

Radiomics-Based Muscle Quality Assessment from Multi-Site Multi-Vendor MRI

Ozkan Cigdem^{1*}, Eros Montin¹, Salim Bin Ghouth¹,
Akshay Chaudhari², Donnie Cameron^{3,4},
Christopher M. Bergeron⁴, Jan Bergeron⁴, Luigi Ferrucci⁴,
Garry E. Gold², Riccardo Lattanzi¹, Cem M. Deniz¹,
Valentina Mazzoli¹

^{1*}Department of Radiology, New York University Grossman School of
Medicine, New York, United States.

²Department of Radiology, Stanford University, Stanford, United States.

³Department of Medical Imaging, Radboudumc, Nijmegen, the
Netherlands.

⁴Intramural Research Program, National Institute on Aging, National
Institutes of Health, Maryland, United States.

*Corresponding author(s). E-mail(s): ozkan.cigdem@nyulangone.org;

S.1. Detailed Inclusion and Exclusion Criteria for Study Cohorts

Across all datasets, participants were required to: (i) be able to provide informed consent; (ii) complete MRI scanning without sedation and complete the strength testing; and (iii) have no contraindications to MRI, including MR-incompatible implants, metallic foreign bodies, claustrophobia, or body size exceeding scanner limits. Individuals were excluded if they met any of the general medical criteria known to affect muscle function or MRI safety, including diabetes, recent corticosteroid use, active hormonal supplementation, clinical diagnoses of cancer, severe osteoarthritis, or any condition deemed unsafe by the study investigators. Additional exclusions included pregnancy and any condition that prevented lying supine for approximately one hour.

Dataset-1 (GESTALT-genetic and epigenetic signatures of translational aging laboratory testing study) [1] followed eligibility rules similar to Dataset-2 but additionally

required participants to have no significant medical history or clinical evidence of genetic, autoimmune, cardiovascular, renal, hepatic, musculoskeletal, or neurological disease. Participants had to perform normal activities of daily living and walk independently for at least 400 meters without symptoms. Chronic drug therapy was not permitted, except for well-controlled antihypertensive monotherapy.

Dataset-2 [2] recruited healthy adults aged ≥ 20 years without muscle pathology. Participants were excluded if they had impaired coagulation. Additional exclusions included pregnancy, recent muscle injury or surgery, osteoporosis, and osteoarthritis.

Dataset-3 included older adults aged ≥ 70 years and classified participants as healthy, pre-frail, or sarcopenic using standardized frailty [3] and sarcopenia [4] assessments.

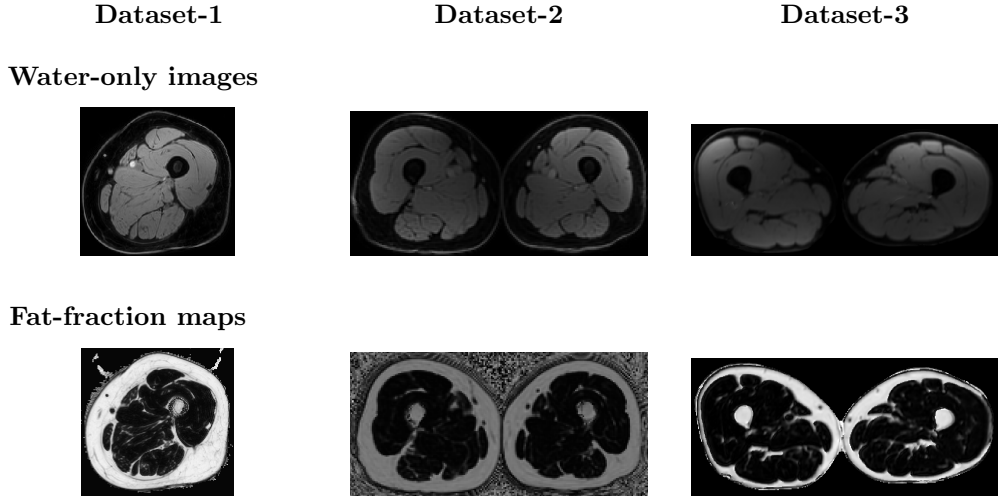


Fig. 1: Representative examples of water-only images and fat-fraction maps from female healthy controls in Dataset-1, Dataset-2, and Dataset-3, aged 56, 56, and 73 years, respectively.

S.2. Model Training

For each cross-validation fold, a multi-stage radiomics feature selection and model optimization procedure was applied on the training subset. First, non-informative features were removed using a median absolute deviation (MAD) filter. Next, univariate feature scoring was performed using a random forest regressor. Each feature was standardized and evaluated individually, where its predictive score was defined as the mean absolute Pearson correlation between predicted and true target values. Features with average score ≥ 0.15 were retained. Redundant predictors were subsequently eliminated via pairwise correlation pruning, discarding the lower-scoring member of any feature pair exceeding a correlation of 0.90. Among the correlated features, the

one with a higher Pearson correlation score was selected. Then, all features were used in XGBoost regressor training step to select the top 10-selected features using only training data.

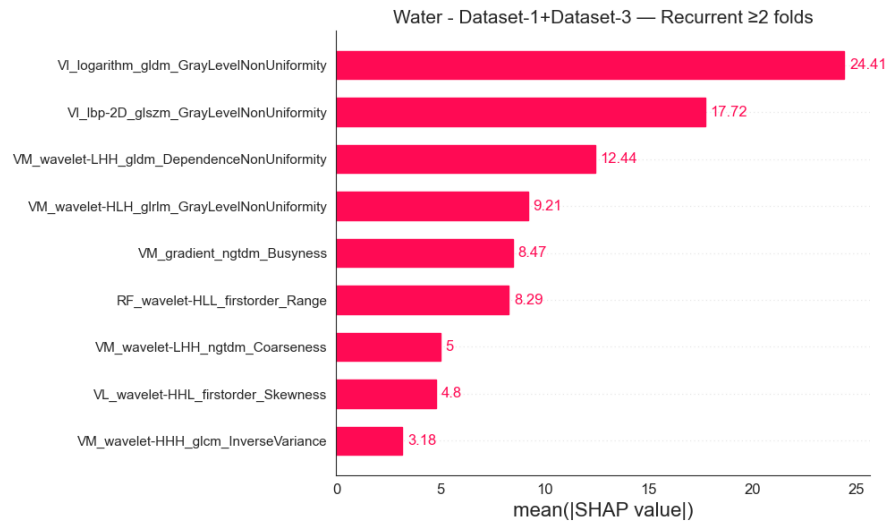
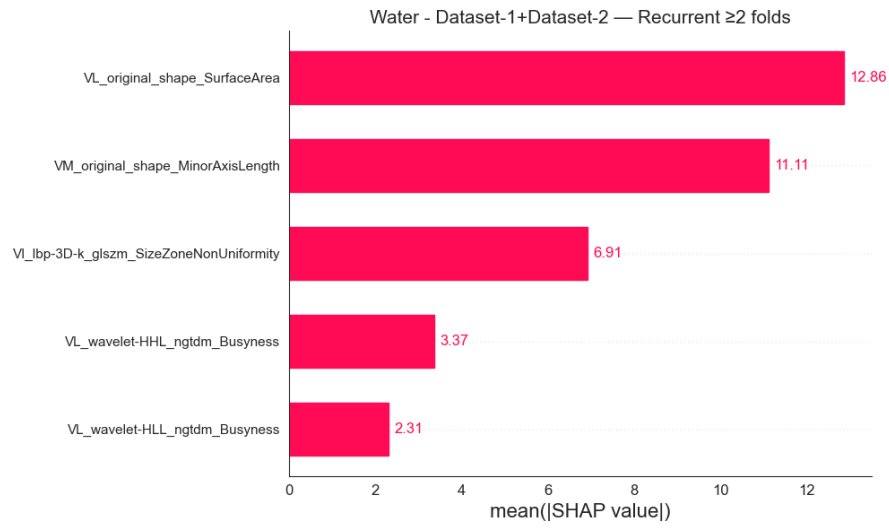
Model fitting used the Optuna hyperparameter optimization [5]. An XGBoost regressor with objective of squared error was optimised over its hyperparameters, including `max_depth` (2–6), `learning_rate` (10^{-3} –0.3), `n_estimators` (200–2000), `subsample` (0.6–0.95), `colsample_bytree` (0.3–0.7), `reg_alpha` (0–5), `reg_lambda` (0–5), and `min_child_weight` (5–15). Within each fold of the 5-fold CV, the model was trained on the training split and hyperparameters were selected based on validation performance. The best model for each fold was retrained on the full training set using the selected hyperparameters. The final performance of the ensemble models were reported on the independent hold-out test set using mean R^2 , RMSE, MAE, and MSE.

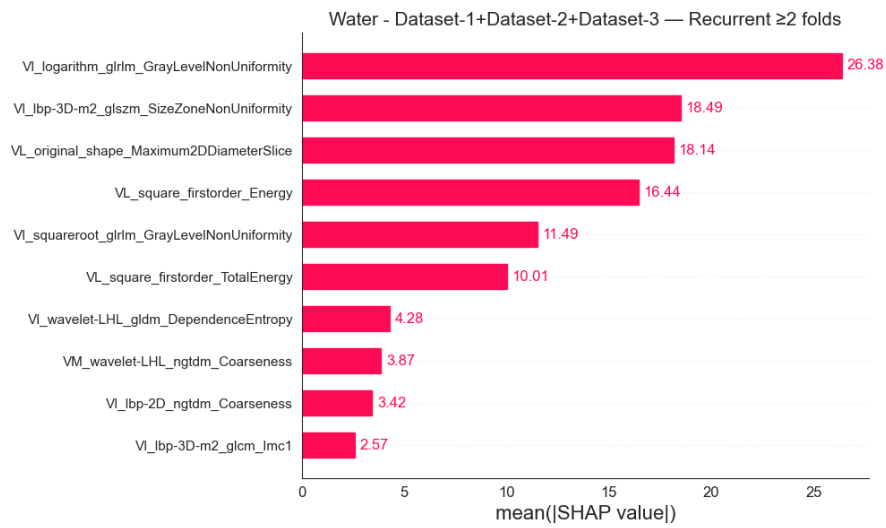
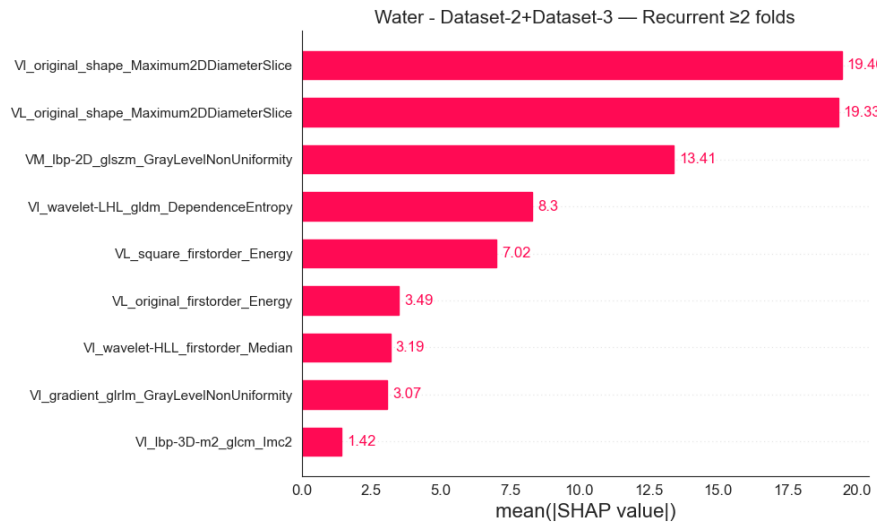
S.3.Selected Features

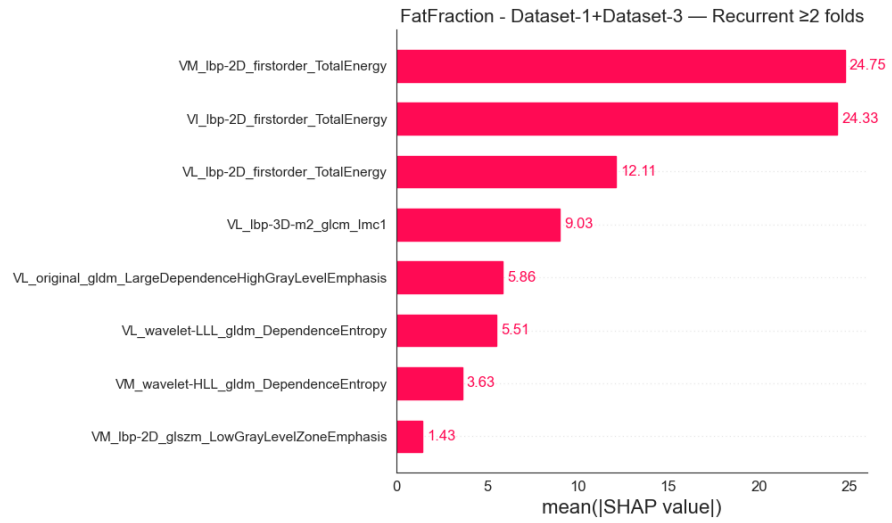
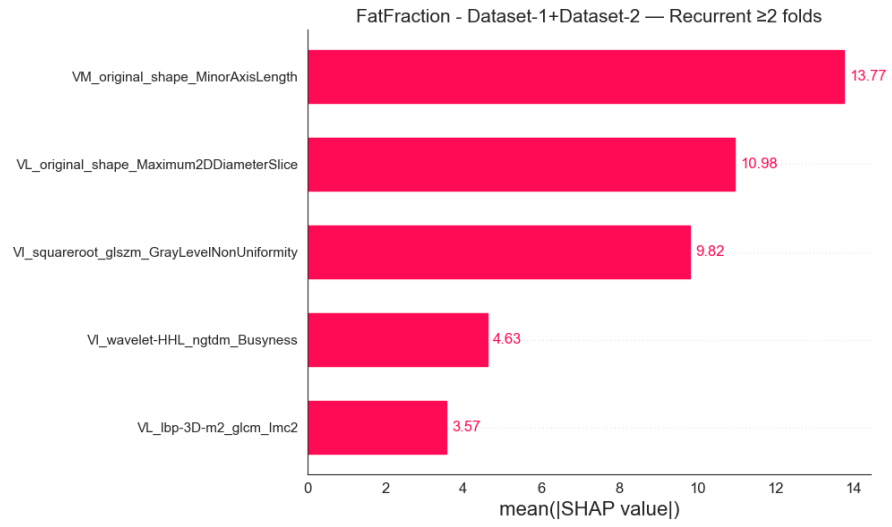
Across all dataset combinations, SHAP analyses provided in Figure 2 indicated that both muscle quantity-related features (shape descriptors) and muscle quality-related features (texture radiomics) contributed to strength prediction, with their relative importance varying by dataset composition and image contrast. Overall, shape descriptors were more often associated with higher predicted strength, consistent with the role of preserved muscle morphology in force generation, whereas texture and heterogeneity features were more frequently associated with lower predicted strength. In particular, features reflecting nonuniformity, entropy, coarseness, and related spatial heterogeneity tended to show negative associations, suggesting that a more heterogeneous imaging phenotype may correspond to reduced muscle quality. Among the Water-only models, Dataset-1+Dataset-3 showed a mixed pattern, whereas Dataset-2+Dataset-3 and especially Dataset-1+Dataset-2+Dataset-3 were more strongly influenced by negatively directed heterogeneity features. Fat fraction-only models showed a similar but more variable pattern, while the combined water+fat fraction models demonstrated a broader feature profile but remained dominated by texture- and heterogeneity-related descriptors. Overall, although the balance between morphology- and texture-related predictors varied across dataset combinations, the general pattern was consistent: higher predicted strength was associated with more preserved morphology and lower internal heterogeneity.

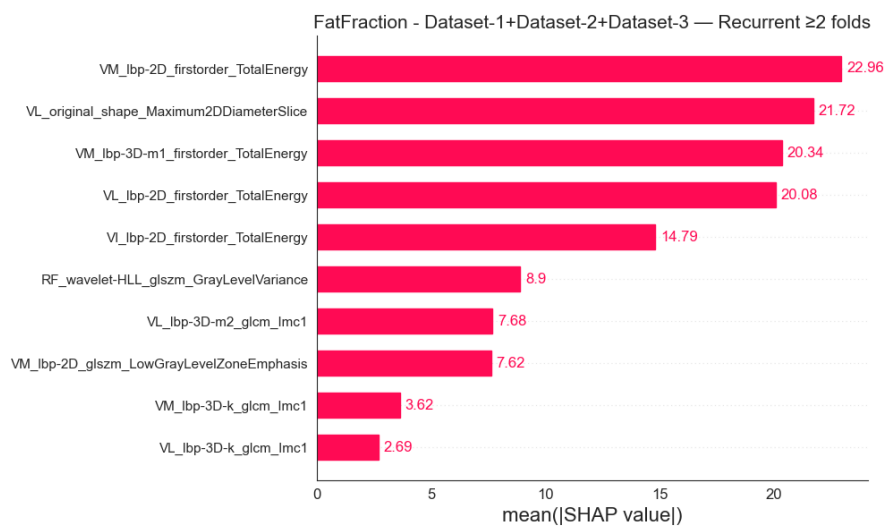
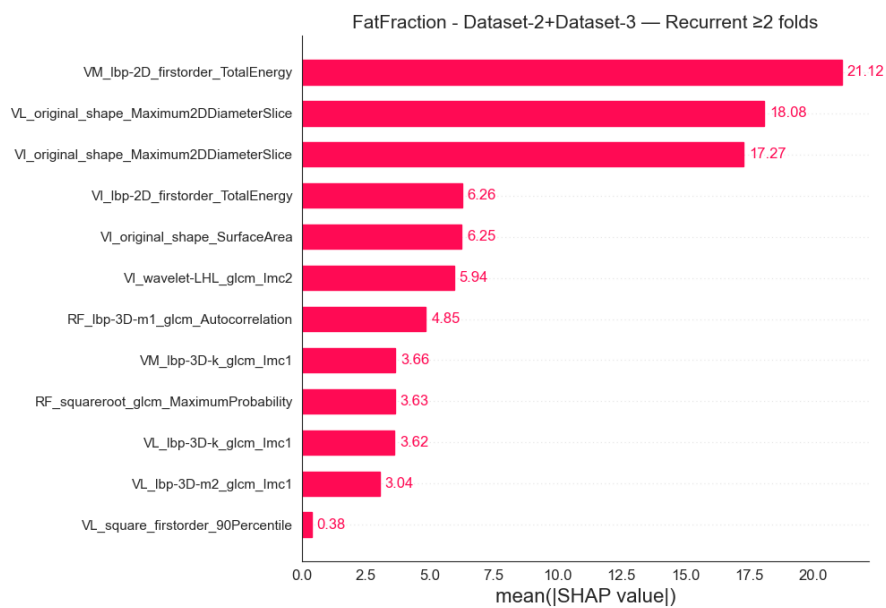
S.4.Fat Fraction Analysis

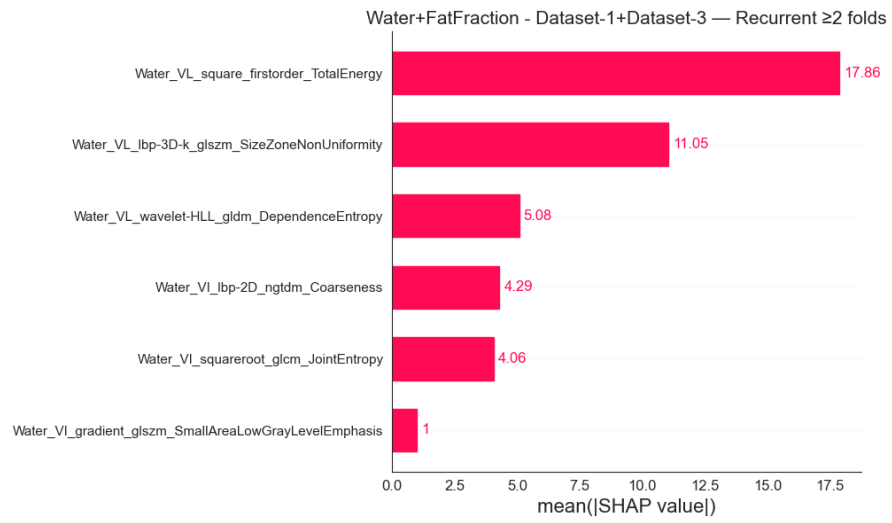
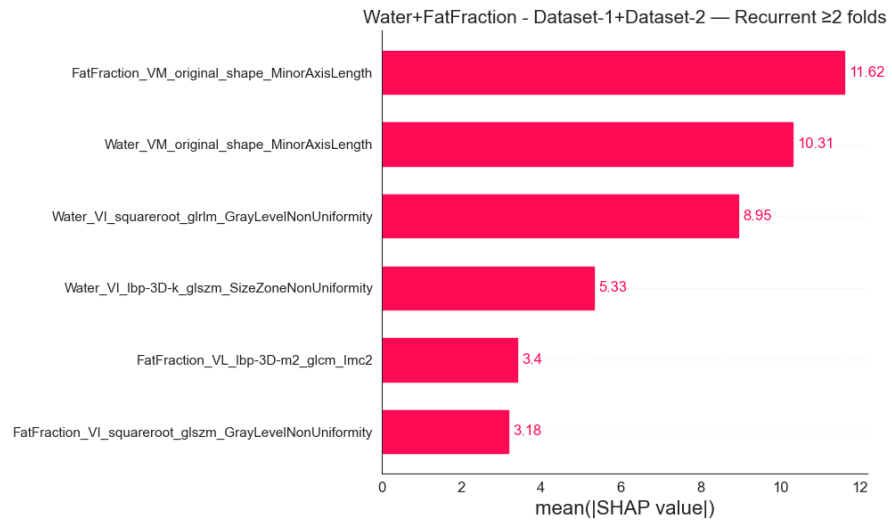
Figure 4 shows that mean fat fraction exhibits a weak-to-moderate negative association with torque across datasets, with noticeable variability between cohorts. Similarly, Figure 3 shows the CSA and torque relationship. Correspondingly, Table 1 compares prediction models and demonstrates that incorporating fat fraction modestly improves performance in Dataset-1 + Dataset-2 but does not substantially enhance accuracy in the other dataset combinations.











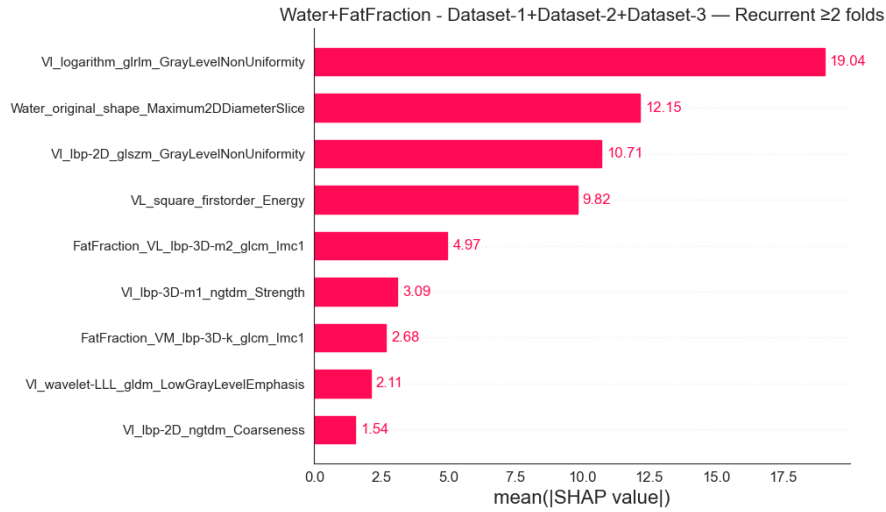
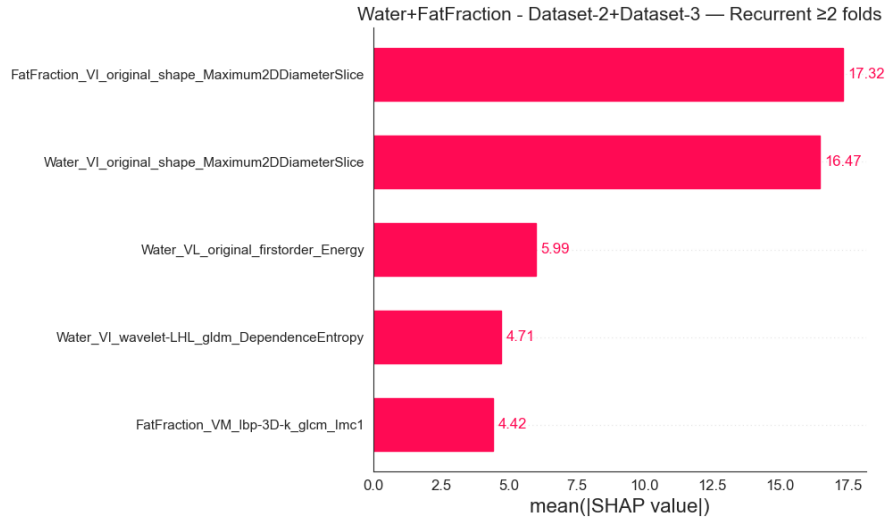


Fig. 2: Mean absolute SHAP value plots showing features that recurred in at least two of five cross-validation folds and their global importance in models built using Water, fat fraction, and combined water+fat fraction radiomic features across different dataset combinations (Dataset-1, Dataset-2, and Dataset-3). Higher values indicate a greater overall contribution of each feature to model predictions.

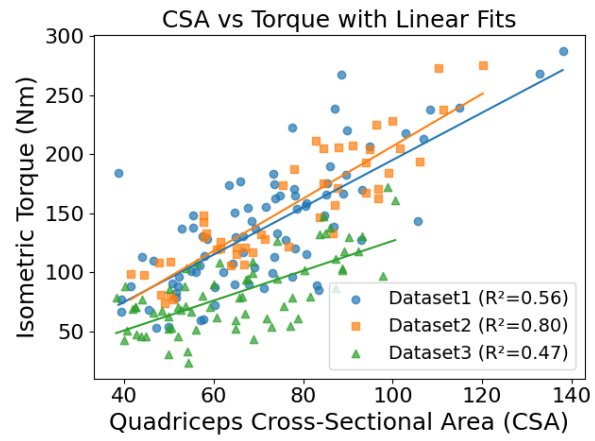
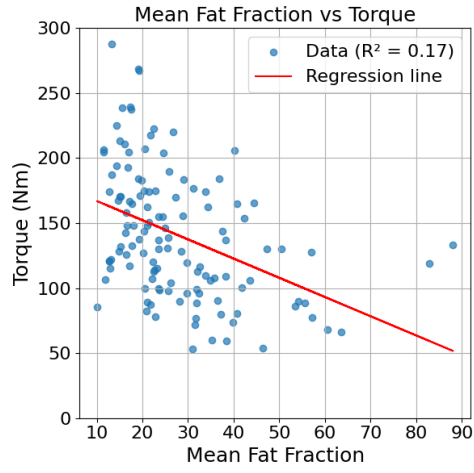
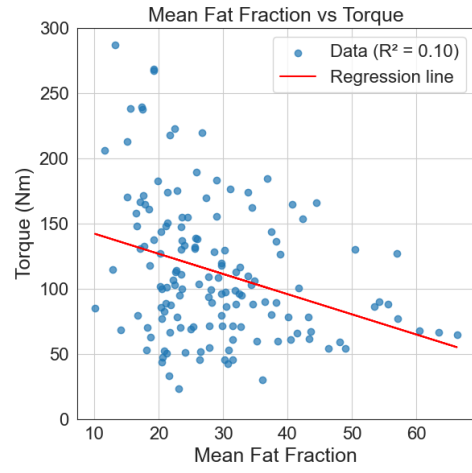


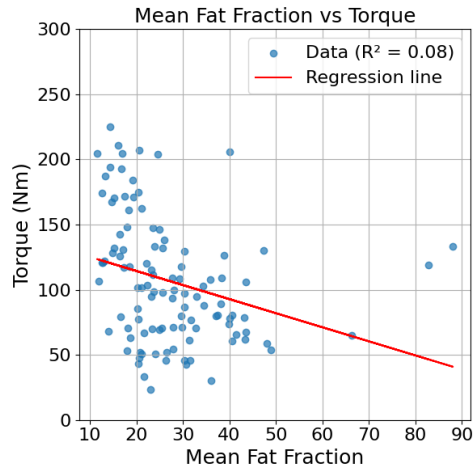
Fig. 3: Relationship between cross-sectional area and isometric knee-extension torque across individual datasets. A linear regression line and corresponding R^2 value are shown for each plot.



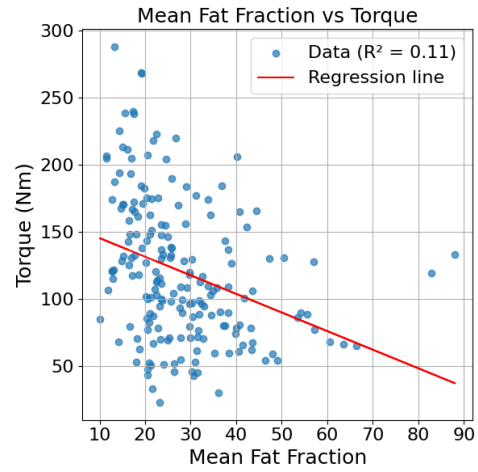
(a)



(b)



(c)



(d)

Fig. 4: Relationship between mean fat fraction and isometric knee-extension torque across dataset combinations: (a) Dataset-1+Dataset-2, (b) Dataset-1+Dataset-3, (c) Dataset-2+Dataset-3, and (d) Dataset-1+Dataset-2+Dataset-3. A linear regression line and corresponding R^2 value are shown for each plot.

Table 1: Comparison of strength prediction performance using cross-sectional area (CSA) from MRI masks and a combined model incorporating CSA and fat fraction values.

Datasets	Model	Inputs	RMSE	MAE	R^2
Dataset-1+Dataset-2	Linear Regression	CSA	30.7	25.8	0.58
		CSA+Fat Fraction	28.9	23.0	0.63
Dataset-1+Dataset-3	Linear Regression	CSA	41.7	33.4	0.24
		CSA+Fat Fraction	42.6	35.8	0.21
Dataset-2+Dataset-3	Linear Regression	CSA	32.4	26.4	0.55
		CSA+Fat Fraction	32.5	27.4	0.55
Dataset-1+Dataset-2+Dataset-3	Linear Regression	CSA	37.9	29.5	0.42
		CSA+Fat Fraction	37.9	30.2	0.42

S.6. Residual Error Characteristics and Agreement Analysis Details

The Figure 5 and Figure 6 show Bland–Altman plots comparing subject-level strength predictions from the CSA-only model and the radiomics-based model across all dataset combinations for water-only and combined water+fat fraction features, respectively. Across models and dataset combinations, mean bias remained small, while the width of the limits of agreement varied according to imaging input and cohort composition. In general, radiomics-based models showed narrower limits of agreement than CSA in dataset combinations including Dataset-3, indicating improved agreement with the reference strength values in older cohort.

The Figure 7, Figure 8, and Figure 9 present boxplots of absolute prediction error for CSA and radiomics models across dataset combinations. These plots complement the Bland–Altman analysis by illustrating the distribution of subject-level absolute errors. Consistent with the main results, radiomics models generally yielded lower absolute errors in dataset combinations including Dataset-3, with the clearest reductions observed for fat fraction-based models in Dataset-1+Dataset-3 and the full combined cohort. In contrast, differences were small in the Dataset-1+Dataset-2 combination.

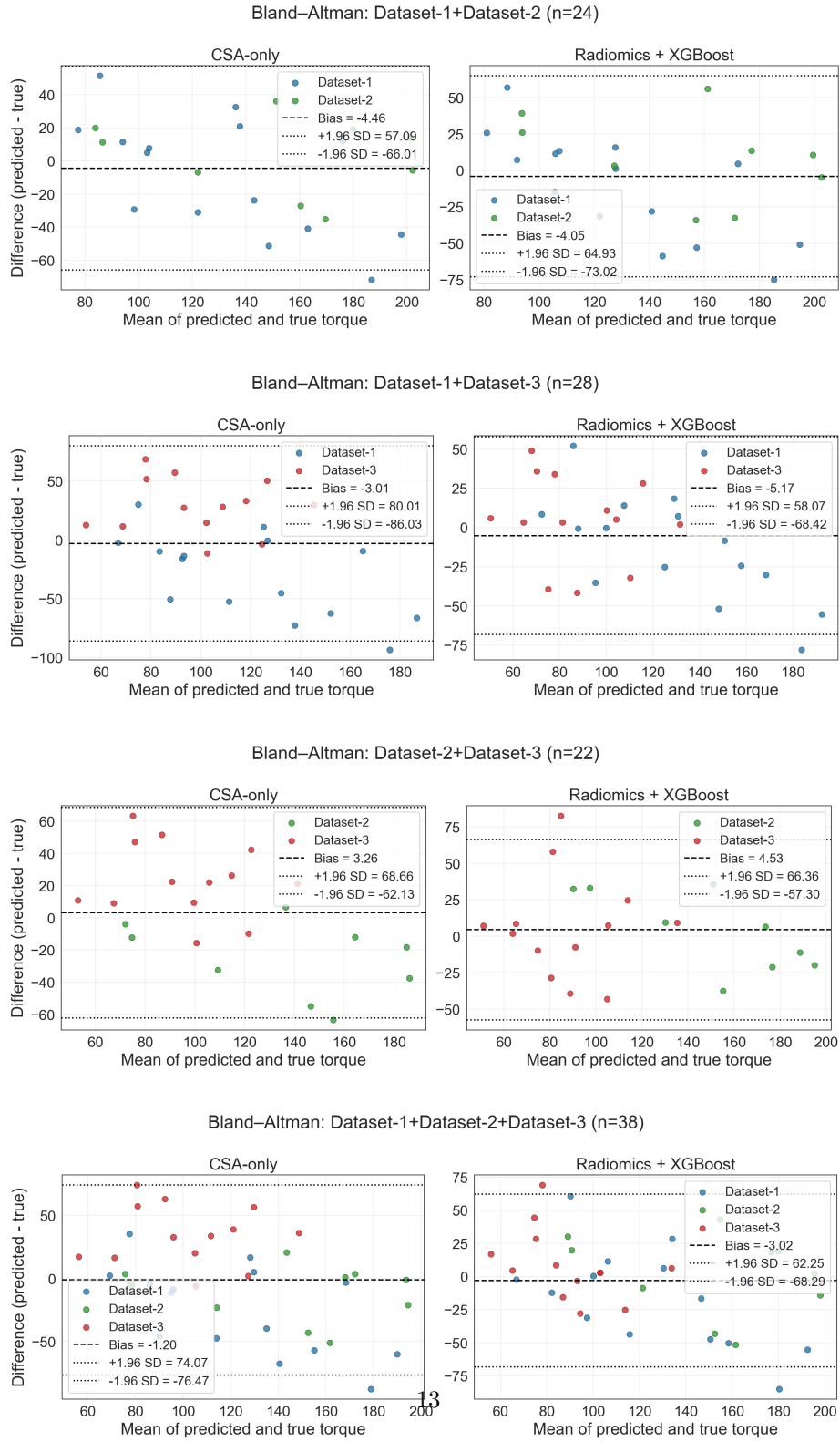


Fig. 5: Bland–Altman plots comparing strength predictions from a CSA-only model (linear mixed model) and a radiomics-based model XGBoost across different dataset combinations using water-only images.

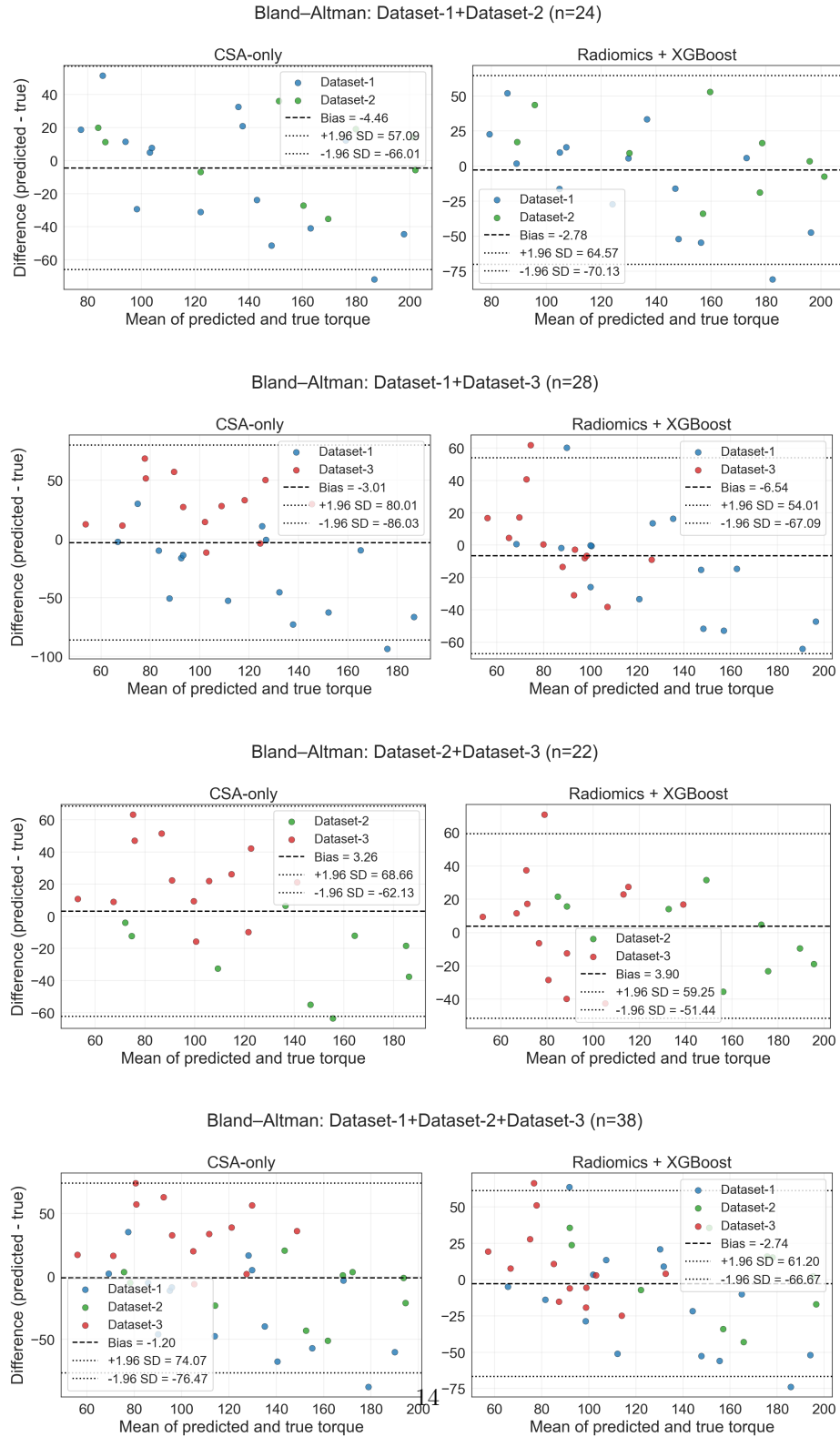


Fig. 6: Bland–Altman plots comparing strength predictions from a CSA-only model (linear mixed model) and a radiomics-based model XGBoost across different dataset combinations using combined water and fat fraction radiomics features.

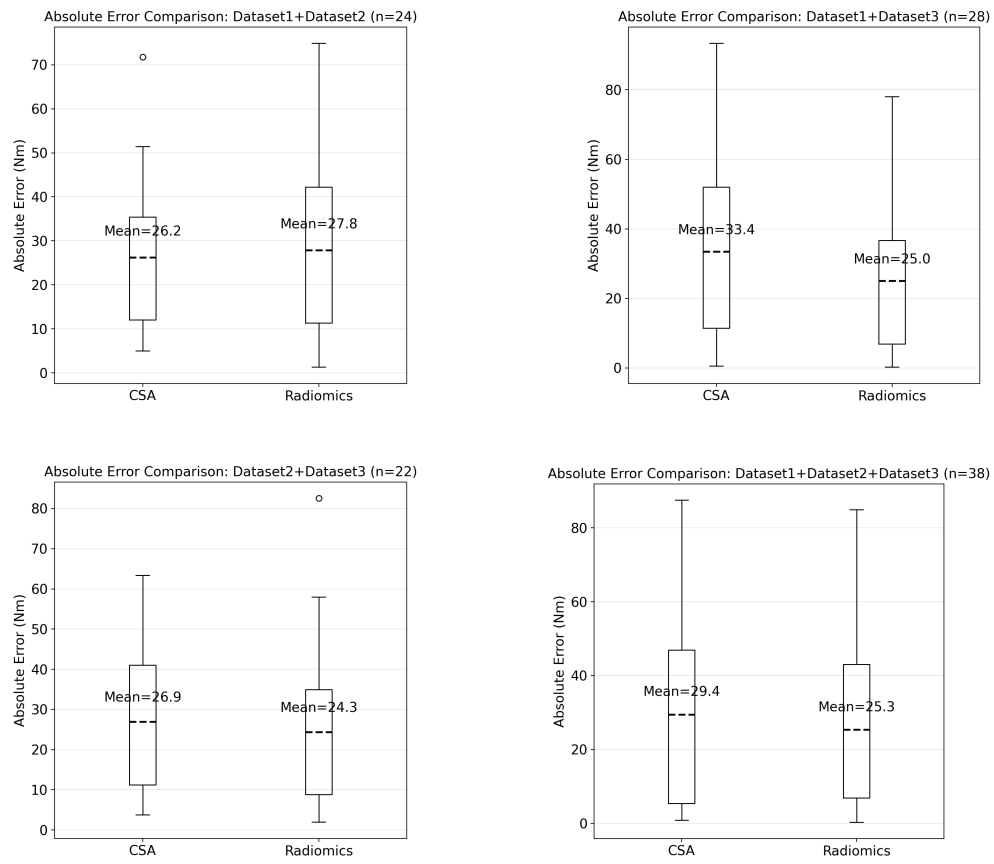


Fig. 7: Bland–Altman and absolute error plots comparing strength predictions from a CSA-only model (linear mixed model) and a radiomics-based model XGBoost across different dataset combinations using water-only images.

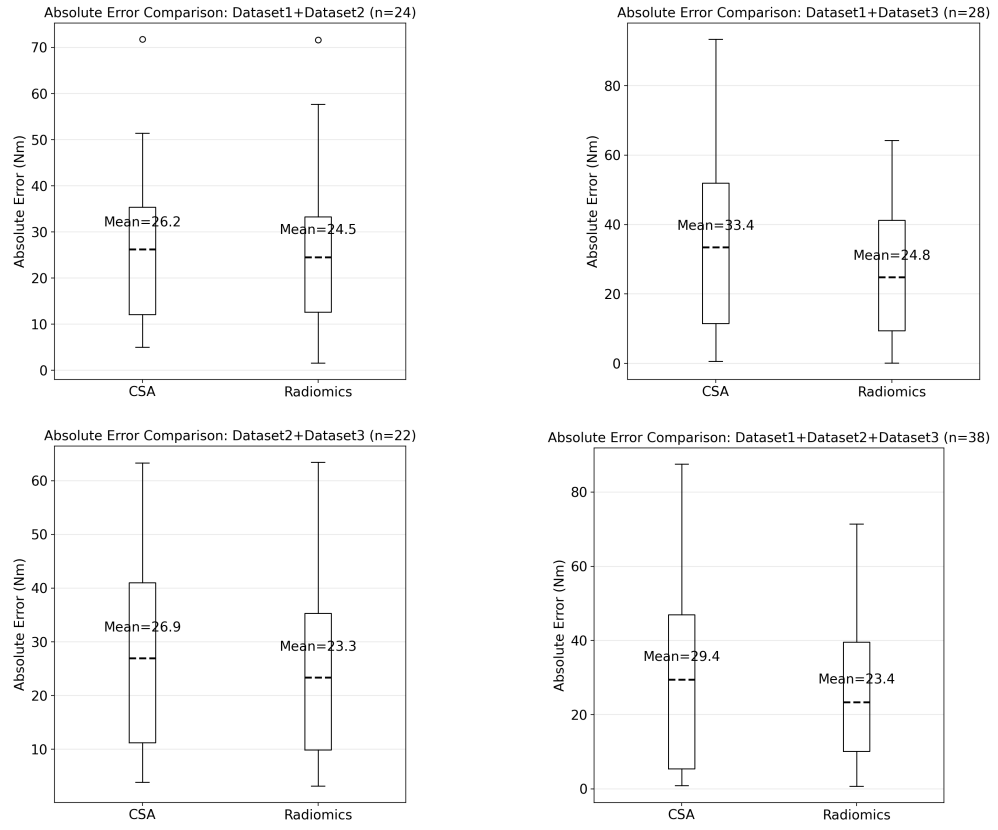


Fig. 8: Bland–Altman and absolute error plots comparing strength predictions from a CSA-only model (linear mixed model) and a radiomics-based model XGBoost across different dataset combinations using fat fraction maps.

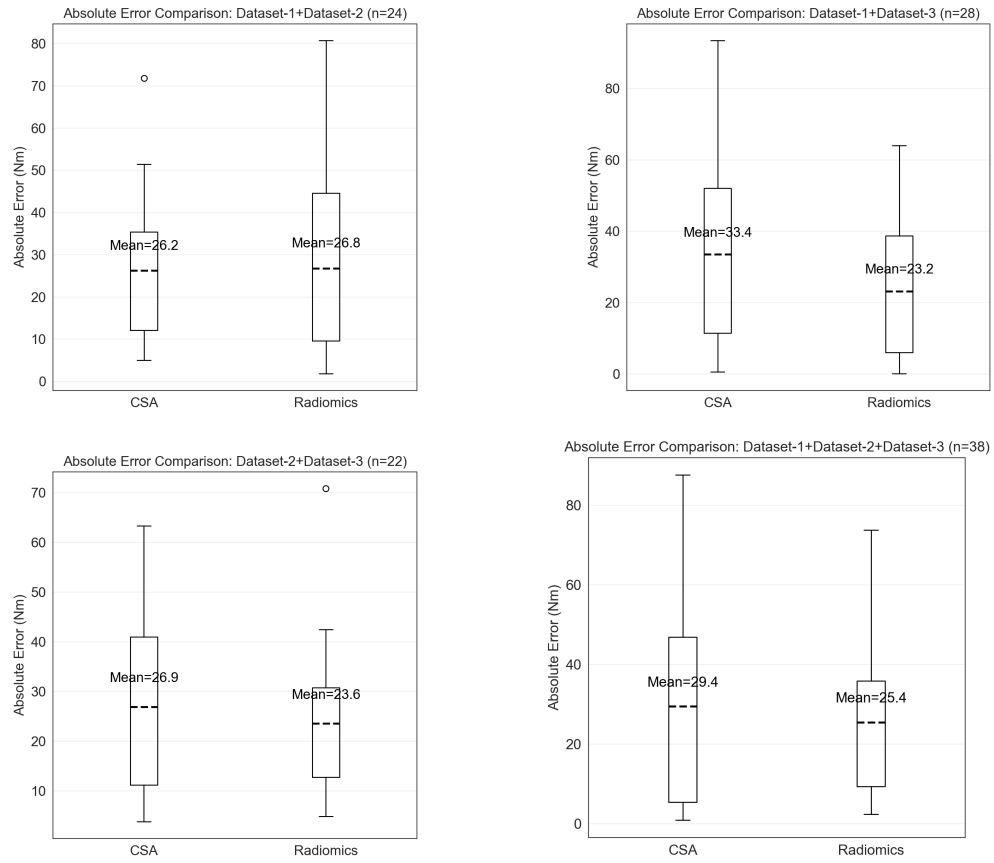


Fig. 9: Bland–Altman and absolute error plots comparing strength predictions from a CSA-only model (linear mixed model) and a radiomics-based model XGBoost across different dataset combinations using combined water and fat fraction radiomics features.

References

- [1] Cameron, D., Reiter, D.A., Adelnia, F., Ubaida-Mohien, C., Bergeron, C.M., Choi, S., Fishbein, K.W., Spencer, R.G., Ferrucci, L.: Age-related changes in human skeletal muscle microstructure and architecture assessed by diffusion-tensor magnetic resonance imaging and their association with muscle strength. *Aging Cell* **22**(7), 13851 (2023) <https://doi.org/10.1111/accel.13851> <https://onlinelibrary.wiley.com/doi/pdf/10.1111/accel.13851>. e13851 ACE-22-0089.R1
- [2] Mazzoli, V., Vainberg, Y., Hall, M., Barbieri, M., Asay, J., Muccini, J., Rosenberg, J., Kogan, F., Delp, S., Gold, G.: Improved strength prediction combining mri biomarkers of muscle quantity and quality. *NMR in Biomedicine* **38**(9), 70112 (2025) <https://doi.org/10.1002/nbm.70112> <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/nbm.70112>. e70112 NBM-25-0417
- [3] Fried, L.P., Tangen, C.M., Walston, J., Newman, A.B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W.J., Burke, G., McBurnie, M.A.: Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **56**(3), 146–156 (2001) <https://doi.org/10.1093/gerona/56.3.M146>
- [4] Cruz-Jentoft, A.J., Bahat, G., Bauer, J., Boirie, Y., Bruyère, O., Cederholm, T., Cooper, C., Landi, F., Rolland, Y., Sayer, A.A., Schneider, S.M., Sieber, C., Topinkova, E., Vandewoude, M., Visser, M., Zamboni, M.: Sarcopenia: Revised european consensus on definition and diagnosis. *Age and Ageing* **48**(1), 16–31 (2019) <https://doi.org/10.1093/ageing/afy169>
- [5] Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework (2019) [1907.10902](https://arxiv.org/abs/1907.10902) [cs.LG]