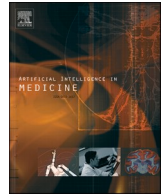




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Hybrid learning method based on feature clustering and scoring for enhanced COVID-19 breath analysis by an electronic nose

Shidiq Nur Hidayat<sup>a,b</sup>, Trisna Julian<sup>a</sup>, Agus Budi Dharmawan<sup>a,c</sup>, Mayumi Puspita<sup>a</sup>, Lily Chandra<sup>d</sup>, Abdul Rohman<sup>e</sup>, Madarina Julia<sup>f</sup>, Aditya Rianjanu<sup>g</sup>, Dian Kesumapramudya Nurputra<sup>f</sup>, Kuwat Triyana<sup>b,\*,1</sup>, Hutomo Suryo Wasisto<sup>a,1</sup>

<sup>a</sup> PT Nanosense Instrument Indonesia, Umbulharjo, Yogyakarta 55167, Indonesia

<sup>b</sup> Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Gadjah Mada, Sekip Utara, BLS 21, Yogyakarta 55281, Indonesia

<sup>c</sup> Faculty of Information Technology, Universitas Tarumanagara, Jl. Letjen S. Parman No. 1, Jakarta 11440, Indonesia

<sup>d</sup> RS Bhayangkara Polda Daerah Istimewa Yogyakarta, Jl. Raya Solo-Yogyakarta KM. 14, Sleman 55571, Indonesia

<sup>e</sup> Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Jl. Farmako Sekip Utara, Yogyakarta 55281, Indonesia

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

<sup>g</sup> Department of Materials Engineering, Institut Teknologi Sumatera, Terusan Ryacudu, Way Hui, Jati Agung, Lampung 35365, Indonesia

### ARTICLE INFO

#### Keywords:

Breath analysis  
Electronic nose  
Machine learning  
Feature permutation importance  
Hierarchical agglomerative clustering  
GeNose C19

### ABSTRACT

Breath pattern analysis based on an electronic nose (e-nose), which is a noninvasive, fast, and low-cost method, has been continuously used for detecting human diseases, including the coronavirus disease 2019 (COVID-19). Nevertheless, having big data with several available features is not always beneficial because only a few of them will be relevant and useful to distinguish different breath samples (i.e., positive and negative COVID-19 samples). In this study, we develop a hybrid machine learning-based algorithm combining hierarchical agglomerative clustering analysis and permutation feature importance method to improve the data analysis of a portable e-nose for COVID-19 detection (GeNose C19). Utilizing this learning approach, we can obtain an effective and optimum feature combination, enabling the reduction by half of the number of employed sensors without downgrading the classification model performance. Based on the cross-validation test results on the training data, the hybrid algorithm can result in accuracy, sensitivity, and specificity values of  $(86 \pm 3)\%$ ,  $(88 \pm 6)\%$ , and  $(84 \pm 6)\%$ , respectively. Meanwhile, for the testing data, a value of 87% is obtained for all the three metrics. These results exhibit the feasibility of using this hybrid filter-wrapper feature-selection method to pave the way for optimizing the GeNose C19 performance.

### 1. Introduction

Over the last few years, diagnostic and monitoring methods for human diseases in clinical medicine have been extended from invasive blood analysis to noninvasive breath pattern analysis [1–4]. Human exhaled breath has a complex composition of gases with various chemical compounds, which include small inorganic compounds (e.g., oxygen (O<sub>2</sub>), carbon dioxide (CO<sub>2</sub>), and nitric oxide (NO)), non-volatile organic compounds (VOCs) (e.g., isoprostanes, leukotrienes, cytosines, and hydrogen peroxide), and VOCs (e.g., hydrocarbons, ketones, alcohols, aldehydes, and esters) [5]. Due to their low solubility in blood, mixed VOCs resulting from cellular metabolism are easily exhaled and

can be used for breath analysis. They have been employed as diagnostic and prognostic response biomarkers for different respiratory diseases, including tuberculosis [6], pneumonia [7], asthma [8], lung cancer [9–12], and chronic obstructive pulmonary disease [13]. This great clinical potential has also led to a growing research area of breathomics, which generally refers to multidimensional analyses of VOCs in exhaled breath [14,15].

Because the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has emerged in late 2019 causing a pandemic of coronavirus disease 2019 (COVID-19) [16–19], the use of breathomics has been extended by researchers and clinicians to provide a fast and noninvasive COVID-19 test [20,21]. This is due to the fact that even though reverse

\* Corresponding author.

E-mail address: [triyana@ugm.ac.id](mailto:triyana@ugm.ac.id) (K. Triyana).

<sup>1</sup> Both authors contributed equally to this work.

transcription-quantitative polymerase chain reaction (RT-qPCR) has become the most recommended laboratory method and gold standard to diagnose COVID-19 with a high accuracy and high specificity, it has several drawbacks (i.e., needs of special equipment, invasive approach, well-trained staffs, and long result delivery) that limit its usage [22]. Thus, simple noninvasive tests with high accuracy are highly demanded to be alternative for RT-qPCR. In terms of the human breath analysis for COVID-19 detection, mass spectrometry (MS), which is versatilely coupled with various chromatographic separation methods, can be employed to better understand the clinical and biochemical processes of COVID-19 [20]. Among the available variants, the selected-ion flow-tube MS (SIFT-MS) [23], proton-transfer reaction MS (PTR-MS) [24,25], and gas chromatography-MS (GC-MS) [26] with thermal desorption or solid-phase microextraction have been attempted to be utilized for COVID-19 diagnosis, in which MS-detected exhaled breath biomarkers can identify RT-qPCR-confirmed positive COVID-19 patients with high sensitivity.

Although the MS approach has already been noninvasive, it is still considered expensive and time-consuming. Thus, an electronic nose (e-nose) can be selected as a rapid COVID-19 detector alternative. It is a device combining broad-spectrum chemical sensor array with a gas sampling chamber and machine learning to mimic human olfactory perception and provide a digital breathprint of the VOCs. Among other various gas sensor types (e.g., gravimetric microelectromechanical system (MEMS) and optical, capacitive, and photoacoustic gas sensors [27–40]), chemoresistive gas sensors based on the metal-oxide semiconductor (MOS) have been widely used as the main components of the e-nose considering their advantageous properties (i.e., high sensitivity, mature material synthesis technology, high robustness, low fabrication cost, short response time, and simple sensing method) [41].

Recently, several proof-of-concept studies for the real-time detection of the SARS-CoV-2 infection with the e-nose have been reported in several countries. First, in Tel Aviv, Israel, a compact PEN3 e-nose (AIRSENSE Analytics GmbH, Schwerin, Germany) had been successfully tested at a drive-through testing station using a one-way disposable sampling valve method to protect tested participants [42]. Second, in the Netherlands, another commercial e-nose so-called aeoNose (The eNose Company, Zutphen, the Netherlands) integrating three types of micro-hotplate MOS sensors (i.e., VOC, carbon monoxide, and nitrogen dioxide sensors) had been used for pre-operative SARS-CoV-2 screening at the Maastricht University Medical Center (MUMC+) [43]. The aeoNose could distinguish positive from negative COVID-19 participants based on VOC patterns in exhaled breath [43]. Third, a cloud-connected e-nose consisting of seven different cross-reactive MOS sensors (SpiroNose® – Breathomix, Leiden, The Netherlands) was also implemented as a SARS-CoV-2 diagnostic test in a public health setting (i.e., at different Amsterdam public test facilities) [44]. Fourth, using multiplexed nanomaterial-integrated sensor array, an exploratory clinical study cohort of the nanosensor-based e-nose concept for COVID-19 detection was conducted in Wuhan, China, in which the test dataset can reach an accuracy of up to 95% in differentiating between patients with COVID-19 and patients with other lung infections [45].

Despite these success stories, in the materials science community, MOS sensors also possess a cross-sensitivity drawback in detecting different types of VOCs and other gases. Basically, two optimization strategies can be opted, focusing on either the sensing active materials (e.g., employing hybrid organic-inorganic functional nanomaterials or molecular imprinting technique to increase the sensor selectivity [46,47]) or post-processing of MOS sensor output signals by machine learning [48–50]. The latter approach has been favorable for commercial e-nose developers because no modification is needed in the hardware assembly and setup, which consequently lowers the product development cost. Furthermore, because a large number of sensors integrated in the system does not necessarily result in a good e-nose performance, an effective method is required to determine and subsequently eliminate the less contributing sensors in certain

applications. In a chemometric study on an electronic tongue consisting of 16 sensors for dairy product discrimination, utilizing a linear discriminant analysis combined with a simulated annealing (SA) feature-selection algorithm could possibly produce an accurate model based on signals from only four sensors [51]. For breath analysis using an e-nose, combining the sparse group lasso (SGL) feature selection with a support vector machine (SVM) improves the classification performance in differentiating patients with lung cancer from healthy subjects and patients with benign pulmonary diseases [52]. However, despite its ability to enhance the classification of up to 12%, this SGL feature selection approach performs dimensionality reduction by eliminating several data rather than grouping and scoring for the whole data, leading to a poor discriminatory predictive performance [52,53]. Moreover, the maintenance of the e-nose performance using few sensors had not been demonstrated in that study.

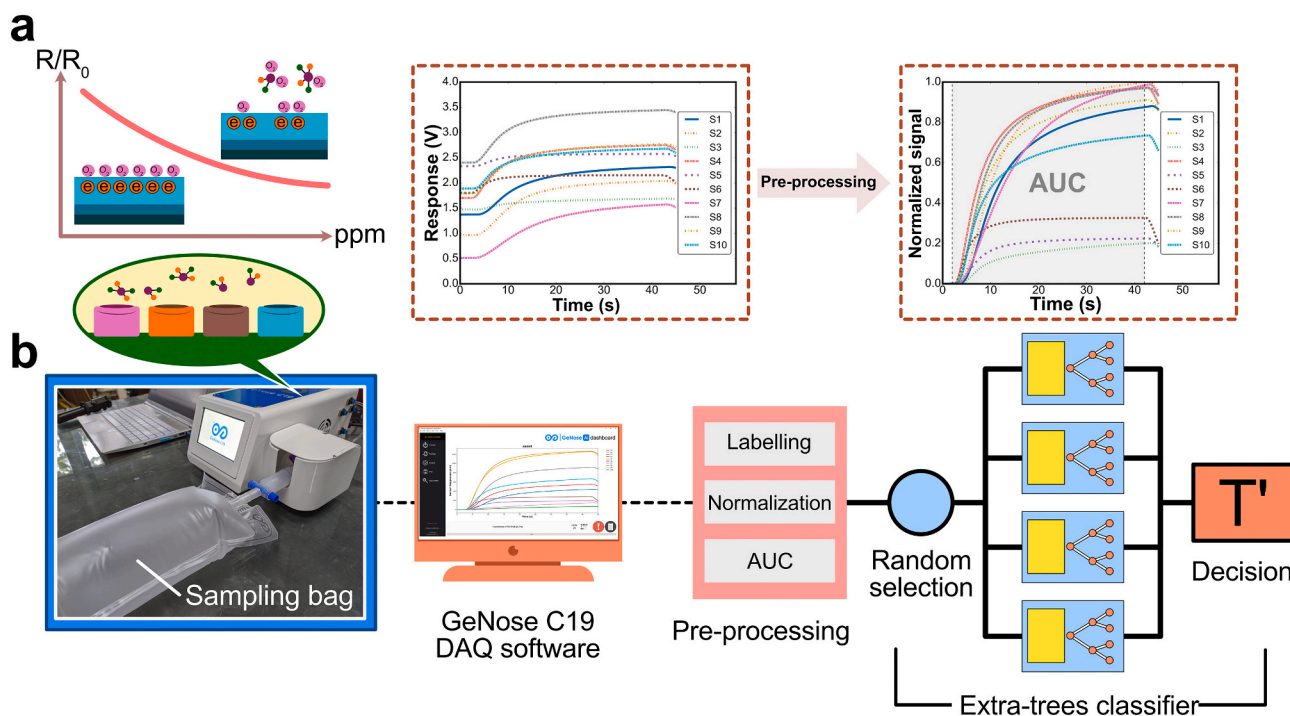
Therefore, herein, we propose a hybrid machine learning-based algorithm integrating hierarchical agglomerative clustering (HAC) analysis with the permutation feature importance method to not only improve the classification performance of a portable e-nose for COVID-19 detection (GeNose C19) but also to reduce the number of contributing sensors in the system. HAC that belongs to a family of unsupervised statistical approaches can classify a set of breath sample data into a hierarchy of groups or clusters according to their characteristic similarities. Meanwhile, the permutation feature importance approach is utilized to discover the characteristic uniqueness of each feature in the dataset and provide importance scoring for prediction. The developed algorithm has been implemented in an exploratory study of COVID-19 tests using GeNose C19 at a public hospital in Sleman, Indonesia, resulting in a promising performance enhancement and lowering the sensor number.

## 2. Results and discussion

### 2.1. GeNose C19 configuration for COVID-19 tests

In the portable e-nose for a rapid COVID-19 test (GeNose C19), main sensing and breath sampling units were developed and integrated into a system, enabling real device implementation at the hospital setting (Figs. 1 and S1). For the sensing module, 10 chemoresistive MOS sensors (S1–S10) with internal heaters and different gas selectivities were employed and arranged as an array inside a sealed chamber of GeNose C19 (Table 1). In the presence of target gases, their conductivity will change as induced by redox reactions between the active MOS materials and adsorbed gas molecules, as illustrated in Fig. 1a. Several other studies have described the sensing principle of MOS gas sensors in detail, which is based on an equilibrium shift of the surface chemisorbed oxygen reaction [54–56]. The output signals then underwent preprocessing steps (i.e., labeling, normalization, and area under the curve (AUC) determination) prior to machine learning-based data assessment (Fig. 1b). Moreover, the sensing unit was equipped with a micropump and three-way solenoid valve to enable and control the alternating flows of reference (ambient) air and exhaled breath to the chamber.

For the breath sampling unit, a high-efficiency particulate air (HEPA) filter and a disposable air sampling bag made of medical-grade polyvinyl chloride (PVC) were employed to filter out virus-containing droplets and store breath samples from patients, respectively. The breath sampling procedure was carefully conducted to ensure human safety and obtain reliable results. In this machine-learning algorithm development for GeNose C19, we used 460 exhaled breath samples from the COVID-19 tests of patients at a public hospital in Sleman, Daerah Istimewa Yogyakarta, Indonesia, from February to March 2021. Among them,  $n_P = 230$  and  $n_N = 230$  samples were confirmed by the gold-standard method of RT-qPCR as positive (P) and negative (N) COVID-19 (see Methods). Table 2 shows the clinical characteristics of the tested patients, including age, sex, and comorbid condition. Most participants are at the age of 6–78 without pre-existing comorbidities. Among the



**Fig. 1. Configuration of a portable e-nose for COVID-19 detection (GeNose C19).** **a** Output signal characteristics of chemoresistive metal-oxide-semiconductor (MOS) gas sensors. The sensor conductivity changes because of redox reactions between the active MOS material and adsorbed gas molecules. The real-time signal monitoring regarding VOC exposure to the sensor surface is performed using data logging software (DAQ software). **b** Procedure to collect the breath samples and process the data utilizing an extra-trees classifier. A hybrid learning algorithm combining hierarchical agglomerative clustering (HAC) analysis and permutation feature importance method enhances the GeNose C19 performance and simultaneously reduces the required sensor number.

**Table 1**  
**Selective target gases for all chemoresistive sensors used in GeNose C19.**  
 Cross-sensitivity toward different gases has been a typical characteristic for such inorganic MOS gas sensors. In terms of selectivity, each sensor sensitively reacts to more than two target gases.

Gas sensor	Selective target gases
S1	Carbon monoxide, ethanol, hydrogen, isobutane, and methane
S2	Ammonia, ethanol, hydrogen, hydrogen sulfide, and toluene
S3	Ethanol, hydrogen, isobutane, and methane
S4	Carbon monoxide, ethanol, hydrogen, isobutane, and methane
S5	Carbon monoxide, ethanol, hydrogen, isobutane, methane, and propane
S6	Carbon monoxide, ethanol, hydrogen, isobutane, methane, and propane
S7	Carbon monoxide, ethanol, hydrogen, and methane
S8	Acetone, benzene, carbon monoxide, ethanol, isobutane, methane, and n-hexane
S9	Ammonia, ethanol, hydrogen, and isobutane
S10	Chlorofluorocarbons, ethanol, and hydrofluorocarbons

positive COVID-19 patients, 64 of them were symptomatic. Moreover, the highest and lowest numbers of comorbidities found from the patients were respiratory (38) and gastroenteritis (11) problems, respectively.

**2.2. Preprocessing of sensor responses**

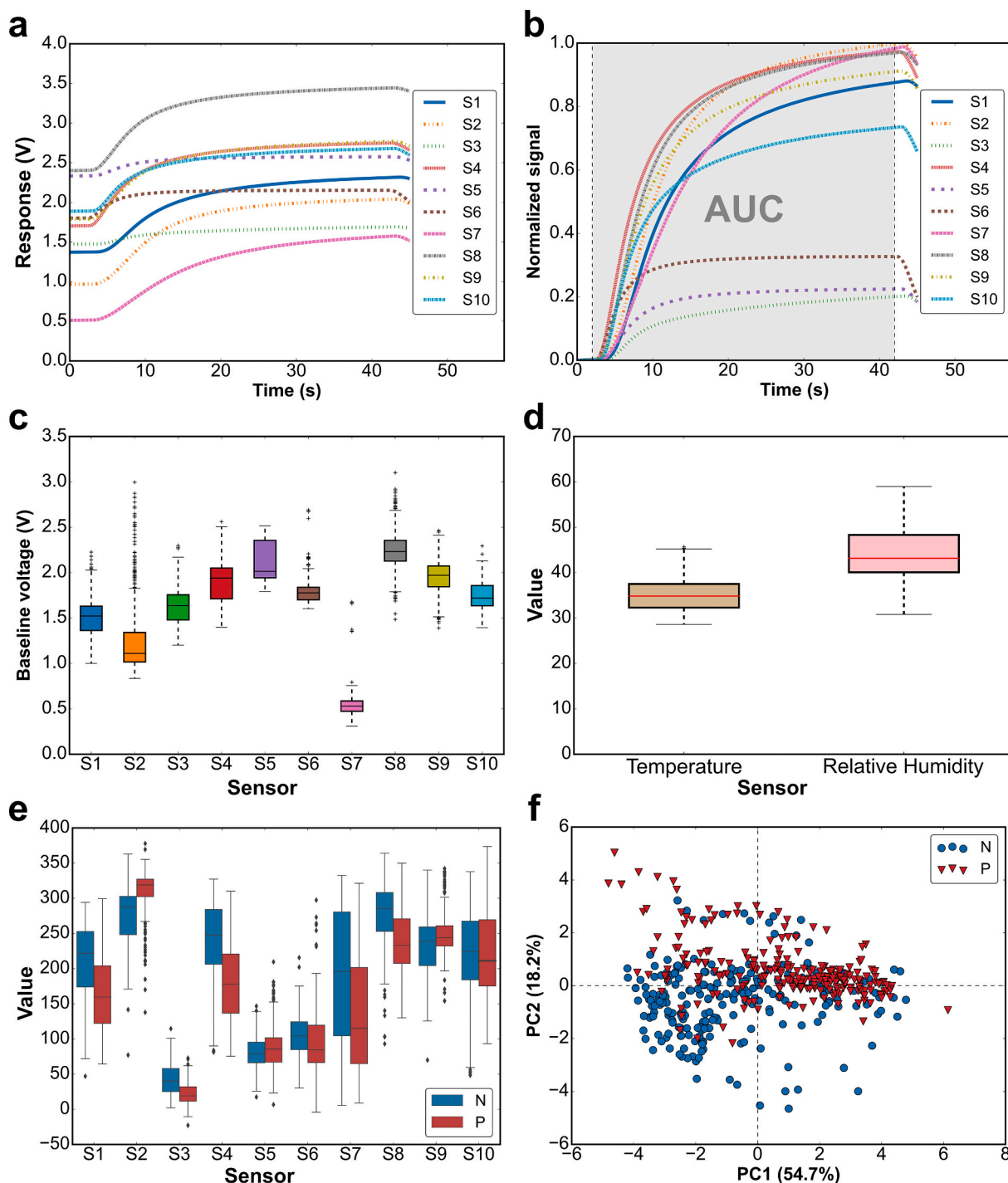
The typical responses of the GeNose C19 sensor array (S1–S10) during exposure to exhaled breath, which was randomly selected from the data distribution, are depicted in Fig. 2a. The sensor signals significantly increased for a few seconds after the sensing chamber was exposed to the breath sample. The signal then started to achieve a steady-state value, in which the maximum sensor response value was defined at a time of 42 s. The various signal shapes and shifts indicate that each sensor provides different and unique characteristics when exposed to breath samples depending on the employed active materials

**Table 2**  
**Clinical characteristics of tested patients, including age, sex, and comorbid condition.**  
 The numbers of the RT-qPCR-confirmed positive and negative COVID-19 patients are  $n_p = 230$  and  $n_N = 230$ , respectively.

Characteristics	RT-qPCR-confirmed positive COVID-19 ( $n_p = 230$ )	RT-qPCR-confirmed negative COVID-19 ( $n_N = 230$ )	Total number
Age distribution (years old)			
0–20	43	52	95
21–40	69	118	187
41–60	85	43	128
61–80	33	17	50
Sex distribution			
Male	127	161	288
Female	103	69	172
Patients with symptoms	64		
Comorbidities			
Respiratory problems	38		
Thermoregulation problems	26		
Anosmia and hypogeusia	16		
Gastroenteritis problems	11		
Systematic problems	14		

and target gases (Fig. 2b). The gray area represents the AUC employed as an input feature for the sensors. Different baseline signal levels were also noticeable among sensors even when the same breath sample was used during the test (Fig. 2a). Moreover, depending on the ambient or environmental condition, a sensor might yield altered baseline values during the continuous measurement of different breath samples.

To fully understand the sensor behavior, we calculated the baseline



**Fig. 2.** Raw and preprocessed data of sensor output signals recorded by GeNose C19 from breath measurements. **a** Typical raw and **b** normalized sensor signals in the breath measurements recorded by the GeNose C19 DAQ software (2 s baseline time, 40 s sampling time, and 3 s purging time). The distributions of the **c** baseline and **d** temperature and relative humidity values calculated from all 460 training data. **e** Calculated feature values of all sensors (S1–S10) based on their area under the curve (AUC) after signal preprocessing steps. **f** PC1 and PC2 plot showing the distributions of all P and N training data.

levels of 10 sensors from all the collected 460 breath data (Fig. 2c). Evidently, each sensor has different baseline values during measurements. Indicated by the significant amount of data outliers, the most and least stable baseline signals were achieved by S7 and S2, respectively. The large baseline signal variation is affected by different ambient conditions. In Fig. 2d, the average temperature and relative humidity values during the measurements were  $(35 \pm 4)^\circ\text{C}$  and  $(44 \pm 7)\%$ , respectively. This characterization was made feasible using a temperature and humidity sensor integrated inside the sensing chamber.

Generally, MOS sensors were highly influenced by the temperature and relative humidity [57–62]. This behavior has become one of the strongest limitations of the sensor array technology based on MOS sensors. This unstable signal baseline, combined with the mixture complexity of VOCs contained in a human breath sample, results in high challenges when developing chemometric model selection to differentiate between positive (P) and negative (N) COVID-19 patterns [52,63–66]. However, to suppress this effect at the minimum level, GeNose C19 can be pre-conditioned and placed in a location where the environmental condition

is relatively stable (e.g., indoor).

The feature value variabilities for P and N samples are displayed as a boxplot in Fig. 2e. A higher distribution of the N classes was observed as compared to the P classes for the S1, S3, S4, S7, and S8 features. However, these results were still insufficient to determine the importance of the sensors for performing the discrimination of P and N samples. Thus, we used principal component analysis (PCA) to thoroughly determine the variability distribution. Fig. 2f shows that PC1 and PC2 represent 72.9% data variability. This finding confirms a significant amount of data overlapping between the P and N samples, especially in the regions where PC1 is higher than 0 ( $PC1 > 0$ ), and PC2 approaches 0 ( $PC2 \rightarrow 0$ ). Nonetheless, the PCA result indicates that to some degree, the clustering of P and N data exists for the training data. The  $PC1 < 0$  region shows a higher trend of N compared to the P sample, which indicates the possibility of developing a classification model that can clearly differentiate between P and N data. Thus, from this point onward, we utilized an extra-tree classifier as the classification method.

### 2.3. Hybrid learning method for the classification and performance optimization

For the feature selection in machine learning, two different approaches are basically available to be implemented (i.e., filter and wrapper methods). Each of them can be either separately used or jointly combined. The filter method is developed using statistical filter algorithms (e.g., Pearson's correlation, information gain, mutual information, ANOVA, chi-squared test, and fast-correlation-based filter). Meanwhile, the wrapper approach utilizes learning algorithms to find the best combination of features that produce high performance (e.g., genetic algorithm (GA) feature selection, simulated annealing (SA) feature selection, recursive feature elimination (RFE), least absolute shrinkage and selection operator (LASSO) algorithm, and ridge regression). Filter-based feature selection has the advantages of low computational complexity, robustness against overfitting, fast processing, and good generalization. However, this method can obtain a classification model performance that is not optimum because it is independent of the used learning algorithm. By contrast, the wrapper method depends on the employed learning algorithm. Hence, it can more likely result in the best combination feature as compared with the filter-based method. However, the use of inappropriate wrapper methods may cause overfitting in the resulting model when the character of the test data changes. In addition, the wrapper approach typically requires a high computational cost and long processing time because of its complexity. This method also depends on the used learning model parameters. To overcome these drawbacks, a hyperparameter tuning procedure needs to be conducted to determine the best parameters of the employed learning algorithm. Nonetheless, altering parameters excessively may lead to excessive different results and their consequent combinations. This method definitely increases the complexity in the process of finding the optimum sensor combination [67,68]. Therefore, in this study, a hybrid learning technique was developed for optimally selecting the features by combining filter and wrapper methods [69].

In e-nose applications, using multiple MOS gas sensors does not always necessarily improve the system performance. Instead, it can result in a high cross-sensitivity of the sensor characteristics [67,68,70]. Thus, we applied a sensor array optimization method to determine the best sensor combination that can distinguish the optimum P and N labels and subsequently eliminate redundant sensors (i.e., sensors with less importance and contribution to the discrimination and classification of the distributed P and N data). Here, a combination of HAC and permutation feature importance was selected and used in the quick and efficient optimization procedure. The HAC was used to find correlations between features in all training data, and permutation feature importance was employed to discover the importance level of each sensor and subsequently distinguish different classes. In this case, we selected the Ward's linkage and extra-tree classifier as HAC distance linkage and

importance feature permutation estimator, respectively, which were iterated over 200 simulation times. The accuracy value was used as a metric in this process. The permutation feature importance method measured the increase in the model prediction error when it randomly changed the value or order of features. It will have an impact on deciding the correlation between the features and actual results.

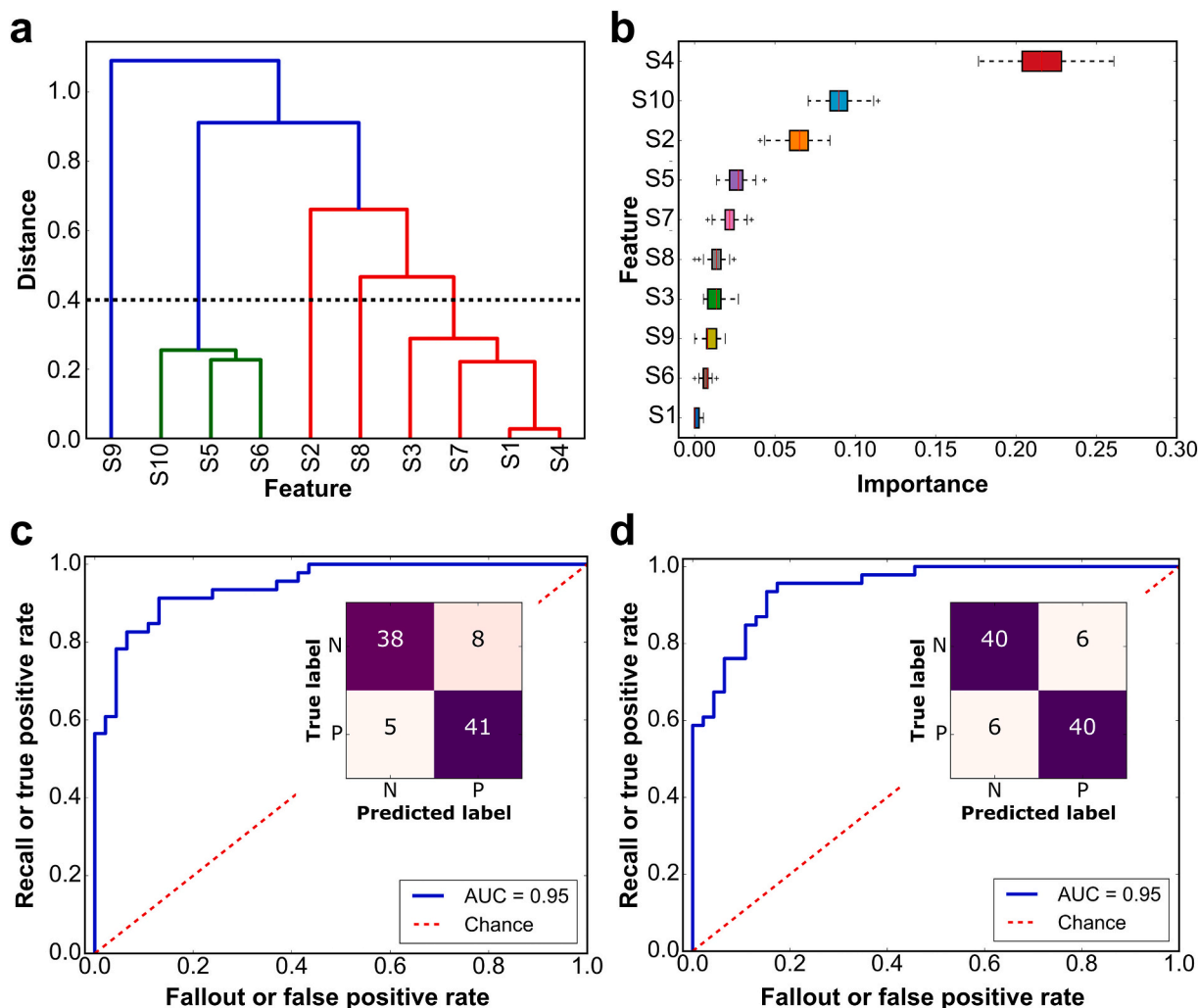
The extra-tree classifier uses a meta estimator that fits a number of randomized decision trees (i.e., extra trees) on various subsamples of the dataset and employs averaging to improve the predictive accuracy and control overfitting [71]. This classifier has several parameters, e.g., criterion, maximum of features (*max\_features*), minimum of the sample leaf (*min\_sample\_leaf*), and minimum of the sample split (*min\_sample\_split*). First, a hyperparameter tuning procedure was needed to determine the best basic extra-tree classifier model parameters. For this, the data were divided into two parts, i.e., training data (80% randomly) and testing data (20% randomly). The training and testing data were used to perform training procedures and validate training results, respectively. The combined parameters that needed to be identified are *criterion*, *max\_features*, *min\_sample\_leaf*, and *min\_sample\_split*. The grid-search method combined with the 10-fold cross-validation was applied to internally validate the selection of the best combination of parameters. In our case, the results show that the optimum basic extra-tree classifier model to distinguish P and N classes possessed *criterion* = entropy, *max\_features* = 1.0, *min\_sample\_leaf* = 2, and *min\_sample\_split* = 3. Implementing the extra-tree classifier model on testing data produced by all 10 sensors (S1–S10) could yield an accuracy of  $(86 \pm 3)\%$  based on the 5-fold cross-validation repeated 10 times (Table 3 and Fig. 3).

Figs. 3a and b display the HAC dendrogram and boxplot permutation importance of all data, respectively. The procedure for selecting the combination and number of sensors was gradually conducted by observing the selected features from the HAC (Fig. 3a) and then selecting the feature with the highest importance in one branch (Fig. 3b). The first combination of two sensors selected S4 and S9. S9 was chosen because it possessed its own branch, whereas the other features were in the same branch. Therefore, based on the boxplot results, S4 was then opted because it had the highest importance factor among the other features in the same branch. Through the same procedure, if the distance values were 0.8, 0.6, and 0.4, then the total sensor numbers of 3 (i.e., S4, S9, and S10), 4 (i.e., S4, S9, S10, and S2), and 5 (i.e., S4, S9, S10, S2, and S8), respectively, could be obtained. These results were supported by the GC–MS outcomes on the breath characteristics of four subjects, in which each subject had been tested three times (Table S1). Subjects A and B were RT-qPCR-confirmed with positive COVID-19, whereas subjects C and D were healthy individuals (RT-qPCR-tested negative COVID-19).

The selection criteria of the employed sensors in GeNose C19 listed in Table 1 were based on their abilities to possibly detect the potential compound-based biomarkers in COVID-19 and their compatibilities to be integrated with other controlling electronic components in the

**Table 3**  
Performance of the hybrid learning-based classification model for different numbers of selected sensors evaluated using a 5-fold cross-validation and repeated 10 times. Similar performances in terms of accuracy, sensitivity, and specificity can be achieved by the models with 5 and 10 sensors, demonstrating the possibility of reducing the used sensor number in GeNose C19.

Number of sensors	Selected sensors	Accuracy (%)	Sensitivity (%)	Specificity (%)
2	S4, S9	$78 \pm 3$	$78 \pm 6$	$78 \pm 7$
3	S4, S9, S10	$83 \pm 3$	$86 \pm 4$	$80 \pm 6$
4	S4, S9, S10, S2	$85 \pm 3$	$89 \pm 5$	$82 \pm 6$
5	S4, S9, S10, S2, S8	$86 \pm 3$	$88 \pm 6$	$84 \pm 6$
6	S4, S9, S10, S2, S8, S3	$85 \pm 3$	$87 \pm 6$	$83 \pm 6$
10	S1–S10	$86 \pm 3$	$87 \pm 6$	$84 \pm 6$



**Fig. 3.** Processing results of the positive (P) and negative COVID-19 data using hybrid learning method. **a** Dendrogram of the hierarchical agglomerative clustering (HAC) on the training data employing Ward's linkage. **b** Boxplot analysis of the permutation feature importance using the extra-tree classifier to obtain the importance value of each sensor for classifying the class labels of positive (P) and negative (N) COVID-19. Confusion matrix results and receiver operating characteristic (ROC) curves demonstrating learning model performances from the testing data when **c** all (10) and **d** 5-selected sensor models are utilized.

system. According to several reports associated with the studies of MS techniques for analyzing breath-borne COVID-19 biomarkers [21,24–26,64,72–76], the discriminant and important compounds for identifying COVID-19 possessed a large variation. Therefore, although the used techniques were quite similar to one another, several factors (e. g., place, setup, and types of breath samples) played important roles to affect the measurement results in these different research works.

One of the most frequently observed COVID-19 symptoms is anosmia, which is basically an indicator of neurodegenerative diseases (i.e., the olfactory system cannot accurately detect or correctly identify odors and is indicated by a loss of smell) [72,77]. Carbon monoxide has been linked with this issue because it is the diffusible intracellular and intercellular biomarker of cyclic nucleotide-gated channel activities in olfactory receptor neurons [78]. In other words, carbon monoxide is an olfactory transduction byproduct related to the reduction of cyclic nucleotide-gated channel activity, in which a loss of olfactory receptor neurons arises for minutes [72]. In our GC–MS results, carbon monoxide was detected by six sensors in GeNose C19 (i.e., S1, S3, S4, S5, S6, and S8). Aside from carbon monoxide, the multivariate analysis of data obtained from parallel COVID-19 breath studies using GC–ion mobility spectrometry (GC–IMS) in Dortmund, Germany and Edinburgh, United Kingdom had indicated that ketones (acetone and butanone), aldehydes (ethanol and octanal), and methanol could discriminate COVID-19 from

other conditions [21]. In Garches, France, by utilizing the proton-transfer reaction quadrupole time-of-flight MS, researchers had discovered four types of VOCs (i.e., methylpent-2-enal, 2,4-octadiene, 1-chloroheptane, and nonanal) that discriminated between COVID-19 and non-COVID-19 acute respiratory distress syndrome [24]. The results are different with those reported from a study conducted in two cities in the USA (Janesville, Wisconsin and Detroit, Michigan), despite the similar characterization technique based on the PTR time-of-flight MS [25]. In that study, Liangou et al. found another set of eight compounds (i.e., nitrogen oxide, butene, methanethiol, acetaldehyde, heptanal, ethanol, methanol water cluster, and propionic acid) as important biomarkers for the identification of COVID-19 in human breath. Meanwhile, a study conducted in Leicester, United Kingdom, employing the thermal desorption coupled GC–MS identified seven exhaled breath features (i. e., benzaldehyde, 1-propanol, 3,6-methylundecane, camphene, beta-cubebene, iodobenzene, and an unidentified compound) that could be used for separating PCR-positive COVID-19 patients from healthy ones [26]. In our case, camphene was also detected only in the negative COVID-19 breath sample by S10.

Moreover, in Beijing, China, Chen et al. reported two sequential research studies that yielded totally different breath-borne biomarkers despite using the same measurement approach (GC–IMS) [64,73]. Their first experiment reported in 2020 indicated that the differentiation

between COVID-19 and non-COVID-19 patients could be conducted by solely monitoring three compounds (i.e., ethyl butanoate, butyraldehyde, and isopropanol) [64]. However, based on their second report in 2021, among many VOC species, acetone was the biomarker because its levels were substantially lower for COVID-19 patients than those of other conditions [73]. In our GeNose C19 sensor array, S8 can detect acetone. Furthermore, in children with SARS-CoV-2 infection in Philadelphia, Pennsylvania, USA, six compound biomarkers (i.e., three aldehydes (octanal, nonanal, and heptanal), decane, tridecane, and 2-pentyl furan) were significantly distinguished in the breath analysis using two-dimensional GC and time-of-flight MS [74]. Another proposed biomarker for COVID-19 was ammonia, whose presence within the body has long been associated to complications stemming from the liver and kidneys affected by SARS-CoV-2 infection [75].

From all the already mentioned examples of MS studies, the identification of specific COVID-19 biomarkers in breath is clearly still challenging and can result in different discriminant compounds depending on several parameters (e.g., measurement technique, filtering approach, location, and breath sample type). Nonetheless, we still performed a GC-MS measurement to gain more insights on the compounds contained in the positive and negative COVID-19 breath samples (Table S1). Here, several hydrocarbons (e.g., ethylene, isoborneol thiocyanatoacetate, and farnesyl acetate) were dominantly sensed in the positive COVID-19 breath samples by S10. Meanwhile, for the negative COVID-19 breath samples, other hydrocarbons (i.e., camphene and octamethylcyclotetrasiloxane) were detected by S10. Moreover, specific esters (i.e., oxalic acid, bis(isobutyl) ester and acetic acid, dimethoxy-, methyl ester) were measured by S2 and S9 in the negative COVID-19 samples. In general, the appearances of the three sensors (S2, S9, and S10) were dominant as compared to those of the others. For instance, S10 was more sensitive toward hydrocarbons, whereas S2 and S9 were more likely to be reactive toward esters and aldehyde. In the case of two alcohols (i.e., ethanol and 1-nonadecanol; TMS derivatives that are detected in the positive and negative samples), all 10 sensors (S1–S10) sensed them. However, regardless of the successful extraction of the compounds, our GC-MS characterization was only performed in a low number of samples (i.e., two positive and two negative COVID-19 patients). Thus, a further investigation with a larger number of breath samples still needs to be carried out in the near future to correlate the measurement results of GeNose C19 and GC-MS methods. However, again, this study has focused more on the development of the hybrid learning method to improve the performance of the portable e-nose (GeNose C19) that can analyze the sensing signal patterns resulting from complex reactions between different MOS sensors and linked VOCs, without the need to analyze single VOC biomarkers in detail.

Based on the clustering results of the sample data, the dendrogram shows that the HAC method can calculate the distance and properly determine the cluster of all sensors. After performing the analysis, the yielded cluster correlates with the characteristics of the used sensors (Table 1). The resulting distance and cluster have a relationship with the sensor similarity. In Fig. 3a, using the HAC technique, S1 and S4 are combined into one cluster where they own a close distance because they have similarities in determining the breath samples from the patients with RT-qPCR-confirmed positive and negative COVID-19. Regarding their selective target gases (i.e., methane, carbon monoxide, isobutane, ethanol, and hydrogen), the two sensors possess a high similarity, despite their slightly different package designs (i.e., S1 and S4 used a 6-pinhole filter and mesh filter in their sensor packages, respectively).

The same phenomenon was also found in the clustering between S5 and S6, where the HAC algorithm makes them in one cluster, even though their distance is not as close as the one between S1 and S4. This is because the two sensors (S5 and S6) possess slightly different patterns in determining the positive and negative conditions of the patient. In the physical inspection, although they can characteristically detect the same target gases (i.e., carbon monoxide, methane, ethanol, propane, isobutane, and hydrogen), they are physically different in terms of sensor

package shapes and sizes. S6 has a dome-like mesh filter with a large cross-sectional area, whereas S5 has a circular-shaped filter with a small area. These different package configurations can consequently lead to different responses; i.e., S6 demonstrates a higher output signal than S5 (Fig. 2a).

The calculated results of the distance and clustering on S2, S8, and S9 demonstrate that the HAC method can estimate a high distance value at the sensors due to the uniqueness in identifying positive and negative models from the database. S9 was specifically designed to be more sensitive toward ammonia than other gases. Therefore, it was not correlated with other sensors. Nonetheless, its dendrogram was indeed closer to that of S2, which could also detect ammonia. Hence, the clustering process using this method can correlate among the breath sample data.

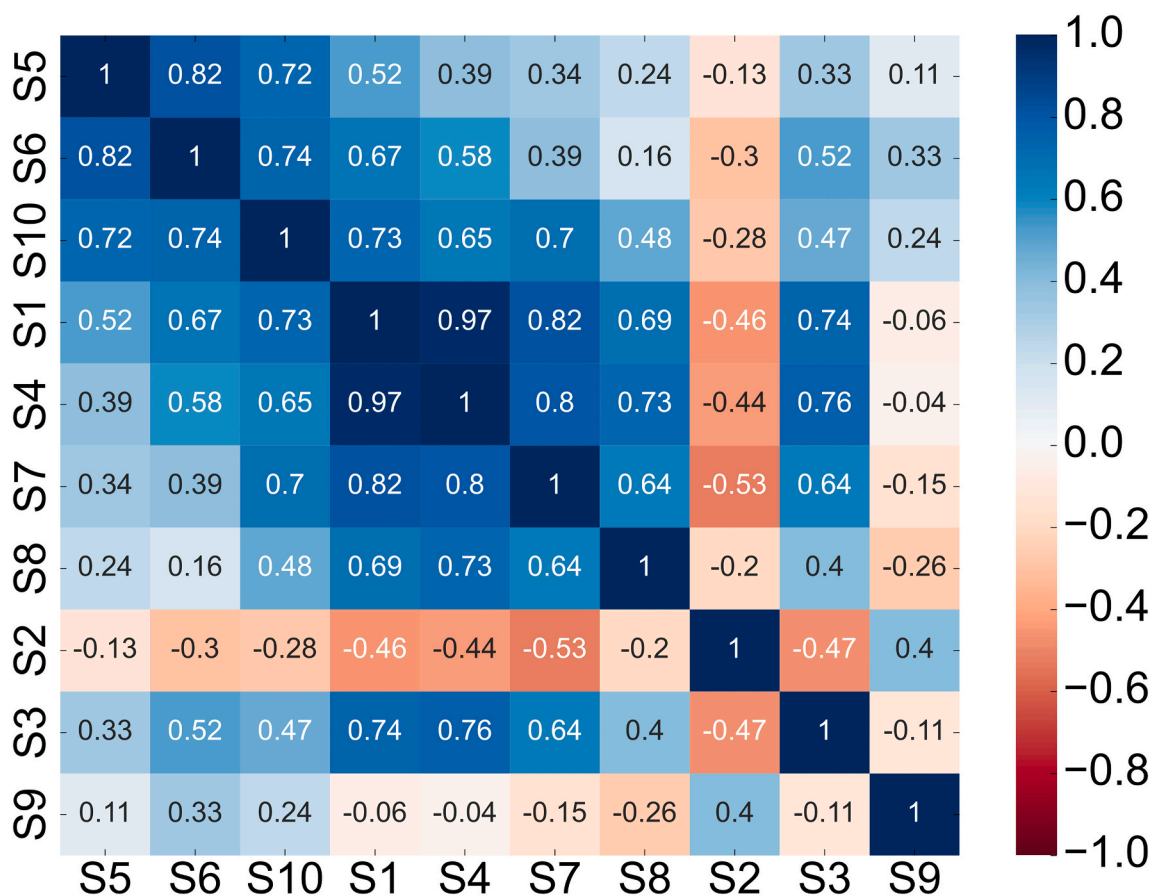
Based on the results of the feature importance permutation method, S4 has the highest feature value as compared to the other sensors. Although S1 and S4 aim at detecting the same target gases (Table 1), they possess different sensitivities and usage applications. While S1 can detect gas with a concentration range of 1–100 ppm, S4 can measure the gas concentration ranging from 60 to 1500 ppm. Moreover, S1 has approximately 10 times higher sensor resistance (i.e., 10–90 k $\Omega$  in air) than S4 (i.e., 1–5 k $\Omega$  in 300 ppm ethanol gas). Therefore, S1 is generally used to determine the environmental air quality, whereas S4 is applied for breath detection (i.e., alcohol-based VOC detection). Normalization has been applied in the analysis of this study. Accordingly, S1 can be balanced with S4 (multicollinear). Fig. 2e and Fig. 4 show that S1 and S4 have similar class label distributions and a correlation value of 0.97. Despite their close distances and correlations on the dendrogram (Fig. 3a) because of the normalization process, they can still be differentiated using the proposed feature selection method. Thus, based on the breath test data, S4 demonstrates a fairly higher importance value than S1, as depicted in Fig. 3b.

Furthermore, because of their similarity on the selective target gases and active materials, S5 and S6 have a high correlation value of 0.82. When they are associated with the closest, more important feature (i.e., S10), the proximity values of S10–S5 and S10–S6 can be searched using the *K*-means algorithm, in which they are measured by the metrics of silhouette and inertia scores. If the silhouette score is close to 1 and the inertia score exhibits a high value, then the data have been well clustered. In other words, the clusters are well apart from one another and clearly distinguished. In the result analysis, the silhouette scores of S10–S5 and S10–S6 are 0.25 and 0.29, respectively. Meanwhile, their inertia scores are 843,265.18 and 1,032,526.17, respectively. These results indicate that S10 is closer to S5 than S6. Based on the proximity analysis, S5 and S6 can be chosen. Therefore, we used a partial dependence analysis method to identify the performances of S5 and S6 to differentiate P and N class labels.

The partial dependence analysis can also be applied to find the feature characteristics. This method attempts to connect or find the relationship between features with the class label given to the selected classification model. In this case, the correlation factor of each feature is ignored. Fig. S2 shows the partial dependence analysis results of all features (S1–S10) on the P and N class labels using the training data and extra-tree classifier model. The results demonstrate that S5 possesses a higher importance factor than S6 because it has an average relationship. Although there are some differences between the P and N patterns in the S5 data distribution, the S6 counterpart cannot obviously distinguish the P and N patterns (i.e., indicated by the flat horizontal line at a value of  $\sim$ 0.5). The class labels of P and N data are represented by the values of 0 and 1, respectively. S5 possesses an average distribution of  $<$ 90 (AUC) indicating a class label of N. In other words, S5 can better identify model differences between P and N in the distribution of data than S6. Therefore, S5 has a higher importance value of the feature as compared to S6.

Table 3 lists the performances of six different combinations of the selected sensors using a 5-fold cross-validation repeated 10 times. As a





**Fig. 4.** Correlation plot among features on the training data using Ward's linkage method. S1 and S4 have a correlation value of 0.97, indicating their strong positive correlation. The high response of S4 results in a high S1 output signal. In addition, S7 possesses a high positive correlation with S1 and S4. S7 is also negatively correlated with S2, where an increase in the response of one sensor is followed by a decrease in the response of the other sensor. S5 and S6 are positively correlated with a value of 0.82, and both are positively correlated with S10 having correlation values of 0.72 and 0.74, respectively. All in all, some sensors possess a high multicollinearity with one another, which leads to a feasibility to optimize feature selection.

standard or basic model, all the 10 sensors (S1–S10) yielded an average accuracy, a sensitivity, and a specificity of  $(86 \pm 3)\%$ ,  $(87 \pm 6)\%$ , and  $(84 \pm 6)\%$ , respectively. In terms of the reduced number of sensors, the five selected sensors showed the most optimum results with the accuracy, sensitivity, and specificity of  $(86 \pm 3)\%$ ,  $(88 \pm 6)\%$ , and  $(84 \pm 6)\%$ , respectively. Here, the metric values of the reduced sensor number model are clearly similar to those of the basic model. This result proves the ability of the hybrid learning method (i.e., HAC combined with permutation feature importance) to not only keep the system performance stable but also reduce the redundant sensors. The results were also confirmed by the confusion matrix and receiver operating characteristic (ROC) curves shown in Figs. 3c and d. Again, here, the 5-selected-sensor model produces a slightly better performance than that of all ten sensors. The basic model produced ROC-AUC, accuracy, sensitivity, specificity, PPV and NPV of 0.95, 86%, 89%, 83%, 84%, and 88%, respectively. Meanwhile, the 5-selected-sensor model resulted in ROC-AUC, accuracy, sensitivity, specificity, PPV and NPV of 0.95, 87%, 87%, 87%, 87%, and 87%, respectively. Compared to the other models (i.e., 2, 3, 4, and 6-selected-sensor models) shown in Fig. S3, the 5-selected-sensor model exhibits the most stable and optimum performances.

All in all, the results may improve the efficiency as only sensors with a significant role will be recommended by our system. Eliminating sensors that play a less important role in target gas detection will reduce redundancy from data models. This can subsequently maintain accuracy when the data quantity increases. Furthermore, our hybrid algorithm has been able to suggest the optimum number and type of the sensors,

which have an interconnection between the digital feature from the database and the physical feature of the sensors.

After demonstrating the hybrid learning method to reduce the required number of sensors without downgrading the system quality, we evaluated the GeNose C19 in terms of its performances in comparison to other commercial and developed e-nose devices (e.g., PEN3 e-nose, aeoNose, SpiroNose, and nanomaterial-based e-nose [42–45]), which have been routinely tested for COVID-19 detection (Table 4). Here, several key parameters are compared (i.e., sensor type, sensor number, breath sample number, positive rate of samples, measurement time, and results during assessment in exhaled breath). As expected, because of their superior characteristics among other sensor techniques and their suitability for developing low-cost portable systems, the MOS sensors have been employed in all the commercial e-nose devices despite their different numbers. The aeoNose possesses the lowest sensor number (i.e., 3 sensors [43,79]). In case of the number and positive rate of tested breath samples, our system is superior compared to aeoNose and nanomaterial-based sensors. Only SpiroNose has been tested with more than 4500 samples [44]. Moreover, the MOS sensors used in GeNose C19 deliver the fastest sensing response (45 s) in comparison to those in other commercial e-nose devices.

It is well known that nanotechnology can support the advancement of sensor performance. Thus, the chemoresistive nanosensor array based on stabilized spherical gold nanoparticles with eight different organic functionalities (i.e., dodecanethiol, 2-ethylhexanethiol, 4-*tert*-methylbenzenethiol, decanethiol, 4-chlorobenzenemethanethiol, 3-ethoxythiophenol, *tert*-dodecanethiol, and hexanethiol) could react to the gas

**Table 4**

Comparison of different e-nose technologies used for COVID-19 detection in exhaled breath. The compared parameters include sensor type, sensor number, breath sample number, positive rate of samples, measurement time, and results during assessment in exhaled breath test.

Electronic nose (e-nose) technology	Number of sensors	Number of samples	Positive rate of samples	Measurement time	Results	Ref.
MOS sensor (PEN3 e-nose)	10	503	5.4%	80 s	66.7% of true positive rate	[42]
MOS sensor (aeoNose)	3	219	26.0%	300 s	86% of sensitivity and 92% of negative predictive value	[43]
MOS sensor (SpiroNose)	7	4510	7.7%	Not available	93.1% of ROC-AUC	[44]
Multiplexed nanomaterial-based chemoresistive sensor	8	130	37.7%	3 s	100% of sensitivity and 61% of specificity	[45]
MOS sensor (GeNose C19)	10 reduced to 5	460	50.0%	45 s	(88 ± 6)% of cross-validation sensitivity and (84 ± 6)% of cross-validation specificity	This work

molecules only within 3 s becoming the fastest sensors in this comparison. Nonetheless, the robustness and repeatability of such organic materials are still questionable. Therefore, despite their good sensitivity and selectivity, the organic ligands used for gas sensors can quickly degrade resulting in low long-term device stability [80–82]. For the results toward COVID-19 assessments, the obtained sensitivity value of GeNose C19 is comparable with that of aeoNose. Meanwhile, a direct result comparison to the other three devices (PEN3 e-nose, SpiroNose, and nanomaterial-based e-nose) is difficult to be created because different analyses were applied in those reported studies. Based on all these evaluations, regardless of the promising device performances in hospital setting, larger and more complex studies are still required to test the long-term reliability and stability of GeNose C19 for rapid identification of COVID-19 in public health setting.

### 3. Conclusions

A portable e-nose (GeNose C19) integrated with a hybrid machine learning method has been developed to be able to differentiate the RT-qPCR-confirmed positive COVID-19 samples from their negative counterparts (i.e., healthy controls). The feature-selection optimization for the gas sensor in the GeNose C19 has been successfully carried out using a combination of hierarchical agglomerative clustering (HAC) and permutation feature importance approaches. HAC has brought fast analysis to calculate multicollinear features, while permutation feature importance can show the importance value and provide scoring of each feature. Combining these two filter and wrapper methods, optimization has succeeded in reducing the number of sensors in the system by 50% without sacrificing the performance of the COVID-19 classification model. The 5-selected-sensor model has exhibited the accuracy, sensitivity, and specificity of (86 ± 3)%, (88 ± 6)%, and (84 ± 6)%, respectively, which are similar to the metrics from the model using all the 10 sensors. Aside from the sensor system miniaturization, further improvements in GeNose C19 can be expected when the adaptive learning method is associated with the currently developed hybrid feature selection approach.

### 4. Methods

#### 4.1. Experimental setup

The machine-learning-based assessment of GeNose C19 test data was performed using 460 breath samples of participants, in which the numbers of samples confirmed as positive (P) and negative (N) COVID-19 were  $n_P = 230$  and  $n_N = 230$ , respectively. COVID-19 infection was confirmed by the RT-qPCR tests on the SARS-CoV-2 ribonucleic acid obtained from the oropharyngeal and nasopharyngeal swabs. Data were collected from February to March 2021 at a public hospital located in Sleman, Daerah Istimewa Yogyakarta, Indonesia. Each individual that participated in this study was requested to store his or her exhaled breath into a single-use sampling bag made of medical-grade PVC, as

shown in Fig. S1.

For obtaining reliable data and ensuring human safety, a breath sampling procedure for GeNose C19 was set and applied during the breath exposure measurements. First, the patients were requested to store their end-tidal breath into a 1 L sampling bag. The taken breath resulted from the third exhalation. Meanwhile, the first two breaths were not sampled to minimize possible contamination sources commonly found in a mixed expiratory breath (e.g., dead space air and mouth-released odor) [83]. In another study, different VOC ratios were yielded from various sampling methods (i.e., mixed expiratory and end-tidal breath sampling approaches) [84,85]. Nonetheless, when blood-borne volatile substances will be employed as disease biomarkers, the end-tidal breath sampling technique is suitable for usage [83]. The ratios describing the differences of expiratory and inspiratory concentrations and the alveolar concentration could be calculated to approximate the alteration of inhaled and exhaled substances. Here, blood-borne volatile substances with clearly exogenous (e.g., 2-butanone and 2-propanol) and endogenous (e.g., acetone, isoprene, and CO<sub>2</sub>) origins typically obtained low (<1) and high (>1.5) ratio values, respectively [23,84]. After inserting the third exhaled breath into the sampling bag, its valve was quickly closed to avoid any air leakages. Lastly, the sealed sampling bag was then connected to the GeNose C19 to perform the breath measurements.

The utilized e-nose consisted of 10 different MOS gas sensors as listed in Table 1 and was equipped with a temperature and humidity sensor inside the chamber. The signals produced by GeNose C19 during the breath assessment were recorded using a data logging software (DAQ software), which also acted as a gas sampling configurator. During the sensor output data acquisition, the times for the baseline setting, breath sensing, and air purging were set to 2 s, 40 s, and 3 s, respectively. Here, each breath sample has a data dimension of 450 × 12 (i.e., 450 time-series data for 12 sensing outputs) because the software recorded the data with an interval of 100 ms.

Despite the different places used during breath collections that depend on the patient locations, the measurements of all the breath samples using GeNose C19 were performed in the same room to ensure high data reliability for the analysis. Here, the collected sampling bags consisting of exhaled breaths distributed from different sampling stations were stored inside a portable container and brought to the same room where an e-nose was located for characterization. Apart from the HEPA filter employed to trap the virus-containing droplets in the front-side breath sampling system, a vacuum filter with a grade filtration of 50 μm was integrated into a GeNose C19 machine to filter the dust microparticles out from ambient (reference) air. However, no specific environmental VOC filter was involved in this case. Again, before analyzing the breath samples, the baseline values and responses of the gas sensors to the reference air were continually measured and monitored to validate the preconditioned system.

During the breath sampling process (February–March 2021), the total number of persons tested for COVID-19 investigation in laboratories located in Daerah Istimewa Yogyakarta Province was

approximately 47,705 according to the statistics from the local government of Yogyakarta [86]. To obtain reliable results and analysis, the number of samples used in the proposed study should be adequate to represent the existing population. Instead of choosing the whole population, the infinite population formula was implemented as a population parameter to prevent high cost, time, and complexity [87]. Here, a confidence level of 95% (Z-value of 1.96), confidence interval of  $\pm 5\%$ , and  $p$ - and  $q$ -values of 50% were set as key parameters for calculating the sample size [88]. As a result, the sample size of 382 was required to estimate the percentage of COVID-19-infected persons. We increased the number of sample size by  $\sim 17\%$  from 382 to 460 to ensure a high precision level in the data analysis. In addition to this technique, a learning curve fitting-integrated post-hoc approach can be employed to simulate the relationship between the training sample size and the mean of the proposed classification model accuracy [89,90]. In Fig. S4, the minimum sample size required to discriminate positive and negative COVID-19 samples with optimum performance is 250. Thus, using a sample size of 460, our study has already met this criterion.

#### 4.2. Feature selection algorithms

Hierarchical agglomerative clustering (HAC) is a method for grouping two clusters by calculating the distance between them. In case the clusters have a high proximity (data with a close distance), they will be combined using an agglomerative algorithm. In this algorithm, each datum is identified as a separate cluster (cluster singleton). During the iteration, the Euclidean distance from each cluster is calculated, and the most similar clusters are sequentially combined. This procedure is repeated until an optimum final cluster is formed. The Euclidean distance  $d(p, q)$ , which measures the shortest straight distance between two points ( $p$  and  $q$ ), is expressed as

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

HAC has different methods for combining clusters, one of which is Ward's linkage. This approach attempts to analyze the variance between clusters rather than directly calculating the distance. The Ward's linkage determines the distance between two clusters of A and B using Eq. (2):

$$\begin{aligned} \Delta(A, B) &= \sum_{i \in A \cup B} \left\| \vec{x}_i - \vec{m}_{A \cup B} \right\|^2 - \sum_{i \in A} \left\| \vec{x}_i - \vec{m}_A \right\|^2 - \sum_{i \in B} \left\| \vec{x}_i - \vec{m}_B \right\|^2 \\ &= \frac{n_A n_B}{n_A + n_B} \left\| \vec{m}_A - \vec{m}_B \right\|^2 \end{aligned} \quad (2)$$

where  $\vec{m}_j$  is the center of cluster  $j$  and  $n_j$  is the number of points in it. Meanwhile,  $\Delta(A, B)$  is the merging cost of combining clusters A and B [91,92].

Permutation feature importance measures the increase in the prediction error of a model after the feature value is permuted [93–95]. This procedure causes the connection between the feature and output to be broken. Features that have a large importance value will produce an increase in a prediction error when they are permuted. For a feature that possesses a low importance score, the predictive value remains stable even though the feature value is randomized because the selected classification model ignores the feature. However, this method assumes that each feature is not correlated with each other. Therefore, the obtained results must be validated first. With a trained model  $\hat{f}$ , feature matrix  $X$ , target vector  $y$ , and error measure  $L(y, \hat{f})$ , feature permutation importance can be performed by employing the following steps:

1. Estimate the original model error  $e_{original} = L(y, \hat{f})$
2. For each feature  $j \in \{1, \dots, p\}$ , do:
  - a. Generate feature matrix  $X_{perm}$  by permuting feature  $j$  in the data  $X$ .
  - b. Estimate error  $e_{perm} = L(y, \hat{f}(X_{perm}))$  based on the predictions of the permuted data.

- c. Calculate permutation feature importance as the quotient  $FI_j = e_{perm}/e_{original}$  or difference  $FI_j = e_{perm} - e_{original}$ .
3. Sort features by descending  $FI$ .

In this study, HAC, which is categorized to the filter-based learning approach, was combined with the permutation feature importance method as the feature-selection method for enhancing the GeNose C19 performance and for reducing the required sensors in the system. The permutation feature importance was used to validate the HAC output results. Using this hybrid method, important features that are not correlated with one another can be found. HAC can quickly measure the correlation value between features. Hence, features that possess a high correlation (multicollinearity) can be obtained. Subsequently, the permutation feature importance method was utilized to validate and select one of the best correlated features. The dendrogram plot was used to visualize the correlation results between the features. Moreover, the boxplot can visualize the importance level of each feature.

The extra-tree classifier was employed to differentiate the positive and negative classes based on the decision tree algorithm, which is similar to the random forest classifier [96–98]. We utilized the normalized AUC during the sensor measurement as the input model. The AUC data were chosen as the feature extractions representing the amount of exhaled VOCs of the participant that interact with the sensing active layer during a 40 s-long sensing phase. In addition, Simpson's rule was used to numerically calculate the AUC of each signal. The signal was then further normalized to eliminate the environmental influence. The normalized AUC value for each sensor ( $x_i$ ) is determined as

$$x_i = \text{AUC} \left( \frac{y_i(t) - y_i(t=0)}{S} \right) \quad (3)$$

where  $y_i(t)$ ,  $y_i(t=0)$ , and  $S$  are the sensor signal (volt), initial sensor signal (volt), and maximum value of the sensor response, respectively. Here, index  $i$  represents the sensor number ( $i = 1$  to 10).

#### 4.3. Analysis of VOCs

The VOCs contained in the RT-qPCR-confirmed negative and positive COVID-19 breath samples were analyzed using a GC–mass spectroscopy (GC–MS) equipment (ISQ 7000 single quadrupole GC–MS system, Thermo Fisher Scientific Inc., Massachusetts, USA).

#### 4.4. Research ethics approval

This study has been 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 has been registered in [clinicaltrials.gov](https://clinicaltrials.gov) (NCT04558372) [99]. All procedures were carried out following the relevant guidelines and regulations based on the Good Clinical Practice and the Helsinki Declaration of 2013 [100].

#### CRediT authorship contribution statement

S.N.H., H.S.W., and K.T. conceived the idea and concept, designed the study, formulated the materials, validated the methods, and interpreted the experimental results. H.S.W. and S.N.H. wrote the initial manuscript. S.N.H. performed the machine learning computation and created the figs. T.J. designed and developed the hardware and electronics for experiments. H.S.W., A.B.D., A.R., and K.T. revised the paper. M.P. performed the characterization and analysis of gas chromatography–mass spectrometry (GC–MS). L.C. and D.K.N. carried out the COVID-19 test experiment using GeNose C19 at the hospitals. A.R. and M.J. provided the inputs on the experiment design. H.S.W. supervised the work, led the project, and acquired the funding at PT Nanosense

Instrument Indonesia. D.K.N. and K.T. supervised the work, led the Universitas Gadjah Mada (UGM) project, and acquired the funding. All authors approved the final manuscript.

### Declaration of competing interest

The authors declare that there are no conflicts of interest.

### Acknowledgment

This work was funded in part by the National Research and Innovation Agency (BRIN) and Indonesia Endowment Fund for Education (LPDP) within the project of “System Optimization of Machine Learning for Early Detection of COVID-19 Using Electronic Nose” under grant agreement numbers of 10/FI/P-KCOVID-19.2B3/X/2020 and 10/FI/P-KCOVID-19.2B3/X/2021 and in part by Universitas Gadjah Mada (UGM) with project numbers of 5230/UN1/DUI/DIT-PUI/HK/2020 and 7262/UN1/DUI/DIT-PUI/HK/2021. The authors are grateful for the administrative and technical support from Muhammad Iqbal Nuriyana, Yuliyani Dwi Prabowo, Iman Rahman, Linda Ardita Putri, Rakha Saputra, Moch Azis Rifa'i, and other employees at PT Nanosense Instrument Indonesia during the development and assessment of GeNose C19. Valuable discussions from Dr. Wahyono (Department of Computer Science and Electronic, UGM) and support from the Science Techno Park (STP) at UGM are also acknowledged. The authors thank all the involved patients, families, hospital directors, research assistants, and healthcare workers (medical doctors, nurses, and interns) for their support during the test of GeNose C19 at the hospital.

### Appendix A. Supplementary data

Components of portable GeNose C19; Partial dependence analysis; Confusion matrix and receiver operating characteristic (ROC) curve; Learning curve fitting-integrated post-hoc analysis; Gas chromatography–mass spectroscopy (GC–MS) analysis. Supplementary data to this article can be found online at doi: <https://doi.org/10.1016/j.artmed.2022.102323>

### References

- [1] Das S, Pal M. Review–Non-invasive monitoring of human health by exhaled breath analysis: a comprehensive review. *J Electrochem Soc* 2020;167:037562. <https://doi.org/10.1149/1945-7111/ab67a6>.
- [2] Yoon J-W, Lee J-H. Toward breath analysis on a chip for disease diagnosis using semiconductor-based chemiresistors: recent progress and future perspectives. *Lab Chip* 2017;17:3537–57. <https://doi.org/10.1039/C7LC00810D>.
- [3] Amann A, Miekisch W, Schubert J, Buszewski B, Ligor T, Jezierski T, et al. Analysis of exhaled breath for disease detection. *Annu Rev Anal Chem* 2014;7:455–82. <https://doi.org/10.1146/annurev-anchem-071213-020043>.
- [4] Lourenço C, Turner C. Breath analysis in disease diagnosis: methodological considerations and applications. *Metabolites* 2014;4:465–98. <https://doi.org/10.3390/metabo4020465>.
- [5] Rodríguez-Aguilar M, Díaz de León-Martínez L, Zamora-Mendoza BN, Comas-García A, Guerra Palomares SE, García-Sepúlveda CA, et al. Comparative analysis of chemical breath-prints through olfactory technology for the discrimination between SARS-CoV-2 infected patients and controls. *Clin Chim Acta* 2021;519:126–32. <https://doi.org/10.1016/j.cca.2021.04.015>.
- [6] Bobak CA, Kang L, Workman L, Bateman L, Khan MS, Prins M, et al. Breath can discriminate tuberculosis from other lower respiratory illness in children. *Sci Rep* 2021;11:2704. <https://doi.org/10.1038/s41598-021-80970-w>.
- [7] van Oort PM, Povoja P, Schnabel R, Dark P, Artigas A, Bergmans DCJJ, et al. The potential role of exhaled breath analysis in the diagnostic process of pneumonia—a systematic review. *J Breath Res* 2018;12:024001. <https://doi.org/10.1088/1752-7163/aaa499>.
- [8] Schleich FN, Zanella D, Stefanuto P-H, Bessonov K, Smolinska A, Dallinga JW, et al. Exhaled volatile organic compounds are able to discriminate between neutrophilic and eosinophilic asthma. *Am J Respir Crit Care Med* 2019;200:444–53. <https://doi.org/10.1164/rccm.201811-2210OC>.
- [9] Chan LW, Anahtar MN, Ong T-H, Hern KE, Kunz RR, Bhatia SN. Engineering synthetic breath biomarkers for respiratory disease. *Nat Nanotechnol* 2020;15:792–800. <https://doi.org/10.1038/s41598-020-0723-4>.
- [10] Sukul P, Schubert JK, Zanaty K, Trefz P, Sinha A, Kamyssek S, et al. Exhaled breath compositions under varying respiratory rhythms reflects ventilatory variations:

- translating breathomics towards respiratory medicine. *Sci Rep* 2020;10:14109. <https://doi.org/10.1038/s41598-020-70993-0>.
- [11] Mertz L. The great exhale: using breath analysis to detect disease. *IEEE Pulse* 2020;11:7–11. <https://doi.org/10.1109/MPULS.2020.2993684>.
  - [12] Janssens E, van Meerbeek JP, Lamote K. Volatile organic compounds in human matrices as lung cancer biomarkers: a systematic review. *Crit Rev Oncol Hematol* 2020;153:103037. <https://doi.org/10.1016/j.critrevonc.2020.103037>.
  - [13] Gupta P, Wen H, Di Francesco L, Ayazi F. Detection of pathological mechano-acoustic signatures using precision accelerometer contact microphones in patients with pulmonary disorders. *Sci Rep* 2021;11:13427. <https://doi.org/10.1038/s41598-021-92666-2>.
  - [14] Khoubnasabjafari M, Mogaddam MRA, Rahimpour E, Soleymani J, Saei AA, Jouyban A. Breathomics: review of sample collection and analysis, data modeling and clinical applications. *Crit Rev Anal Chem* 2021:1–27. <https://doi.org/10.1080/10408347.2021.1889961>.
  - [15] Haworth JJ, Pitcher CK, Ferrandino G, Hobson AR, Pappan KL, Lawson JLD. Breathing new life into clinical testing and diagnostics: perspectives on volatile biomarkers from breath. *Crit Rev Clin Lab Sci* 2022:1–20. <https://doi.org/10.1080/10408363.2022.2038075>.
  - [16] Hu B, Guo H, Zhou P, Shi Z-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol* 2021;19:141–54. <https://doi.org/10.1038/s41579-020-00459-7>.
  - [17] Comito C, Pizzuti C. Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: a focused review. *Artif Intell Med* 2022;128:102286. <https://doi.org/10.1016/j.artmed.2022.102286>.
  - [18] Bhatia M, Manocha A, Ahanger TA, Alqahtani A. Artificial intelligence-inspired comprehensive framework for Covid-19 outbreak control. *Artif Intell Med* 2022;127:102288. <https://doi.org/10.1016/j.artmed.2022.102288>.
  - [19] Adjei G, Enuameh YA, Thomford NE. Prevalence of COVID-19 genomic variation in Africa: a living systematic review protocol. *JBIM Evid Synth* 2022;20:158–63. <https://doi.org/10.111124/JBIES-20-00516>.
  - [20] Yuan Z-C, Hu B. Mass spectrometry-based human breath analysis: towards COVID-19 diagnosis and research. *J Anal Test* 2021. <https://doi.org/10.1007/s41664-021-00194-9>.
  - [21] Ruszkiewicz DM, Sanders D, O'Brien R, Hempel F, Reed MJ, Riepe AC, et al. 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>.
  - [22] Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn* 2020;20:453–4. <https://doi.org/10.1080/14737159.2020.1757437>.
  - [23] Belluono I, Boshier PR, Myridakis A, Vadhvana B, Markar SR, Spanel P, et al. Selected ion flow tube mass spectrometry for targeted analysis of volatile organic compounds in human breath. *Nat Protoc* 2021;16:3419–38. <https://doi.org/10.1038/s41596-021-00542-0>.
  - [24] Grassin-Delye S, Roquencourt C, Moine P, Saffroy G, Carn S, Heming N, et al. 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>.
  - [25] Liangou A, Tasoglou A, Huber HJ, Wistrom C, Brody K, Menon PG, et al. A method for the identification of COVID-19 biomarkers in human breath using proton transfer reaction time-of-flight mass spectrometry. *EclinicalMedicine* 2021;42:101207. <https://doi.org/10.1016/j.eclinm.2021.101207>.
  - [26] Ibrahim W, Cordell RL, Wilde MJ, Richardson M, Carr L, Sundari Devi Dasi A, et al. Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry. *ERJ Open Res* 2021;7:00139–2021. <https://doi.org/10.1183/23120541.00139-2021>.
  - [27] Casals O, Markiewicz N, Fabrega C, Gràcia I, Cané C, Wasisto HS, et al. A parts per billion (ppb) sensor for NO<sub>2</sub> with microwatt ( $\mu$ W) power requirements based on micro light plates. *ACS Sens* 2019;4:822–6. <https://doi.org/10.1021/acssens.9b00150>.
  - [28] Markiewicz N, Casals O, Fabrega C, Gràcia I, Cané C, Wasisto HS, et al. Micro light plates for low-power photoactivated (gas) sensors. *Appl Phys Lett* 2019;114:053508. <https://doi.org/10.1063/1.5078497>.
  - [29] Wasisto HS, Prades JD, Güllink J, Waag A. Beyond solid-state lighting: miniaturization, hybrid integration, and applications of GaN nano- and micro-LEDs. *Appl Phys Rev* 2019;6:041315. <https://doi.org/10.1063/1.5096322>.
  - [30] Wasisto HS, Merzsch S, Stranz A, Waag A, Uhde E, Salthammer T, et al. Silicon resonant nanopillar sensors for airborne titanium dioxide engineered nanoparticle mass detection. *Sens Actuators B Chem* 2013;189:146–56. <https://doi.org/10.1016/j.snb.2013.02.053>.
  - [31] Wasisto HS, Merzsch S, Waag A, Uhde E, Salthammer T, Peiner E. Airborne engineered nanoparticle mass sensor based on a silicon resonant cantilever. *Sens Actuators B Chem* 2013;180:77–89. <https://doi.org/10.1016/j.snb.2012.04.003>.
  - [32] Wasisto HS, 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>.
  - [33] Bindra P, Hazra A. Capacitive gas and vapor sensors using nanomaterials. *J Mater Sci Mater Electron* 2018;29:6129–48. <https://doi.org/10.1007/s10854-018-8606-2>.
  - [34] Rianjanu A, Fauzi F, Triyana K, Wasisto HS. Electrospun nanofibers for quartz crystal microbalance gas sensors: a review. *ACS Appl Nano Mater* 2021;4:9957–75. <https://doi.org/10.1021/acsnm.1c01895>.
  - [35] Julian T, Hidayat SN, Rianjanu A, Dharmawan AB, Wasisto HS, Triyana K. Intelligent mobile electronic nose system comprising a hybrid polymer-functionalized quartz crystal microbalance sensor array. *ACS Omega* 2020;5:29492–503. <https://doi.org/10.1021/acsomega.0c04433>.



- [87] Rodríguez del Águila M, González-Ramírez A. Sample size calculation. *Allergol Immunopathol (Madr)* 2014;42:485–92. <https://doi.org/10.1016/j.aller.2013.03.008>.
- [88] MdF Jubayer, MdTI Limon, MdM Rana, MdS Kayshar, MdS Arifin, Uddin AM, et al. COVID-19 knowledge, attitude, and practices among the Rohingya refugees in Cox's Bazar, Bangladesh. *Public Health Pract* 2022;3:100227. <https://doi.org/10.1016/j.puhip.2022.100227>.
- [89] Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. *BMC Med Inform Decis Mak* 2012;12:8. <https://doi.org/10.1186/1472-6947-12-8>.
- [90] Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Can Assoc Radiol J* 2019;70:344–53. <https://doi.org/10.1016/j.carj.2019.06.002>.
- [91] Sasirekha K, Baby P. Agglomerative hierarchical clustering algorithm - a review. *Int J Sci Res Publ* 2013;3:3.
- [92] Tokuda EK, Comin CH, Costa Lda F. Revisiting agglomerative clustering. *Phys Stat Mech Its Appl* 2022;585:126433. <https://doi.org/10.1016/j.physa.2021.126433>.
- [93] Casalicchio G, Molnar C, Bischl B. Visualizing the feature importance for black box models. In: Berlingerio M, Bonchi F, Gärtner T, Hurley N, Ifrim G, editors. *Mach. Learn. Knowl. Discov. Databases*. 11051. Cham: Springer International Publishing; 2019. p. 655–70. [https://doi.org/10.1007/978-3-030-10925-7\\_40](https://doi.org/10.1007/978-3-030-10925-7_40).
- [94] Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res* 2019;20:1–81.
- [95] Molnar C. *Interpretable machine learning (second edition)*. leanpub.com; 2022.
- [96] Kiala Z, Mutanga O, Odindi J, Masemola C. Optimal window period for mapping parthenium weed in South Africa, using high temporal resolution imagery and the ExtraTrees classifier. *Biol Invasions* 2021;23:2881–92. <https://doi.org/10.1007/s10530-021-02544-1>.
- [97] Fang G, Annis IE, Elston-Lafata J, Cykert S. Applying machine learning to predict real-world individual treatment effects: insights from a virtual patient cohort. *J Am Med Inform Assoc* 2019;26:977–88. <https://doi.org/10.1093/jamia/ocz036>.
- [98] Czajkowski M, Grześ M, Kretowski M. Multi-test decision tree and its application to microarray data classification. *Artif Intell Med* 2014;61:35–44. <https://doi.org/10.1016/j.artmed.2014.01.005>.
- [99] Genosvid Diagnostic Test for Early Detection of COVID-19 - Full Text View - ClinicalTrials.gov. <https://clinicaltrials.gov/ct2/show/NCT04558372>. [Accessed 25 April 2021].
- [100] World Medical Association. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013; 310:2191. <https://doi.org/10.1001/jama.2013.281053>.