Bio-intelligence for Lung Cancer Prediction

Zong-Lin Lin[1,3], Yu-Chiang Pan[1,2], Kuan-Ting Chiang[1,2], Guan-Hua Huang[1,3], Yaw-Kuen Li[1,2]

[1]*Center for Emergent Functional Matter of Science, National Chiao Tung University, Taiwan*
[2] *Department of Applied Chemistry, National Chiao Tung University, Taiwan*
[3] *Institute of Statistics, National Chiao Tung University, Taiwan*

Superstar147896325@yahoo.com.tw

A select ion flow tube mass spectrometry (SIFT-MS) has been employed for the quantitative analysis of human volatile organic compounds (VOCs). By integration of SIFT-MS analysis and classical liquid biopsy, we aim to identify a group of gaseous biomarkers of lung cancer and to establish a model for disease prediction. A databank containing more than 1500 VOCs data (including lung cancer and other diseases) has been constructed and analyzed. The analyses were performed based on breath samples from 148 patients with histologically confirmed lung cancers and 168 healthy volunteers. We used the XGBoost, a machine learning method, to build a prediction model that used VOC measurements to predict the disease state of lung cancer. The proposed prediction model achieved better performance than those from other previous approaches with accuracy, sensitivity, specificity, and AUC (the area under the ROC curve) being 0.89, 0.82, 0.94, and 0.95, respectively. When further considering the effect of environmental VOCs, our model can be improved to reach an outcome with 0.91 accuracy, 0.93 sensitivity, 0.88 specificity, and 0.98 AUC. This outcome is potentially useful for further biomedical applications.