



OPEN ACCESS

EDITED BY

Rajib Biswas,
Tezpur University, India

REVIEWED BY

Antonio Sasso,
University of Naples Federico II, Italy
Ganesan Angarai,
Indian Institute of Technology Madras,
India

*CORRESPONDENCE

Rao Tatavarti,
✉ rtatavarti@gmail.com

SPECIALTY SECTION

This article was submitted
to Optics and Photonics,
a section of the journal
Frontiers in Physics

RECEIVED 10 December 2022

ACCEPTED 20 March 2023

PUBLISHED 13 April 2023

CITATION

Tatavarti R, Nadimpalli S, Mangina GVK,
Kiran Machiraju N, Pachiyappan A,
Hiremath S, Jagannathan V and
Viswanathan P (2023), Photonic system
for real-time detection, discrimination,
and quantification of microbes in air.
Front. Phys. 11:1118885.
doi: 10.3389/fphy.2023.1118885

COPYRIGHT

© 2023 Tatavarti, Nadimpalli, Mangina,
Kiran Machiraju, Pachiyappan, Hiremath,
Jagannathan and Viswanathan. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Photonic system for real-time detection, discrimination, and quantification of microbes in air

Rao Tatavarti^{1*}, Sridevi Nadimpalli²,
Gowtham Venkata Kumar Mangina³, Naga Kiran Machiraju³,
Arulmozhivarman Pachiyappan⁴, Shridhar Hiremath⁵,
Venkateshan Jagannathan⁵ and Pragasam Viswanathan⁵

¹GVP-SIRC & GVP College of Engineering, Visakhapatnam, Andhra Pradesh, India, ²CASTLE Advanced Technologies and Systems, Visakhapatnam, Andhra Pradesh, India, ³CATS Eco Systems Pvt Ltd, Nashik, Maharashtra, India, ⁴Centre for Clean Environment, and Office of Academic Research, VIT University, Vellore, Tamil Nadu, India, ⁵School of Biosciences and Technology, Pearl Research Park, VIT University, Vellore, Tamil Nadu, India

We report the results of the non-invasive photonic system AUM for remote detection and characterization of different pathogenic bacterial strains and mixtures. AUM applies the concepts of elastic light scattering, statistical mechanics, artificial intelligence, and machine learning to identify, classify and quantify various microbes in the scattering volume in real-time and, therefore, can become a potential tool in controlling and managing diseases caused by pathogenic microbes.

KEYWORDS

photonic system, real-time, detection, discrimination, quantification, microbes in air

1 Introduction

1.1 Impetus

Pathogenic bacteria, fungi, viruses, and parasites can travel through various routes (air, water, physical contact, and bodily fluids) and enter their hosts (plants, animals, and humans), causing infectious diseases [1–5]. Airborne microorganisms transmitted through human speech, coughing, sneezing, and exhalation are known to have long residence times in the environment [6–12] and are responsible for the rapid and extensive spread of diseases [13–15]. The increasing incidences of infectious diseases worldwide are a significant cause for concern, pointing out the need for better control and management of diseases. The interplay of pathogenic microbes, their hosts, and the environment, coupled with the complex dynamics of disease emergence, demand rapid, direct detection and identification of airborne microorganisms to prevent transmission and the outbreak of diseases [16]. As a result, various technologies based on nucleic acid, mass spectrometry, cell structure, and optics were reported and are now in use.

Traditional pathogen detection employs morphological and biochemical techniques involving culturing with an enrichment step to isolate the target organisms on an agar plate or to detect pathogens directly from the enriched samples using advanced, sophisticated technologies and molecular techniques [17–25]. In addition, diversified technology combinations in biosensors [18], especially optical biosensors [22], have shown improved capabilities in directly detecting airborne microorganisms. However, the current molecular technologies in use for the identification of components and signal

amplification of sensors [17–19], including advanced, sophisticated technologies and techniques (genomic fingerprinting [20, 21], real-time quantitative PCR [24], mass spectrometry [25, 26]), still require on-site sampling and further testing in the laboratory, involves complex operation processes, having long detection times [1–5]. Interestingly, even though the current techniques have high detection efficiency and accuracy, the sampling strategy's subjectivity and the sampling mechanisms employed are known to have limitations affecting the accuracy of microbe detection [2].

Although on-site air sample detection features a short detection time, flexibility, and convenience, the challenges posed by the effects of *in situ* environmental parameters, the low concentrations of airborne microbes, and the possibility of occurrence of a broad mix of species at a particular location, lead to practical difficulties. Hence, notwithstanding their limitations and long overturn times, conventional methods and technologies are still widely used for their high specificity and sensitivity in identifying and quantifying pathogens in the air. In addition, traditional culturing is still considered a gold standard for detecting microbes due to its ability to obtain an isolated pure form of the pathogen accurately [27], even though there are difficulties associated with the isolation and maintenance of a novel unidentified or un-culturable number of bacteria and viruses present in the real-world samples.

The growing pathogenic strain variations and antimicrobial resistances demand quick and broad prevention strategies and improved countermeasures to ensure more directed therapeutic interventions through point-of-care testing for the prevention and control of infections [28–30]. Optical technologies have generated significant interest due to their high speed, high throughput, non-destructive nature, and amplification-free measurements that require minimal sample preparation [31].

1.2 Optical technologies

The optical technologies for the study of microbes can further be categorized based on types of interaction that occur between microbes and light: UV-Visible Spectroscopy, Fluorescence (Bulk Fluorescence Spectroscopy, Fluorescence Imaging), Flow Cytometry, Vibrational Spectroscopy (Raman, Mid-Infrared, Near-Infrared), Scattering (Elastic Scattering, Dynamic Light Scattering, Dynamic Laser Speckle (Bio Speckle), Quantitative Phase Imaging, Differential Dynamic Microscopy, Video Microscopy), and Optical Coherence Tomography.

The principles of Elastic Light Scattering (ELS), which use the characteristics of the spatial distribution of the scattered light with the same wavelength of the illuminating light source, were earlier used to find bioaerosols among complex, diverse atmospheric aerosols with high sensitivity due to its increased signal strength, and the sizeable scattering cross-section, *via* single-particle or multiple-particle interrogation [32–36], without using specific labelling reagents [32].

All optical technologies either detect a signal for a defined wavelength using a photodiode or other broadband detector or resolve the output as a function of wavelength using a spectral sensor. The collective molecular signature from the sample (cell or a colony) makes up the spectrum, and the level of chemical specificity within a range depends on the spectroscopy type.

Imaging systems, especially time-lapse imaging systems based on CMOS imaging sensors—were used to perform early detection and classification of bacteria. However, limitations in scanning the samples due to the constraints dictated by the field of view of the imaging sensors posed difficulties in probing the volume of samples quickly and effectively. Recent research [37] overcame the limitations by using a Thin Film Transistor (TFT) based imaging sensor to build a real-time Colony Forming Unit (CFU) detection system to count bacterial colonies automatically and rapidly identify their species using deep learning.

Addressing the need for a portable and cost-effective device for long-term monitoring and quantification of various types of pollen which affect human health, Lou et al. [38] presented a label-free sensor that takes holographic images of flowing particulate matter (PM) concentrated by a virtual impactor, which selectively guides particles of a specific size (larger than $6\ \mu\text{m}$) to fly through an imaging window. The inline holograms cast on a CMOS image sensor due to the flowing particles illuminated by a pulsed laser diode enable particle detection based on deep learning techniques. However, observing single particles flowing in a single file through the light beam eliminates the complexity of multiple scattering and the challenging task of deconvoluting the signals from an array of particles in size, shape, and materials [33]. Although an exciting sampling strategy, the technique adopted is still not *in situ*, with the inability to detect smaller ($<6\ \mu\text{m}$) particles remaining a practical limitation, as the smaller particles (PM_{2.5} and lesser) are known to cause many adverse effects on human health.

Optics-based technologies comprise a variety of instrumental configurations by which the signal from bacteria is probed, with or without spatial information, generally from a relatively small sample volume. However, to our knowledge, all the technologies and systems using optics for bacterial identification and characterization involve *in-vitro* sampling methods, not *in situ* ones that are desirable. Moreover, the devices are neither portable nor economical for wider deployment for emerging real-world applications. Therefore, a rapid, robust, mobile, low-cost, and *in situ* device is necessary for bacterial detection, discrimination, and quantification for wider deployment for operational uses in the real world.

2 Materials and methods

2.1 AUM photonic system

Against this backdrop, Tatavarti [39] designed and developed a photonic system AUM (Air Unique-quality Monitor) (Figure 1) capable of detecting, discriminating, and quantifying various air molecules *in situ*. AUM photonic system comprises a continuous wave laser light source located at the bottom on the system's front side to illuminate the target volume of air and a position-sensing photodetector positioned at the top on the front side, to detect the backscattered laser light from the target volume. Optical filters in front of the laser source and the photodetector ensure that only the particular wavelength of light from the source passes through and falls back on the photodetector after scattering from aerosols/molecules in the target air volume. AUM system's backscatter data observations enabled the scatterers' (dust, particulate matter,



FIGURE 1

AUM photonic system [comprising of a continuous wave laser light source at the bottom on the front side, to illuminate the target volume of air, and, a position sensing photodetector at the top on the front side, to detect the back scattered laser light from the target volume. Optical filters in front of both laser source and the photodetector ensures that only the particular wavelength of light from the source passes through and falls back on the photodetector after scattering from microbes in target volume of air [Picture adopted from Tatavarti, 2021 [38]].

and different air molecules - which constitute the pollutants) detection, differentiation, and quantification. In the AUM photonic system, the backscattered light from a laser (635 nm, <5 mW, circular beam, TEM₀₀, unpolarized) source is designed to fall on the small measurement area on a duo-lateral position sensing photodetector. The position sensing detector (PSD) is a Lateral Effect Diode, having a sufficiently active sensing area with a wide dynamic range, whose measurements are independent of the light spot profile and intensity distribution. The centroid of the light beam is computed and supplied as electrical output signals proportional to the displacement from the centre of the detector. The PSD's resolution is detector/circuit signal-to-noise ratio-dependent having a resolution of better than 100 nm. Analog data from the photodetector can be digitized at any sampling frequency, up to 10 ???, and processed using proper digital signal processing circuitry and software. After appropriate signal conditioning, the photodetector's analogue positional and light intensity signals are digitized and stored against their observation timestamp. The data from AUM can be ported onto storage devices with wireless connectivity through a remote cloud server or wired connection to a nearby device. Therefore, AUM photonic system can yield the backscattered beam's position in the (x, y) plane of the detector and the intensity with a sampling frequency as high as 10 kHz.

2.2 AUM: Remote, real-time, *in situ* measurement of ELS from bio-aerosols

The continuous interrogation of the laser light beam from AUM, with the microbial aerosols in a small air volume, results in optical scattering. The backscattered light characteristics induced by microbial aerosols in the small interrogation volume are captured by the photodetector and monitored by AUM as a function of time.

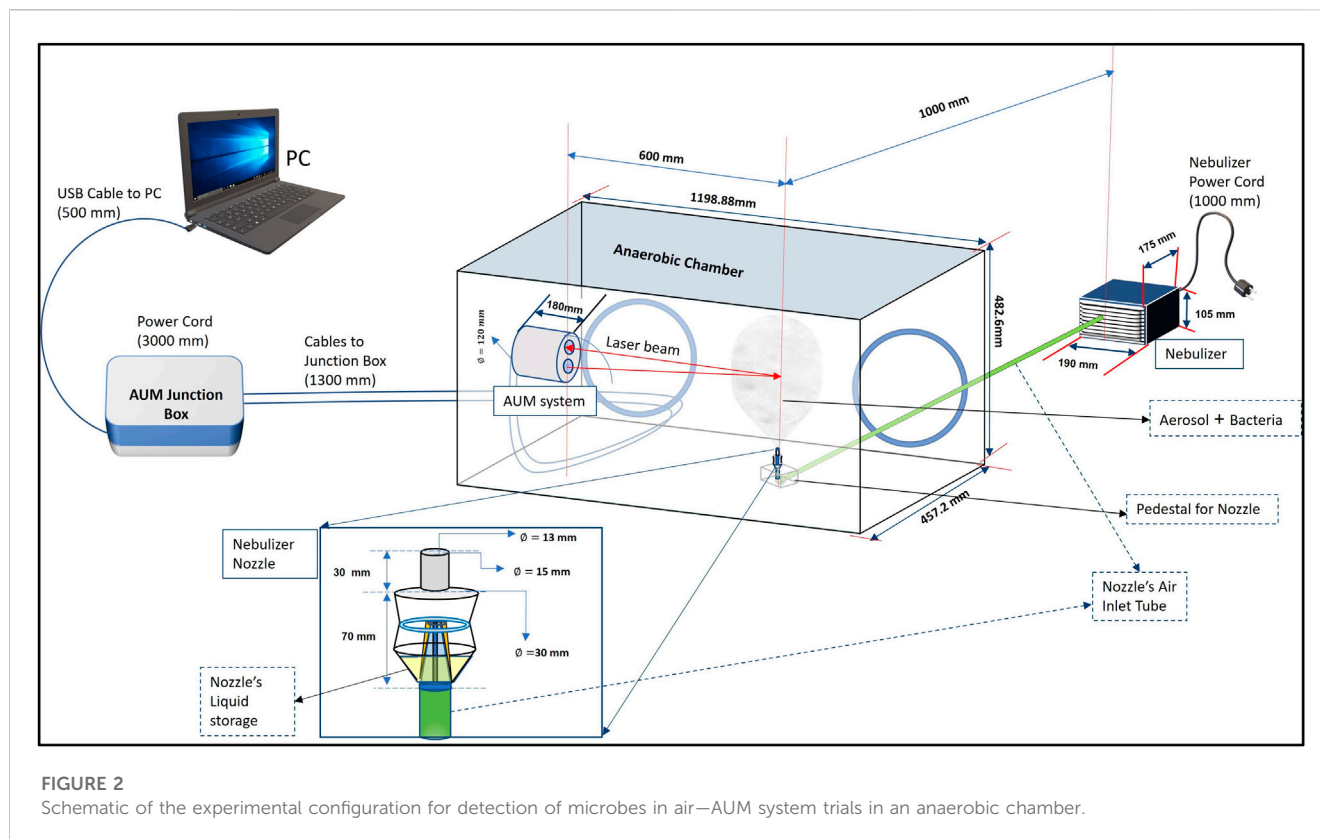
Thus, the backscattered light falling on the PSD carries an imprint of the *in situ* characteristics of scatterers (bioaerosols) present in the scattering volume.

As ELS can change the spatial distribution and propagation direction of the illuminating light, the information regarding scattering could be monitored and analyzed in real-time analysis for the detection and characterization of aerosol particles (encompassing the range of sizes from 100 nm to 10 mm and concentrations from parts per trillion to parts per million) [39–43], in both point and remote-sensing systems [39, 42]. Furthermore, the principles of ELS suggest that the backscattered light signature patterns carry an imprint of the scatterers' size, concentration, molecular weight, and refractive index *per se* in multidimensional space. Therefore, the random patterns of points in *n*-dimensional space, i.e., spatial point processes [44], known to play a unique role in stochastic geometry, can be a valuable model for gaining insights into the phenomena causing them.

2.3 Signal processing techniques for ELS measurements

The identification of bacteria from optical data measurements from unknown samples generally utilizes multivariate statistical methods (Least Squares, Multiple Regression, Principal Components, Support Vector Machines, Neural Networks, etc.), along with database measurements taken from known samples. Many crucial factors, like the classification level (e.g., genus, species, and strain), the range of bacterial types, and the sample preparation's influence on variation, affect these databases.

The method most used when seeking a linear model between two variables is the standard Minimum Least-Squared approximation (MLS), also known as linear regression. However, this method has an underlying assumption that the variable chosen



as independent is assumed to be noiseless. Furthermore, the underlying belief that the base series is free of noise produces a bias in favour of the base series. Thus, there is a need for an analysis technique that considers the uncertainty in both variables under comparison. One such technique, which involves the Eigen modal decomposition of the covariance matrix, is the Empirical Orthogonal Function analysis (EOF). As conventionally applied to data, empirical orthogonal function (EOF) analysis decomposes spatial and temporally distributed data into modes ranked by their temporal variances. In addition, EOF analysis allows partitioning of large data sets into signal-like and noise-like parts.

EOF analysis is also known as the Principal Component Analysis (PCA) and is quite popular in many fields of study. There are several reasons for this popularity. Probably the most important is that this method often enables a description of the variations of a complex field with a relatively small number of functions and associated time coefficients. This property is fundamental in developing statistical prediction schemes relying on multiple linear regression. The skill and statistical confidence of prediction depend heavily upon *a priori* methods of reducing the number of available predictors. Secondly, EOF/PCA analysis is popular because the derived empirical functions are often amenable to physical interpretation, which may give substantial insight into complex processes.

PCA accomplishes an orthogonal linear transformation that transforms the data to a new coordinate system such that the most significant variance by some scalar projection of the data comes to lie on the first coordinate (called the first principal component), the second largest variance on the second coordinate, and so on. PCA is a linear dimensionality reduction technique that transforms a set of correlated variables into a smaller number of uncorrelated variables

called principal components while retaining as much variation as possible in the original dataset.

Tatavarti and Andrade [45] have demonstrated that the Eigen decomposition technique of Empirical Orthogonal Function (EOF) analysis (or PCA), and a modified Minimum Least Square (MLS) approximation technique is the same. Furthermore, the modified minimum least square approximation technique—involving a revised definition of error, compared to the standard description of the error—is independent of bias on any variable in the analysis, suggesting that PCA is effective on linear data. Of course, one can still do a PCA computation on non-linear data. Still, the results will be meaningless beyond decomposing to the dominant linear modes and providing a global linear representation of the spread of the data.

The principles of ELS suggest that the backscattered light signature patterns carry an imprint of the scatterers' size, concentration, molecular weight, and refractive index *per se* in multidimensional space [32, 33, 39, 41–43]. Therefore, the random patterns of points in *n*-dimensional space, i.e., spatial point processes [44], which are known to play a unique role in stochastic geometry, can be a valuable model for gaining insights into the phenomena causing them.

A random medium's scattering characteristics depend on the medium's (air, pollutants, and microbes) composition (the geometry, morphology, number, and spatial distribution of its scatterers and their refractive indices). The systems governed by non-linear interactions are ubiquitous in nature and biology. For example, anisotropy can significantly optimize the optical scattering efficiency of a volume of air laden with diverse pollutants and microbes of varying concentrations [42].

TABLE 1 Photonic system AUM trials with injection details of the samples investigated (bacterial strains and mixtures of bacteria) and their pre-injection concentrations before injection. Cell culture particulars, protocols, and procedures adopted for their classifications using gold standards along with durations of each run are highlighted. For trials with Individual microbe strains as well as with mixtures, concentrations introduced into the anaerobic chamber increased with time during each run. During the various Runs with different bacterial strains, the minimum concentrations of bacterial strains injected, were <10 CFU, while the maximum concentrations ranged from 1.6×10^7 CFU to 2.3×10^8 CFU. During the trial runs with mixtures of five bacterial species, the minimum concentrations of bacterial species injected were <10 CFU, while the maximum concentrations ranged from 0.1×10^8 CFU to 3.2×10^8 CFU in the mixtures.

Pathogens/data acquisition in trials	S. No	Bacteria (strain)	Runs (20 min Run, 5 ml vol.)	
			Initial pre-injection concentrations of bacterial strain	
			Cell numbers (CFU/ml)	
9 pathogenic bacterial strains	1	<i>Escherichia coli</i> (ATCC25922)	2.08×10^8	
two mixtures comprising five bacteria species each	2	<i>Escherichia coli</i> (ATCC35218)	1.42×10^8	
Separate runs for each strain and mixture (Trial run—20 min duration)	3	<i>Staphylococcus aureus</i> (ATCC29213)	1.54×10^8	
1 kHz sampling frequency for each run	4	<i>Staphylococcus aureus</i> (ATCC25923)	1.50×10^8	
12,00,000 observations in each run	5	<i>Pseudomonas aeruginosa</i> (ATCC27853)	2.3×10^8	
Good repeatability and robustness	6	<i>Enterococcus faecalis</i> (ATCC29212)	1.02×10^8	
	7	<i>Staphylococcus hominis</i> (H77)	0.16×10^8	
	8	<i>Bacillus cereus</i> (BY44)	1.66×10^8	
<u>Cultures</u> Luria—bertani medium, laminar flow, incubators, phosphate buffered saline (PBS) solution for constant pH of 7.4 <u>Measurements</u> UV Spectrophotometer for optical density (OD _{600nm})/ concentration measurements <u>Standard protocols and safety procedures</u> US CDC norms	9	<i>Enterococcus faecium</i> (MB224)	0.88×10^8	
	S. No	Bacteria (Mixture)	Runs (20 min Run, 5 ml vol.)	
			Initial pre-injection average concentrations of bacterial mixture	
			Cell numbers (CFU/ml)	
	1	Mixture-1	3.16×10^8	
			<i>E. coli</i> (ATCC25922), <i>S. aureus</i> (ATCC29213)	
			<i>P. aeruginosa</i> (ATCC27853), <i>E. faecalis</i> (ATCC29212)	
	2	Mixture-2	3.26×10^8	
				<i>S. hominis</i> (H77)
				<i>E. coli</i> (ATCC35218), <i>S. aureus</i> (ATCC25923)
<i>P. aeruginosa</i> (ATCC27853), <i>B. cereus</i> (BY44)				
		<i>E. faecium</i> (MB224)		




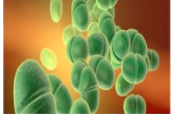
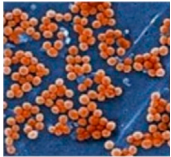

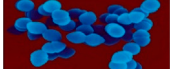
The elastic scattering of light by air laden with diverse pollutants and microbes is a function of their size, shape, molecular structure, refractive index, and concentrations of various microbes in the measurement volume and the distance of scatterers from the detector. The interactions of the scattered radiation due to all these parameters are non-linear and can only be considered linear at best under overly simplistic optimal conditions [42]. Therefore, it becomes more prudent to consider the higher-order statistical moments of scatter data to explore the relationships between optical scattering parameters and the scatterers.

Tatavarti [39] demonstrated the real-world application and use of the AUM technology to detect, discriminate and quantify the various pollutants in the air when AUM was collocated with EPA-approved state-of-the-art air quality monitoring stations (comprising multiple sensors for monitoring different parameters) in India at various geographical locations during

varying environmental conditions over a long period. The outputs from the collocated AUM and the EPA-approved stations yielded comparable results indicating the effectiveness and utility of real-world applications of AUM in detecting, discriminating, and quantifying the various pollutants in the air.

Recent statistical and machine-learning methods for detecting interactions among features include decision trees and their ensembles: CART [46], Random Forests (RFs) [47], Node Harvest [48], Forest Garrote [49], and Rulefit3 [50], as well as methods more specific to gene-gene interactions with categorical features, such as logic regression [51], multifactor dimensionality reduction [52], and Bayesian epistasis mapping [53]. However, except for RFs, the above tree-based procedures grow shallow trees to prevent overfitting, excluding the possibility of detecting high-order interactions without affecting predictive accuracy [54]. RFs are an attractive alternative, looking at high-order interactions to

TABLE 2 The physical and biological characteristics of microbes studied along with their SEM pictures.

S.No	Sample	Gram positive or gram negative	Size	Shape	Anaerobic or aerobic	Scanning electron microscope
1	<i>Escherichia coli</i> (ATCC 25922)	Gram Negative	2 μm -long 0.25–1.0 $\text{\AA}\mu\text{m}$. in diameter	Rod shaped coliform bacteria	Facultative anaerobic	
2	<i>Escherichia coli</i> (ATCC 25922)					
3	<i>Staphylococcus aureus</i> (ATCC 29213)	Gram Positive	0.5–1.0 $\text{\AA}\mu\text{m}$ in diameter	Round shaped bacterium	Facultative anaerobic	
4	<i>Staphylococcus aureus</i> (ATCC 25923)	Gram Negative				
5	<i>Pseudomonas aeruginosa</i> (ATCC 27853)	Gram Positive	0.5–0.8 $\text{\AA}\mu\text{m}$ by 1.5–3.0 μm	Rod shaped bacterium	Strict aerobic	
6	<i>Enterococcus faecalis</i> (ATCC 29212)	Gram positive	0.6–2.0 $\text{\AA}\mu\text{m}$ by 0.6–2.5 $\text{\AA}\mu\text{m}$	Oval shaped cells	Facultative anaerobe	
7	<i>Staphylococcus hominis</i> (H77)	Gram Positive	1.2–1.4 $\text{\AA}\mu\text{m}$ in diameter	Spherical cells in clusters	Acidic aerobic	
8	<i>Bacillus cereus</i> (8Y44)	Gram Positive	1 by 3–4 $\text{\AA}\mu\text{m}$	Rod shaped bacterium, motile, spore forming bacteria	Facultative anaerobic	
9	<i>Enterococcus faecium</i> (MB224)	Gram Positive	1–2 mm (colony size)	Non-haemolytic bacterium	Aerobic and anaerobic	

Ref: 1. Medical Microbiology, University of Texas Medical Branch at Galveston, Galveston, Texas. Editor: Samuel Baron, Link: <https://www.ncbi.nlm.nih.gov/books/NBK8326/>

2. SEM Images: Public Domain, Content Provider: https://en.wikipedia.org/wiki/Centers_for_Disease_Control_and_Prevention's_Public_Health_Image_Library.

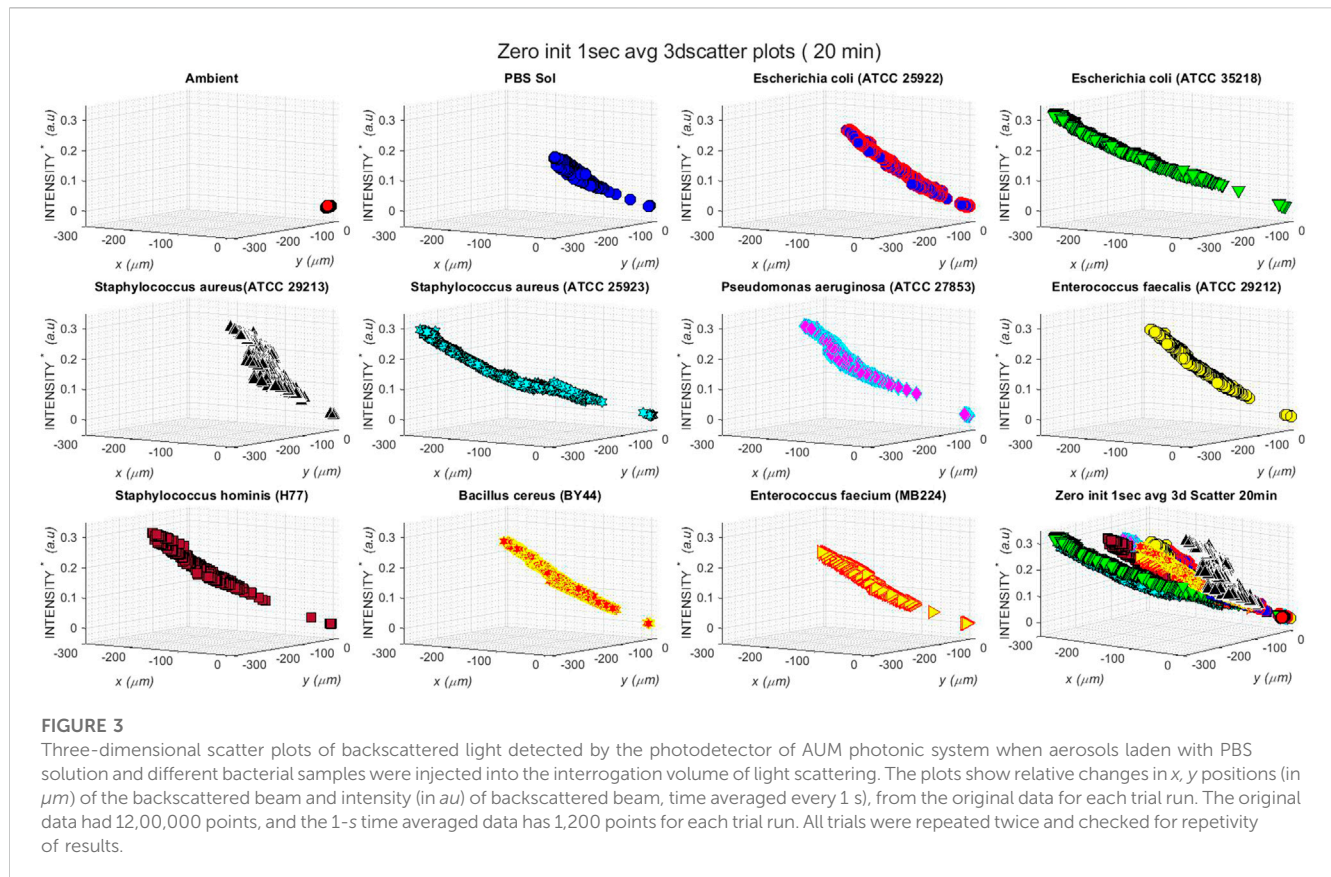
obtain state-of-the-art prediction accuracy, even though interpreting interactions in the resulting tree ensemble remains challenging.

2.4 AUM: ELS, statistical mechanics, artificial intelligence, machine learning principles

The AUM photonic system monitors backscattering characteristics in real-time. It determines the attributes of *in situ* microbes, utilizing a unique algorithm that combines concepts of stochastic geometry with those of artificial intelligence and machine learning. An extended backscatter data matrix is constructed to represent various statistical features of backscattered light data from the observations of the AUM photonic system. From an initial data matrix of M observations of three parameters [i.e., data matrix of size $(M \times 3)$]

of the backscattered light Intensity and positional information at desired sampling; N observations of backscattered light data of seven statistical parameters (four statistical moments, and the three statistical features—minimum, maximum and range), over a chosen averaging interval, are computed to form an extended, scattered data matrix of size $(N \times 35)$, containing multidimensional information. Data analytical tools and a robust machine learning algorithm (Random Forests algorithm) [47, 54, 55] enabled the extraction of information from large and high-dimensional datasets, thereby identifying structures and enabling detection and discrimination of the various microbial characteristics present in the volume of interrogation of light and matter, which were responsible for the scattering process.

We employed an innovative experimental configuration (Figure 2), where AUM was positioned in an anaerobic chamber to make optical



backscattering observations from a small interrogation volume, into which microbial aerosols were introduced continuously as a function of time. To deliver microbial aerosols into the interrogation volume, we used a piston compressor nebulizer which converts the liquid in the machine's carousel (PBS solution laden with microbes) into a fine mist due to controlled pressurized air. Ensuring that the liquid sample is homogeneously mixed and that the nebulization rate is constant, we determined the change in concentration of bioaerosols injected into the interrogation volume as a function of time.

2.5 Microbes for study

The microbes for this study comprised nine bacterial strains and two mixtures of five bacterial species combinations. Table 1 summarizes the pathogenic bacteria, pre-injection concentrations, and data acquisition details. Table 2 highlights the characteristics of pathogenic bacteria considered for this study. The selected strains and mixtures are known to cause various infections and are resistant to antimicrobial drugs. The samples of pathogenic bacteria were cultured by adopting the gold standard protocols and procedures of cell culture. The selected pathogenic strains were brought to 37°C after retrieving from -80°C and placing them in sterile autoclaved distilled water. Luria-Bertani (LB), a widely used bacterial culture medium for various facultative organisms, was used for this study. LB broth medium was prepared and autoclaved at 121°C for 15 min. The sterile medium was inoculated with individual stains under aseptic conditions using a Laminar Airflow cabinet and incubated

aerobically in the incubator at 37°C for 18–24 h. 2% of the overnight grown seed culture was inoculated in a sterile medium and set further to count the cell numbers. The grown cultures, harvested by centrifugation at 10,000 rpm for 10 min, and the cell pellets were washed twice with Phosphate Buffered Saline (PBS) solution and resuspended in 2 ml of PBS to measure optical density ($\text{OD}_{600\text{nm}}$) using UV Spectrophotometer (Microplate Reader, 800 TS, United States). Approximate cell numbers were calculated by the standard value of OD_{600} ($0.1 \text{ OD} = 10^8 \text{ cells/ml}$). Table 1 details the individual bacteria strains and the mixtures of bacterial species, their initial pre-injection concentrations (final cell numbers measured after 24 h incubation), cell culture particulars, protocols, and procedures adopted for their classifications using gold standards, along with durations of each trial run (Table 1). Pathogenic bacterial species and their initial concentrations for aerosol injection cover a wide range of detection sensitivity and specificity [2].

2.6 AUM system trials for microbe detection and characterization

Tests were conducted by continuously injecting different microbial samples (nine different bacterial strains and two different mixtures of five bacterial species) through the nebulizer for each trial run. During that time, the AUM photonic system recorded the optical backscattered data at a sampling frequency of 1 kHz, for a duration of 20 min resulting in 12,00,000 observations for each trial. The large number of observations ensured tests for

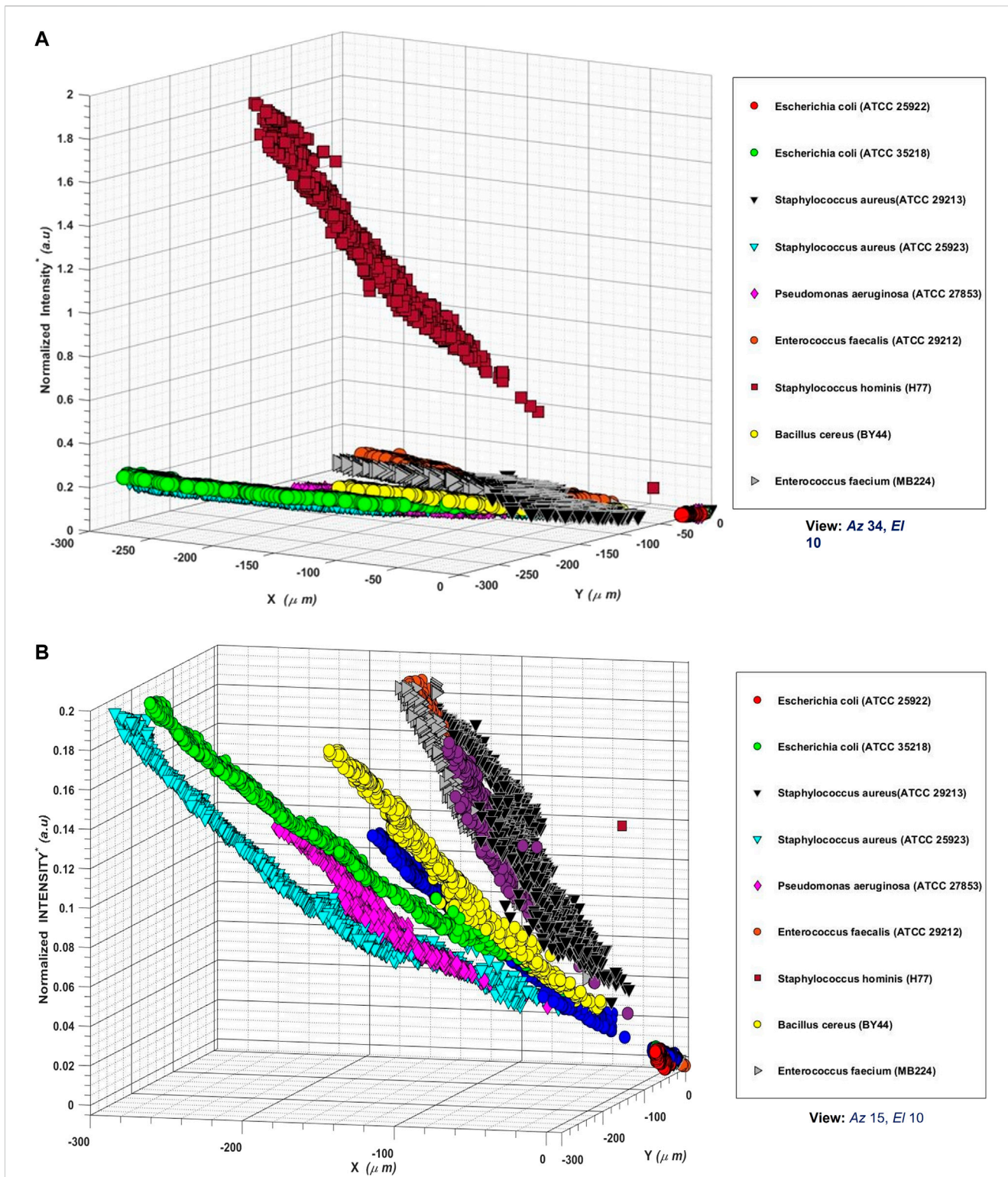


FIGURE 4

(A) Three-dimensional scatter plots of backscattered light detected by the photodetector of AUM photonic system when aerosols laden with different bacterial samples were injected into the interrogation volume of light scattering. The plots show x, y positions (in μm) of the backscattered beam and normalized intensity (in au) of backscattered beam, time averaged (for every 1 s). The plots show the relative changes in x, y positions (in μm) of the backscattered beam and the normalized intensity (in au) of backscattered beam, time averaged (for every 1 s). Intensity values were normalized by dividing them with the pre-injection concentration values of the species for each trial. The multiplication factor of 10^{-8} in the normalized intensity values is not shown on the vertical axes, to avoid clutter in the plot. (B) Three-dimensional scatter plots of backscattered light detected by the photodetector of AUM photonic system when aerosols laden with different bacterial samples were injected into the interrogation volume of light scattering. This figure is a zoomed in (in the Z axis) version of Figure 4A, from a different view.

TABLE 3 Best fit equations of the zero-initialized intensity (backscattered) observations of different samples (ambient, PBS solution, nine bacterial strains) with time (i.e., with varying concentrations). The pre-injection concentrations of species and their respective volumes are shown. Data pertain to the trials of 20 min duration runs. The start and stop of injection times (in seconds) of bioaerosol samples during different runs are indicated in the last column.

Sample injected during trial run	Best fit equation	Goodness of fit	Relative start and stop times (in sec) of sample injection during runs
Initial pre-injection conc. (CFU/ml)		(R^2)	
Ambient Air	$I = 3.204e^{-07}t^2 - 5.275e^{-05}t + 0.001115$	0.7136	5–235 s data only
PBS solution	$I = 9.16400e^{-05}t + 0.07732$	0.8820	5–898 s injection
S1, <i>Escherichia coli</i> (ATCC 25922) [2.08×10^8]	$I = 0.0001893t + 0.0654$	0.9685	200–1,180 s injection
S2, <i>Escherichia coli</i> (ATCC 35218) [1.42×10^8]	$I = 0.0001772t + 0.101$	0.9847	11–1,190 s injection
S3, <i>Staphylococcus aureus</i> (ATCC 29213) [1.54×10^8]	$I = 0.0001597t + 0.08181$	0.9683	11–1,189 s injection
S4, <i>Staphylococcus aureus</i> (ATCC 25923) [1.50×10^8]	$I = 0.0001879t + 0.04601$	0.9473	11–1,189 s injection
S5, <i>Pseudomonas aeruginosa</i> (ATCC 27853) [2.30×10^8]	$I = 0.0001224t + 0.1300$	0.9497	11–1,189 s injection
S6, <i>Enterococcus faecalis</i> (ATCC 29212) [1.02×10^8]	$I = 0.0001597t + 0.07891$	0.9808	11–1,189 s injection
S7, <i>Staphylococcus hominis</i> (H77) [0.16×10^8]	$I = 0.0001455t + 0.1245$	0.9609	11–1,189 s injection
S8, <i>Bacillus cereus</i> (BY44) [1.66×10^8]	$I = 0.0001706t + 0.06605$	0.9813	11–1,189 s injection
S9, <i>Enterococcus faecium</i> (MB224) [0.88×10^8]	$I = 0.0001329t + 0.07161$	0.9165	11–1,130 s injection

statistical repeatability, robustness, and a high degree of statistical confidence. As a result of using a nebulizer during each run, the microbial concentrations inside the anaerobic chamber continuously increased with time during each run. The increase of concentrations as a function of time was computed based on the initial concentrations of the samples, the injection rate of the nebulizer, the pre-injection volume of the sample in the carousel after dilution in PBS solution, and the time of the trial durations. It was established that the increase of microbe concentration in the anaerobic chamber was linear with time. For each sample, trials with each sample run for a duration of 20 min were conducted. All the experiments were video recorded to gain additional insights during analysis. During the tests with individual microbe strains, as the pre-injection concentrations of the samples were different, the minimum concentrations of bacterial strains injected were <10 CFU, while the maximum concentrations injected ranged from 1.6×10^7 CFU to 2.3×10^8 CFU, whereas in trials with two mixtures comprising five different bacterial species with the individual injection concentrations ranging from 10 CFU to 10^8 CFU, the average initial pre-injection mixture concentrations varied from 3.1×10^8 CFU to 3.2×10^8 CFU. In summary, the whole range of aerosol bacterial strains, generally reported in various environments (from low concentrations to high concentrations and low specificity to high specificity), were covered during the trials.

3 Results and discussion

Figure 3 shows the 3D (three-dimensional) scatter plots of backscattered light detected by the photodetector of the AUM photonic system in ambient conditions and when aerosols laden with PBS solution and nine different bacterial samples were injected into the interrogation volume of light scattering. The 3-D plots reflect the (x, y) location of the backscattered beam's centroid position on the photodetector, and the light intensity of the

backscattered beam, as a function of time for the trial duration. Although not shown explicitly, the plots have time information embedded in them. For clarity and statistical robustness, only the time-averaged (1 s) observations for each trial are shown. For each trial run of 20 min duration, in the corresponding 3D scatter plots, there are 1,200 averaged backscattered data points depicting the (x, y) location of the backscattered beam's centroid position on the photodetector, and the light intensity of the backscattered beam. In plotting Figure 3, the first values of the 1-s averaged data of x, y positions, and the light intensity were deducted from all the other consecutive values of x, y , and the light intensity respectively, for each trial, thus ensuring the elimination of the effects (if any) of the prevailing ambient air in the measurement volume and the artifacts (if any) of the changes in the light source characteristics. All subplots have the same axes and limits for visual intercomparison.

The 3D backscattered data plots of Figure 3 demonstrate the effects of 1) ambient conditions in the chamber, i.e., ambient air only, when no aerosols were injected (top left subplot); 2) the injection of the aerosolized PBS solution into the anaerobic chamber, with concentration increasing with time (subplot on the top row, second from left); and, 3) the injection of aerosolized pathogenic bacterial strains (nine different strains) into the anaerobic observation chamber of increasing concentration with time (see the nine subplots starting from the top row, third from left to right, to the bottom row third from left).

The static position of the scatter at the origin of the three axes (top left subplot, Figure 3); indicates that there is no influence of ambient air conditions, and there is no optical scattering as there are no aerosols introduced into the measurement volume. With increasing time and therefore increased concentrations of the injected PBS and the nine different bioaerosols in the observation chamber, the data scatter moved diagonally from the bottom right towards the top left direction in each subplot of Figure 3. The 3D backscattered data plot (subplot on the top row, second from left,

Figure 3) due to increasing concentrations of PBS solution and the nine bioaerosols (the nine subplots starting from the top row, third from left to right, to the bottom row third from left, Figure 3) in the chamber shows the backscattered light intensity increases with time as the concentrations of scatterers increase in the measurement volume, suggesting a direct linear relationship between the concentration of scatterers and the backscattered light intensity.

The effects of the start and stop of bioaerosol injection in each plot are evident as the initial gap in scatter near the origin of axes and the varying extents of scatter in the negative (x, y) axes due to the different concentrations of injected aerosols. In addition, we see that the location of the scattered beam positions projected on the 2D photodetector also increases with increasing concentrations of injected aerosols (of PBS and the nine bacterial strains). Assuming that the different slopes of the scatter plots of various aerosol injections into the observation volume, seen in Figure 3, are artifacts of the sizes, shapes, and structures of the injected aerosols in the scattering volume of interrogation, we addressed the question of resolvability of these artifacts from the backscatter observations by superposing all the backscattered data on one plot.

To further investigate the role of the characteristics of the scatterers, we normalized the relative scattered light intensity values by the pre-injection concentration values of the respective aerosols injected into the observation chamber and plotted in Figure 4. The multiplication factor of 10^{-8} in the normalized intensity values is not shown on the vertical axes to avoid clutter in the plot. Again, the first values of the 1-s averaged data were deducted from all the consecutive values of x, y , and the normalized scattered intensity, for each trial while plotting Figure 4, with 1,200 averaged backscattered data points for each subplot to ensure statistical robustness and the elimination of 1) the effects

(if any) of the prevailing ambient air in the measurement volume, 2) the artifacts (if any) of the changes in the light source characteristics, and 3) the effects of pre-injection concentrations of the microbes. Figures 4A, B show the superposed 3D backscattered data plots of runs encompassing ambient air (no injection), injection of PBS solution, and each of the nine pathogenic bacterial strains. Figure 4B is the zoomed-in version of Figure 4A, but from a different view angle [in Figure 4A, the 3D view is from an Azimuth of 34° and an Elevation of 10° ; while in Figure 4B, the Azimuth is 15° and the Elevation is 10°].

Several insights become apparent in Figure 4: 1) the light scattering effects due to different aerosols in the interrogation volume are both detectable and resolvable, underlying the role of size, shape, and structure of the aerosols in the scattering process, 2) not only the different bacterial species but also the different strains of the same species, are detectable and resolvable, underlying the role of the structure and composition of the aerosols, 3) the effect of aerosol size is apparent on the observed scattered intensities, with increased aerosol size resulting in increased scattered light intensity, 4) the effect of aerosol concentration is apparent on the observed scattered intensities, with increased aerosol concentration resulting in increased scattered light intensity, 5) the location of observed light scatter position (x, y) is indicative of the shape of the bioaerosols in the interrogation volume with spherical or round shaped aerosols resulting in light intensity scatter with steeper slopes in an (x, y) plane compared to rod or rectangular)-shaped aerosols, 6) the spread of the observed scatter positions with a skewness towards x, y , and z directions with a particular slope of scatter, is indicative of the spectral sizes and shapes of the aerosols in the interrogation volume, and 7). the gap in the scatter plots (at the lower right corner closer to the origin [$x, y, z = 0$] is the artifact of

S. No	Bacterial strain	Pre-injection initial concentration (CFU/ml)	Slope of linear fit equation	Normalized slope of linear fit equation (Per unit initial conc.)
		A	(Scattered intensity vs. Time)	B/A
			B	
1	<i>Enterococcus coli</i> (ATCC 25922)	2.08E + 08	1.89E + 04	9.10E-13
2	<i>Enterococcus coli</i> (ATCC 35218)	1.42E + 08	1.77E + 04	1.248E-12
3	<i>Staphylococcus aureus</i> (ATCC 29213)	1.54E + 08	1.59E + 04	1.03E-12
4	<i>Staphylococcus aureus</i> (ATCC 25923)	1.50E + 08	1.87E + 04	1.25E-12
5	<i>Pseudomonas aeruginosa</i> (ATCC 27853)	2.30E + 08	1.22E + 04	5.32E-13
6	<i>Enterococcus faecalis</i> (ATCC 29212)	1.02E + 08	1.59E + 04	1.56E-12
7	<i>Staphylococcus hominis</i> (H77)	1.60E + 07	1.45E + 04	9.09E-12
8	<i>Bacillus cereus</i> (BY44)	1.66E + 08	1.70E + 04	1.02E-12
9	<i>Enterococcus faecium</i> (MB224)	8.80E + 07	1.32E + 04	1.51E-12

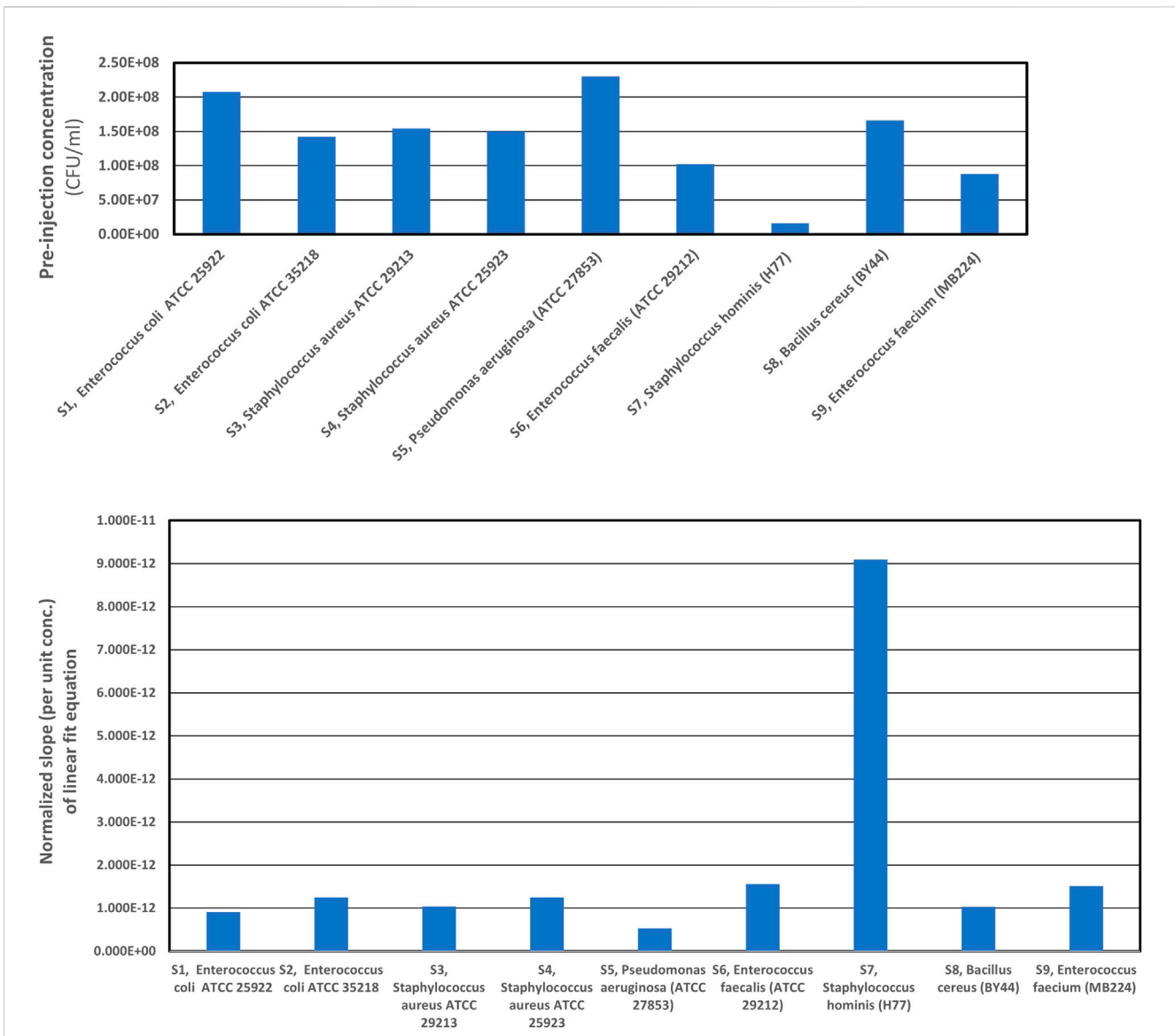


FIGURE 5 Normalized (per unit concentrations) linear slopes of the best fit equations, from the back scattered intensity data for individual bacteria strains computed for 20 min trial runs. We observe that the normalized slopes per unit concentrations are different, for different strains having varied physical characteristics and concentrations. Therefore, one can infer that the relationship between best fit equation's normalized slope of backscattered light intensity data (see Table 4), the physical characteristics of different bacterial strains (scatterers) as well as the concentration of scatterers in the scattering volume of interrogation is imprinted in the optical back scatter information.

TABLE 4 Trial details highlighting the two mixtures of bacteria, their composition, pre-injection concentrations, for each of the two trial runs of 20 min durations each. The relative start and stop times of sample injection during trial runs with mixtures, are 10 sec and 1140 sec respectively.

Mixture 1 (five species mixed with equal volume proportions)	Initial pre-injection species concentrations (CFU/ml)	Mixture 2 (5 species mixed with equal volume proportions)	Initial pre-injection species concentrations (CFU/ml)
Initial pre-injection concentration 3.16×10^8 CFU		Initial pre-injection concentration 3.26×10^9 CFU	
<i>Escherichia coli</i> (ATCC 25922)	3.18×10^8	<i>Escherichia coli</i> (ATCC 35218)	3.30×10^8
<i>Staphylococcus aureus</i> (ATCC 29213)	2.92×10^8	<i>Staphylococcus aureus</i> (ATCC 25923)	3.10×10^8
<i>Pseudomonas aeruginosa</i> (ATCC 27853)	3.16×10^8	<i>Pseudomonas aeruginosa</i> (ATCC 27853)	3.16×10^8
<i>Enterococcus faecalis</i> (ATCC 29212)	2.64×10^8	<i>Bacillus cereus</i> (BY44)	2.78×10^8
<i>Staphylococcus hominis</i> (H77)	1.60×10^7	<i>Enterococcus faecium</i> (MB224)	2.72×10^7

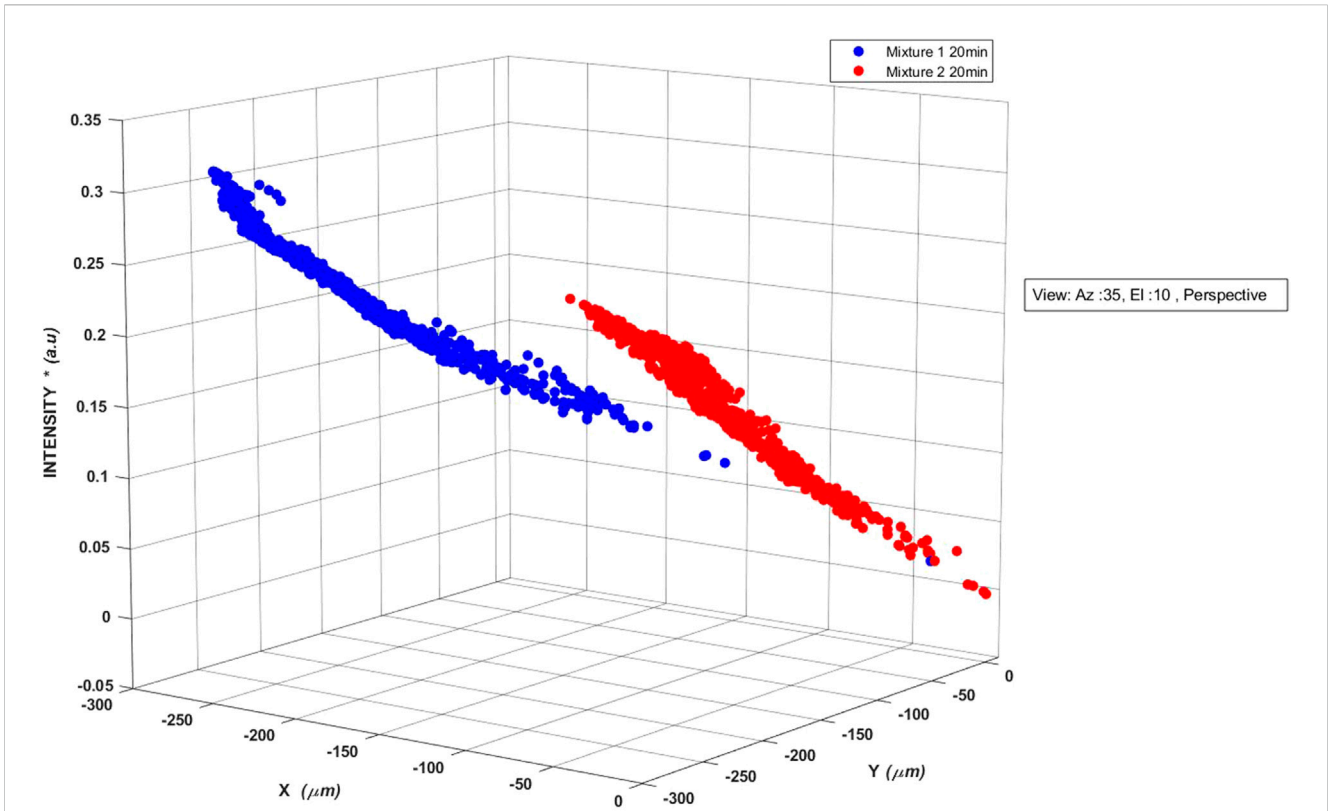


FIGURE 6

Three-dimensional scatter plots of backscattered light detected by the photodetector of AUM photonic system when aerosols laden with different bacterial mixtures were injected into the interrogation volume of light scattering. The multiplication factor of 10^{-7} in the normalized intensity values is not shown. The superposed 3D backscattered data indicate that backscattered data pertaining to the two different microbial mixtures of species still occupy unique Euclidian space but raises questions regarding resolvability into their constituent individual microbial species. Therefore, the need for appropriate models for resolving the mixture scatter into different individual species present in the mixture.

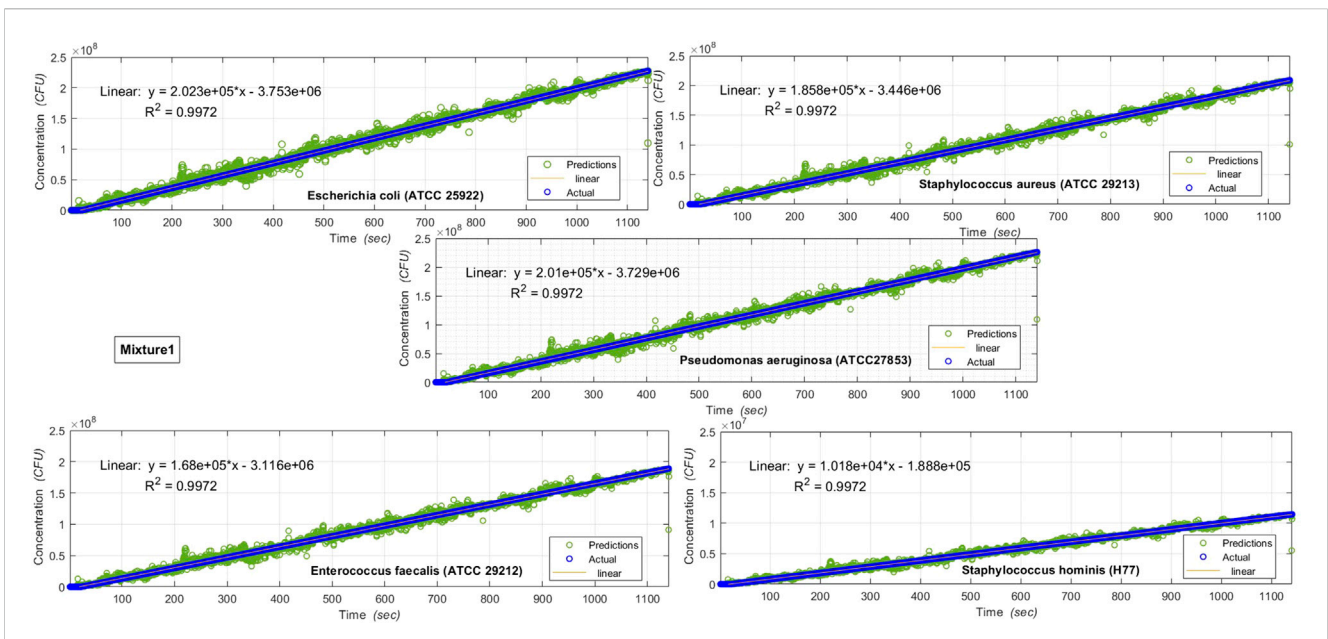
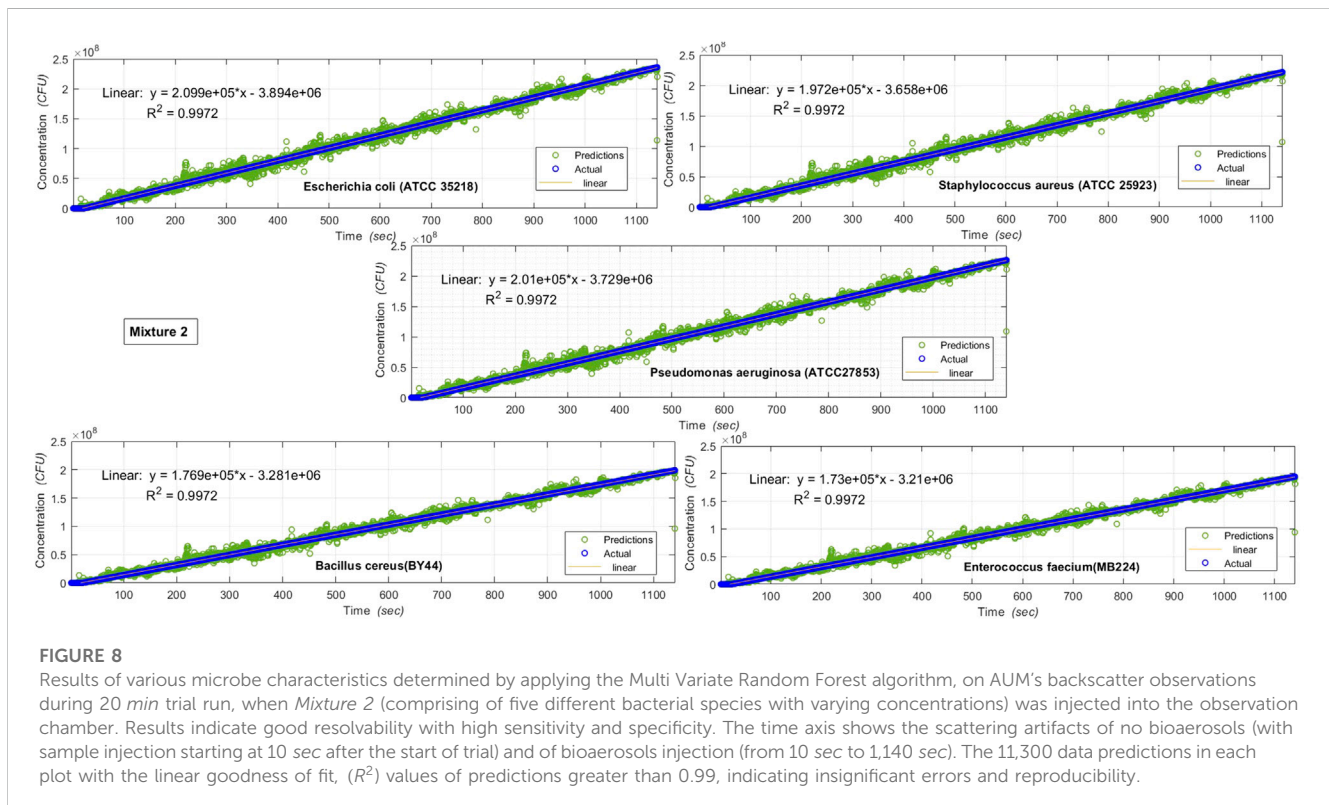


FIGURE 7

Results of various microbe characteristics determined by applying the Multi Variate Random Forest algorithm, on AUM's backscatter observations during 20 min trial run, when *Mixture 1* (comprising of five different bacterial species with varying concentrations) was injected into the observation chamber. Results indicate good resolvability with high sensitivity and specificity. The time axis shows the scattering artifacts of no bioaerosols (with sample injection starting at 10 sec after the start of trial) and of bioaerosols injection (from 10 sec to 1,140 sec).



the effect of lag in starting the nebulization, i.e., the injection of aerosols into observation chamber.

The superimposed 3D backscattered data plots, therefore, indicate that backscattered data (from PBS and microbe-laden aerosols having distinct characteristics of size, shape, and structure, with varying concentrations) occupy a unique Euclidian space dictated by the features of the scatterers in the interrogation volume of air; suggesting that backscattered light data analyses can reveal microbe characteristics. Figures 4A, B thus showcase the results of the scatterers' features (size, shape, structure) on the observed backscattered data.

Effects of increasing time and, therefore, increasing concentrations of the microbe-laden aerosols can be seen as the scatter go diagonally from the bottom right (origin of axes) towards the top left direction in each subplot. The artifacts of the start of injection of the bioaerosol sample, and the stop of injection of the bioaerosol samples of varying initial concentrations for each sample run, are reflected as different start and stop positions of scattering in all the 3 days scatter plots. We notice that for each sample run, the 3D scatter begins at the origin of the axes (when there is no injection of bioaerosols). With the inception of bioaerosol injection at a particular concentration, the location of the 3D scatters shifts from the origin of axes towards the negative (x, y) direction with varying scattered light intensity, consistent with the initial concentration of the bioaerosols. Gradually, the shift increases in the negative (x, y) direction with increasing intensity values, consistent with increasing concentrations of bioaerosols of the sample in the scattering volume. As the injection of bioaerosols stopped, the scatter position fell back towards the axes' origin after some time. Although not in the present study's scope, we expect this

would be in sync with the residence times and free fluid dynamics of injected bioaerosol samples. For each of the different trials (involving different strains of the same species and various species), the slopes of the scatter plots were different.

In short, Figures 4A, B show that the backscattered beam positions projected on a 2D planar area of ($300 \times 300 \mu\text{m}$) on the photodetector, caused by the scatterers (different strains of bacteria, with concentrations ranging from 10–109 CFU) and the relative normalized scattered light intensity changes due to increasing concentrations of scatterers occupying a range of ($0-2 \times 10^{-8} \text{ au}$), are not only detectable but also resolvable, indicating the high sensitivity, high resolution, and wide dynamic range of the AUM system's detectability and discrimination capabilities.

The superposed 3D backscattered plot from all bacterial strains (Figure 4) shows a bouquet of scatter points resolvable in time and space, indicating that the backscattered data about each of the microbe strains/species (having varying sizes, shapes, and concentrations) does indeed occupy a unique n -dimensional Euclidian space, and therefore, using appropriate algorithms can reveal information about the microbe characteristics in real-time.

Table 3 highlights the bacterial species, their initial concentrations, slopes, and normalized (per unit concentrations) slopes of linear fit equations. Figure 4 with Table 3 demonstrates that there is a direct linear relationship between microbe concentration and the backscattered light intensity (with R^2 values greater than 0.9 for all nine bacterial strains during the 20 min (1200 s) trials.

Figure 5 shows the plots of the pre-injection concentrations of different bacterial strains investigated and the normalized linear slopes (per unit pre-injection concentrations) of the best-fit equations from the observed backscattered intensity data for

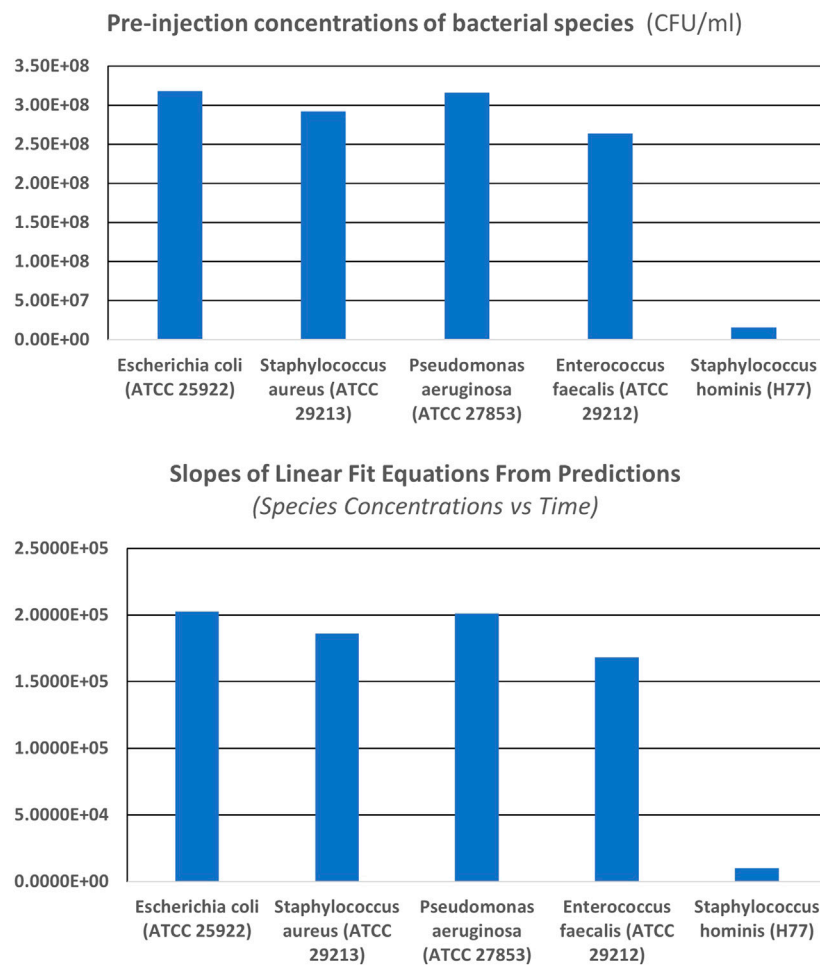


FIGURE 9

Pre-injection concentration (CFU/ml) plot of the individual species in *Mixture 1* (top) and the plot of the slopes of the linear fit equations (of species concentrations (CFU) as a function of time) from predictions for *Mixture 1*, computed by the machine learning algorithm based on the back scattered data from AUM system (bottom). The predicted slopes of the linear fit equations of species concentrations as a function of time are consistent with the pre-injection concentrations of the individual species of the *Mixture 1*. The linear goodness of fit, (R^2) values of predictions are greater than 0.99.

different bacterial strains during the 20 min run trials. We observed that the normalized slopes per unit concentrations are different for different strains of the same species and other species having varied physical characteristics. For example, *Staphylococcus hominus* (H77), which are known to be spherical cells (4–4.5 μm in diameter when cultured in Agar medium) and are known to occur in clusters (typically in tetrads and sometimes in pairs), have distinctly different scattering characteristics with higher normalized slopes indicating that the higher the size, higher the slope (per unit concentration) of the scattered intensity.

Figure 5 also showcases the special relationship between the normalized slopes of best-fit linear equations of scattered intensity and the concentrations of the scatterers (i.e., as a function of time). Looking at the composite scattered plot of 3-D scattering characteristics of all the individual bacterial strains, we summarize that the aerosols' size, shape, and concentration govern the location of the scattered light beam and its intensity. The plots, therefore, indicate the robust relationship between the

best-fit equation's normalized slope of backscattered light intensity data, physical characteristics, and the concentration of scatterers in the scattering volume of interrogation, further validating the principle that backscattered data is uniquely representative of the physical characteristics and concentration of the scatterers present in the interrogation volume. Thus, AUM's utility for determining the microbe characteristics (the individual microbial strains/species) in real-time with high sensitivity and specificity is justified.

Table 4 shows the trial details highlighting the two mixtures of bacteria, their compositions, and concentrations for each of the two runs of 20 min duration. Each mix had five different bacterial species (of equal proportions in volume) with varying initial pre-injection concentrations. The superposed 3D scatter plots from the AUM photonic system's observations (x, y positions, and intensity of scattered beam), of the trials of two different mixtures, are showcased in Figure 6. The plots show the one second time averaged, normalized backscattered beam intensity (per unit concentration in arbitrary units, au) and (x, y) positions of the

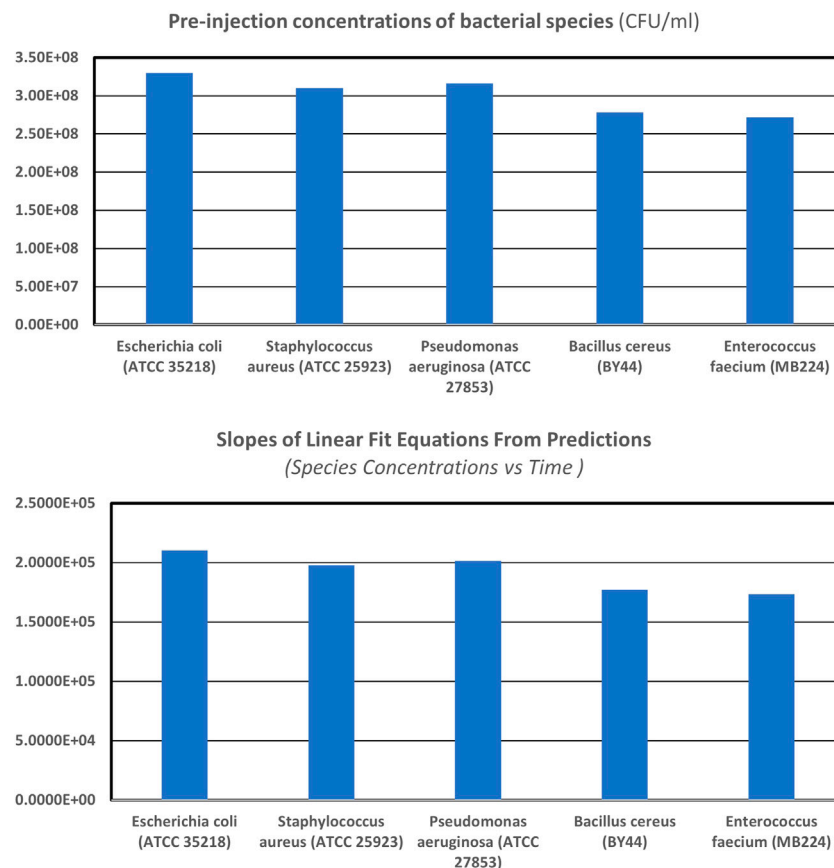


FIGURE 10

Pre-injection concentration (CFU/ml) plot of the individual species in *Mixture 2* (top) and the plot of the slopes of the linear fit equations (of species concentrations (CFU) as a function of time) from predictions for *Mixture 2*, computed by the machine learning algorithm based on the back scattered data from AUM system (bottom). The predicted slopes of the linear fit equations of species concentrations as a function of time are consistent with the pre-injection concentrations of the individual species of the *Mixture 2*. The linear goodness of fit, (R^2) values of predictions are greater than 0.99.

backscattered beam, as a function of increasing time (i.e., increasing concentrations of scatterers in the scattering interrogation volume). The scattering results due to the two mixtures, comprising five different pathogenic bacterial species each, are shown in different colors. The effects of the prevailing ambient air in the observation chamber were removed by initializing the first values of scattered data to zero for each run. The superposed 3D backscattered data plots (Figure 6) indicate that the backscattered data of the two different microbial mixtures of species occupy unique Euclidean spaces, unique to the mixture composition.

Nevertheless, the resolvability of the mix into various microbial species warrants additional signal processing. Figure 6 indicates that when a combination of microbes is present, the mixture characteristics can mask the individual species information and validates the necessity for applying the machine learning algorithm on the extended scattering matrix data for extracting microbe information with high sensitivity and specificity. Thus, appropriate signal processing models are needed to resolve the mixture details into the contributions of different species.

The AUM system trials demonstrated that although the backscattered light signature patterns carried an imprint of the

scatterers' size, concentration, molecular weight, and refractive index *per se* in multidimensional space (Figures 3–6), there is a need to apply appropriate signal processing techniques on the observed data to gain additional insights on the scatterers *per se*. Therefore, we chose Random Forrest machine learning algorithms to tap the information in the non-linear structure of scatter in multidimensional space [45, 46].

By initially training a subset of the optical backscattered data sets with corresponding subsets of microbe data of known features, we created machine learning algorithms for estimating microbe characteristics. Each subset comprised 40% of the total data sets of 12,00,000 data points (i.e., 4,80,000). We have used 80% of the chosen subset of observations for training (3,84,000 data points) and tested the data on the remaining 20% of the subset observations (96,000 data points). The models thus trained and tested not only yielded high predictive performance (correlation coefficients >0.95) but also revealed the feature importance based on Gini impurity values, indicating how much each feature contributed to class prediction, allowing us to iterate on our selection of the features and validate the choice of the statistical features chosen in our Random Forest algorithm [45, 46, 52–54]. Finally, we deployed the trained and tested models (which demonstrated high predictive

TABLE 5 Trial details highlighting the two mixtures of bacteria, their composition, pre-injection concentrations, for each of the runs along with the slopes of the linear fit equations from predicted results of concentrations of individual species as a function of time, and the normalized slopes (per unit initial concentrations) of linear fit equations. The slopes of linear fit equations are consistent with the pre-injection initial concentrations of individual species of the mixture of scatterers present in the scattering volume of interrogation. The normalized slopes of individual species in the mixtures are identical, showing the dependence only on the average concentration of the mixture in each run.

S. No	Mixture 1 (Bacterial strain samples)	Pre-injection initial concentrations (CFU/ml)	Slopes of linear fit equations from computed results (concentration vs. time)	Normalized slopes (per unit initial concentration) of linear fit equations
1	<i>Escherichia coli</i> (ATCC 25922)	3.18E + 08	2.02E + 05	6.37E-04
2	<i>Staphylococcus aureus</i> (ATCC 29213)	2.92E + 08	1.85E + 05	6.37E-04
3	<i>Pseudomonas aeruginosa</i> (ATCC 27853)	3.16E + 08	2.01E + 05	6.37E-04
4	<i>Enterococcus faecalis</i> (ATCC 29212)	2.64E + 08	1.68E + 05	6.37E-04
5	<i>Staphylococcus hominis</i> (H77)	1.60E + 07	1.01E + 04	6.37E-04
S. No	MIXTURE 2 (Bacterial strain samples)	Pre-injection initial concentrations (CFU/ml)	Slopes of linear fit equations from computed results (concentration vs. time)	Normalized slopes (per unit initial concentration) of linear fit equations
1	<i>Escherichia coli</i> (ATCC 35218)	3.30E + 08	2.09E + 05	6.36E-04
2	<i>Staphylococcus aureus</i> (ATCC 25923)	3.10E + 08	1.97E + 05	6.37E-04
3	<i>Pseudomonas aeruginosa</i> (ATCC 27853)	3.16E + 08	2.01E + 05	6.37E-04
4	<i>Bacillus cereus</i> (BY44)	2.78E + 08	1.76E + 05	6.37E-04
5	<i>Enterococcus faecium</i> (MB224)	2.72E + 08	1.73E + 05	6.37E-04

performance with correlation coefficients >0.95) for further testing and evaluation. Such models were applied to the total (100%) trial run data (12,00,000 data points) to derive microbe characteristics, as shown in Figures 7–10. We hypothesize that such frozen models, for individual species or mixtures of species, can be applied in real-time on any optical backscattered data obtained from the AUM system to determine the relevant microbe characteristics in real-time in the air.

On applying the machine learning algorithms on the backscattered optical data obtained from AUM, the individual microbe species characteristics (in the mixtures investigated) can be determined, as shown in Figures 7, 8.

Figures 7, 8 show the results of various individual microbe characteristics determined by applying the machine learning (multivariate Random Forest) algorithm [54] on AUM's backscatter observations during the 20 min trials, when mixtures 1 and 2 (comprising of five different bacterial species with varying concentrations) were used. The results demonstrated the ability of our photonic system AUM for real-time detection, discrimination, and quantification of microbes in the air. Figures 7, 8 contain a significantly high number of points (original observations for 20 min (i.e., 12,00,000 original data at 1 kHz sampling, were resampled (averaged) at 10 Hz sampling, i.e., 12,000 data points). Each of the plots in Figures 7, 8 has 11,300 model predictions of

concentrations for each of the five species overlain against the actual concentrations introduced into the observation chamber due to the injection of the mixture of species of equal volume but different concentrations. The R^2 (Goodness of Fit, between the actual vs. the predicted) values in each linear scatter plot was >0.99, indicating that the errors in estimation/prediction were insignificant and the results reproducible. We have not shown the error bars on these plots to avoid cluttering the figures.

Figures 9, 10 show that the AUM system's results for different species, as well as other strains of the same species, are accurate for a wide range of concentrations (0 to 10^8 CFU). Results indicate good resolvability of all five different bacterial species in both the mixtures, with high sensitivity and specificity, with R^2 values >0.99, demonstrating the efficacy of the AUM photonic system for real-time monitoring of microbe characteristics in air.

Table 5 details the two mixtures of bacteria, their composition, and pre-injection concentrations for each of the runs, along with the slopes of the linear fit equations from predicted results of concentrations of individual species as a function of time and the normalized slopes (per unit initial concentrations) of linear fit equations. The slopes of linear fit equations are consistent with the pre-injection initial concentrations of individual species of the mixture of scatterers present in the scattering volume of interrogation.

Figures 9, 10 show the pre-injection concentration (*CFU/ml*) plots of the individual species in sample *mixtures* (top) and slopes of the linear fit equations (of species concentrations as a function of time) from predictions for sample *mixtures*, computed by the machine learning algorithm, based on the backscattered data from AUM system (bottom). The predicted slopes of the linear fit equations of species concentrations as a function of time are consistent with the pre-injection concentrations of the individual species of the sample *mixtures*. Thus, demonstrating the ability of the AUM system to detect and discriminate the species/strains comprising the mixes.

4 Conclusion

The experiments conducted in an anaerobic chamber, covering a range of microbial species with varying concentrations, demonstrated AUM's capabilities to detect, discriminate, and quantify the individual microbes and their mixtures in the air in real-time from a remote location. We showed that the results from gold standard culturing methods and conventional sampling strategies, when compared with those obtained by AUM, yielded (R^2) values greater than 0.99.

Since air is a central reservoir for microorganisms in controlled environments such as operating theatres, regular microbial monitoring by AUM helps measure air quality and identify critical situations. The rapidly emerging novel airborne viral diseases, with their evolving variants and expedited global spread due to faster modes of transportation for hosts, necessitate a sensitive and specific rapid screening technology, like the AUM photonic system. Therefore, the deployment of AUM for the detection of microbes in the air, with its capability of good resolvability with high specificity and sensitivity, across various ports of entry can be effective and imperative in not only controlling and managing diseases but also in opening new pathways in understanding the complex disease dynamics.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

References

- Ibacache-Quiroga C, Romo N, Díaz-Viciedo R, Alejandro Dinamarca M. Detection and control of indoor airborne pathogenic bacteria by biosensors based on quorum sensing chemical language: Bio-tools, connectivity apps and intelligent buildings. In: *Biosensing technologies for the detection of pathogens - a prospective way for rapid analysis* (2017). p. 73–88. Chapter 6. doi:10.5772/intechopen.72390
- Ma J, Du M, Wang C, Xie X, Wang H, Zhang Q. Advances in airborne microorganisms detection using biosensors: A critical review. *Front Environ Sci Eng* (2021) 15(3):47. doi:10.1007/s11783-021-1420-8
- Chen Y, Wang Z, Liu Y, Wang X, Li Y, Ma P, et al. Recent advances in rapid pathogen detection method based on biosensors. *Eur J Clin Microbiol Infect Dis* (2018) 37:1021–37. doi:10.1007/s10096-018-3230-x
- Lazcka O, Javier Del Campo F, Munoz FX. Pathogen detection: A perspective of traditional methods and biosensors. *Biosens Bioelectron* (2007) 22:1205–17. doi:10.1016/j.bios.2006.06.036
- Va'radi L, Jia L, Luo DEH, Perry JD, Anderson RJ, Oregae S, et al. Methods for the detection and identification of pathogenic bacteria: Past, present, and future. *Chem Soc Rev* (2017) 46:4818–32. doi:10.1039/c6cs00693k
- Després VR, Huffman JA, Burrows SM, Hoose C, Safatov AS, Buryak G, et al. Primary biological aerosol particles in the atmosphere: A review. *Tellus B: Chem Phys Meteorology* (2012) 64(1):15598. doi:10.3402/tellusb.v64i0.15598
- Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol, and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New Engl J Med* (2020) 382(16):1564–7. doi:10.1056/nejmc2004973
- Hoehl S, Rabenau H, Berger A, Kortenbusch M, Cinatl J, Bojkova D, et al. Evidence of SARS-CoV-2 infection in returning travelers from Wuhan, China. *New Engl J Med* (2020) 382(13):1278–80. doi:10.1056/nejmc2001899
- Meselson M. Droplets and aerosols in the transmission of SARS-CoV-2. *New Engl J Med* (2020) 382(21):2063. doi:10.1056/nejmc2009324
- Anfinrud P, Stadnytskyi V, ChristinaBax PAB. Visualizing speech-generated oral fluid droplets with laser light scattering. *New Engl J Med* (2020) 382(21):2061–3. doi:10.1056/nejmc2007800
- Setti L, Passarini F, de Gennaro G, Barbieri P, Perrone MG, Borelli M, et al. SARS-CoV2RNA found on particulate matter of Bergamo in Northern Italy: First evidence. *Environ Res* (2020) 188:109754. doi:10.1016/j.envres.2020.109754

Author contributions

RT conceived the research idea and applications after designing and developing AUM. RT, with inputs from SN, AP, and PV designed and conducted the experiments, developed the algorithms and the model for characterization. SH and VJ prepared the microbe samples and characterized them before experimentation with AUM. GM, NK, SN, and PV, helped in data acquisition and analysis. The manuscript was prepared by RT, reviewed, and revised by SN, AP, and PV. All authors have approved the submission of the final version of the manuscript.

Acknowledgments

The authors acknowledge CATS Eco-Systems Pvt., Ltd., Nashik, India, for providing material support, VIT University for facilities provided at the Centre for Microbial Research during the trials, and GVP-SIRC for research facilities offered, for this study.

Conflict of interest

GKM and NKM (i.e., 3rd and 4th Authors) were provided fellowships during the tenure of this work, by CATS Eco Systems Pvt Ltd, Nashik, India. CATS Eco Systems Pvt Ltd also facilitated microbial samples for the study. Author SN was employed by the company CASTLE Advanced Technologies and Systems.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

12. Razzini K, Castrica M, Menchetti L, Maggi L, Negroni L, Orfeo NV, et al. SARS-CoV-2 RNA detection in the air and on surfaces in the COVID-19 ward of a hospital in Milan, Italy. *Sci Total Environ* (2020) 742:140540. doi:10.1016/j.scitotenv.2020.140540
13. Zheng Y, Chen H, Yao M, Li X. Bacterial pathogens were detected from human exhaled breath using a novel protocol. *J Aerosol Sci* (2018) 117:224–34. doi:10.1016/j.jaerosci.2017.12.009
14. Peláez EC, Estevez MC, Mongui A, Menéndez MC, Toro C, HerreraSandoval OL, et al. Detection, and quantification of HspX antigen in sputum samples using plasmonic biosensing: Toward a real point-of-care (POC) for tuberculosis diagnosis. *ACS Infect Dis* (2020) 6(5):1110–20. doi:10.1021/acscinfdis.9b00502
15. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature* (2004) 430:242–9. doi:10.1038/nature02759
16. Engering A, Hogerwerf L, Jan S. Pathogen–host–environment interplay and disease emergence. *Emerging Microbes & Infections* (2013) 2(1):1–7. doi:10.1038/emi.2013.5
17. Shen Z, Wang J, Qiu Z, Jin M, Wang X, Chen Z, et al. Detection of *Escherichia coli* O157:H7 with piezoelectric immune sensor based on enhancement with immunonanoparticles. *Acta Microbiologica Sinica* (2009) 49(6):820–5.
18. Mavrikou S, Moschopoulou G, Tsekouras V, Kintzios S. Development of a portable, ultra-rapid and ultra-sensitive cell-based biosensor for the direct detection of the SARS-CoV-2 S1 spike protein antigen. *Sensors* (2020) 20(11):3121. doi:10.3390/s20113121
19. Xu Y, Xie X, Duan Y, Wang L, Cheng Z, Cheng J. A review of impedance measurements of whole cells. *Biosens Bioelectron* (2016) 77:824–36. doi:10.1016/j.bios.2015.10.027
20. Freije CA, Myhrvold C, Boehm CK, Lin AE, Welch NL, Carter A, et al. Programmable inhibition, and detection of RNA viruses using Cas13. *Mol Cell* (2019) 76(5):826–37.e11. doi:10.1016/j.molcel.2019.09.013
21. Wang Y, Wang C, Song L. Distribution of antibiotic resistance genes and bacteria from six atmospheric environments: Exposure risk to human. *Sci Total Environ* (2019) 694:133750. doi:10.1016/j.scitotenv.2019.133750
22. Damborský P, Svitel J, Katrlík J. Optical biosensors. *Essays Biochem* (2016) 60: 91–100. doi:10.1042/ebc20150010
23. Boxrud D. Advances in subtyping methods of foodborne disease pathogens. *Curr Opin Biotechnol* (2010) 21:137–41. doi:10.1016/j.copbio.2010.02.011
24. Espy MJ, Uhl JR, Sloan LM, Buckwalter SP, Jones MF, Vetter EA, et al. Real-time PCR in clinical microbiology: Applications for routine laboratory testing. *Clin Microbiol Rev* (2006) 19(3):165–256. doi:10.1128/cmr.19.1.165-256.2006
25. Sauer S, Kliem M. Mass spectrometry tools for the classification and identification of bacteria. *Nat Rev Microbiol* (2010) 8:74–82. doi:10.1038/nrmicro2243
26. Sandrin TR, Goldstein JE, Schumaker S. MALDI tof ms profiling of bacteria at the strain level: A review. *Mass Spectrom Rev* (2013) 32:188–217. doi:10.1002/mas.21359
27. Rajapaksha P, Elbourne A, Gangadoo S, Brown R, Cozzolino D, Chapman J. A review of methods for the detection of pathogenic microorganisms. *Analyst* (2019) 144: 396–411. doi:10.1039/c8an01488d
28. Davies SC, 2011. London (2013). Infections and the rise of antimicrobial resistance *Annu Rep Chief Med Officer*
29. Bassetti M, Merelli M, Temperoni C, Astilean A. New antibiotics for bad bugs: Where are we? *Ann. Clin Microbiol Antimicrob* (2013) 12:22. doi:10.1186/1476-0711-12-22
30. Alarcón-Angeles G, Arturo G, Álvarez-Romero G, Rosati G, Parolo C, Merkoçi A. Point-of-Care sensors in clinical environments: Potential and challenges. In: *Reference module in biomedical sciences* (2022).
31. Mc Govern C, Steed C, Essan A, Robertson J, Swift S, Vanholsbheec F. Optical Methods for bacterial detection and characterization. *APL Photon* (2021) 6:080903. doi:10.1063/5.0057787
32. Wyatt PJ. Differential light scattering: A physical method for identifying living bacterial cells. *Appl Opt* (1968) 7(10):1879–96. doi:10.1364/ao.7.001879, No.
33. Kerker M. Light scattering instrumentation for aerosol studies: An historical overview. *Aerosol Sci Tech* (1997) 27(4):522–40. doi:10.1080/02786829708965492
34. Holler S, Pan YL, Chang RK, Bottiger JR, Hill SC, Hillis DB. Two-dimensional angular optical scattering for the characterization of airborne microparticles. *Opt Lett* (1998) 23(18):1489–91. doi:10.1364/ol.23.001489
35. Yong-Le Pan KBA, Chang RK, Hart M, Eversole JD, Eversole JD. Characterizing and monitoring respiratory aerosols by light scattering. *Opt Lett* (2003) 28:589–91. doi:10.1364/ol.28.000589
36. Robinson JP, Rajwa BP, Bae E, Patsek V, Roumani AM, Bhunia AK, et al. *Using scattering to identify bacterial pathogens*. Washington DC: OPN Optics & Photonics News (2011). 21–7.
37. Li Y, Liu T, Wang H, O'Riordan K, Bai B, Haga Y, et al. Hatice Ceylan Koydemir Deep learning-enabled detection and classification of bacterial colonies using a thin-film transistor (TFT) image sensor. *ACS Photon* (2022) 9, 2455–66. doi:10.1021/acsp Photonics.2c00572
38. Luo Y, Zhang Y, Liu T, Yu A, Wu Y, Ozcan A. Virtual impactor-based label-free pollen detection using holography and deep learning. *ACS Sensors* (2022) 7 (12), 3885–94. doi:10.1021/acssensors.2c01890
39. Rao T. Ambient air quality monitoring: Impetus, complexities, challenges and solutions. *Glob J Sci Frontier Res* (2021) 21(4):41–58. doi:10.34257/gjsfrhvol21is4pg41
40. Bae E, Bhunia AK. *Label-free light-scattering sensors for high throughput screening of microbes in food. High throughput screening for food safety assessment. Biosensor technologies, hyperspectral imaging and practical applications*. Sawston, Cambridge: Woodhead Publishing Series in Food Science, Technology and Nutrition (2015). 149–63.
41. Pan Y, Kevin A, Arnold J, Cheng S, Kalume A, Piedra CW, et al. (2022). Review of elastic light scattering from single aerosol particles and application in bioaerosol detection. *J Quantitative Spectrosc Radiative Transfer*, 279, 108067–23. doi:10.1016/j.jqsrt.2022.108067
42. James Davis E, Schweiger G. Elastic light scattering. In: *The airborne microparticle: Its Physics, chemistry, optics, and transport phenomena*. Berlin: Springer (2002).
43. Ogendal LH. *Light scattering demystified: Theory and practice*. Copenhagen: University of Copenhagen (2017). 125.
44. Baddeley A, Bárány I, Schneider R. Spatial point processes and their applications. In *Stochastic geometry: Lectures given at the CIME summer school held in martina franca Italy, september* (2007) 13–18, 1–75.
45. Tatavarti R, Andrade Y. The eigenvalue problem as a form of minimum least-squared approximation. *Ocean Engng* (1992) 19(6):511–26. doi:10.1016/0029-8018(92)90021-u, No
46. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and regression trees*. Belmont, CA: CRC Press (1984).
47. Breiman L. Random forests. *Machine Learn* (2001) 45:5–32. doi:10.1023/a:1010933404324
48. Meinshausen N. Node harvest. *Ann Appl Stat* (2010) 4:2049–72. doi:10.1214/10-aos367
49. Meinshausen N. Forest garrote. *Electron J Stat* (2009) 3:1288–304. doi:10.1214/09-ejs434
50. Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat* (2008) 2:916–54. doi:10.1214/07-aos148
51. Ruczinski CKI, LeBlanc ML, Hsu L. Sequence analysis using logic regression. *Genet Epidemiol* (2001) 21:S626–31. doi:10.1002/gepi.2001.21.s1.s626
52. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am J Hum Genet* (2001) 69: 138–47. doi:10.1086/321276
53. Zhang Y, Liu JS. Bayesian inference of epistatic interactions in case-control studies. *Nat Genet* (2007) 39:1167–73. doi:10.1038/ng2110
54. Hoffmann J, Bar-Sinai Y, Lee LM, Andrejevic J, Mishra S, Rubinstein SM, et al. Machine learning in a data-limited regime: Augmenting experiments with synthetic data uncovers order in crumpled sheets. *Sci Adv* (2019) 5:eaau6792. doi:10.1126/sciadv.aau6792
55. Basu S, Kumbier K, Brown JB, Yu B. Iterative random forests to discover predictive and stable high-order interactions. *Proc Natl Acad Sci USA* (2018) 115:1943–8. doi:10.1073/pnas.1711236115