Supplementary Information

Geometric and Quantum Kernel Methods for Predicting Skeletal Muscle Outcomes in chronic obstructive pulmonary disease

Azadeh Alavi, Hamidreza Khalili, Stanley H. Chan,
Fatemeh Kouchmeshki, Muhammad Usman, Ross Vlahos

Corresponding authors: Azadeh Alavi (azadeh.alavi@rmit.edu.au)
and Stanley M. H. Chan (stanley.chan@rmit.edu.au)

This supplementary file reports a focused sensitivity analysis within the SPD branch for the muscle-weight task using the same three biomarkers as in the main-text biomarker-only comparison. The purpose of Supplementary Table S1 is to show how performance changes when the *SPD representation* is changed while the outcome, biomarker set, downstream regressor, and outer cross-validation splits are kept fixed. The first row is the plain ridge-regression baseline using only the original biomarkers. The remaining rows augment ridge regression with SPD prototype-distance features derived from alternative SPD constructions. Differences between rows therefore isolate the contribution of the SPD representation itself rather than changes in the data split or prediction model. The row retained in the main manuscript is the best-performing outer-product SPD configuration, whereas the local-covariance row is included as an exploratory alternative. All values are averaged over the same five outer folds used in the main manuscript. All runs shown here are "no synthetic" variants, meaning that unlabeled synthetic SPD interpolation was switched off so that this table isolates the effect of the base SPD construction. Lower RMSE and %RMSE, and higher R^2, indicate better performance. ROC-AUC was not computed for these secondary ablations because they were used as regression-focused sensitivity analyses within the SPD family rather than as headline screening comparisons.

Supplementary Table S1: Sensitivity analysis of SPD representations for muscle-weight prediction using three biomarkers. All models were evaluated on the same five outer folds used in the main manuscript. The ridge baseline uses only the original biomarkers. The SPD rows augment ridge regression with distances from each sample to training-derived SPD prototypes under different SPD constructions. The table therefore tests whether the gain over baseline arises from the SPD representation itself and which SPD variant is most effective in this task. Percent RMSE denotes $100 \times \mathrm{RMSE}/\bar{y}_{\mathrm{test}}$ computed within each held-out fold and then averaged. Lower RMSE and %RMSE, and higher R^2, indicate better performance.

Model	RMSE	%RMSE	R^2
Ridge baseline (biomarkers only)	4.7873	10.60	0.3517
Ridge + SPD distances (outer-product, K=3, no synthetic)	4.6180	10.21	0.3962
Ridge + SPD distances (outer-product; best, no synthetic)	**4.5509**	**10.08**	**0.4123**
Ridge + SPD distances (local covariance, K=6, k=8, no synthetic)	4.8437	10.70	0.33768

Taken together, the table shows that the outer-product SPD representations improve on the biomarker-only ridge baseline, with the best outer-product configuration reducing RMSE from 4.7873 to 4.5509 and increasing R^2 from 0.3517 to 0.4123. In contrast, the local-covariance SPD variant does not improve this particular three-biomarker muscle-weight task, which supports the choice of the outer-product SPD representation in the main manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the National Health and Medical Research Council of Australia under Project Grant Numbers APP1084627 and APP1138915.

Additional Information

Correspondence and requests for materials should be addressed to Azadeh Alavi (azadeh.alavi@rmit.edu.au) and stanley.chan@rmit.edu.au.