

Weekly Meeting with Dr. Hannah



Presenter: Hasan Shaikh
Quantitative Imaging Research and
Artificial Intelligence Lab (QIRAIL)

Progress Report Submission Timeline

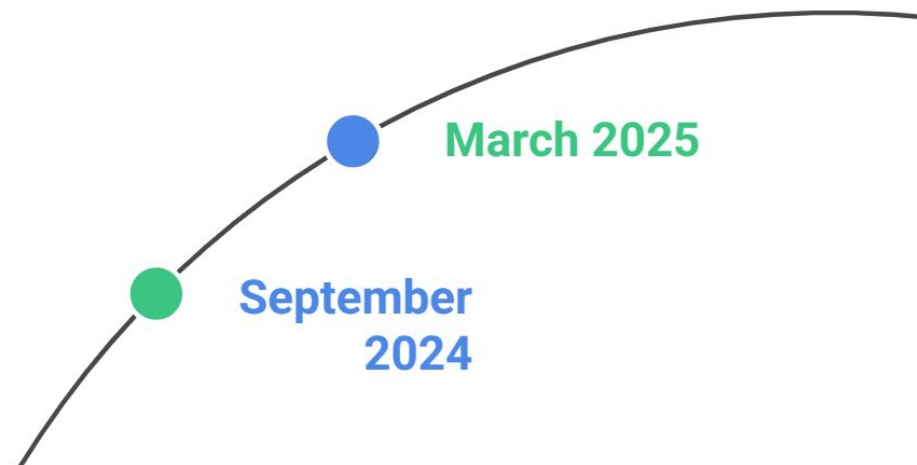
Report Submission

Progress report completed and submitted for review



Project Start

Fellowship work commenced under Dr. Hannah's supervision



Data Recipients 

Dr. Jeny (NITK)

Dr. Balu

Dr. Hannah



Data Sharing
with NITK

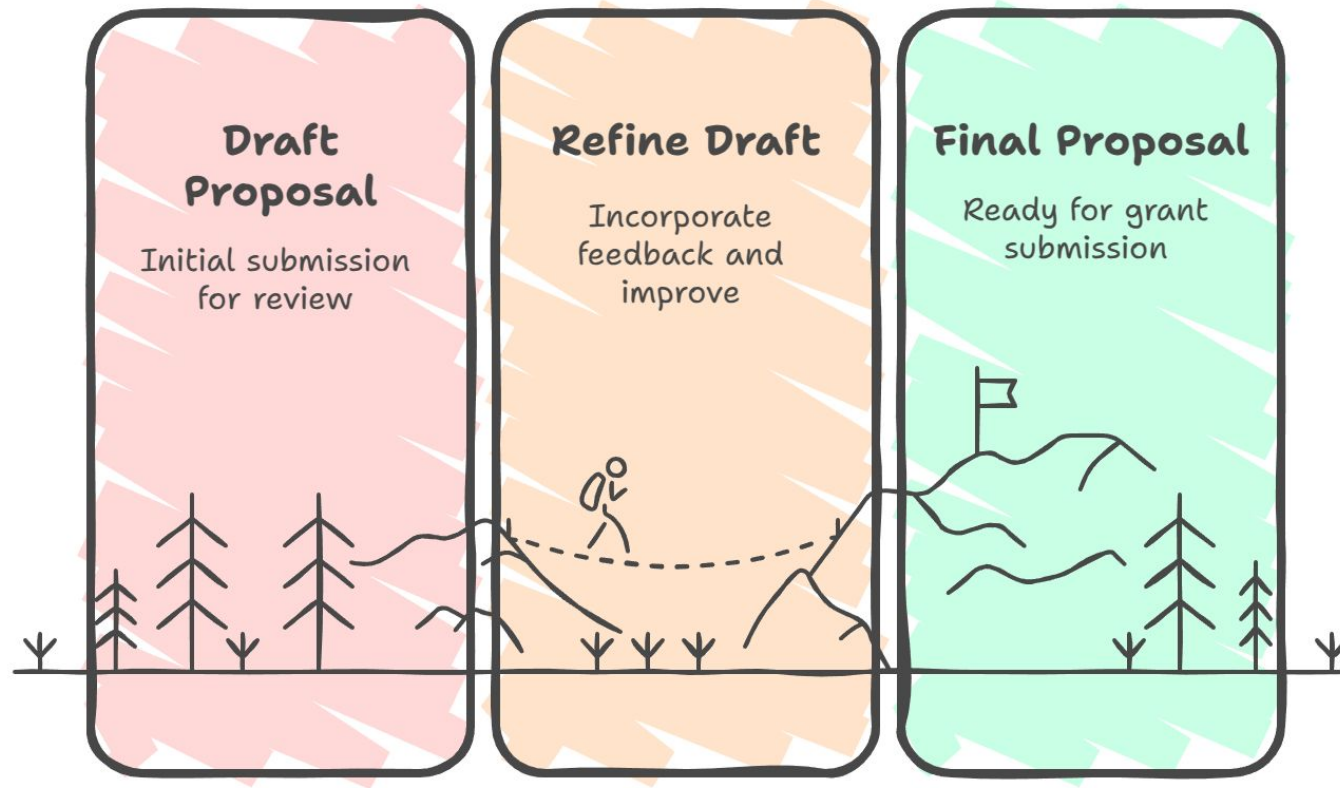


Data Preparation

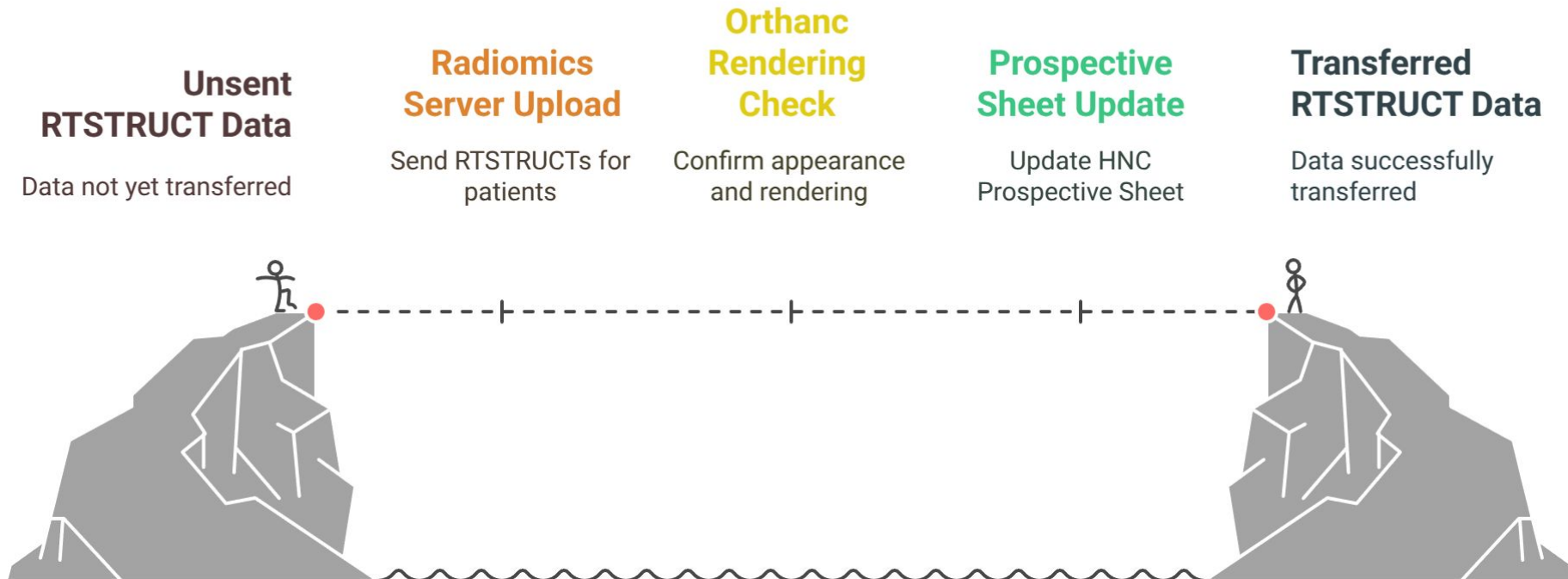
De-identified CT Images

RTSTRUCT (GTV-1) Files

NVIDIA Academic Grant Proposal



RTSTRUCT Upload & Prospective Sheet Update



***Credit:** Thanks to Sathya who helped to send RTSTRUCT, also he shared lot of data.

CHAVI Client & De-identification System

Request Again Data Upload

Asked Joanna to re-upload clinical and DICOM data to CHAVI-client.

Re-anonymize Data

Re-applied de-anonymization to DICOM and RTSTRUCT files.

Learn DICOMFx

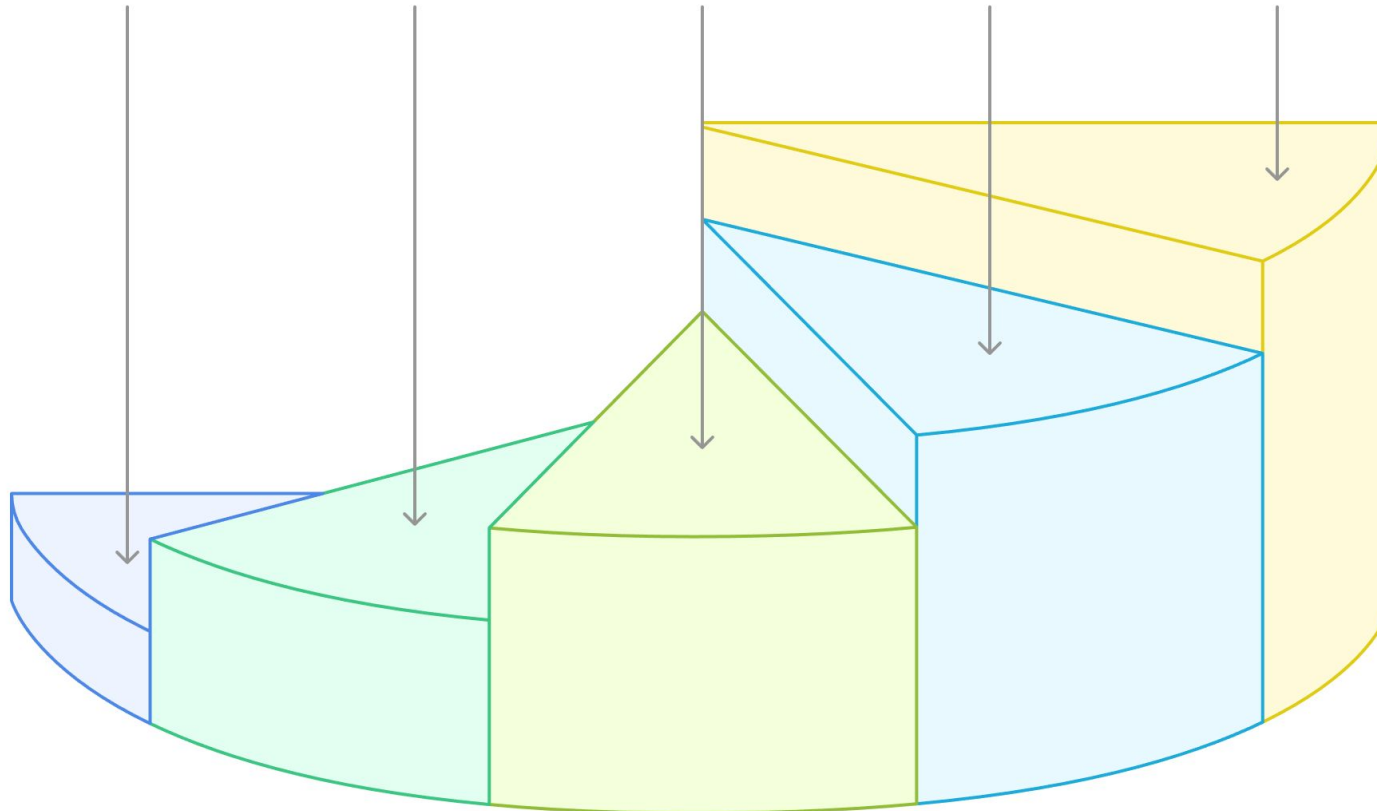
Explore and learn about the DICOMFx extension in VS Code.

View Metadata

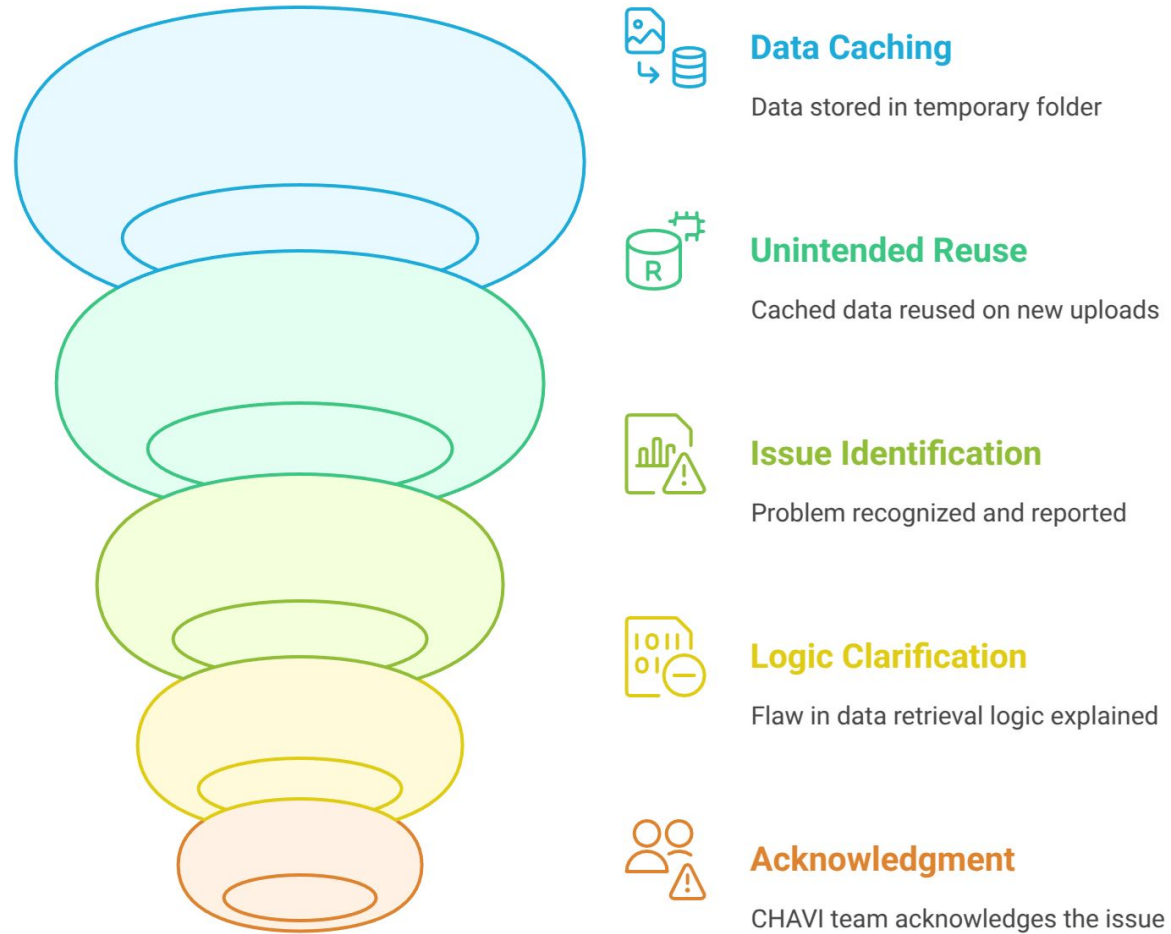
Using DICOMFx to view and understand DICOM metadata.

Debug Metadata

Utilizing DICOMFx to debug and correct DICOM metadata.



Critical Issue in the CHAVI De-anonymization Infrastructure - Data Caching



Classifier Implementation and Performance Comparison

Implemented the following 49 classifier-feature selection combinations (7 classifiers × 7 feature selection techniques):

Classifiers to Evaluate:

1. Logistic Regression
2. Naive Bayes
3. Linear Support Vector Machine (Linear SVM)
4. Radial Basis Function Kernel SVM (RBF SVM)
5. Decision Tree
6. Random Forest
7. Voting Classifier

Selected Feature Selection Techniques:

1. SelectKBest (Univariate statistical selection)
2. Least Absolute Shrinkage and Selection Operator (LASSO)
3. Particle Swarm Optimization (PSO)
4. Whale Optimization Algorithm (WOA)
5. Grey Wolf Optimizer (GWO)
6. Genetic Algorithm (GA)
7. Simulated Annealing (SA)

All combinations have been successfully implemented and tested. Results were evaluated using ROC AUC and Accuracy on both training and test data..

How to optimize radiomics studies?

Apply Pure Metaheuristics

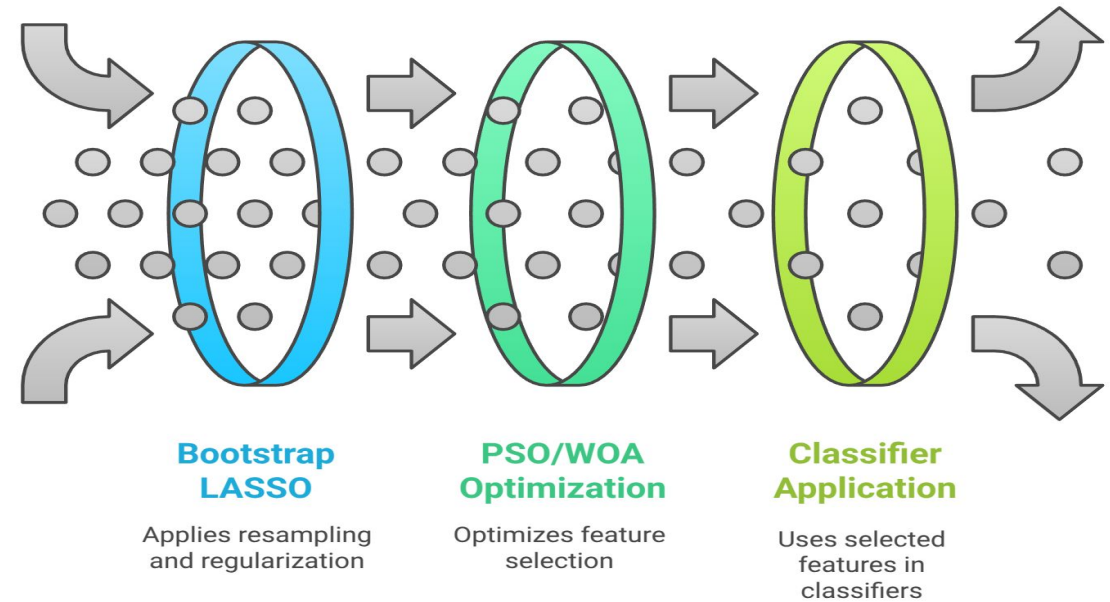
May lead to overfitting
and less interpretable
results



Apply Filtering First

Reduces overfitting and
improves reproducibility

Feature Selection Process



“A hybrid feature selection algorithm combining information gain and grouping particle swarm optimization for cancer diagnosis” (DOI: [10.1371/journal.pone.0290332](https://doi.org/10.1371/journal.pone.0290332))

Selected Feature Techniques: Least Absolute Shrinkage and Selection Operator (LASSO)

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.69	0.68	0.35	0.63
Naive Bayes	0.63	0.68	0.70	0.68
Linear Support Vector Machine (Linear SVM)	0.61	0.65	0.73	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.64	0.66	0.68	0.63
Decision Tree	0.87	0.80	0.27	0.47
Random Forest	0.99	0.92	0.33	0.53
Voting Classifier	0.99	0.82	0.48	0.63

Selected Feature Techniques: Particle Swarm Optimization (PSO)^{1,2}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.63	0.54	0.57	0.59
Naive Bayes	0.77	0.51	0.58	0.42
Linear Support Vector Machine (Linear SVM)	0.87	0.65	0.48	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.87	0.64	0.44	0.63
Decision Tree	0.84	0.82	0.64	0.63
Random Forest	0.99	0.95	0.49	0.47
Voting Classifier	0.99	0.82	0.48	0.63

Hybrid Feature Selection Approach: Bootstrap LASSO + Particle Swarm Optimization (PSO)^{1,2}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.70	0.70	0.61	0.68
Naive Bayes	0.80	0.72	0.70	0.58
Linear Support Vector Machine (Linear SVM)	0.78	0.66	0.51	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.78	0.66	0.51	0.63
Decision Tree	0.90	0.85	0.65	0.63
Random Forest	0.98	0.95	0.80	0.84
Voting Classifier	0.99	0.95	0.54	0.53

Selected Feature Techniques: Whale Optimization Algorithm (WOA)^{3,4}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.64	0.53	0.58	0.63
Naive Bayes	0.78	0.53	0.45	0.53
Linear Support Vector Machine (Linear SVM)	0.87	0.65	0.49	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.13	0.65	0.67	0.63
Decision Tree	0.85	0.77	0.56	0.58
Random Forest	0.99	0.93	0.48	0.47
Voting Classifier	0.99	0.82	0.47	0.63

Hybrid Feature Selection Approach: Bootstrap LASSO + Whale Optimization Algorithm (WOA)^{3,4}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.69	0.73	0.58	0.63
Naive Bayes	0.77	0.69	0.82	0.63
Linear Support Vector Machine (Linear SVM)	0.36	0.65	0.63	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.75	0.65	0.61	0.63
Decision Tree	0.92	0.88	0.57	0.53
Random Forest	0.97	0.91	0.52	0.58
Voting Classifier	0.99	0.95	0.54	0.53

Selected Feature Techniques: Grey Wolf Optimization Algorithm (GWOA)^{5,6}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.66	0.57	0.58	0.63
Naive Bayes	0.80	0.54	0.48	0.53
Linear Support Vector Machine (Linear SVM)	0.87	0.65	0.37	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.87	0.65	0.38	0.63
Decision Tree	0.85	0.76	0.52	0.47
Random Forest	0.99	0.95	0.56	0.53
Voting Classifier	0.99	0.82	0.48	0.63

Hybrid Feature Selection Approach: Bootstrap LASSO + Grey Wolf Optimization Algorithm (GWOA)^{5,6}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.71	0.68	0.48	0.47
Naive Bayes	0.77	0.73	0.73	0.74
Linear Support Vector Machine (Linear SVM)	0.73	0.65	0.54	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.76	0.65	0.40	0.63
Decision Tree	0.93	0.85	0.50	0.47
Random Forest	0.97	0.95	0.80	0.84
Voting Classifier	0.99	0.95	0.54	0.53

Selected Feature Techniques: Genetic Algorithm (GA)

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.63	0.58	0.71	0.68
Naive Bayes	0.78	0.51	0.65	0.53
Linear Support Vector Machine (Linear SVM)	0.15	0.65	0.74	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.82	0.65	0.68	0.63
Decision Tree	0.79	0.80	0.54	0.53
Random Forest	0.99	0.95	0.56	0.47
Voting Classifier	0.99	0.82	0.48	0.63

Hybrid Feature Selection Approach: **Bootstrap LASSO + Genetic Algorithm (GA)**^{7,8}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.60	0.69	0.87	0.63
Naive Bayes	0.72	0.69	0.93	0.68
Linear Support Vector Machine (Linear SVM)	0.64	0.65	0.89	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.46	0.65	0.11	0.63
Decision Tree	0.93	0.85	0.80	0.79
Random Forest	0.91	0.89	0.90	0.79
Voting Classifier	0.99	0.95	0.54	0.53

Selected Feature Techniques: Simulated Annealing (SA)^{9,10}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.60	0.60	0.88	0.74
Naive Bayes	0.73	0.53	0.80	0.53
Linear Support Vector Machine (Linear SVM)	0.14	0.65	0.68	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.22	0.65	0.27	0.63
Decision Tree	0.84	0.74	0.77	0.74
Random Forest	0.99	0.96	0.60	0.47
Voting Classifier	0.99	0.82	0.48	0.63

Hybrid Feature Selection Approach: **Bootstrap LASSO + Simulated Annealing (SA)**^{9,10}

Classifier	Training ROC AUC	Training Accuracy	Test ROC AUC	Test Accuracy
Logistic Regression	0.60	0.70	0.90	0.63
Naive Bayes	0.72	0.70	0.94	0.63
Linear Support Vector Machine (Linear SVM)	0.65	0.65	0.91	0.63
Radial Basis Function Kernel SVM (RBF SVM)	0.68	0.65	0.12	0.63
Decision Tree	0.91	0.81	0.85	0.84
Random Forest	0.99	0.91	0.81	0.68
Voting Classifier	0.99	0.95	0.54	0.53

REFERENCES

1. Particle Swarm Optimization (PSO) Original Paper
2. Implemented Particle Swarm Optimization (PSO) for HNC paper
3. Whale Optimization Algorithm (WOA) Original Paper
4. Implemented Whale Optimization Algorithm (WOA) for HNC Paper
5. Grey Wolf Optimizer (GWO) Original Paper
6. Implemented Grey Wolf Optimizer (GWO) for HNC Paper
7. Genetic Algorithm (GA) Original Paper
8. Implemented Genetic Algorithm (GA) for HNC Paper
9. Simulated Annealing (SA) Original Paper
10. Implemented Simulated Annealing (SA) for HNC Paper