

Fast and noninvasive electronic nose for sniffing out COVID-19 based on exhaled breath-print recognition

Dian Kesumapramudya Nurputra (✉ dian.k.nurputra@ugm.ac.id)

Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281

Ahmad Kusumaatmadja

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara PO Box BLS 21, Yogyakarta 55281

Mohamad Saifudin Hakim

Department of Microbiology, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281

Shidiq Nur Hidayat

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara PO Box BLS 21, Yogyakarta 55281

Trisna Julian

PT Nanosense Instrument Indonesia, Umbulharjo, Yogyakarta 55167

Budi Sumanto

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara PO Box BLS 21, Yogyakarta 55281

Yodi Mahendradhata

Department of Health Policy and Management, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta

Antonia Morita Iswari Saktiawati

Department of Internal Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281

Hutomo Suryo Wasisto

PT Nanosense Instrument Indonesia, Umbulharjo, Yogyakarta 55167

Kuwat Triyana

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara PO Box BLS 21, Yogyakarta 55281

Keywords: electronic nose, breathalyzer, SARS-CoV-2, COVID-19, GeNose, profiling test, pattern recognition, gas sensor, machine learning

Posted Date: July 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-750988/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Despite its high accuracy to detect the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the reverse transcription-quantitative polymerase chain reaction (RT-qPCR) approach possesses several limitations (e.g., the lengthy invasive procedure, the reagent availability, and the requirement of specialized laboratory, equipment, and trained staffs). We developed and employed a low-cost, noninvasive method to rapidly sniff out the coronavirus disease 2019 (COVID-19) based on a portable electronic nose (GeNose C19) integrating metal oxide semiconductor gas sensor array, optimized feature extraction, and machine learning models. This approach was evaluated in profiling tests involving a total number of 615 breath samples (i.e., 333 positive and 282 negative COVID-19 confirmed by RT-qPCR) obtained from 83 patients in two hospitals located in the Special Region of Yogyakarta, Indonesia. Four different machine learning algorithms (i.e., linear discriminant analysis (LDA), support vector machine (SVM), stacked multilayer perceptron (MLP), and deep neural network (DNN)) were utilized to identify the top-performing pattern recognition methods and to obtain high system detection accuracy (88–95%), sensitivity (86–94%), specificity (88–95%) levels from the testing datasets. Our results suggest that GeNose C19 can be considered a highly potential breathalyzer for fast COVID-19 screening.

1. Introduction

Contagious coronaviruses can cause intestinal and respiratory infections in both humans and animals.^{1,2} The emergence of a novel coronavirus, officially termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed serious challenges to global health. SARS-CoV-2 infection, causing coronavirus disease 2019 (COVID-19), was found in late 2019 in Wuhan, Hubei province, China, and subsequently spreading to be a causative agent for an ongoing and escalating pandemic in more than 200 countries and territories across the world.^{3–5} While the other previously found human coronaviruses (e.g., HCoV-OC43, HCoV-NL63, HCoV-229E, and HKU1) only caused mild upper respiratory diseases in immunocompetent patients, SARS-CoV-2 has been considered to be the third deadly pathogenic coronavirus over the past two decades after the appearances of SARS-CoV (2002 – 2003) in Guangdong province, China and the Middle East respiratory syndrome coronavirus (MERS-CoV, 2012) in Middle Eastern countries.^{1,6} COVID-19 pandemic is associated with significant fatalities, especially in the elderly and immunocompromised populations.

The reverse transcription-quantitative polymerase chain reaction (RT-qPCR) method has been routinely utilized to confirm the diagnosis of COVID-19 since the beginning of the pandemic. So far, this diagnostic technique detecting ribonucleic acid (RNA) of SARS-CoV-2 has become the most widely accepted test for SARS-CoV-2 detection.⁷ Various clinical samples (i.e., nasal and pharyngeal swabs, sputum, faeces, blood, bronchoalveolar lavage fluid, and urine) can be employed.⁸ Several companies and laboratories have developed PCR-based detection kits targeting at least two regions (genes) of the SARS-CoV-2 genome. Despite its high specificity compared to other diagnostic methods (e.g., antibody and nucleocapsid protein antigen detection assays), RT-qPCR has several drawbacks from the clinical

perspective (i.e., the need to be carried out by professionally trained healthcare technicians, the requirement to be performed in a specialized laboratory, invasive sampling procedure, high cost, lengthy processing time, and complicated RNA extraction)⁹. All those limitations potentially lead to delay in establishing the diagnosis of SARS-CoV-2 infection.

Another critical issue is that some commercially available PCR kits do not publicly disclose their primer sequence sets, leading to difficulties in validating each primer set's sensitivity and specificity. Thus, in most cases, the detection protocol and primer sets have to be first optimized because the unoptimized primer sets could potentially yield false-positive results.¹⁰ All those limitations do not only potentially lead to delay in establishing the diagnosis, but also to arise a question on whether the RT-qPCR should still be an operational gold standard in the diagnosis of COVID-19.¹¹ To overcome these issues, several clinicians and researchers have attempted to add another diagnostic component of chest computed tomography (CT) alongside the RT-qPCR for validating the diagnosis results.^{12,13} However, it should be noted that CT alone without RT-qPCR cannot be used as a definitive diagnostic tool because some patients may have normal radiological conditions at the early stage of COVID-19, which can then affect the negative predictive values. In addition to RT-qPCR, more than 100 kits with different working principles (e.g., isothermal amplification and lateral flow-based detection for nucleic acid targets, chest CT imaging, and immunoassays) for COVID-19 diagnosis produced by several companies have been available on the market.¹⁴

Since the COVID-19 primarily involves the respiratory tract, breath analysis offers an alternative as a fast and non-invasive detection approach for sniffing out this disease. It is also more comfortable to perform and more convenient to the patients, including pediatric and elderly populations.¹⁵ Exhaled human breath contains hundreds of volatile organic compounds (VOCs), which result from various metabolic pathways. They could be detected by gas chromatography-mass spectroscopy (GC-MS) and proton transfer reaction-mass spectrometry (PTR-MS).^{16,17} The most abundant VOCs from exhaled human breath are acetone, methanol, ethanol, propanol, and isoprene. Additionally, a few other compounds are also detected in exhaled breath (e.g., benzene, acetonitrile, diallyl sulfide, allyl methyl sulfide, and diallyl disulfide).^{18,19} It is well known that several VOCs with their specific compositions can be utilized as noninvasive biomarkers for various respiratory diseases, including oesophageal-gastric cancer¹⁶, lung cancer²⁰, asthma²¹, rhinovirus-induced wheeze²², influenza infection in swine²³, and tuberculosis²⁴. Recent evidence indicated that COVID-19 could also be diagnosed using a VOC-based breath analysis approach by means of near-patient gas chromatography-ion mobility spectrometry (GC-IMS).²⁵ Regardless of its high accuracy, ability to precisely identify the VOC types, and usefulness in pathophysiological research, the mass spectrometry analysis is still time-consuming, expensive, and not user friendly for bedside clinical practice. Moreover, for rapid COVID-19 testing of a large-scale human population, this technique alone is inappropriate, especially if the target detection time should be only in a few minutes.

Therefore, an electronic nose, an artificial olfactory system consisting of an integrated gas sensor array with different active layers and artificial intelligence (AI) to discriminate complex odors, can be opted as an alternative path to assess the VOC mixtures in the exhaled breath. During its exposure to the breath, the sensors respond in a specific way to the various fractions of VOCs.^{26,27} Each odor, representing a unique VOC mixture, can then result in a sensor signal pattern that is distinctive to that odor. For the exhaled air, it is called breath-print. Complex VOC mixtures using pattern recognition algorithms like machine learning, can therefore be discriminated and classified at high throughput without identifying the individual molecular components. An electronic nose is relatively inexpensive, mostly portable, fast (creating results within a few minutes), and easy to use. Considering patient perspective in the case of COVID-19, exhaled breath analysis using an electronic nose is attractive because it is noninvasive, rapid, safe, and simple to operate. Previously, the electronic nose has been implemented in clinical settings to diagnose and monitor respiratory and urologic diseases (e.g., ventilator-associated pneumonia, tuberculosis, kidney failure), and recently was used as a pre-operative screening for COVID-19.²⁸⁻³¹ VOC profiles are expected to change following the continuous progression of COVID-19 within the patient respiratory system. Their alteration is complex, which is associated with the SARS-CoV-2 and its unique interaction with the host.

Depending on their working principles and materials, several thin-film and micro-/nanoscale gas sensors can be employed as main detecting components in the electronic nose, which include gravimetric sensors (e.g., surface acoustic wave resonators, resonant micro-/nanocantilevers, and quartz crystal microbalances)³²⁻³⁹, chemoresistive sensors (e.g., metal oxide semiconductor, polymer, two-dimensional material, and carbon-based sensors)⁴⁰⁻⁴³, colorimetric sensors (e.g., printable pastes and optical dye-functionalized washable threads)^{44,45}, and optical sensors (e.g., visible and infrared micro-/nanoLEDs)⁴⁶⁻⁴⁹. Among them, chemoresistive metal oxide semiconductor gas sensors have been favorable, especially for VOC detection, due to their excellent characteristics (i.e., low cost, short response time, simple measurement setup, high durability, and long lifetime).⁵⁰ The drawback of such sensors concerning the moderate selectivity can be overcome when different types of active materials are used in a simultaneous fashion. Their unique signals can then be identified and combined in AI-based data post-processing. Hence, electronic nose comprising chemoresistive metal oxide semiconductor sensors has been used in many environmental monitoring fields.⁴⁰

Nonetheless, to the best of our knowledge, there is no study so far that investigated the potency of electronic noses in detecting COVID-19 patients at bedside clinical setting. The existing study of electronic nose application (Aeonose) on SARS-CoV2 detection was on pre-operative screening.³¹ Therefore, in this work, we developed a portable breathalyzer, so-called GeNose C19, by integrating metal oxide semiconductor gas sensor array, machine learning analysis, and breath sampling setup (see **Figure 1**). The custom-built system was employed in profiling clinical tests involving two different subject groups (i.e., RT-qPCR-confirmed positive and negative COVID-19 patients) in two hospitals located in Indonesia to investigate its potency for differentiating the exhaled breath patterns of both groups. Four

different machine learning algorithms were studied to find the highest possible accuracy of the developed device.

2. Results And Discussion

2.1 Integrated electronic nose for COVID-19 detection (GeNose C19)

The electronic and mechanical components integrated into an electronic nose for sniffing out COVID-19 (GeNose C19) were grouped into two main parts (i.e., sensing and breath sampling units), as depicted in Figs. 2a **and b**. The former consists of a chemoresistive sensor array sealed in a miniaturized chamber, a micropump, a three-way solenoid valve, a power supply, and a data acquisition system. All of them are placed in a three dimensional (3D)-printed housing. Meanwhile, the latter comprises a high-efficiency particulate air (HEPA) filter and a disposable air sampling bag made of medical-grade polyvinyl chloride (PVC). Both units were designed separately, which were then connected utilizing a flexible medical-grade polytetrafluoroethylene (PTFE) tube with an outer diameter of 4 mm to enable airflow during gas sensing (see Fig. 2c). Here, the HEPA filter was attached between the reservoir bag and the GeNose C19 inlet to filter out the virus-containing droplets. Moreover, it possesses a water absorber element to eliminate the water molecules known as one of the most interfering factors for gas sensors (Fig. 2d). Hence, only the target VOCs will enter the sensing chamber as the viruses are expected to be trapped by the filter containing fibrous mats (see Figs. 2e **and f**) and are not allowed to contaminate the whole air trajectory in the GeNose C19 machine.

To ensure human safety during the breath exposure test and obtain reliable results, the breath sampling procedure was carefully set and controlled (Fig. 2a). First, the patients were asked to store their end-tidal breath during the third exhalation into 1 L sampling bags (see Fig. 2a**(1)**). Here, the first two breaths were not taken to minimize contamination from the dead space air (i.e., ventilated air that does not participate in gas exchange) and mouth-released odor, which are commonly found in the mixed expiratory breath.⁵¹ From another breath analysis report, different sampling procedures (i.e., mixed expiratory and end-tidal breath sampling methods) were discovered to result in varied VOC ratios.⁵² In case the blood-borne volatile substances will be assigned as biomarkers of disease, like in COVID-19 detection, the end-tidal breath sampling technique is, however, more appropriate to be used.⁵¹ The ratio values between the difference of expiratory and inspiratory concentrations and the alveolar concentration were normally calculated to estimate content alteration of both inhaled and exhaled substances, where the blood-borne volatile substances with clearly endogenous (e.g., CO₂, isoprene, and acetone) and exogenous origins (e.g., 2-propanol and 2-butanone) obtained high (> 1.5) and low (< 1) ratios, respectively.^{52,53} After the thirdly exhaled breath had been inserted into the sampling bag, its valve had to be closed quickly to avoid any air leakages (see Fig. 2a**(2)**). At the last step, the upper end of the reservoir tube was connected to the HEPA filter. Hence the valve could subsequently be reopened, allowing the airflow to occur from the bag to the GeNose C19 machine (Fig. 2a**(3)**). Such a similar breath sampling concept has already been employed in various off-line breathomics pipelines.⁵¹

Depending on the used filter, having multiple stacks of combined micro-/nanofibers with diameters in the range of $< 1-50 \mu\text{m}$ (see Figs. 2d-f) can ensure the successful filtration of bioaerosols carrying RNA copies of SARS-CoV-2. Different mechanisms are responsible for the final sizes of the emitted virus-containing particles, which were found to be ranging from < 1 to $> 100 \mu\text{m}$ when humans performed different actions (i.e., breathing, talking, sneezing, and coughing).⁵⁴ For normal breathing, particles with diameters of $< 0.8-2.0 \mu\text{m}$ were mostly observed.^{54,55} From the other clinical investigation, a similar HEPA filter with high grade has been placed between the bag valve and the patient mask during preoxygenation in intubation procedure for patients with COVID-19.⁵⁶ This filtering technique was also used to safely deliver aerosolized medications to the COVID-19 patients avoiding the aggravation of the novel coronavirus spreading.⁵⁷ Despite the adequate proofs of the HEPA filter effectiveness in trapping the aerosolized virus-containing particles from the previous reports^{56,57}, we had carried out additional investigation for the air trajectory part (pneumatic PTFE tube) after the GeNose C19 was used to measure the exhaled breath of a confirmed positive COVID-19 patient. As a result, the taken sample of the inner tube part was confirmed negative in the RT-qPCR test, which indicates the successful removal of the coronaviruses from the airways in the fiber filter. Thus, they did not contaminate the other components inside the sensing unit. In other words, the exhaust air produced by the GeNose C19 can be safely released to the ambient environment and will not further transmit airborne viruses.

In the sensing unit, ten different gas sensors based on metal oxide semiconductors (i.e., S1 – S10) are integrated with internal heaters as a detection module, where each of them has unique cross-sensitivity characteristics to several VOCs resulting in a specific signal pattern that represents the exhaled breath from either RT-qPCR-confirmed positive or negative COVID-19 patients. Since each sensor requires varied power consumptions between 280 and 835 mW, a total power of $\sim 6 \text{ W}$ was supplied to operate the whole sensor module. Besides, the GeNose C19 is also equipped with an environmental sensor to monitor the possible interferences from water molecules (humidity) and temperature change inside the sensing chamber. For enabling and controlling the alternating flows of reference (ambient) air and exhaled breath to the chamber, a micropump and a three-way solenoid valve are integrated into the GeNose C19 machine, respectively. While the micropump could create an air flowrate of $(1 \pm 0.2) \text{ L/min}$, the solenoid valve could be configured into three modes (i.e., delay, sampling, and purging phases). First, during the delay phase of 10 s, the ambient air was drawn by the micropump into the test chamber and subsequently characterized by the sensor array as the baseline voltages. Second, a VOC-containing breath sample was flown from the reservoir bag into the test chamber in the sampling phase. It will then react with the sensors within 40 s, yielding various response signals. Afterwards, at the last step (purging phase), the ambient air was again re-inserted into the chamber for 120 s to remove the remaining VOCs. The produced exhaust air was flown out to the ambient via the outlet located at the rear of the GeNose C19 machine. Both microcontroller and analog to digital converter (ADC, 16 bit external) were employed in the data acquisition system module to control the data exchange within electronics and convert the analog sensor responses (in voltage) to digital values, respectively. Hence, the digital data could finally be transferred to a homebuilt data logging software in the personal computer through a bluetooth or universal serial bus (USB) connection for further analysis. As the sensors are sensitive to variations in

atmospheric conditions, the initial placement of the GeNose C19 system can vary the output sensing results. In this case, indoor environments with good-ventilated indoor spaces or outdoor areas poor air circulation (i.e., rooms with strong odor backgrounds like fragrance or alcohol-based hand sanitizer) will result in bad or poor signals. Thus, the machine has to be carefully pre-conditioned before its use.

2.2 Sensor characteristics to exhaled breaths

As mentioned above, the sensor baseline values were first measured where no breath was flown into the test chamber (i.e., the sensor signals corresponding to ambient condition). These baseline values would then be used as subtrahends for the output response signals to the VOC-containing breath. After the gas molecules had interacted with the active layer surfaces of the chemoresistive sensors, their output voltages increased due to the reduced electrical resistances. The sensing measurements were conducted simultaneously for all the sensors within 40 s to reach their saturation stages. It is well known that for metal oxide semiconductor gas sensors, their conductivity will change in the presence of target gases due to a redox reaction between the active material (e.g., *n*-type tin oxide (SnO₂)) and the gas molecules. The detailed sensing mechanism towards VOCs based on an equilibrium shift of the surface chemisorbed oxygen reaction has been described in other studies.^{40, 58–60} Here, the depletion regions at material surfaces are controlled by target VOCs, leading to a change in the movement of free charge carriers from the metal oxide semiconductor to the oxygens or vice versa (i.e., free electrons in case of the *n*-type SnO₂ active layers).

Initially, when the activated sensor is exposed to ambient air, the oxygen molecules are adsorbed on the metal oxide semiconductor surfaces. The free electrons inside the sensing material are then attracted and attached to those oxygens. As a result, the measured electrical resistance increases, and the sensor output voltage drops into a lower value. This condition can be seen when the GeNose C19 sensors were pre-heated for at least 20–50 min in ambient air (see **Figure S1 in the Supplementary Information**). At this elevated temperature, sensor output voltage reductions ranging from 0.24 to 2.74 V were yielded depending on the initial measured baseline values and type of the functional materials. Afterwards, in the presence of reducing gases or VOCs (e.g., propane and formaldehyde (HCHO)), the VOC molecules will react with the adsorbed oxygen ions resulting in a lower number of oxygen on the semiconductor surfaces. Consequently, the previously trapped electrons are released back to the metal oxide semiconductor, leading to higher free electron concentration and lower sensor electrical resistance. As for the sensor output voltages, arise in their values was also observed while the sensors were interacting with the VOCs contained in the exhaled breaths of RT-qPCR-confirmed negative and positive COVID-19 patients (see Figs. 3a **and b**, respectively). After ~ 10–40 s sensing, the sensor signals had become stable, indicating the material surface saturation. A higher VOC concentration injected into the chamber leads to a lower concentration of the adsorbed oxygen on the semiconductor surface, which then further reduces the sensor resistance. Such a phenomenon was employed to realize real-time monitoring of the VOCs in the air.⁶¹ Once exposure of the sensor to the exhaled breath had ended (i.e., the sensing chamber inside the GeNose C19 machine was flushed with ambient air), the free electrons inside the metal oxide semiconductor returned, attaching to the adsorbed oxygens on surfaces. This was followed by the

recovery of the response signals to their initial values (baselines). Here, it could be noticed that the recovery duration is longer than the response time, which can be attributed to the required time to break the existing bonding between the remaining VOCs and oxygen molecules on the sensing surfaces.

Figures 3a **and b** show the typical responses of the gas sensor array in the GeNose C19 to the breaths of negative and positive COVID-19 patients, respectively. From these typical sensing signals, it is not easy to distinguish them directly and visually. Therefore, we extracted the features of each sensor based on the combination of maximum, median, standard deviation, and variance values, which resulted in 40 features for each data of a patient. Figures 3c - f depict boxplots of the distribution for the positive and negative label classes. A non-parametric statistical test (i.e., the Kruskal-Wallis test) was used to determine the significance of the difference between the negative and positive patient group data. This test is commonly used as an alternative to the one-way ANOVA test, where the assumption of normality of the data is not met. Sensors showing significant group population differences are indicated by an asterisk (*) symbol in the boxplot label. More detailed feature values are listed in **Table S1 in the Supplementary Information**. The test results demonstrated that the positive and negative groups could be significantly distinguished (p -value < 0.05) by the most features of the maximum, median, standard deviation, and variance of S1, S5, S6, S9, and S10.

The sensitivity of each sensor during exposure to VOC concentration depends on the used active material. Moreover, it should be noted that the sensor behaviors might be slightly altered when they were tested to the breath samples from different patients, although they were from the same group (either positive or negative COVID-19). This occurrence could be understood because the content and complexity of the exhaled VOCs are diverse, as discovered in another breath analysis study using GC-IMS.²⁵ Several VOC biomarkers could be identified as the best discriminants for distinguishing between positive and negative COVID-19 patients (e.g., ethanal, acetone, acetone/2-butanone cluster, 2-butanone, methanol monomer and dimer, octanal, feature 144, isoprene, heptanal, propanol, and propanal).²⁵ Nonetheless, the compounds observed from two different hospitals (i.e., Edinburgh, the United Kingdom (UK), and Dortmund, Germany) in their study were dissimilar for the same case of COVID-19 patients, which then added more complexity in analyzing the obtained breath data. These limitations were due to uncertainties in instrument setup, operating conditions, and background contamination levels.

A more detailed study in those matters has not been performed yet so far. Meanwhile, the other clinical GC-IMS study conducted by researchers in Beijing, China, suggested several other potential breath-borne VOC biomarkers for COVID-19 (i.e., acetone (C₃H₆O), ethyl butanoate, butyraldehyde, and isopropanol).⁶² They found that the decrease and rise of acetone (C₃H₆O) and ethyl butanoate levels, respectively, due to the changes of metabolites resulting from SARS-CoV-2 infections, were distinctive for COVID-19 patients.^{62,63} Moreover, the average measured isopropanol and butyraldehyde for the COVID-19 patients were lower than those for the healthy control, lung cancer, and non-COVID-19 respiratory infection patients. Metabolomics of exhaled breath in critically ill COVID-19 patients were also investigated from a research consortium in France using a proton-transfer-reaction quadrupole time-of-flight mass

spectrometer, where they observed four prominent VOCs (i.e., methylpent-2-enal, 2,4-octadiene, 1-chloroheptane, and nonanal) that could discriminate between COVID-19 and non-COVID-19 acute respiratory distress syndrome patients.⁶⁴ Overall, those previously reported mass spectrometry studies in several different countries (i.e., UK, Germany, France, and China) indicate that the distinctive VOC biomarkers for COVID-19 may vary across the world and should be further investigated based on community, races, and case-by-case with larger cohorts.⁶⁵

In contrast to a mass spectrometry method that attempts to find and identify the exact VOC biomarkers from the exhaled breaths quantitatively, our technique used in the GeNose C19 has focused more on the AI-based pattern analysis of the integrated sensor responses to those complex VOCs qualitatively resulting from the combined extra-pulmonary metabolic and gastrointestinal manifestations of COVID-19.⁶⁶ Thus, the breath data analysis and decision making procedure can be carried out in a simpler way and a shorter time, respectively, with high detection accuracy. To enable this, besides owning high sensitivity, the chemoresistive sensors should ideally be designed to possess high selectivity to a specific analyte in a gas mixture and zero cross-sensitivity to other compounds in the operating background. Such sensors were normally constructed in hybrid organic/inorganic structures with three-dimensional nano-architectures (e.g., nanofibers, nanowires, and nanofins), enhancing the active surface-area-to-volume ratios.^{67,68} There, the surfaces of semiconductor nanostructures were often functionalized with certain self-assembled monolayers (SAMs) or polymers to detect the target gas molecules specifically.^{33,35,69} Nevertheless, those organic materials suffer from low robustness. They are all well-known to degrade within a short duration of use (i.e., their chemical compositions will alter downgrading the sensor performance). As a result, pure inorganic materials (metal oxide semiconductors) are still preferably manufactured by sensor companies and widely used in gas sensing applications, including in the GeNose 19 system. Here, a single sensor alone was not sufficient for performing specific breath pattern recognition because the exhaled VOCs might have similar characteristics from one to another. This selectivity drawback could be alleviated by employing an array of ten sensors with different sensitivities and integrating the machine learning-based breath pattern recognition algorithms.

Furthermore, to demonstrate a proof-of-concept ability of the GeNose C19 for detecting VOCs, we performed additional sensing assessments for acetone vapors in a modified test setup (see **Figure S2 in the Supplementary Information**). This gas sensing configuration utilizing microsyringe for vapor injection has already been used in our former experiments for other VOC sensor types (e.g., nanofiber-functionalized QCMs).^{36,70,71} Acetone was chosen as a VOC model because it is not only produced in the rebreathed breath (0.8 to 2.0 ppm)⁷² but also indicated as one of the significant breath-borne COVID-19 biomarkers based on the study by Chen et al.⁶² Moreover, in the clinical practices, breath-containing acetone has also been extensively examined to diagnose some other diseases (i.e., lung cancer, diabetes mellitus, starvation, and ketogenic diet).⁷³ From the obtained measurement results shown in **Figures S2b and c**, the S3 and S7 sensors (or their extracted features of F3 and F7) demonstrated the poorest responses toward acetone vapors. Whereas the S2, S8, and S9 sensors were exhibiting higher sensitivities among others. The sensor output signals given by GeNose C19 data acquisition system agree well with

those measured by a calibrated digital voltmeter. Increasing the acetone vapor concentrations from 0.04 to 0.1 μL with 0.02 intervals resulted in higher responses of those three sensors (S2, S8, and S9), while the S3 and S7 sensors kept being irresponsive (see **Figure S2d** in the Supplementary Information). It is worth mentioning that each vapor concentration was measured ten times to acquire quantitative results. Lastly, as depicted in **Figure S2e**, linear discriminant analysis (LDA) was able to discriminate the output voltages produced by the sensors during their exposures to four different acetone concentrations (i.e., 0.04–0.1 μL).

In terms of ambient conditions, temperature and humidity might influence the performances of metal oxide semiconductor sensors.⁷⁴ Thus, we also performed cross-sensitivity assessments in respect to those two parameters for all the employed GeNose C19 sensors (see **Figures S3 and S4 in the Supplementary Information**). From three repeated measurements, the sensors yielded similar responses indicating their reliable sensing results. The sensor resistance decreased (i.e., higher output voltage was obtained) when the temperature was ramped up from 40 to 46°C, and the humidity was kept stable at (30 \pm 1) %RH (see **Figure S3 in the Supplementary Information**). Differentiating from silicon micromechanical resonant sensors that have frequency shift interference caused by temperature-induced Young's modulus change (material softening),^{38,75} the resistance decrease of the employed metal oxide semiconductor sensors (e.g., *n*-type SnO₂ with a bandgap of 3.6 eV) at higher temperature is caused by the increasing number of electrons that have sufficient energy crossing to the conduction band and thus contributing to the conductivity.⁷⁶ As this is a natural characteristic of semiconductor materials, we overcome this effect in GeNose C19 by controlling the temperature inside the test chamber at relatively stable values (i.e., (42 \pm 2) °C) during sensing phase of the exhaled breath.

Similar to the trend shown in the cross-temperature test, the sensor resistance also dropped to a lower value resulting in higher output voltage when the relative humidity (RH) was raised from 30 to 35% and the temperature was set constant at (40 \pm 1) °C (see **Figure S4 in the Supplementary Information**). The electrical characteristics of metal oxide semiconductors change due to water-adsorption on their surface while being exposed to humid air. Two different mechanisms of chemisorption and physisorption processes take place to create the first layer (i.e., chemisorbed layer) and its subsequent films of water molecules (i.e., physisorbed water layers), respectively.⁷⁷ If the first chemisorbed layer has been formed, the successive layers of water molecules will be physically adsorbed on the first hydroxyl layer. Because of the high electrostatic fields in the chemisorbed layer, the dissociation of physisorbed water can easily occur, producing hydronium ion (H₃O⁺) groups. Here, it should be noted that the conduction mechanism relies on the coverage of adsorbed water on the metal oxide semiconductor. First, in the event only hydroxyl ions exist on the metal oxide surface, the charge carriers of protons (H⁺) resulting from hydroxyl dissociation will hop between adjacent hydroxyl groups. Second, after the water molecules have been adsorbed but not fully covered the oxide surfaces, the charge transfer will be dominated by H₃O⁺ diffusion on hydroxyl groups, despite the occurring proton transfer between adjacent water molecules in clusters. Finally, once the continuous film of the first physisorbed water has been formed (i.e., full coverage of metal oxide by physisorbed water layer), proton hopping between neighboring water

molecules in the continuous film will be responsible for the charge transport.⁷⁸ More detailed explanations of the sensing mechanism and adsorption of water molecules on metal oxide semiconductor surfaces have been described elsewhere.^{74,77,78} Again, from the conducted cross-sensitivity measurements (**Figure S4 in the Supplementary Information**), the signal changes of the GeNose C19 sensors affected by humidity are relatively low (i.e., < 100 mV) compared to those exposed to exhaled breaths (i.e., around 1 V, as shown in Figs. 3a **and b**). Thus, it can be concluded that both temperature and humidity will insignificantly influence the system performance during breath measurements.

2.3 GeNose C19 performance in COVID-19 profiling tests

GeNose C19 was employed in COVID-19 profiling tests at two hospitals (*i.e.*, Bhayangkara General Hospital in Sleman District (RS Bhayangkara) and Bambanglipuro COVID-19 Special Field Hospital in Bantul District (RSLKC Bantul)) to investigate its potency and functionality for COVID-19 detection. Eighty-three subjects were recruited, consisting of 43 and 40 subjects confirmed as positive COVID-19 (case group) and negative COVID-19 (control group), respectively (see Fig. 4). Two subjects from the case group were excluded due to the deterioration of their clinical conditions, and they were referred to the higher level of health facilities.

In terms of clinical characteristics of the tested patients, Table 1 shows that 79–80% of patients were asymptomatic. The estimated mean of length-of-stay (LoS) was (9.0 ± 1.6) days. Moreover, most participants are aged 18–44 without pre-existing comorbidities. Breath samples were obtained by excluding the invalid ones measured by GeNose C19, in which the total confirmed positive and negative COVID-19 breath samples were $n_{b(+)} = 333$ and $n_{b(-)} = 282$, respectively. Here, even though most patients were asymptomatic, the difference in the VOC pattern could be distinguished. However, it is important to consider the possible presence of confounding VOCs that arise from existing comorbid and confounders. Previous studies reported several comorbid and confounding factors (e.g., chronic obstructive pulmonary disease, asthma, tuberculosis, lung cancer, food, beverage, and smoking cigarette), which might affect the composition of VOCs.⁷⁹ Taken such consideration into account, we evaluated the possible difference in the VOC pattern emitted/detected from the breaths of patients with cardiovascular problem (coronary heart disease) and endocrine disorder (diabetes), which showed no significant difference compared to VOC emitted from patients without comorbid. Nonetheless, due to the less comorbid case obtained from our subjects, which is our limitation in the present study, the influence of existing comorbidities to the VOC pattern will be evaluated further in the next study.

Table 1

Clinical characteristics of the tested patients, including age, sex, comorbid condition, symptoms, and breath sample type. All the positive and negative COVID-19 patients have been tested and confirmed using reverse transcription-quantitative polymerase chain reaction (RT-qPCR).

Characteristics	RT-qPCR-confirmed positive COVID-19 ($n_{p(+)} = 43$)	RT-qPCR-confirmed negative COVID-19 ($n_{p(-)} = 40$)	Total number	<i>p</i> -value
Age distribution (years old)				0.239
0–17	2	0	2	
18–44	27	22	49	
45–64	13	18	31	
65–74	1	0	1	
Sex distribution				0.000
Male	30	6	36	
Female	13	34	47	
Comorbidities				0.554
Cardiovascular problems	4	3	7	
Allergy and atopic conditions	0	1	1	
Endocrine metabolic problem	1	0	1	
Respiratory problems	0	0	0	
No pre-existing comorbid	38	36	74	
Symptoms				0.667
Patients with symptoms	9	6	15	
Asymptomatic patients	34	34	68	
Breath samples				
Positive COVID-19 detected	318	12	330	

Characteristics	RT-qPCR-confirmed positive COVID-19 ($n_{p(+)} = 43$)	RT-qPCR-confirmed negative COVID-19 ($n_{p(-)} = 40$)	Total number	<i>p</i> -value
Negative COVID-19 detected	15	270	285	

Table 2

Key parameters of GeNose C19 obtained from the optimum machine learning algorithm (i.e., DNN) of all breath samples (training and testing datasets). These include sensitivity, specificity, and area under the curve (AUC).

2 × 2 Tables		RT-qPCR		Total	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
		(Positive)	(Negative)				
GeNose C19	(Positive)	318	12	330	95.5 (92.7–97.3)	95.7 (92.7–97.5)	95.6 (93.7–97.1)
	(Negative)	15	270	285			
Total		333	282	615			

During machine learning and AI optimization, 70% of breath samples were used as training materials for the AI-based software validation. The remaining 30% of samples were employed for accuracy testing using the first 70%. Later, all breath samples (100%) were processed to measure the sensitivity and specificity levels of all machine learning algorithms (i.e., support vector machine (SVM), stacked multilayer perceptron (MLP), deep neural network (DNN), and linear discriminant analysis (LDA)). The ones that were previously employed in training were re-tested using all those four AI models. Table 2 lists the most important parameters (i.e., sensitivity, specificity, and area under the curve (AUC)) of GeNose C19 obtained from all datasets processed with DNN model, which is considered as the most stable and optimum machine learning algorithm among others. The sensitivity and specificity levels were 95.5% (95% CI: 92.7–97.3%) and 95.7% (95% CI: 92.7–97.5%), respectively. Meanwhile, the AUC was found to be 95.6% (95% CI: 93.7–97.1%) It is important to note that those calculated values here were to show the reliability of DNN algorithm in predicting the trained and untrained data of the breath samples. It was not meant to express the real clinical accuracy, which has to be demonstrated in a double-blinded fashion in the subsequent study.

Table 2

Key parameters of GeNose C19 obtained from the most optimum machine learning algorithm (i.e., DNN) of all breath samples (training and testing datasets). These include sensitivity, specificity, and area under the curve (AUC).

2 × 2 Tables		RT-qPCR		Total	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)
		(Positive)	(Negative)				
GeNose C19	(Positive)	318	12	330	95.5 (92.7–	95.7 (92.7–	95.6 (93.7–
	(Negative)	15	270	285	97.3)	97.5)	97.1)
Total		333	282	615			

From **Figure S5 in the Supplementary Information**, the principal component analysis (PCA) plots of the first three principal components (PCs) naturally did not show any separation between positive and negative groups. Thus, the focus of the next analysis was to use supervised machine learning models of LDA, SVM, MLP, and DNN. Classification of two different exhaled breath datasets (positive and negative) was first performed using linear discriminant analysis (LDA) to simplify their high-dimensional data complexity while keeping both patterns and trends. We transformed the obtained data into fewer dimensions acting as feature summaries. LDA is a supervised learning method that can also reduce the dataset dimension. LDA has focused more on the characteristics of the different classes to discriminate them, in which each class can then be optimized.⁸⁰ Fig. 5a shows the plot of the optimized LDA analysis for all the exhaled breath data. These results indicate that the most two classes can be well distinguished with small portion are still overlapped.

All datasets obtained during the sensing phase in the COVID-19 profiling test were utilized as inputs for all four classification models. To control the learning process, the hyper-parameters were optimized for SVM, MLP, and DNN models. Furthermore, the grid-search optimization algorithm (i.e., linear kernel function and radial basis function) was applied to tune the SVM parameters. From the optimized results of dataset preprocessing by StandardScaler, we found the cost value of 10 while the kernel function was linear. We also applied a genetic algorithm for tuning hyper-parameters of MLP and DNN. In the case of the MLP model, a tree-based pipeline optimization tool (TPOT) was implemented, which is basically an automated machine learning library. This TPOT library allowed the selection of a stacked MLP model with MinMaxScaler for preprocessing the dataset. Finally, to optimize the DNN model, we tuned the epochs, the number of batches, hidden layer, and neuron, dropout value, optimizer function, and activation function. A sigmoid function was used for the last activation function with binary cross-entropy loss. The applied procedure allowed us to employ the epochs of 500, the batch size of 5, the first hidden layer with 500 neurons and the “relu” activation function, the first 0.1 dropout layer, the second hidden layer with 250 neurons, and the “relu” activation function, the second 0.1 dropout layer, and the “adam” optimizer function. The 0.2 validation size and five early stopping procedures were used in the DNN training procedure to avoid over-fitting. In the hyper-parameter tuning procedure, data training and

10-fold cross-validation were applied. After obtaining the best parameters, the classification model was trained using training data and subsequently evaluated by testing data.

The performance of each machine learning model was represented by three types of metrics (i.e., sensitivity, specificity, and overall accuracy). Sensitivity and specificity are the proportion of correctly classified of actual positives and negatives, respectively. Meanwhile, accuracy is correctly classified proportion of both combined actual positives and negatives. Figure 5b-d show the accuracy, sensitivity, and specificity of all datasets measured by GeNose C19 during the profiling tests. The internal validation of the machine learning model was carried out by applying a repeated 10-fold cross-validation technique, which ensures that 70% of the training data is used for internal validation purposes in each iteration, being the other 30% of the data used for training purposes at each run. The DNN model possessed the most stable performance among others, which was indicated by high and similar accuracy values between training and testing data. In addition to those two performance parameters of sensitivity and specificity, we also investigated the receiver operating characteristic (ROC) curves with their resulting area under the curve (AUC) values of the training (Fig. 5e) and testing data (Fig. 5f) for all machine learning models. Among the machine learning models, DNN showed the most stable performance indicated by high and similar AUC values between training and testing data of 98.76% and 96.87%, respectively.

Currently, diagnostic methods used to screen COVID-19 are antigen test, rapid molecular test, and chest CT scan. Antigen tests have average sensitivity of 56.2% (95% CI: 29.5–79.8%) and average specificity of 99.5% (95% CI: 98.1–99.9%).⁸¹ The average sensitivity and specificity for the rapid molecular tests are 95.2% (95% CI: 86.7–98.3%) and 98.9% (95% CI: 97.3–99.5%), respectively.⁸¹ Meanwhile, chest CT scan possesses the average sensitivity and specificity of 87.9% (95% CI: 84.6–90.6%) and 80.0% (95% CI: 74.9–84.3%), respectively.⁸² These diagnostic methods have their drawbacks. The average sensitivity of antigen tests was not high, as shown by the study above, and it declines when the viral load decreases, which often happens to COVID-19 patients. Besides, the sample collection is invasive (by a nasopharyngeal or oropharyngeal swab). Rapid molecular test also employs an invasive sample collection method (by a nasopharyngeal or oropharyngeal swab), and the turn-around-time of point-of-care rapid molecular tests still take at least 20 minutes.⁸³ Chest CT scan exposes patients to radiation and is not specific. Compared to these diagnostic methods, GeNose C19 has the potency to be a screening test. A breath test with portable GeNose C19 is non-invasive and easy-to-use as it only requires patients to breathe to a sampling bag, has a fast turn-around time (~ 3 minutes), and does not have radiation concerns. The sensitivity and specificity results of GeNose C19 from these profiling tests show that combining GeNose C19 with an optimum machine learning algorithm can accurately distinguish between positive and negative COVID-19 patients, which then opens an opportunity for using this developed breathalyzer as a rapid, noninvasive COVID-19 screening device based on the exhaled breath-print identification.

We are aware that several factors may influence the breath-prints, i.e., pathological and disease-related conditions (smoking, comorbidities, medication), physiological factors (age, sex, food, beverages), and

sampling-related issues (bias with VOCs in the environment).⁸⁴ A previous study revealed that older age altered breath-prints in patients with lung cancer.⁸⁵ Food and beverages (e.g., poultry meat and plant oil) can influence the breath-prints, while smoking may increase the levels of benzene, 2-butanone and pentane and simultaneously decrease the level of butyl-acetate in the exhaled breaths.^{86–88} In our study, none of the patients smoked before the breath test. The comorbidities were also comparable between the case and control groups. There was no significant difference in food and beverages between the case and control groups. The measurements were conducted in the same environment for all participants. Thus, there was no bias with other interfering VOCs. However, it was important also to consider the possible presence of physiological variation resulting from physiological and biochemical changes of the body due to alteration in the respiratory rhythm affected by manipulated breathing technique.⁵² Therefore, in our work, the breath sampling was performed in such a defined protocol to collect only the third exhaled end-tidal breath, so that natural breathing pattern and rhythm can be preserved, resulting in minimal change of VOCs. We avoided excessive effort or repeated sampling in each breath collection where previous studies reported that it might alter the quality of collected VOCs.⁸⁹ The disturbance from other factors to breath test results is minimal. However, it is important to note that such confounding factors are most likely present in the real implementation and can affect at least the breath-print to a certain degree. Further study is now conducted to reveal the effects of various confounders.

Again, our study using GeNose C19 did not evaluate the distinctive concentration on each VOC found in breath samples of patient with positive or negative COVID-19. However, to investigate the types of VOCs produced in exhaled breaths of the positive and negative COVID-19 patients, we conducted another characterization based on GC-MS for several exhaled breaths of patients (see **Table S2 in the Supplementary Information**). From the extracted results, it was found that there was no significant difference in the composition of VOCs between patients with positive and negative COVID-19, suggesting that the difference in the breath-print pattern may be contributed more by the variation in concentration or proportion of several VOCs rather than the presence of one or two signature VOCs. For example, previous study reported that acetone was one of the VOCs with highest concentration emitted by healthy human.⁸⁹ However, in patients with positive COVID-19, acetone was reported to be in a lower proportion, compared to health care worker or healthy control group.⁶² Meanwhile, other VOC (i.e., ethyl butanoate) has also been reported as one of signature VOCs in COVID-19 patients where its concentration is slightly higher compared to healthy control.⁶²

Our breath analysis study will be extended in combination with GC-MS to identify the specific VOC biomarkers that are distinctive for COVID-19 cases in Indonesia. It has been found that varied VOC biomarkers for COVID-19 were detected in different countries across the world (i.e., Germany, UK, France, and China).^{25,62,64} Once they can be determined, a molecular imprinting method could be applied to generate highly selective sensors that target these VOC markers. Hence, the sensitivity and specificity of GeNose C19 and its overall accuracy can be further enhanced. Another critical step for the system development is conducting a diagnostic test with a large cohort to strongly elucidate its potency as a diagnostic tool in the near future.

3. Conclusions

A compact electronic nose (GeNose C19) has been developed, integrating metal oxide semiconductor gas sensors and machine learning analysis to enable rapid and noninvasive detection of coronavirus disease 2019 (COVID-19). Based on the profiling test results obtained at two hospitals in Indonesia with a total number of 615 breath samples (i.e., RT-qPCR-confirmed 333 positive and 282 negative COVID-19 samples) from 83 patients, the developed smart sensing system could clearly distinguish the difference of exhaled breath-prints between case (positive) and control (negative) groups. The sensitivity and specificity of the GeNose C19 using deep neural network (DNN) model were 95.5% (95% CI: 92.7–97.3%) and 95.7% (95% CI: 92.7–97.5%), respectively. This report has focused on the feasibility study in COVID-19 breath pattern analysis obtained by the electronic nose instead of finding specific volatile organic compound (VOC) biomarkers. A combined measurement of GeNose C19 with gas chromatography–mass spectrometry (GC-MS) will be carried out in the near future to answer the questions related to the distinctive VOCs for COVID-19. Moreover, a study on the clinical diagnosis of COVID-19 with a larger number of exhaled breath samples is being conducted to prove the device potency as a rapid COVID-19 screening tool.

4. Methods

4.1 Ethical statement

This study was approved by the Medical and Health Research Ethics Committee of Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada/Dr.Sardjito General Hospital, Yogyakarta, Indonesia, with the reference number KE/0489/05/2020 and had been registered in clinicaltrials.gov (NCT04558372).⁹⁰ All procedures were carried out following the relevant guidelines and regulations based on Good Clinical Practice and the Helsinki Declaration of 2013.⁹¹

4.2 Study design and patient recruitment

The study was designed as an open-label case-control prospective cohort study. Both of the healthcare workers and patients were informed regarding the given treatment. Subjects were recruited consecutively, and written informed consent was obtained before admission. Subjects, who were eligible in the study, were classified into two groups (i.e., case and control groups), as depicted in **Fig. 4**. The case group consisted of patients with confirmed positive COVID-19, who were admitted to the isolation rooms of two hospitals in the Special Region of Yogyakarta, Indonesia (i.e., Bhayangkara General Hospital in Sleman District (RS Bhayangkara) and Bambanglipuro COVID-19 Special Field Hospital in Bantul District (RSLKC Bantul)). COVID-19 infection was confirmed by the reverse transcription-quantitative polymerase chain reaction (RT-qPCR) tests on the SARS-CoV-2 ribonucleic acid (RNA) extracted from their nasopharyngeal and oropharyngeal swabs. Swab samples were taken in respective hospitals according to World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC) standards and sent to a certified national laboratory in Indonesia (i.e., Balai Besar Teknik Kesehatan Lingkungan dan

Pengendalian Penyakit (BBTKLPP)) for RT-qPCR tests. Subjects, who had worsening clinical symptoms and were unable to provide breath samples, were excluded from the study. All patients admitted to the isolation room received standard treatment based on the WHO guideline. The control group consisted of patients with diseases other than COVID-19, with two negative RT-qPCR results.

4.3 Medical treatment

All clinical data were recorded daily, including anthropometric data, laboratory results, and chest x-ray images. All patients received medical treatment according to the WHO and hospital treatment standard guideline.

4.4 GeNose C19 preconditioning

Preconditioning of GeNose C19 system needs to be performed before its usage since this developed sensor system possesses several limitations, especially related to the other interfering odor sources. Thus, several points designated as a standard operation procedure (SOP) should be followed to ensure a stable and proper working of the equipment to sniff out COVID-19. First, decent fresh air circulation should be assured during the operation of the GeNose C19 in both indoor and outdoor environments. Second, as the sensing system comprises thermally activated semiconductor metal oxide sensors (S1 – S10), it needs to be turned on and pre-heated for at least 30 min before its operation to obtain stable signal baselines. Each sensor is integrated with an internal heater enabling activation temperature at a few hundred degrees Celsius. Third, the gas monitoring chamber must be cleared away from the possible rest contaminants and the previously tested VOCs by selecting the “flushing mode” for 30 min. Here, the chamber will be circulated with the reference air from the ambient. Fourth, the patients should neither eat nor drink anything other than water for 1–3 h before exhaling their breath to air sampling bags. Short fasting assists the GeNose C19 to record a more accurate measure of the sensor signal pattern. Finally, the surrounding environment should not possess excessive interfering odor sources. The system needs to be analyzed in the absence of a sampling bag, in which the sensor signal increase should not be more than 400 mV during “analyze mode”. If the signal level reaches above that value, the GeNose C19 has to be moved to other more suitable and cleaner places.

4.5 Exhaled breath sampling procedure

Breath samples were collected from two different patient groups (i.e., RT-qPCR-confirmed positive and negative COVID-19) using a modified single-use non-rebreathing mask connected to sampling bags (see **Fig. 2**). Breath samples were taken every day during hospitalization in the isolation room until the day of discharge. We applied a safe breath collection procedure to ensure the safety of healthcare workers and other patients. The collectors (nurses), who previously received specialized, standardized training for breath sampling, were evaluated by Cohen's kappa method to assess their agreement procedure.⁹² In addition, they wore personal protective equipment (level 3) for coronavirus disease according to the standard and recommendation from the WHO, which includes respirator N95, gown, gloves, eye protection (face shield), and apron.⁹³ Subjects were instructed to take two initial normal breaths and exhale per

usual using their protective masks. Afterwards, they were requested to exhale their third breath, in the fashion of end-expiratory volume breath⁵¹, to the mask-integrated sampling bag entry-hole. The sampling bag was subsequently sealed and connected to the GeNose C19 machine via collecting hose and high-efficiency particulate air (HEPA)-filter (see **Fig. 2c**). Before being analyzed by the electronic nose, the breath sample was purified by a commercial HEPA-filter through polytetrafluoroethylene (PTFE) medical tube. It should be noted that the invalid breath samples were excluded from the data processing.

4.6 Pattern recognition algorithms

Before being analyzed by machine learning models, the obtained gas sensor array responses were subtracted to their initial baseline values and their features in the time domain were then extracted by four parameters, i.e., the maximum, median, standard deviation, and variance values (see **Fig. 3 and S6 in the Supplementary Information**). Pattern recognition was applied not only to obtain relevant data but also to eliminate redundant information. Hence, the classification model performance can be increased. In this study, those four feature extractions were employed as input parameters for the machine learning models, in which four classification algorithms (i.e., linear discriminant analysis (LDA), support vector machine (SVM), stacked multilayer perceptron (MLP), dan deep neural network (DNN)) were assigned. The receiver operating characteristic (ROC) graphs were also added to yield the sensitivity (true positive ratio function) and specificity (true negative ratio function) plots for all the parameter points. Each point in a ROC curve represents the coupled sensitivity/specificity related to a particular boundary decision. The area under the ROC curve (AUC) values were determined as a performance measure for machine learning algorithms related to their abilities in differentiating between positive and negative COVID-19 groups.

All datasets were divided into two groups (i.e., training and testing data) to evaluate each classification algorithm performance. The groups were randomly split with ratios of 70% and 30% for training and testing data, respectively. Ten-fold cross-validation was also employed to prevent overfitting. Moreover, LDA, which could be carried out without changing the parameters, acted as a baseline model. The SVM model has several significant parameters determining its classification performance (e.g., cost value, gamma-value, and kernel function). A grid-search procedure was applied to choose the optimum SVM parameters. A tree-based pipeline optimization tool (TPOT) library was also utilized to find the finest combination of MLP parameters using a genetic algorithm, in which 100 generations, 100 population size, 0.9 mutation rate, and 0.1 cross-over rate were applied in the computation. Furthermore, based on several layers having connected neurons, DNN with feed-forward and back-propagation algorithms was also attempted to be applied in GeNose C19 data analysis. Here, the main hyperparameters (e.g., the number of hidden layers and neuron, activation function, dropout layer, and optimizer function) were involved during data processing to yield the optimum model.

4.7 Characterizations of filter and volatile organic compounds

The morphology of fiber filter was analyzed using a scanning electron microscope (SEM, JSM-6510, JEOL Ltd., Tokyo, Japan). The types of volatile organic compounds (VOCs) contained in both RT-qPCR-

confirmed positive and negative COVID-19 breath samples were investigated employing a gas chromatography mass spectroscopy (GC-MS) machine (ISQ 7000 single quadrupole GC-MS system, Thermo Fisher Scientific Inc., Massachusetts, USA).

Declarations

Associated content

Supplementary information

Pre-conditioning of GeNose C19 to ambient air; GeNose C19 assessment toward acetone vapors; Effects of ambient temperature and humidity on sensor array; Principal component analysis (PCA); Exhaled breath data processing steps; Non-parametric ANOVA test; Gas chromatography-mass spectroscopy (GC-MS).

Author information

Corresponding authors:

Dian Kesumapramudya Nurputra

Department of Child Health, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281, Indonesia

Postgraduate Program in Clinical Medicine Science, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281, Indonesia

ORCID ID: <https://orcid.org/0000-0001-6262-3586>, E-mail: dian.k.nurputra@ugm.ac.id

Hutomo Suryo Wasisto

PT Nanosense Instrument Indonesia, Umbulharjo, Yogyakarta, 55167, Indonesia

ORCID ID: <https://orcid.org/0000-0002-4522-3625>, E-mail: h.wasisto@nanosense-id.com

Kuwat Triyana

Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara PO Box BLS 21, Yogyakarta 55281, Indonesia

ORCID ID: <https://orcid.org/0000-0002-1466-4364>, E-mail: triyana@ugm.ac.id

Author contribution

D.K.N., A.K., M.S.H., A.M.I.S., H.S.W., and K.T. conceived the idea and concept, designed the experiments, formulated the materials, validated the methods, interpreted the experimental results, and revised the

paper. H.S.W., D.K.N., and K.T. wrote the initial manuscript. S.N.H. and T.J. built the software and hardware of GeNose C19, performed the device quality control during tests in the hospital, and calibrated the sensors. S.N.H., A.H., and K.T. developed the artificial intelligence algorithms and analyzed their resulting patterns. H.S.W. and S.N.H created the artworks and plotted figures in the paper. B.S. supported the logistics during the test in hospitals. Y.M. provided input on the clinical characteristic analysis. D.K.N. collected the exhaled breath samples and carried out the administration of patients during test and hospitalization. D.K.N., M.S.H., Y.M., and A.M.I.S. collected and analyzed the data that were examined by RT-qPCR. H.S.W. supervised the work, led the project, and acquired the funding in PT Nanosense Instrument Indonesia. D.K.N. and K.T. supervised the work, led the project in the university, and acquired the funding. All authors discussed the results and approved the final manuscript.

Conflict of interest

Results presented in this paper (i.e., method and components of GeNose C19) have been partially patented in Indonesia. The patent application of GeNose C19 technology with a number of IDP000074761 entitled “*Unit Hidung Elektronik yang Dilengkapi Dehumidifier untuk Meningkatkan Unjuk Kerja terhadap Sampel Cairan*” has been submitted on November 29th, 2017 and subsequently granted on February 1st, 2021. The claims include stable sampling and purging system using micropump, sensor array arrangement, control and data acquisition system, and data analysis system.

Acknowledgement

This work was funded in part by the Indonesian State Intelligence Agency (BIN) under grant agreements no. 5471/UN1.P.V/DIT-KAUI/HK/2020 and PK-008/IX/2020, in part by the Indonesian Ministry of Research and Technology/National Research and Innovation Agency (Kemenristek/BRIN) and Indonesia Endowment Fund for Education (LPDP) within the project of “E-nose COVID-19: Electronic Nose Innovation for COVID-19 Screening in Indonesia” under grant agreement no. 10/FI/P-KCOVID-19.2B3/X/2020, and in part by Directorate of Business Development and Incubation (PUI) at Universitas Gadjah Mada (UGM) with project no. 1687/UN1.P.V/DIT-PUI/KA/2020. Shidiq Nur Hidayat, Trisna Julian, and Hutomo Suryo Wasisto are grateful for the administrative support from Muhamad Iqbal Nuriyana in PT Nanosense Instrument Indonesia during the development of GeNose C19. Support from the Science Techno Park (STP) at UGM is also acknowledged. Kuwat Triyana thanks Alfian Nur Wicaksono and James Covington from the University of Warwick, United Kingdom for the constructive discussion related to breath sampling procedure and Wahyono from the Computer Science Faculty of UGM for his expertise in artificial intelligence (AI) development. The authors thank all the involved patients, families, hospital directors (Theresia Lindawaty and Tarsisius Glory), research assistants (Diah Ariesa, Sania Rahmawati, Olivia R. Wiguna, and Stefanus Purwanto), and healthcare workers (medical doctors, nurses, and interns) for their technical support during the assessment of GeNose C19 in the hospitals (i.e., Bhayangkara General Hospital in Sleman District (RS Bhayangkara) and Bambanglipuro COVID-19 Special Field Hospital in Bantul District (RSLKC Bantul)).

Data availability

All data generated or analyzed during this study are included in this published article (and its Supplementary Information file).

References

- (1) Cui, J.; Li, F.; Shi, Z.-L. Origin and Evolution of Pathogenic Coronaviruses. *Nat. Rev. Microbiol.* **2019**, *17* (3), 181–192. <https://doi.org/10.1038/s41579-018-0118-9>.
- (2) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2020**, No. December. <https://doi.org/10.1038/s41579-020-00459-7>.
- (3) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727–733. <https://doi.org/10.1056/nejmoa2001017>.
- (4) Krammer, F. SARS-CoV-2 Vaccines in Development. *Nature* **2020**, *586* (7830), 516–527. <https://doi.org/10.1038/s41586-020-2798-3>.
- (5) Dhama, K.; Khan, S.; Tiwari, R.; Sircar, S.; Bhat, S.; Malik, Y. S.; Singh, K. P.; Chaicumpa, W.; Bonilla-Aldana, D. K.; Rodriguez-Morales, A. J. Coronavirus Disease 2019–COVID-19. *Clin. Microbiol. Rev.* **2020**, *33* (4). <https://doi.org/10.1128/CMR.00028-20>.
- (6) Yang, J.; Chen, X.; Deng, X.; Chen, Z.; Gong, H.; Yan, H.; Wu, Q.; Shi, H.; Lai, S.; Ajelli, M.; Viboud, C.; Yu, P. H. Disease Burden and Clinical Severity of the First Pandemic Wave of COVID-19 in Wuhan, China. *Nat. Commun.* **2020**, *11* (1), 1–10. <https://doi.org/10.1038/s41467-020-19238-2>.
- (7) Mathuria, J. P.; Yadav, R.; Rajkumar. Laboratory Diagnosis of SARS-CoV-2 - A Review of Current Methods. *J. Infect. Public Health* **2020**, *13* (7), 901–905. <https://doi.org/10.1016/j.jiph.2020.06.005>.
- (8) Tang, Y.-W.; Schmitz, J. E.; Persing, D. H.; Stratton, C. W. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *J. Clin. Microbiol.* **2020**, *58* (6). <https://doi.org/10.1128/JCM.00512-20>.
- (9) Ji, T.; Liu, Z.; Wang, G. Q.; Guo, X.; Akbar khan, S.; Lai, C.; Chen, H.; Huang, S.; Xia, S.; Chen, B.; Jia, H.; Chen, Y.; Zhou, Q. Detection of COVID-19: A Review of the Current Literature and Future Perspectives. *Biosens. Bioelectron.* **2020**, *166* (March), 112455. <https://doi.org/10.1016/j.bios.2020.112455>.
- (10) Park, M.; Won, J.; Choi, B. Y.; Lee, C. J. Optimization of Primer Sets and Detection Protocols for SARS-CoV-2 of Coronavirus Disease 2019 (COVID-19) Using PCR and Real-Time PCR. *Exp. Mol. Med.* **2020**, *52* (6), 963–977. <https://doi.org/10.1038/s12276-020-0452-7>.

- (11) Dramé, M.; Tabue Teguo, M.; Proye, E.; Hequet, F.; Hentzien, M.; Kanagaratnam, L.; Godaert, L. Should RT-PCR Be Considered a Gold Standard in the Diagnosis of COVID-19? *J. Med. Virol.* **2020**, *92* (11), 2312–2313. <https://doi.org/10.1002/jmv.25996>.
- (12) Harmon, S. A.; Sanford, T. H.; Xu, S.; Turkbey, E. B.; Roth, H.; Xu, Z.; Yang, D.; Myronenko, A.; Anderson, V.; Amalou, A.; Blain, M.; Kassin, M.; Long, D.; Varble, N.; Walker, S. M.; Bagci, U.; Ierardi, A. M.; Stellato, E.; Plensich, G. G.; Franceschelli, G.; Girlando, C.; Irmici, G.; Labella, D.; Hammoud, D.; Malayeri, A.; Jones, E.; Summers, R. M.; Choyke, P. L.; Xu, D.; Flores, M.; Tamura, K.; Obinata, H.; Mori, H.; Patella, F.; Cariati, M.; Carrafiello, G.; An, P.; Wood, B. J.; Turkbey, B. Artificial Intelligence for the Detection of COVID-19 Pneumonia on Chest CT Using Multinational Datasets. *Nat. Commun.* **2020**, *11* (1), 4080. <https://doi.org/10.1038/s41467-020-17971-2>.
- (13) Khatami, F.; Saatchi, M.; Zadeh, S. S. T.; Aghamir, Z. S.; Shabestari, A. N.; Reis, L. O.; Aghamir, S. M. K. A Meta-Analysis of Accuracy and Sensitivity of Chest CT and RT-PCR in COVID-19 Diagnosis. *Sci. Rep.* **2020**, *10* (1), 22402. <https://doi.org/10.1038/s41598-020-80061-2>.
- (14) Pokhrel, P.; Hu, C.; Mao, H. Detecting the Coronavirus (COVID-19). *ACS Sensors* **2020**, *5* (8), 2283–2296. <https://doi.org/10.1021/acssensors.0c01153>.
- (15) Sola Martínez, R. A.; Pastor Hernández, J. M.; Yanes Torrado, Ó.; Cánovas Díaz, M.; de Diego Puente, T.; Vinaixa Crevillent, M. Exhaled Volatile Organic Compounds Analysis in Clinical Pediatrics: A Systematic Review. *Pediatr. Res.* **2020**. <https://doi.org/10.1038/s41390-020-01116-8>.
- (16) Chin, S.-T.; Romano, A.; Doran, S. L. F.; Hanna, G. B. Cross-Platform Mass Spectrometry Annotation in Breathomics of Oesophageal-Gastric Cancer. *Sci. Rep.* **2018**, *8* (1), 5139. <https://doi.org/10.1038/s41598-018-22890-w>.
- (17) Capuano, R.; Khomenko, I.; Grasso, F.; Messina, V.; Olivieri, A.; Cappellin, L.; Paolesse, R.; Catini, A.; Ponzi, M.; Biasioli, F.; Di Natale, C. Simultaneous Proton Transfer Reaction-Mass Spectrometry and Electronic Nose Study of the Volatile Compounds Released by Plasmodium Falciparum Infected Red Blood Cells in Vitro. *Sci. Rep.* **2019**, *9* (1), 12360. <https://doi.org/10.1038/s41598-019-48732-x>.
- (18) Lindinger, W.; Hansel, A.; Jordan, A. On-Line Monitoring of Volatile Organic Compounds at Pptv Levels by Means of Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) Medical Applications, Food Control and Environmental Research. *Int. J. Mass Spectrom. Ion Process.* **1998**, *173* (3), 191–241. [https://doi.org/10.1016/s0168-1176\(97\)00281-4](https://doi.org/10.1016/s0168-1176(97)00281-4).
- (19) Phillips, M.; Herrera, J.; Krishnan, S.; Zain, M.; Greenberg, J.; Cataneo, R. N. Variation in Volatile Organic Compounds in the Breath of Normal Humans. *J. Chromatogr. B Biomed. Sci. Appl.* **1999**, *729* (1–2), 75–88. [https://doi.org/10.1016/S0378-4347\(99\)00127-9](https://doi.org/10.1016/S0378-4347(99)00127-9).
- (20) Phillips, M.; Cataneo, R. N.; Cummin, A. R. C.; Gagliardi, A. J.; Gleeson, K.; Greenberg, J.; Maxfield, R. A.; Rom, W. N. Detection of Lung Cancer with Volatile Markers in the Breath. *Chest* **2003**, *123* (6), 2115–

2123. <https://doi.org/10.1378/chest.123.6.2115>.

(21) Delfino, R. J.; Gong, H.; Linn, W. S.; Hu, Y.; Pellizzari, E. D. Respiratory Symptoms and Peak Expiratory Flow in Children with Asthma in Relation to Volatile Organic Compounds in Exhaled Breath and Ambient Air. *J. Expo. Sci. Environ. Epidemiol.* **2003**, *13* (5), 348–363.

<https://doi.org/10.1038/sj.jea.7500287>.

(22) van der Schee, M. P.; Hashimoto, S.; Schuurman, A. C.; Repelaer van Driel, J. S.; Adriaens, N.; van Amelsfoort, R. M.; Snoeren, T.; Regenboog, M.; Sprikkelman, A. B.; Haarman, E. G.; van Aalderen, W. M. C.; Sterk, P. J. Altered Exhaled Biomarker Profiles in Children during and after Rhinovirus-Induced Wheeze. *Eur. Respir. J.* **2015**, *45* (2), 440–448. <https://doi.org/10.1183/09031936.00044414>.

(23) Traxler, S.; Bischoff, A. C.; Saß, R.; Trefz, P.; Gierschner, P.; Brock, B.; Schwaiger, T.; Karte, C.; Blohm, U.; Schröder, C.; Miekisch, W.; Schubert, J. K. VOC Breath Profile in Spontaneously Breathing Awake Swine during Influenza A Infection. *Sci. Rep.* **2018**, *8* (1), 1–10. <https://doi.org/10.1038/s41598-018-33061-2>.

(24) MacLean, E.; Broger, T.; Yerlikaya, S.; Fernandez-Carballo, B. L.; Pai, M.; Denking, C. M. A Systematic Review of Biomarkers to Detect Active Tuberculosis. *Nat. Microbiol.* **2019**, *4* (5), 748–758. <https://doi.org/10.1038/s41564-019-0380-2>.

(25) Ruszkiewicz, D. M.; Sanders, D.; O'Brien, R.; Hempel, F.; Reed, M. J.; Riepe, A. C.; Bailie, K.; Brodrick, E.; Darnley, K.; Ellerkmann, R.; Mueller, O.; Skarysz, A.; Truss, M.; Wortelmann, T.; Yordanov, S.; Thomas, C. L. P.; Schaaf, B.; Eddleston, M. Diagnosis of COVID-19 by Analysis of Breath with Gas Chromatography-Ion Mobility Spectrometry - a Feasibility Study. *EClinicalMedicine* **2020**, *29–30*, 100609. <https://doi.org/10.1016/j.eclinm.2020.100609>.

(26) Röck, F.; Barsan, N.; Weimar, U. Electronic Nose: Current Status and Future Trends. *Chem. Rev.* **2008**, *108* (2), 705–725. <https://doi.org/10.1021/cr068121q>.

(27) Wilson, A. Application of Electronic-Nose Technologies and VOC-Biomarkers for the Noninvasive Early Diagnosis of Gastrointestinal Diseases. *Sensors* **2018**, *18* (8), 2613.

<https://doi.org/10.3390/s18082613>.

(28) Chen, C.-Y.; Lin, W.-C.; Yang, H.-Y. Diagnosis of Ventilator-Associated Pneumonia Using Electronic Nose Sensor Array Signals: Solutions to Improve the Application of Machine Learning in Respiratory Research. *Respir. Res.* **2020**, *21* (1), 45. <https://doi.org/10.1186/s12931-020-1285-6>.

(29) Sánchez, C.; Santos, J.; Lozano, J. Use of Electronic Noses for Diagnosis of Digestive and Respiratory Diseases through the Breath. *Biosensors* **2019**, *9* (1), 35.

<https://doi.org/10.3390/bios9010035>.

(30) Saktiawati, A. M. I.; Stienstra, Y.; Subronto, Y. W.; Rintiswati, N.; Sumardi; Gerritsen, J.-W.; Oord, H.; Akkerman, O. W.; van der Werf, T. S. Sensitivity and Specificity of an Electronic Nose in Diagnosing

- Pulmonary Tuberculosis among Patients with Suspected Tuberculosis. *PLoS One* **2019**, *14* (6), e0217963. <https://doi.org/10.1371/journal.pone.0217963>.
- (31) Wintjens, A. G. W. E.; Hintzen, K. F. H.; Engelen, S. M. E.; Lubbers, T.; Savelkoul, P. H. M.; Wesseling, G.; van der Palen, J. A. M.; Bouvy, N. D. Applying the Electronic Nose for Pre-Operative SARS-CoV-2 Screening. *Surg. Endosc.* **2020**. <https://doi.org/10.1007/s00464-020-08169-0>.
- (32) Di Pietrantonio, F.; Benetti, M.; Cannatà, D.; Verona, E.; Palla-Papavlu, A.; Fernández-Pradas, J. M.; Serra, P.; Staiano, M.; Varriale, A.; D'Auria, S. A Surface Acoustic Wave Bio-Electronic Nose for Detection of Volatile Odorant Molecules. *Biosens. Bioelectron.* **2015**, *67*, 516–523. <https://doi.org/10.1016/j.bios.2014.09.027>.
- (33) Roto, R.; Rianjanu, A.; Rahmawati, A.; Fatyadi, I. A.; Yulianto, N.; Majid, N.; Syamsu, I.; Wasisto, H. S.; Triyana, K. Quartz Crystal Microbalances Functionalized with Citric Acid-Doped Polyvinyl Acetate Nanofibers for Ammonia Sensing. *ACS Appl. Nano Mater.* **2020**, *3* (6), 5687–5697. <https://doi.org/10.1021/acsanm.0c00896>.
- (34) Rianjanu, A.; Julian, T.; Hidayat, S. N.; Yulianto, N.; Majid, N.; Syamsu, I.; Wasisto, H. S.; Triyana, K. Quartz Crystal Microbalance Humidity Sensors Integrated with Hydrophilic Polyethyleneimine-Grafted Polyacrylonitrile Nanofibers. *Sensors Actuators B Chem.* **2020**, 128286. <https://doi.org/10.1016/j.snb.2020.128286>.
- (35) Julian, T.; Hidayat, S. N.; Rianjanu, A.; Dharmawan, A. B.; Wasisto, H. S.; Triyana, K. Intelligent Mobile Electronic Nose System Comprising a Hybrid Polymer-Functionalized Quartz Crystal Microbalance Sensor Array. *ACS Omega* **2020**, *5* (45), 29492–29503. <https://doi.org/10.1021/acsomega.0c04433>.
- (36) Triyana, K.; Rianjanu, A.; Nugroho, D. B.; As'ari, A. H.; Kusumaatmaja, A.; Roto, R.; Suryana, R.; Wasisto, H. S. A Highly Sensitive Safrole Sensor Based on Polyvinyl Acetate (PVAc) Nanofiber-Coated QCM. *Sci. Rep.* **2019**, *9* (1), 15407. <https://doi.org/10.1038/s41598-019-51851-0>.
- (37) Wasisto, H. S.; Merzsch, S.; Uhde, E.; Waag, A.; Peiner, E. Handheld Personal Airborne Nanoparticle Detector Based on Microelectromechanical Silicon Resonant Cantilever. *Microelectron. Eng.* **2015**, *145*, 96–103. <https://doi.org/10.1016/j.mee.2015.03.037>.
- (38) Wasisto, H. S.; Merzsch, S.; Waag, A.; Uhde, E.; Salthammer, T.; Peiner, E. Airborne Engineered Nanoparticle Mass Sensor Based on a Silicon Resonant Cantilever. *Sensors Actuators B Chem.* **2013**, *180*, 77–89. <https://doi.org/10.1016/j.snb.2012.04.003>.
- (39) Wasisto, H. S.; Steib, F.; Merzsch, S.; Waag, A.; Peiner, E. Vertical Silicon Nanowire Array-Patterned Microcantilever Resonators for Enhanced Detection of Cigarette Smoke Aerosols. *Micro & Nano Lett.* **2014**, *9* (10), 676–679. <https://doi.org/10.1049/mnl.2014.0249>.

- (40) Park, S. Y.; Kim, Y.; Kim, T.; Eom, T. H.; Kim, S. Y.; Jang, H. W. Chemoresistive Materials for Electronic Nose: Progress, Perspectives, and Challenges. *InfoMat* **2019**, *1* (3), 289–316. <https://doi.org/10.1002/inf2.12029>.
- (41) Utari, L.; Septiani, N. L. W.; Suyatman; Nugraha; Nur, L. O.; Wasisto, H. S.; Yulianto, B. Wearable Carbon Monoxide Sensors Based on Hybrid Graphene/ZnO Nanocomposites. *IEEE Access* **2020**, *8*, 49169–49179. <https://doi.org/10.1109/ACCESS.2020.2976841>.
- (42) Casals, O.; Markiewicz, N.; Fabrega, C.; Gràcia, I.; Cané, C.; Wasisto, H. S.; Waag, A.; Prades, J. D. A Parts Per Billion (Ppb) Sensor for NO₂ with Microwatt (MW) Power Requirements Based on Micro Light Plates. *ACS Sensors* **2019**, *4* (4), 822–826. <https://doi.org/10.1021/acssensors.9b00150>.
- (43) Markiewicz, N.; Casals, O.; Fabrega, C.; Gràcia, I.; Cané, C.; Wasisto, H. S.; Waag, A.; Prades, J. D. Micro Light Plates for Low-Power Photoactivated (Gas) Sensors. *Appl. Phys. Lett.* **2019**, *114* (5), 053508. <https://doi.org/10.1063/1.5078497>.
- (44) Engel, L.; Benito-Altamirano, I.; Tarantik, K. R.; Pannek, C.; Dold, M.; Prades, J. D.; Wöllenstein, J. Printed Sensor Labels for Colorimetric Detection of Ammonia, Formaldehyde and Hydrogen Sulfide from the Ambient Air. *Sensors Actuators B Chem.* **2021**, *330*, 129281. <https://doi.org/10.1016/j.snb.2020.129281>.
- (45) Oweyung, R. E.; Panzer, M. J.; Sonkusale, S. R. Colorimetric Gas Sensing Washable Threads for Smart Textiles. *Sci. Rep.* **2019**, *9* (1), 5607. <https://doi.org/10.1038/s41598-019-42054-8>.
- (46) Thepudom, T.; Kladsomboon, S.; Pogfay, T.; Tuantranont, A.; Kerdcharoen, T. Portable Optical-Based Electronic Nose Using Dual-Sensors Array Applied for Volatile Discrimination. In *2012 9th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*; IEEE, 2012; pp 1–4. <https://doi.org/10.1109/ECTIcon.2012.6254208>.
- (47) Wasisto, H. S.; Prades, J. D.; Gülink, J.; Waag, A. Beyond Solid-State Lighting: Miniaturization, Hybrid Integration, and Applications of GaN Nano- and Micro-LEDs. *Appl. Phys. Rev.* **2019**, *6* (4), 041315. <https://doi.org/10.1063/1.5096322>.
- (48) Wu, W.; Stapelfeldt, F.-N.; Kroker, S.; Wasisto, H. S.; Waag, A. A Compact Calibratable Pulse Oximeter Based on Color Filters: Towards a Quantitative Analysis of Measurement Uncertainty. *IEEE Sens. J.* **2020**, 1–1. <https://doi.org/10.1109/JSEN.2020.3048118>.
- (49) Mariana, S.; Gülink, J.; Hamdana, G.; Yu, F.; Stempel, K.; Spende, H.; Yulianto, N.; Granz, T.; Prades, J. D.; Peiner, E.; Wasisto, H. S.; Waag, A. Vertical GaN Nanowires and Nanoscale Light-Emitting-Diode Arrays for Lighting and Sensing Applications. *ACS Appl. Nano Mater.* **2019**, *2* (7), 4133–4142. <https://doi.org/10.1021/acsanm.9b00587>.

- (50) Kononov, A.; Korotetsky, B.; Jahatspanian, I.; Gubal, A.; Vasiliev, A.; Arsenjev, A.; Nefedov, A.; Barchuk, A.; Gorbunov, I.; Kozyrev, K.; Rassadina, A.; Iakovleva, E.; Sillanpää, M.; Safaei, Z.; Ivanenko, N.; Stolyarova, N.; Chuchina, V.; Ganeev, A. Online Breath Analysis Using Metal Oxide Semiconductor Sensors (Electronic Nose) for Diagnosis of Lung Cancer. *J. Breath Res.* **2019**, *14* (1), 016004. <https://doi.org/10.1088/1752-7163/ab433d>.
- (51) Lawal, O.; Ahmed, W. M.; Nijsen, T. M. E.; Goodacre, R.; Fowler, S. J. Exhaled Breath Analysis: A Review of 'Breath-Taking' Methods for off-Line Analysis. *Metabolomics* **2017**, *13* (10), 110. <https://doi.org/10.1007/s11306-017-1241-8>.
- (52) Miekisch, W.; Kischkel, S.; Sawacki, A.; Liebau, T.; Mieth, M.; Schubert, J. K. Impact of Sampling Procedures on the Results of Breath Analysis. *J. Breath Res.* **2008**, *2* (2), 026007. <https://doi.org/10.1088/1752-7155/2/2/026007>.
- (53) Phillips, M. Method for the Collection and Assay of Volatile Organic Compounds in Breath. *Anal. Biochem.* **1997**, *247* (2), 272–278. <https://doi.org/10.1006/abio.1997.2069>.
- (54) Guzman, M. I. An Overview of the Effect of Bioaerosol Size in Coronavirus Disease 2019 Transmission. *Int. J. Health Plann. Manage.* **2020**, hpm.3095. <https://doi.org/10.1002/hpm.3095>.
- (55) Morawska, L.; Johnson, G. R.; Ristovski, Z. D.; Hargreaves, M.; Mengersen, K.; Corbett, S.; Chao, C. Y. H.; Li, Y.; Katoshevski, D. Size Distribution and Sites of Origin of Droplets Expelled from the Human Respiratory Tract during Expiratory Activities. *J. Aerosol Sci.* **2009**, *40* (3), 256–269. <https://doi.org/10.1016/j.jaerosci.2008.11.002>.
- (56) Lin, L.-W.; Hung, T.-Y. Swivel-HEPA-ETT (SHE) Bougie and HEPA-ETT (HE) Methods for Safe Intubation While Managing Patients with COVID-19. *Emerg. Med. J.* **2020**, emermed-2020-209625. <https://doi.org/10.1136/emered-2020-209625>.
- (57) Ari, A. Practical Strategies for a Safe and Effective Delivery of Aerosolized Medications to Patients with COVID-19. *Respir. Med.* **2020**, *167*, 105987. <https://doi.org/10.1016/j.rmed.2020.105987>.
- (58) Lin, T.; Lv, X.; Hu, Z.; Xu, A.; Feng, C. Semiconductor Metal Oxides as Chemoresistive Sensors for Detecting Volatile Organic Compounds. *Sensors* **2019**, *19* (2), 233. <https://doi.org/10.3390/s19020233>.
- (59) Wang, S.; Yang, J.; Zhang, H.; Wang, Y.; Gao, X.; Wang, L.; Zhu, Z. One-Pot Synthesis of 3D Hierarchical SnO₂ Nanostructures and Their Application for Gas Sensor. *Sensors Actuators B Chem.* **2015**, *207*, 83–89. <https://doi.org/10.1016/j.snb.2014.10.032>.
- (60) Huang, J.; Wu, J. Robust and Rapid Detection of Mixed Volatile Organic Compounds in Flow Through Air by a Low Cost Electronic Nose. *Chemosensors* **2020**, *8* (3), 73. <https://doi.org/10.3390/chemosensors8030073>.

- (61) Mirzaei, A.; Leonardi, S. G.; Neri, G. Detection of Hazardous Volatile Organic Compounds (VOCs) by Metal Oxide Nanostructures-Based Gas Sensors: A Review. *Ceram. Int.* **2016**, *42* (14), 15119–15141. <https://doi.org/10.1016/j.ceramint.2016.06.145>.
- (62) Chen, H.; Qi, X.; Ma, J.; Zhang, C.; Feng, H.; Yao, M. Breath-Borne VOC Biomarkers for COVID-19. *medRxiv* **2020**, 20136523. <https://doi.org/10.1101/2020.06.21.20136523>.
- (63) Shen, B.; Yi, X.; Sun, Y.; Bi, X.; Du, J.; Zhang, C.; Quan, S.; Zhang, F.; Sun, R.; Qian, L.; Ge, W.; Liu, W.; Liang, S.; Chen, H.; Zhang, Y.; Li, J.; Xu, J.; He, Z.; Chen, B.; Wang, J.; Yan, H.; Zheng, Y.; Wang, D.; Zhu, J.; Kong, Z.; Kang, Z.; Liang, X.; Ding, X.; Ruan, G.; Xiang, N.; Cai, X.; Gao, H.; Li, L.; Li, S.; Xiao, Q.; Lu, T.; Zhu, Y.; Liu, H.; Chen, H.; Guo, T. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **2020**, *182* (1), 59-72.e15. <https://doi.org/10.1016/j.cell.2020.05.032>.
- (64) Grassin-Delyle, S.; Roquencourt, C.; Moine, P.; Saffroy, G.; Carn, S.; Heming, N.; Fleuriet, J.; Salvator, H.; Naline, E.; Couderc, L.-J.; Devillier, P.; Thévenot, E. A.; Annane, D. Metabolomics of Exhaled Breath in Critically Ill COVID-19 Patients: A Pilot Study. *EBioMedicine* **2021**, *63*, 103154. <https://doi.org/10.1016/j.ebiom.2020.103154>.
- (65) Davis, C. E.; Schivo, M.; Kenyon, N. J. A Breath of Fresh Air – the Potential for COVID-19 Breath Diagnostics. *EBioMedicine* **2021**, *63*, 103183. <https://doi.org/10.1016/j.ebiom.2020.103183>.
- (66) Gupta, A.; Madhavan, M. V.; Sehgal, K.; Nair, N.; Mahajan, S.; Sehrawat, T. S.; Bikdeli, B.; Ahluwalia, N.; Ausiello, J. C.; Wan, E. Y.; Freedberg, D. E.; Kirtane, A. J.; Parikh, S. A.; Maurer, M. S.; Nordvig, A. S.; Accili, D.; Bathon, J. M.; Mohan, S.; Bauer, K. A.; Leon, M. B.; Krumholz, H. M.; Uriel, N.; Mehra, M. R.; Elkind, M. S. V.; Stone, G. W.; Schwartz, A.; Ho, D. D.; Bilezikian, J. P.; Landry, D. W. Extrapulmonary Manifestations of COVID-19. *Nat. Med.* **2020**, *26* (7), 1017–1032. <https://doi.org/10.1038/s41591-020-0968-3>.
- (67) Kaushik, A.; Kumar, R.; Arya, S. K.; Nair, M.; Malhotra, B. D.; Bhansali, S. Organic–Inorganic Hybrid Nanocomposite-Based Gas Sensors for Environmental Monitoring. *Chem. Rev.* **2015**, *115* (11), 4571–4606. <https://doi.org/10.1021/cr400659h>.
- (68) Wang, S.; Kang, Y.; Wang, L.; Zhang, H.; Wang, Y.; Wang, Y. Organic/Inorganic Hybrid Sensors: A Review. *Sensors Actuators B Chem.* **2013**, *182*, 467–481. <https://doi.org/10.1016/j.snb.2013.03.042>.
- (69) Hoffmann, M. W. G.; Mayrhofer, L.; Casals, O.; Caccamo, L.; Hernandez-Ramirez, F.; Lilienkamp, G.; Daum, W.; Moseler, M.; Waag, A.; Shen, H.; Prades, J. D. A Highly Selective and Self-Powered Gas Sensor Via Organic Surface Functionalization of p-Si/n-ZnO Diodes. *Adv. Mater.* **2014**, *26* (47), 8017–8022. <https://doi.org/10.1002/adma.201403073>.
- (70) Rianjanu, A.; Nurfani, E.; Arif, M. F.; Triyana, K.; Wasisto, H. S. Stability Evaluation of Quartz Crystal Microbalances Coated with Polyvinyl Acetate Nanofibrous Mats as Butanol Vapor Sensors. *Mater. Today Commun.* **2020**, 101770. <https://doi.org/10.1016/j.mtcomm.2020.101770>.

- (71) Rianjanu, A.; Aflaha, R.; Khamidy, N. I.; Djamal, M.; Triyana, K.; Wasisto, H. S. Room-Temperature Ppb-Level Trimethylamine Gas Sensors Functionalized with Citric Acid-Doped Polyvinyl Acetate Nanofibrous Mats. *Mater. Adv.* **2021**, *2* (11), 3705–3714. <https://doi.org/10.1039/D1MA00152C>.
- (72) Anderson, J. C.; Lamm, W. J. E.; Hlastala, M. P. Measuring Airway Exchange of Endogenous Acetone Using a Single-Exhalation Breathing Maneuver. *J. Appl. Physiol.* **2006**, *100* (3), 880–889. <https://doi.org/10.1152/jappphysiol.00868.2005>.
- (73) Ruzsányi, V.; Péter Kalapos, M. Breath Acetone as a Potential Marker in Clinical Practice. *J. Breath Res.* **2017**, *11* (2), 024002. <https://doi.org/10.1088/1752-7163/aa66d3>.
- (74) Wang, C.; Yin, L.; Zhang, L.; Xiang, D.; Gao, R. Metal Oxide Gas Sensors: Sensitivity and Influencing Factors. *Sensors* **2010**, *10* (3), 2088–2106. <https://doi.org/10.3390/s100302088>.
- (75) Wasisto, H. S.; Merzsch, S.; Waag, A.; Uhde, E.; Salthammer, T.; Peiner, E. Evaluation of Photoresist-Based Nanoparticle Removal Method for Recycling Silicon Cantilever Mass Sensors. *Sensors Actuators A Phys.* **2013**, *202*, 90–99. <https://doi.org/10.1016/j.sna.2012.12.016>.
- (76) Butera, R. A.; Waldeck, D. H. The Dependence of Resistance on Temperature for Metals, Semiconductors, and Superconductors. *J. Chem. Educ.* **1997**, *74* (9), 1090. <https://doi.org/10.1021/ed074p1090>.
- (77) Blank, T. A.; Eksperiandova, L. P.; Belikov, K. N. Recent Trends of Ceramic Humidity Sensors Development: A Review. *Sensors Actuators B Chem.* **2016**, *228*, 416–442. <https://doi.org/10.1016/j.snb.2016.01.015>.
- (78) Anderson, J. H.; Parks, G. A. Electrical Conductivity of Silica Gel in the Presence of Adsorbed Water. *J. Phys. Chem.* **1968**, *72* (10), 3662–3668. <https://doi.org/10.1021/j100856a051>.
- (79) Hanna, G. B.; Boshier, P. R.; Markar, S. R.; Romano, A. Accuracy and Methodologic Challenges of Volatile Organic Compound–Based Exhaled Breath Tests for Cancer Diagnosis. *JAMA Oncol.* **2019**, *5* (1), e182815. <https://doi.org/10.1001/jamaoncol.2018.2815>.
- (80) Yilmaz, M. F.; Danisman, Y.; Larour, J.; Arantchouk, L. Linear Discriminant Analysis Based Predator-Prey Analysis of Hot Electron Effects on the X-Pinch Plasma Produced K-Shell Aluminum Spectra. *Sci. Rep.* **2019**, *9* (1), 11867. <https://doi.org/10.1038/s41598-019-47997-6>.
- (81) Dinnes, J.; Deeks, J. J.; Adriano, A.; Berhane, S.; Davenport, C.; Dittrich, S.; Emperador, D.; Takwoingi, Y.; Cunningham, J.; Beese, S.; Dretzke, J.; Ferrante di Ruffano, L.; Harris, I. M.; Price, M. J.; Taylor-Phillips, S.; Hooft, L.; Leeflang, M. M.; Spijker, R.; Van den Bruel, A. Rapid, Point-of-Care Antigen and Molecular-Based Tests for Diagnosis of SARS-CoV-2 Infection. *Cochrane Database Syst. Rev.* **2020**. <https://doi.org/10.1002/14651858.CD013705>.

- (82) Islam, N.; Ebrahimzadeh, S.; Salameh, J.-P.; Kazi, S.; Fabiano, N.; Treanor, L.; Absi, M.; Hallgrimson, Z.; Leeflang, M. M.; Hooft, L.; van der Pol, C. B.; Prager, R.; Hare, S. S.; Dennie, C.; Spijker, R.; Deeks, J. J.; Dinnes, J.; Jenniskens, K.; Korevaar, D. A.; Cohen, J. F.; Van den Bruel, A.; Takwoingi, Y.; van de Wijgert, J.; Damen, J. A.; Wang, J.; McInnes, M. D. Thoracic Imaging Tests for the Diagnosis of COVID-19. *Cochrane Database Syst. Rev.* **2021**. <https://doi.org/10.1002/14651858.CD013639.pub4>.
- (83) Hansen, G.; Marino, J.; Wang, Z.-X.; Beavis, K. G.; Rodrigo, J.; Labog, K.; Westblade, L. F.; Jin, R.; Love, N.; Ding, K.; Garg, S.; Huang, A.; Sickler, J.; Tran, N. K. Clinical Performance of the Point-of-Care Cobas Liat for Detection of SARS-CoV-2 in 20 Minutes: A Multicenter Study. *J. Clin. Microbiol.* **2021**, *59* (2). <https://doi.org/10.1128/JCM.02811-20>.
- (84) Bikov, A.; Lázár, Z.; Horvath, I. Established Methodological Issues in Electronic Nose Research: How Far Are We from Using These Instruments in Clinical Settings of Breath Analysis? *J. Breath Res.* **2015**, *9* (3), 034001. <https://doi.org/10.1088/1752-7155/9/3/034001>.
- (85) Bikov, A.; Hernadi, M.; Korosi, B. Z.; Kunos, L.; Zsamboki, G.; Sutto, Z.; Tarnoki, A. D.; Tarnoki, D. L.; Losonczy, G.; Horvath, I. Expiratory Flow Rate, Breath Hold and Anatomic Dead Space Influence Electronic Nose Ability to Detect Lung Cancer. *BMC Pulm. Med.* **2014**, *14* (1), 202. <https://doi.org/10.1186/1471-2466-14-202>.
- (86) Wojnowski, W.; Majchrzak, T.; Dymerski, T.; Gębicki, J.; Namieśnik, J. Portable Electronic Nose Based on Electrochemical Sensors for Food Quality Assessment. *Sensors* **2017**, *17* (12), 2715. <https://doi.org/10.3390/s17122715>.
- (87) Kushch, I.; Schwarz, K.; Schwentner, L.; Baumann, B.; Dzien, A.; Schmid, A.; Unterkofler, K.; Gastl, G.; Spaněl, P.; Smith, D.; Amann, A. Compounds Enhanced in a Mass Spectrometric Profile of Smokers' Exhaled Breath versus Non-Smokers as Determined in a Pilot Study Using PTR-MS. *J. Breath Res.* **2008**, *2* (2), 026002. <https://doi.org/10.1088/1752-7155/2/2/026002>.
- (88) Chen, X.; Wang, F.; Lin, L.; Dong, H.; Huang, F.; Ghulam Muhammad, K.; Chen, L.; Gorlova, O. Association of Smoking with Metabolic Volatile Organic Compounds in Exhaled Breath. *Int. J. Mol. Sci.* **2017**, *18* (11), 2235. <https://doi.org/10.3390/ijms18112235>.
- (89) Sukul, P.; Schubert, J. K.; Zanaty, K.; Trefz, P.; Sinha, A.; Kamysek, S.; Miekisch, W. Exhaled Breath Compositions under Varying Respiratory Rhythms Reflects Ventilatory Variations: Translating Breathomics towards Respiratory Medicine. *Sci. Rep.* **2020**, *10* (1), 14109. <https://doi.org/10.1038/s41598-020-70993-0>.
- (90) Nurputra, D. K. Genosvid Diagnostic Test for Early Detection of COVID-19 <https://clinicaltrials.gov/show/NCT04558372> (accessed Jan 3, 2021).
- (91) World Medical Association. World Medical Association Declaration of Helsinki. *JAMA* **2013**, *310* (20), 2191. <https://doi.org/10.1001/jama.2013.281053>.

(92) Watson, P. F.; Petrie, A. Method Agreement Analysis: A Review of Correct Methodology. *Theriogenology* **2010**, *73* (9), 1167–1179. <https://doi.org/10.1016/j.theriogenology.2010.01.003>.

(93) World Health Organization. Rational use of personal protective equipment for coronavirus disease (COVID-19): interim guidance https://apps.who.int/iris/bitstream/handle/10665/331498/WHO-2019-nCoV-IPCPPE_use-2020.2-eng.pdf (accessed Jan 4, 2021).

Figures

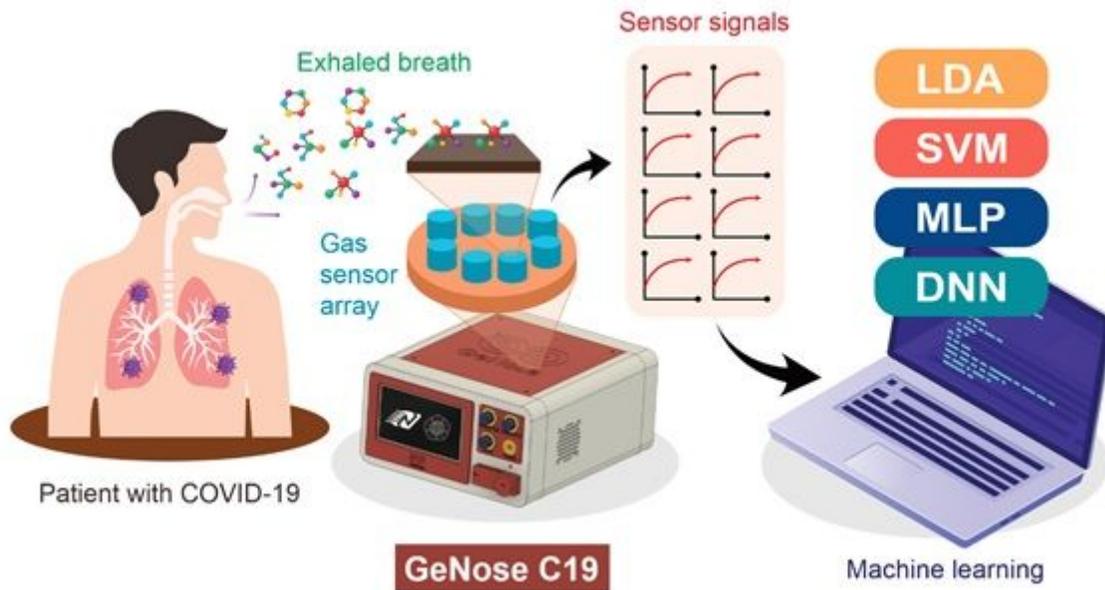


Figure 1

Illustration of the fast and noninvasive COVID-19 detection utilizing portable electronic nose (GeNose C19) integrated with artificial intelligence (AI). Four different machine learning algorithms (i.e., linear discriminant analysis (LDA), support vector machine (SVM), stacked multilayer perceptron (MLP), and deep neural network (DNN)) are employed to differentiate and classify the exhaled breath patterns of the patients, which are measured by ten different metal oxide semiconductor gas sensors.

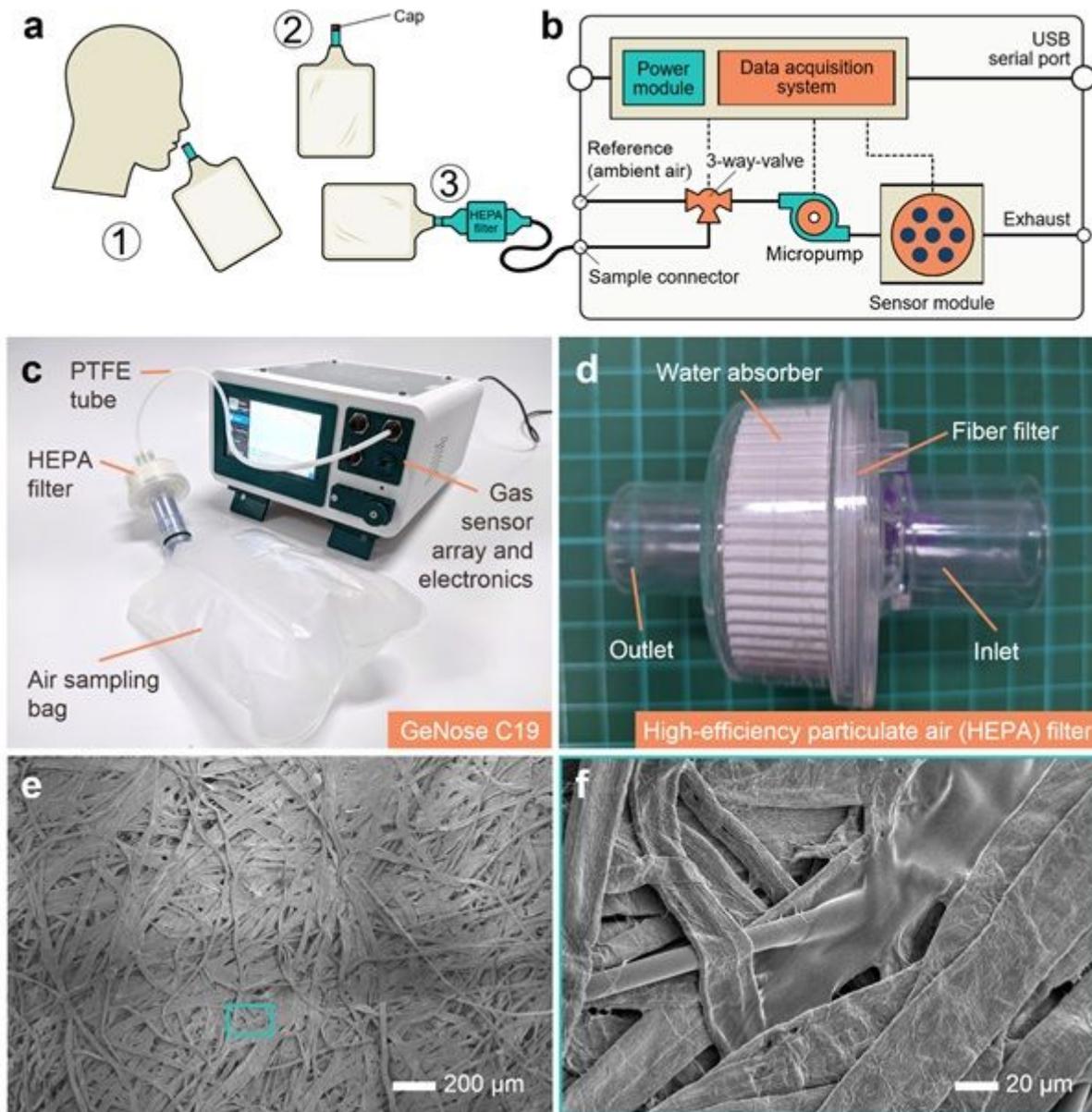


Figure 2

Integrated GeNose C19 system and its components. a Scheme of the off-line breath sampling pipeline for GeNose C19: (1) air inhaled through the nose and subsequently exhaled through the mouth to a sampling bag, (2) sealing or closing of the sampling bag cap to avoid the collected air leakage, and (3) direct plugging of the sampling bag into the electronic nose inlet (sample connector). b Diagram and c photograph of the GeNose C19 integrated with a high-efficiency particulate air (HEPA) filter and an air sampling bag through a flexible medical-grade polytetrafluoroethylene (PTFE) tube with an outer diameter of 4 mm. The electronic nose consists of several main electronic and mechanical parts (i.e., power module, data acquisition system, 3-way-valve, micropump, and sensor module inside the sealed gas chamber). The sensor module comprises ten different sensing devices arranged as an array. d HEPA filter for filtering out the particulate matters and trapping the SARS-CoV-2 available from the exhaled

breath of patient confirmed with positive COVID-19. e-f Scanning electron micrographs of fiber filter used in GeNose C19 filtering system.

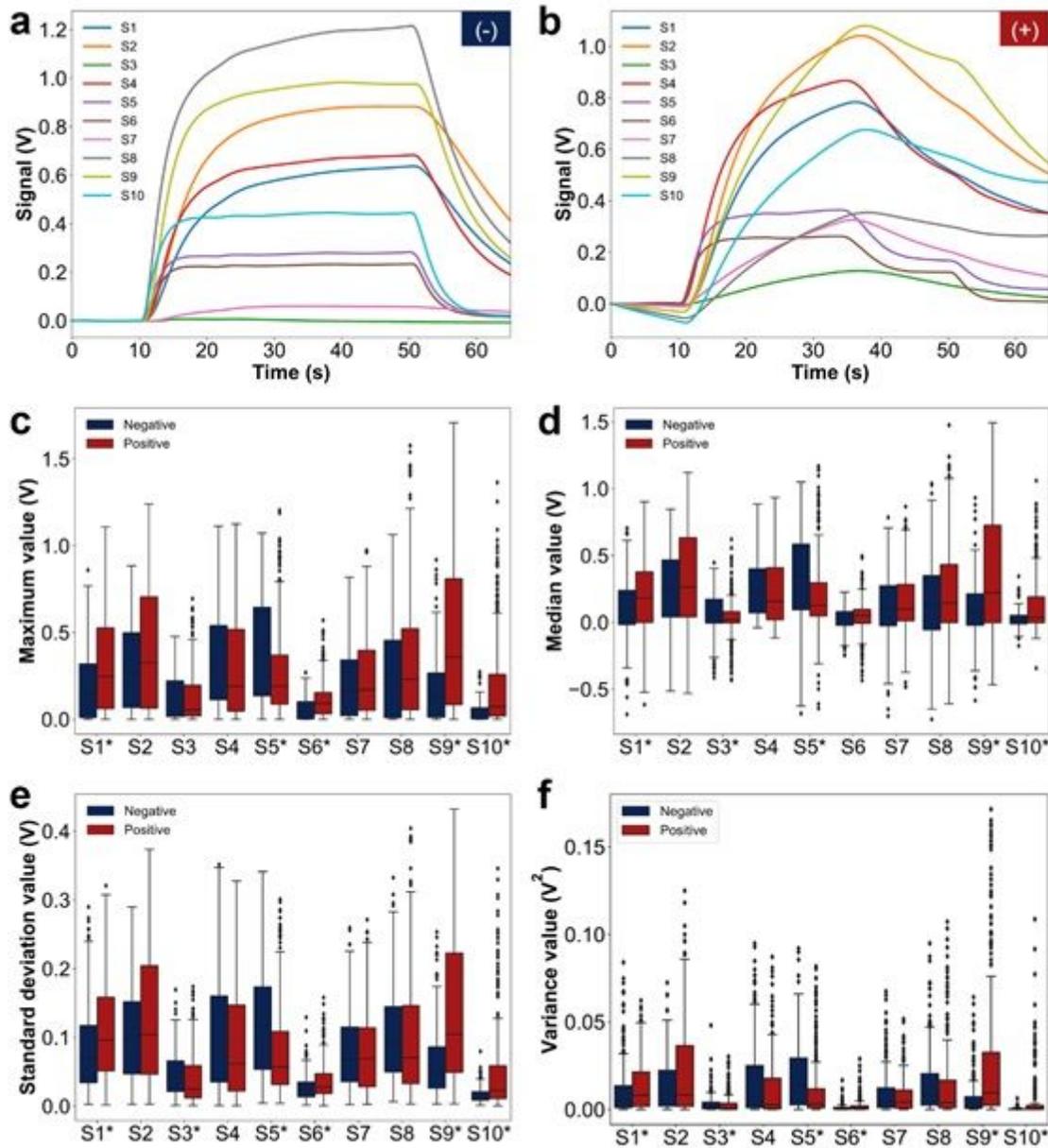


Figure 3

GeNose 19 sensor responses and their extracted features from the exhaled breaths of RT-qPCR-confirmed negative and positive COVID-19 patients. Typical sensing responses obtained from ten different conductometric gas sensors (S1 – S10), which are integrated into a portable GeNose C19 system for the exhaled breaths of RT-qPCR-confirmed a negative and b positive COVID-19 patients. Boxplots of the distribution for negative and positive COVID-19 samples based on feature extraction: c maximum, d median, e standard deviation, and f variance values.

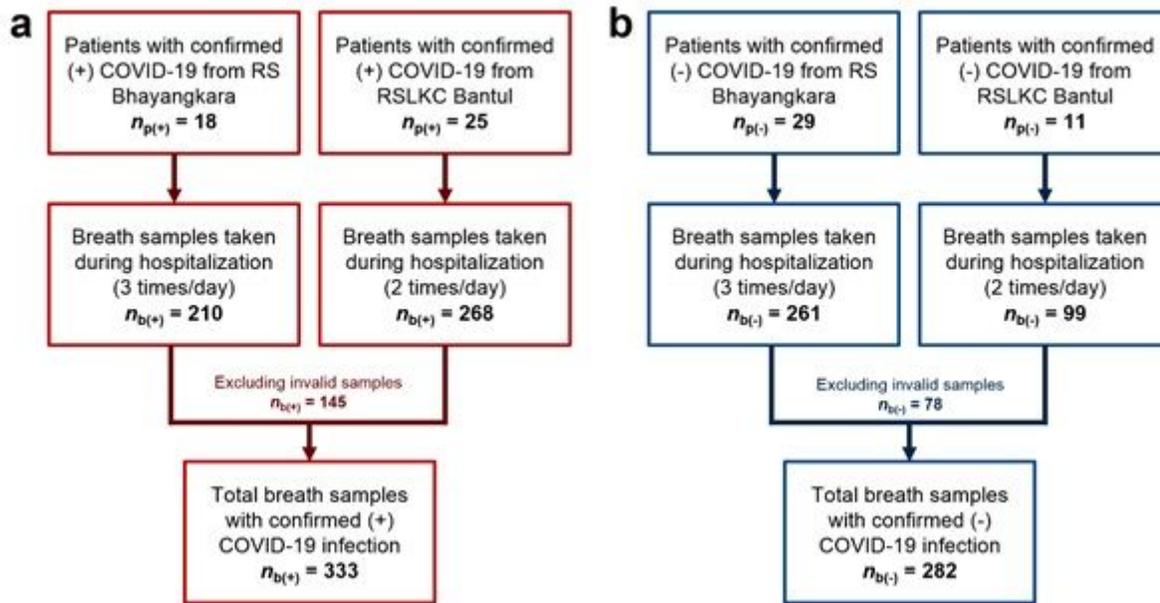


Figure 4

Exhaled breath collection procedure performed in the profiling test. Breath sampling procedure comprising patients with RT-qPCR-confirmed a positive ($n_{p(+)}$) and b negative ($n_{p(-)}$) COVID-19 infection in two hospitals (i.e., Bhayangkara General Hospital in Sleman District (RS Bhayangkara) and Bambanglipuro COVID-19 Special Field Hospital in Bantul District (RSLKC Bantul)). Both are located in the Special Region of Yogyakarta, Indonesia. Total breath samples were obtained by excluding the invalid ones measured by GeNose C19, in which the total confirmed positive and negative COVID-19 samples were $n_{b(+)} = 333$ and $n_{b(-)} = 282$, respectively.

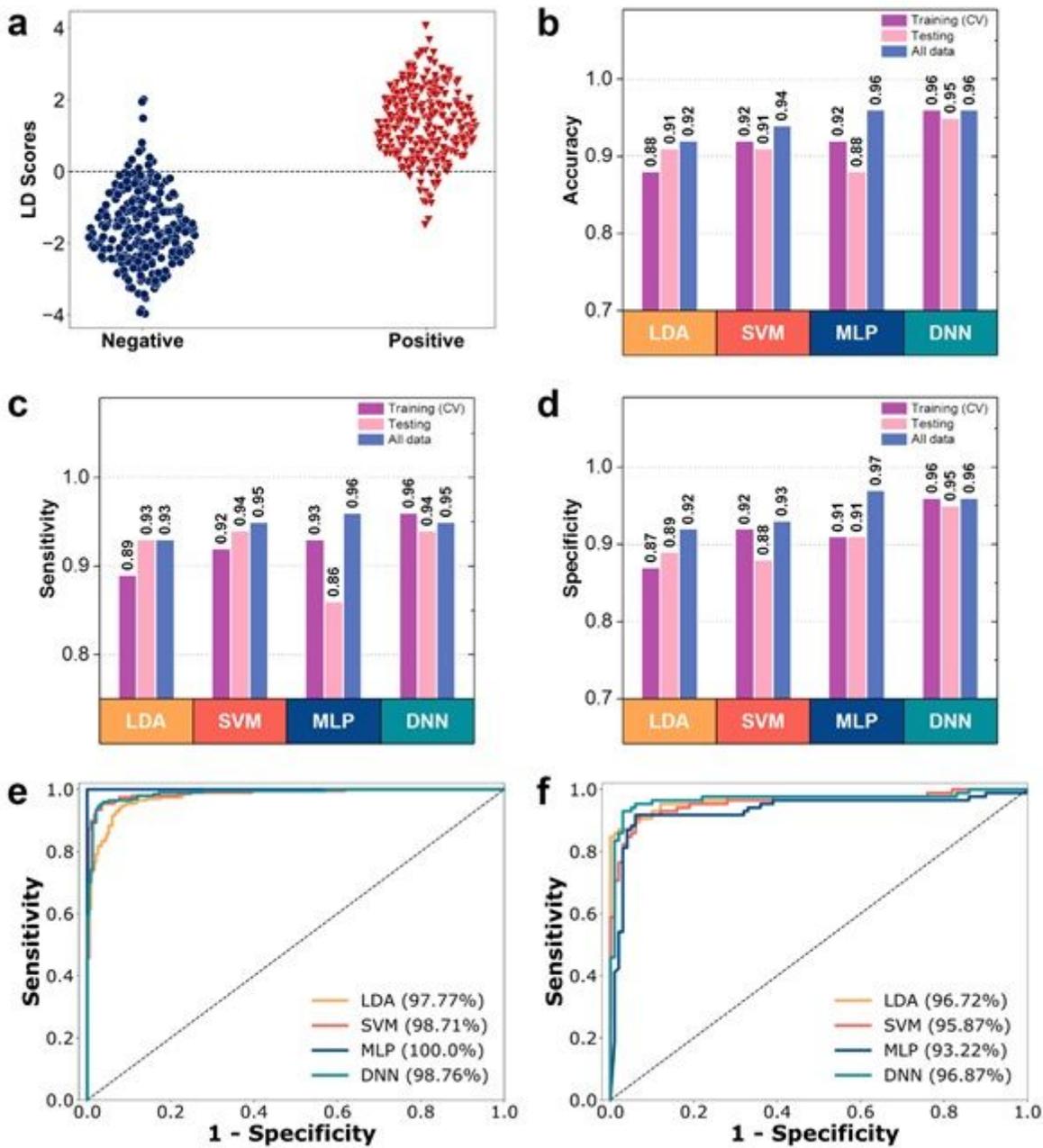


Figure 5

Measured breath data analysis using machine learning. a Classification of two different exhaled breath samples (i.e., positive and negative COVID-19 samples) using LDA model, b Overall accuracy (micro-averaged F1-score), c and d are sensitivity and specificity, respectively, of training (10-fold cross-validation), testing, and all datasets obtained by four different machine learning algorithms (LDA, SVM, MLP, and DNN), e and f are receiver operating characteristic (ROC) curves of the training and testing data, respectively, obtained by four different machine learning algorithms (LDA, SVM, MLP, and DNN).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [05ListoftablesGeNoseC19.docx](#)
- [03SIGeNoseC19Final.docx](#)
- [03SIGeNoseC19Final.docx](#)