ChestVision: Deep Learning-Based Multi Disease Detection from Chest X-Rays

Soham Chaudhari[1], Vedant Sawadh[1], Shreya Maurya[1], Dr. Naziya Hussain[2]

[1]UG Student, Department of Computer Engineering, MPSTME, SVKM's NMIMS Deemed to be University, Shirpur, Maharashtra, India

[2]Assistant Professor, Department of Computer Engineering, MPSTME, SVKM's NMIMS Deemed to be University, Shirpur, Maharashtra, India

Abstract

The advancement of deep learning within recent years has radically transformed the analysis of the chest X-rays (CXR) by making thoracic abnormality detection, classification, and interpretation notably accurate. This paper presents a dedicated review and integrated synthesis of the recent methods applied in the diagnosis of diseases by CXR), including convolutional neural networks (CNNs), recurrent networks (e.g., LSTM), and the idea of an ensemble. We talk about the state-of-the-art systems such as systems based on hybrid CNN-LSTM networks, performance based image annotation networks, self denoising autoencoders, CGAN, and generative art models. The systems seek to address such critical issues of research problems as data imbalance, lack of interpretability, and challenges of deploying these systems into real clinical practices. We have designed a specific evaluation of a combined architecture based on DenseNet to extract spatial features and LSTM layers to refine the features sequentially and enhance multi-label disease prediction. Comparative observations suggest that ensemble models have the potential to increase the accuracy and minimize classification errors, as long as explainability tools (e.g., Grad-CAM) are incorporated to enhance clinical reliability and trust. Nevertheless, even with promising performance, some gaps are present: a lack of performance in the diversity of the population groups, the inability to cover the entire range of thoracic diseases, and the inability to smoothly connect with the information processes of the hospital. In general, the discussion indicates that CXR diagnostic systems built with the help of deep learning have a strong potential of being good decision-support tools in radiology. However, further studies are required to come up with scalable, interpretable, and clinically implementable solutions. The review can thus act as a valuable source to the researchers and practitioners in the healthcare field who need to take AI-enabled radiology to the next level of practical, real-world maturity.

Keywords— *Chest X-ray, Radiology Report Generation, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Disease Classification Support (DCS), Deep Learning, Medical Image Analysis, Natural Language Processing (NLP), Multimodal Learning, Clinical Decision Support Systems*

1.INTRODUCTION

The life of a radiologist in one such rural hospital involves reviewing up to 200 chest X-rays (CXR) per day, all the while managing the weighted stress and fatigue of diagnosing pneumonia or tuberculosis. Considering there are only 10,000– 15,000 radiologists in the country for over 1.4 billion people, this is especially true for Tier 2 and Tier 3 cities, where the lack of specialists leads to misreported reporting—delays in the reading of X- rays—and at times, misreported diagnoses. Even more government health missions and advanced diagnostics complicate these systems and with this socioeconomic disparity, AI systems are needed to automate what overburdened and stretched health systems cannot.

At the heart of such automatization are Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks—modern AI systems used for medical imaging applications. CNNs assess images via convoluted filters to assess hierarchically—edges to higher elements like nodules in the lung or borders of the heart [1]—and are especially effective with thoracic pathologies as they learn scalable representations from diverse populations with high data volumes [2]. LSTMs, a type of recurrent neural network, assess temporal data by incorporating memory cells that model long- term dependencies [3] and convert learned image features into language through deep learning through their mathematically formalized processes [4]. This is the means by which stable reports are generated over time.ChestVision combines these approaches in an AI- based system that automatically generates chest X- ray reports. First, it leverages a CNN-derived language model which determines features of the image to classify pathology and generates relevant probability vectors from diseases expected through CXRs (tuberculosis, pneumonia, nodules, etc.). This derived feature vector with a likelihood assessment is then interpreted by an LSTM-derived language model which generates a comprehensive, clinically relevant report generated by pathology classification from the feature image assessment. Written in Python with use of standard libraries for AI and optional deployment through Streamlit, ChestVision improves diagnostic access, reduces interpreter error and turnaround time in low resource settings [5][6], reduces the burden on overworked radiologists, and shows the potential of clinical augmentation through AI.

2.REVIEW METHODOLOGY

This review compiles and analyzes studies published between 2021 and 2025 on AI-based chest X-ray interpretation and automated report generation.Relevant publications from IEEE Xplore, PubMed, ScienceDirect, and SpringerLink were found by using search terms such as "chest X-ray," "radiology report generation," "CNN-LSTM," and "deep learning in healthcare."Only peer-reviewed studies employing benchmark datasets such as IU X-ray, MIMIC-CXR, and NIH ChestXray14 were considered in order to guarantee data comparability and reliability.Each study was assessed using model architecture (e.g., CNN, LSTM, transformer-based), performance metrics (e.g., BLEU, ROUGE, F1), and clinical integration potential.The new architectures, evaluation methods, and clinical uses of AI-powered radiology solutions are fairly summarized by this methodical approach.

3.RELATED WORKS

Interpretations of the CXR in the rural Indian clinics with a backlog of cases where a CXR would have benefited the treatment at that time through a detailed report with limited time and resources with many overlooking results in the initial examination (e.g. early tuberculosis). The application of AI- based applications such as ChestVision aids in the review of images and the production of reports using natural language processing, which generates an initial report to support overworked doctors. According to the latest publications, there has been a progress in automated CXR interpretation in terms of accuracy, explainability, and deployment; the three values motivated the development of ChestVision. GIT-CXR is a report generation system based on the use of a GenerativeImage2Text transformer which uses curriculum learning to generate reports; a single image and multiple view CXR system to report on a patient which also takes into consideration patient demographics. GIT-CXR is trained on MIMIC-CXR-JPG, it achieves F1- macro 0.72, F1-micro 0.75, METEOR 0.28, BLEU-4 0.35 and ROUGE-L 0.62. The transformer is better at long-range dependencies but uses stunning amounts of resources thus not being applicable in low-resource scenarios (Sîrbu et al., 2025, [7]). IHRAS combines multi-label CNN prediction of 14 thoracic diseases with Grad-CAM based abnormality visualization, SAR-Net segmentation of reports and DeepSeek-R1 structural generation of reports. In NIH ChestX-ray (112 120 images), F1 0.68 with segmentation provides a 12%

specificity growth in pneumonia and neoplasm (ChestVision- partial pathologies). The design of the component can be associated with the final objective of the ChestVision, which is clinical relevance; the IHRAS dataset cannot be generalized as it involves a biased national population (Rodrigues et al., 2025, [8]).

Convolutional Block Attention Modules (CBAM) are used in transformers with abnormal region attention and multimodal alignment to predict the abnormality and generate the report; that is, in the case of abnormality prediction, 15% higher than CNN-LSTM baselines; BLEU-4 0.38 and ROUGE-L 0.65 (Zhao et al., 2025, [9]). The generators of follow-up reports developed using trained transformers are used to follow the disease progress over IU X-ray (7,470 pairs) longitudinally, and the algorithm generates ROUGE-L 0.70 and F1 0.67 to produce disease-specific reports indicating the need to observe more than one day, but the size of the dataset is usually insufficient to cover sufficient time (Wang et al., 2025, [10]). ReXGradient-160K includes multi-label classifications of 14 pathologies in images of >273,047 CXRs of 109,487 patients across multi-institutions; therefore, it is the first U.S.-based standardized data on CXRs with multi- label classifications but again requires a second round of datasets based in India with the view to indicate interest in a non-western population (Rajpurkar et al., 2025, [11]).

The TB, pneumonia, COVID-19, and cardiomegaly are analyzed with an F1 of 94, a recall of 99, a precision of 95, and an accuracy of 89.3 by a CNN- LSTM trained on 10,000 images (94, 99, 95, 89.3), which is 5-10 times better than other ensembles. It is important due to the interpretability gap of explainable AI, including the use of Grad-CAM on ChestVision (Nair and Singh, 2025, [12]). A review paper in CNN discusses features-to-deep learning method-based CXR report transformation and its effects on clinical adoption, data scarcity, and interpretable higher-order NLP features according to the LSTM methodology of ChestVision (Ouis and Akhloufi, 2024, [13]).The range of problems that DenseNet201 can identify is smaller than that of ChestVision that uses SHAP, LIME, and Grad-Cam to identify COVID- 19, pneumonia, and TB with a 99.2-percent accuracy on 12,000 images (Mahamud et al., 2024, [14]). CLAHE preprocessing gains 8 percent of efficiency. In spite of integration and dataset limitations, a YOLOv4 + attention-LSTM model achieves a higher coherence score on PEIR (3,500 images) with BLEU 81.78% and METEOR 78.56% (Ravinder and Srinivasan, 2024, [15]). The domain-specific metrics, CNN-transformer hybrids, and reinforcement learning are reviewed on various datasets regarding Automatic Radiology Report Generation (ARRG) (Sloan et al., 2024, [16]).

CheXNet Transformer LSTM generated BLEU-1 0.4636 and BLEU-4 0.3575 on MIMIC-CXR with a combination of image feature extraction and report generation. Attention further improves the focus of ChestVision, though it tackles a problem of dataset bias, a generalization issue, by using rigorous training (Adel Aborizka, 2024, [17]). Multi-scale feature fusion models, based on channel attention and memory matrices, are able to combine both local and global characteristics with a ROUGE-L score of 0.68 on IU X-ray. However, the computational needs of these models do not suit well the resource- heavy environment (Pan et al., 2024, [18]). Although it is not as valuable in smaller clinics, vision encoder-decoder models in MIMIC-CXR augmented by reinforcement learning and text augmentation can achieve an F1 score of 66.2 (CheXbert) and 37.8 (RadGraph) and produce more diverse reports (Parres et al., 2024, [19]). Multimodal ARRG aims to be scalable proposing self-supervised learning and human-in-the-loop systems and also explains the encoder-decoder, retrieval-based, and hybrid encoder-decoder states in classifying integrated methods and datasets such as MIMIC-CXR and IU X-Ray and metrics such as BLEU and CheXbert F1 (Wang et al., 2024, [20]). Accuracy in terms of terminologies is better in CXR- IRGEN because it attains an F1 of 0.70 on MIMIC- CXR but does not excel in uncertainty which can be enhanced by ChestVision

(Shentu Al Moubayed, 2024, [21]. Based on BioGPT and co-attention, ChestBioX performs well in medical language fluency with a score of 0.30 on Open-i and still has problems with long reports (Ouis Akhloubi, 2024, [22]. The multimodal models incorporating demographics, vitals, and clinical history perform well at putting out the reports with the ROUGE-L score of 0.67 on the IU X-ray set but what they lack is what they do not have in rural clinics, namely, structured data (Aksoy et al., 2024, [23]. CXR- AGENT has a feature of included confidence levels, which it applies to perform well on F1 at 0.69 on MIMIC-CXR and also assists in adoption but has high computational requirements Sharma, 2024, [24].

Compared to ChestVision, multi-model binary classifiers cover fewer conditions but achieve 95% accuracy on cardiomegaly, effusion, and consolidation (Muharram et al., 2023, [25]).ChestVision's evaluation approach is shaped by a PRISMA-guided review of 41 studies that reveals issues with data imbalance and report generation across 14 datasets and metrics like BLEU, ROUGE, and CIDEr (Liao et al., 2023, [26]). On MIMIC- CXR and IU X-ray, CvT with DistilGPT2 enhances CE-F1 by 8.3%, BLEU-4 by 1.8%, and ROUGE-L by 1.6%, demonstrating pretraining benefits (Nicolson et al., 2023, [27]). For variable-length reports, GRU-based models capture long-range dependencies with BLEU-4 0.36 on Open-i, which is appropriate for environments with limited resources but has a modest accuracy (Akbar et al., 2023, [28]). Knowledge graph-based techniques improve concept extraction (F1 0.65 on IU X-ray) despite requiring manual graph creation; RadGraph uses multi-head attention to achieve BLEU-4 0.40 and ROUGE-L 0.66 (Zhang et al., 2022; Yang et al., 2022, [29–30]). VGG16 has a 100% accuracy rate in differentiating between COVID-19 and pneumonia on 2,000 images, which makes it suitable for rapid diagnosis but limited in its application (Islam & Tarique, 2022, [31]). Diagnostic content evaluation, which shows that BLEU alone is inadequate and encourages clinical metrics (F1 0.62, IU X-ray), served as the basis for ChestVision's design (Babar et al., 2021, [32]).

4.CONCLUSION

This review summarizes the state of the art in deep learning for CXR diagnostics and highlights the effective approaches of convolutional neural networks (CNNs), recurrent architectures (e.g., LSTM) and ensemble methods for disease identification and report generation. ChestVision harnesses these approaches through DenseNet's feature-rich patterns and LSTM's sequential generation for clinically relevant reports in a bid to decrease the deficit of trained radiologists in resource-poor areas, like in rural India. There is much promise in the literature with increasing accuracy from transformers to attention and innovative interpretability approaches (e.g., Grad- CAM) but limitations still exist, including biased datasets, diseases which can only be detected from CXR via limited aggregations, and lack of compute feasibility. ChestVision's aims with interpretability, with hospital deployment in mind, suggest extensive plans for addressing these limitations and those ancillary to building clinical trust. Future direction should focus on a larger validation set to include a diverse cumulative dataset for regionally relevant findings across the globe, as well as multimodal input (i.e., historical patient data to supplement) for additional diagnosis relevance. Finally, true radiological implementation requires clear thresholds and vetted medical questions for reporting as well as uncertainty quantification and medical meta-measurements to further support findings. ChestVision hopes to act as a safety net for trained radiologists, improving diagnosis quality and turnaround time via such secure, trustworthy AI approaches which render human oversight still necessary yet partnered. Ultimately, implementing such systems can radically change access to radiological assistance when addressing these pertinent findings.

5.REFERENCE

[1] Ouis, Mohammed Yasser, and Moulay A. Akhloufi. "Deep learning for report generation on chest X-ray images." *Computerized Medical Imaging and Graphics* 111 (2024): 102320.

[2] Mahamud, Eram, et al. "An explainable artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine- tuned transfer learning." *Decision Analytics Journal* 12 (2024): 100499.

[3] Nair, Rekha R., and Tripty Singh. "Exploring Ensemble Architectures for Lung X-Ray Multi-Class Image Classification Using CNN-LSTM." *Procedia Computer Science* 258 (2025): 852–861.

[4] Sirshar, Mehreen, et al. "Attention based automated radiology report generation using CNN and LSTM." *PLOS One* 17.1 (2022): e0262209.

[5] Yan, Fengqi, et al. "Combining LSTM and DenseNet for Automatic Annotation and Classification of Chest X-Ray Images." *IEEE Access* 7 (2019): 74181–74194.

[6] Visuña, Lara, et al. "Computer-Aided Diagnostic for Classifying Chest X-Ray Images Using Deep Ensemble Learning." *BMC Medical Imaging* 22.1 (2022): 178.

[7] Sîrbu, Iustin, et al. "GIT-CXR: End-to-End Transformer for Chest X-Ray Report Generation." *Information* 16.7 (2025): 524.

[8] Rodrigues, Gabriel Arquelau Pimenta, et al. "IHRAS: Automated Medical Report Generation from Chest X-Rays via Classification, Segmentation, and LLMs." *Bioengineering* 12.8 (2025): 795.

[9] Zhao, Jian, et al. "Automated Chest X-Ray Diagnosis Report Generation with Cross-Attention Mechanism." *Applied Sciences* 15.1 (2025).

[10] Wang, Zhichuan, et al. "Disease probability- enhanced follow-up chest X-ray radiology report summary generation." *Scientific Reports* 15.1 (2025): 26930.

[11] Rajpurkar, Pranav, et al. "ReXGradient-160K: A Large-Scale Publicly Available Dataset of Chest Radiographs with Free-text Reports." *arXiv preprint arXiv:2505.00228* (2025).

[12] Ouis, Mohammed Yasser, and Moulay A. Akhloufi. "Deep learning for report generation on chest X-ray images." *Computerized Medical Imaging and Graphics* 111 (2024): 102320.

[13] Ravinder, Paspula, and Saravanan Srinivasan. "Automated medical image captioning with soft attention-based LSTM model utilizing YOLOv4 algorithm." *Journal of Computer Science* 20.1 (2024): 52-68.

[14] Sloan, Phillip, et al. "Automated radiology report generation: A review of recent advances." *IEEE Reviews in Biomedical Engineering* 18 (2024): 368-387.

[15] Adel, Mohamed, and Mohamed Aborizka. "Automated radiology report generation from chest X-ray images using CheXNet and Transformer- LSTM architecture." (2024).

[16] Pan, Yu, et al. "Chest radiology report generation based on cross-modal multi-scale feature fusion." *Journal of Radiation Research and Applied Sciences* 17.1 (2024): 100823.

[17] Parres, Daniel, Alberto Albiol, and Roberto Paredes. "Improving radiology report generation

quality and diversity through reinforcement learning and text augmentation." *Bioengineering* 11.4 (2024): 351.

[18] Wang, Xinyi, et al. "A survey of deep learning- based radiology report generation using multimodal data." *arXiv preprint arXiv:2405.12833* (2024).

[19] Shentu, Junjie, and Noura Al Moubayed. "CXR-IRGEN: An integrated vision and language model for the generation of clinically accurate chest X-ray image-report pairs." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2024).

[20] Ouis, Mohammed Yasser, and Moulay A. Akhloufi. "ChestBioX-Gen: Contextual biomedical report generation from chest X-ray images using BioGPT and co-attention mechanism." *Frontiers in Imaging* 3 (2024): 1373420.

[21] Aksoy, Nurbanu, et al. "Beyond images: an integrative multi-modal approach to chest X-ray report generation." *Frontiers in Radiology* 4 (2024): 1339612.

[22] Sharma, Naman. "CXR-AGENT: Vision- language models for chest X-ray interpretation with uncertainty aware radiology reporting." *arXiv preprint arXiv:2407.08811* (2024).

[23] Muharram, Arief Purnama, et al. "Automated Chest X-Ray Report Generator Using Multi-Model Deep Learning Approach." *2023 IEEE International Conference on Data and Software Engineering (ICoDSE)*. IEEE, 2023.

[24] Liao, Yuxiang, Hantao Liu, and Irena Spasić. "Deep learning approaches to automatic radiology report generation: A systematic review." *Informatics in Medicine Unlocked* 39 (2023): 101273.

[25] Nicolson, Aaron, Jason Dowling, and Bevan Koopman. "Improving chest X-ray report generation by leveraging warm starting." *Artificial Intelligence in Medicine* 144 (2023): 102633.

[26] Akbar, Wajahat, et al. "Automated report generation: A GRU based method for chest X-rays." *2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE, 2023.

[27] Zhang, Dehai, et al. "Improving medical X-ray report generation by using knowledge graph." *Applied Sciences* 12.21 (2022): 11111.

[28] Yang, Shuxin, et al. "Knowledge matters: Chest radiology report generation with general and specific knowledge." *Medical Image Analysis* 80 (2022): 102510.

[29] Visuña, Lara, et al. "Computer-Aided Diagnostic for Classifying Chest X-Ray Images Using Deep Ensemble Learning." *BMC Medical Imaging* 22.1 (2022): 178.

[30] Islam, Rumana, and Mohammed Tarique. "Chest X-Ray Images to Differentiate COVID-19 from Pneumonia with Artificial Intelligence Techniques." *International Journal of Biomedical Imaging* (2022): 5318447.

[31] Babar, Zaheer, et al. "Evaluating diagnostic content of AI-generated radiology reports of chest X-rays." *Artificial Intelligence in Medicine* 116 (2021): 102075.

[32] Alfarghaly, Omar, et al. "Automated radiology report generation using conditioned transformers." *Informatics in Medicine Unlocked* 24 (2021): 1005.