**1. Machine Learning and Deep Learning for predicting the chemical substance**

Hiện nay đã tồn tại nhiều kỹ thuật machine learning dự đoán thành phần hóa học của một mẫu dựa trên phổ near infrared spectroscopy (NIR), một số kỹ thuật machine learning phổ biến [ ] trong bài toán dự đoán thành phần dựa trên phổ NIR như là Partial Least Squares Regression, Principal Components Regression, Ridge Regression,... .

Các nghiên cứu gần đây cho thấy rằng các kỹ thuật Ensemble Learning [ ] [ ] kết hợp các mô hình machine learning giúp cải thiện performance của mô hình cho ra output vượt trội với yêu cầu về mặt dữ liệu và thời gian training thấp so với deep learning [ ]

Ở paper này, chúng tôi đề xuất một mô hình Ensemble Learning mới với performance tốt hơn so với các kĩ thuật machine learning trước đây.

**2. Giải thích sự liên quan giữa chất và bước sóng**

Sự liên quan giữa chất và bước sóng còn được gọi là wavelength selection. Các phương pháp feature selection chúng ta thường biết như Variance Thresholding,Correlation-based methods,Genetic algorithm,... đây là nhóm các phương pháp Feature Selection chỉ dựa trên dữ liệu input, một nghiên cứu gần đây của NIRPY Research [ ] đã sử dụng thuật toán Genetic algorithm cho bài toán wavelength selection.Tuy nhiên trong bài toán tìm kiếm sự biến đổi wavelengths phụ thuộc vào thành phần hóa học, một mẫu sẽ thường chứa nhiều thành phần hóa học nên chúng ta cần tìm kiếm mối liên hệ giữa input với mỗi output, vì vậy các thuật toán feature selection này sẽ không phù hợp. Nghiên cứu gần đây của Zhifeng Chen et al. [ ] dựa trên F-statistic tính toán mối liên hệ giữa input và output để từ đó loại bỏ đi những đặc trưng không liên quan. Ở paper này, chúng tôi đề xuất một kỹ thuật mới dựa trên nguyên tắc này gọi là Backward Elimination.

Machine Learning:

[] NIRPY Research

[] Near-infrared spectroscopy and machine learning-based technique to predict quality-related parameters in instant tea

Ensemble Learning:

[] A Novel Interpretable Ensemble Learning Method for NIR-Based Rapid Characterization of Petroleum Products

[] Using Near-Infrared Spectroscopy and Stacked Regression for the Simultaneous Determination of Fresh Cattle and Poultry Manure Chemical Properties

Feature Selection:

[] A Novel Interpretable Ensemble Learning Method for NIR-Based Rapid Characterization of Petroleum Products

[ ] A Guide to NIR: Understanding NIR spectra

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

[ ] Wavelength selection with a genetic algorithm - NIRPY Research