**Development of feature extraction method for near infrared spectroscopy using stepwise bayesian linear regression**

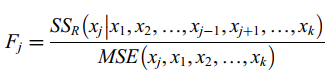
Paper này đề xuất một method gọi là stepwise bayesian linear regression (SBLR)

Cho NIR spectral data có N samples:



Sử dụng thuật toán bayesian linear regression (BLR) để xây dựng 1 data-driven model để tính toán F-statistic.

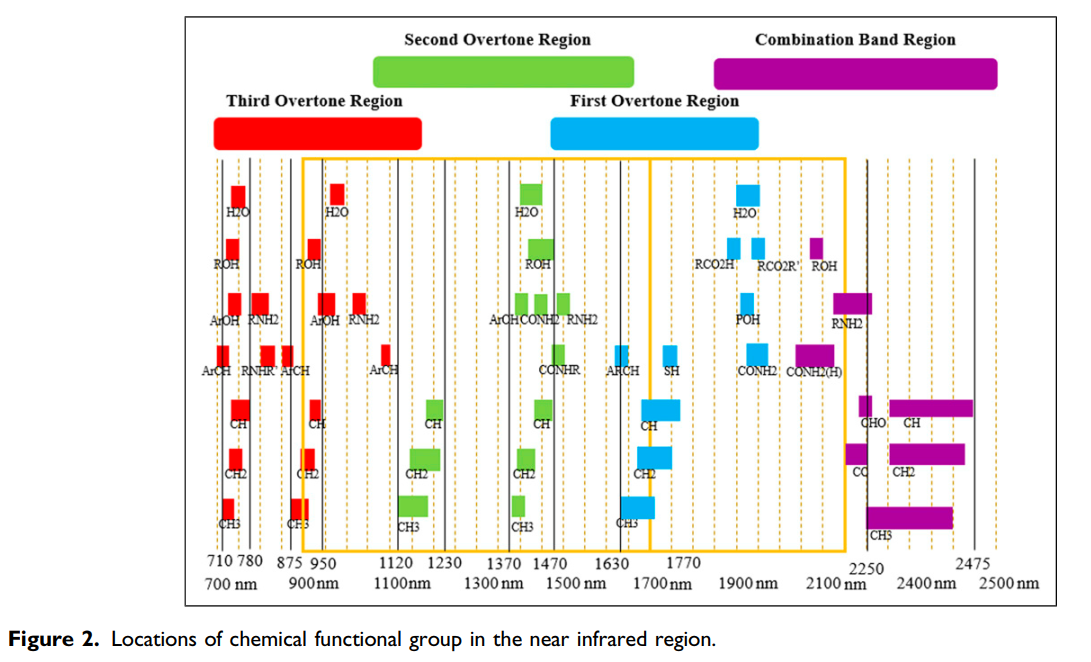
Giá trị **F-statistic** được tính toán bởi công thức:



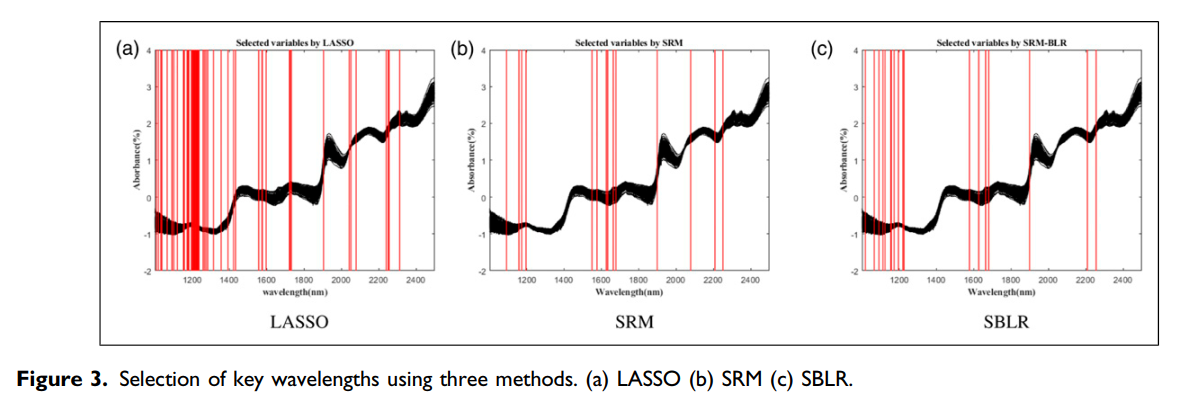
Chọn ngưỡng 

Nếu giá trị F > 0.95 thì sẽ được add vào optimal set (tập tối ưu) , ngược lại nếu bé hơn sẽ bị loại bỏ. -> đây là cách method lựa chọn đặc trưng (feature selection)

Thực nghiệm:



Các đường màu đen biểu thị khoảng bước sóng hấp thụ của các hợp chất C-H và O-H: 710–780, 850–950, 1120 - 1230, 1370-1470, 1630-1770, 2050-2100, 2250-2475.



So sánh thuật toán SBLR so với 2 phương pháp trước đây LASSO và SRM

LASSO + F-statistic

SRM + F-statistic

SBLR + F-statistic -> best performance