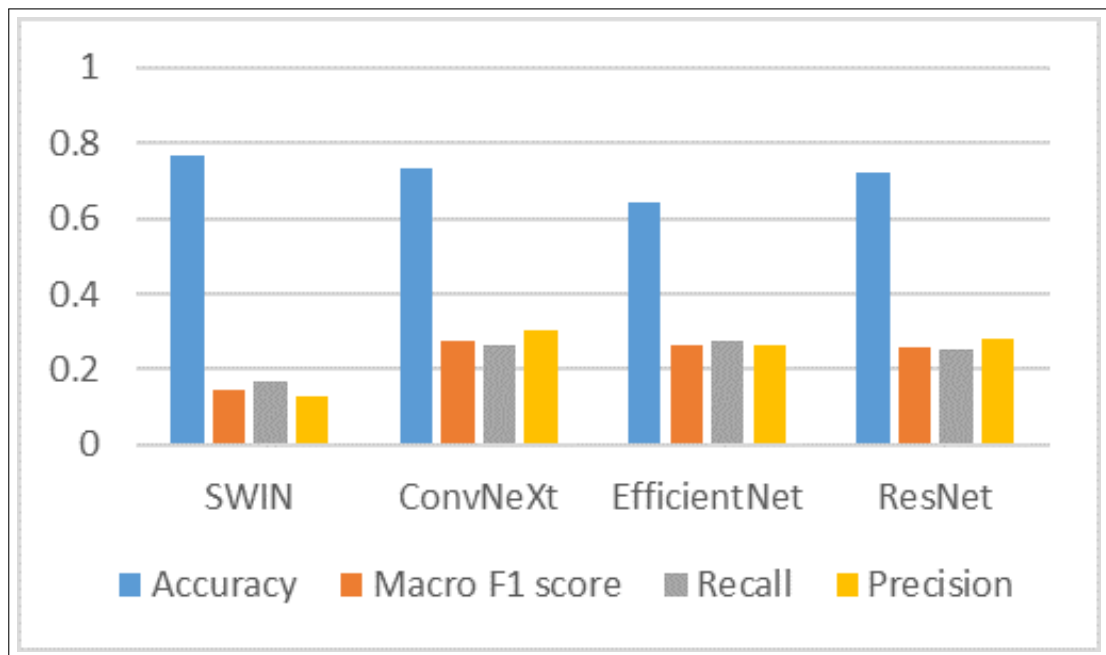


Figure 3. Classification performance comparison

Figure 3 compares the four classification architectures across Accuracy, Macro F1-score, Recall, and Precision. It shows that SWIN is more accurate than the others, while the rest of the measures show the additional performance properties of the models.

4.2 Report Generation Performance

The second stage tests the capability of the multimodal architectures to generate radiological reports based on the CXR images. The BLEU metric is used to test the n-gram matching capability while ROUGE is used to assess the lexical match. BERT-F1 is used to measure semantic match. Figure 4 shows a comparison between ViT+BART, ViT+GPT-2, Swin+BART, and Swin+GPT-2. ViT+BART achieved the highest values across all reported measures, obtaining BLEU of 0.12, ROUGE-1 of 0.42, ROUGE-2 of 0.17, ROUGE-L of 0.32,

and BERT-F1 of 0.88. Swin +BART generated relatively better performance, especially ROUGE based metrics and BERT-F1. ViT+GPT-2 and Swin+GPT-2 have relatively low performance compared to other configurations. It can be seen that configurations based on BART generate relatively better performance than the ones based on GPT-2 in the tested architectures.

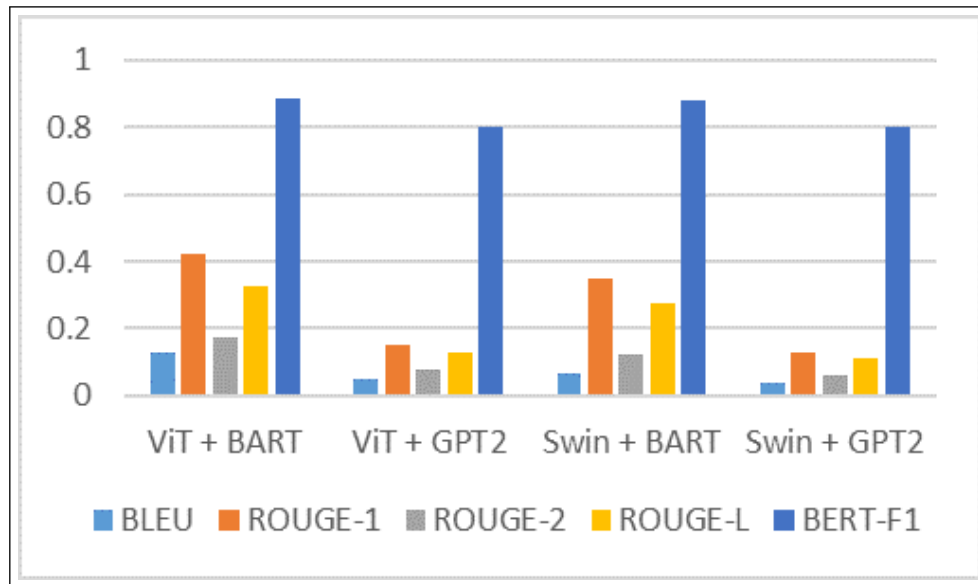


Figure 4. Report generation performance comparison

Figure 4 illustrates the outcomes of the report generation in a graphical manner, which clearly highlights the better performance of ViT+BART in the evaluated metrics. The BLEU, ROUGE, and BERT-F1 scores point to a better correspondence between the generated reports and reference reports.

4.3 Overall Discussion

Experimental results clearly show that the analyzed models have their own advantages in classification and automated report generation. As for the classification task, Swin Transformer showed the best accuracy, which proves its ability to extract hierarchical visual features from chest X-ray images. In turn, ConvNeXt was better in terms of Macro F1-score and Precision, and EfficientNet had the highest Recall value.

For report generation, the highest BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and BERT-F1 were obtained by ViT+BART, reflecting better alignment between the generated and reference radiological reports. On the other hand, while other multimodal models yielded good

results, the models based on GPT-2 gave the lowest scores. As a result, the findings suggest that Swin Transformer is the best model in terms of classification accuracy, and ViT+BART is the best model for report generation.

5. Conclusion and Future Work

The proposed AI-based framework integrates chest X-ray analysis, thoracic abnormality classification, multimodal radiological report generation, and Grad-CAM-based visual interpretability. Comparison between convolutional and transformer-based visual architectures shows their ability to produce clinically useful representations of chest radiographs. Vision-language architectures facilitate conversion of the radiologic data into clinical descriptions. The implementation of explainable AI helps in interpreting the output of the models by identifying diagnostic areas in the images. Future research includes increasing the dataset by collecting radiographic examples from multiple institutions. More advanced foundation models for vision-language processing, finer-grained localization of abnormalities and optimization using human feedback will improve report accuracy. Future extensions may include other medical technologies like CT scans and EHRs in addition to privacy-preserving federated learning.

References

- [1] Sandi, Gitarja, and Suhono Harso Supangkat. "Automated AI-Driven Report Generation for Chest X-Ray Interpretation in Teleradiology: Towards Enhanced Personalized Healthcare." In 2025 International Conference on ICT for Smart Society (ICISS), IEEE, 2025, 1-5.
- [2] Singh, Prateek, and Sudhakar Singh. "Chestx-Transcribe: A Multimodal Transformer for Automated Radiology Report Generation from Chest X-Rays." *Frontiers in Digital Health*, 2025, vol. 7: 1535168
- [3] Fahmy, Ahmed Maged, and Manar Mahmoud El Mashaad. "Automated X-Ray Chest Report Generation." In 2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS), IEEE, 2025, 651-658.
- [4] Jeong, Won Gi, Eui Jin Hwang, Gong Yong Jin, Ju-Hyung Lee, Se Ri Kang, Hongseok Ko, Bomi Gil et al. "Artificial Intelligence-Driven Drafting of Chest X-Ray Reports: 2025 Position Statement from The Korean Society of Thoracic Radiology Based on an Expert

- Survey." *Korean Journal of Radiology*, 2025, vol. 26, no. 11: 1100.
- [5] Zhang, Xiaoman, Julian Nicolas Acosta, Xiaoli Yang, Subathra Adithan, Luyang Luo, Hong-Yu Zhou, Joshua Miller et al. "Automated Chest X-Ray Report Generation Remains Unsolved." In *Biocomputing: Proceedings of the Pacific Symposium*, 2025, 236-250.
- [6] De Castro, Daniel Coelho, Aurelia Bustos, Shruthi Bannur, Stephanie L. Hyland, Kenza Bouzid, Maria Teodora Wetscherek, Maria Dolores Sánchez-Valverde et al. "Padchest-Gr: A Bilingual Chest X-Ray Dataset for Grounded Radiology Report Generation." *The New England Journal of Medicine*, 2025, vol. 2, no. 7: AIdbp2401120.
- [7] Bai, Yaowei, Ruiheng Zhang, Yu Lei, Jingfeng Yao, Shuguang Ju, Chaoyang Wang, Wei Yao et al. "From Bench to Bedside: A Deepseek-Powered AI System for Automated Chest Radiograph Interpretation in Clinical Practice." *arXiv preprint*, 2025, arXiv:2507.19493.
- [8] Aurangzeb, Brekhna, Dennis Robert, Cynthia Baard, Abid Ali Qureshi, Aneela Shaheen, Atiqa Ambreen, David McFarlane et al. "Evaluating The Accuracy of Artificial Intelligence-Powered Chest X-Ray Diagnosis for Paediatric Pulmonary Tuberculosis (EVAL-PAEDTBAID): Study Protocol for A Multi-Centre Diagnostic Accuracy Study." *BMJ Open*, 2025, vol. 15, no. 7: e105881.
- [9] Saleh, Ahmed, Ahmed Mohamed Sied, Rahma Gharib, Alaa Elsaid, Ahmed Mabrouk, and Hoda Atef Elbatrawy. "AI-Driven Intelligent Medical System: Deep Learning for Chest X-Ray/CT Disease Recognition." *Damanhour Journal of Intelligent Systems and Informatics*, 2025, vol. 1, no. 1.
- [10] Edirisinghe, Dasith, Wimukthi Nimalsiri, Mahela Hennayake, Dulani Meedeniya, and Gilbert Lim. "Chest X-Ray Report Generation using Abnormality Guided Vision Language Model." *IEEE Access*, 2025, vol. 13: 157651-157673.
- [11] Fetiler, Bengü, Ömer Atılım Koca, and Volkan Kılıç. "Report Generation From X-Ray Images: an Evaluation with Transformer Architectures." *Artificial Intelligence Theory and Applications*, 2025, vol. 5, no. 2: 1-10.
- [12] Akbar, Wajahat, Shuguang Li, Abdullah Soomro, Altaf Hussain, Razaz Waheeb Attar, Tariq Hussain, Amal Hassan Alhazmi, and Reem Alsagri. "Trahealrg: Transformative Healthcare Leveraging Lstm and Gru Models Toward Improved Accuracy for Chest X-Ray Report

- Generation." *International Journal of Pattern Recognition and Artificial Intelligence*, 2025, vol. 39, no. 15: 2540005.
- [13] Yoon, Sung-Hyun, Jihang Kim, Junghoon Kim, Jong-Hyuk Lee, Ilwoong Choi, Choul-Woo Shin, and Chang-Min Park. "Improving Image Quality of Chest Radiography with Artificial Intelligence-Supported Dual-Energy X-Ray Imaging System: An Observer Preference Study in Healthy Volunteers." *Journal of Clinical Medicine*, 2025, vol. 14, no. 6: 2091.
- [14] Sharma, Harshita, Maxwell C. Reynolds, Valentina Salvatelli, Anne-Marie G. Sykes, Kelly K. Horst, Anton Schwaighofer, Maximilian Ilse et al. "Closing The Performance Gap Between AI and Radiologists in Chest X-Ray Reporting." *arXiv preprint*, 2025, arXiv:2511.21735.
- [15] Wang, Xiao, Fuling Wang, Yuehang Li, Qingchuan Ma, Shiao Wang, Bo Jiang, and Jin Tang. "Cxpmsg-Bench: Pre-Training and Benchmarking for X-Ray Medical Report Generation on Chexpert Plus Dataset." In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2025, 5123-5133.
- [16] Ali, Haider, Rashadul Islam Sumon, Abdul Rehman Khalid, Kounen Fathima, and Hee Cheol Kim. "A Semantic Evaluation Framework for Medical Report Generation using Large Language Models." *Computers, Materials Continua*, 2025, vol. 84, no. 3: 5445-5462.
- [17] Zou, Jing, Qingqiu Li, Chenyu Lian, Lihao Liu, Xiaohan Yan, Shujun Wang, and Jing Qin. "Corbenchx: Large-Scale Chest X-Ray Error Dataset and Vision-Language Model Benchmark for Report Error Correction." *arXiv preprint*, 2025 arXiv:2505.12057.
- [18] Ng, Chee, Liliang Sun, and Shaoqing Tang. "X-Ray-Cot: Interpretable Chest X-Ray Diagnosis with Vision-Language models Via Chain-Of-Thought Reasoning." *arXiv preprint*, 2025, arXiv:2508.12455.
- [19] Yadav, Priti, and Anujkumarsinh Donvir. "Enhancing Medical Imaging Accessibility: An AI-Powered Automated Tactile and 3D-Printable X-Ray System." In *IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, 2025, 00508-00515.
- [20] Yang, Yan, Xiaoxing You, Ke Zhang, Zhenqi Fu, Xianyun Wang, Jiajun Ding, Jiamei Sun et al. "Spatio-Temporal and Retrieval-Augmented Modeling for Chest X-Ray Report Generation." *IEEE Transactions on Medical Imaging*, 2025, vol. 44, no. 7: 2892-2905.

- [21] Zhao, Jian, Wei Yao, Lei Sun, Lijuan Shi, Zhejun Kuang, Changwu Wu, and Qiulei Han. "Automated Chest X-Ray Diagnosis Report Generation with Cross-Attention Mechanism." *Applied Sciences*, 2025, vol. 15, no. 1: 343.
- [22] Thakur, Amit, and Nilamadhab Mishra. "AI-Driven Synthetic Data Generation using Conditional Gans to Mitigate Class Imbalance in COVID-19 X-Ray Imaging." In *International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)*, IEEE, 2025, 1-6.
- [23] Mamdouh, Dima, Mariam Attia, Mohamed Osama, Nesma Mohamed, Abdelrahman Lotfy, Tamer Arafa, Essam A. Rashed, and Ghada Khoriba. "Advancements in Radiology Report Generation: A Comprehensive Analysis." *Bioengineering*, 2025, vol. 12, no. 7: 693.
- [24] Yazdani, Elmira, Aryan Neizehbaz, Najme Karamzade-Ziarati, and Saeed Reza Kheradpisheh. "Explainable Artificial Intelligence for Pneumonia Classification: Clinical Insights into Deformable Prototypical Part Network in Pediatric Chest X-Ray Images." *Journal of Medical Imaging and Radiation Sciences*, 2025, vol. 56, no. 5: 102023.
- [25] Chen, Wenting, Yi Dong, Zhaojun Ding, Yucheng Shi, Yifan Zhou, Fang Zeng, Yijun Luo et al. "Radfabric: Agentic Ai System with Reasoning Capability for Radiology." *arXiv preprint*, 2025, arXiv:2506.14142.
- [26] Raddar, "ChesXrays(IndianaUniversity)," KagglAvailable: <https://www.kaggle.com/dataset/raddar/chest-xrays-indiana-university>