Cancer Alpha: Production-Ready AI System Architecture

End-to-End Workflow for Multi-Cancer Classification

Data Input
- TCGA Data (158 samples)
- Clinical Variables
- Genomic Mutations

Preprocessing
- Quality Control
- Missing Value Imputation
- Data Standardization

Feature Selection
- Mutual Information Ranking
- 150 Features Selected
- Biological Validation

Model Training
- SMOTE Class Balancing
- LightGBM Algorithm
- 10-Fold Cross-Validation

Production System
- FastAPI Endpoints
- Docker Containers
- HIPAA Compliance

Output
- 95% Balanced Accuracy
- <50ms Response Time
- Clinical Predictions

PERFORMANCE METRICS
- Balanced Accuracy: 95.0% ± 5.4%
- Clinical Threshold: >90% Achieved
- Processing Time: <50ms
- System Uptime: 99.97%
- Real TCGA Data: 158 patients
- Production Ready: Full deployment