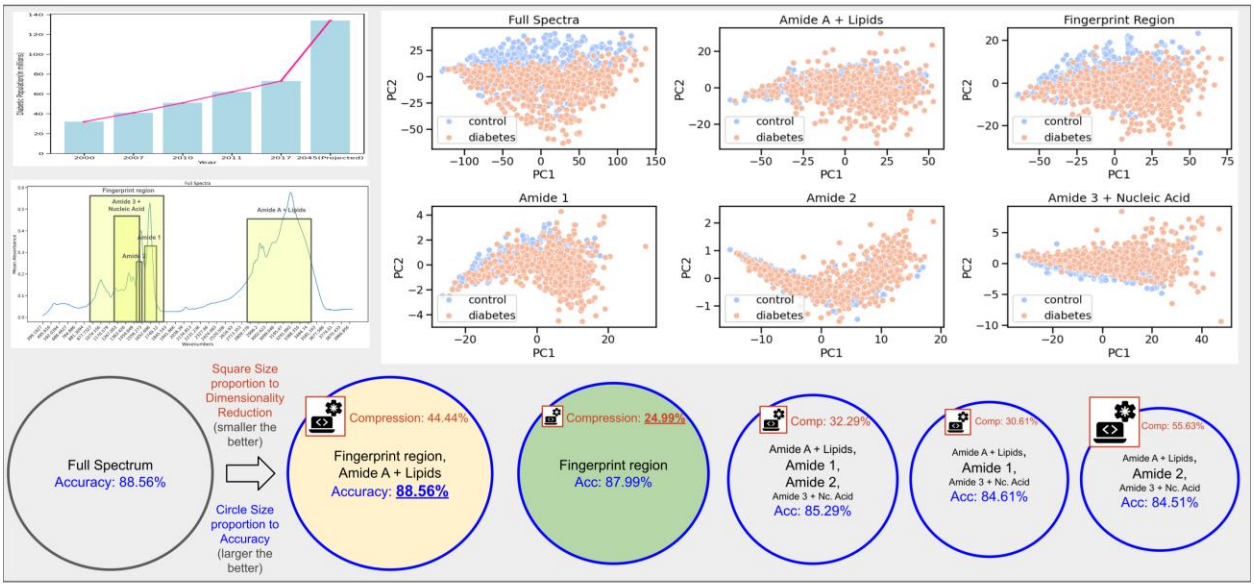
AI-Assisted Spectral Analysis for Diabetes Prediction

Veer Gajarlawar[1,3], Anjali Devi J S[2], Mahesh Mohan M R[3*]

[1]Department of Chemistry, IIT Kharagpur; [2]Department of Chemistry, Kannur University;
[3]Department of Artificial Intelligence, IIT Kharagpur
*Email:mahesh.mohan@cai.iitkgp.ac.in



Diabetes is nearing the proportions of a global epidemic [1]. Traditional diagnosis of diabetes follows an invasive approach, by pricking the body for blood samples. This approach causes great inconvenience, which results in improper monitoring and a large population with undiagnosed diabetes. This emphasizes a need for a non-invasive method. One promising solution in this regard is via analysis of saliva using Fourier transform infrared (FTIR) spectroscopy with machine learning (ML) [2]. However, this approach faces a limitation due to the very high dimension of the FTIR spectrum, viz., the curse of dimensionality that leads to higher computation and loss of interpretability [3]. One major challenge is that different regions of the FTIR have overlapping information that carry critical but complex information (as shown by principal components in Figure). We address this problem by identifying regions of FTIR that contribute to diabetes diagnosis, while eliminating unwanted regions by employing ML techniques: K-nearest neighbor and support vector machine. Our results suggest that the combination of fingerprint region (<1500 cm^{-1}) and Amide A + Lipid Region (2800-3500 cm^{-1}) are optimal for diabetes diagnosis, which leads to a 44.4% dimensionality reduction, while delivering the same performance as that of full FTIR spectrum.

References

1. Tabish S. A. Is Diabetes Becoming the Biggest Epidemic of the Twenty-first Century? *International Journal of Health Science*, **2007,** *1(2)*, V–VIII.
2. Sanchez-Brito, M.; Luna-Rosas, F. J.; Mendoza-Gonzalez, R.; Vazquez-Zapien, G. J.; Martinez-Romo, J. C.; Mata-Miranda, M. M. Type 2 Diabetes Diagnosis Assisted by Machine Learning Techniques through the Analysis of FTIR Spectra of Saliva. *Biomedical Signal Processing and Control* **2021**, *69*, 102855.
3. Altman, N.; Krzywinski, M. The Curse(s) of Dimensionality. *Nature Methods* **2018**, *15* (6), 399–400.