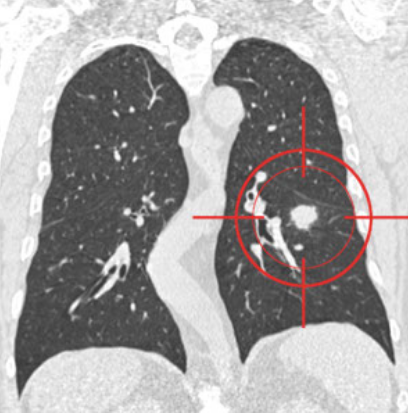


**ML Challenge 2 Milestone 3 Report**



Aleix Ramon Brugada Iglesias, 41655484T

Adrián Redondo Fernández, 48181224L

María Fernández Barrera, 46837717B

Soufiane Aatab, X6132088N

Júlia Casanovas i Saumell, 26071980P

* 24-25 MACHINE LEARNING (150400)

23/05/2025

**1. Introduction**

In Milestone 3, we develop and evaluate an SVM-based classification pipeline using slice-level Gray-Level Co-occurrence Matrix (GLCM) features to distinguish benign from malignant lung nodules. This work focuses on understanding the impact of validation schemes and hyperparameter optimization on model robustness.

**1.1 Goals**

* Establish baseline SVM performance with slice-level cross-validation.
* Assess generalization with StratifiedGroupKFold at the nodule level.
* Optimize SVM hyperparameters through both Grid and Random Search.

**2. Methodology**

We loaded 7,414 CT slices represented by 24 GLCM textural features and corresponding benign/malignant labels. The pipeline applies standard scaling, trains a Support Vector Machine classifier, and evaluates performance using slice-level and group-level cross-validation schemes. Hyperparameter tuning was conducted via Grid Search and Random Search to identify optimal regularization and kernel parameters.

**3. Experimental Design**

**3.1 Dataset Description**

The dataset comprises 7,414 slice samples with 24 GLCM-derived features. Each slice is labeled as benign or malignant, and samples are grouped by nodule ID to enable group-wise validation reflective of patient-level generalization.

**3.2 Validation Strategies**

Baseline evaluation uses 5-fold StratifiedKFold at the slice level and 5-fold StratifiedGroupKFold at the nodule level. Hyperparameter tuning employs 5-fold StratifiedKFold with exhaustive Grid Search and randomized parameter sampling via Random Search.

**4. Results**

**4.1 Baseline Classification**

Baseline accuracies under different cross-validation schemes are summarized in Table 1.

| CV Scheme | Mean Accuracy | Std Dev |
| --- | --- | --- |
| StratifiedKFold (slice-level) | 0.909 | 0.005 |
| StratifiedGroupKFold (nodule-level) | 0.935 | 0.062 |

*(Table 1: Baseline Classification results)*

**4.2 Hyperparameter Tuning**

Optimal parameters from Grid Search and Random Search are listed in Table 2. ANOVA confirms significant performance differences across parameter combinations (p < 0.001).

| Method | Best Parameters | Accuracy/sd | ANOVA (F/p-value) |
| --- | --- | --- | --- |
| Grid Search | C=10, γ=0.1 | 0.923/0.005 | 13.37/2.1893e-14 |
| Random Search | C≈4.53, γ≈0.33 | 0.925/0.009 | 2.1/2.7860e-03 |
| Optuna | C≈3.77, γ≈0.43 | 0.925/0.01 | 8.58/6.1382e-18 |

*(Table 2: Hyperparameter tuning results)*

**5. Discussion and Conclusions**

The baseline evaluation demonstrates that nodule‐level cross‐validation (StratifiedGroupKFold) yields higher average accuracy than slice‐level CV (93.5% vs. 90.9%), albeit with increased variability, underscoring the importance of assessing model generalization at the patient level. Hyperparameter tuning further improves performance: both Random Search and Optuna slightly outperform Grid Search, achieving around 92.5% accuracy compared to 92.3%, with ANOVA confirming that tuning choices produce statistically significant differences (Table 2). Overall, these results highlight two key insights: (1) group‐aware validation is critical for reliable malignancy prediction, and (2) efficient exploration of the hyperparameter space can yield modest but meaningful gains.

**6. References**

[1] Haralick, R.M., Shanmugam, K., Dinstein, I. (1973). Textural Features for Image Classification.

[2] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python.

[3] Akiba, T. et al. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework.