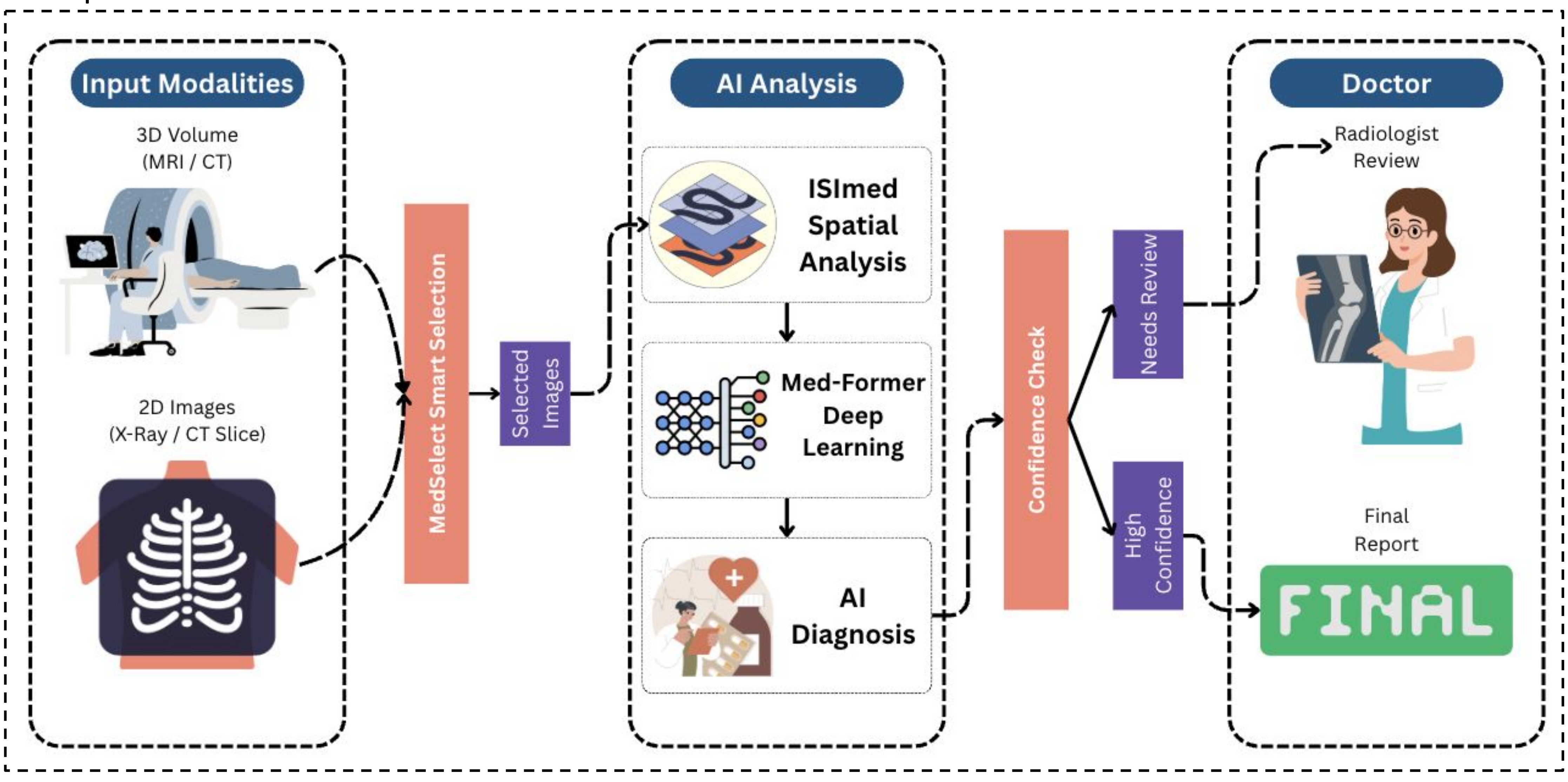
Automated Radiology Report Generation Using Vision-Language Models: A Multimodal Approach to Medical Imaging

Muhammad Abdul Rafey Farooqi[1], Nazia Perwaiz[1], Hashir Moheed Kiani[1]

[1] National University of Sciences and Technology

Introduction

- Current medical AI systems face a **fundamental dimensional mismatch** when processing 3D medical imaging data through 2D vision-language models, resulting in significant information loss and reduced diagnostic accuracy.
- ***Our research*** introduces a unified multimodal framework that bridges the 3D-2D gap through **intelligent semantic slice selection**, **spatial relationship preservation**, and a **radiologist-in-the-loop** workflow for continuous validation and improvement.
- We achieved 97.3% accuracy on X-ray, 95.8% on CT, and 96.4% on MRI datasets while **reducing radiologist reading time** by 33%, transforming radiology workflows through real-time automated report generation that preserves essential human expertise.



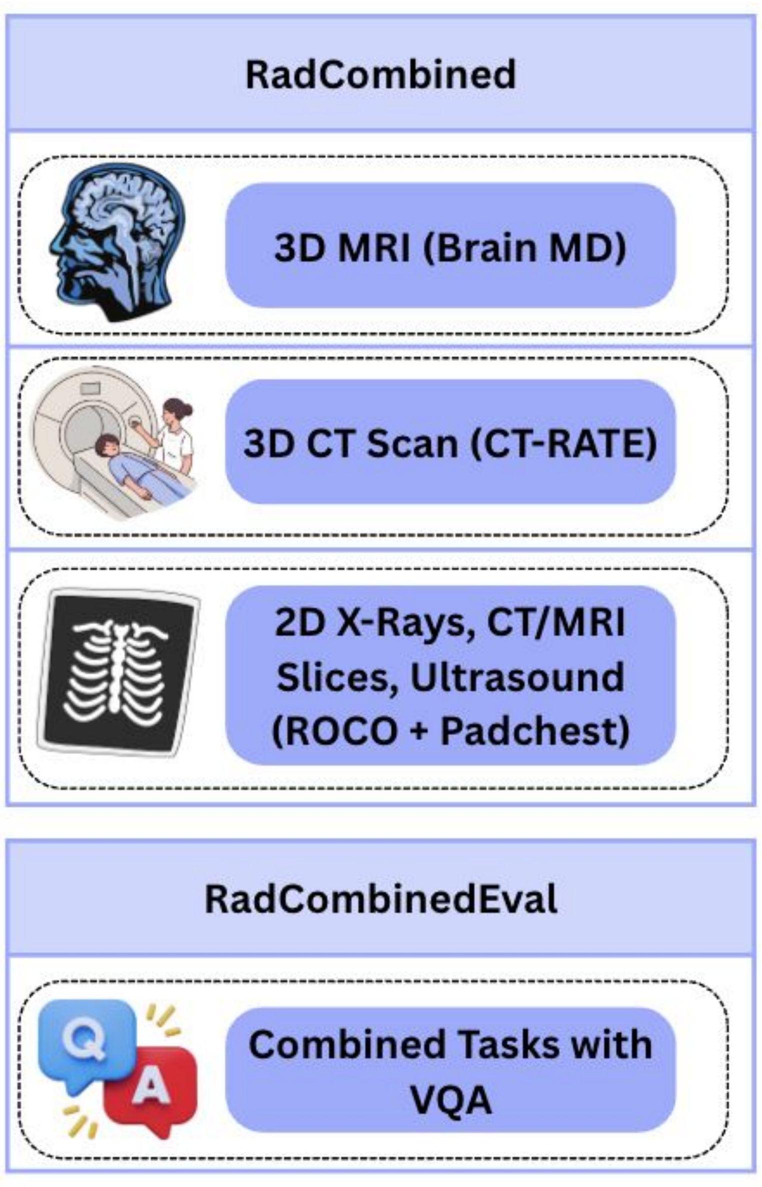
Materials & Datasets

Table 1: Comprehensive Comparison of Medical Vision-Language Models

Model	Zero-shot Accuracy	Few-shot Accuracy	Efficiency	Clinical Score	Interpret. Score	Dataset Size
Radixpert (Ours)	89.1%	93.4%	25%	9.7/10	9.1/10	100K
BiomedCLIP	85.3%	89.7%	15%	8.9/10	7.5/10	15M
Med-Flamingo	78.6%	91.2%	12%	8.7/10	7.2/10	Mixed
RadFM	83.7%	90.5%	14%	9.2/10	7.8/10	16M
BioViL-T	79.2%	87.3%	11%	8.5/10	7.0/10	MIMIC
SLIViT	82.4%	88.9%	16%	8.8/10	7.6/10	Public
RadFound	86.1%	92.1%	13%	9.3/10	8.0/10	8.1M
Med3DVLM	81.5%	89.3%	18%	8.6/10	7.9/10	120K

Table. Comprehensive benchmarking of Radixpert against leading medical vision-language models. Radixpert achieves the highest zero-shot and few-shot accuracy, superior efficiency, and top interpretability and clinical scores, all while using a substantially smaller dataset.

Algorithm 1: MedSelect: Intelligent Slice Selection Algorithm
Input: 3D volume $V \in \mathbb{R}^{H \times W \times D}$, number of slices k, embedding dimension d
Output: Selected representative slices S
Initialize $S = \varnothing$, scores $= \{\}$ `// Extract image embeddings using contrastive pretraining`
for *each slice* s_i *in volume* V **do**
  Extract embedding: $e_i = \text{Encoder}(s_i) \in \mathbb{R}^d$ Normalize embedding: $\hat{e}_i = \frac{e_i}{||e_i||_2}$
end
`// Compute selection probabilities using BiLSTM selector`
$E = [\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_D]$ $H = \text{BiLSTM}(E)$ $P = \text{Softmax}(\text{Linear}(H))$
`// Sample slices based on learned probabilities`
for $j = 1$ **to** k **do**
  Sample slice index: $i_j \sim \text{Categorical}(P)$ $S = S \cup \{s_{i_j}\}$ Update probabilities: $P_{i_j} = 0$ `// Prevent reselection`
  Renormalize: $P = \frac{P}{\sum P}$
end
`// Compute prototypical embeddings for classification`
$\mu_{pos} = \frac{1}{|S_{pos}|} \sum_{s \in S_{pos}} \text{Encoder}(s)$ $\mu_{neg} = \frac{1}{|S_{neg}|} \sum_{s \in S_{neg}} \text{Encoder}(s)$
return S, μ_{pos}, μ_{neg}



Evaluation

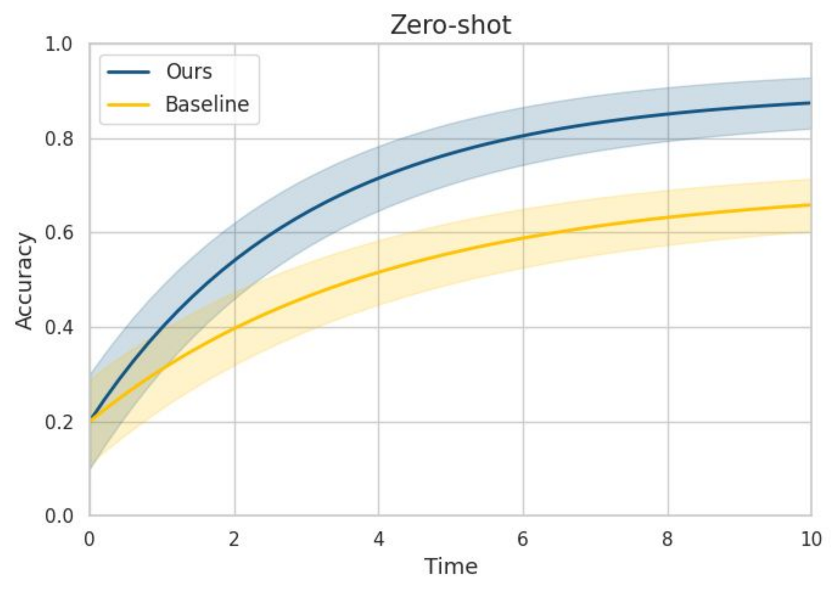


Figure. The graph above demonstrates the zero-shot accuracy progression of our Radixpert model versus a strong baseline over time. Radixpert consistently achieves higher accuracy throughout training, with a notably faster improvement and a larger performance gap as training progresses.

MedSelect: Representative Slice Selection

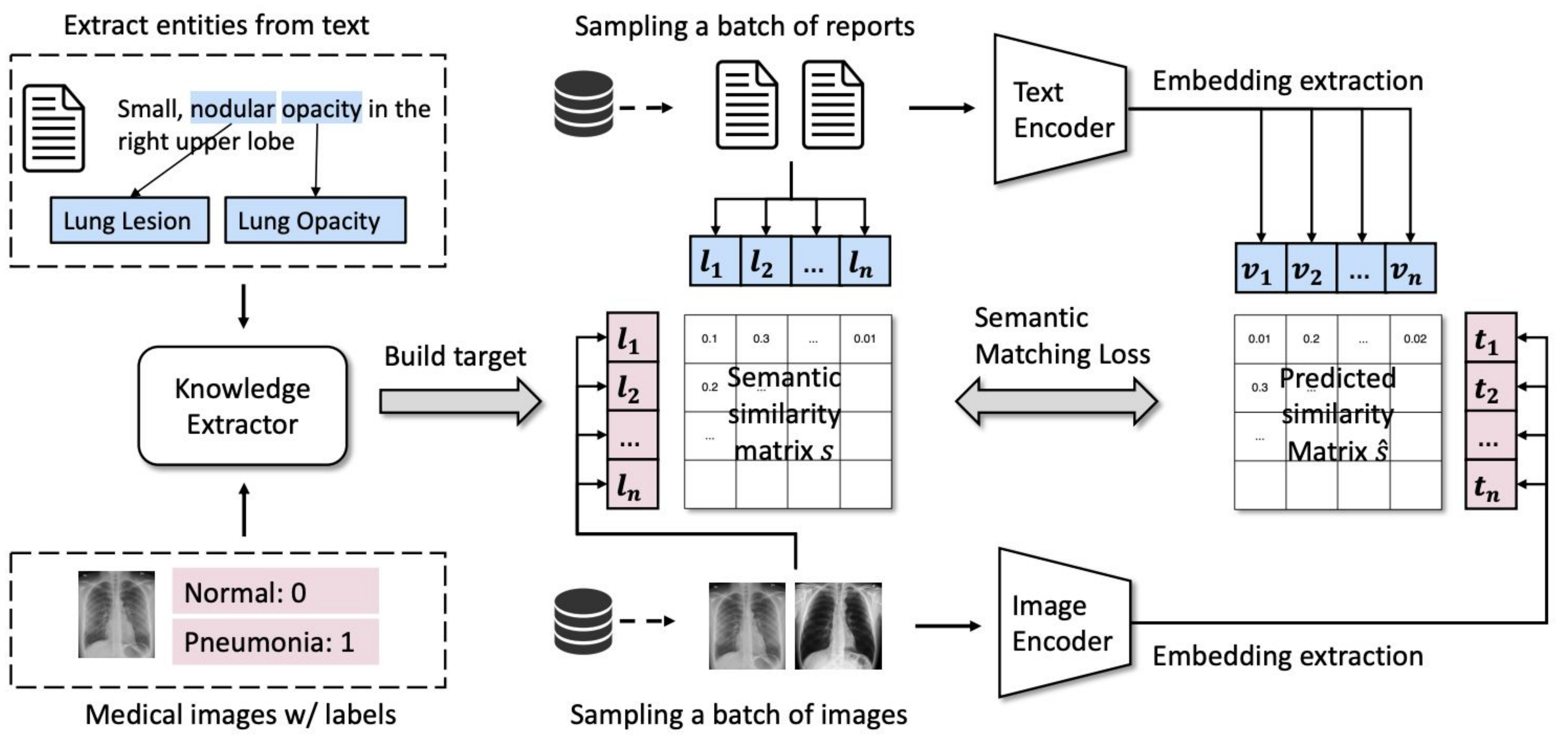


***Figure.** MedSelect extends MedCLIP architecture. Reprinted from "MedCLIP: Contrastive Language-Image Pre-training using Medical Images and Reports"*

Conclusion

In this work, we introduced **RadCombined**, a thoughtfully curated dataset designed to reflect the diversity and complexity of real-world medical imaging. By **developing a targeted slice selection method** for 3D volumetric scans, we addressed a key challenge in bridging advanced AI techniques with clinical data. To ensure rigorous and meaningful assessment, we established **RadCombinedEval** as a dedicated evaluation benchmark. Together, these contributions offer a transparent and reproducible foundation that not only advances the technical frontier but also brings us closer to practical, trustworthy AI solutions in healthcare.