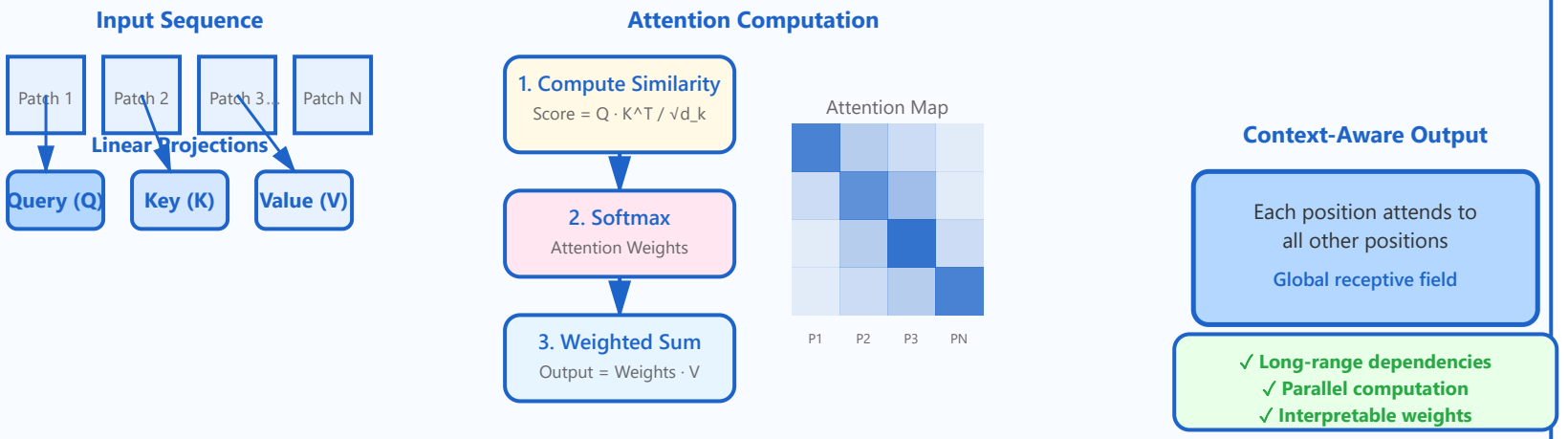
Attention Mechanisms

Self-Attention Mechanism



Self-Attention

Capture long-range dependencies. Every position attends to all others

Cross-Attention

Attend between different modalities or sequences. Query from one, key/value from another

Vision Transformers

Pure attention-based architecture. ViT, Swin Transformer for medical imaging

Hybrid Architectures

CNN backbone + Transformer head. CoAtNet, TransUNet combine local and global context

Interpretability Benefits

Attention maps show model focus. More intuitive than CNN activation maps