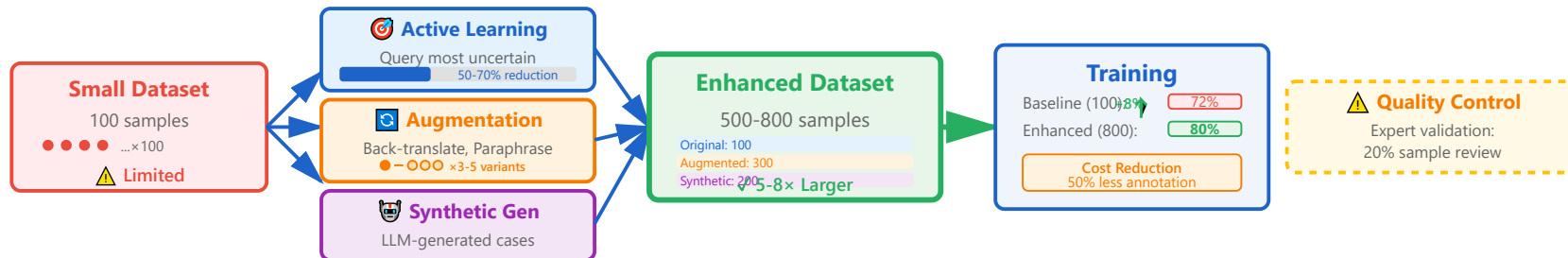


Data Efficiency Techniques

Maximize Learning from Limited Medical Data Active learning, augmentation, and synthesis



3-5x

Data Efficiency

50%

Cost Reduction

+8%

Accuracy Gain

Active Learning

- **Uncertainty Sampling:** Select most uncertain predictions
- **Diversity Sampling:** Cover different case types
- **Query-by-Committee:** Multiple model disagreement
- Reduce annotation needs by 50-70%



Data Augmentation

- **Back-translation:** Translate to another language and back
- **Synonym replacement:** Medical term substitution
- **Paraphrasing:** GPT-4 based rephrasing
- **Mixup:** Interpolate between examples