Extraction and identification of information from Mass Spectra of the breath of patients infected with SARS-CoV-2

Abstract

After the pandemic of COVID-19, rapid COVID screening plays an essential role of control the spread. However, almost all the current screening methods are invasive and bring discomfort and potential harm to subjects, especially for vulnerable patients such as infants and the aged. Therefore, a non-invasive detection method based on volatile organic components (VOC) is researched. Using specific equipment proposed by NanoTech Analysis S.r.l. (NTA) and the collected volatile gas plastic bag from cooperated health care center, the Gas Chromatography Mass Spectrometry (GC-MS) of VOC are generated and further processed as fingerprint of each subject. Consequently, with 89 positive and 230 negative patients' data, several machine learning algorithms (random forest, support vector machine, gradient boosting machine, etc.) are combined to acquire accuracy ranging from 80% to 90%, and recall ranging from 70% to 85%. The developed technology provides a novel concept for non-invasive rapid test screening for COVID-19 in various scenarios, although more positive data is required to tune the model.

Introduction

Problem description

The COVID-19 is an noval coronavirus appears firstly in 2019, causing 6.5 million death and even more people live with sequela of different level. There are 2 proven effective way to reduce the damage that the COVID-19 cause, the widely inoculated vaccine and eliminated social contact. However, none of these 2 ways can be efficient without the accurate and fast covid test.

There are several COVID-19 test available in the market, in which most of them requires qualified skillers to ensure the test accuracy and all of them are invasive, which means the tests shall bring discomfort or even potential damage to subjects' health. This situation is more severe to those who are in volunerable status, for example the infants and the aged. Consequently, some people may be COVID-19 positive but are spared to be tested, which will leads to more infections. Thus, a non-invasive and easy-use COVID test technic is required to fill the last gap. Therefore, NanoTech Analysis S.r.l. (NTA), an Italian innovative company operating with MEMS and NEMS technologies, proposed a possible technology that can fullfill the requirement. With a portable instrument and some volatile gas from the subjects, a Gas Chromatography Mass Spectrometry (GC-MS) of VOC can be generated and the further data analysis can be implemented. After the data cleaning and feature selection, the data can be feed to the machine learning model and finally we can have a classification result of subject.