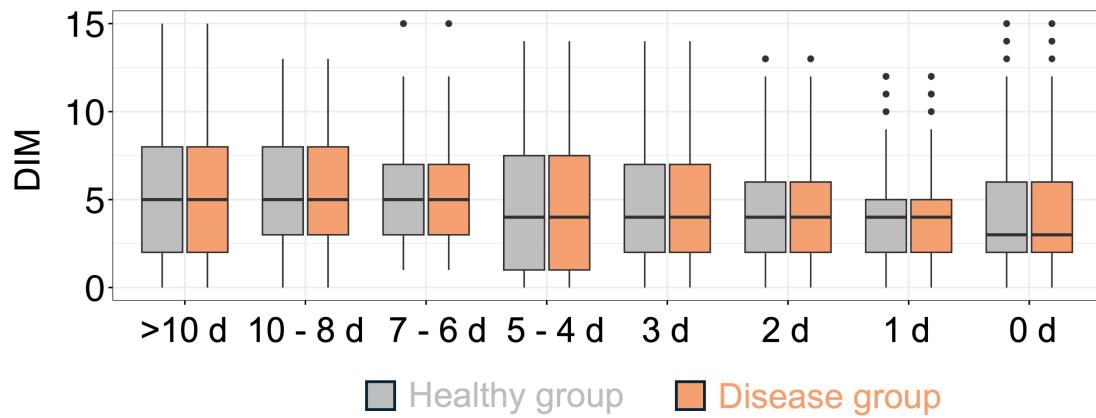
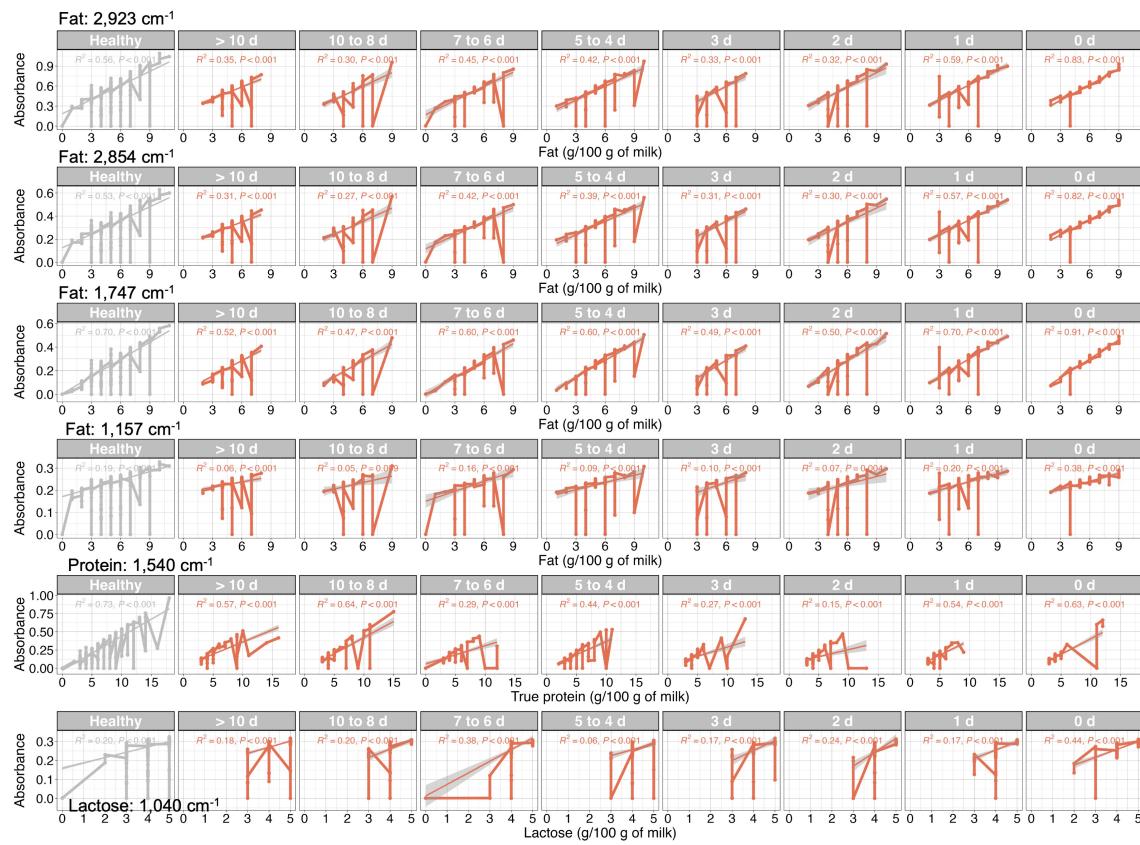


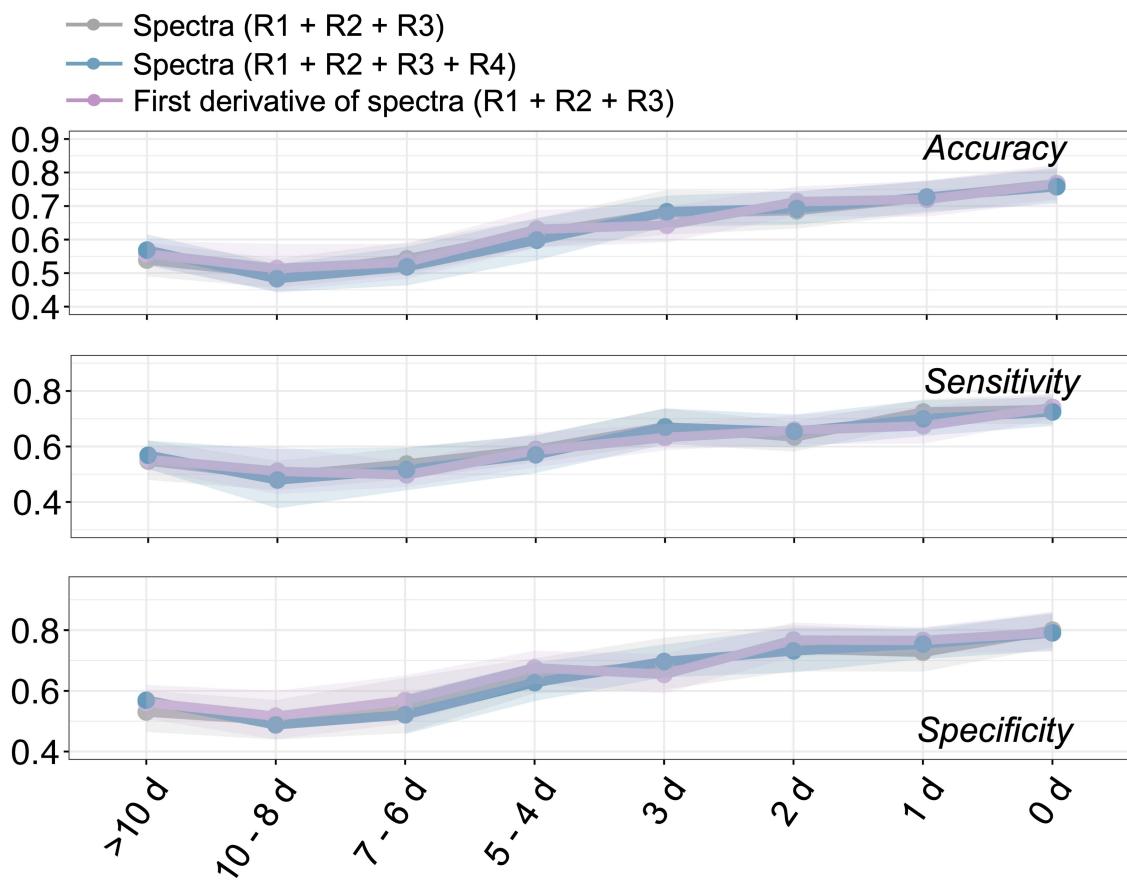
Supplemental Figure 1. The DIM distribution between healthy and diseases groups in each time period.



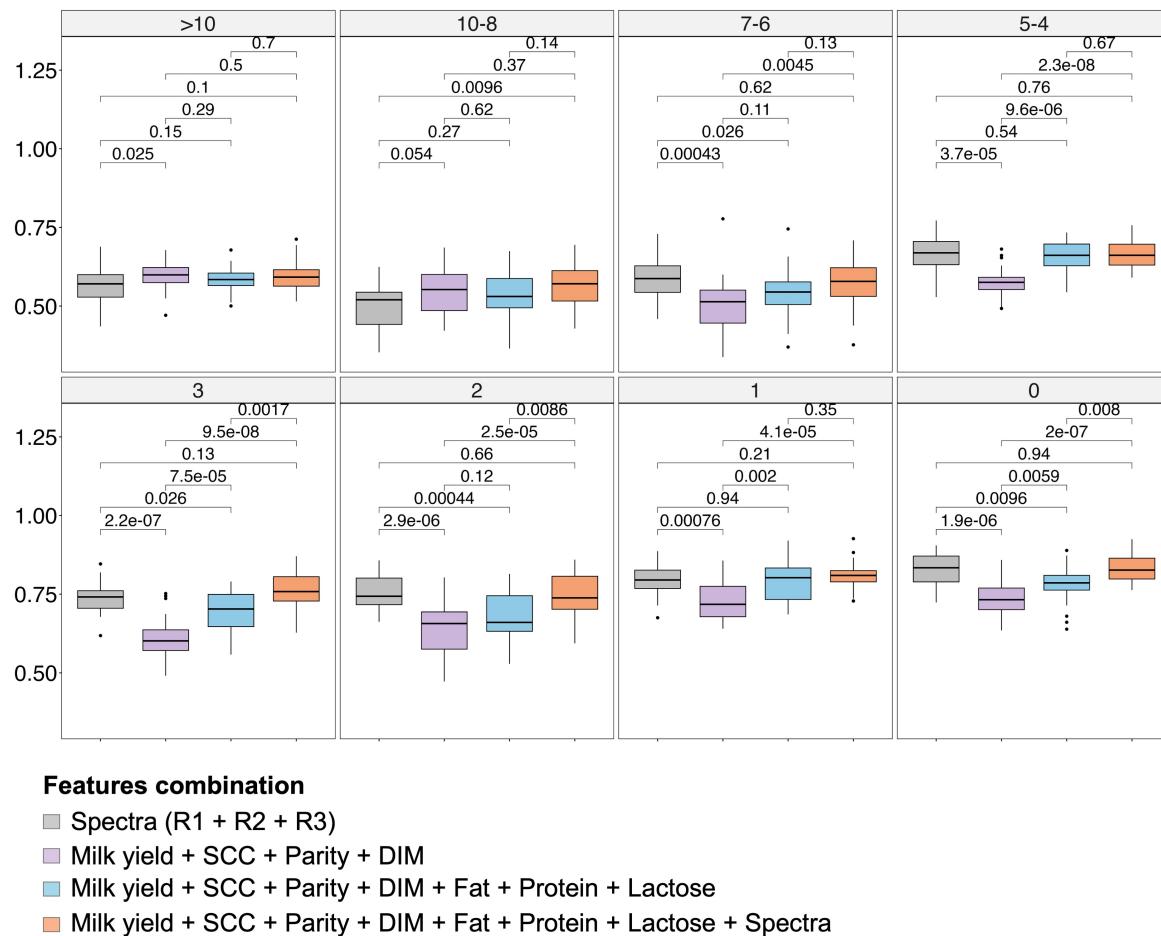
Supplemental Figure 2. Line plots with 95% confident interval showing the linear relationship between the predictive milk components and the absorption peaks of these components (fat, true protein and lactose) at eight separate time periods prior to disease diagnosis (>10 d, 10 – 8 d, 7 – 6 d, 5 – 4 d, 3 d, 2 d, 1 d, and 0 d). Coefficient of determination(R^2) which represents the percentage of the variance in the absorption peaks that the predictive milk components explain collectively, and P-value which indicates overall significance of the regression model were present.



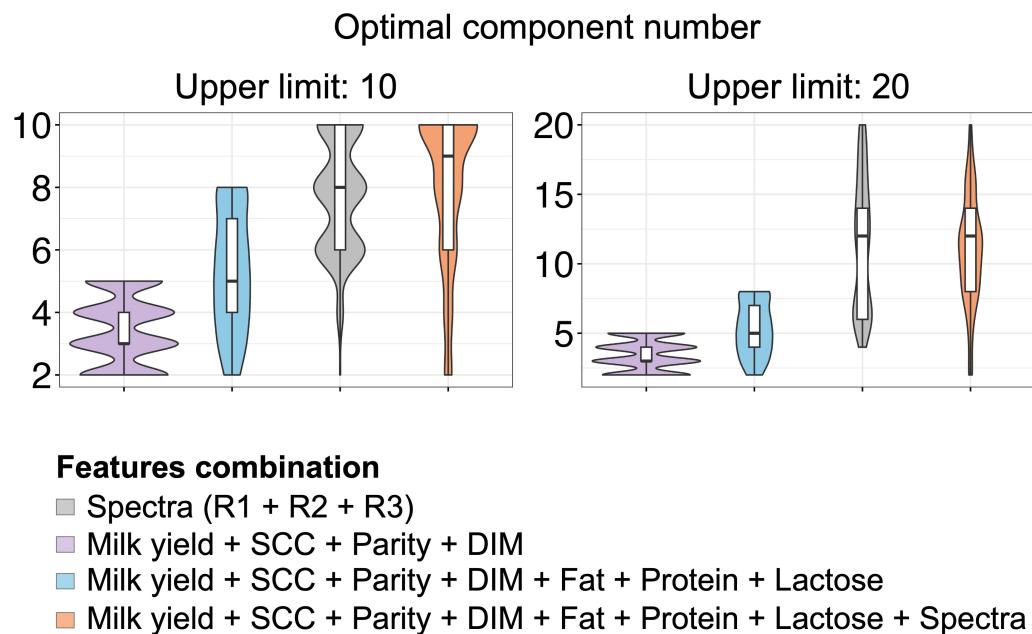
Supplemental Figure 3. Predictive performance (accuracy, sensitivity and specificity) of PLS-DA using milk FTIR spectra ($R_1+R_2+R_3$ and $R_1+R_2+R_3+R_4$) and the first derivative.



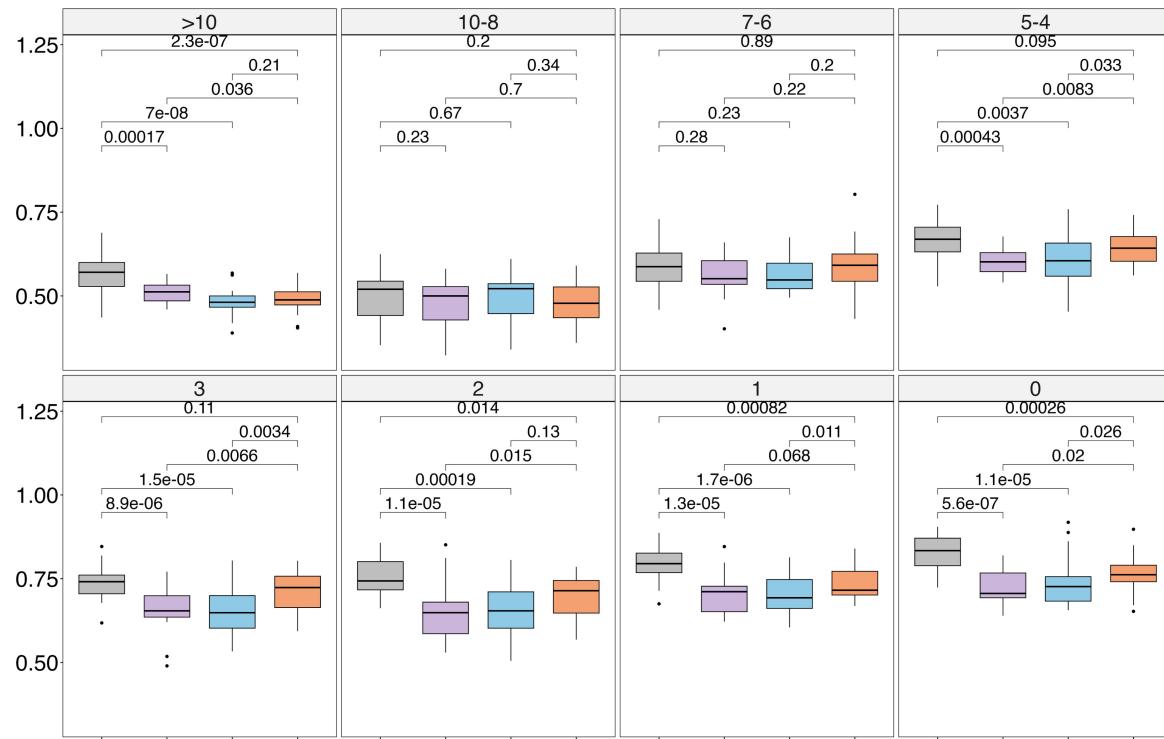
Supplemental Figure 4. Comparisons of predictive performance (AUROCs) of PLS-DA using milk FTIR spectra (R1+R2+R3) and various feature combinations (DIM, parity, milk yield, SCC, spectra-predicted fat%, protein% and lactose%). * adjusted $P < 0.05$. ** adjusted $P < 0.01$. *** adjusted $P < 0.001$.



Supplemental Figure 5. Optimal principal components number searched by double cross-validation framework of PLS-DA using milk FTIR spectra (R1+R2+R3) and various feature combinations (DIM, parity, milk yield, SCC, spectra-predicted fat%, protein% and lactose%).



Supplemental Figure 6. Comparisons of predictive performance (AUROCs) of PLS-DA, ridge regression, random forest and LSTM, using milk FTIR spectra (R1+R2+R3). * adjusted $P < 0.05$. ** adjusted $P < 0.01$. *** adjusted $P < 0.001$.



Models

- PLS-DA
- Ridge regression
- Random forest
- LSTM