Hussain et al. *Light: Science & Applications* (2020)9:21
https://doi.org/10.1038/s41377-020-0255-6

Official journal of the CIOMP 2047-7538
www.nature.com/lsa

ARTICLE Open Access

An ultra-compact particle size analyser using a CMOS image sensor and machine learning

Rubaiya Hussain[1], Mehmet Alican Noyan[1,2], Getinet Woyessa[3], Rodrigo R. Retamal Marín[4], Pedro Antonio Martinez[1], Faiz M. Mahdi[5], Vittoria Finazzi[1], Thomas A. Hazlehurst[5], Timothy N. Hunter[5], Tomeu Coll[1], Michael Stintz[4], Frans Muller[5], Georgios Chalkias[6] and Valerio Pruneri[1,7]

Abstract

Light scattering is a fundamental property that can be exploited to create essential devices such as particle analysers. The most common particle size analyser relies on measuring the angle-dependent diffracted light from a sample illuminated by a laser beam. Compared to other non-light-based counterparts, such a laser diffraction scheme offers precision, but it does so at the expense of size, complexity and cost. In this paper, we introduce the concept of a new particle size analyser in a collimated beam configuration using a consumer electronic camera and machine learning. The key novelty is a small form factor angular spatial filter that allows for the collection of light scattered by the particles up to predefined discrete angles. The filter is combined with a light-emitting diode and a complementary metal-oxide-semiconductor image sensor array to acquire angularly resolved scattering images. From these images, a machine learning model predicts the volume median diameter of the particles. To validate the proposed device, glass beads with diameters ranging from 13 to 125 μm were measured in suspension at several concentrations. We were able to correct for multiple scattering effects and predict the particle size with mean absolute percentage errors of 5.09% and 2.5% for the cases without and with concentration as an input parameter, respectively. When only spherical particles were analysed, the former error was significantly reduced (0.72%). Given that it is compact (on the order of ten cm) and built with low-cost consumer electronics, the newly designed particle size analyser has significant potential for use outside a standard laboratory, for example, in online and in-line industrial process monitoring.

Introduction

Particle size analysis based on light scattering has widespread application in many fields, as it allows relatively easy optical characterisation of samples enabling improved quality control of products in many industries including pharmaceutical, food, cosmetic, polymer production, etc.[1–3]. Recent years have seen many advancements in light scattering technologies for particle characterisation. For submicron particle measurement, dynamic light scattering (DLS)[4] has now become an industry standard technique. This method analyses the fluctuations of scattered light by particles in suspension when illuminated with a laser to determine the velocity of the Brownian motion, which can then be used to obtain the hydrodynamic size of particles using the Stokes-Einstein relationship. Although DLS is a useful approach to determine the size distribution of many nano- and biomaterials systems, it does suffer from several disadvantages. For example, DLS is a low-resolution method that is not suitable for measuring polydisperse samples, while the presence of large particles can affect the size accuracy[4]. Other scattering techniques have emerged, such as nanoparticle tracking analysis (NTA)[5], which tracks individual particle movement through scattering using image recording. NTA also measures the hydrodynamic size of particles from the diffusion coefficient but

Correspondence: Valerio Pruneri (valerio.pruneri@icfo.eu)
[1]ICFO- Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain
[2]Ipsumio B.V., High Tech Campus, 5656 AE, Eindhoven, Netherlands
Full list of author information is available at the end of the article.

© The Author(s) 2020
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

is capable of overcoming some of the limitations posed by DLS[5,6].

While the above-mentioned techniques are best suited for measuring particles typically in the submicron region, particle size analysers (PSAs) based on static light scattering or laser diffraction (LD)[7,8] have become the most popular and widely used instruments for measuring particles from hundreds of nanometres to several millimetres. Similar scattering theory is also utilised in systems based on non-electromagnetic wave propagation, such as ultrasonic analysers[9,10]. In LD PSAs, a laser beam is used to irradiate a dilute suspension of particles. The light scattered by the particles in the forward direction is focused by a lens onto a large array of concentric photodetector rings. The smaller the particle is, the larger the scattering angle of the laser beam is. Thus, by measuring the angle-dependent scattered intensity, one can infer the particle size distribution using Fraunhofer or Mie scattering models[11,12]. In the latter case, prior knowledge of the refractive index of the particle being measured as well as the dispersant is required.

Commercial LD PSAs have gained popularity due to their broad dynamic range, rapid measurement, high reproducibility and the capability to perform online measurements. However, these devices are generally large in size (~700 × 300 × 450 mm), heavy (~30 kg) and expensive (in the 50–200 K€ range). On the one hand, the large size of common devices is due to the large distance needed between the sample and the detectors to provide the desired angular resolution. Furthermore, their high price is mainly due to the use of expensive laser sources and a large number of detectors, i.e., one sensor for each scattering angle to be monitored. Some commercial devices contain up to twenty sensors. This complexity of commercial LD PSAs, together with the fact that they often require maintenance and highly trained personnel, make them impractical in the majority of online industrial applications, which require the installation of probes in processing environments, often at multiple locations.

The application of LD PSAs is also normally restricted to dilute suspensions. This is because the optical models used to estimate the particle size distribution (PSD) are based on a single scattering approximation. In practice, most industrial processes require measuring concentrated suspensions, where multiple scattering becomes a prominent effect. Multiple scattering in dense media leads to an underestimation of the particle size since the light scattered by the particles encounters diffraction points multiple times before reaching the detector, which in turn increases the apparent scattering angle[13]. To overcome this issue, LD PSAs require appropriate sampling and dilution systems, which increase capital investments and operational costs. Another approach is to apply multiple scattering correction models together with the optical models to compute the PSD. A large number of algorithms for multiple scattering correction can be found in the literature[14–16]. However, these algorithms typically require implementing a complex correction, which increases the computation time and is often not suitable for online measurements[16].

An alternative approach to compute the PSD without the use of optical models and complex correction factors is to apply machine learning (ML) techniques[17]. Machine learning is a valuable tool that relies on pattern recognition to learn and adapt to changes in processes and provide reliable results. It has been previously shown that, given the concentration and the angular distribution of scattered light, ML models can predict particle size even at high concentrations[18,19]. Such optimisations open up new opportunities for the use of LD PSAs in industrial processes without the need for time-consuming and cumbersome sample preparation. However, the low integration level for multiple sensor configuration and high cost of current commercial LD PSAs still remain significant barriers for their widespread implementation in online industrial monitoring.

Particles have also been measured using imaging techniques. More specifically, lens-free imaging systems that use complementary metal-oxide-semiconductor (CMOS) image sensors can perform direct imaging of particle holograms[20] or particle diffraction patterns[21]. These systems allow the measurement of individual particles, differentiating them by geometrical shape, and do not require a significant refractive index difference between the particles and the containing medium[22]. In this work, we propose a novel low-cost and miniaturised PSA in a collimated beam configuration using a CMOS image sensor and an ML model based on a random forest algorithm[23]. In contrast to other lens-free imaging systems using CMOS sensors, we analyse the angular distribution of scattered light from an ensemble of particles, similar to the LD PSA. The proposed PSA device enables the measurement of samples with high concentrations. The key innovation in our proposed device is a small form factor (5 mm diameter, 17 mm long) angular spatial filter (ASF) made with an array of holes with different diameters that are extruded from a polymer rod. Light collected from different sized holes is representative of a different set of scattering angles. Upon illumination of the target sample with a light-emitting diode (LED), the ASF allows characterising the angular dependence of the scattered light by performing angle-resolved cumulative light power measurements. The patented ASF technology[24] enables setting a specific design for each working size range. The rest of the analyser consists of off-the-shelf consumer electronic products, such as a CMOS image sensor array and LED light source. This design significantly reduces the cost and size compared to those

of commercial LD PSAs, which require several detectors to obtain an adequately resolved angular scattering distribution.

To validate the new PSA, glass beads of different size distributions were measured, ranging from 13 µm to 150 µm at several concentrations in liquid dispersions. The random forest algorithm enables overcoming the current understanding of the theoretical limitations due to multiple scattering, enlarging the working size range and application possibilities, especially for measurements in liquid. By analysing the raw ASF images obtained from the CMOS image sensor array, we show how multiple scattering becomes prominent at high concentrations depending on the particle size being measured and how the random forest algorithm can correct this issue. Thus, the proposed PSA has great potential to become a cost-effective and compact solution for a broad range of industrial applications.

Results

Design and fabrication of ASF

The ASF is the core of the proposed particle size analyser, which is capable of distinguishing different spatial frequencies scattered from the sample by means of a low-pass angular filter array. The ASF used in this work is an array of holes of different diameters that function as apertures (Fig. 1a). The angular acceptance— we call it the cut-off angle, θ_c—for the scattered light of the apertures is determined by the hole's diameter (D) and length (L):

$$\theta_c = \arctan\left(\frac{D}{L}\right) \tag{1}$$

The light scattered up to predefined θ_c values (shown with dashed vertical lines in Fig. 1b, c) is measured using a CMOS image sensor array that can simultaneously acquire power from multiple apertures (holes). This design allows the reconstruction of the cumulative angular scattering profile, as shown in Fig. 1c. In this description of the PSA and ASF working principle, we assume, for simplicity, that the inner walls of the ASF do not reflect, there is no crosstalk between the holes and the hole filtering has a square-like response up to the corresponding cut-off angle. In addition, Eq. 1 does not include effects on the calculation of θ_c due to light diffraction in the filter holes. In our work, as we will show later, there are instances where we observed residual reflection and diffraction effects through the ASF holes. However, the angular dependence of the ASF holes and the capability of the device to discriminate particle size and concentration are preserved. If necessary, light diffraction in the filter holes can be strongly reduced by increasing D and L proportionally, i.e., still maintaining the same θ_c, as the typical diffraction angle is inversely proportional to D.

The larger the particle size is, the smaller the scattering angle is. Thus, a smaller θ_c is required, which means a larger L/D ratio. For example, to measure particles of hundreds of microns, we estimate that a minimum L/D ratio of 200 is required. For a typical length of several mm, this would mean a maximum D on the order of 50 µm. Making hole apertures with such dimensions and length is very challenging, even for the latest generation of 3D micro-printers using layer-by-layer fabrication. Other sophisticated techniques, such as mask-less photolithography, offer submicron resolution but cannot produce features with such high L/D. Additive manufacturing with micro-machining, e.g., laser sintering, selective laser melting and laser drilling, may achieve high L/D with micron resolution, but they impose significant constraints on the ASF, such as the combination of multiple pieces requiring tight alignment tolerances.

In this study, an interesting approach to overcome these fabrication hurdles and produce a highly optimised ASF including large arrays of holes with high L/D was to use a polymer extrusion technique. Such techniques have been widely used in fabricating micro-structured polymer optical fibres (mPOFs)[25], for example. To construct the ASF, a micro-structured cane was fabricated using a drill-and-draw technique from a commercially available poly (methyl methacrylate) (PMMA) rod from Nordisk Plast. A cane preform was prepared by machining the rod to 60 mm in diameter and 100 mm in length, which was followed by drilling the desired hole patterns. The preform was then annealed for a week at 80 °C and drawn to canes of 5 mm in diameter and 50 mm in length. A complete description of the experimental methodologies involved in the drill-and-draw technique can be found in[26]. This method of fabricating the ASF allows high flexibility in design since both D and L for the holes can be easily adjusted to collect scattering angles required for specific applications.

The fabricated ASF used in this work consists of 23 holes with diameters ranging from 112 to 800 µm. The length is selected to be 17 mm so that the PSA incorporating such ASF can measure scattering angles from 0.38 to 2.7°. However, for measuring particles in suspension, these angles need to be corrected because the rays from the particles undergo refraction at the flow cell wall, i.e., water-glass and glass-air interfaces. The relation between the detected (θ_c) and the actual (θ) scattering angles is given by:

$$\sin\theta = \frac{\sin\theta_c}{n_w} \tag{2}$$

where n_w is the refractive index of the water, giving θ from 0.29 to 2.02°. Using Mie theory, we can approximate this angular range of the current ASF to be suitable for measuring particles from approximately 10 to 125 µm. A

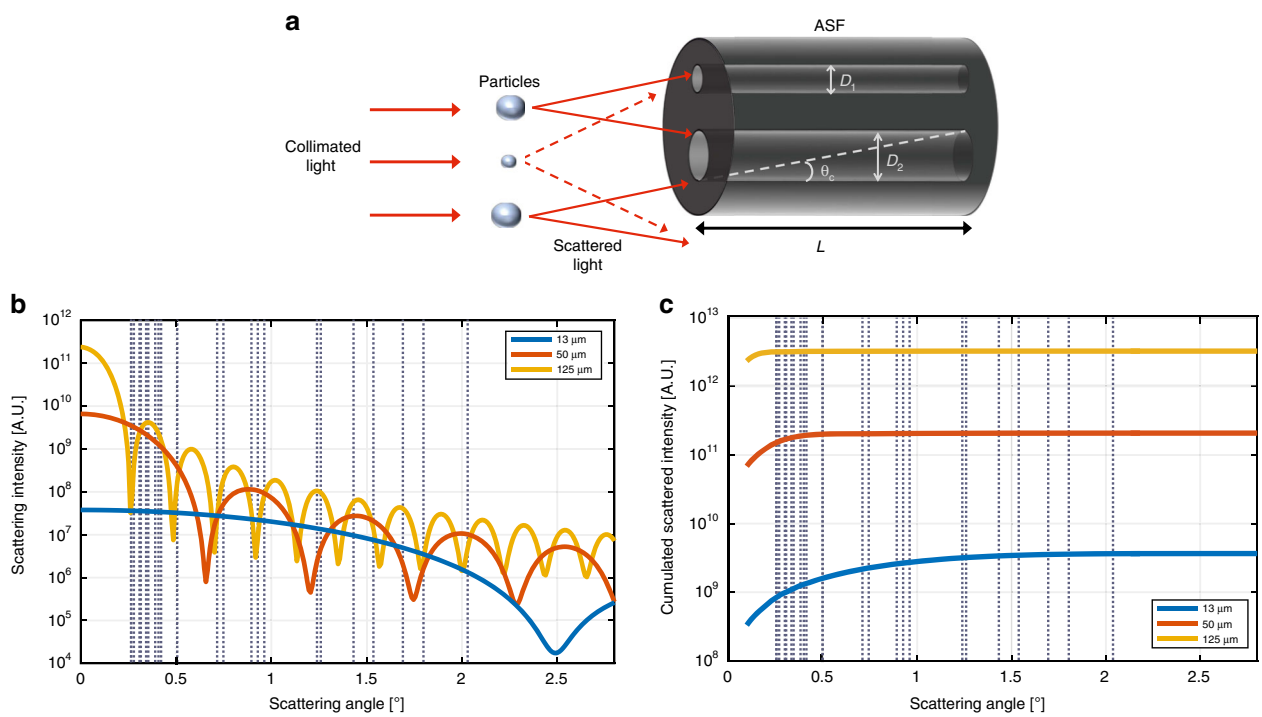

Fig. 1 Concept of the new particle size analyser. **a** Schematic diagram of the ASF showing how the cut-off angle, θ_c, is dependent on the diameter (D) and the length (L) of the holes. Light rays scattered from particles entering at angles larger than θ_c will be absorbed by the sidewalls. **b** The angular scattering profiles in water for three different glass beads of diameters 13, 50 and 125 µm with refractive index of 1.51 at a wavelength of 632.8 nm, simulated using the Mie algorithm[30] in MATLAB. **c** Cumulative scattering intensity for the three particle sizes. Instead of sampling the scattering profile at each angle, the ASF apertures perform a cumulative scattering power measurement from zero to a predetermined θ_c. The corresponding θ_c for each ASF hole for $L = 17$ mm, derived from Eq. 1 and converted to that in water using Eq. 2, is indicated by dashed vertical lines in **b** and **c**. We plot here results for single-particle scattering, but similar working principle description can be applied to the multiple-particle case

smaller and larger hole diameter-to-length ratio is required to measure particles above and below this range, respectively. Note that for very small particles (i.e., below 10 µm), the signal-to-noise ratio becomes a limiting factor due to the weak scattering signal intensity. A more sensitive image sensor array, such as a commercially available single-photon camera, can be used to improve the measurement at low scattering intensity. The implementation of such a camera will be a topic of further study.

To account for the multiple scattering effect at high concentrations that causes widening of the scattering lobe, we polish one side of the ASF along the entire length. This process leaves an empty space inside the holder, which acts as a large aperture. Such an aperture allows the entire angular spectrum of the forward scattered light to be collected from the sample.

The mPOF polymer used for fabricating the ASF is only partly absorbing in the working wavelength range in the visible spectrum. The inner walls of the ASF are thus covered with a black acrylic paint to increase their absorption and reduce reflection and crosstalk between adjacent holes.

Design of the PSA using the ASF

Figure 2a depicts the schematic diagram of the proposed PSA design based on the ASF. A fibre-coupled and collimated red LED—with a wavelength of 632.8 nm—is used as the light source. A 10 mm beam illuminates the sample containing particles dispersed in water. The scattered and unscattered light from the sample is collected by the ASF and the holder attached to the CMOS image sensor array. Additional details on the CMOS can be found in the Materials and Methods section.

All the data analysis in this work is conducted using MATLAB and Python. A typical raw image from the CMOS image sensor is shown in Fig. 2b and the fabricated ASF in Fig. 2c. The corresponding lab prototype is shown in Fig. 2d.

Measurements of particle suspensions

The experiments using the proposed PSA were carried out with samples listed in Table 1 for concentrations ranging from 1 to 40 mg ml^{-1}. For the smallest particle size range, i.e., 13–20 µm, the highest concentration that could be measured was 10 mg ml^{-1}, where above this concentration the light intensity reaching the CMOS

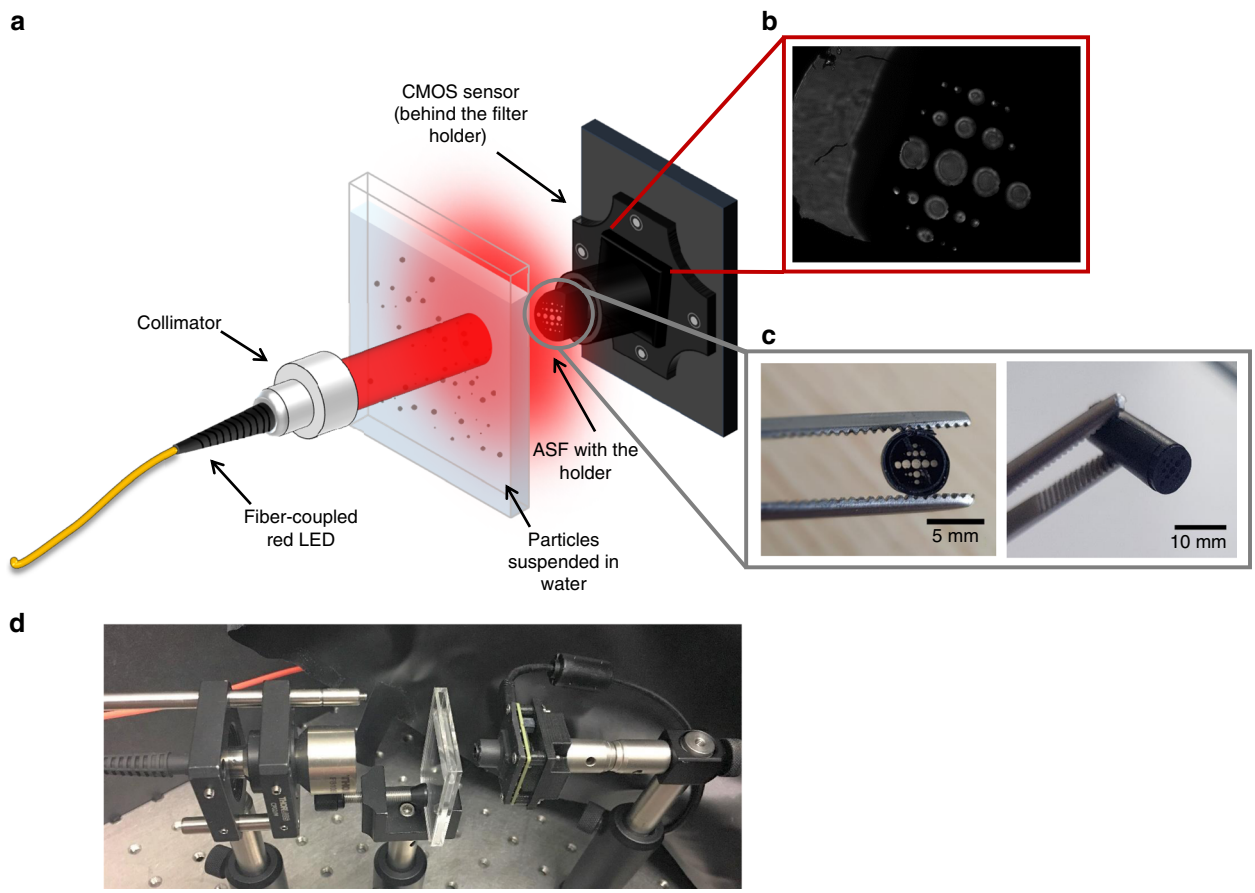

Fig. 2 Design of the proposed PSA using the ASF. **a** Schematic diagram of the PSA with a novel ASF that allows angle-resolved forward scattering measurements, in combination with a CMOS image sensor array and a collimated LED source, **b** An example raw image of sample with a volume median diameter of 44 μm at a concentration of 15 mg ml^{-1} obtained from the CMOS image sensor array. **c** Photograph of the fabricated ASF and **d** laboratory prototype showing the compactness of the proposed PSA

Table 1 Sample characteristics and concentrations measured.

Sample	Density (gcm^{-3})	Refractive index @ $\lambda = 632.8$ nm	Size range (μm)	Commercial LD PSA (HELOS/KR-H2487)			Concentrations measured (mg ml^{-1})
				D10 (μm)	D50 (μm)	D90 (μm)	
Guyson	2.5	1.51	80	55	74	92	1,5,10,15,20,25,30,40,50
			40	24	39	56	1,5,10,15,16,18,20,25,30
Cp5000	2.56	1.51	13–20	6	11.9	21	1,2,3,4,5,6,7,8,9,10
Sovitec	2.46	1.51	0–50	18	34.8	51	1,5,10,15,18,20,22,25,30,40
			40–50	33	43.6	51	
			40–70	46	62.3	80	
			70–110	68	87.5	108	
			90–150	97	125.5	157	

image sensor array becomes too low and would require longer integration times to achieve reliable results. At the beginning of each sample measurement, 200 ml of water was circulated through the flow cell, and a set of five images was obtained with a time gap of between 20 and 60 s. For each concentration, a sample suspension was

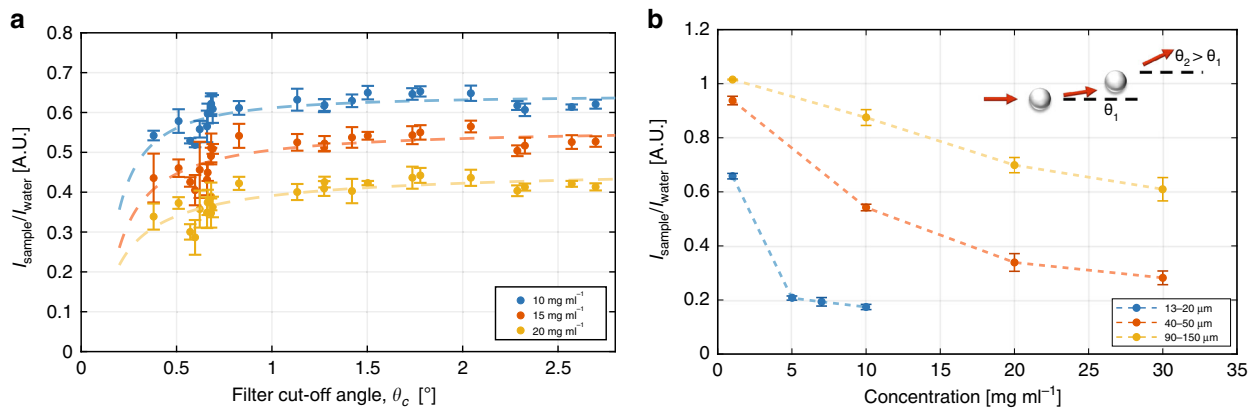

Fig. 3 Measurements performed with glass beads at different concentrations. **a** The average intensities normalised to water of the filter holes for glass beads with a 40–50 µm diameter distribution are plotted as a function of filter cut-off angles (θ_c)—calculated from the holes' diameters and length of ASF using Eq. 1—for three different concentrations. The error bars represent the 95% confidence interval. The dashed lines guiding the eyes represent a least square fit. **b** The average intensity normalised to water of the 112 µm diameter hole against concentration for three different glass bead diameter distributions, 13–20, 40–50 and 90–150 µm. The dependence on concentration, increasing for smaller glass beads, is a signature of multiple scattering

added from the stock solution (100 mg ml^{-1} concentration) to the water, and images were captured. The flow cell was cleaned with deionized water prior to measuring each new sample. A schematic diagram of the experimental setup is shown in Supplementary Fig. S1, and the raw images obtained from the CMOS sensor of the samples measured at a certain concentration are shown in Supplementary Fig. S2.

The light distribution between the ASF holes depends on the concentration. In Fig. 3a, we show this dependence for glass beads with a 40–50 µm size distribution. The intensity values plotted are the average of the five images calculated using the ¨regionprops¨ function in MATLAB. For the same concentration, smaller particles present a larger effect on the measured intensity and its dependence on the scattering angle (see Figure S3). This phenomenon can be explained in terms of the multiple scattering effect, where particles undergo several scattering events before reaching the CMOS image sensor array[14–16]. The result is the widening of the scattering angle and hence a decrease in the forward scattering intensity. This finding is also confirmed by Fig. 3b, where the average intensity of a small hole for three different particle size distributions is plotted against particle concentration.

Particle size prediction using a machine learning algorithm

While a high concentration leads to multiple scattering effects, an excessively low concentration leads to a poor signal-to-noise ratio. Therefore, a certain working concentration range must be defined for different particle size distributions. To avoid this concentration dependence and facilitate a wide working concentration range, we developed a machine learning algorithm using a random forest model, as explained within the Materials and Methods, to predict D50 from a given image and concentration value.

The image processing and the machine learning steps are summarised as a flowchart in Fig. 4. The mean and the standard deviation as a function of data points for one set of measurements were first monitored. After 100 repeats, no significant improvement in the predicted error was observed. Hence, the model was trained and tested 100 times. The mean of 100 mean absolute percentage error (MAPE) on the test sets was found to be 2.52%, with a standard deviation of 0.73%. The performance of the model on only one of these test sets is depicted in Fig. 5a, b. The random forest model can therefore correct the dependence on particle concentration that leads to the multiple scattering effect (Fig. 5b) and predict the particle size with high accuracy (Fig. 5a).

So far, we tested Model 1 to predict D50 using concentration as one of the inputs. Since in practice, the particle size should be provided as an independent parameter from concentration, we tested Model 2, which relies only on filter sizes and intensities to predict D50. Upon testing the model, the MAPE was found to be 5.09 ± 1.56%. The predicted D50 vs nominal D50 and D50 vs concentration plots are given in Fig. 5c, d, respectively. As expected, Model 1 has a higher precision than Model 2 when concentration information is given as input, but the prediction error without concentration is still acceptable.

We also performed separate training of the model using intensity values from the large hole only for all the particle sizes measured (Supplementary Fig. S4). Note that the big hole analysis is performed on the same images as the ASF holes. The large hole intensity includes the entire angular

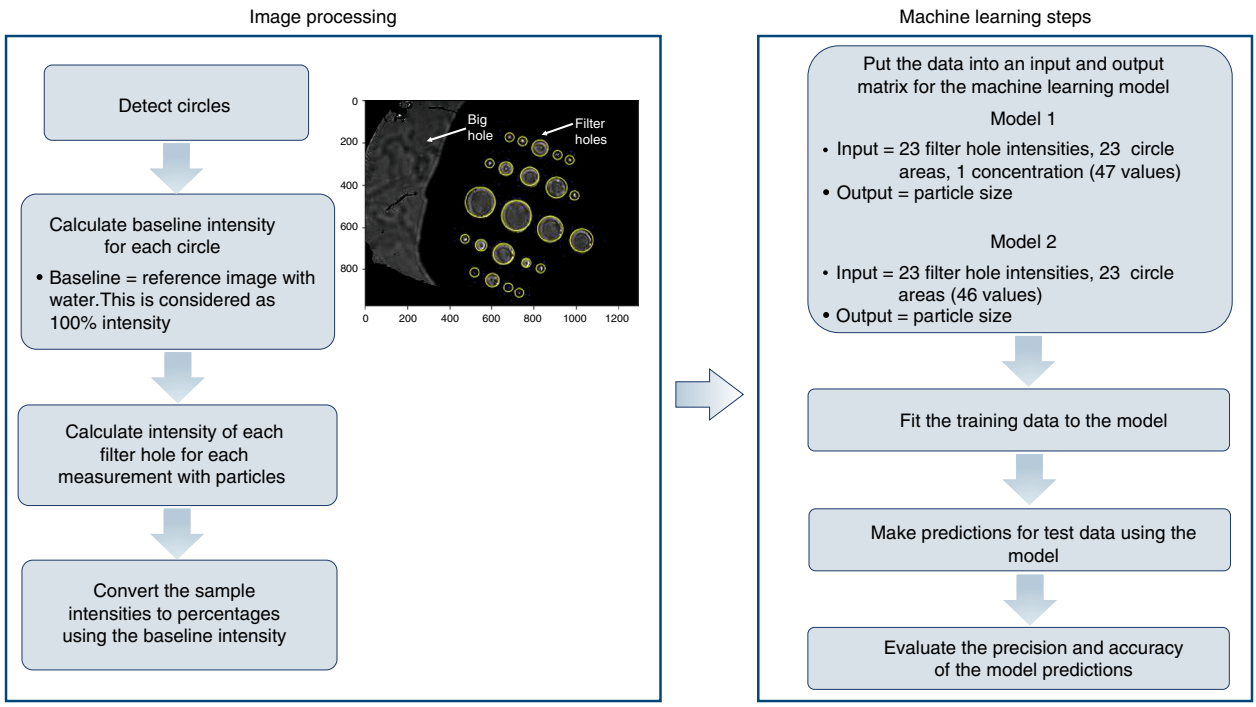

Fig. 4 Flowchart of the particle size detection algorithm using machine learning

spectrum for the scattered light and scales as the absorption due to the particle solution. The model was tested 100 times with different test sets. The mean predictions were found to deviate significantly from the nominal values with varying concentrations, and the mean error was increased to 23.03 ± 5.61%. This result confirms that absorption analysis is not enough and that scattering and the ASF play crucial roles in predicting particle size with high accuracy using the random forest model. In addition, the big hole analysis indicates that there are no hidden correlations between the different images, other than those related to the particles (e.g., size, concentration) on which scattering depends. If there were hidden correlations, then the MAPE for the large hole would not have given such large errors compared to those obtained with the ASF analysis.

In particle size analysis, D50 is certainly one of the most important parameters. However, some applications require knowledge of the distribution, as given by the D10 and D90 parameters, which correspond to fine and coarse particles, respectively, in the sample. The same machine learning algorithm can also be trained to predict additional percentile values for the volume median diameter, e.g., D5, D10, D15 to D95, without the need to modify the experimental set-up. We performed a trial training of the random forest model with the D10, D50 and D90 values measured using a commercial LD PSA; on testing the model, the MAPE was found to be 4.27 ± 1.64%, 3.02 ± 1.07% and 2.4 ± 0.8%, respectively. Even though the results are quite promising with only one set of D10 and D90 data for each size, they can be further improved by measuring samples with the same D50 but varying distribution spread. Future development will include experiments with different refractive indices and different size range particles.

In addition to the above-mentioned batch measurements, we performed a test flow-through measurement (described in supplementary information) to demonstrate the capability of our ML model for such measurements. We collected data with two samples, 13-20 µm and 40-70 µm, for different concentrations and calibrated our previous model with these data. We then tested the model on a new set of data for the same samples collected on a separate day. The MAPE for Model 1 was found to be 1.77 ± 0.25% (Supplementary Fig. S5). Though only two samples were measured, this preliminary result suggests that our system can be used to predict the change in particle size for different samples. With further optimisation, the flow measurement procedure and performance can be improved, and the accuracy can be increased.

We also note that larger deviations from the nominal value are observed for the Guyson beads (D50: 39 and 74 µm). Microscope images (see Supplementary Fig. S6a and b) of these beads reveal the presence of some non-spherical particles, the shape of which has an influence on their scattering pattern. Supplementary Fig. S6c and d

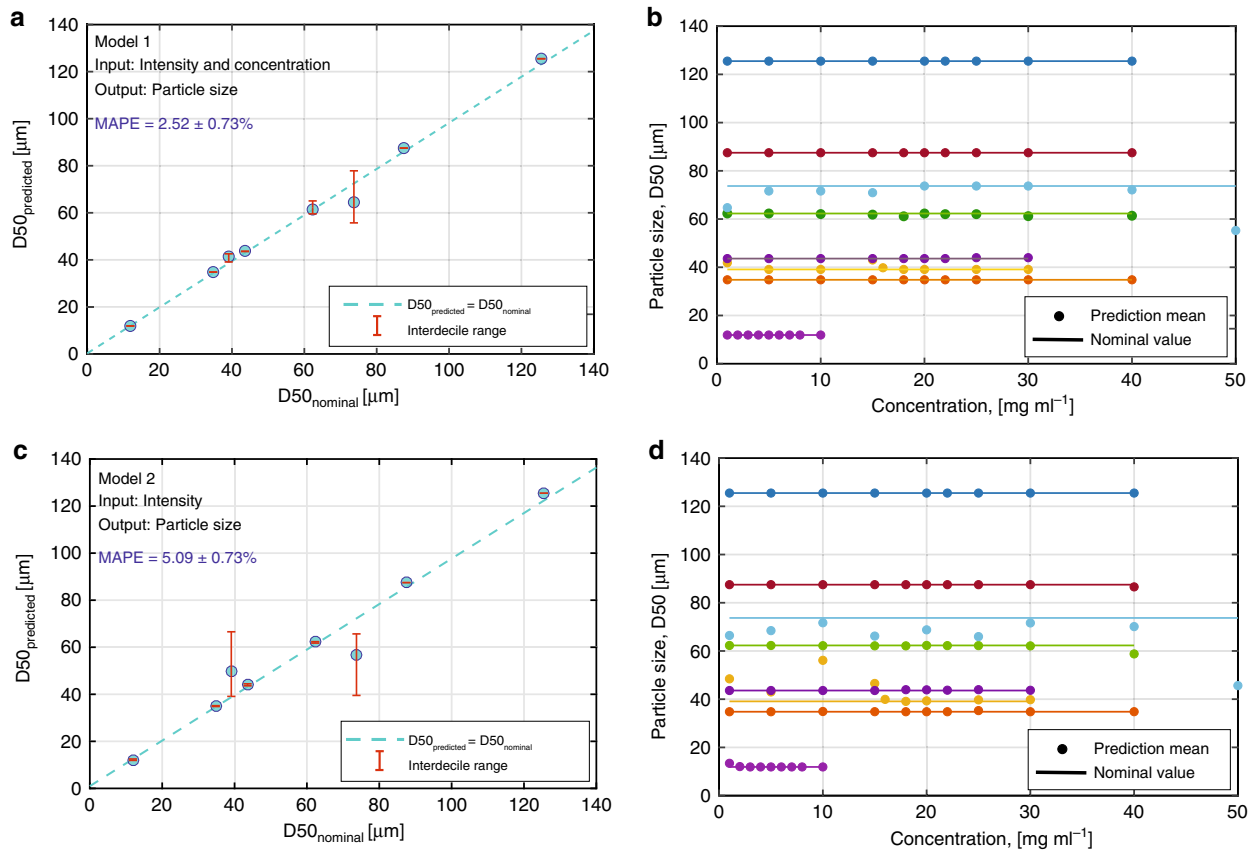

Fig. 5 Performance of the machine learning algorithm in the prediction of particle size (glass bead diameter) when trained and tested with images obtained from the CMOS image sensor array. Two models are used for training and testing purposes. Model 1 used the intensity and diameter of the 23 holes together with the concentration information for training, whereas Model 2 used only the intensity and diameter (46 features) for inference. **a** The mean predicted D50 values using Model 1 for one of the test sets are compared to the nominal D50 values measured using a commercial LD PSA (HELOS/KR-H2487). The dashed line represents predicted diameter = nominal diameter. The interdecile range is also shown for each predicted D50. **b** The D50 prediction values from Model 1 are plotted against concentration. Despite the multiple scattering effects that produce a strong dependence on concentration, the predicted diameters are close to the nominal diameters (straight lines). **c** The mean predicted D50 against nominal diameter using Model 2 and **d** D50 prediction against concentration using Model 2. When only spherical particles were analysed, the error for Model 2 was significantly reduced from 5.09 to 0.72% (see Supplementary Fig. S6)

shows the performance of Model 2 for all glass beads except the Guyson beads and for only the Guyson beads, respectively, confirming that the MAPE is strongly increased by the Guyson beads. By removing the particles with a non-spherical shape from the sample analysis, the MAPE becomes much smaller (0.72%). Therefore, our device performs according to ISO13320[27], which requires that for polydisperse spherical particles, the measured D50 should be within 2.5% of the quoted maximum or minimum values for the reference materials. In future work, by collecting more data, including on non-spherical particles, one can expect that the precision of the device will increase further.

Discussion

In this work, we proposed a novel design of a compact, portable and cost-effective particle size analyser (PSA) in a collimated beam configuration using a CMOS image sensor and machine learning. Unlike commercially available counterparts, such as laser diffraction-based systems that use several detectors to measure the scattering signature of particles, the proposed PSA uses an innovative design including a novel, very small angular spatial filter. The ASF combined with an LED and a CMOS image sensor array allows the acquisition of angle-dependent scattering images that are used by an ML model to predict the median diameter of particles.

The proposed PSA was validated by measuring glass beads of various size distributions at different concentrations. The results obtained from the ML model showed that, given the particle concentration, the median particle size could be measured, with a low mean absolute percentage error of 2.5%, even in the presence of significant multiple scattering. When the concentration is

not an input parameter, this error increases to 5%. However, by removing samples with non-spherical particles, we achieve a MAPE of 0.72% for Model 2, i.e., without predefining concentration as an input parameter. These performances compare well with those of commercially available laser diffraction-based counterparts, with reported device accuracies for monomodal latex standards of approximately 0.6%.

While future improvements in the optical hardware and a larger quantity of data for the ML algorithm, including non-spherical particles collected with well-designed sample feeding systems for dry and wet measurements, will lead to higher precision, we intend to utilise the inherent flexibility of the simple design and low hardware cost of our proposed PSA for incorporation in online or at-line applications. For online operations, such instrumentation is mostly used for quality assurance (QA) and control purposes, which are often focused on measuring system changes rather than necessarily exact values. We have shown such an example of a real-time change response with the proof-of-concept measurements in a flow cell system using our proposed PSA. As such, the performance is a trade-off between the hardware cost and the required level of accuracy, where the exact error limits will likely not be as low as those for commercial *ex situ* PSAs. The additional benefit of online analysis is the lower degree of sample intrusion, and thus, for many particle processes, measurements are often actually more representative of the system, even if the absolute instrument error is higher.

Therefore, we believe our proposed PSA is an attractive solution for online monitoring of particles in different industrial processes without the need to perform dilution operations. We also note that our proposed PSA device is sensitive to the refractive index difference between the particles and the surrounding medium. In principle, the system may thus also be used in relevant biological applications, for example, in the detection of microorganisms in water, such as *Escherichia coli* and *Legionella*, and red cells in blood.

Materials and methods

Glass bead characterisation

To test the functionality of the newly designed PSA, we measured various size distributions of glass beads at different concentrations, which are summarised in Table 1.

The sample suspensions in water at each concentration are measured using a commercial LD PSA (Model HELOS/KR-H2487, Sympatec GmbH, Clausthal-Zellerfeld, Germany) for angular ranges below 35° (i.e., forward scattering). In this work, an angular range of 0.1° to 9° is used since this range is sensitive to particle sizes from 0.5 μm to 175 μm[28]. The volume median diameter D50 together with the volume-weighted 10th and 90th percentiles, D10 and D90, respectively, of the particle distributions for each sample measured with HELOS/KR-H2487 are also listed in Table 1.

Electron micrographs of the particles dispersed in water are also taken using a scanning electron microscope (SEM), model DSM 982 Gemini (Zeiss/Germany). The SEM is a low-voltage electron microscope (30 kV) with a maximum magnification of 200,000, i.e., a resolution of approximately 10 nm. The instrument detects both types of electrons, namely, backscattered primary and backscattered secondary electrons, and therefore can provide high sizing accuracy and three-dimensional impression.

The particle size distributions obtained from the HELOS/KR-H2487 system together with the SEM images of the glass beads used in the experiments are shown in Supplementary Fig. S7.

Particle suspension preparation

For each size range to be measured, a known mass of powder samples is taken and dispersed in a known volume of deionized water to make a suspension. An overhead stirrer at 300 rpm is used to prevent agglomeration or deposition of the particles in the beaker containing the suspension. The suspension is then circulated into a flow cell (component of HELOS-KR-SUCELL—Sympatec GmbH, Clausthal-Zellerfeld, Germany; measurement volume ~ 6 ml) with a path length of 4 mm using a peristaltic pump. The pressure of the pump is controlled to prevent air bubble formation while maintaining a homogeneous flow of suspension in the measurement cell.

Image acquisition and processing

The CMOS image sensor array used to capture the images is a Micron MT9P0311, and the images are displayed using DevWare software. The active area of the image sensor array is $5.7 \times 4.28\ \text{mm} = 24.4\ \text{mm}^2$, which is also the field-of-view of the proposed PSA when the ASF is in close proximity to the flow cell; it consists of 2592 × 1944 pixels, each of which has a size of 2.2 × 2.2 μm. The array has four colour channels, of which only red is used for data processing in our experiments. The frame rate to obtain a full-resolution image is 14 frames per second (fps).

Prior to each measurement, a dark image in the absence of LED illumination is captured and subsequently subtracted from the sample images. For each particle size range, first, a reference image with only water is measured, and then, images of suspensions at different concentrations are measured.

Machine learning algorithm

The data from the sensor include an image per test condition (concentration and standard particle size). The

following procedure was developed to correlate the images to the median volume particle size D50 used in each test. First, the location of the 23 filter holes is established using an image processing library (scikit-image, blob detection) in Python. The pixel intensities and diameters are calculated for each hole. The pixel intensities are then converted to relative intensities, as percentages, using the reference images to which pixel intensities of 100% are assigned. A set of five images is obtained for each combination of concentration and particle size, and the average of these images corresponds to a single data point: (a) the known concentration and D50 and (b) the relative intensities and diameters for 23 holes. Second, the dataset comprises 459 images that are randomly partitioned into two sets: (a) the training set (344 images) and (b) the test set (115 images).

Third, the random forest algorithm is used to find the correlation between the input variables (the concentration, the 23 relative intensities and 23 hole diameters) and the output variable (D50). Among the different ML algorithms available[29], we chose the random forest because it is suitable for structured data, as in our case. We have also made a preliminary comparison between gradient boosting and random forest and confirmed that the latter provides slightly better predictions for the number of data points used in the analysis. The random forest consists of multiple decision trees. Each tree is a tree-like model of decisions. Each decision (splitting of the data) uses one feature and its threshold value. Learning (i.e., training) includes choosing the features, threshold values and when to stop the tree. The model is developed using a scikit learn machine learning library, with the hyper-parameters given in Supplementary Table S1.

Fourth, the generation of the training/test sets is a random process; therefore, model performance can change from split to split. To handle this fluctuation, steps 2 and 3 are repeated 100 times. MAPE of model predictions on the test set is used to assess the performance of the model. It is defined as:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=0}^{n}\left|\frac{\text{Actual value}_i - \text{Predicted value}_i}{\text{Actual value}_i}\right| \quad (3)$$

where n is the number of images in the test set.

The mean MAPE of 100 models and their standard deviations are reported as the final figure of merit. To visualise the predictions of one of the models on the test set, predictions are plotted per particle size.

Until now, we developed a model that predicts D50 using concentration as one of the inputs. We refer to this as Model 1. For a truly functional sensor, however, predicting D50 only from intensity, without any input concentration, is essential. Therefore, we developed another random forest model, Model 2, which uses only filter sizes and intensities (i.e., 23 hole diameters and 23 intensity values) to predict D50.

Acknowledgements

This work is funded by the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 637232 (ProPAT project). R.H. and V.P. acknowledge financial support from the Spanish Ministry of Economy and Competitiveness through the 'Severo Ochoa' Programme for Centres of Excellence in R&D (SEV-2015-0522), from Fundació Privada Cellex, and from Generalitat de Catalunya through the CERCA programme, from AGAUR 2017 SGR 1634. V.P. acknowledges financial support from the Spanish Ministry of Economy and Competitiveness through the project OPTO-SCREEN (TEC2016-75080-R). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665884. The authors acknowledge the Chemometrics group at the Universitat de Barcelona, especially Adrián Gómez Sánchez and Rodrigo Rocha de Oliveira, for their contribution in the helpful discussions on measurement optimisation and background correction.

Author details

[1]ICFO- Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain. [2]Ipsumio B.V., High Tech Campus, 5656 AE, Eindhoven, Netherlands. [3]Department of Photonics Engineering, Technical University of Denmark, DK-2800 Kgs Lyngby, Denmark. [4]Research Group Mechanical Process Engineering, Institute of Process Engineering and Environmental Technology, Technische Universität Dresden, Münchner Platz 3, D-01062 Dresden, Germany. [5]School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, UK. [6]IRIS Technology Solutions, SL, 08860 Castelldefels (Barcelona), Spain. [7]ICREA- Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain

Author contributions

V.P. proposed the device and experiments. G.W. fabricated the angular spatial filter designed by R.H. P.A.M. and T.C. built the prototype. R.R.R.M. and M.S. contributed to the development of the sample preparation procedures. M.A.N. developed the machine learning models. R.H., P.A.M., R.R.R.M., F.M, and T.H. performed measurements with glass beads. R.H. and V.P. wrote the manuscript with contributions from all other authors. All authors contributed to the data analysis and interpretation of the results.

Conflict of interest

P.A.M. and V.P. are inventors of US Patent US9857300B2, Apparatus for measuring light scattering (2018). The remaining authors declare that they have no conflict of interest.

Supplementary information is available for this paper at https://doi.org/10.1038/s41377-020-0255-6.

Received: 6 June 2019 Revised: 9 January 2020 Accepted: 27 January 2020
Published online: 12 February 2020

References

1. Valsangkar, A. J. Principles, methods and applications of particle size analysis. *Can. Geotech. J.* **29**, 1006 (1992).
2. Shekunov, B. Y. et al. Particle size analysis in pharmaceutics: principles, methods and applications. *Pharm. Res.* **24**, 203–227 (2007).
3. Servais, C., Jones, R. & Roberts, I. The influence of particle size distribution on the processing of food. *J. Food Eng.* **51**, 201–208 (2002).
4. Stetefeld, J., McKenna, S. A. & Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophysical Rev.* **8**, 409–427 (2016).
5. Kim, A. et al. Validation of size estimation of nanoparticle tracking analysis on polydisperse macromolecule assembly. *Sci. Rep.* **9**, 2639 (2019).
6. Kim, A., Bernt, W. & Cho, N. J. Improved size determination by nanoparticle tracking analysis: influence of recognition radius. *Anal. Chem.* **91**, 9508–9515 (2019).
7. Blott, S. J. et al. Particle size analysis by laser diffraction. Geological Society, London, Special Publications. **232**, 63–73 (2004).

8. Xu, R. L. Light scattering: a review of particle characterization applications. *Particuology* **18**, 11–21 (2015).
9. Bux, J. et al. Measurement and density normalisation of acoustic attenuation and backscattering constants of arbitrary suspensions within the Rayleigh scattering regime. *Appl. Acoust.* **146**, 9–22 (2019).
10. Povey, M. J. W. Ultrasound particle sizing: a review. *Particuology* **11**, 135–147 (2013).
11. Vargas-Ubera, J., Aguilar, J. F. & Gale, D. M. Reconstruction of particle-size distributions from light-scattering patterns using three inversion methods. *Appl. Opt.* **46**, 124–132 (2007).
12. Ye, Z. & Jiang, X. P. Wang, Z. C. Measurements of particle size distribution based on Mie scattering theory and Markov chain inversion algorithm. *J. Softw.* **7**, 2309–2316 (2012).
13. Mishchenko, M. I., Travis, L. D. & Lacis, A. A. *Multiple Scattering of Light by Particles: Radiative Transfer and Coherent Backscattering*. (Cambridge University Press, Cambridge, 2006).
14. Gomi, H. Multiple scattering correction in the measurement of particle size and number density by the diffraction method. *Appl. Opt.* **25**, 3552–3558 (1986).
15. Quirantes, A., Arroyo, F. & Quirantes-Ros, J. Multiple light scattering by spherical particle systems and its dependence on concentration: a T-matrix study. *J. Colloid Interface Sci.* **240**, 78–82 (2001).
16. Wei, Y. H., Shen, J. Q. & Yu, H. T. Numerical calculation of multiple scattering with the layer model. *Particuology* **7**, 76–82 (2009).
17. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
18. Nascimento, C. A. O., Guardani, R. & Giulietti, M. Use of neural networks in the analysis of particle size distributions by laser diffraction. *Powder Technol.* **90**, 89–94 (1997).
19. Guardani, R., Nascimento, C. A. O. & Onimaru, R. S. Use of neural networks in the analysis of particle size distribution by laser diffraction: tests with different particle systems. *Powder Technol.* **126**, 42–50 (2002).
20. Wu, Y. C. et al. Air quality monitoring using mobile microscopy and machine learning. *Light Sci. Appl.* **6**, e17046, https://doi.org/10.1038/lsa.2017.46 (2017).
21. Roy, M. et al. Low-cost telemedicine device performing cell and particle size measurement based on lens-free shadow imaging technology. *Biosens. Bioelectron.* **67**, 715–723 (2015).
22. Seo, S. et al. High-throughput lens-free blood analysis on a chip. *Anal. Chem.* **82**, 4621–4627 (2010).
23. Cutler, A., Cutler, D. R. & Stevens, J. R. Random forests. In *Ensemble Machine Learning* (eds. Zhang, C. & Ma, Y. Q.) Ch. 5, 157–175 (Boston: Springer, 2012). https://doi.org/10.1007/978-1-4419-9326-7_5.
24. Pruneri, V., Martïnez Cordero, P. A. & Jofre Cruanyes, M. Apparatus for measuring light scattering. US Patent 9857300 (2018).
25. Barton, G. et al. Fabrication of microstructured polymer optical fibres. *Optical Fiber Technol.* **10**, 325–335 (2004).
26. Large, M. C. J. et al. *Microstructured Polymer Optical Fibres*. (Boston, Springer, 2008).
27. ISO 13320:2009 Particle size analysis-laser diffraction methods (2009).
28. Retamal Marín, R. R. et al. Effects of sample preparation on particle size distributions of different types of silica in suspensions. *Nanomaterials* **8**, 454 (2018).
29. Franklin, J. The elements of statistical learning: data mining, inference and prediction. *Math. Intell.* **27**, 83–85 (2005).
30. Bohren, C. F. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles*. (New York, John Wiley & Sons, 2008).