TOWARDS AI-DRIVEN DRUG QUALITY CONTROL: A SPECTRAL AUGMENTATION FRAMEWORK FOR FOURIER TRANSFORM INFRARED SPECTROSCOPY (FTIR) BASE CHEMOMETRIC QUANTIFICATION

B.A.M.C.L. Abeysinghe[1] & S. Bandaranayake[1]*

[1]University of Colombo, Sri Lanka

*Corresponding author: savinisb@chem.cmb.ac.lk

Abstract

Pharmaceutical quality control traditionally relies on analytical techniques that require extensive sample preparation, significant analysis time, and substantial chemical consumption. Fourier Transform Infrared (FTIR) spectroscopy combined with machine learning offers a rapid and sustainable alternative. However, the development of robust artificial neural network (ANN) models is often constrained by the limited availability of large calibration datasets. This study presents a spectral augmentation framework designed to generate realistic synthetic FTIR calibration data for the development of AI-driven pharmaceutical analytical methods. As a proof-of-concept, the framework was applied to the quantitative analysis of vitamins B1, B2, and B3 in locally manufactured vitamin B tablets. Although the commercial formulations contain multiple excipients, cellulose and starch were selected as representative excipients for method development. FTIR spectra were collected from 27 calibration samples containing active pharmaceutical ingredients (APIs) and excipients at concentrations ranging from 80–120% w/w of nominal formulation levels. A linear spectral mixing model was constructed using pure-component spectra and known formulation compositions to generate simulated FTIR spectra. The agreement between simulated and experimentally acquired spectra was evaluated using Pearson correlation coefficients, Hit Quality Index (HQI), Spectral Angle Mapper (SAM), and residual analysis. Correlation coefficients ranged from 0.9897 to 0.9979 (mean = 0.9954), HQI values from 0.9989 to 1.0000 (mean = 0.9998), demonstrating high spectral fidelity. Residual analysis was used to characterize experimental variability and incorporate realistic noise into the augmentation process. Using Latin Hypercube Sampling, 1800 synthetic spectra were generated to represent formulation variability and measurement uncertainty. The augmented dataset was then used to train ANN models for quantitative prediction of vitamin concentrations in multicomponent formulations. These results demonstrate the potential of extending this work to physics-informed spectral augmentation with machine learning to support sustainable, data-efficient pharmaceutical quality control.

Keywords: Sustainable, FTIR, ANN, Chemometrics, Pharmaceutical