

Topic:

ReelSense:Advanced High-Performance Recommender System

TEAM NAME: 2BIT ENGINEERS

TEAM MEMBERS:

- Anubhab Rakshit
- Bodhisatwa Dutta

Executive Summary:

ReelSense is an advanced hybrid recommender system designed to address one of the core challenges in modern recommendation engines: the balance between accuracy, diversity, and novelty. Traditional systems often optimize only for prediction accuracy, which can lead to repetitive suggestions and the creation of “filter bubbles.” ReelSense tackles this limitation by combining multiple recommendation paradigms into a unified ensemble architecture that understands both long-term user preferences and short-term behavioral patterns.

The system integrates three complementary models: matrix factorization (SVD) for stable preference memory, a graph neural network (LightGCN) for discovering hidden relational patterns, and a context-aware Transformer (SASRec) for modeling sequential user behavior. To further enhance user experience, ReelSense applies Maximal Marginal Relevance (MMR), an optimization strategy that promotes diverse and engaging recommendations while preserving relevance.

1. Introduction:

Modern recommendation systems often suffer from the filter bubble problem, where users are repeatedly exposed to highly similar content, limiting discovery and reducing engagement. Traditional recommenders prioritize accuracy alone, frequently ignoring diversity and novelty.

ReelSense is an high-performance hybrid recommender system designed to optimize three core objectives:

- Accuracy – Deliver highly relevant recommendations
- Diversity – Avoid repetitive suggestions
- Novelty – Encourage discovery of new content.

Instead of relying on a single model, ReelSense introduces an ensemble architecture that combines classical machine learning, graph neural networks, and deep learning sequence modeling to understand both user preferences and behavioral patterns.

The guiding principle of the project is:

“Move beyond simple ratings. Understand the complete user journey.”

2. System Architecture:

ReelSense is built around a three-component ensemble architecture referred to as the “Holy Trinity.” Each component represents a different mathematical perspective on recommendation.

2.1 Memory Layer — Matrix Factorization (SVD)

The Memory layer uses Singular Value Decomposition (SVD) to capture global and long-term user preferences. This classical collaborative filtering technique models latent relationships between users and items.

Its main functions are:

- Learning stable preference patterns
- Providing a robust baseline prediction
- Handling sparse rating matrices efficiently

SVD serves as the foundation of the ensemble by preserving historical preference memory.

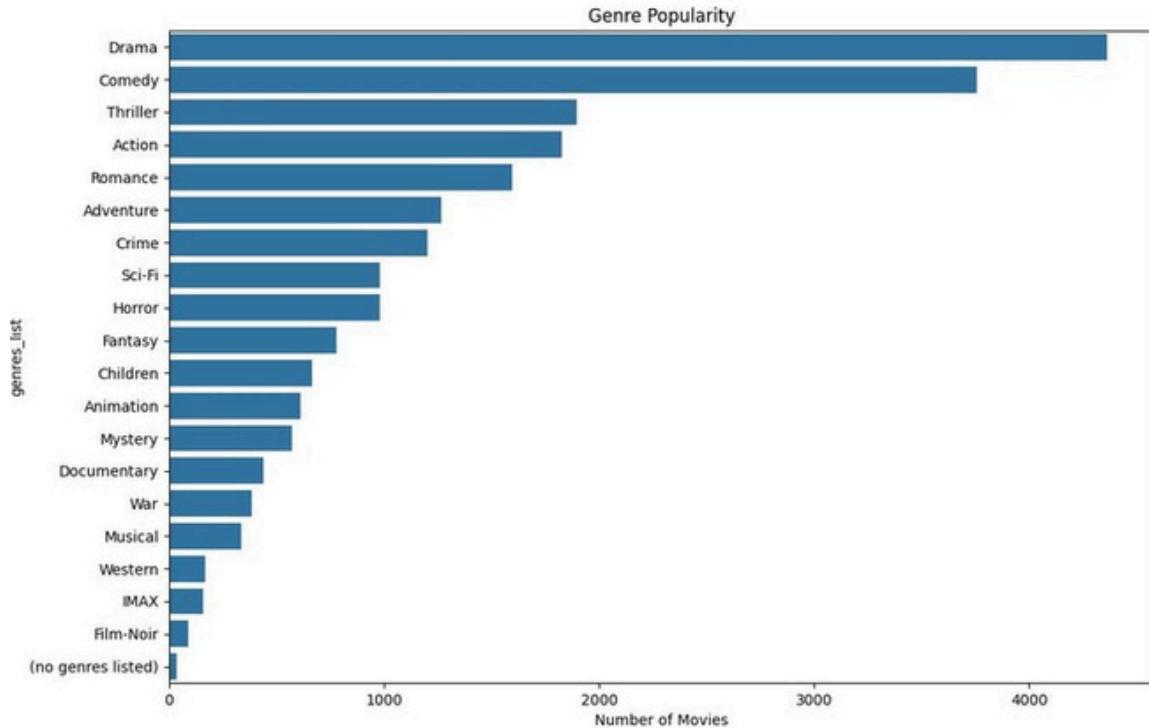


Fig1: Genre Popularity of Different Movies

2.2 Connector Layer — Graph Neural Network (LightGCN)

The Connector layer uses LightGCN, a graph neural network designed for collaborative filtering.

In this model:

- Users and movies form a bipartite interaction graph
- Embeddings propagate through graph layers
- Hidden relationships emerge from network topology

This approach enables ReelSense to detect implicit connections between users and items that are not visible in traditional matrix factorization.

LightGCN improves recommendation quality by leveraging structural relationships within the dataset.

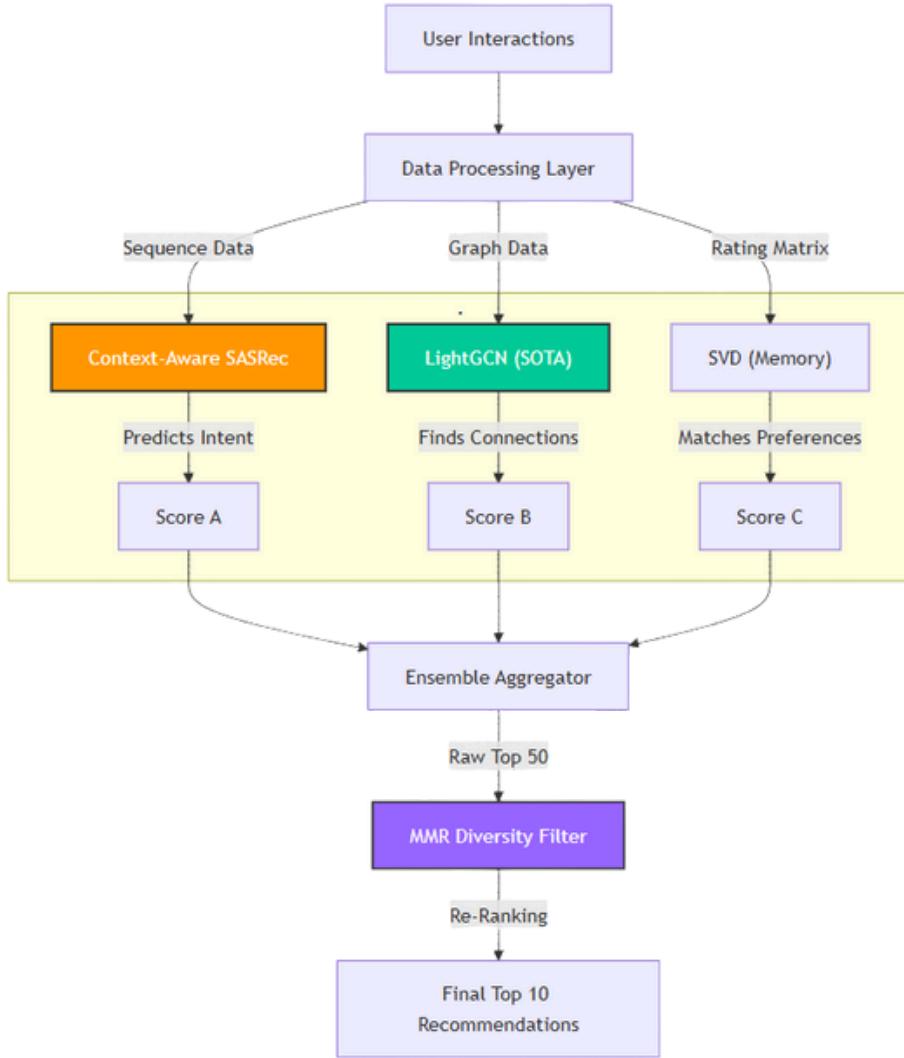


Fig 2: WorkFlow Architecture of Reel Sense

2.3 Predictor Layer — Context-Aware Transformer (SASRec)

The Predictor layer is based on SASRec, a Transformer-based sequential recommendation model.

Unlike static recommenders, SASRec analyzes:

- Temporal user behavior
- Sequential interaction patterns
- Contextual genre information

A key innovation is the integration of genre embeddings into the attention mechanism. This allows the model to generalize to unseen or rare items, addressing the cold-start problem.

3. Optimization Strategy

ReelSense is designed to balance recommendation accuracy with diversity and computational efficiency. While high prediction accuracy is essential, an effective recommender system must also prevent redundancy and maintain user engagement. ReelSense achieves this through a combination of algorithmic optimization and hardware-aware performance tuning.

3.1 Maximal Marginal Relevance (MMR)

Accuracy alone often leads to redundant recommendations. To address this, ReelSense incorporates Maximal Marginal Relevance (MMR).

The scoring function balances relevance and diversity:

$$\text{Score} = \text{X. Accuracy (1}\lambda\text{). Similarity - }$$

MMR penalizes highly similar items, ensuring that recommendation lists contain varied and engaging content. This improves:

- Content coverage
- User satisfaction
- Exploration capability

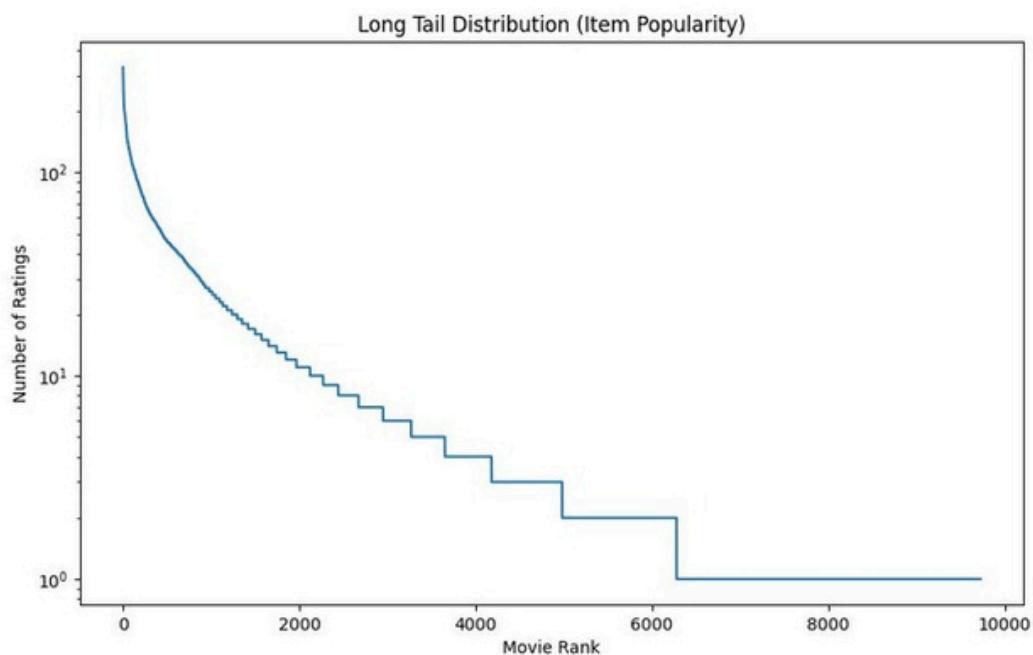


Fig 3: Long Tail Distribution vs No.of Rating Graph Based on Movie Rank

3.2 Hardware Accelerations

The deep learning components are optimized for Apple Silicon (M-series) using Metal Performance Shaders (MPS).

This platform-specific optimization enables:

- 10× faster training compared to CPU execution
- Efficient 50-epoch deep learning training loops
- Rapid experimentation without cloud infrastructure

4. Implementation and Technical Stack

The system is implemented in Python with a modular architecture.

Core technologies include:

- Python 3.10+
- PyTorch for neural network training
- Scikit-learn for evaluation metrics
- Pandas, NumPy, and SciPy for data processing
- Matplotlib and Seaborn for visualization

All recommendation models (LightGCN and SASRec) are implemented from scratch to ensure full architectural control and customization.

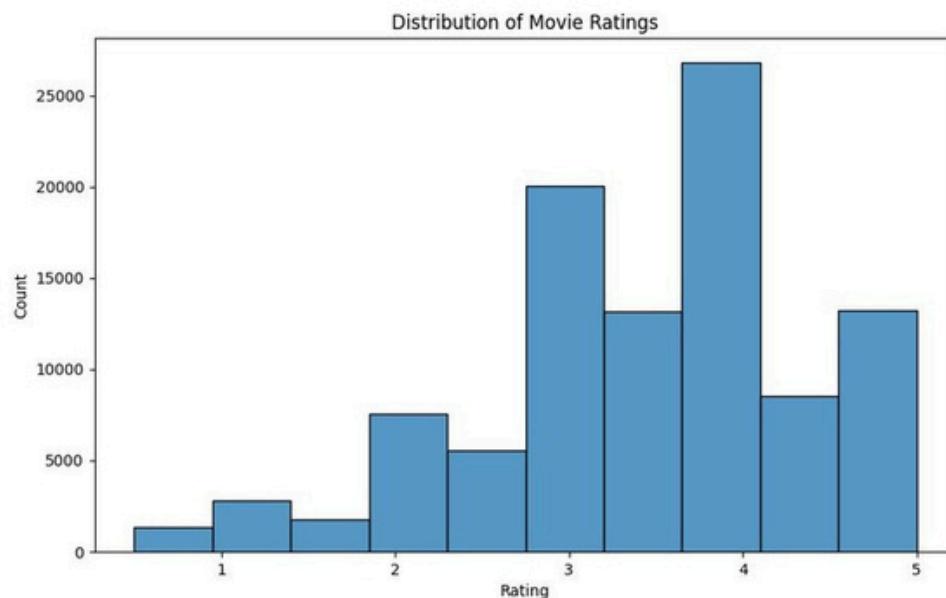


Fig 4: Distribution of Movies Rating Per Count

5. Evaluation and Results

System performance was evaluated using ranking and precision metrics.

Key results:

- NDCG (Ranking Quality): ~0.0157 – achieved by SASRec
- Precision: ~0.0034 – achieved by the Hybrid + MMR ensemble

These results demonstrate that no single model dominates across all objectives. Instead, the ensemble framework allows ReelSense to adapt to different recommendation goals.

6. Conclusion

ReelSense demonstrates that combining multiple recommendation paradigms leads to a more balanced and effective system.

By integrating matrix factorization, graph learning, sequential transformers, and diversity optimization, the system achieves:

- Strong ranking performance
- Improved recommendation diversity
- Enhanced user experience

The ensemble architecture ensures robustness and flexibility, making ReelSense suitable for scalable real-world recommendation platforms.

7. Future Work

Potential improvements include:

- Real-time online learning
- Reinforcement learning-based personalization
- Cross-domain recommendation
- Larger-scale distributed training

Topic:

Cognitive Radiology: Automated Medical Report Generation using Hierarchical Vision Transformers

TEAM NAME: 2BIT ENGINEERS

TEAM MEMBERS:

- Anubhab Rakshit
- Bodhisatwa Dutta

Executive Summary:

Cognitive Radiology is an advanced artificial intelligence system designed to automatically generate clinically meaningful radiology reports from chest **X-ray images**. The project moves beyond conventional image captioning by modeling a structured diagnostic workflow similar to that of a human radiologist: visual inspection, pathology detection, and narrative synthesis. The system integrates a hierarchical Vision Transformer backbone, an explicit disease classification module, and a transformer-based report generator into a unified end-to-end architecture. This design ensures that visual evidence and detected clinical findings directly guide the generated medical text, improving both interpretability and diagnostic relevance. Trained on a large-scale medical imaging dataset with hardware-accelerated optimization, the model achieves efficient large-batch training and scalable performance. Cognitive Radiology demonstrates the feasibility of combining hierarchical visual representation learning with structured language generation to support automated clinical documentation and assistive diagnostic workflows.

1. Introduction:

Medical imaging plays a critical role in modern healthcare, with radiology reports serving as the primary medium for communicating diagnostic findings. However, generating detailed and accurate radiology reports is a time-intensive process that requires expert knowledge and careful interpretation. With the rapid growth of medical imaging data, there is an increasing need for intelligent systems that can assist clinicians by automating parts of the reporting workflow while maintaining clinical reliability.

Recent advances in deep learning, particularly in computer vision and natural language processing, have enabled significant progress in image-to-text generation tasks. Despite these developments, many existing medical report generation systems behave like generic image captioning models. They often describe superficial visual patterns without capturing the structured reasoning process used by radiologists, leading to incomplete or clinically inconsistent reports.

2. System Architecture:

The Cognitive Radiology framework is composed of three tightly coupled modules that simulate the stages of clinical reasoning: perception, diagnosis, and report synthesis.

2.1 Visual Backbone: Hierarchical Vision Transformer

The visual backbone is built on a large-scale Vision Transformer architecture that extracts multi-level image representations. Instead of relying on a single global embedding, the system applies progressive feature alignment to capture information at different semantic scales:

- Pixel-level features encode fine-grained textures and edges
- Region-level features represent anatomical structures
- Organ-level features provide global contextual understanding

This hierarchical representation enables the model to localize subtle abnormalities while maintaining an holistic view of the chest anatomy.

2.2 Disease Classification Module

A dedicated classification module processes the hierarchical visual features to detect common thoracic pathologies. Using a multi-layer perceptron architecture, the classifier produces a probability distribution over clinically relevant disease categories.

Explicit disease prediction serves two key purposes:

- Anchoring the report generation process in interpretable clinical labels
- Reducing hallucinations by constraining the language model with diagnostic evidence

The classifier acts as an intermediate reasoning step between perception and language generation.

2.3 Report Generation Decoder

The report generator is an autoregressive Transformer decoder that synthesizes structured medical narratives. It attends to three contextual memory sources:

- Visual memory: Encoded image features
- Label memory: Disease embeddings weighted by classifier confidence
- Text memory: Global linguistic context

Through multi-head cross-attention, the decoder generates reports token by token, aligning textual output with visual findings and predicted pathologies. This mechanism ensures semantic consistency between detected abnormalities and the final narrative.

3. Training Methodology

3.1 Hardware-Accelerated Optimization

Training is optimized using modern GPU acceleration techniques to enable large-scale experimentation and efficient convergence. Mixed-precision computation reduces memory overhead while maintaining numerical stability, allowing the system to process large batches and complex transformer architectures.

Performance optimizations focus on maximizing throughput, minimizing memory bottlenecks, and ensuring stable gradient flow during deep model training.

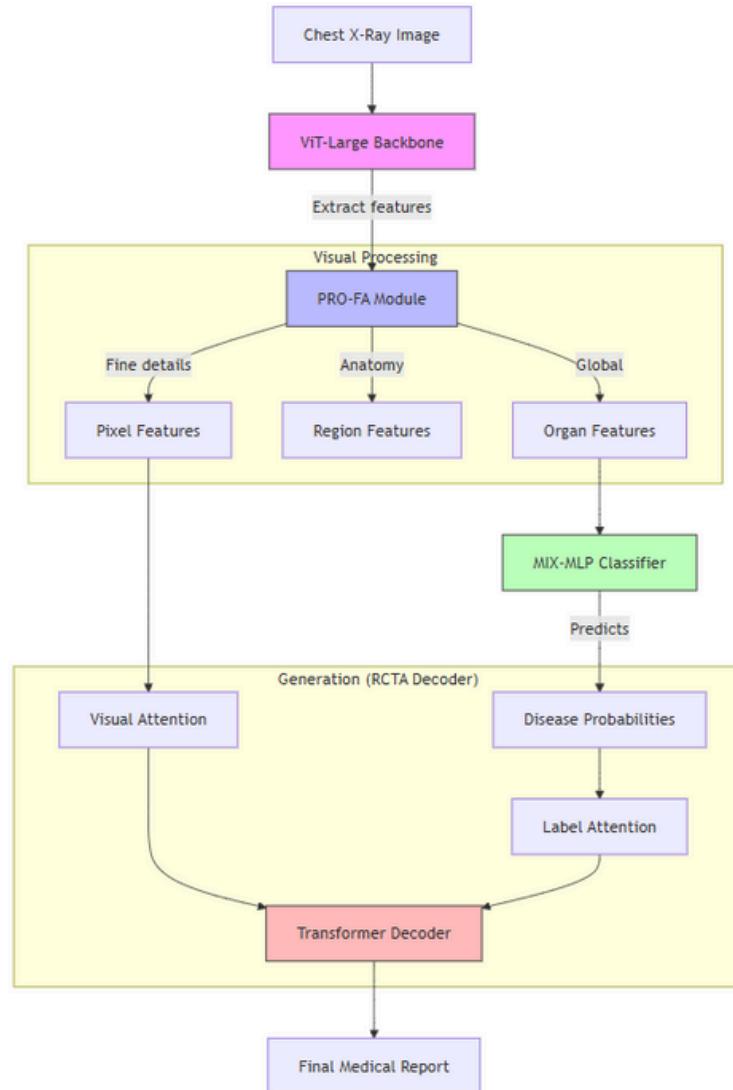


Fig: Working Architecture of Cognitive Radiology

3.2 Training Configuration

The model is trained on a large-scale chest X-ray dataset using a carefully tuned optimization strategy. The training pipeline incorporates:

- Adaptive weight optimization with decoupled regularization
- Learning rate scheduling with gradual warmup and smooth decay
- Regularization techniques to prevent overfitting and overconfidence
- Gradient stabilization for reliable large-batch updates

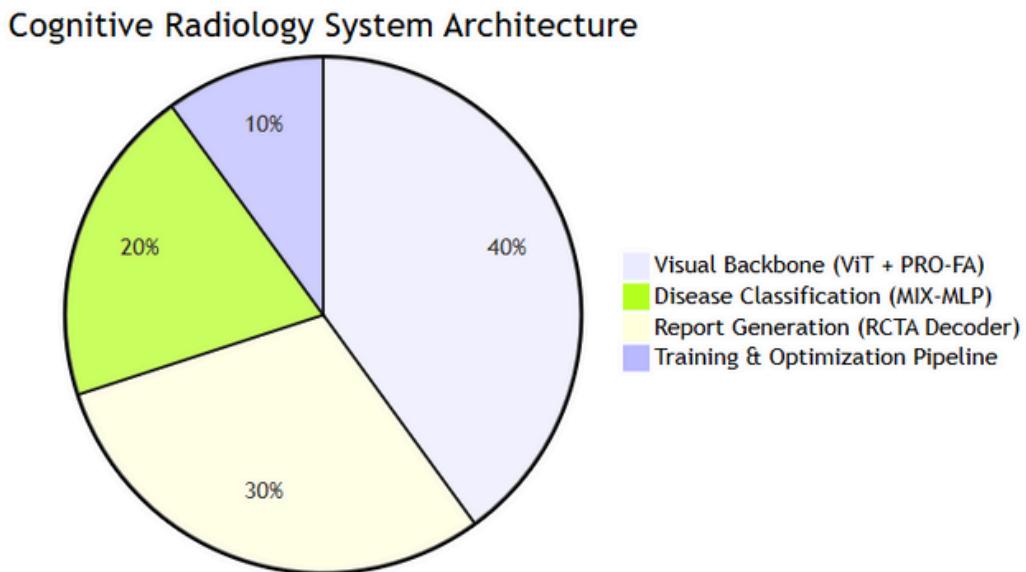
These techniques collectively improve convergence speed, generalization, and training stability.

4. Implementation Details

The system is implemented using a modern deep learning framework with a scalable cloud-based training environment. The data pipeline employs parallelized loading and memory optimizations to eliminate input bottlenecks and maintain consistent GPU utilization.

Key implementation features include:

- Modular architecture for independent experimentation with components
- Efficient multi-process data loading
- GPU-optimized tensor operations
- Reproducible training configuration



5. Applications and Impact

Cognitive Radiology has potential applications in:

- Automated clinical documentation assistance
- Decision support for radiologists
- Large-scale medical data annotation
- Research in medical image-language modeling

By integrating hierarchical visual understanding with structured language generation, the system represents a step toward intelligent clinical AI assistants capable of supporting healthcare professionals.

6. Conclusion

This methodology demonstrates a comprehensive framework for automated radiology report generation that combines hierarchical visual representation learning, explicit disease reasoning, and transformer-based language synthesis. The architecture bridges the gap between perception and clinical interpretation, offering a scalable approach to medical image-to-text systems. Future work may explore domain adaptation, multimodal fusion with electronic health records, and real-time clinical deployment.

7. Future Work

Future work will focus on improving the clinical reliability and scalability of the Cognitive Radiology system. Integrating multimodal data such as electronic health records and patient history can enable more context-aware report generation. Enhancing model interpretability through visual explanations and attention maps will increase clinician trust. Optimization for real-time deployment will support seamless integration into hospital workflows. Additionally, exploring self-supervised learning can leverage large unlabeled medical datasets to improve generalization. Expanding the framework to other imaging modalities, including CT and MRI, will broaden its applicability and move toward a comprehensive medical image-to-text platform.