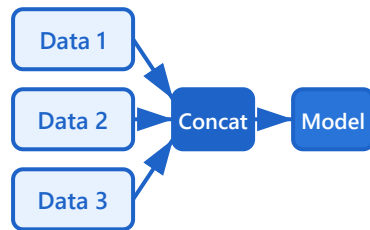


Early vs Late Fusion

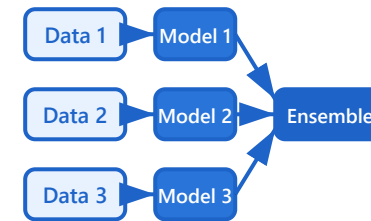
Early Fusion



Feature-level concatenation

- ✓ Captures feature interactions
- ✓ Joint representation learning
- ✗ Computationally expensive

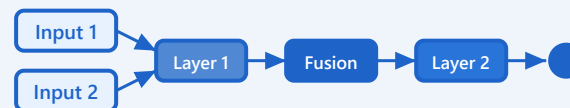
Late Fusion



Decision-level combination

- ✓ Flexible and modular
- ✓ Independent training
- ✗ Misses feature interactions

Intermediate Fusion



Combines features at intermediate layers, balancing both approaches

Trade-offs

Early: captures interactions but computationally expensive | Late: flexible but misses feature interactions