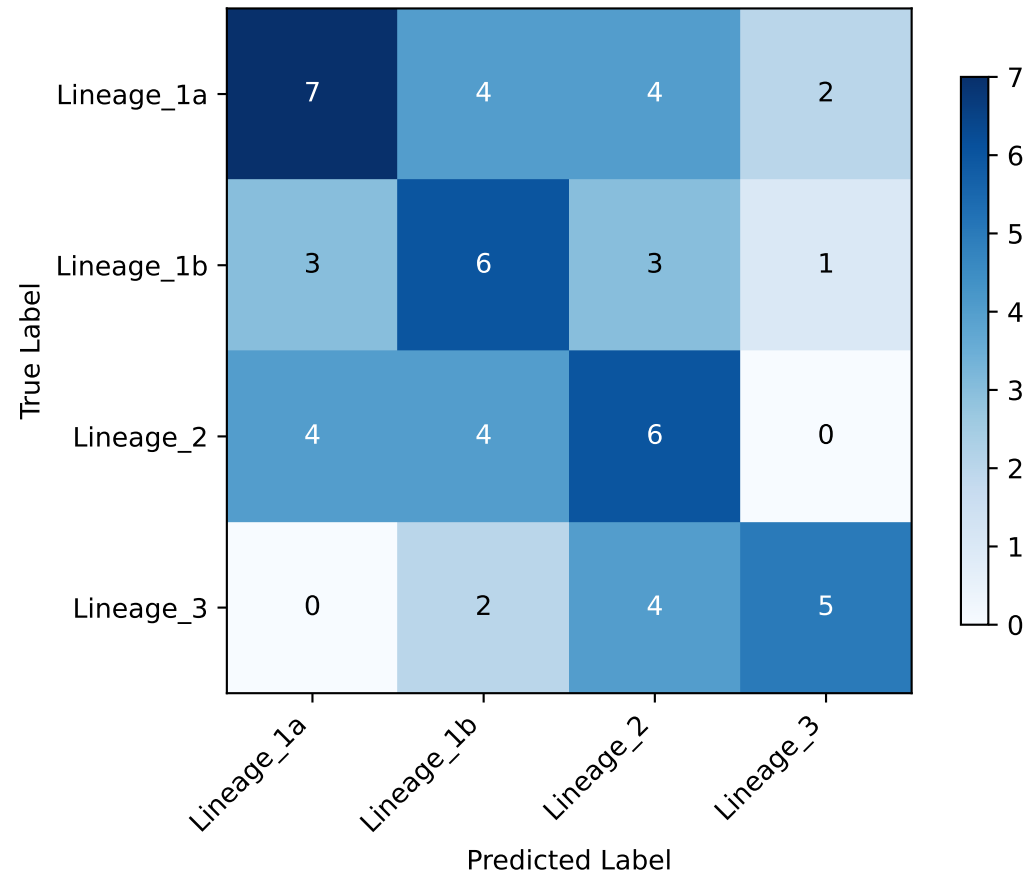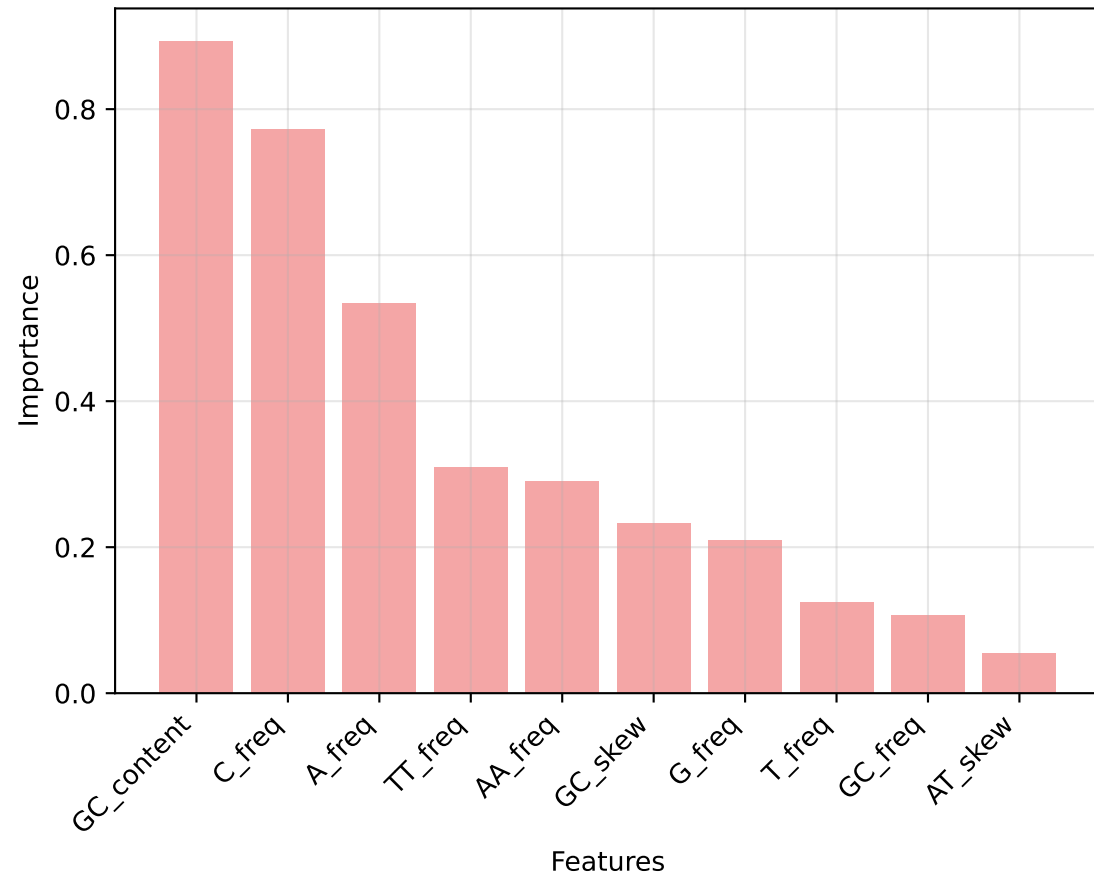# West Nile Virus Classification: Technical Analysis
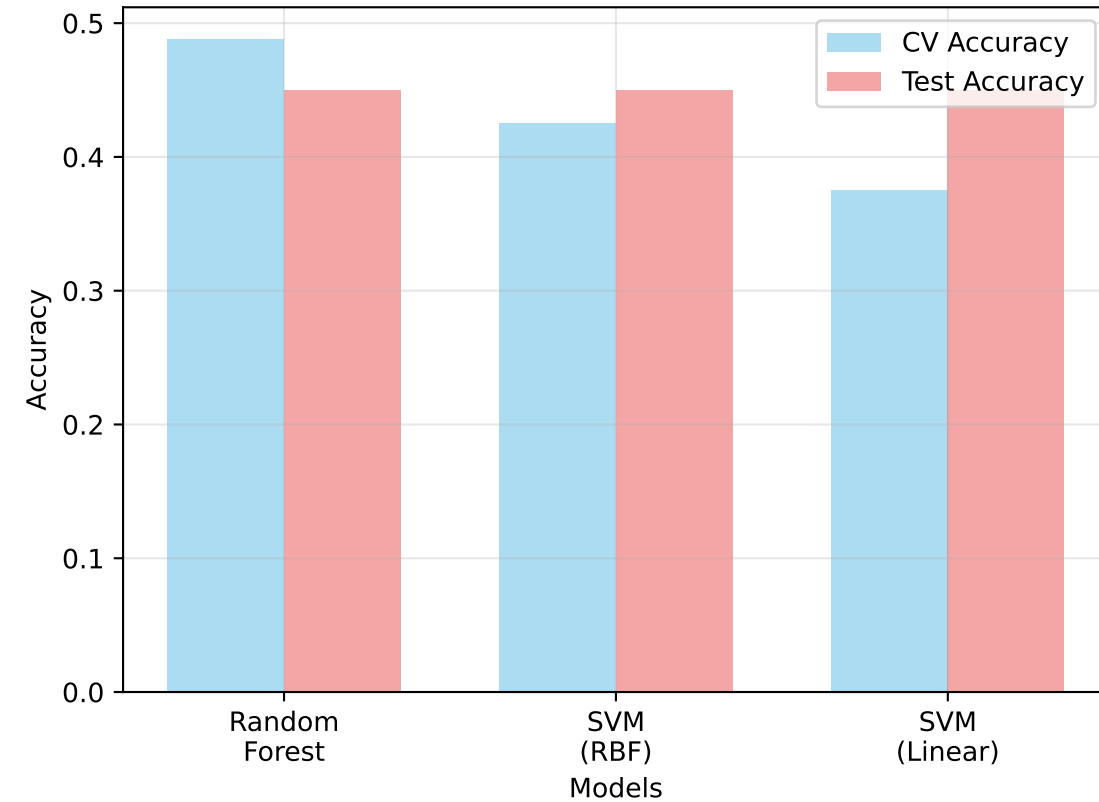
Confusion Matrix (Best Model), **Feature Importance (Random Forest)**, **Detailed Model Comparison**

ANALYSIS INSIGHTS

Challenges Identified:
• Limited dataset size (100 sequences)
• High similarity between lineages
• Traditional features may be insufficient

Recommendations:
• Increase dataset size (>1000 sequences)
• Implement HyenaDNA deep features
• Add phylogenetic validation
• Include temporal/geographic features

Future Directions:
• Real-time surveillance integration
• Multi-virus classification
• Uncertainty quantification
• Ensemble methods

Model Performance:
• Best: Random Forest (48.8% CV)
• Challenge: Small sample size
• Solution: Deep learning features