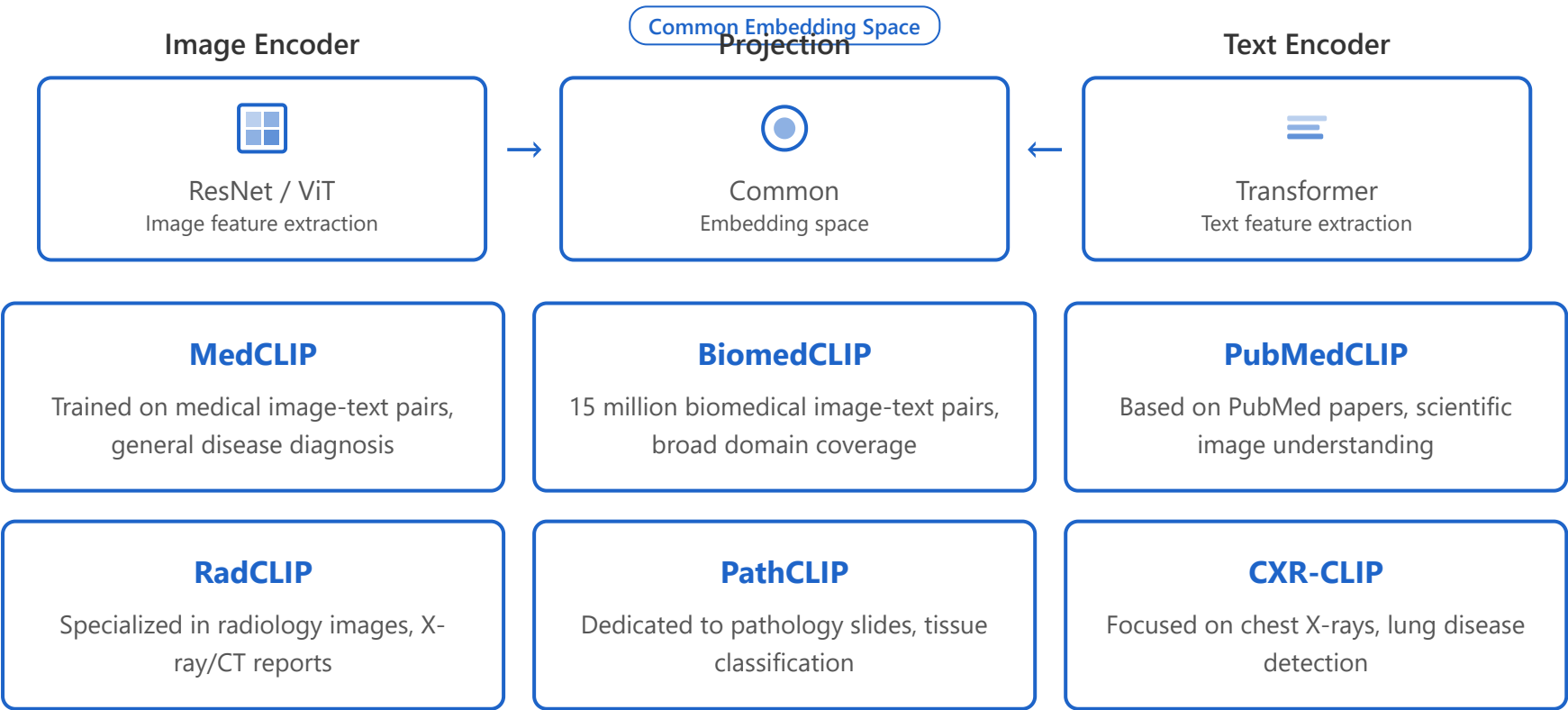# CLIP Medical Variants: Contrastive Learning

CLIP (Contrastive Language-Image Pre-training): Aligns images and text in a common embedding space, enabling **zero-shot learning**

Common Embedding Space

Image Encoder	Projection	Text Encoder
ResNet / ViT — Image feature extraction	→ Common — Embedding space	← Transformer — Text feature extraction

MedCLIP
Trained on medical image-text pairs, general disease diagnosis

BiomedCLIP
15 million biomedical image-text pairs, broad domain coverage

PubMedCLIP
Based on PubMed papers, scientific image understanding

RadCLIP
Specialized in radiology images, X-ray/CT reports

PathCLIP
Dedicated to pathology slides, tissue classification

CXR-CLIP
Focused on chest X-rays, lung disease detection

Contrastive Learning
- Matching image-text pairs: high similarity

Image-Text Alignment Effects
- Enables zero-shot classification

- Non-matching pairs: low similarity
- Discriminative learning with InfoNCE loss

- Image retrieval with text queries
- Multimodal representation learning