Breast Cancer Classification Using SVM and Decision Trees

Sarfo K. Frimpong

Virginia State University – Graduate Student

Background: Why Breast Cancer Classification?

- Breast cancer is one of the most commonly diagnosed cancers.
- Early detection improves survival outcomes.
- Machine learning improves diagnostic accuracy.
- Automated classification reduces human error.

Literature Review: ML in Medical Diagnosis

- SVM widely used for biomedical classification.
- Decision Trees valued for interpretability.
- Preprocessing improves model performance.
- Studies show SVM >90% accuracy on WDBC.

Dataset Overview

- 569 samples with 30 numerical features.
- Classes: Benign (357) and Malignant (212).
- Features derived from FNA cell images.
- Goal: classify tumor as benign or malignant.

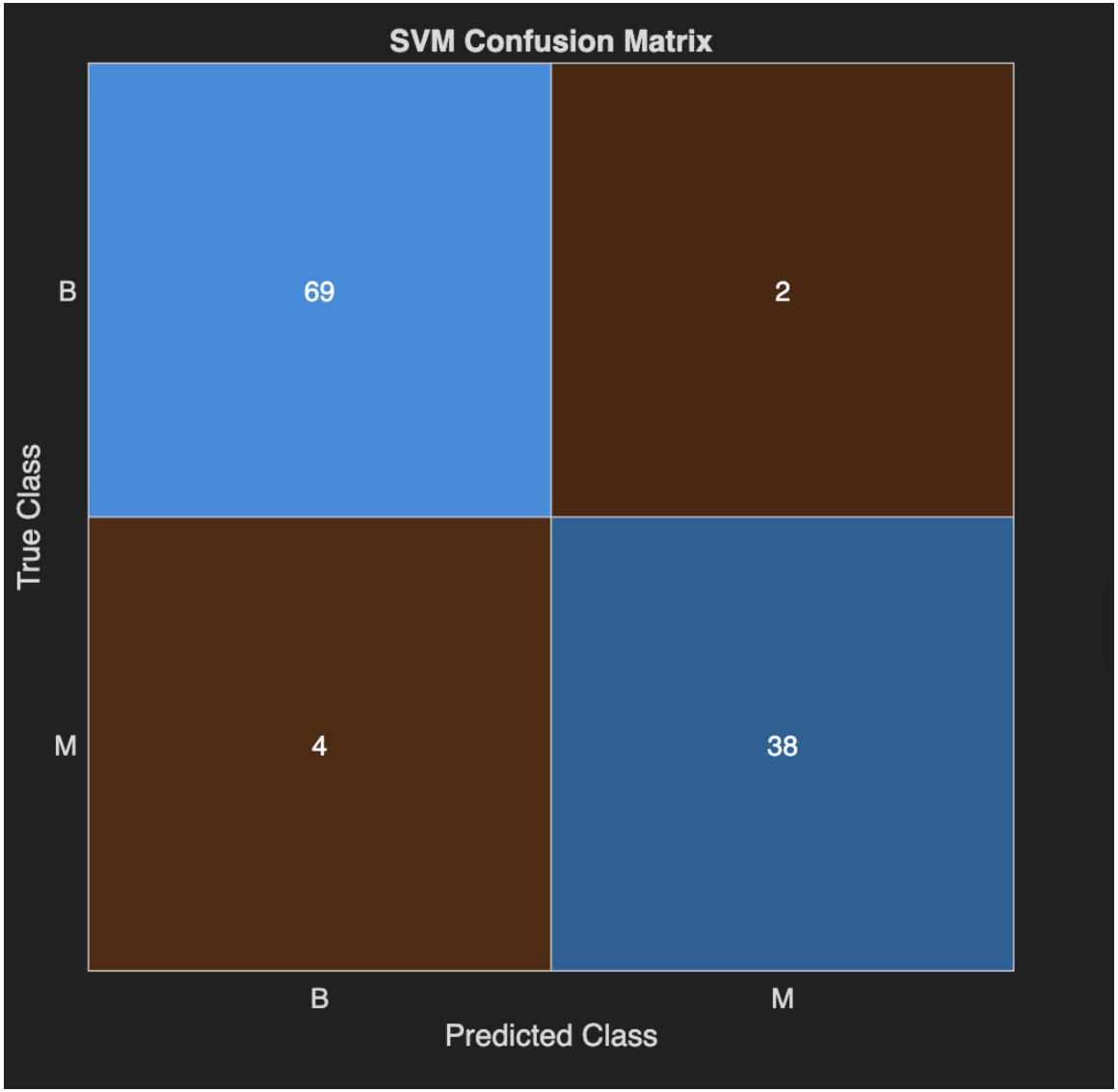
Method: Classification Pipeline

- Load and preprocess dataset.
- Encode labels (B/M).
- Standardize features for SVM.
- 80/20 stratified train-test split.
- 10-fold cross-validation for SVM.
- Train SVM and Decision Tree models.

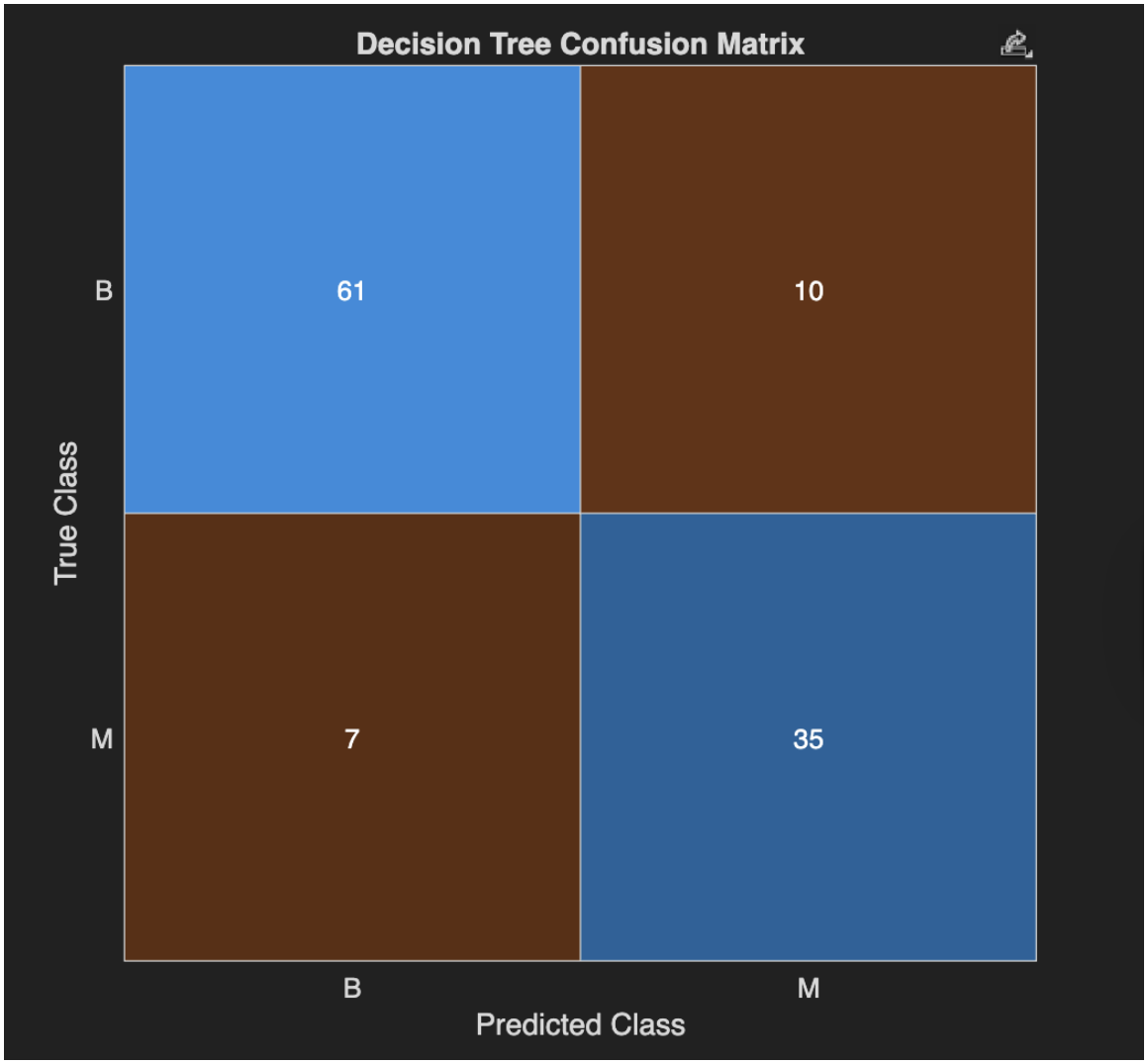
Model Training

- SVM (RBF Kernel) with tuned hyperparameters.
- Grid search: C in {0.1, 1, 10}, gamma in {0.01, 0.001}.
- Best parameters: C=1, gamma=0.01.
- Decision Tree: baseline model.
- Metrics: accuracy, precision, recall, F1, AUC.

Results: SVM Performance



Results: Decision Tree Performance



Comparison: SVM vs Decision Tree

- SVM Accuracy: 0.9469, AUC: 0.9896.
- DT Accuracy: 0.8496, AUC: 0.8266.
- SVM has lower false negatives.
- SVM is better suited for clinical diagnostics.

Conclusion & Future Work

- SVM outperforms Decision Tree on WDBC dataset.
- Handles high-dimensional features better.
- Future work: Random Forest, XGBoost, PCA.
- Deep learning for FNA image diagnosis.

Demo Instructions

- Open MATLAB script.
- Load dataset and preprocess.
- Train SVM with cross-validation.
- Train Decision Tree.
- Show confusion matrices and metrics.
- Explain performance differences.

References

- [1] Cortes & Vapnik, Support-vector networks, 1995.
- [2] Dua & Graff, UCI Machine Learning Repository, 2019.
- [3] Quinlan, Induction of Decision Trees, 1986.