LINEAR PROGRAMMING AND QUADRATIC PROGRAMMING APPROACHES FOR CLASSIFICATION PROBLEMS

By: Jayques Nelson

Natural Science & Mathematics Department

Advisor: Dr Wei Wan

ABSTRACT

- **Purpose and Approach**
  The study compares linear programming (LP) and quadratic programming (QP) methods for classification across three medical tasks: tumor classification, medical triage, and COVID-19 case identification. LP focuses on minimizing classification errors, while QP (using SVM) focuses on maximizing the margin for better generalization.

- **Key Findings**
  Both LP and QP achieved 100% accuracy on all datasets, demonstrating their strong performance in linearly separable medical classification problems. However, the methods differ in how they optimize decisions: LP provides simpler and more computationally efficient solutions, while QP explicitly optimizes margins.

- **Insights and Implications**
  QP produced consistently larger margins (1.38 to 21.48), offering stronger theoretical generalization to unseen data. LP, on the other hand, is more computationally efficient and yields sparser, more interpretable models. These results highlight important trade-offs practitioners should consider when choosing between LP and QP for healthcare applications.

INTRODUCTION

- **Classification** is a core task in machine learning with major uses in medicine, finance, and industry. The goal is to find an effective decision boundary that separates different types of data.
- **Linear Programming** (LP) has been a foundation for classification since the 1960s, offering simple, interpretable models that minimize classification errors through violation variables. Its efficiency makes it suitable for real-time and large-scale applications.
- **Quadratic Programming** (QP), particularly through Support Vector Machines (SVM), introduced margin maximization as a superior strategy for generalization. By balancing error minimization with margin size, QP provides stronger performance on unseen data.

MOTIVATION FOR THE STUDY

- Although LP and QP have both proven effective independently, direct comparative studies are limited—especially within medical applications where accuracy and interpretability both matter.
- Healthcare contexts often require models that not only perform well but also provide understandable decision boundaries for clinical use.
- The introduction emphasizes the need for a systematic comparison of LP and QP performance, characteristics, and trade-offs across medical classification tasks to guide practitioners in method selection.

METHODOLOGY

RESEARCH DESIGN & APPROACH

- The study uses an experimental design comparing LP and QP across three medical classification tasks: tumor classification, medical triage assessment, and COVID-19 case identification. Each task is solved using both LP and QP under consistent evaluation metrics for accuracy, margin width, and computational time.
- Data preprocessing included feature scaling, formulation of mathematical models, and implementation using Python libraries: PuLP for LP and CVXPY for QP. Visualizations were generated using Matplotlib, and NumPy/scikit-learn were used for data handling.
- Each classification problem was mathematically formulated with LP minimizing classification errors through violation variables, while QP used the SVM framework to maximize margin size with slack variables controlling misclassification.

EVALUATION PLAN & LIMITATIONS

- The analysis incorporated both quantitative metrics (accuracy, objective values, margin width, runtime) and qualitative assessments (interpretability, implementation complexity, clinical relevance).
- Limitations included small dataset sizes, use of synthetic data for the triage problem, and the assumption of linear separability for both LP and QP models. Computational results may vary across hardware and solver configurations.
- Ethical procedures were followed by using anonymized or synthetic data, stating assumptions clearly, and ensuring reproducibility through transparent implementation and publicly accessible code.

DATASETS USED IN LP & QP CLASSIFICATION

TUMOR CLASSIFICATION DATASET

■ MALIGNANT TUMORS (R1) 10 SAMPLES × 2 FEATURES

Sample	Feature 1	Feature 2
1	3.3807	3.7832
2	1.1479	1.1443
3	2.2768	3.6643
4	3.1458	1.0959
5	4.0091	4.1817
6	4.3168	1.5749
7	2.4625	0.8847
8	0.6238	1.1299
9	0.6718	2.7722
10	1.1588	2.1298

BENIGN TUMORS (R2)
10 SAMPLES × 2 FEATURES

Sample	Feature 1	Feature 2
1	7.0825	5.8413
2	9.2387	5.7428
3	8.1337	7.8886
4	7.9738	9.0063
5	9.6274	9.7030
6	6.7863	6.0846
7	8.9074	8.0597
8	8.8918	7.6123
9	7.2120	5.5536
10	8.0552	7.0171

COVID-19 DATASET

POSITIVE CASES (8 SAMPLES)

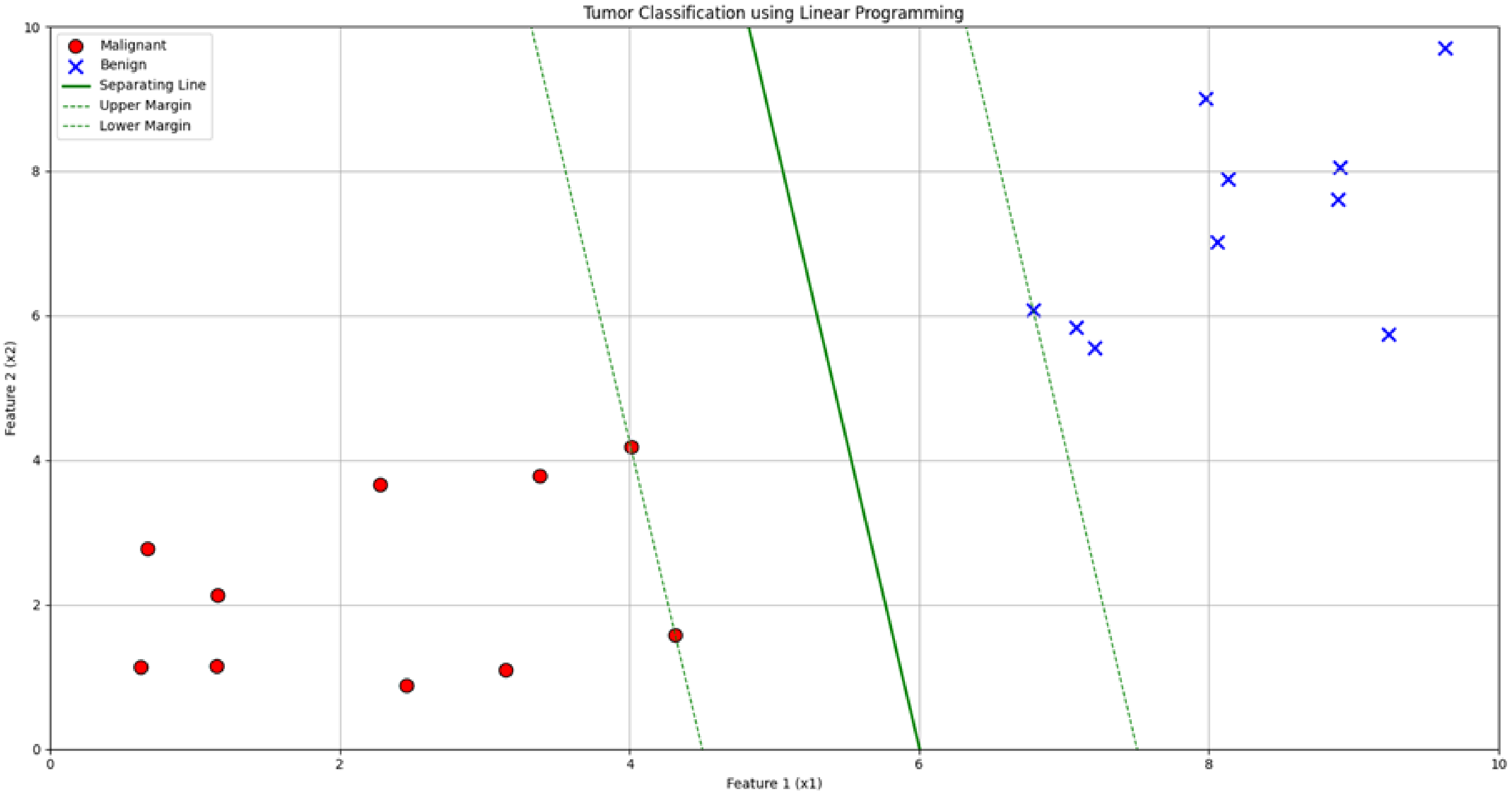
Case	Temperature	Oxygen Saturation	Age	Cough Severity
1	38.5	92	65	3
2	37.8	94	72	2
3	39.1	89	58	3
4	38.2	93	61	2
5	37.9	95	45	1
6	38.7	91	70	3
7	38.0	96	52	1
8	39.2	88	68	3

NEGATIVE CASES (8 SAMPLES)

Case	Temperature	Oxygen Saturation	Age	Cough Severity
1	36.8	98	32	0
2	36.5	99	28	0
3	37.1	97	41	1
4	36.9	98	35	0
5	37.0	99	29	0
6	36.7	98	38	1
7	37.2	97	44	1
8	36.6	99	31	0

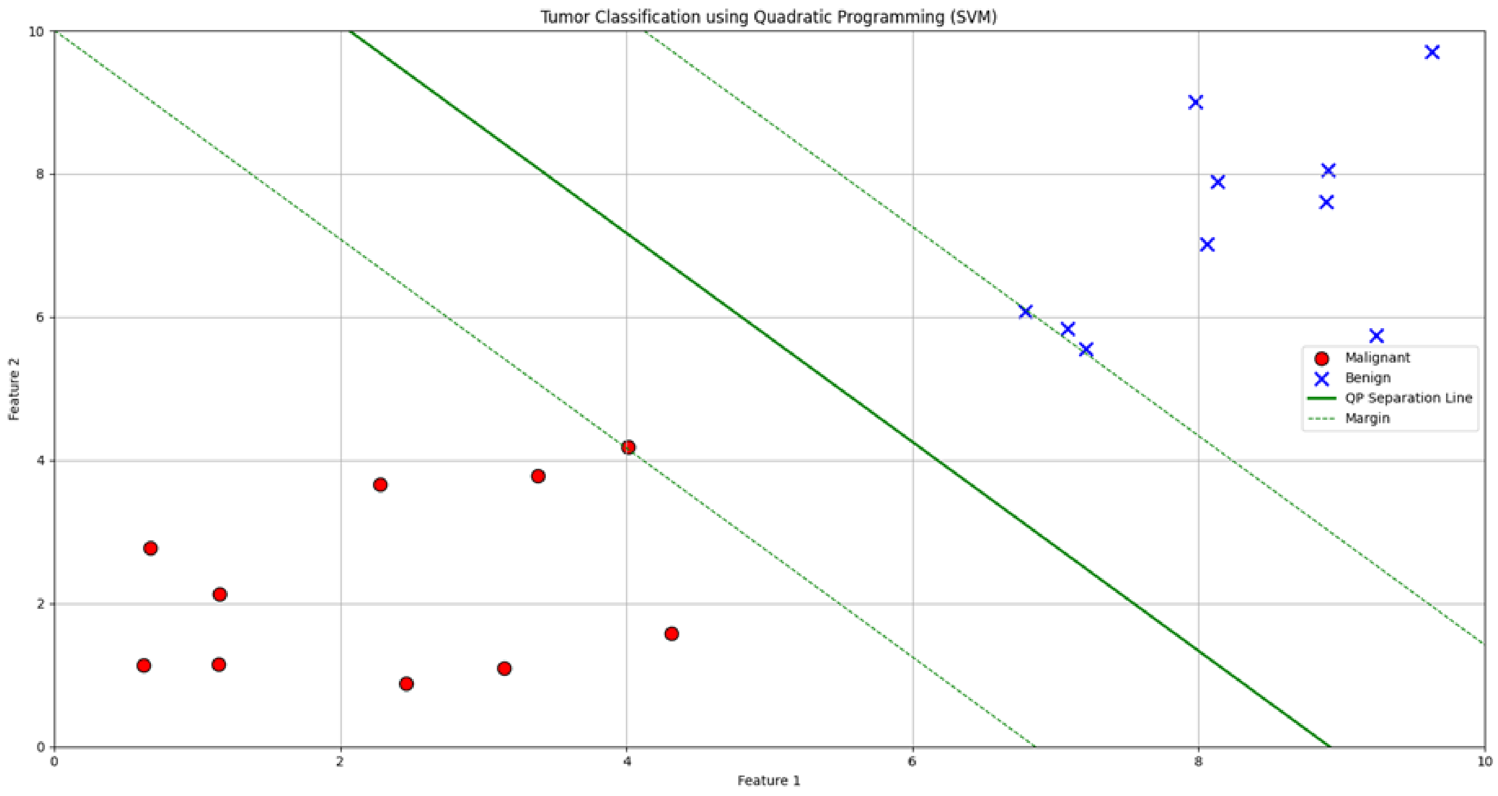
RESULTS

TUMOR CLASSIFICATION RESULTS
LINEAR PROGRAMMING APPROACH



- The LP formulation for tumor classification achieved perfect separation with the following optimal parameters:
- w_1 = -0.666264
- w_2 = -0.078644
- γ = -3.999985
- Objective Value = 0.0
- Accuracy = 100.0%

TUMOR CLASSIFICATION RESULTS
QUADRATIC PROGRAMMING APPROACH

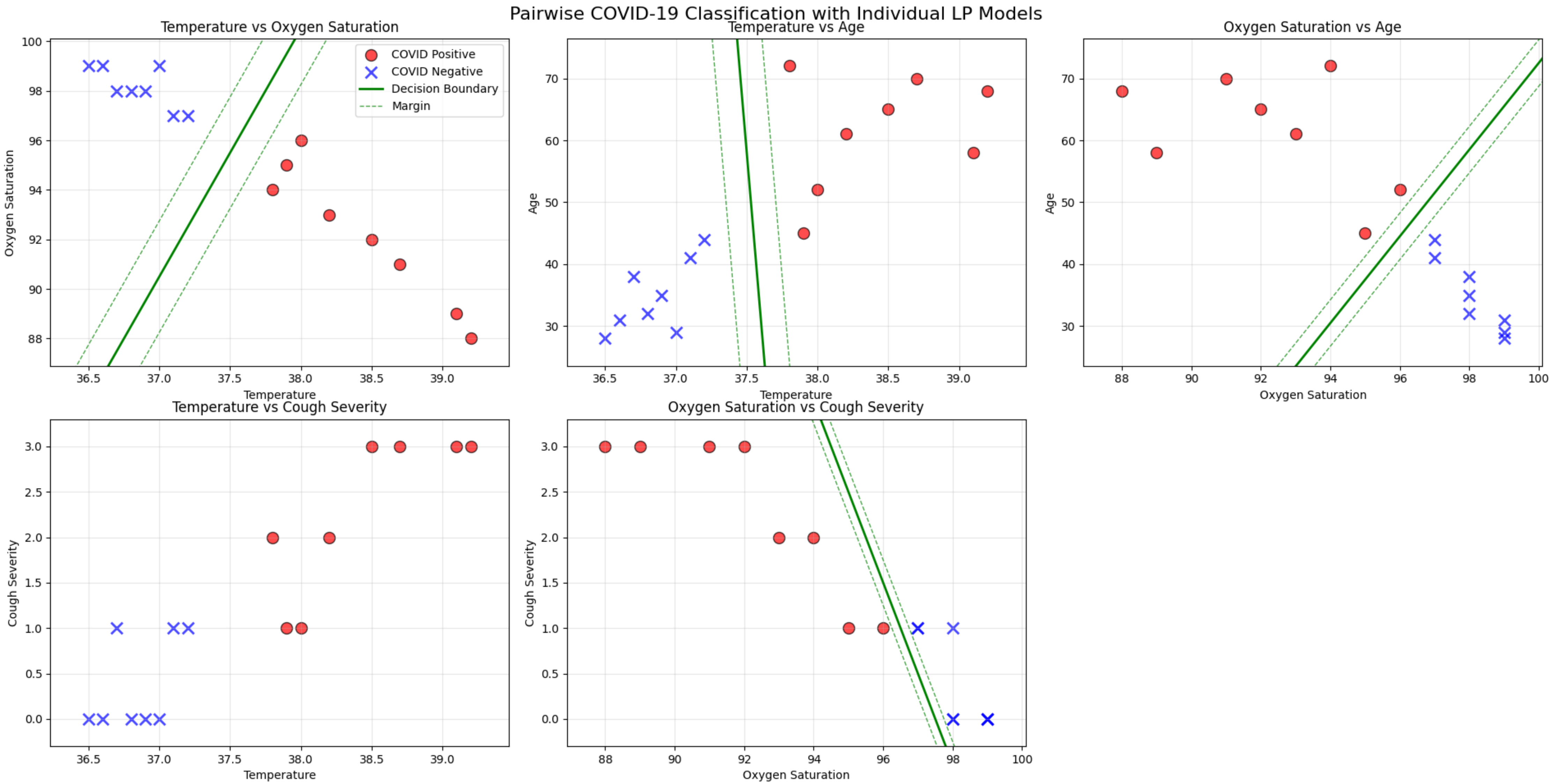


- The QP-SVM approach achieved perfect classification with margin optimization:
- $w_1 = -0.485785$
- $w_2 = -0.332853$
- $\gamma = 4.330706$
- Margin Width = 2.894
- Accuracy = 100.0%

SUMMARY OF COVID-19 LP CLASSIFICATION MODELS

Feature Pair	w_1	w_2	γ	Objective Value
Temperature - Oxygen Saturation	1.195220	-0.711248	0.374661	0.0
Temperature - Age	1.245036	0.378898	0.187856	0.0
Oxygen Saturation - Age	-1.070702	0.885370	0.466515	0.0
Temperature - Cough Severity	1.403793	0.378607	0.296649	0.0
Oxygen Saturation - Cough Severity	-1.411151	0.194684	0.399593	0.0

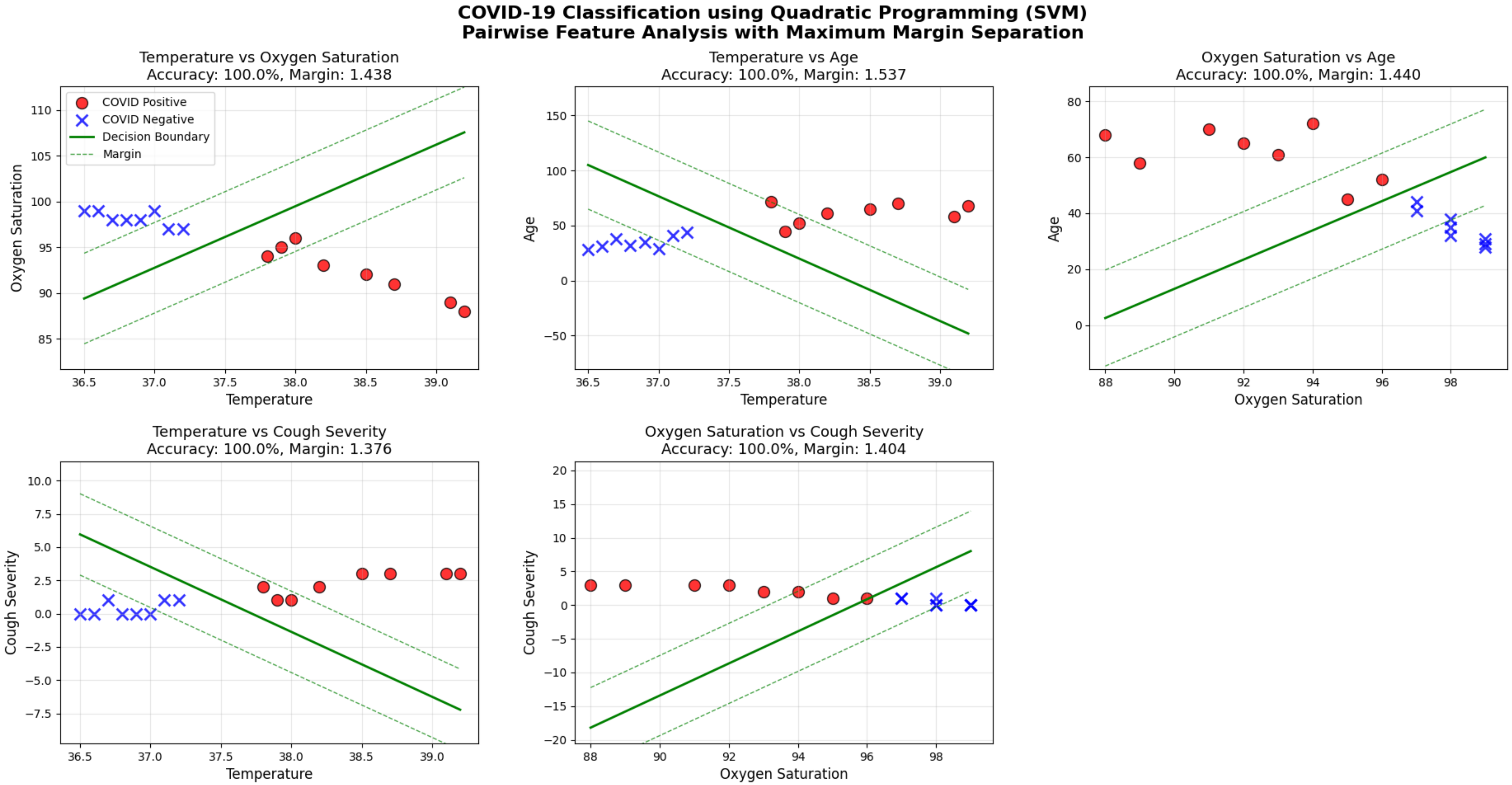
COVID-19 CLASSIFICATION RESULTS FOR FIVE FEATURE PAIRS USING LINEAR PROGRAMMING



SUMMARY OF COVID-19 QP CLASSIFICATION MODELS

Feature Pair	w_1	w_2	γ	Accuracy	Margin Width
Temperature - Oxygen Saturation	1.195220	-0.711248	0.374661	100.0%	1.4380
Temperature - Age	1.245036	0.378898	0.187856	100.0%	1.5368
Oxygen Saturation - Age	-1.070702	0.885370	0.466515	100.0%	1.4395
Temperature - Cough Severity	1.403793	0.378607	0.296649	100.0%	1.3756
Oxygen Saturation - Cough Severity	-1.411151	0.194684	0.399593	100.0%	1.4040

COVID-19 CLASSIFICATION RESULTS FOR FIVE FEATURE PAIRS USING QUADRATIC PROGRAMMING



DISCUSSION & CONCLUSIONS

KEY FINDINGS & PERFORMANCE

- **Both LP and QP achieved perfect accuracy (100%)** across all medical classification tasks (Tumor, Triage, COVID-19).
- **LP is computationally faster** (approx. 2-3x) than QP, making it ideal for real-time or large-scale applications.
- **QP provides explicit margin optimization**, with margins ranging from 1.38 to 21.48, offering stronger theoretical guarantees for generalization on unseen data.
- **Both models are highly interpretable**, allowing clinicians to validate feature importance.

RECOMMENDATIONS & GUIDELINES

Choose Linear Programming (LP) when:

- Computational speed and efficiency are critical.
- Model simplicity and interpretability are top priorities.
- Dealing with large-scale datasets or real-time classification.
- Resources for parameter tuning are limited.

Choose Quadratic Programming (QP/SVM) when:

- Generalization performance on future data is the main goal.
- Robust separation and margin maximization are important.
- Theoretical performance guarantees are desired.
- Sufficient computational resources are available.

OVERALL CONCLUSION

- **No single "best" method**—the choice is a **trade-off**.
- **LP excels** in **efficiency, simplicity, and interpretability**.
- **QP excels** in **generalization and robust performance**.
- Both are powerful, mathematically sound approaches that can significantly improve medical classification systems and healthcare decision-making.

THANK YOU FOR LISTENING

Any Questions?