Predicting Heart Failure

Ashish Singh, Tian Qiu, Jun Ryu, Yuki Yu, Teresa Bui, Sesha Chalamalasetti

Introduction & Objective

Using a dataset from Kaggle with **60 predictors** and **369 observations**, we assessed algorithms such as logistic regression, Random Forest, and SGD and SVM for **optimal predictive accuracy in predicting mortality based on a number of predictors.**

Feature Selection

Utilizing VIF, we detected **multicollinearity** and removed highly correlated predictors.

PREDICTORS

White Blood Cell Count	Lymphocyte Ratio
Platelet Count	Monocyte Percentage
Number of Major Vessels	ST-segment depression
Number of Follow-Ups	

Models Tested

Applied the following linear classifiers:

- **Logistic Regression**
- **Random Forest**
- **SGD Support Vector Machine**

Model Tuning

Utilized **Gridsearchcv function** to test a range of hyperparameters and choose the optimal ones.

Initial Results

Random Forest: 98.2% accuracy

Logistic Regression: 82.9% accuracy

SGD SVM: 82.0% accuracy

Final Model: Random Forest

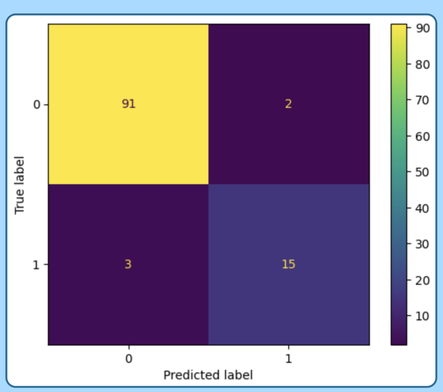
Hyperparameter Tuning

	params	rank_test_score	mean_test_score
0	{'rf_cv__n_estimators': 5}	5	0.933786
1	{'rf_cv__n_estimators': 10}	1	0.945551
2	{'rf_cv__n_estimators': 25}	3	0.933861
3	{'rf_cv__n_estimators': 50}	2	0.941629
4	{'rf_cv__n_estimators': 100}	3	0.933861

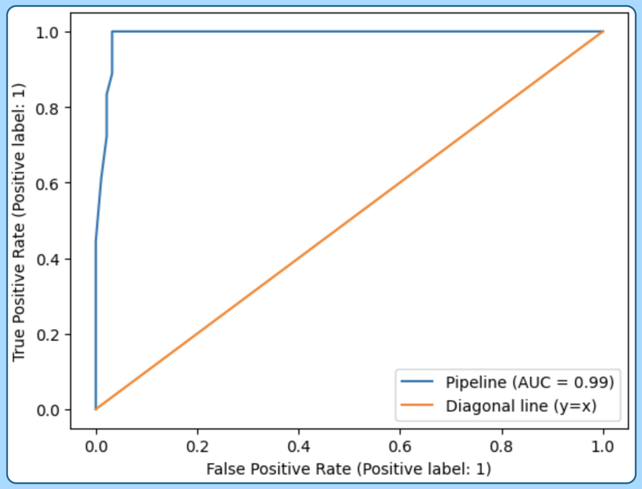
Classification Report

		precision	recall	f1-score	support
Mortality 0 = Died 1 = Alive	0	1.00	0.98	0.99	93
	1	0.90	1.00	0.95	18
accuracy				0.98	111
macro avg		0.95	0.99	0.97	111
weighted avg		0.98	0.98	0.98	111

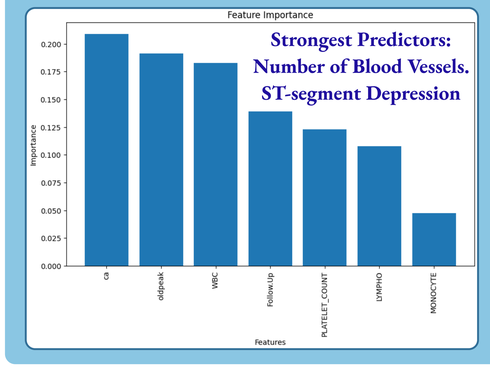
Model Evaluation



ROC Curve



Feature Importance & Insights



Higher follow-up frequency, 'ca', and 'WBC' are linked to **lower mortality risk**, whereas a higher 'Lymphocyte ratio' **increases it.**

Conclusions & Limitations

- Random Forest was the top-performing model at 98.2%
- Overfitting concerns due to a small dataset
- Hyperparameter tuning with grid search mitigated overfitting issues
- **More balanced data is needed for a better, less biased model!**

Acknowledgements

Khan, Asghar Ali. “Heart Failure Prediction Dataset.” Faisalabad Hospital Named Institute of Cardiology, 2022.