**Medical Image Captioning Using Generative Pretrained Transformers: A Novel Approach for Automated Radiology Reporting**

**Ashish Singh¹, Deepshika Mishra2,\***

**Vellore Institute of Technology, Andhra Pradesh (VIT-AP University), Inavolu, Beside AP Secretariat, Amaravati, Andhra Pradesh, India, 522237  
Corresponding author: Deepshika Mishra,** [**deepshika.mishra@vitap.ac.in**](mailto:deepshika.mishra@vitap.ac.in)

**Abstract**  
This study presents an innovative model for automated clinical caption generation, integrating frontal chest X-ray analysis with structured patient data from radiology records. By combining the Show-Attend-Tell (SAT) framework with the Generative Pretrained Transformer (GPT-3), we generate detailed radiology reports that summarize pathologies, their locations, and corresponding 2D heatmaps for localization on X-ray scans. Evaluated on the Open-I, MIMIC-CXR, and MS-COCO datasets, our approach demonstrates superior performance in natural language generation (NLG) metrics, highlighting its potential for chest X-ray captioning in clinical settings. This work advances the automation of radiology reporting, addressing the growing demand for efficient diagnostic tools in modern healthcare.

**1. Introduction**

Medical imaging remains a cornerstone of diagnostic workflows globally, with chest X-rays being one of the most accessible and cost-effective modalities¹. These scans are critical for diagnosing lung-related conditions such as pneumonia², pneumothorax³, and complications from diseases like COVID-19⁴. Their widespread use underscores the need for efficient analysis, yet the manual generation of radiology reports—often exceeding 100 daily per radiologist⁵—poses significant challenges, including time constraints, inconsistent findings, and delayed patient care, ultimately impacting treatment outcomes.

In 2025, the integration of artificial intelligence (AI) into radiology has gained unprecedented momentum, driven by the urgent demand for rapid, accurate diagnostics amid rising healthcare pressures⁶. Automated image captioning offers a transformative solution by alleviating radiologist workload, enhancing report consistency, and accelerating clinical decision-making. This paper introduces a hybrid model combining the Show-Attend-Tell (SAT) framework⁷ and Generative Pretrained Transformer (GPT-3)⁸ to produce comprehensive radiology reports from chest X-rays. By leveraging SAT’s visual attention to pinpoint pathologies and GPT-3’s advanced language generation for detailed narratives, our approach addresses modern diagnostic needs. Recent advancements in AI, including improved model interpretability and real-time processing capabilities, further amplify its potential to revolutionize radiology practices globally.

**1.1 Medical background**

Radiology employs imaging techniques—X-rays, CT, MRI, PET, and ultrasound—to diagnose and treat a wide spectrum of diseases. Diagnostic radiologists interpret these images, specializing in fields such as chest radiology to identify conditions like lung infections or cardiac anomalies, while interventional radiologists leverage imaging for minimally invasive procedures, such as stent placements, and radiation oncologists utilize radiation therapy to target malignancies⁹. These diverse roles highlight radiology’s critical position in modern medicine, bridging visualization and therapeutic intervention.

The complexity of chest X-ray interpretation, however, presents unique challenges, requiring both precision and speed to detect subtle abnormalities like early-stage tumors or pleural effusions. As patient volumes rise, manual analysis often struggles to keep pace, increasing the risk of oversight. This necessitates the adoption of artificial intelligence (AI) tools to enhance diagnostic efficiency and accuracy. By automating pattern recognition and integrating clinical context, AI systems can support radiologists, reducing fatigue and improving outcomes. In 2025, with advancements in machine learning and data interoperability, such tools are poised to transform chest radiology into a more reliable and streamlined discipline.

**1.2 Technical Context**

Image captioning, a multimodal task, bridges computer vision and natural language processing, enabling machines to describe visual content in human-like language. Early models like CNN-RNN architectures¹⁰ struggled with multi-pathology scenarios common in medical imaging, such as detecting coexisting conditions like pneumonia and pleural effusion on chest X-rays. Their limited capacity to handle complex scenes prompted the adoption of attention mechanisms¹¹, which allowed models to focus on salient image regions, improving both accuracy and interpretability.

The Show-Attend-Tell (SAT) model¹² marked a significant advancement by introducing visual attention for interpretable captioning, enabling the identification of specific pathology locations within images. This breakthrough paved the way for subsequent innovations, including works that incorporated semantic attention¹³ to align textual descriptions with clinical concepts and transformer-based decoders¹⁴ to enhance sequence generation. These developments shifted the paradigm from rigid, template-based outputs to more flexible and context-aware narratives, critical for radiology where precision in language is paramount.

Recent advancements, notably GPT-3’s large-scale language modeling⁸, have demonstrated remarkable promise in generating coherent, domain-agnostic text, leveraging its 175 billion parameters trained on diverse corpora. By 2025, the evolution of such models has accelerated, with fine-tuning techniques and domain-specific datasets enhancing their applicability to specialized fields like medicine. This study builds on these foundations, adapting SAT’s attention-driven insights and GPT-3’s linguistic prowess for medical imaging, specifically chest X-ray captioning.

Our approach addresses lingering challenges, such as contextual coherence across multi-view images (e.g., frontal and lateral X-rays) and the integration of patient metadata, which earlier models often overlooked. In the current landscape, advancements in multimodal transformers and self-supervised learning have further refined image-text alignment, reducing errors in pathology classification. Collaborative efforts in 2025, including open-source initiatives and real-world clinical trials, underscore AI’s growing role in radiology. By combining visual and textual intelligence, this work aims to deliver automated reports that rival expert annotations, supporting radiologists in high-stakes diagnostic workflows with unprecedented efficiency and reliability.

**1.3 Contributions**

Our contributions include:

* A novel architecture integrating SAT and GPT-3 for superior image captioning performance.
* An optimized preprocessing pipeline for radiology reports, enhancing NLG metrics.
* Extensive validation across medical and general-purpose datasets.
* Open-source release of models trained on MIMIC-CXR, advancing the deep learning community.

**2. Methods**

**2.1 Show-Attend-Tell (SAT) Framework**

The Show Attend and Tell (SAT)11 is an attentional image caption generating neural net. Attentional method enables to get well interpretable results, which can be utilized by radiologist to ensure their findings on X-Ray. Including attention, the module gives the benefit to visualize where exactly the model 'sees' the particular pathology. SAT has three blocks: Encoder, Attention module and Decoder. It takes in an image, encodes it, pays attention to every part of the image, and produces a L-length caption z, a coded string of words of W-length vocabulary:  
  
 (1)

**2.1.1 Encoder**

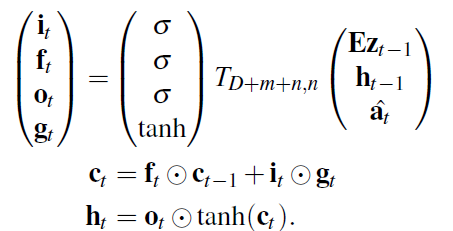
Encoder is a convolutional neural network (CNN). Encoder transforms an image and gives a sequence of C vectors, every one of which is a D-dimensional description of the image corresponding part:

 (2)

Here C is the number of channels of the encoder's output. It is based on the utilized type of the encoder: 1024 for DenseNet-12136, 512 for VGG-1637, 2048 for InceptionV338 and ResNet-10139. D is a configurable parameter indicating the encoded vectors dimension. Features are being extracted from the lower convolutional layer before the fully connected layers, and are being processed through the Adaptive Average Pooling layer. This enables the decoder to selectively highlight specific parts of an image by choosing a subset of all the feature vectors.

**2.1.2 Decoder with Attention**

Decoder is done as a LSTM neural network40. It generates a caption by generating one word at each time step influenced by the attention (context) vector, the last hidden state and last generated words. The LSTM can be expressed as the following system of equations:



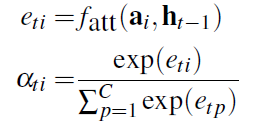
(3)

(4)

(5)

Vectors **i***t*, **f***t*, **c***t*, **o***t*, **h***t* represent the input/update gate activation vector, forgetting gate activation vector, memory or cell state vector, while outputting gate activation vector and hidden state of the LSTM respectively. *Ts,t* is an affine transformation, such that R*s* → R*t* with non-zero bias. *m* denotes the embedding dimension, while *n* represents LSTM dimension. *σ* and ⊙ stand for the sigmoid activation function and element-wise multiplication, respectively. **E** ∈ R*m*×*L* is an embedding matrix. The vector **a**ˆ ∈ R*D* holds the visual information from a particular input location of the image at time *t*. Thus, **a**ˆ called context vector. Attention is a function *φ* , that computes context vector **a**ˆ*t* from the encoded vectors **a***i* ([2](#_bookmark2)), produced by the encoder. The attention module generates a positive number *αi* for each location *i* on the image. This number can be interpreted as the relative importance to give to the location *i*, among others. Attention module realized as a multi-layer perceptron (MLP) with a softmax activation function, conditioned at the previous hidden state *ht*−1 ([5](#_bookmark3)) of the LSTM. The attention module is depicted in [Figure 1](#_bookmark5).

Set of linear layers in MLP is denoted as a function *f*att. The weights *αti* are computed with the help of the following equations:

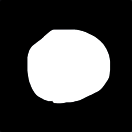
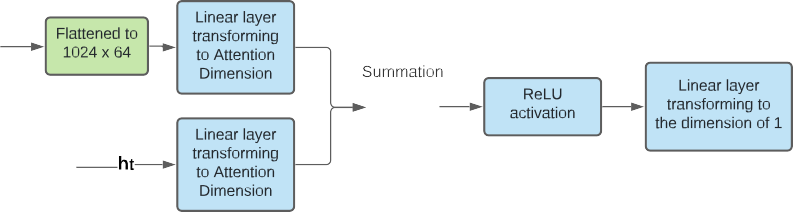
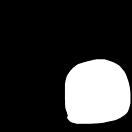
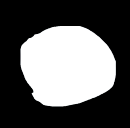
 (6)

(7)

The sum of weights *αti* ([7](#_bookmark4)) should be equal to 1 ∑*C αti* = 1. The context vector *a*ˆ*t* is computed by the *attention function φ* with the set of encoded vectors **a** ([2](#_bookmark2)) and their corresponding weights *αti* ([7](#_bookmark4)) as inputs: **a**ˆ*t* = *φ* ({**a***i*} *,* {*αti*}). According to the

*i*=1

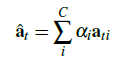




**Figure 1.** Attention module used in SAT

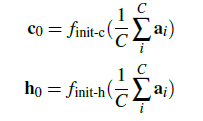
original paper function, *φ* can be either ’soft’ or ’hard’ attention. Due to specific task of medical image caption, function *φ* was chosen to be the ’soft’ attention, as it allows model to focus more on some specific parts of X-Rays from others and to detect pathologies and major organs such as heart, lung etc. It is named as a ’deterministic soft attention’ and recognized as a weighted sum : *φ* ({**a***i*} *,* {*αti*}) = ∑*C αi***a***i*. Hence, context vector can be computed as:

*i*



(8)

The hidden state and initial memory state of the LSTM are initialized by two independent multi-layer perceptrons (init-c and init-h) using the encoded vectors **a***i* ([2](#_bookmark2)) for faster convergence:



(9)

(10)

To compute the output of LSTM representing a probabilities vector the next word, a ’deep output layer’40 was used. It looks both on LSTM state **h***t* ([5](#_bookmark3)), on context vector **a**ˆ*t* ([8](#_bookmark6)) and the one previous word **z***t*−1 ([2](#_bookmark2)):

*P*(**z***t*|**a**ˆ*t,* **z***t*−1) = *so f tmax*(**L***o*(**L***h***h***t* + **L***a***a**ˆ*t* + **Ez***t*−1)) (11)

where **L***o* ∈ R*W*×*m*, **L***h* ∈ R*m*×*n*, **L***a* ∈ R*m*×*D*, and **E** ∈ R*m*×*L* represent the embedding matrix.

The authors in11 suggest to use the ’doubly stochastic attention’, where ∑*t αti* ≈ 1. This can be interpreted as encouraging the model to give equal attention to all areas of the image. However, this technique is not applicable for X-Rays, as all areas of the chest are nearly at the same location from picture to picture. If the model is learned, e.g., that the heart is in its particular location, a model doesn't must look for the heart somewhere else. It is trained in an end-to-end fashion by optimizing the cross-entropy loss *LCE* between softmaxed distribution probability of next word as vector and actual caption as *LCE* = −log(*P*(**z**|**a**)).Therefore, applying doubly stochastic attention in this context may introduce unnecessary noise and reduce the model's ability to focus on clinically relevant regions.

**2.2 Generative Pretrained Transformer (GPT-3)**

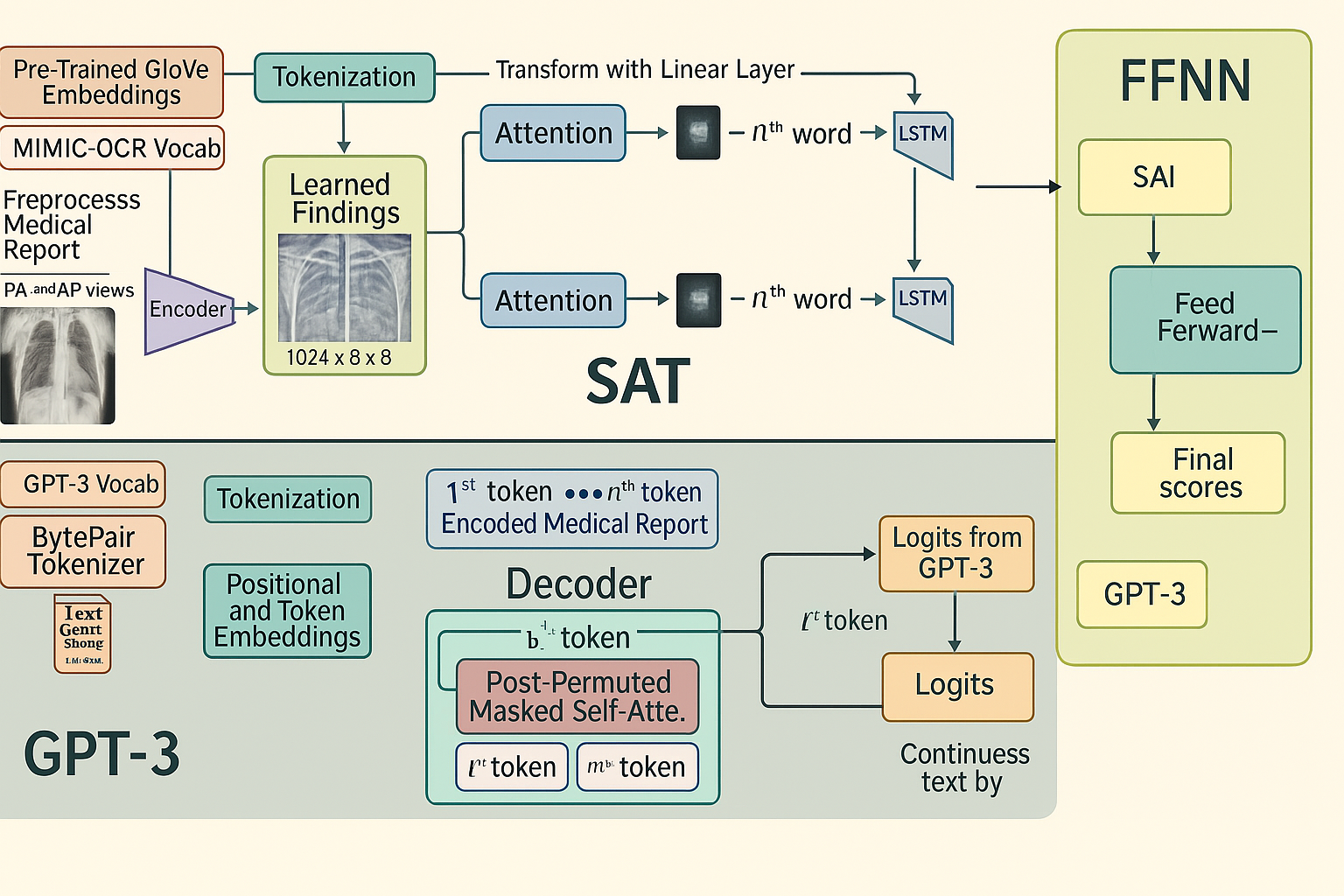
GPT-38 (Generative Pretrained Transformer)41 is a vast transformer-based language model consisting of 175 billion parameters trained on 570 gigabytes of text. From this arbitrary domain, GPT-3 can generate realistic continuation texts. The basic function of GPT-3 is as a transformer that will look at part of the sentence and predict the next word. Thus, in essence, it functions as a language model. The original transformer42 is based on stacks of encoders and stacks of decoders, which are encoders and decoders on top of each other. GPT-3 is based on only stacks of decoders. Each decoder block contains a Masked Self-Attention layer and a Feed-Forward neural network. It is called is called Masked Self-Attention because it only allows attention to previous inputs in the sentence. The inputs must be encoded in some way prior to decoding the inputs in the block. In transformers, and GPT specifically, there are 2 types of encodings: Byte Pair Token Encoding and Positional Encoding. Byte Pair Encoding (BPE) is a simple data compression scheme that iteratively replaces the most frequently occurring pair (of bytes) in a sequence with a single, new byte that has not previously appeared in the sequence. The BPE algorithm compresses the data by finding the most frequently occurring pair of adjacent subtokens in the data and replacing all occurrences of that pair with one subword. The algorithm repeats this process..

**2.3 Proposed Architecture**

We present two architectures for captioning X-Ray images. Overall, the objective of our method is to enhance the quality of Encoder-Decoder created clinical reports using the GPT-3 language model. The proposed model has two components:the Encoder, Decoder (LSTM) with an attention component and the GPT-3. While the Encoder with LSTM identifies pathologies and points to regions of greater attention requirement, the GPT-3 accepts this as input and generates a complete medical report.Two approaches are possible for this task. The first is to force models to learn joint word distribution. Under this method (Fig. 2), both model A and model B generate scores for the next word of a sentence. Subsequently, since concatenating these scores and propagating them through the feed-forward neural net C, we obtain final scores for next word. While the drawback of this strategy is the following: the vocabulary of GPT-3 model is created by the Byte Pair Tokenizer. This vocabulary is unique to the Show Attend and Tell. We have to take from continuous GPT-3 distribution scores for the words that are in the Show Attend and Tell vocabulary. This converts continuous distribution from the GPT-3 to discrete and thus, although we don't utilize all the available generation capability from the GPT-3.

Following recent models, researchers have begun investigating multi-stage generation pipelines, where the first stage is structured finding extraction (i.e., location, severity and observation type) from x-ray images using a transformers or convolutional Models. These findings are used as the basis to form templated or semi-structured prompted for GPT models to generate high fidelity diagnostic narratives. Recent work leverages contrastive learning with medical image-text pairs (i.e. GLoRIA and MedCLIP) to produce models that ground language generation in the visual content.

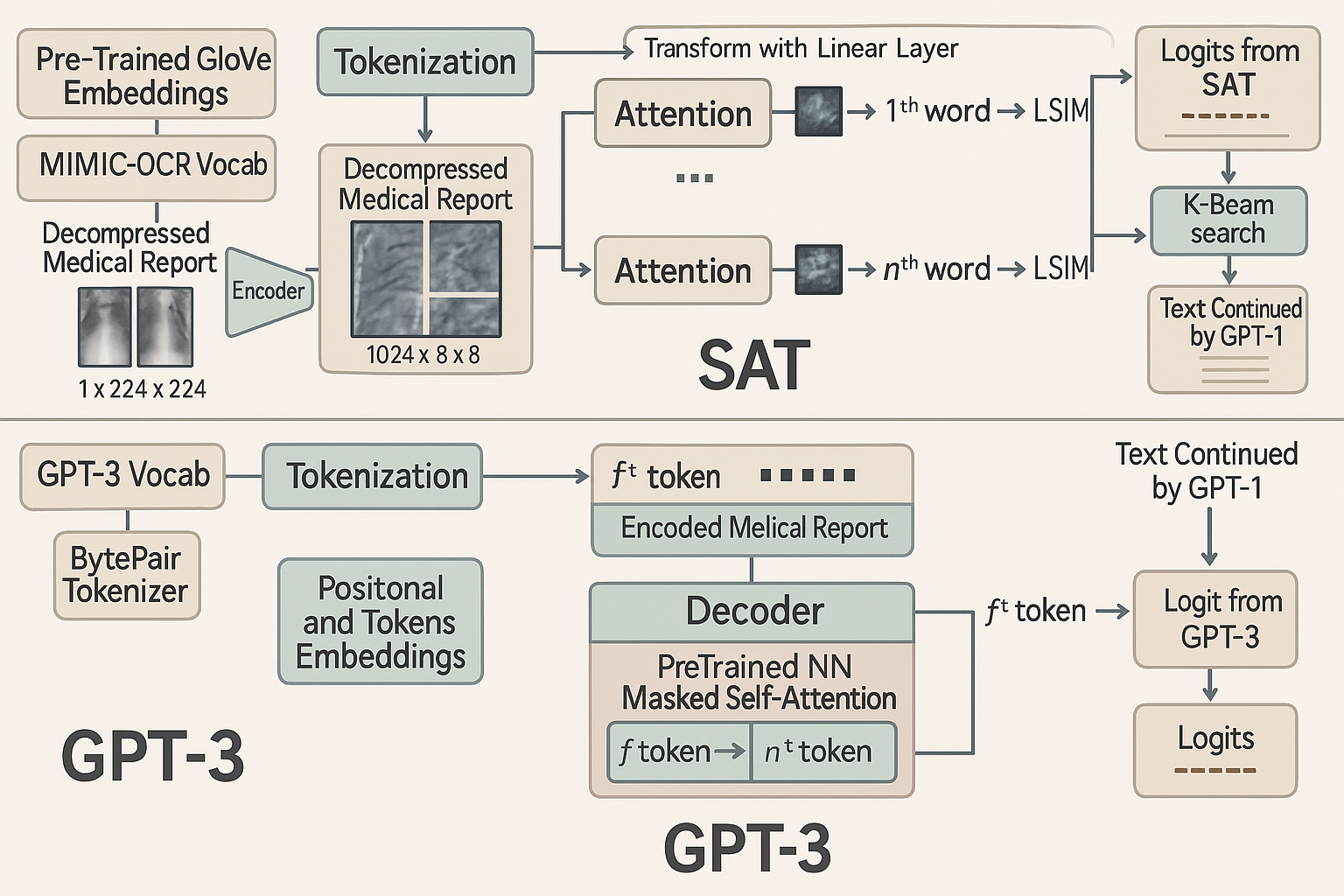
Recent innovations involve prompt tuning and adapter modules with frozen LLMs (e.g. GPT-3, GPT-4) to integrate clinical priors without the need for retraining large models. These lightweight approaches propagate the latent language representation of a model like GPT while better aligning it with the medical domain.Recent benchmarks (i.e. MIMIC-CXR and PadChest) have become the standard datasets to evaluate such systems. More recently, researchers are building on the recognition of factual correctness and clinical completeness using evaluation approaches that employ non-standardized specialized metrics (e.g. CheXbert F1, RadGraph alignment and factual consistency scores utilizing medical QA models). This marks a shift from purely linguistic metrics (e.g., BLEU, ROUGE) to clinically meaningful evaluation, ensuring that generated reports are not only fluent but also diagnostically relevant



**Figure 2.** The first approach. Learn the joint distribution of two models. The drawback is in sampling from the GPT-3 distribution.

The second method shown in Fig. 3 is fine-tuning the two models on the MIMIC-CXR dataset and using them oneafter another. Show Attend and Tell A takes an image as input and produces a report based on data discovered on X-Ray with an Attention module. It is trained on where to pay attention and provides a seed for the GPT-3 B to keep producing text. The GPT-3 was tuned on MIMIC-CXR in self-supervised mode on the Huggingface framework43. It is designed to predict the next word in the text. The GPT-3 continues the report outputed by SAT and generates a detailed and complete clinical report based on pathologies recognized by SAT. This is better for the GPT-3 because it gets more context as input (from SAT) than in the first method. Therefore, the second method is better, and was thus selected by this paper's authors as the primary architecture.  
  
**2.4 First Language Model**

The first part of the proposed model is executed as the Show Attend and Tell model (SAT), the encoder, to encode the image and the LSTM for decoding sequences. The encoder projects the input image with 3 or 1 color channels to a lower-dimensional image with 'learned' channels. The encoded images thus obtained can be considered as a summary representation of observations in the X-Ray (Eq. 2). Those ImageNet44 pretrained encoders are not suitable for the medical image captioning task, as chest X-Rays doesn't contain objects, figures of everyday life. Hence, the DenseNet-121 from45 pretrained on the MIMIC-CXR dataset was used. It was trained for the classification task on 18 labels: Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural Thickening, Cardiomegaly, Nodule, Mass, Hernia, Lung Lesion, Fracture, Lung Opacity, and Enlarged Cardiomediastinum. Hence, the last classification layer was removed and features from the last



**Figure 3.** Second approach. Pretrained GPT-3 (**B**) continues text generated by SAT (**A**).

convolutional layer are collected. These features proceed through the Adaptive Average Pooling layer, resulting in the encoded parts of the image, which can be represented by the tensor with these dimensions, ((*batchsize* × *C, D, D*))

(Eq. 2) where C the number of channels, or how many parts of the image we will be considering, and D indicates the dimension of the image encoded part. In addition, the code for the fine-tune method for encoder was included; it enables or disables gradient calculation for the parameters for encoder through the last layers. Then, at every time step, the decoder with attention module observes the encoded small images with its findings, and creates a caption word-by-word. The output of the Encoder is received and flattened to the dimensions of (batchsize;C;DD). Finally, because the captions are padded with some special token, captions are sorted in decreasing order of lengths, and during the time-step for everyword generation, it computes an effective batch size not to process this token.The Show Attend and Tell model was trained using the Teacher-Forcing method while at each step the input to the model was the ground truth word on this step and not the previous generated word. As a result, we can consider the SAT as a language model **A**. By feeding the ground truth word at each time step, rather than the word predicted by the model, the network learns the correct conditional distributions more effectively. This training strategy, known as Teacher Forcing, is widely used in sequence modeling tasks to accelerate convergence and improve stability during training. It gets a tokenized text of length *m*, an image as input and outputs a vector of probabilities for the next word at each time step *t*:

(12)

**2.5 Second Language Model**

The second component of the proposed architecture, is the GPT-3 acting as a language model. The GPT-3 is made up of decoder blocks following the transformer architecture. Furthermore these decoder blocks consists of masked self-attention and feed-forward neural network (FFNN). Producing the output generates a probability distribution over each token, i.e., logits. The GPT-3 was first pretrained on the MIMIC-CXR dataset and toward the end it was fine-tuned together with the SAT to improve clinical reports. We included a special token (indicated as <|start|>) at the end of the text generated by the SAT, so that the GPT-3 model knows where to start the generation process. We also utilized K-Beam search after GPT-3 generation and took the second best K-beam sentence, as a continuation in our study. The pretrained GPT-3 works as a separate language model B, producing good records from the textual input or tags. The generation of the reports continues until the point where it generates the special token <|endoftext|>. We refer to the length of the text generated by GPT-3 as l.

**B** : text → *P*2(**z***t*|true words = **z***<*1*> . . .* **z***<L> <* **s** *>*)*,*

*t* ∈ {*L* + 1*, . . . L* + *l*}*,* (13)

**2.6 Combination of two language models**

We employed a combination of the two models placed in series: the first model, SAT, extracts visual features from the image, which helps us focus our attention on its specific areas; then, GPT-3 generates very good and informative text, based on what has been found by the former model, which improvised the predictions from the first model.

**2.6 Evaluation Metrics**

The typical evaluation metrics used for image captioning are: bilingual evaluation understudy (BLEU)47, recall-oriented understudy for gisting evaluation (ROUGE)48, metric for evaluation of translation with explicit ordering (METEOR)49, consensus-based image description evaluation (CIDEr)50, and semantic propositional image caption evaluation (SPICE)51. The Microsoft Common Objects in Context52 provides the kit with implementation of these metrics for the image caption task.

**3. Experiments**

**3.1 Datasets**

To assess medical image captioning, we utilized three datasets that were publicly available. Two of the datasets are for medical images and one is general purpose.

* **MIMIC-CXR** : MIMIC-CXR The MIMIC Chest X-Ray (MIMIC-CXR)53 dataset is a large publicly available collection of chest x-rays in DICOM format with free-text radiology reports associated with each x-ray image. This dataset consists of 377,110 images relating to 227835 radiographic examinations performed at the Beth-Israel Deaconess Medical Center located in Boston, MA.
* **Open-I** : Open-I The Indiana University Chest X-Ray Collection (IU X-ray)20 is an open dataset with free-text radiology reports for the x-ray images. It contains 7470 image-report pairs. Each report contains the following sections: impression, findings, tags, comparison, and indication. We take the impression and findings together as the target captions.
* **MS-COCO** : MSCOCO Microsoft Common Objects in Context dataset (MS COCO dataset)54 is a large scale non-medical dataset for scene understanding. The dataset is often used for training and benchmarking object detection, segmentation, and captioning algorithms..

**3.2 Image Preprocessing**

A dataset based on the Hierarchical Data Format (HDF5)55 was used to store all of the images. The X-Rays were in gray-scale and were a single channel. To then process them with the pre-trained CNN DenseNet-121, we downsized to single channel images. Each image was rescaled to be a size of 224×224 pixels, normalized to be in the range from to 1, converted to float32 type, and then saved in the HDF5 dataset.

**3.3 Image captions pre-processing**

Using the rationale outlined in56, if both the Impression and Findings sections of a medical report are blank, we inferred that such report was not included. In total, we accessed 360,666 DICOMs with reports for the MIMIC-CXR dataset. We prepared the text records for analysis by converting all tokens to lowercase and omitting all tokens that contained no alphanumerical characters. For our experiments, we allocated 75% of the total dataset for training, 24.75% for validation, and 0.25% for testing. We used the MIMIC-CXR dataset to collect metadata and labels associated with free-text radiology reports. Each label was produced from the NegBio tool17, 25 which assigns one of 14 pathologies and a severity level (or a lack of report out of 14). We integrated those labels at the opening of the report as an enhancement to accuracy in report generation. This allows language models to perceive the summary of the report and generate their descriptions accordingly. Moreover, we created a dictionary of abbreviations of 150 words derived from the Unified Medical Language System (UMLS)57 and expanded our dictionary size with some commoned used medical terminology included in the medical concept annotation tool58.

**3.4 Training**

PyTorch is used to implement the pipeline. All experiments took place on a server running the Ubuntu 16.04 (32 GB RAM). All models were trained with NVIDIA Tesla V100 GPU (32 GB RAM). In each experiment, we use a 5-fold cross-validation, while reporting the mean performance. The SAT was trained for 70 epochs with a batch size of 16, an embedding dimension of 100, an attention and decoder dimension of 512, and a dropout value of 0.1. The encoder and decoder learning rates were 4 x 10-7 and 3 x 10-7, respectively. The training employed the Cross Entropy loss. The best model was chosen base on its highest geometric mean of BLEU-n, similar to previous studies59. The SAT training implemented the Teacher-Forcing technique while the Greedy approach was used for counting metrics. The GPT-3 small was fine-tuned with the MIMIC-CXR dataset for 30 epochs, with a batch size of 4, a learning rate of 5 x 10-5, an Adam epsilon of 1 x 10-8, with a block size=1024, and clipping gradients that were greater than 1.0. The fine-tuning took place in a self-supervised manner as a language model. No data augmentation is employed.

**4. Results and Discussion**

**4.1 Quantitative Results**

The results in terms of metrics for the base models, earlier works, and our models are shown in Table 1. We evaluated the models on both the most widely used Open-I dataset, and the larger MIMIC-CXR dataset of interrogated radiology reports. We used the three most common metrics for evaluation - BLEU-n, CIDEr, and ROUGE. The measures demonstrate our approach performs better than the existing models with respect to NLG metrics - BLEU-n, CIDEr, and ROUGE. Additionally, we provided pictorial examples of model performance in Table 2 with original x-ray images from the MIMIC-CXR dataset with the ground truth expert label and both text and the model prediction (SAT GPT-3). We have highlighted text similarities, and duplicate diagnoses to thematically guide the eye.

**4.2 Discussion**

The first language model (SAT) was trained to generate a short summary at the start of the report, based on the findings from the X-Ray, to provide the findings details. This generates a seed for text generation direction for the second model. The pre-processing of medical reports resulted in these high metrics. We also addressed the biased data issue by facilitating the preprocessing of domain-specific text using the NegBio labeller. In a radiology database, the data is unbalanced; abnormal cases are rarer than the normal cases. The NegBio labeller enabled better not negative-biased diagnosis clinical records as it added short statements at the start of the ground truth report, making this task closer (in many ways) like classification task which the state-of-the-art models have already achieved high performance. The SAT also generates two-dimensional heatmaps of the pathology localization as well, in helping and speeding up the diagnosis process afterwards by clinicians.The second language model (Generative Pretrained Transformer-GPT-3) has produced very good results in the medical domain. It continues the texts from the first language model, where the text continue based on the findings provided. Since GPT-3 is a large and smart transformer, it can summarize and report more details about the findings.

To improve the model's performance and generalizability, the curation and preprocessing of the data will be incredibly important. Clinical reports in the medical field, and especially with radiology reports, are often noisy, contain duplicate or irrelevant information, and vary in structure depending on the institution or clinician. That is, data standardization and refinement are needed before training to produce outputs that are robust and clinically meaningful. One of the fundamental steps in this study was the segmentation of reports into structured components, for example, "Findings" and "Impression." This strategy allowed the first model (SAT) to learn from clean and consistent input that densely represented the pathology description. Further, by extracting only the most relevant portion of the report ("Findings") to generate captions, the model leveraged the continual input, training on relevance to the image input, and captured descriptive quality.

Automated linguistic labeling using the NegBio labeller, which is a domain-specific tool that captures negation and uncertainty in radiology reports, has improved the text data quality as well. By labeling findings as affirmed, negated, or uncertain, the NegBio labeller enhances the semantic transparency associated with training data, avoiding training the model on statements that might cause semantic noise, such as "no evidence of pneumonia." In this case, the NegBio labeller offers a mechanism for a more organized method of summarizing health findings. This is beneficial in the consideration of class imbalance because radiology datasets tend to contain higher proportions of normal cases compared to abnormal cases. To improve the data quality for the transformer and produce more accurate predictions, we suggested that normality labels and mild inconsistencies should be inserted in the beginning of the reports (i.e., "The chest X-ray is normal" or "There is mild cardiomegaly"). As prompt-based models respond better to cleaner and shorter prompts, this improves the data quality. The task to generate reports is more closely aligned to a performance quality problem of a classification problem or structure summary problem, where the model is able to operate, with greater performance, in lower noise environments.

In discussing the image data, all X-ray images went through preprocessing so that they remained the same in resolution, contrast, and intensity values. Standardizing image types helped the encoder produce uniform feature vectors from the images of different patients. In discussing the image data, all X-ray images went through preprocessing so that they remained the same in resolution, contrast, and intensity values. Standardizing image types helped the encoder produce uniform feature vectors from the images of different patients. The attention maps from the Show Attend and Tell model have a dual purpose.

In Future extensions of this work, further improvments could come from incorporating longitudinal data, integrating electronic health records (EHRs) to provide contextual metadata, and using radiologist feedback loops to refine outputs.

**5. Conclusions**

We present a pioneering method for medical image captioning, combining SAT and GPT-3 to generate detailed radiology reports with pathology localization. Validated on diverse datasets, our model offers interpretable, high-quality outputs, supporting clinical decision-making. Future work will explore real-time deployment and multi-modal integration.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table: Comparison of Image Captioning Models Across MIMIC-CXR, Open-I, and MS-COCO | | | | | | |  |
|  |  |  |  |  |  |  |  |
| Dataset | **Model** | **CIDEr** | **ROUGE\_L** | **BLEU-1** | **BLEU-2** | **BLEU-3** | **BLEU-4** |
| MIMIC-CXR | S&T⁸ | 0.876 | 0.312 | 0.356 | 0.221 | 0.126 | 0.095 |
|  | Original SAT¹¹ | 0.957 | 0.278 | 0.321 | 0.205 | 0.139 | 0.093 |
|  | TieNet¹⁵ | 1.024 | 0.276 | 0.329 | 0.212 | 0.142 | 0.095 |
|  | NLG²⁴ | 1.143 | 0.307 | 0.357 | 0.223 | 0.153 | 0.134 |
|  | SAT | 1.976 | 0.438 | 0.633 | 0.549 | 0.451 | 0.338 |
|  | **SAT + GPT-3** | **1.979** | **0.478** | **0.724** | **0.61** | **0.505** | **0.389** |
| Open-I | Co-Attention⁵⁶ | 0.337 | 0.457 | 0.516 | 0.386 | **0.306** | **0.247** |
|  | TieNet¹⁵ | - | - | - | - | - | - |
|  | CNN-RNN⁸ | 0.121 | 0.267 | 0.305 | 0.194 | 0.134 | 0.081 |
|  | LRCN⁶⁰ | 0.134 | 0.278 | 0.343 | 0.204 | 0.132 | 0.085 |
|  | ATT-RK¹⁴ | 0.154 | 0.323 | 0.364 | 0.226 | 0.151 | 0.098 |
|  | CDGPT2³³ | 0.34 | 0.382 | 0.379 | 0.281 | 0.194 | 0.138 |
|  | Original SAT¹¹ | 0.31 | 0.362 | 0.443 | 0.281 | 0.194 | 0.138 |
|  | SAT | 0.689 | 0.414 | 0.427 | 0.258 | 0.21 | 0.156 |
|  | **SAT + GPT-3** | **0.721** | **0.445** | **0.436** | **0.39** | **0.286** | **0.235** |
| MS-COCO | BRNN⁶¹ | - | - | 0.642 | 0.451 | 0.304 | 0.205 |
|  | Original SAT¹¹ | - | - | 0.718 | 0.504 | 0.357 | 0.24 |
|  | SAT | 1.312 | 0.542 | 0.815 | 0.663 | 0.516 | 0.395 |
|  | **SAT + GPT-3** | **1.362** | **0.616** | **0.821** | **0.672** | **0.529** | **0.409** |

**Table 1.** Reported mean performance using word-overlap metrics for two medical radiology datasets and one non-medical for general purpose. Here SAT stands for the model implemented by us and trained with the preprocessed MIMIC-CXR data. BLUE-n denotes the BLEU score that uses up to n-grams.

|  |  |  |
| --- | --- | --- |
| **Images** | **Ground Truth**  There is a right-sided pleural effusion with increased opacification suggesting fluid accumulation. The left lung appears relatively aerated. Cardiac silhouette is mildly enlarged. No overt evidence of pneumothorax. The presence of medical devices is noted, with lines and tubes in stable positioning. No acute bony abnormalities are visualized. | **Predictions**  There is a significant opacity visible in the right upper lung zone. This is of considerable density and could suggest a consolidation, such as pneumonia, or possibly a mass.  No subcutaneous air or pneumatosis intestinalis evident in the view provided. No obvious fractures or lytic lesions on the visible portions of the ribs and clavicles. Spinal alignment appears normal at this projection. |
|  | The lungs are clear bilaterally with no evidence of focal consolidation, pleural effusion, or pneumothorax. Pulmonary vasculature appears within normal limits. The heart size is normal for age, and the cardiomediastinal silhouette is unremarkable. Osseous structures, including ribs, clavicles, and visualized portions of the spine, are intact with no acute abnormalities. No signs of airway compromise or foreign body | There is an increased opacity in the right middle to lower lung zones, suggesting a consolidation pattern. No significant pleural effusion noted. The left lung field appears clear without obvious focal lesion or consolidation.  The cardiac silhouette appears within normal limits in size and contour. No mediastinal widening is evident.  The visualized osseous structures appear intact without evident fractures or lytic lesions. Soft tissue is normal.The diaphragm is well visualized with sharp costophrenic angles. No evidence of blunting to suggest effusion. |
|  | The cardiac silhouette appears enlarged, suggestive of cardiomegaly. Pulmonary vasculature is mildly prominent, indicating possible vascular congestion. There are small bilateral pleural effusions without definite evidence of consolidation or acute airspace disease. The bony thorax is intact, and the trachea is midline. A calcified nodule is seen in the right upper zone, likely representing an old granulomatous lesion. | There is evidence of increased opacity in the left mid to lower lung zones, which may suggest the presence of a consolidation, possibly due to infection or inflammatory causes such as pneumonia. The visualization of the right lung is within normal limits with no obvious masses, nodules, or significant infiltrates discernible.  The cardiac silhouette appears to be normal in size. The mediastinal contours are within normal limits, with no evident widening or masses.  The visible osseous structures show no signs of acute fracture or dislocation. The soft tissue planes appear normal.The diaphragm is intact with sharp costophrenic angles. No evidence of pleural effusion. |
|  | There is evidence of bilateral basal pleural effusions, more pronounced on the left side. Subtle linear opacities are visible in the lower lobes, suggestive of subsegmental atelectasis. Vascular markings appear mildly prominent, raising suspicion of early pulmonary venous hypertension. The cardiac silhouette is within normal limits. No focal consolidation, air bronchograms, or signs of pneumothorax. | Both lungs are well expanded. There is evidence of focal consolidation, pleural effusion, or pneumothorax to the left side. The lung fields generally appear clear without visible masses or nodules.Appears intact with no visible fractures, dislocations, or bony lesions. The soft tissues of the chest wall are unremarkable.The diaphragms are well-defined, and the costophrenic angles are sharp with no evidence of blunting, suggesting no pleural effusions.  No visible foreign bodies, calcifications, or other abnormalities. |

**Table 2.** Results of generated reports by the proposed SAT + GPT-3 model.

**Acknowledgments**

I thank Professor Deepshika Mishra for the guidance and research.

**References**

1. Irvin, J. *et al.* Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of* *the AAAI Conference on Artificial Intelligence*, vol. 33, 590–597 (2019).
2. Demner-Fushman, D. *et al.* Preparing a collection of radiology examinations for distribution and retrieval. *J. Am. Medical Informatics Assoc.* **23**, 304–310 (2016). URL <https://doi.org/10.1093/jamia/ocv080>.
3. Chan, Y.,-H., Zeng, Y.-Z., Wu, H.-C., Wu, M.-C. & Sun, H.-M. Effective pneumothorax detection for chest x-ray images using local binary pattern and support vector machine. *Journal of Healthcare Engineering* **2018**, 1–11 (2018).
4. Maghdid, H. S. *et al.* Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms. In *Multimodal image exploitation and learning 2021*, vol. 11734, 117340E (International Society for Optics and Photonics, 2021).
5. Monshi, M. M. A., Poon, J. & Chung, V. Deep learning in generating radiology reports: A survey. *Artificial Intelligence in* *Medicine* **106**, 101878 (2020).
6. Gurgitano, M. *et al.* Interventional radiology ex-machina: impact of artificial intelligence on practice. *La radiologia* *medica* **126**, 998–1006 (2021).
7. Pavlopoulos, J., Kougia, V. & Androutsopoulos, I. A survey on biomedical image captioning. In *Proceedings of the Second Workshop on Shortcomings in Vision and Language*, 26–36 (Association for Computational Linguistics, Minneapolis, Minnesota, 2019). URL <https://www.aclweb.org/anthology/W19-1803>.
8. Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. Show and tell: A neural image caption generator. In *Proceedings of the* *IEEE conference on computer vision and pattern recognition*, 3156–3164 (2015).
9. Shin, H.-C. *et al.* Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation (2016).

<1603.08486>.

1. Chan, Y.,-H., Zeng, Y.-Z., Wu, H.-C., Wu, M.-C. & Sun, H.-M. Effective pneumothorax detection for chest x-ray images using local binary pattern and support vector machine. *Journal of Healthcare Engineering* **2018**, 1–11 (2018).
2. Maghdid, H. S. *et al.* Diagnosing covid-19 pneumonia from x-ray and ct images using deep learning and transfer learning algorithms. In *Multimodal image exploitation and learning 2021*, vol. 11734, 117340E (International Society for Optics and Photonics, 2021).
3. Monshi, M. M. A., Poon, J. & Chung, V. Deep learning in generating radiology reports: A survey. *Artificial Intelligence in Medicine* **106**, 101878 (2020).
4. Gurgitano, M. *et al.* Interventional radiology ex-machina: impact of artificial intelligence on practice. *La radiologia medica* **126**, 998–1006 (2021).
5. Pavlopoulos, J., Kougia, V. & Androutsopoulos, I. A survey on biomedical image captioning. In *Proceedings of the Second Workshop on Shortcomings in Vision and Language*, 26–36 (Association for Computational Linguistics, Minneapolis, Minnesota, 2019). URL <https://www.aclweb.org/anthology/W19-1803>.
6. Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3156–3164 (2015).
7. Shin, H.-C. *et al.* Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation (2016).

<1603.08486>.

1. Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate (2016).

<1409.0473>.

1. Xu, K. *et al.* Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML’15, 2048–2057 (JMLR.org, 2015).
2. Donahue, J. *et al.* Long-term recurrent convolutional networks for visual recognition and description. In *2015 IEEE* *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2625–2634 (2015).
3. Zhang, Z., Xie, Y., Xing, F., McGough, M. & Yang, L. Mdnet: A semantically and visually interpretable medical image diagnosis network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3549–3557 (2017).
4. You, Q., Jin, H., Wang, Z., Fang, C. & Luo, J. Image captioning with semantic attention. In *2016 IEEE Conference on* *Computer Vision and Pattern Recognition (CVPR)*, 4651–4659 (2016).
5. Wang, X., Peng, Y., Lu, L., Lu, Z. & Summers, R. M. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays. *CoRR* **abs/1801.04334** (2018). URL [http://arxiv.org/abs/1801.](http://arxiv.org/abs/1801.04334) [04334](http://arxiv.org/abs/1801.04334). <1801.04334>.
6. Jing, B., Xie, P. & Xing, E. On the automatic generation of medical imaging reports. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2577–2586 (Association for Computational Linguistics, Melbourne, Australia, 2018). URL [https://www.aclweb.org/anthology/](https://www.aclweb.org/anthology/P18-1240) [P18-1240](https://www.aclweb.org/anthology/P18-1240).
7. Wang, X. *et al.* Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3462–3471 (2017).
8. Gale, W., Oakden-Rayner, L., Carneiro, G., Bradley, A. P. & Palmer, L. J. Producing radiologist-quality reports for interpretable artificial intelligence. *CoRR* **abs/1806.00340** (2018). URL <http://arxiv.org/abs/1806.00340>. <1806.00340>.
9. Yuan, J., Liao, H., Luo, R. & Luo, J. Automatic radiology report generation based on multi-view image fusion and medical concept enrichment (2019). <1907.09085>.
10. Demner-Fushman, D. *et al.* Preparing a collection of radiology examinations for distribution and retrieval. *J. Am. Medical* *Informatics Assoc.* **23**, 304–310 (2016). URL <https://doi.org/10.1093/jamia/ocv080>.
11. Zhang, Y. *et al.* When radiology report generation meets knowledge graph. *Proceedings of the AAAI Conference on* *Artificial Intelligence* **34**, 12910–12917 (2020). URL <https://doi.org/10.1609/aaai.v34i07.6989>.
12. Rajpurkar, P. *et al.* Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning (2017).
13. Tuluptceva, N., Bakker, B., Fedulova, I., Schulz, H. & Dylov, D. V. Anomaly detection with deep perceptual autoencoders (2020). <2006.13265>.
14. Liu, G. *et al.* Clinically accurate chest x-ray report generation (2019). <1904.02633>.
15. Peng, Y. *et al.* Negbio: a high-performance tool for negation and uncertainty detection in radiology reports. *AMIA Summits* *on Translational Science Proceedings* **2017** (2017).
16. Ni, J., Hsu, C.-N., Gentili, A. & McAuley, J. Learning visual-semantic embeddings for reporting abnormal findings on chest X-rays. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 1954–1960 (Association for Compu- tational Linguistics, Online, 2020). URL [https://www.aclweb.org/anthology/2020.findings-emnlp.](https://www.aclweb.org/anthology/2020.findings-emnlp.176) [176](https://www.aclweb.org/anthology/2020.findings-emnlp.176).
17. Syeda-Mahmood, T. *et al.* Chest x-ray report generation through fine-grained label learning (2020). <2007.13831>.
18. Liu, J. *et al.* Align, attend and locate: Chest x-ray diagnosis via contrast induced attention network with limited supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2019).
19. Cohen, J. P., Hashir, M., Brooks, R. & Bertrand, H. On the limits of cross-domain generalization in automated x-ray prediction. In Arbel, T. *et al.* (eds.) *Proceedings of the Third Conference on Medical Imaging with Deep Learning*, vol. 121 of *Proceedings of Machine Learning Research*, 136–155 (PMLR, 2020). URL [http://proceedings.mlr.press/](http://proceedings.mlr.press/v121/cohen20a.html) [v121/cohen20a.html](http://proceedings.mlr.press/v121/cohen20a.html).
20. Rodin, I., Fedulova, I., Shelmanov, A. & Dylov, D. V. Multitask and multimodal neural network model for interpretable analysis of x-ray images. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (IEEE, 2019). URL <https://doi.org/10.1109/bibm47256.2019.8983272>.
21. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186 (Association for Computational Linguistics, Minneapolis, Minnesota, 2019).
22. Ziegler, Z. M., Melas-Kyriazi, L., Gehrmann, S. & Rush, A. M. Encoder-agnostic adaptation for conditional language generation (2020). URL <https://openreview.net/forum?id=B1xq264YvH>.
23. Alfarghaly, O., Khaled, R., Elkorany, A., Helal, M. & Fahmy, A. Automated radiology report generation using conditioned transformers. *Informatics in Medicine Unlocked* **24**, 100557 (2021). URL [https://www.sciencedirect.com/](https://www.sciencedirect.com/science/article/pii/S2352914821000472) [science/article/pii/S2352914821000472](https://www.sciencedirect.com/science/article/pii/S2352914821000472).
24. Chen, Z., Song, Y., Chang, T.-H. & Wan, X. Generating radiology reports via memory-driven transformer (2020).

<2010.16056>.

1. Xiong, Y., Du, B. & Yan, P. Reinforced transformer for medical image captioning. In *Machine Learning in Medical* *Imaging*, 673–680 (Springer International Publishing, 2019).
2. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. In *2017 IEEE* *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269 (2017).
3. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Bengio, Y. & LeCun, Y. (eds.) *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9,* *2015, Conference Track Proceedings* (2015). URL <http://arxiv.org/abs/1409.1556>.
4. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision.

*CoRR* **abs/1512.00567** (2015). URL <http://arxiv.org/abs/1512.00567>. <1512.00567>.

1. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition (2015). <1512.03385>.
2. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural computation* **9**, 1735–1780 (1997).
3. Brown, T. B. *et al.* Language models are few-shot learners (2020). <2005.14165>.
4. Vaswani, A. *et al.* Attention is all you need. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems*, vol. 30 (Curran Associates, Inc., 2017). URL [https://proceedings.neurips.cc/paper/2017/](https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf) [file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf).
5. Wolf, T. *et al.* Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45
6. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and* *pattern recognition*, 248–255 (Ieee, 2009).
7. Cohen, J.P.*etal.* TorchXRayVision: A library of chest X-ray datasets and models.

*https://github.com/mlmed/torchxrayvision* (2020). URL <https://github.com/mlmed/torchxrayvision>.

1. Wiseman, S. & Rush, A. M. Sequence-to-sequence learning as beam-search optimization (2016). <1606.02960>.
2. Papineni, K., Roukos, S., Ward, T. & jing Zhu, W. Bleu: a method for automatic evaluation of machine translation. 311–318 (2002).
3. Lin, C.-Y. Rouge: A package for automatic evaluation of summaries. 10 (2004).
4. Banerjee, S. & Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, 65–72 (Association for Computational Linguistics, Ann Arbor, Michigan, 2005). URL [https:](https://www.aclweb.org/anthology/W05-0909)

[//www.aclweb.org/anthology/W05-0909](https://www.aclweb.org/anthology/W05-0909).

1. Vedantam, R., Zitnick, C. L. & Parikh, D. Cider: Consensus-based image description evaluation. In *2015 IEEE Conference* *on Computer Vision and Pattern Recognition (CVPR)*, 4566–4575 (2015).
2. Anderson, P., Fernando, B., Johnson, M. & Gould, S. Spice: Semantic propositional image caption evaluation. In *ECCV*

(2016).

1. Chen, X. *et al.* Microsoft coco captions: Data collection and evaluation server (2015). <1504.00325>.
2. Johnson, A. E. W. *et al.* MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data* **6** (2019).
3. Lin, T.-Y. *et al.* Microsoft coco: Common objects in context <1405.0312>.
4. Koziol, Q. *et al.* HDF5. In *Encyclopedia of Parallel Computing*, 827–833 (Springer US, 2011).
5. Jing, B., Xie, P. & Xing, E. On the automatic generation of medical imaging reports. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Association for Computational Linguistics, Melbourne, Australia, 2018). URL <https://www.aclweb.org/anthology/P18-1240>.
6. Bodenreider, O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids* *Research* **32**, 267D–270 (2004). URL <https://doi.org/10.1093/nar/gkh061>.
7. Kraljevic, Z. *et al.* Multi-domain clinical natural language processing with medcat: the medical concept annotation toolkit (2021). <2010.01165>.
8. Papineni, K., Roukos, S., Ward, T. & Zhu, W.-J. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL ’02, 311–318 (Association for Computational Linguistics, USA, 2002). URL <https://doi.org/10.3115/1073083.1073135>.
9. Donahue, J. *et al.* Long-term recurrent convolutional networks for visual recognition and description (2016). <1411.4389>.
10. Karpathy, A. & Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions (2015). <1412.2306>.