MULTIVARIATE ANALYSIS OF MOLECULAR SPECTROSCOPY DATA FOR COVID-19 DETECTION

QIZHONG LIANG, YA-CHU CHAN, JUTTA TOSCANO, *JILA and NIST, University of Colorado, Boulder, CO, USA*; KRISTEN K. BJORKMAN, LESLIE A. LEINWAND, ROY PARKER, *BioFrontiers Institute , University of Colorado Boulder, Boulder, CO, USA*; DAVID J. NESBITT, JUN YE, *JILA and NIST, University of Colorado, Boulder, CO, USA*.

In exhaled human breath, there exist hundreds of sparse molecular species and many contain rich information about various health conditions or diseases. When associated with a specific medical response, a co-variation in concentrations for multiple molecular species can occur, thereby facilitating diagnosis. A recent technological improvement to the cavity-enhanced frequency comb spectroscopy (CE-DFCS) has enabled broadband molecular spectra to be collected at the parts-per-trillion detection sensitivity [a], allowing unambiguous and objective detection of multiple molecular species in a simultaneous manner. Here, we show how the breath spectroscopy data collected by CE-DFCS can realize non-invasive medical diagnostics [b]. The key to such realization comes from the use of supervised machine learning to process the comb spectroscopy data in parallel with extreme-dimensional data channel inputs. Using a total of 170 individual breath samples, we report cross-validated results with excellent discrimination capability for COVID-19. At the same time, significant differences are identified for several other personal attributes, including smoking, abdominal pain, and biological sex difference. Our demonstrated approach can be extended immediately to investigate the diagnostic potential for a number of other disease states, including breast cancer, asthma, and intestinal problems. We discuss how further development in machine learning and frequency comb-based breath analysis can benefit significantly from enriching the absorption database to include more molecular species.

---

[a] Q. liang, et al., "Ultrasensitive multispecies spectroscopic breath analysis for real-time health monitoring and diagnostics," PNAS 118(40) (2021).

[b] Q. liang, et al., "Frequency comb and machine learning-based breath analysis for COVID-19 classification," arXiv:2202.02321 (2022).