UDC 004.93

# AUTOMATED CHEST X-RAY REPORT GENERATION USING ARTIFICIAL INTELLIGENCE BASED ATTENTION-ENHANCED GOOGLENET-LSTM ARCHITECTURE

**Paracha Muhammad Faheem** – Post-graduate student of the Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/my-orcid?orcid=0009-0009-7147-9593.

**Mahmood Mudasir** – Assistant Professor, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/0009-0004-3398-5384.

**Farhan Muhammad** – Lecture, Faculty of Computing, Gomal University, Dera Ismail Khan, Pakistan. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/my-orcid?orcid=0000-0001-6214-3013.

## ABSTRACT

**Context.** Proper and effective diagnosis of chest diseases is vital in timely treatment and efficient clinical decision-making. Chest X-rays (CXRs) are commonly utilized in the detection of chest diseases because they are readily accessible and cost-effective. Nevertheless, radiograph interpretation is still a time-consuming, subjective, and error-prone process, especially in health care settings where resources are limited. Radiologists require automated systems capable of producing consistent diagnostic information that will facilitate quicker and standardized patient care.

**Objective.** The proposed research will develop and assess a deep learning model with the ability to produce valid and explainable diagnostic reports using chest radiographs. The main aim is to minimize human error, save time in the diagnostic process, and deliver uniform findings, which will guide clinicians to make sound decisions within a short period of time.

**Method.** The images are then extracted using a convolutional neural network called GoogleNet that extracts high-level visual features, which contain important structural and anatomical information. The features extracted are then fed to a Long Short-Term Memory network, which represents the sequential character of the report generation process by conditioning itself on relationships between words and phrases in diagnostic text. To enhance its accuracy and interpretability, an attention mechanism is added to allow the system to concentrate on the most clinically valuable parts of the image when producing every part of the report. The Indiana University Chest X-ray dataset was employed to train and evaluate the proposed system, while several experiments were carried out to evaluate the performance of the proposed system regarding its performance against the existing benchmark models.

**Results.** The GoogleNet-LSTM-Attention model was shown to be more effective at generating high-quality diagnostic reports. It has performed well compared to benchmark models on various natural language evaluation measures, such as BLEU, ROUGE, and CIDEr scores. These improvements indicate that the quality of clinical data and fluency of text generated are correlated and that CNN, RNN, and attention mechanisms are effective in medical image reporting.

**Conclusions.** The study presented shows that a combination of CNNs, LSTMs, and attention in a single architecture has the potential to transform the process of interpreting chest X-rays. The system proposed not only improves the precision of the diagnosis but also provides clinical assistance, as it allows for performing radiographic assessment rapidly, consistently, and interpretably. Such AI-driven systems can bring about the potential to reduce workloads, decrease diagnostic errors, and enhance patient outcomes in various healthcare facilities.

**KEYWORDS:** Convolution Neural Networks, Deep Learning, Chest X-ray, Radiology Report Generation, GoogleNet, LSTM, Attention Mechanism.

## ABBREVIATIONS

AD is an Alzheimer's disease;
CNN is a Convolutional Neural Network;
ROC is a Receiver Operating Characteristic;
LSTM is a Long Short-term Memory;
AUC is an Area Under the Curve;
RNN is a Recurrent Neural Network;
BLUE is a Bilingual Evaluation Understudy;
ROUGH is a Recall-Oriented Understudy for Gisting Evaluation;
CIDEr is a Consensus-based Image Description Evaluation;
CXR is a Chest X-Ray.

## NOMENCLATURE

$S$ is a diagnostic report;
$I$ is an Images;
$M$ is a mapping;
$R$ is an expert-written reports.

## INTRODUCTION

Chest radiographs are still one of the most common and economically efficient tools of imaging in medicine. They are very important in the detection and monitoring of diseases like pneumonia, tuberculosis, pleural effusion, pneumothorax, and lung cancer [1]. Although CXRs are clinically important, their proper interpretation is difficult. It involves knowledge of thoracic anatomy, identification of subtle abnormalities, and correlation of results with patient history [2, 3]. Delays, poor reporting, and human error are common in cases of a shortage of trained radiologists, particularly in health care systems that are limited in resources. Manual reporting is also a time-consuming process, and inter-observer variability. These challenges are further exaggerated by the global burden of chest diseases. The World Health Organization explains that respiratory diseases lead to the death of millions of people per year [4, 5]. Low-resource countries like Pakistan have a higher prevalence rate of tuberculosis


DOI 10.15588/1607-3274-2026-2-9

because of poor medical facilities and overcrowded conditions, whereas lung cancer and COPD are more typical of developed countries, usually due to lifestyle and work-related exposures. These differences bring to light the critical requirement for diagnostic systems that are both scalable and dependable.

Recent improvements in artificial intelligence (AI) and deep learning (DL) have provided the possibilities of solving these problems [6–13]. CNNs can obtain higher-level semantic information in images, LSTMs can produce structured text, and attention mechanisms are able to enhance concentration on clinically significant areas. The given research suggests using a GoogleNet - LSTM-Attention model to produce meaningful and accurate reports and enhance the efficiency, minimize error rates, and assist radiologists working in a high-volume and underserved environment.

**The object of study** will be the creation of an AI-driven diagnostic system that combines GoogleNet, LSTM, and attention-based models to automatically create trustworthy and clinically relevant chest X-ray reports.

**The subject of the study** is the analysis of the chest radiographs (CXRs) and their interpretation with the deep learning methodology, with particular emphasis on the automated report writing.

**The purpose of the study** is to determine whether coherent, accurate, and clinically interpretable diagnostic reports can be produced by an integrated GoogleNet-LSTM-Attention framework. The main aim is to improve the diagnostic accuracy, lessen the radiologist workload, decrease reporting inconsistencies, and enable expert-level interpretation, especially in health facilities with limited resources.

## 1 PROBLEM STATEMENT

The Automated Chest Radiograph Diagnostic Report Generation Problem can be stated as follows: Train an Automated Chest Radiograph Diagnostic Report Generation system such that it can automatically convert chest X-ray images I to meaningful diagnostic reports S that are comparable in quality and accuracy to those produced by expert writers of diagnostic reports R. Formally, this means finding a mapping M: I → S that uses well-known natural language evaluation metrics like BLEU, ROUGE, and CIDE to make the generated and reference reports as close as possible. The problem is not only to create linguistically fluent text only but also medically accurate, so that what is generated in the reports complies with the medical terms and clinical utility. The problem is complicated further by other constraints: first, datasets of chest X-ray images frequently have a severe class imbalance where normal images vastly outnumber abnormal ones, which can bias the system into less informative results; second, generated reports have a fixed limit on length and require a restricted set of medical words to prevent ambiguity; and thirdly, the attention mechanism that drives the system should pay attention to clinically relevant parts of the image and render valid probability distributions across the regions. Overcoming these concerns is important in creating reliable, interpretable, and clinically useful automated reporting systems that can assist radiologists in practice.

## 2 REVIEW OF THE LITERATURE

The recent years have witnessed significant improvements in the generation of diagnostic reports directly from chest X-rays. Advanced encoder-decoder systems serve as the main basis for improvements that enhance image and text processing capabilities. Certain systems combine image features with text features through repeating patterns to generate more reliable reports with clear meaning. BERT serves as an advanced language model that analyzes word meaning and context to improve diagnostic report text understanding [14]. Other smart systems utilize a fusion of the features of deep learning models and text templates to produce well-formatted reports. One such example is ChestBioX-Gen [15], an architecture that links images and text and utilizes a memory network to improve the model's perception of what it observes and what it writes. It also increases the accuracy by the use of weak labels, or basic textual cues, in case the data has not been labeled correctly. The quality of these reports has significantly increased due to the inclusion of attention mechanisms that aid the model in focusing on specific image areas. These mechanisms can aid the model in paying attention to small but critical features within images and texts to improve the outcome.

The major challenge when working with chest X-ray datasets stems from their unbalanced nature, which results in more occurrences of normal findings than diagnostic ones. Such a mismatch creates challenges for the model to learn disease recognition capabilities. A summary of the primary chest X-ray research datasets appears in Table 1.

The authors of the paper [21] proposed "MedWriter", a hierarchical retrieval-based framework designed to automatically extract and use report and sentence-level templates for generating clinically accurate medical reports. MedWriter employs Visual-Language Retrieval (VLR) and Language-Language Retrieval (LLR) modules to maintain logical coherence and template accuracy. The method integrates multi-query attention to fuse visual and retrieved textual features effectively. Evaluations on the Open-I and MIMIC-CXR datasets demonstrated its superior performance, verified through CIDEr, ROUGE-L, and BLEU scores. Chen, Shen, Yan, and Wan [22] and Kaur and Mittal [23] work on an encoder-decoder framework presenting Cross-modal Memory Networks (CMN) with multi-attention and one-shot global pruning for efficient chest X-ray report generation. Tests conducted on the IU X-Ray, MIMIC-CXR, and OpenI datasets demonstrate competitive accuracy with significantly lower computational demands, verified by BLEU and ROUGE, and CIDER evaluation metrics.


DOI 10.15588/1607-3274-2026-2-9

Table 1 – Overview of Available Chest X-Ray Datasets

| Dataset Name | Source Institution | Labeled Diseases | Total Images | Total Reports | Patient Count |
|---|---|---|---|---|---|
| Chest X-ray14 [16] | National Institutes of Health (NIH) | 14 common thoracic diseases | 112.120 | Not publicly available | 30.805 |
| CheXpert [17] | Stanford University | 14 observations with uncertainty labels | 224.316 | 223.462 | 65.240 |
| MIMIC-CXR [18] | Massachusetts Institute of Technology (MIT) | Multiple thoracic conditions | 377.110 | 227.835 | 65.379 |
| IU Chest X-Ray [19] | Indiana Network for Patient Care | Expert | 8.121 | 3.996 | 3.996 |
| PadChest [20] | Hospital Universitario de San Juan, Alicante (Spain) | Human Evaluation + Neural Network | 1.60.000 | 206000 | 67.000 |

The paper [24] leverages the Cross-View Attention Modules (CVAM) and Medical Visual-Semantic LSTMs (MVSL) with co-attention to promote the contextual accuracy of the chest X-ray report generation. The IU-X-Ray dataset was evaluated with high scores on BLEU and ROUGE, which proves its clinical relevance and coherence. Yang, Wu, Ge, Zheng, Zhou, Xiao [25] proposed a model incorporating a learned knowledge base and multi-modal alignment. The method automatically minimizes medical knowledge based on textual embeddings, which makes it possible to match image data, reports, and disease labels in a specified way. The analysis performed on IU-Xray and MIMIC-CXR datasets depicted superior performances in a range of clinical and natural language generation metrics, which demonstrates that the model could produce clinically useful radiology reports. The model is an innovative use of hierarchical image mapping, which improves perception without raising the computational load compared with standard encoder-decoder models that equalize attention to all features used in the encoding process. Evaluation across multilingual datasets in English and Portuguese indicated promising results, as demonstrated by ROUGE-L and METEOR metrics, particularly on NIH Chest X-ray datasets. Zhao, Yao, Sun, Shi, Kuang, Wu, and Han [27] combined Convolutional Block Attention Module (CBAM) with a cross-attention mechanism into an integrated model for their analysis. The visual encoder employs ResNet-101 with CBAM capabilities to identify abnormal areas in chest X-rays, alongside the cross-attention mechanism, which enhances the alignment between image features and text sequences for better report generation. Table 2 is the list of the latest related studies on medical image report generation in summary format.

Table 2 – Summary of Existing Studies

| Serial No. | Reference paper | Model/ | Dataset | Metric Used | Limitation |
|---|---|---|---|---|---|
| 1 | [21] 2021 | MedWriter | Open-I & MIMIC-CXR | BLEU, ROUGE,CIDER | Retrieval dependency |
| 2 | [22] 2021 | Cross Model Memory Network (CMN) | IU X-Ray & MIMIC-CXR | BLEU, ROUGE, METEOR | Alignment gaps |
| 3 | [23] 2023 | CheXPrune | Open-i | BLEU, ROUGE,CIDER | Pruning losses |
| 4 | [24] 2023 | CVAM+MVSL | Chexpert & IU X-ray | BLEU, ROUGE,CIDER | semantic ambiguity |
| 5 | [25] 2023 | Learned Knowledge Base (LKB) and Multi-Model Alignment (MMA) | IU X-Ray & MIMIC-CXR | BLEU, ROUGE,CIDER | Static knowledge |
| 6 | [26] 2024 | XRaySwinGen [27] | IU X-ray & NIH Chest X-ray | BLEU,ROUGE, SPICE,METEOR | High complexity |
| 7 | [27] 2025 | Convolutional Block Attention Module (CBAM) with a cross-attention mechanism | IU X-Ray | BLUE,ROUGE, METEOR | Attention limitations |
| 8 | [28] 2025 | ChestX-Transcribe | IU X-Ray | BLUE,ROUGE, METEOR | Computational Constraints |


DOI 10.15588/1607-3274-2026-2-9

## 3 MATERIALS AND METHODS

This study introduces a modified theoretical framework that unifies GoogleNet, Long Short-Term Memory (LSTM), and an attention mechanism into an integrated architecture for automated diagnostic report generation from chest X-ray images. The theoretical advancement of this work is based on defining the mapping $M: I \rightarrow S$, where I represents the space of input chest X-ray images and S denotes the space of generated diagnostic sentences. This mapping formalizes the transformation of visual medical data into semantically meaningful textual interpretations, forming the core theoretical contribution of the proposed method.

The architecture extends the conventional encoder-decoder paradigm by embedding a probability-normalized attention scoring function capable of assigning adaptive weights to different image feature vectors. Unlike standard encoder-decoder models that treat all encoded features equally, the proposed attention formulation can selectively emphasize parts of the image that are of clinical interest. This has improved interpretability and accuracy by giving more priority to anatomical constructions, like lungs, diaphragm, and the cardiac silhouette, during sentence generation. The attention mechanism is mathematically implemented by the use of a feedforward scoring network consisting of a tanh activation and a softmax normalization step. These normalized scores are used as the coefficients to obtain a context vector, which is calculated by taking the weighted sum of all the encoded feature vectors.

This study was done on the publicly accessible Indiana University (IU) Chest X-ray data, encompassing 7.470 images of chest X-ray and 3.955 diagnosis reports prepared by professional radiologists. Figure 1 presents the general procedure of the offered model.

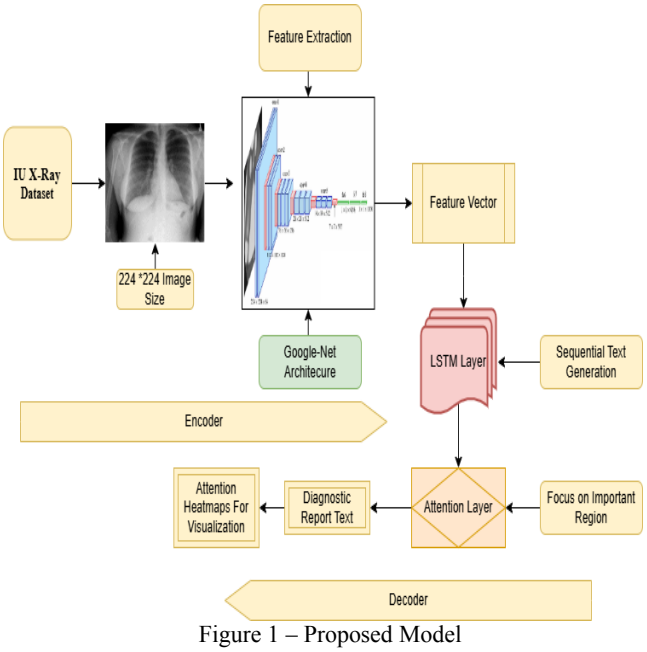


Figure 1 – Proposed Model

Each reports have a series of sections where rich textual data is given, including Findings, Impression, and Tags, which can be used to train. To prepare the X-ray images before training, all the X-ray images had to be resized to the standardized resolution of 224 x 224 pixels to fit into the input specifications of GoogleNet. The feature extractor was GoogleNet, a convolutional neural network that is pre-trained with the ImageNet dataset and can capture both local and global visual information of the chest X-ray images. GoogleNet generated high-dimensional feature maps, which were converted into fixed-length feature vectors that reflected the semantic content of every image. Such coded features of visual representations were subsequently sent to an LSTM-based decoder to generate a sequence of texts. The decoder was developed as a hierarchical unrolled LSTM network equipped with the capability to learn the temporal dependencies that exist between words. The network employed the ReLU activation function, and the size of the hidden state will be 256 units. The word embedding of the token that had been generated earlier and the context vector that was derived from the attention mechanism were the two primary inputs that were received by the decoder at each time step individually.

This allowed the model to produce coherent, contextually relevant sentences and to be grammatically consistent across the report. An attention mechanism was added between the encoder and decoder to increase the interpretability. The module dynamically gave attention weights on various areas of the image, which enabled the model to target medically important organs like the lungs or heart while producing diagnostic statements. The attention mechanism took each image region and calculated a score with the help of a feedforward neural network using the tanh activation, where the weights were normalized in a softmax layer. The resulting context vector was computed as being the weighted additive sum of all the image feature vectors, focusing on the most informative parts of the image. A batch size of 64 and multiple epochs were used to train the model until it converged. To avoid overfitting, dropout regularization was employed, and the difference between the generated and reference reports was minimized with the help of the cross-entropy loss function. Adam optimizer was used as a model optimization method because of its ability to handle sparse gradients and adaptive learning efficiently.

## 4 EXPERIMENTS

The experiment was performed to assess the efficiency of the proposed GoogleNet-LSTM with attention mechanisms to produce automatic diagnostic reports based on chest X-ray images. It was based on the Indiana University (IU) Chest X-Ray dataset which included 7.470 images and 3.955 diagnostic reports. The textual description of each X-ray was given by a radiologist, which was used as a reference during training and assessment. All images were scaled to 224x224 pixels, and pixel values were brought to a range of 0 to 1 to provide a similarity in feeding the model. Data


DOI 10.15588/1607-3274-2026-2-9

augmentation methods were implemented to avoid overfitting and make the model more general, which included rotation, zooming, shifting, and horizontal flipping. To extract features, the pretrained weights of GoogleNet were used in ImageNet. This network encoded the X-ray images into feature vectors that showed important visual details like textures, lung borders, structural patterns, and so on. The features that were extracted were subsequently fed to an LSTM-based decoder, which was tasked with the production of textual diagnostic sentences. The model also incorporated an attention mechanism to enable the decoder to give attention to certain areas of the image that were significant to the clinic. The attention network focused on different regions of the features at every time step, giving weight to areas that were more relevant to the diagnosis. These attention weights have been obtained by using a softmax function, which transformed raw attention scores into a probability distribution across the image pixels. The LSTM used the context vector, which was a weighted average of the image features, in generating the next word in the report. This method enabled the model to dynamically highlight important areas of the lungs as well as decrease attention to the unimportant areas, like ribs or background. The model aimed to reduce the gap between the expected and actual reports. The prediction errors were calculated by a loss function formed on the categorical cross-entropy, and these errors were backpropagated through the network to update weights during training. While training the model, batches of 64 photos were used for each iteration over the course of numerous epochs. A dropout mechanism was incorporated to prevent the model from being overfit.

The ReLU activation function was used in intermediate layers, and the word embedding size was set to 256. To assess the quality of generated medical reports, three standard evaluation metrics were used: BLEU, ROUGE, and CIDEr. BLEU measured how closely the machine-generated text matched the reference reports at various n-gram levels (BLEU-1 through BLEU-4). ROUGE emphasized recall, evaluating how important content from the reference report was captured in the generated one. CIDEr measured the consensus between the generated and human-written reports, focusing on semantic and contextual alignment rather than mere word overlap.

## 5 RESULTS

A GoogleNet-LSTM framework with attention mechanisms evaluated the IU Chest X-Ray.

BLEU (Bilingual Evaluation Understudy), ROUGE (Recall Oriented Understudy for Gisting Evaluation), and CIDEr (Consensus-based Image Description Evaluation) scores are used as the evaluation methods to measure the consistency between human expert reports and machine-generated reports based on our model. In this study, BLEU measures precision at the n-gram level, which includes single words (BLEU-1), two-word phrases (BLEU-2), three-word chains (BLEU-3), and four-word chains (BLEU-4). For instance, the Bleu-1 focuses on key word matching, while Bleu-4 focuses on the completeness of the matching phrases. ROUGE also evaluates the overlap of n-grams, word sequences, and word pairs between the candidate and reference sentences.

It particularly focuses more on recall than precision. CIDEr is another metric we used in our research, which measures the correlation and similarity between the ground truth caption and the predicted caption. It works on the concept that the generated caption should not only be similar to the ground truth caption in terms of word choice and grammar but also in terms of meaning and content. Moreover, we compare the proposed Model with numerous existing architectures that are currently working and are represented in Table 2.

Table 3 summarizes various methods for creating medical reports from chest X-ray images through

Table 3 – Key outcomes for BLEU, ROUGE and CIDEr sores calculated across multiple n-grams using the suggested approach for medical report generation on the IU X-Ray dataset

| Dataset | Methods | BLEU-1 | BLEU-2 | BLEU-3 | BLEU-4 | ROUGE | CIDER |
|---|---|---|---|---|---|---|---|
| IU-X Ray | MedWriter [21] | 0.471 | 0.336 | 0.238 | 0.166 | 0.382 | 0.345 |
| | Cross Model Memory Network (CMN) [22] | 0.475 | 0.309 | 0.222 | 0.170 | 0.375 | – |
| | CheXPrune[23] | 0.543 | 0.446 | 0.374 | 0.320 | 0.598 | 0.322 |
| | CVAM+MVSL [24] | 0.460 | 0.294 | 0.207 | 0.152 | 0.385 | 0.409 |
| | Learned Knowledge Base (LKB) and Multi Model Alignment (MMA) [25] | 0.497 | 0.319 | 0.230 | 0.174 | 0.399 | 0.407 |
| | XRaySwinGen [26] | 0.470 | 0.304 | 0.219 | 0.165 | 0.371 | – |
| | Convolutional Block Attention Module (CBAM) with a cross-attention mechanism [27] | 0.456 | 0.294 | 0.205 | 0.152 | 0.364 | – |
| | ChestX-Transcribe [28] | 0.675 | 0.585 | 0.523 | 0.472 | 0.72 | – |
| | **GoogleNet – LSTM (Without Attention)** | **0.569** | **0.406** | **0.316** | **0.276** | **0.492** | **0.290** |
| | **GoogleNet – LSTM (With Attention)** | **0.795** | **0.542** | **0.417** | **0.336** | **0.510** | **0.364** |


DOI 10.15588/1607-3274-2026-2-9

evaluation with BLEU, ROUGE, and CIDEr score metrics on the IU X-Ray dataset. The results are compared with other methods like MedWriter, CheXPrune, XRaySwinGen, CBAM, and ChestX-Transcribe, etc., against two GoogleNet-LSTM variants (with and without attention). The ChestX-Transcribe achieves superior scores in BLEU-2 through BLEU-4 (0.585, 0.523, 0.472), which demonstrates better phrase cohesion, but the GoogleNet-LSTM with attention obtains the best BLEU-1 (0.795) score because of its strong matching to reference reports at the single-word level. The scores of other n-grams are promising as well, which shows the good performance of the proposed model. The GoogleNet-LSTM without attention produces high BLEU scores in 1-gram and 2-gram (0.569, 0.406) as compared to other mentioned models. But achieves poor results above gram-2 (0.316, 2.276), which indicates lower consistency across multiple words. The test results show that attention-based models produce superior outcomes than non-attention techniques since they effectively target key image areas to build better medical reports. Experiments also show that the good results of other two metrices ROUGE and CIDEr. The Results of ROUGE are better in ChestX-Transcribe and CheXPrune then our model which shows the more room of research and betterment in the proposed mode as future work.

The next figures display the comparison of training and validation loss and accuracy throughout the epochs.

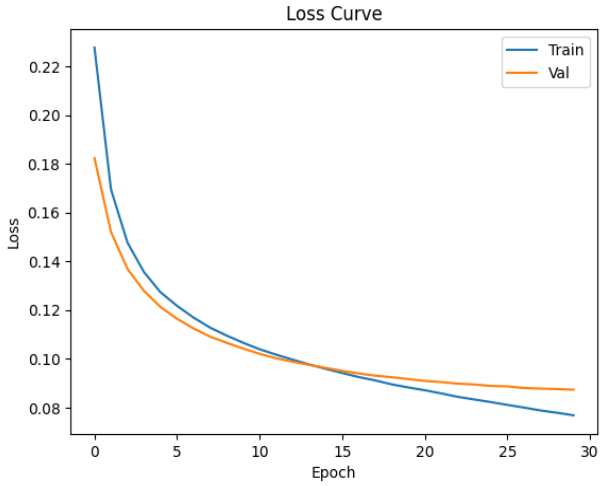


Figure 2 – Loss curve of the proposed Model

The accuracy of the proposed GoogleNet-LSTM model with attention mechanisms appears in Figure 3 throughout the training process. The performance of the model over time is tracked through the graph showing its success at producing precise diagnostic reports from chest X-rays. The model's accuracy increases throughout the training period, which demonstrates its effective learning of the provided information. Figure 2 shows the training loss of the same model during epochs. The loss measurement shows how well the model matches actual written reports during prediction. The training loss begins high but steadily declines since the model enhances its performance during training. After training for several epochs, the model shows no further loss reduction as it has achieved peak performance.

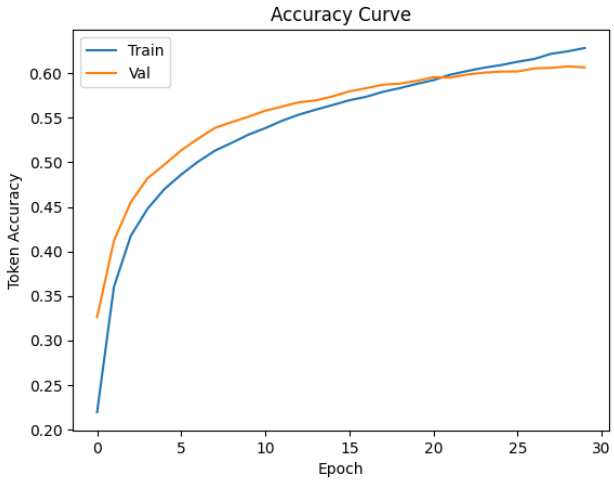


Figure 3 – Learning Curve accuracy of the proposed Model

Furthermore, some of the qualitative results with medical report generation utilize the conventional Encoder-Decoder model. The output contains the radiograph name along with the ground truth or findings based on the doctor's findings and the predicted results with conventional Encoder-Decoder.

Several qualitative examples illustrate the model's performance in generating diagnostic reports from chest X-ray images. For radiograph CXR2805, the ground truth describes clear lungs bilaterally, a normal heart size, and no evidence of pneumothorax or focal consolidation, while the predicted caption reflects similar findings but includes repeated words and minor structural inconsistencies. For CXR901, the radiologist's report states that there is no pneumothorax, no focal air space opacity suggesting pneumonia, and a normal heart, whereas the model-generated caption identifies clear lungs and intact bony structures but again contains repetition. In the case of CXR6, the ground truth indicates a normal cardiac silhouette, unremarkable mediastinum and perihilar regions, clear lungs, and normal osseous structures; the model's prediction captures the normal heart size and unremarkable mediastinum but repeats terms and lacks full detail. For CXR3999, the radiologist notes normal lung inflation without focal airspace disease, a normal heart size, and normal mediastinal contours, while the predicted caption conveys similar observations but includes duplicated words and simplified phrasing. These examples collectively demonstrate that the model is capable of capturing core diagnostic elements, though with some limitations in linguistic fluency and repetition.

## 6 DISCUSSION

Some qualitative results of medical report generation using the Proposed Attention-based Encoder-Decoder Architecture are depicted in below in Figure 4.


DOI 10.15588/1607-3274-2026-2-9

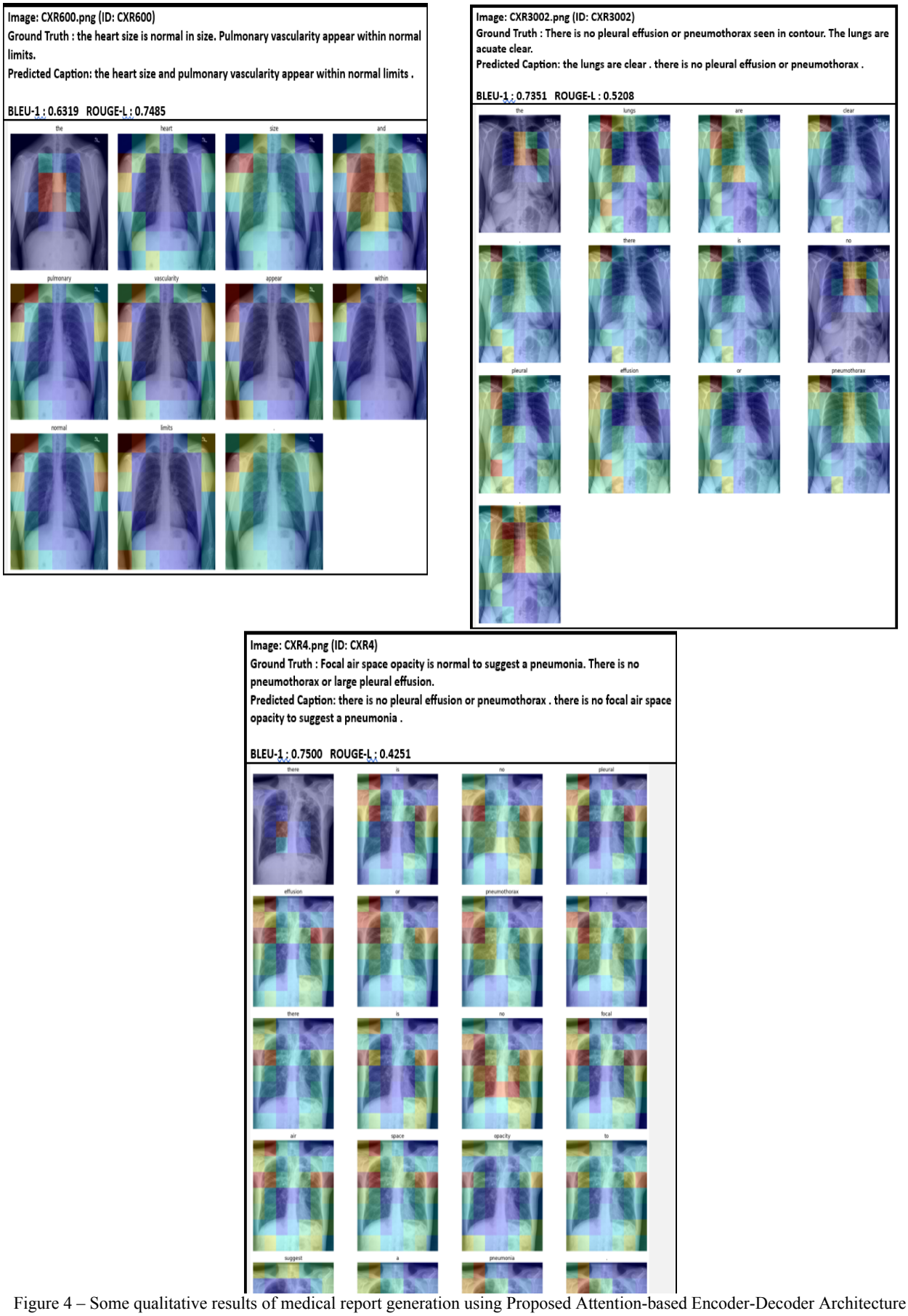


Figure 4 – Some qualitative results of medical report generation using Proposed Attention-based Encoder-Decoder Architecture


DOI 10.15588/1607-3274-2026-2-9

Figure 4 shows the text generation results of an attention-based encoder-decoder system which generate medical reports from chest X-ray images. The figure shows that the model matches what radiologists write and creates professional diagnostic reports. The results show the model's superiority over conventional encoder-decoder methods regarding its capability in retrieving underlying findings and its consistency with ground truth report.

## CONCLUSIONS

The suggested Model generate automatic textual reports from chest X-rays to help medical practitioners save time and effort. The system relies on a convolutional neural network (CNN) feature extraction model, which is used as an encoder to transform images into vector representations of fixed size. Then, using the learned features of the images, an RNN decoder is employed to produce matching phrases. The Model's efficacy is examined in the Indiana dataset through quantitative and qualitative analyses. The impact of various components on medical report creation has been examined through comparative studies of alternative methodologies, which have also showcased numerous use cases for the proposed system. The study used GoogleNet as CNN (encoder) and LSTM as RNN (decoder) in our model to which gives state of the art results which shows the efficiency of our model. Further enhancement of the performance could be achieved by expanding the dataset size and training the model on more images. The use of the deep learning models, including the GoogleNet-LSTM with the attention mechanism, to optimize the chest radiograph diagnostics. Through the addition of attention mechanisms, the Model paid maximum attention to essential portions of the radiographs, making the interpretation of the radiographs more justifiable and emphasizing the high importance of the features. The strategy enhances the accuracy of the diagnoses because it helps the Model to learn various areas of the chest radiograph that are more indicative of specific conditions. Moreover, this Model is effective in controlling bulk data in practice and best-fit applications, including supporting radiologists in their busy clinical setting and providing diagnostic support units with a limited approach to healthcare workers. The combination of GoogleNet-LSTM and attention mechanisms is a major step in the development of the automated diagnostic systems. Further testing and fine-tuning of the different sets of the chest radiographs could greatly enhance the outcome of patients, their diagnostic, and the probability of human error.

## DECLARATIONS

**Conflict of interest**: The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

**Authors' contributions**: **Muhammad Faheem Paracha:** Conceptualization of the study, model design, data preprocessing, implementation of the GoogleNet-LSTM-Attention architecture, experimentation, and original manuscript preparation. **Mudasir Mahmood:** Supervision, methodological guidance, validation of experimental results, critical review, and editing of the manuscript. **Muhammad Farhan:** Literature review, result analysis, performance comparison with existing methods, manuscript revision, and final proofreading.

**Data availability**: The manuscript has associated data in a data repository https://github.com/muhammadfarhan01-hub/checst-reposrt-generation .

**Software availability:** The manuscript has no associated software.

**Use of artificial intelligence tools:** The authors confirm that they did not use artificial intelligence technologies in creating the submitted work.

## REFERENCES

1. Wang X., Peng Y., Lu L. et al. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *IEEE Conf. Comput. Vis. Pattern Recogni,* 2017, pp. 3462–3471.
2. Brady A. P. Measuring radiologist workload: How to do it, and why it matters. *Eur. Radiol.,* 2011, Vol. 21, № 11, pp. 2315–2317.
3. Cowan I. A., MacDonald S. L. S., Floyd R. A. Measuring and managing radiologist workload: Measuring radiologist reporting times using data from a Radiology Information System. *J. Med. Imaging Radiat. Oncol.,* 2013, Vol. 57, № 5, pp. 558–566.
4. Geel van K., Lameijer I., Wielaard S. et al. Chest X-ray evaluation training: impact of normal and abnormal image ratio and instructional sequence. *Med. Educ.,* 2019, Vol. 53, № 2, pp. 153–164.
5. Islam M. S., Rahman M. M., Mahmud S. et al. Challenges issues and future recommendations of deep learning techniques for SARS-CoV-2 detection utilising X-ray and CT images: a comprehensive review. *PeerJ Comput. Sci.,* 2024, Vol. 10.
6. Li Q., Zhang W., Zhang X. et al. A Survey on Text Classification: From Shallow to Deep Learning. *ACM Trans. Intell. Syst. Technol. Association for Computing Machinery*, 2020, Vol. 37, № 4.
7. Sajed S., Inayat M., Qadir A. et al. The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review. *Appl. Soft Comput.,* 2020, Vol. 147.


DOI 10.15588/1607-3274-2026-2-9

8. Hwang E. J., Park S. Y., Kim J. H. et al. Deep learning for chest radiograph diagnosis in the emergency department. *Radiology,* 2019, Vol. 293, № 3, pp. 573–580.
9. Rajpurkar P., Irvin J., Zhu K.et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med.,* 2018, Vol. 15, № 11, pp. 1–17.
10. Sahlol A. T., Abdulkadir M., Elaziz A. A. et al. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features. *Symmetry (Basel),* 2020, Vol. 12, № 7.
11. Majkowska A., Pham C., He K. et al. Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology*, 2020, Vol. 294, № 2. pp. 421–431.
12. Li X., Wang Y., Xu S. et al. Deep Learning in Chest Radiography: Detection of Pneumoconiosis. *Biomed. Environ. Sci.,* 2021, Vol. 34, № 10, pp. 842–845.
13. Uçar M. Deep neural network model with Bayesian optimization for tuberculosis detection from X-Ray images. *Multimed. Tools Appl. Springer US,* 2023, Vol. 82, № 24, pp. 36951–36972.
14. Devlin J., Chang M., Lee K. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 – 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. – Proc. Conf. Association for Computational Linguistics (ACL),* 2018, Vol. 1, pp. 4171–4186.
15. Ouis M. Y., Akhloufi M. A. ChestBioX-Gen: contextual biomedical report generation from chest X-ray images using BioGPT and co-attention mechanism. *Front. Imaging.,* 2024, Vol. 3.
16. Wang X., Peng Y., Lu L. et al. ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases. *Adv. Comput. Vis. Pattern Recognit.,* 2019, № September, pp. 369–392.
17. Irvin J., Rajpurkar P., Ko M. et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, 2019, pp. 590–597.
18. Johnson A. E. W., Pollard T. J., Berkowitz S. J. et al. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs, 2019, Vol. 14, pp. 1–7.
19. Nicolson A., Dowling J., Koopman B. Improving chest X-ray report generation by leveraging warm starting. *Artif. Intell. Med. Elsevier B.V.*, 2023, Vol. 144, № April 2022, P. 102633.
20. Bustos A., Pertusa Y., Salinas J. M. et al. PadChest: A large chest x-ray image dataset with multi-label annotated reports. *Med. Image Anal.,* 2020, Vol. 66, pp. 1–35.
21. Yang X., Zhou J., Li Y. et al. Writing by memorizing: Hierarchical retrieval-based medical report generation. *ACL-IJCNLP 2021 – 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.,* 2021, pp. 5000–5009.
22. Chen Z., Peng Y., Liu X. et al. Cross-modal memory networks for radiology report generation. A*CL-IJCNLP 2021 – 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf.,* 2021, № 2018, pp. 5904–5914.
23. Kaur N., Mittal A. CheXPrune: sparse chest X-ray report generation model using multi-attention and one-shot global pruning. *J. Ambient Intell. Humaniz. Comput. J Ambient Intell Humaniz Comput,* 2023, Vol. 14, № 6, pp. 7485–7497.
24. Gu Y., Li X., Zhang L. et al. Automatic Medical Report Generation Based on Cross-View Attention and Visual-Semantic Long Short Term Memorys. *Bioengineering. Multidisciplinary Digital Publishing Institute (MDPI),* 2023, Vol. 10, № 8.
25. Yang S., Yin H., Wang X. et al. Radiology report generation with a learned knowledge base and multi-modal alignment. *Med. Image Anal.,* 2023, Vol. 86.
26. Veras Magalhães G., Santos dos L., Ferreira M. et al. XRaySwinGen: Automatic medical reporting for X-ray exams with multimodal model. *Heliyon, Elsevier Ltd*, 2024, Vol. 10, № 7.
27. Zhao J., Chen Y., Li Q. et al. Automated Chest X-Ray Diagnosis Report Generation with Cross-Attention Mechanism. *Appl. Sci.*, 2025, Vol. 15, № 1.
28. Singh P., Singh S. ChestX-Transcribe: a multimodal transformer for automated radiology report generation from chest x-rays. *Front. Digit. Heal. Frontiers Media SA,* 2025, Vol. 7, P. 1535168.

Received 12.10.2025.
Accepted 02.04.2026.
Published 26.06.2026.


DOI 10.15588/1607-3274-2026-2-9

УДК 004.93

# АВТОМАТИЗОВАНЕ СТВОРЕННЯ ЗВІТІВ ПРО РЕНТГЕНОГРАМИ ГРУДНОЇ КЛІТКИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ НА ОСНОВІ АРХІТЕКТУРИ GOOGLE NET-LSTM З ПІДВИЩЕНОЮ УВАГОЮ

**Парача Мухаммад Фахім** – доктор філософії, факультет обчислювальної техніки, Університет Гомаля, Дера Ісмаїл Хан, Пакистан. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/my-orcid?orcid=0009-0009-7147-9593.

**Махмуд Мудасір** – доцент, факультет обчислювальної техніки, Університет Гомаля, Дера Ісмаїл Хан, Пакистан. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/0009-0004-3398-5384.

**Фархан Мухаммад** – лекція, факультет обчислювальної техніки, Університет Гомаля, Дера Ісмаїл Хан, Пакистан. ROR: https://ror.org/0241b8f19. ORCID: https://orcid.org/my-orcid?orcid=0000-0001-6214-3013.

## АНОТАЦІЯ

**Актуальність.** Точна та ефективна діагностика захворювань грудної клітки має вирішальне значення для своєчасного лікування та прийняття ефективних клінічних рішень.

Рентген грудної клітки (РГК) широко використовується для виявлення захворювань грудної клітки завдяки своїй доступності та економічній ефективності. Однак ручна інтерпретація рентгенограм залишається трудомісткою, суб'єктивною та схильною до помилок, особливо в медичних закладах з обмеженими ресурсами. Автоматизовані системи, які можуть генерувати надійні діагностичні дані, необхідні для підтримки радіологів у наданні швидшої та стандартизованої медичної допомоги.

**Мета.** Це дослідження спрямоване на розробку та оцінку системи на основі глибокого навчання, здатної генерувати точні та інтерпретовані діагностичні звіти з рентгенограм грудної клітки. Основною метою є зменшення людських помилок, економія часу діагностики та забезпечення послідовних результатів, які допомагають клініцистам швидко приймати обґрунтовані рішення.

**Метод.** Запропонована структура інтегрує три основні компоненти глибокого навчання для генерації точних діагностичних звітів з рентгенограм грудної клітки. GoogleNet, надійна згорткова нейронна мережа, використовується для вилучення високорівневих візуальних ознак із зображень, фіксуючи важливі структурні та анатомічні деталі. Ці вилучені ознаки потім передаються до мережі довгострокової пам'яті, яка моделює послідовний характер генерації звітів, вивчаючи зв'язки між словами та фразами в діагностичному тексті. Вбудовано механізм уваги, який дозволяє системі зосереджуватися на найбільш клінічно значущих областях зображення під час генерації кожної частини звіту, покращуючи як точність, так і інтерпретованість. Набір даних рентгенографії грудної клітки Університету Індіани був використаний для навчання та оцінки, а також було проведено численні експерименти для порівняння продуктивності запропонованої системи з існуючими еталонними моделями.

**Результати.** Модель GoogleNet-LSTM-Attention продемонструвала чудову продуктивність у створенні високоякісних діагностичних звітів. Вона перевершила еталонні моделі за кількома показниками оцінки природною мовою, включаючи оцінки BLEU, ROUGE та CIDEr. Ці покращення відображають як точність клінічної інформації, так і плавність згенерованого тексту, підкреслюючи ефективність поєднання механізмів CNN, RNN та уваги у звітності про медичні зображення.

**Висновки.** Це дослідження доводить, що інтеграція CNN, LSTM та уваги в єдину архітектуру може революціонізувати інтерпретацію рентгенографії грудної клітки. Запропонована система не тільки підвищує точність діагностики, але й пропонує практичну клінічну підтримку, забезпечуючи швидку, послідовну та інтерпретовану рентгенографічну оцінку. Такі системи на базі штучного інтелекту є перспективними для зменшення робочого навантаження, мінімізації діагностичних помилок та покращення результатів лікування пацієнтів у різних медичних закладах.

**КЛЮЧОВІ СЛОВА:** згорткові нейронні мережі, глибоке навчання, рентген грудної клітки, генерація радіологічних звітів, GoogleNet, LSTM, механізм уваги.

## ЛІТЕРАТУРА

1. ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases/ [X. Wang, Y. Peng, L. Lu et al.] // IEEE Conf. Comput. Vis. Pattern Recognit. – 2017. – P. 3462–3471.
2. Brady A. P. Measuring radiologist workload: How to do it, and why it matters / A. P. Brady // Eur. Radiol. – 2011. – Vol. 21, № 11. – P. 2315–2317.
3. Cowan I. A. Measuring and managing radiologist workload: Measuring radiologist reporting times using data from a Radiology Information System / I. A. Cowan, S. L. S. MacDonald, R. A. Floyd // J. Med. Imaging Radiat. Oncol. – 2013. – Vol. 57, № 5. – P. 558–566.
4. Chest X-ray evaluation training: impact of normal and abnormal image ratio and instructional sequence/ [K. van Geel, I. Lameijer, S. Wielaard et al. ] // Med. Educ. – 2019. – Vol. 53, № 2. – P. 153–164.
5. Challenges issues and future recommendations of deep learning techniques for SARS-CoV-2 detection utilising X-ray and CT images: a comprehensive review/[M. S. Islam, M. M. Rahman, S. Mahmud et al.] // PeerJ Comput. Sci. –2024. – Vol. 10.
6. A Survey on Text Classification: From Shallow to Deep Learning/ [Q. Li, W. Zhang, X. Zhang et al.] // ACM Trans. Intell. Syst. Technol. Association for Computing Machinery. – 2020. – Vol. 37, № 4.


DOI 10.15588/1607-3274-2026-2-9

7. The effectiveness of deep learning vs. traditional methods for lung disease diagnosis using chest X-ray images: A systematic review / [ S. Sajed, M. Inayat, A. Qadir et al.] // Appl. Soft Comput. – 2020. – Vol. 147.
8. Deep learning for chest radiograph diagnosis in the emergency department / [ E. J. Hwang, S. Y. Park, J. H. Kim et al.] // Radiology. – 2019. – Vol. 293, № 3. – P. 573–580.
9. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists / [P. Rajpurkar, J. Irvin, K. Zhu et al.] // PLoS Med. – 2018. – Vol. 15, № 11. – P. 1–17.
10. A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features / [A. T. Sahlol, M. Abdulkadir, A. A. Elaziz et al.] // Symmetry (Basel). – 2020. – Vol. 12, № 7.
11. Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation / [A. Majkowska, C. Pham, K. He et al.] // Radiology. – 2020. – Vol. 294, № 2. – P. 421–431.
12. Deep Learning in Chest Radiography: Detection of Pneumoconiosis / [ X. Li, Y. Wang, S. Xu et al.] // Biomed. Environ. Sci. – 2021. – Vol. 34, № 10. – P. 842–845.
13. Uçar M. Deep neural network model with Bayesian optimization for tuberculosis detection from X-Ray images / M. Uçar // Multimed. Tools Appl. Springer US. – 2023. – Vol. 82, № 24. – P. 36951–36972.
14. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / [J. Devlin, M. Chang, K. Lee et al.] // NAACL HLT 2019 – 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. – Proc. Conf. Association for Computational Linguistics (ACL). – 2018. – Vol. 1. – P. 4171–4186.
15. Ouis M. Y. ChestBioX-Gen: contextual biomedical report generation from chest X-ray images using BioGPT and co-attention mechanism / M. Y. Ouis, M. A. Akhloufi // Front. Imaging. – 2024. – Vol. 3.
16. ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases / [X. Wang, Y. Peng, L. Lu et al.] // Adv. Comput. Vis. Pattern Recognit. – 2019. – № September. – P. 369–392.
17. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison / [ J. Irvin, P. Rajpurkar, M. Ko et al.]// 33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019. – 2019. – P. 590–597.
18. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs / [ A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz et al.]. – 2019. –Vol. 14. – P. 1–7.
19. Nicolson A. Improving chest X-ray report generation by leveraging warm starting / [A. Nicolson, J. Dowling, B. Koopman] // Artif. Intell. Med. Elsevier B.V. – 2023. –Vol. 144, № April 2022. – P. 102633.
20. PadChest: A large chest x-ray image dataset with multi-label annotated reports / [A. Bustos, Y. Pertusa, J. M. Salinas et al.] // Med. Image Anal. – 2020. – Vol. 66. – P. 1–35.
21. Writing by memorizing: Hierarchical retrieval-based medical report generation / [X. Yang, J. Zhou, Y. Li et al.] // ACL-IJCNLP 2021 – 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf. – 2021. – P. 5000–5009.
22. Cross-modal memory networks for radiology report generation/ [Z. Chen, Y. Peng, X. Liu et al.] // ACL-IJCNLP 2021 – 59th Annu. Meet. Assoc. Comput. Linguist. 11th Int. Jt. Conf. Nat. Lang. Process. Proc. Conf. – 2021. – № 2018. – P. 5904–5914.
23. Kaur N., Mittal A. CheXPrune: sparse chest X-ray report generation model using multi-attention and one-shot global pruning // J. Ambient Intell. Humaniz. Comput. J Ambient Intell Humaniz Comput. – 2023. – Vol. 14, № 6. – P. 7485–7497.
24. Automatic Medical Report Generation Based on Cross-View Attention and Visual-Semantic Long Short Term Memorys / [Y. Gu, X. Li, L. Zhang et al.] // Bioengineering. Multidisciplinary Digital Publishing Institute (MDPI). –2023. – Vol. 10, № 8.
25. Radiology report generation with a learned knowledge base and multi-modal alignment/ [S. Yang, H. Yin, X. Wang et al.] // Med. Image Anal. – 2023. – Vol. 86.
26. XRaySwinGen: Automatic medical reporting for X-ray exams with multimodal model/ [G. Veras Magalhães, L. dos Santos, M. Ferreira et al. ] // Heliyon. Elsevier Ltd. – 2024. – Vol. 10, № 7.
27. Automated Chest X-Ray Diagnosis Report Generation with Cross-Attention Mechanism / [J. Zhao, Y. Chen, Q. Li et al.]// Appl. Sci. – 2025. – Vol. 15, № 1.
28. Singh P. ChestX-Transcribe: a multimodal transformer for automated radiology report generation from chest x-rays / P. Singh, S. Singh // Front. Digit. Heal. Frontiers Media SA. – 2025. – Vol. 7. – P. 1535168.


DOI 10.15588/1607-3274-2026-2-9