# SUPPLEMENTARY MATERIAL

Supplementary Table 1 – Hyperparameters defined using random grid search for elastic net (EN), gradient boosting machine (GBM), random forest (RF) and partial least square regression (PLS), in each cross-validation scenario for body condition score and milk traits.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Parameter | Samples-out random CV | | |  | Herd/Date-out random CV | | |
| BCS | BHB | k-CN |  | BCS | BHB | k-CN |
| EN | alpha | 0.089 | 0.134 | 0.148 |  | 0.105 | 0.144 | 0.221 |
| lambda | 0.141 | 0.220 | 0.180 |  | 0.263 | 0.309 | 0.317 |
| GBM | Ntree | 585 | 420 | 470 |  | 520 | 690 | 770 |
| learning rate | 0.228 | 0.428 | 0.139 |  | 0.118 | 0.108 | 0.187 |
| max tree depth | 27 | 29 | 22 |  | 17 | 26 | 20 |
| min samples | 130 | 76 | 111 |  | 116 | 131 | 116 |
| RF | Ntree | 889 | 773 | 625 |  | 1120 | 850 | 991 |
| mtries | 785 | 550 | 150 |  | 825 | 716 | 330 |
| PLS | NLV | 19 | 15 | 11 |  | 21 | 20 | 17 |

lambda – defines the amount of shrinkage; alpha - regularization parameter between ridge regression (alpha = 0) and lasso regression (alpha = 1); Ntree - the number of decision trees to be used in the model; learning rate - determines the contribution of each tree on the final model; max tree depth - specifies the maximum depth to which each tree will be built; min samples - specifies the minimum number of observations for a leaf in order to split; mtries - specify the number of predictors to randomly select at each level; NLV – number of latent variables.



**Supplementary Figure 1.** Heatmap of the Pearson correlations matrix for of Fourier-transform infrared (FTIR) milk spectra in Holstein-Friesian (468 cows). A total of 1,060 wavelengths of FTIR spectra covering a region from 5,011 to 925 cm−1 was used.