Attention Based Automated Radiological Report Generation Using Multimodel Architecture

ADVANCED COLLEGE
OF ENGINEERING & MANAGEMENT

Prepared by:

Ankit Chhetri 34459

Bhaskar Subedi 34469

Biplov Belbase 34477

Under The Supervision of :

Er. Utsav Pokhrel

OVERVIEW

- Introduction
- Problem Statement
- Project Objective
- Significance of Study
- System Requirements
- Literature Review
- Methodology
- Results, Analysis and Conclusion
- Limitations and Future Enhancements
- References

INTRODUCTION

ADVANCED COLLEGE
OF ENGINEERING & MANAGEMENT

"Transforming Radiology Through AI"

- Automated Report Generation
- Attention-Based Multimodel
- Image-Text Integration
- Efficiency & Workload Reduction
- Radiologist Diagnostic Support

PROBLEM STATEMENT

- **Improve Access:** Increase chest X-ray facilities in rural Nepal for timely diagnosis.
- **Use Technology:** Leverage AI to analyze X-ray images where radiologists are scarce.

PROJECT OBJECTIVE

- Using AI for enhancing chest X-ray reports in Nepal can speed up diagnosis and treatment delays, leading to better patient outcomes.

SIGNIFICANCE OF STUDY

- Enhance Accessibility
- Faster Diagnosis
- Specialist Support
- Better Outcomes
- Reduce Burden
- Transform Healthcare

SYSTEM REQUIREMENTS

3.1 Functional Requirements

- Image Input
- Image Preprocessing
- Feature Extraction
- Text Generation
- Report Formatting

3.2 Non-Functional Requirements

- Performance
- Security
- Usability
- Training and Support

3.3 Hardware Requirements

1. **GPUs**: High-end GPUs for training deep learning models.
2. **Memory**: Minimum 8 GB RAM to manage datasets and support GPU operation.
3. **Storage**: At least 256 GB of SSD storage to store datasets and model checkpoints efficiently.

3.4 Software Requirements

1. **Operating System**: Windows
2. **Deep Learning Frameworks**:
   - **Tensorflow and Keras**: Deep learning framework.
3. **Libraries and Dependencies**:
   - **NumPy**: For numerical operations.
   - **Pandas**: For data manipulation and analysis.
   - **NLTK**: For natural language processing tasks.

4. **Pre-trained Models and Tokenizers**:
    - **CheXNet** is based on DenseNet-121 but customized for medical image analysis
5. **CUDA**: For GPU acceleration.
6. **Jupyter Notebooks**: For interactive development and testing.

LITERATURE REVIEW

- Medical report generation process proposed a CNN-RNN architecture to generate captions for images whose results were too simple and lacked details[1,2,3].
- As more work was done, attention was introduced with model’s attention with RNN and CNN [4].
- CNNs were shown to be capable of classifying view orientations of chest radiographs with excellent accuracy [5,6,7].

LITERATURE REVIEW

- The attention mechanism enhances neural networks by focusing on key input features and sequential understanding for radiological report generation task. This boosts both accuracy and interpretability in medical imaging.[8]

- Jing et al. developed a multi-task framework with co-attention and hierarchical LSTM to predict tags, localize abnormalities, and generate radiology reports. They tested it on IU CXR and PEIR Gross dataset.[9]

LITERATURE REVIEW

- Cho et al. introduced the Gated Recurrent Unit (GRU) as a simpler alternative to LSTMs, addressing the vanishing gradient problem in RNNs and enabling better learning of long-term dependencies.[10]

- Bahdanau et al. introduced the additive attention mechanism, allowing the decoder to dynamically focus on relevant input parts, improving long-sequence handling through context vectors.[11]

METHODOLOGY

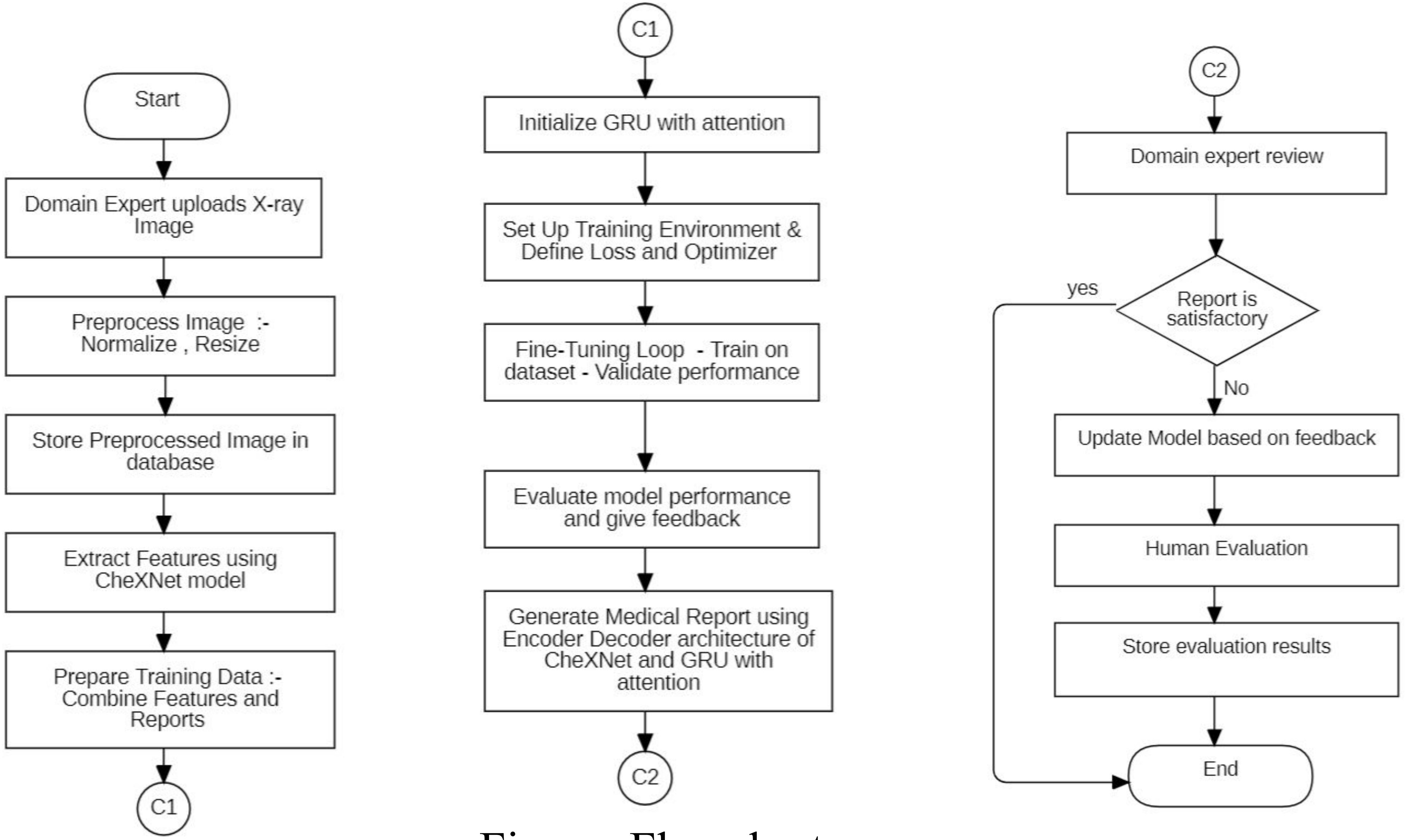

Figure: Flowchart

METHODOLOGY

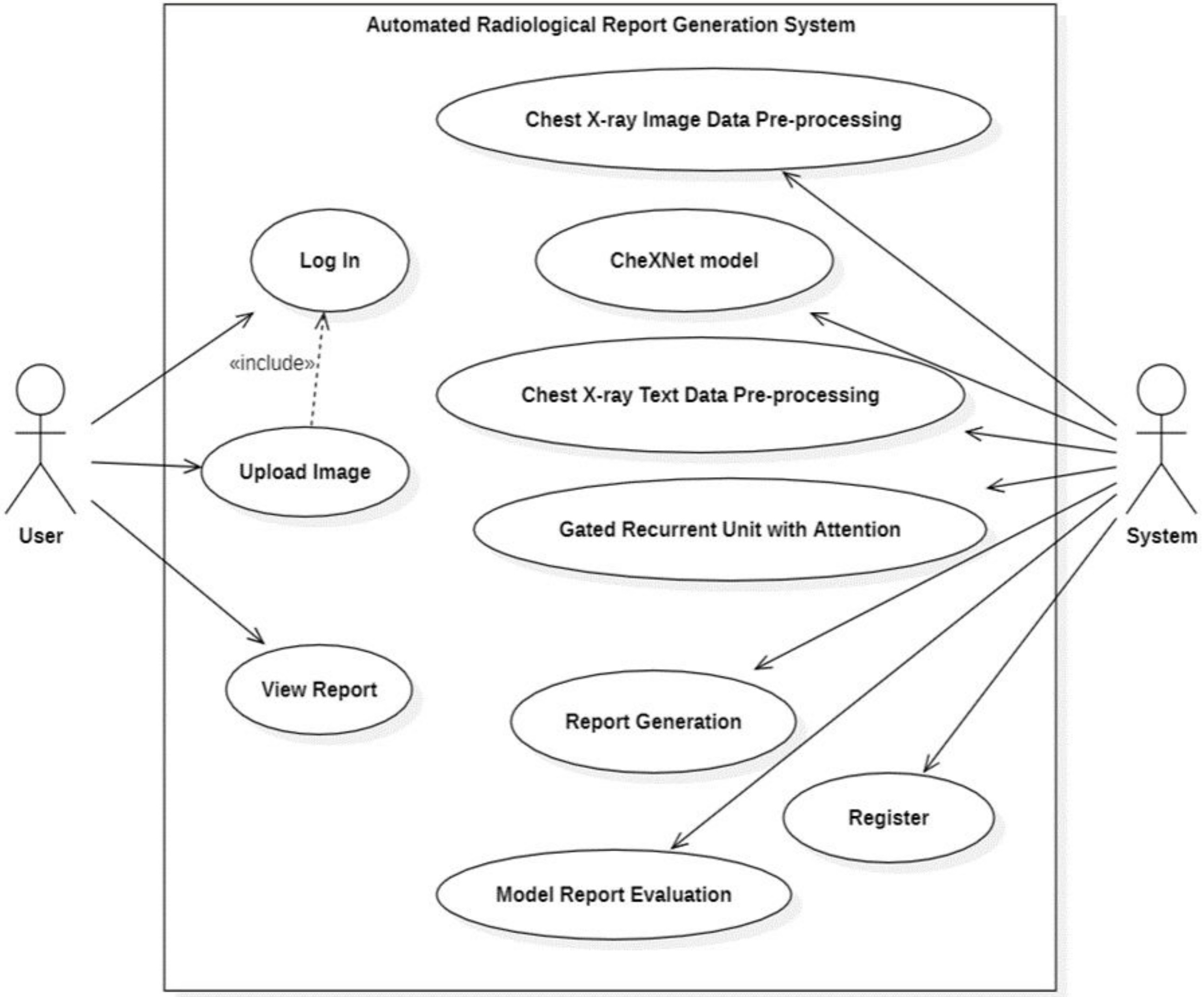

Figure: Use Case Diagram

METHODOLOGY

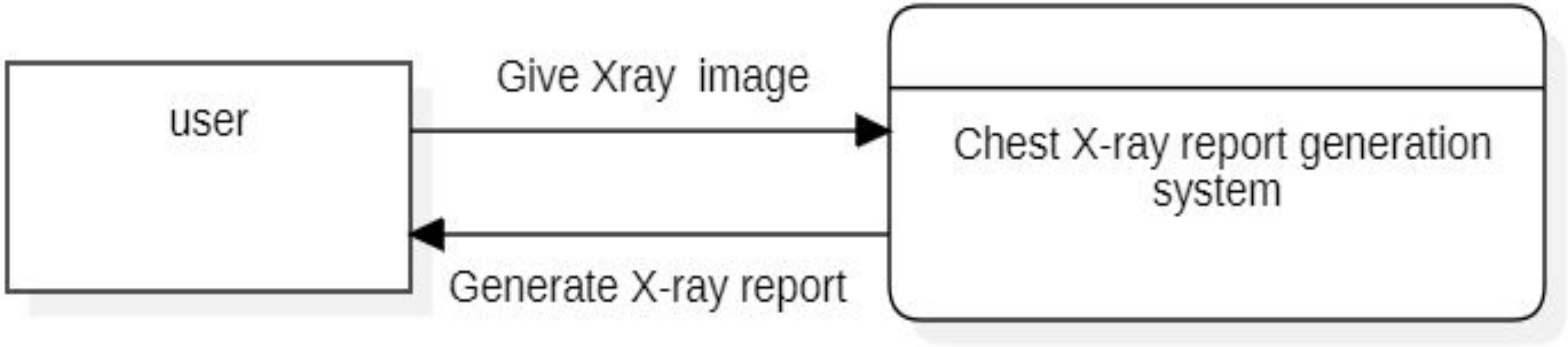

Figure: DFD level 0

METHODOLOGY

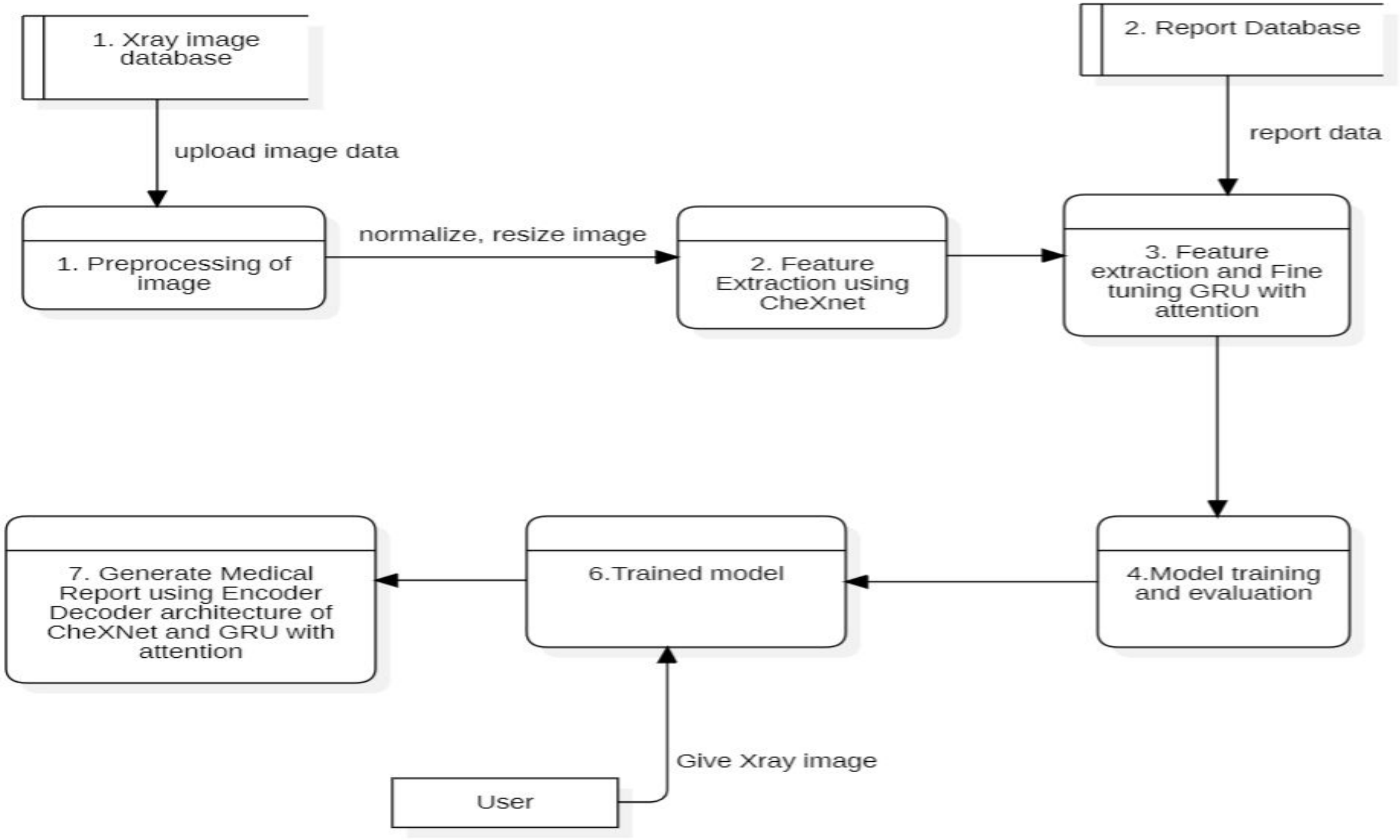

Figure: DFD level 1

METHODOLOGY

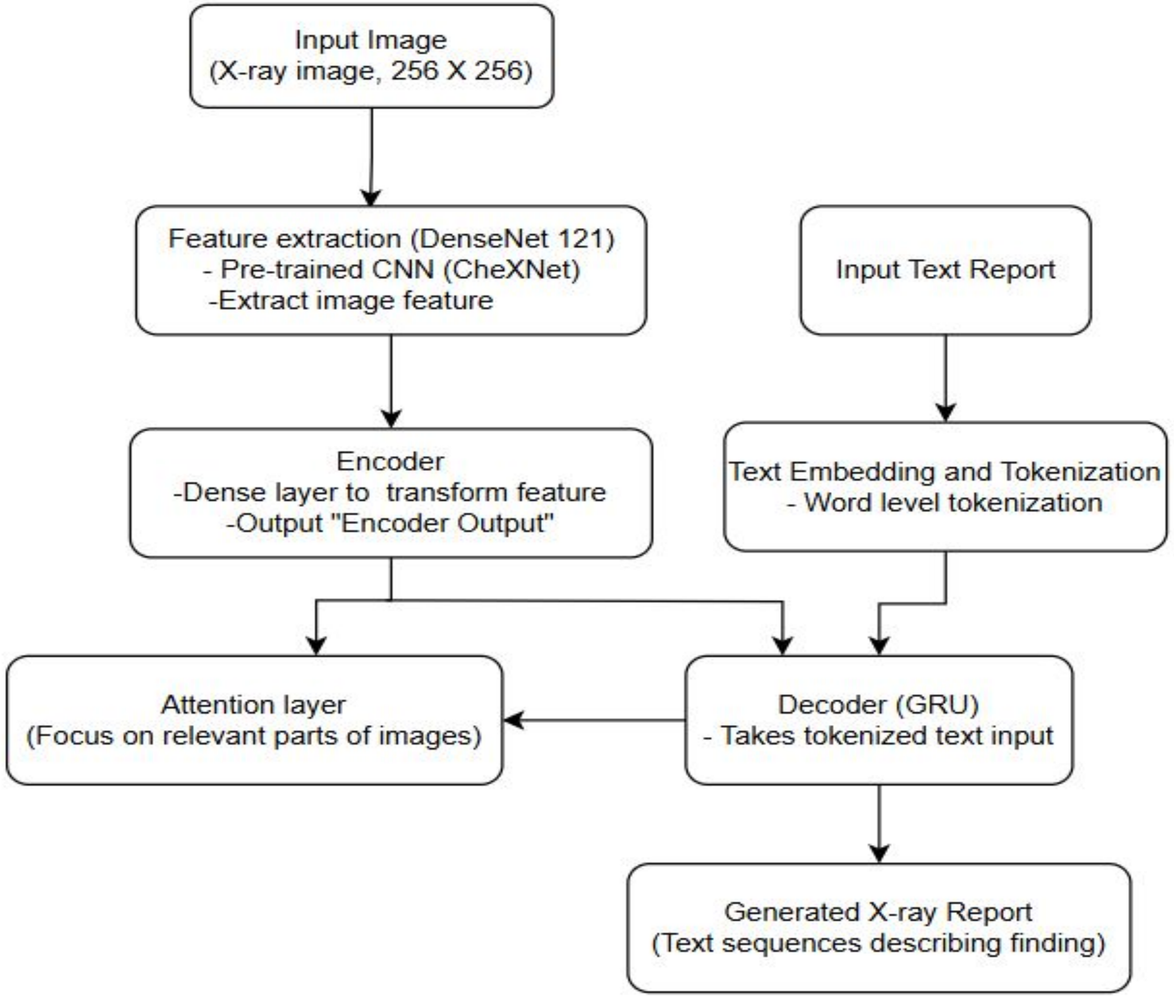

Figure: System Architecture

Data Collection

Data Source: Indiana University (X-ray images and radiology reports).

Chest X-ray: 7,471 .png images (front and lateral views).

Radiology Reports: 3,955 patient reports in .XML format.

Image-Report Pairing: Each image has four captions (Comparison, Indication, Findings, Impressions).

Goal: Predict the findings (key medical information) from the chest X-ray images.

Data Preprocessing

XML Parsing & Data Points:

- XML files containing patient information (image IDs, captions like comparison, indication, findings, impression) were parsed into csv file.
- The "findings" section was extracted as the report text, and image IDs were used to link reports to corresponding X-rays.

EDA and Data Preprocessing:

- Text data was cleaned (removing tags, punctuation, numbers, performing decontractions).
- Empty image names were handled by dropping rows.
- Word counts were calculated for text columns.
- Empty text entries were replaced with "No Impression".
- The final dataset contained 3851 rows.

Structured Data (Image Pairing):

Patients had varying numbers of X-rays (0-5). To create consistent input, each report was paired with two images. The strategy was:

- 5 images: Create 4 data points (1st+5th, 2nd+5th, 3rd+5th, 4th+5th).
- 4 images: Create 3 data points (1st+4th, 2nd+4th, 3rd+4th).
- 3 images: Create 2 data points (1st+3rd, 2nd+3rd).
- 2 images: Use as it is.
- 1 image: Duplicate the image.

Baseline Model

- **Encoder-Decoder Architecture:**
  - A sequence-to-sequence model was used.
  - The encoder processes image features into a context vector, which the decoder uses to generate the report text.
- **Add Tokens:**
  - <start> and <end> tokens were added to the text report.
- **Tokenization:**
  - Text was converted to numerical tokens using a word-level tokenization.

- **Encoder-Decoder Details:**
    - The encoder used a dense layer and dropout.
    - In the decoder part, an embedding layer, a dropout layer, and an LSTM layer are included.
    - Encoder and decoder outputs were combined using an "Add" layer, followed by a time-distributed dense layer.

- **Model Inference (Baseline):**
    - Greedy search (selecting the most probable word at each step) was used for generating reports.

Main Model (with Attention)

- **Input:** Image vectors and report text .
- **Encoder:** Same as the baseline model.
- **Additive Attention:** Calculates attention weights (alpha) for each word in the input sequence, creating a context vector.
- **Decoder:** Uses a GRU.

- **Decoder (One-Step Decoder):** Takes decoder input, encoder output, and state value. Uses an embedding layer, attention layer (to produce the context vector), and a GRU.
- **Loss Function:** Sparse Categorical Cross-entropy.
- **Model Inference (Main Model):** Beam search (keeping track of the top K most likely sequences) was used for report generation.

- **Model Evaluation:**
    - **Metrics:** BLEU score (precision of generated text).
    - **Validation:** A separate validation set was used.

RESULT ANALYSIS AND CONCLUSION

Result and Analysis :

- The initial encoder-decoder model had a low BLEU-1 score of 0.106, indicating poor report accuracy.
- Adding an attention mechanism helped the model focus on important areas of medical images.
- Replacing the decoder with a GRU improved text coherence and generation, increasing BLEU-1 score to 0.3787.
- Beam search enhanced the capability of model to generate more accurate report during report generation.

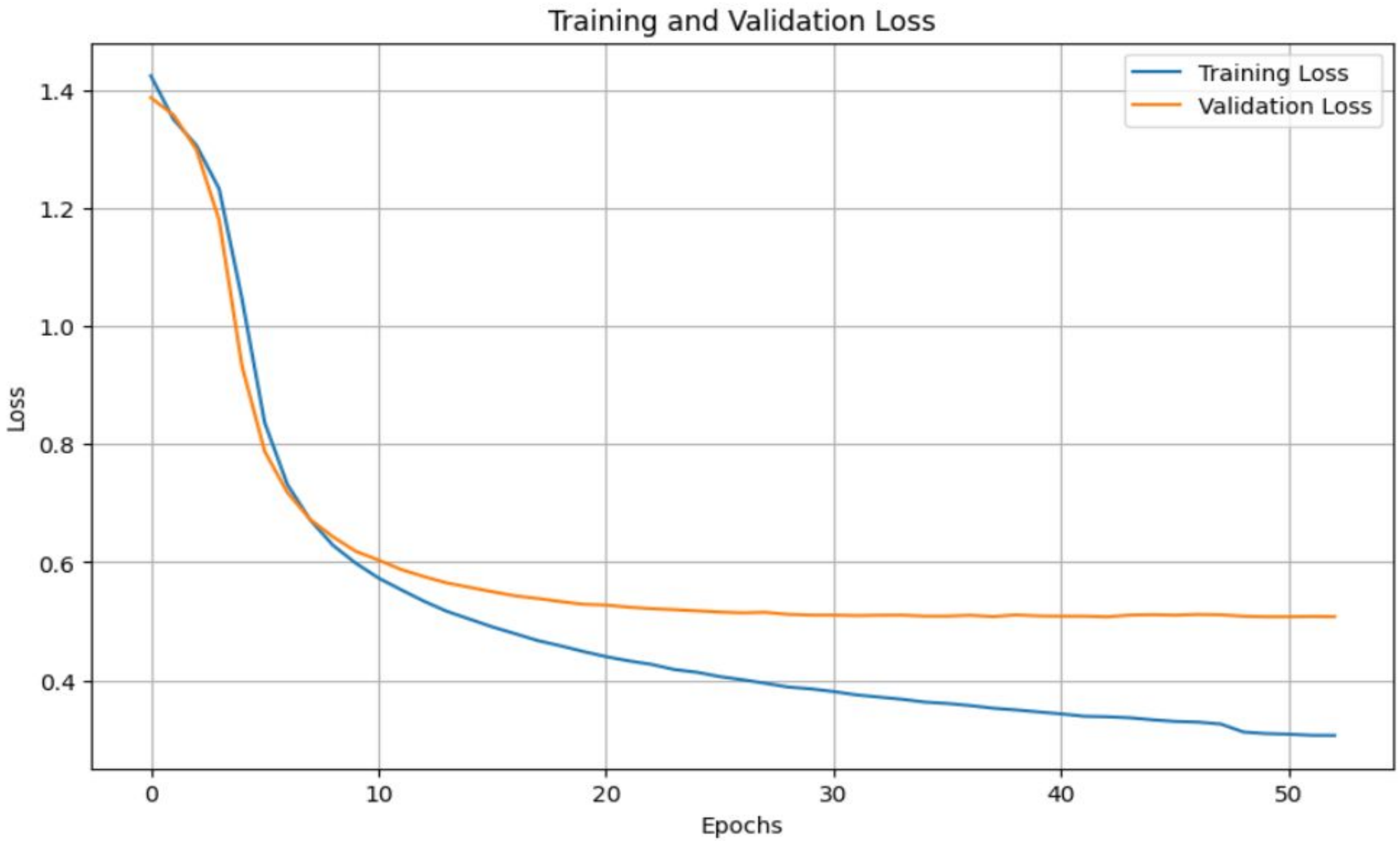

Fig: Training and Validation Loss

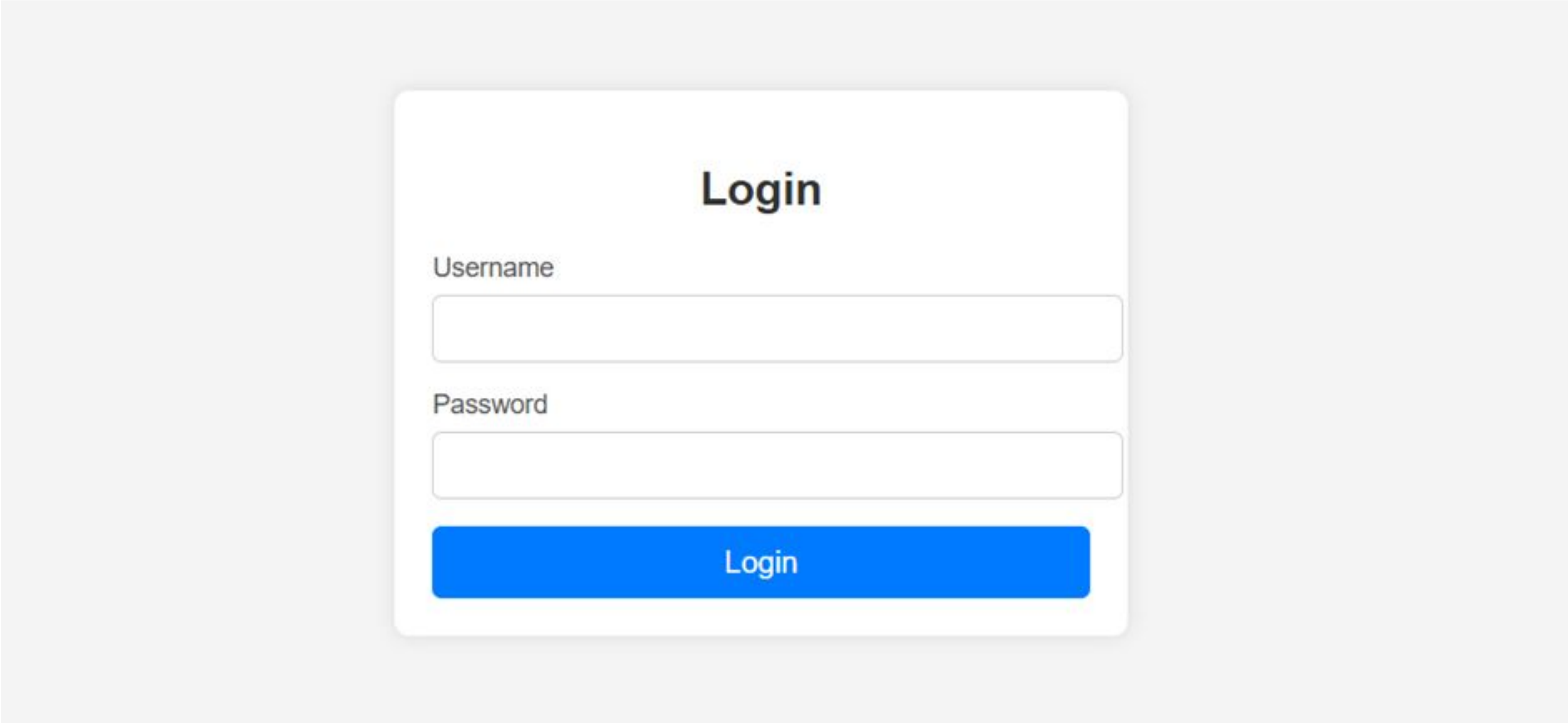

Figure : Login Page

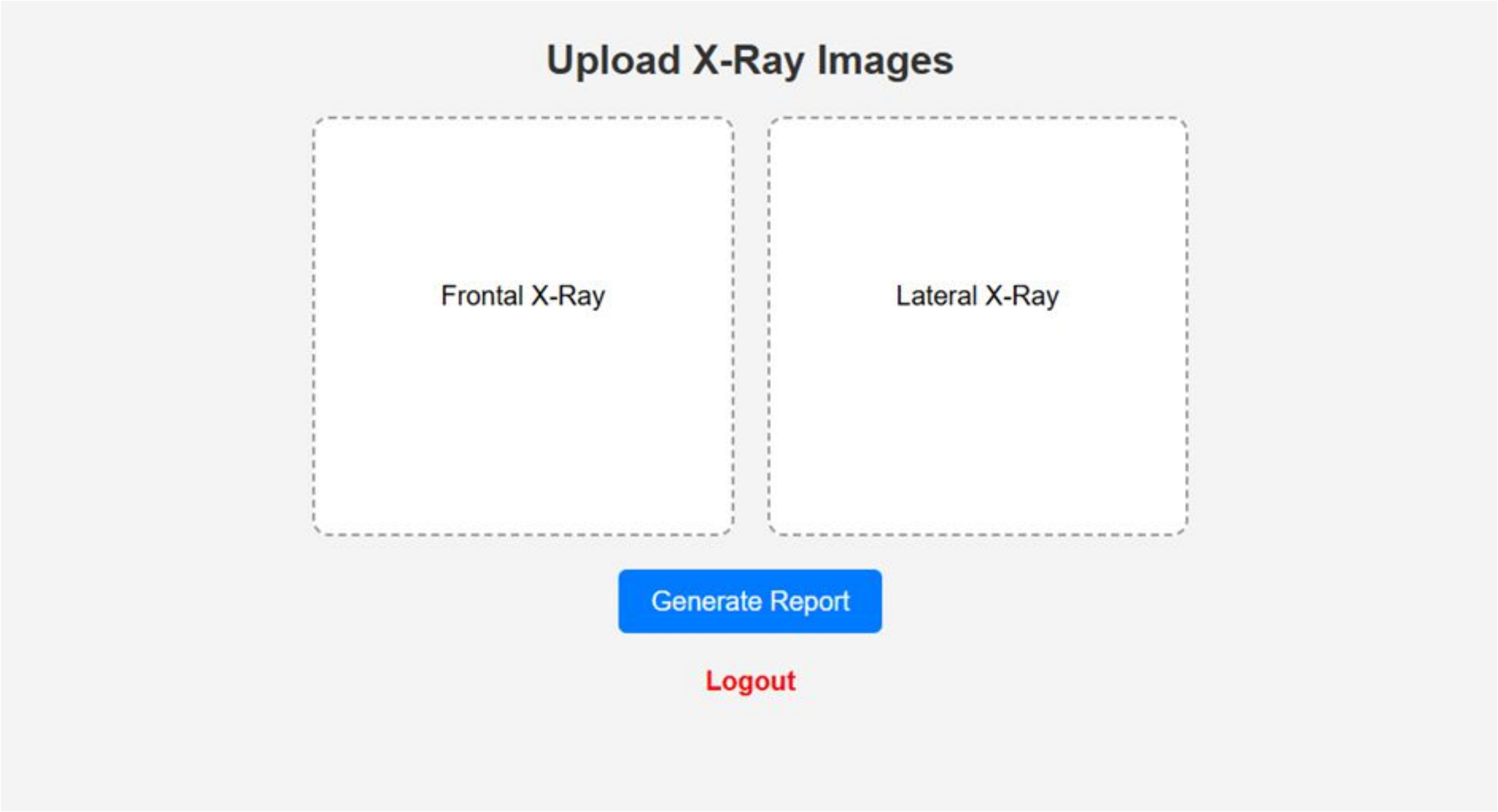

Figure : Home Page

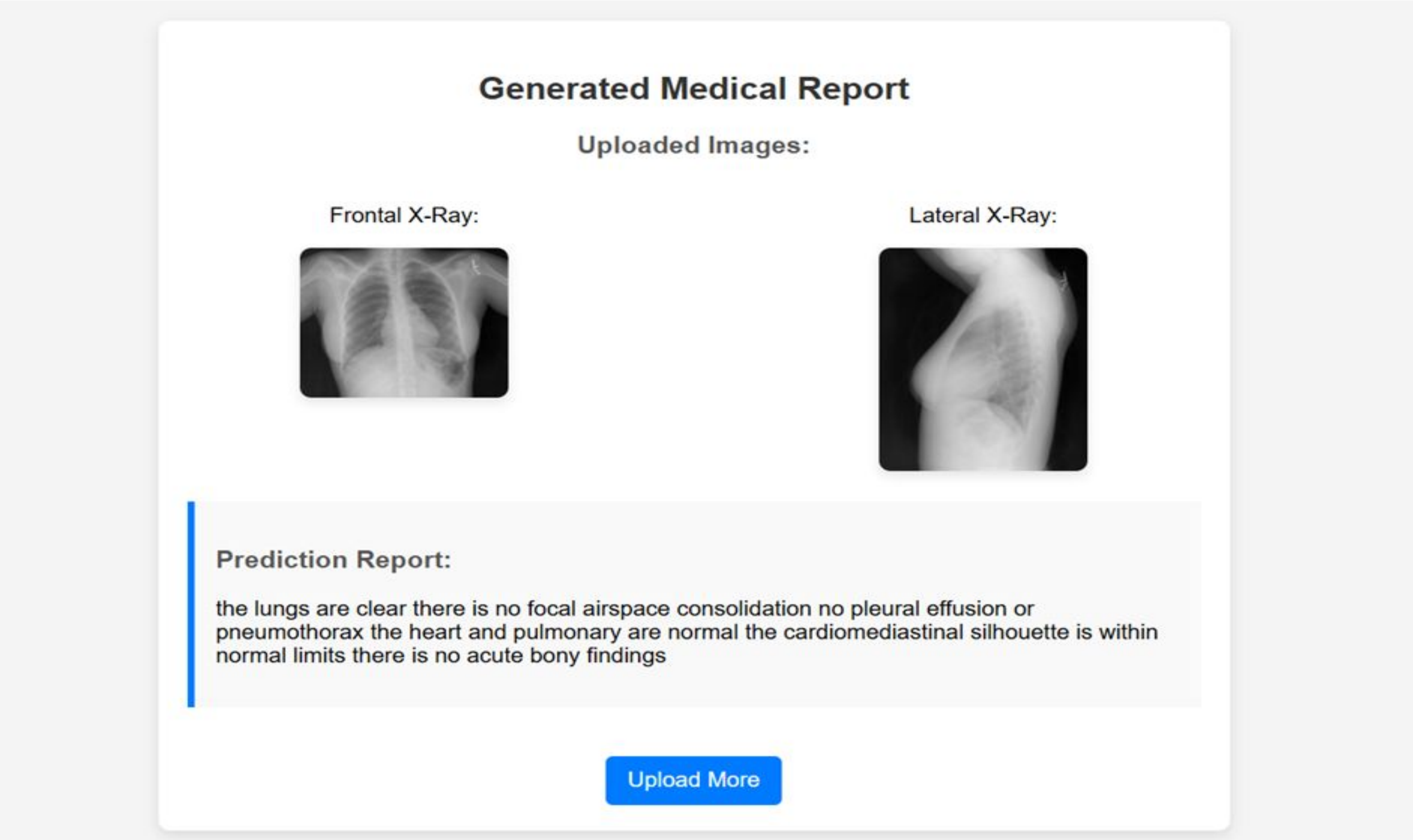

Figure : Output

RESULT ANALYSIS AND CONCLUSION

Conclusion:

- The project aims to generate accurate chest X-ray reports using frontal and lateral images.
- Data was collected from Indiana University's public chest X-ray dataset, with EDA and preprocessing.
- The baseline encoder-decoder model had poor performance.
- The attention-based model showed significant improvement, generating more meaningful reports.

LIMITATIONS AND FUTURE ENHANCEMENTS

Limitations

- The model cannot generate accurate chest x-ray due to the limited number of training samples.
- The model is more biased towards generating no disease (normal case) because dataset contains majority data points of normal case.

LIMITATIONS AND FUTURE ENHANCEMENTS

ADVANCED COLLEGE
OF ENGINEERING & MANAGEMENT

Future Enhancement

- Training the model on larger dataset like MIMIC-CXR to improve the model prediction.
- Developing complex architectures by using VT, CVT, and LLM to better capture information of images and reports while training on larger datasets.

REFERENCES

[1] "Hurdles to hospitals," The Rising Nepal, Mar. 13, 2024. [Online]. Available: https://risingnepaldaily.com/news/23880#:~:text=The%20doctor-patient%20ratio%20in,patient%20ratio%20of%201%3A1%2C000.

[2] F. F. Alqahtani, M. M. Mohsan, K. Alshamrani, J. Zeb, S. Alhamami, and D. Alqarni, "CNX-B2: A novel CNN-transformer approach for chest X-ray medical report generation," IEEE Access, vol. 12, 2024, pp. 26626–26635.

REFERENCES

[3] I. Allaouzi, M. B. Ahmed, B. Benamrou, and M. Ouardouz, “Automatic caption generation for medical images,” in Proc. 3rd Int. Conf. Smart City Appl., New York, NY, USA: Association for Computing Machinery, Oct. 2018, pp. 1–6.

[4] J. Yuan, H. Liao, R. Luo, and J. Luo, “Automatic radiology report generation based on multi-view image fusion and medical concept enrichment,” in Medical Image Computing and Computer Assisted Intervention – MICCAI 2019 (Lecture Notes in Computer Science), vol. 11769. Midtown Manhattan, NY, USA: Springer, 2019, pp. 721–729.

REFERENCES

[5] A. Rajkomar, S. Lingam, A. G. Taylor, M. Blum, and J. Mongan, "High-throughput classification of radiographs using deep convolutional neural networks," J. Digital Imaging, vol. 30, no. 1, Feb. 2017, pp. 95–101.

[6] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," Radiology, vol. 284, no. 2, Aug. 2017, pp. 574–582.

REFERENCES

[7] M. Cicero, A. Bilbily, E. Colak, D. Kontos, and J. Mermelstein, "Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs," *Investigative Radiology*, vol. 52, no. 5, May 2017, pp. 281–287.

[8] D. Soydaner, "Attention mechanism in neural networks: where it comes and where it goes," *Neural Computing and Applications*, vol. 34, no. 16, pp. 13371–13385, May 2022.

[9] J. B. Jing, P. Xie, and E. Xing, "On the automatic generation of medical imaging reports," *arXiv preprint arXiv:1711.08195*, 2017.

REFERENCES

[10] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in Proc. 2014 Conf. Empirical Methods Natural Lang. Process. (EMNLP), Doha, Qatar, pp. 1724–1734, Oct. 2014.

[11] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

Thank you!!!!!