Integration of light scattering with machine learning for label free cell detection

Wendy Yu Wan,[1,4] **Lina Liu,**[1,4] **Xiaoxuan Liu,**[1] **Wei Wang,**[1] **Md. Zahurul Islam,**[1,2] **Chunhua Dong,**[3] **Craig R. Garen,**[3] 🆔 **Michael T. Woodside,**[3] **Manisha Gupta,**[1] **Mrinal Mandal,**[1] **Wojciech Rozmus,**[3] **and Ying Yin Tsui**[1,*]

[1]*Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada*
[2]*Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh*
[3]*Department of Physics, University of Alberta, Edmonton, AB, Canada*
[4]*Authors with equal contribution*
[*]*ytsui@ualberta.ca*

Abstract: Light scattering has been used for label-free cell detection. The angular light scattering patterns from the cells are unique to them based on the cell size, nucleus size, number of mitochondria, and cell surface roughness. The patterns collected from the cells can then be classified based on different image characteristics. We have also developed a machine learning (ML) method to classify these cell light scattering patterns. As a case study we have used this light scattering technique integrated with the machine learning to analyze staurosporine-treated SH-SY5Y neuroblastoma cells and compare them to non-treated control cells. Experimental results show that the ML technique can provide a classification accuracy (treated versus non-treated) of over 90%. The predicted percentage of the treated cells in a mixed solution is within 5% of the reference (ground-truth) value and the technique has the potential to be a viable method for real-time detection and diagnosis.

1. Introduction

Conventional bench-top flow cytometer, also known as the fluorescence-activated cell sorter (FACS), is a valuable tool for cell identification in many biological and health-related applications [1]. It is based on measurement of fluorescence of molecules that are attached to the illuminated target cells [1]. However, these fluorescent biomarkers can interfere with the function of the cells they bind to, hindering further potential analyses and complicating interpretation [2]. Additionally, adequate biomarkers are not available for many cell types, and they can be difficult or expensive to develop for organisms such as microbes and viruses [2]. Development of label-free techniques of cell identification is thus of high interest to the scientists to avoid such complications.

Light scattering has been studied as a label-free technique for single-cell analysis by several research groups [3–7]. In earlier studies in our group, we have used angular light scattering patterns as signatures for label-free cell identification [8–16]. This technique was applied to study yeast, human raji [10] and hematopoietic stem cells [14–16]. Single cells were identified by comparing the experimental 2D scattered light patterns measured using a charge-coupled device (CCD) camera with the numerical simulation results. In these studies, we used a speckle count technique for distinguishing the cells. The simulations were carried out by solving the Maxwell Equations using the Finite Difference Time Domain (FDTD) technique and simplified models for optical properties of the cells [8]. The cells have been defined in simulations as three-dimensional dielectrics of spherical or oval shapes, containing different cell organelles of varying indices of refraction [15,16]. In these papers we have also investigated the effect of different degrees of surface roughness. Numerical simulations identify the main scattering

centers in cells such as nuclei, small organelles, e.g. mitochondria or lysosomes, and reproduce the characteristic features of the experimental angular scattering spectra.

The cell light scattering patterns are dominated by the small scale 2D speckle patterns originating from light scattering from the mitochondria as they have higher refractive index than other organelles [10,14–16]. These patterns are different due to the variation in the shape, number and distribution of the mitochondria. Statistical analysis of the spatial distribution of the speckles in the scattered light patterns allows us to effectively distinguish one cell type from the other. The difference in these patterns in our previous works [14,15] can be considered a signature for use in cell identification. Two speckle properties - the number and the average area of speckles in the scattering pattern, can be used to discriminate scattering patterns among different types of blood cells. This technique was applied to distinguish between the cancerous blood cells with randomly distributed mitochondria and healthy blood cells with aggregated mitochondria [15]. Pattern recognition based on machine learning has been explored in several recent works [17–19]. The Adaptive boosting (AdaBoost) classifier for the discrimination of normal cervical cells and HeLa cells, yeast cell clusters with different numbers of cells was proposed [17]. The image gray level co-occurrence matrix (GLCM) features along with support vector machine (SVM) have been used to classify scattering patterns of prostate cancer cell structures [18]. Histogram of oriented gradients (HOG) has also been used to discriminate scattering patterns of ovarian cancer cells from normal cells [19].

The classification techniques mentioned above can be termed as 'hand-crafted' feature-based techniques, which are popular in traditional image classification. However, these techniques may not be very effective for scattering pattern analysis due to the lack of visibly distinguishable image characteristics. Superior results may be obtained if the features are learned from the images using techniques such as deep learning neural networks to perform image classification [20,21]. In this paper, we have utilized a convolutional neural network (CNN) based machine learning (ML) technique for the pattern classification. However, as the image database is very small, we use a pretrained CNN, known as DenseNet-201 [20], as the deep feature extractor. The DenseNet-201 takes the scattering images as inputs and generates deep feature vectors. The obtained deep feature vectors are then fed to an SVM for classification of the scattering patterns.

In this study, we have used the light scattering technique in combination with ML techniques to distinguish non-treated and staurosporine-treated SH-SY5Y neuroblastoma cells. The SH-SY5Y neuroblastoma cells are used in neuroscience for better understanding of human neurological disorders such as Parkinson's disease [22–25]. Staurosporine-treated SH-SY5Y neuroblastoma cells are used in the fundamental studies of apoptosis [26,27] and neuroprotection studies [28–30]. Experimentally, we have used green and red lasers to study the light scattering patterns from these cells. The light scattering patterns were collected at different angles to obtain more information. Mie and FDTD simulations of light scattering were carried out for these cells to validate the experimental results. In addition to these cells, polystyrene beads of known refractive indices and sizes are used in experiments and simulation to obtain reference light scattering patterns to distinguish the treated and non-treated cells. The ML techniques were then applied to these experimental patterns distinguish the treated and non-treated cells.

2. Methods

Figure 1 shows the schematic diagram of experimental setup for collecting scattered laser light patterns from beads and cells. The setup is similar to our experiments reported previously [16] but with the additional capability of measurement of scattered light in three different directions. Key components of the system include a probing laser, a sample holder, and CCD cameras with microscope objectives. The beads/cells were irradiated with one of two wavelengths individually (632.8 nm red light from a He-Ne laser or 532 nm green light from a second-harmonic diode-pumped solid-state laser). A plano-convex lens with a focal length of 5 cm was used to focus each

laser beam to a diameter of approximately 0.1 mm at the focal position. Three CCD cameras were placed in forward, side and backward directions, respectively. Microscope objectives with long working distance (10x Mitutoyo infinity-corrected long working distance objective with a numerical aperture (NA) of 0.28) were used with the forward and backward CCD cameras, while a normal microscope objective (10x microscope objective with a NA of 0.25) was used with the CCD camera placed in the side direction. Each microscope objective and CCD camera were connected by a tube and the system was placed on three-directional translation stages. The central lines of the forward, side and backward microscope objectives were 30°, 90° and 150° away from laser beam direction, respectively. The ranges of light collection angles were between 18° and 42° in the forward direction, between 79° and 101° in the side direction, and, between 141° and 159° in the backward direction, with respect to the laser beam direction (z direction). The overlap of the laser beam and the observation region of each CCD camera with microscope objectives defines a small measurement volume of approximately 0.002 mm^3 [31]. Typically a 10 ml solution containing spherical polystyrene microbeads or SH-SY5Y cells with a concentration of approximately 3000 particles/ml was prepared for each experiment. The micro-particles move freely inside the solution due to Brownian motion and several micro-particles in a minute enter the measurement volume, leading to their 2D light scattered patterns being recorded by the CCD cameras.
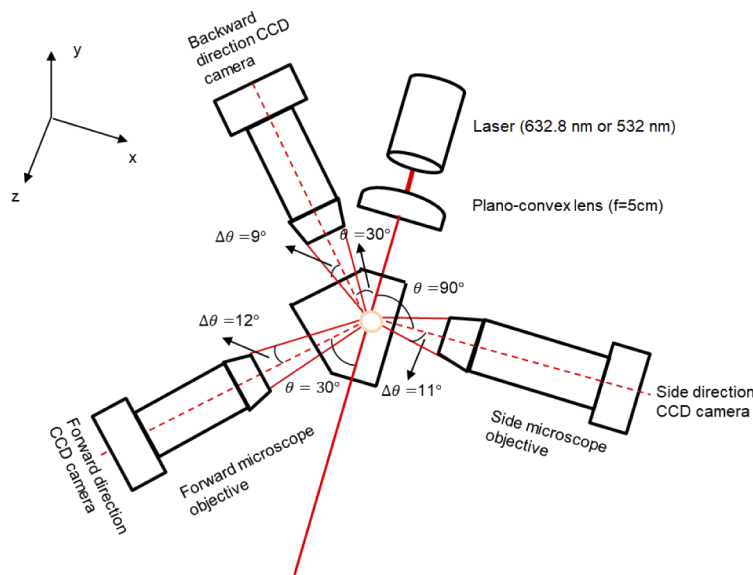
Fig. 1. A schematic diagram of the experimental setup used to obtain two-dimensional light scattered patterns of spherical beads and cells in three directions. The probing laser and the side, forward and backward CCD cameras that collect scattered light are shown in the schematic.

A multimode fiber with diameter of 125 μm was used for optics alignment. The fiber was placed in the sample holder and a focused fiber image was first obtained in all three CCD cameras simultaneously, and then a defocused image was obtained in the CCD cameras when each microscope objective was moved several hundred micrometers away from the sample holder. With the setup, two-dimensional scattered light patterns of spherical beads and cells can be obtained in three directions with illumination with two different wavelengths. The CCD has 1294x964 pixels and each pixel has a size of 3.75 μm. A typical experimental scattering pattern

image is circular and has a diameter of approximately several hundred pixels, giving an angular resolution of approximately several minutes of arc.

Spherical polystyrene beads of two different sizes with diameters of 10 μm and 15 μm were used in the study. The concentration of the beads was reduced from around 10^7 particles/ml to around 3000 particles/ml using phosphate-buffered saline (PBS) so that light scattering patterns of single beads could be obtained. These initial measurements with plastic microbeads of sizes similar to SH-SY5Y cells were helpful in calibrating the collection optics and optimizing the concentration of cells. Our measurements rely on the thermal motions of cells for introducing different scattering centers in the focus of the laser probe and collection optics.

SH-SY5Y cells were from the American Type Culture Collection (ATCC, Manassas, Va, USA) and cultured using a 1:1 mixture of Eagle's Minimum Essential Medium supplemented with F12 Medium containing 10% v/v fetal bovine serum and penicillin/streptomycin mixture. Cultures were seeded at 20% cell density and then allowed to grow for approximately 48 h until they had reached 70% confluence of adherent cells. At this point, staurosporine was added to a final concentration of 5 μM to make the treated SH-SY5Y cells group, while an equivalent volume of PBS was added to the control SH-SY5Y cells group without staurosporine. Both groups were allowed to incubate for 48 h, and then cells were fixed with 10% p-formaldehyde for 15 minutes. Cells were sampled for each of the two conditions and were counted using a hemocytometer to determine cell density present for each aliquot. Undiluted samples had a concentration of approximately 10^6 to 10^7 cells/ml and were diluted to a final assay concentration of around 3000 cells/ml so that light scattering patterns of single cells could be obtained.

3.   Results and discussion with microbeads

3.1.   Measurements of scattered light from beads

Two-dimensional scattered light patterns of single spherical polystyrene microbeads have been successfully obtained experimentally in all three directions with laser lights of two different wavelengths. These basic measurements tested the laser probe and the detection system and identified the scatter patterns generated along with the number of fringes in the patterns corresponding to the diameter of the scattering beads, as shown in Fig. 2. Fringe patterns were observed for both 10 μm (top row) and 15 μm (bottom row) spherical beads in forward, side and backward directions, and for red (Fig. 2(a)) and green and (Fig. 2(b)) illuminations. Similar number of fringes was observed in all three directions for the beads of same diameter at each illumination. More fringes were observed with shorter (green) wavelength illumination for the
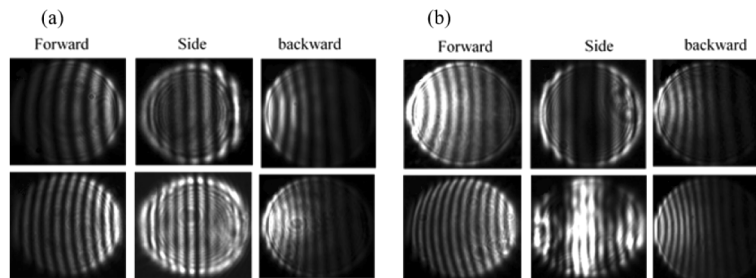

Fig. 2. 2D scattered light patterns with (a) red and (b) green laser illuminations for polystyrene beads with diameters of 10 μm (upper rows) and 15 μm (lower rows). The pictures in a row from left to right are the forward, side and backward scattered light patterns of a bead, respectively. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).

beads of same diameter. In addition, a greater number of fringes were observed with larger diameter (15 µm) beads.

3.2. Analysis of bead scattering measurement

Scatterings from spherical beads of known refractive index were performed to establish the connection between measured angular spectra and Mie theory and to verify the concentration of beads that allows scattering from a single bead at the time when the scattering object enters the focal volume of the optical system in Fig. 1. Note that in our basic setup we rely upon Brownian motion of beads (or cells) to obtain angular scattering spectra of statistically significant number of
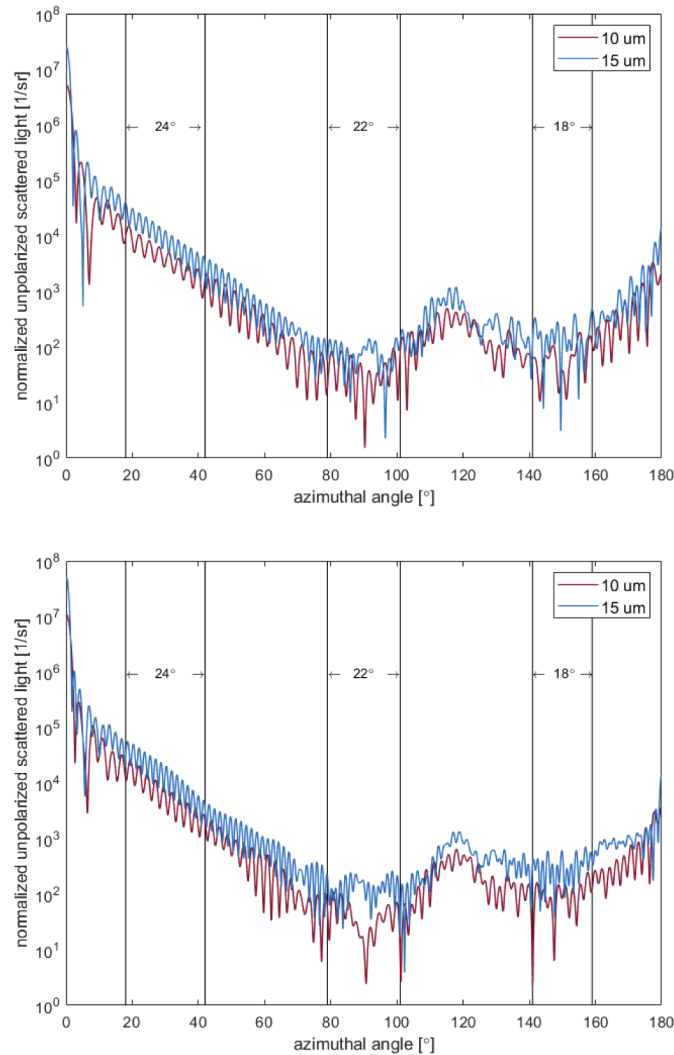
Fig. 3. Angular spectra from Mie theory calculations. The red curve corresponds to the scattering intensity from spherical beads of diameter 10 µm; the blue curve corresponds to beads of the size defined by a diameter of 15 µm. The vertical lines define the angular ranges corresponding to the three directions of measurements: 18°-42° (forward), 79°-101° (side), and 141°-159° (backward). Above: illumination using 632.8 nm light; below: illumination using 532 nm light.

cells and to identify the concentration of cells of given characteristics. Two parameters required by Mie calculation, the size parameter x and the relative index of refraction m, can be described by equations:

$$x = \frac{2\pi r}{\lambda_{vac}/n_{medium}} \tag{1}$$

$$m = \frac{n_{sphere}}{n_{medium}} \tag{2}$$

where r is the radius of the spherical object, λ_{vac} is the vacuum wavelength, n_{sphere} and n_{medium} are the refractive indices of the spherical object and medium under this wavelength, respectively. For the red laser with 632.8 nm wavelength, the refractive indices of polystyrene latex beads and PBS are 1.587 and 1.332 respectively. For the green laser with 532 nm wavelength, they are 1.598 and 1.3337 respectively. These parameters are used in the Mie calculation, whose result is shown in Fig. 3 in the form of line plots that express the angular distribution of logarithm scattered intensities. Mie calculation was done using external software [32].

The theoretical result shows the angular variation of the intensities, which results in the fringe patterns in the experimental result. Three angle windows are applied in the plot in Fig. 3, which correspond to the three observation angle ranges from the forward, side and backward directions. The number of fringes in the experimental results and the number of peaks in the calculation results shown in Fig. 3 are within 1 fringe of one another, and considered to be in good agreement.

4. Results and discussion: SH-SY5Y cells

### 4.1. Measurements of non-treated and staurosporine-treated SH-SY5Y cells

Experimental scattered light patterns of both treated and non-treated SH-SY5Y cells in all three directions have been successfully obtained using our experimental setup of Fig. 1. The angular distribution of scattered light displayed complicated patterns with no easily discernible characteristics and no apparent differences between samples that contained treated and non-treated cells, cf. Figs. 4–7. An analysis of such experimental spectra is usually based on relations between scattering objects in cells and features of the scattered light spectra within different angular ranges. Numerical simulations of light scattering from dielectrics have been helpful in identifying such relations. For example, the nucleus, the largest organelle in a cell, scatters light predominantly in the forward directions. If the regular fringe scatter pattern is observed, the Mie theory can be used to estimate the overall size of the nucleus [16]. Nanoscale structures such as mitochondria play an important role in the side scattering. If mitochondria are optically thick they dominate the scattering cross-section and produce randomly distributed speckle-like patterns of the scattered light intensity [10].
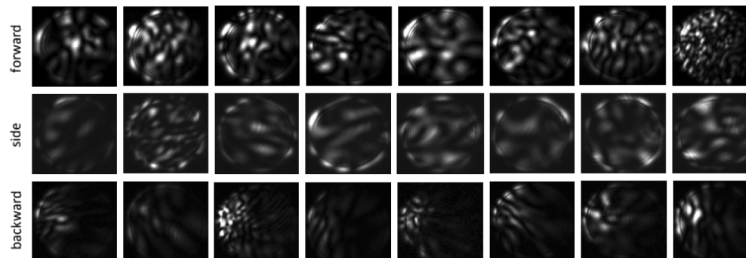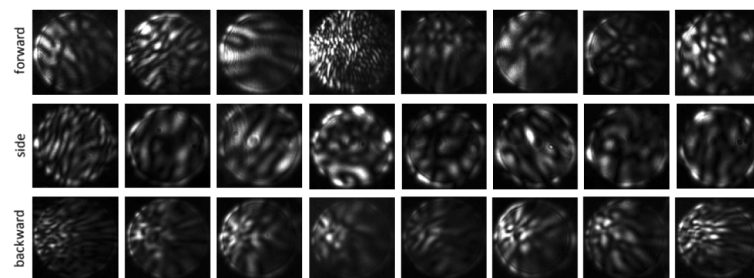

Fig. 4. Scattered light patterns from treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).
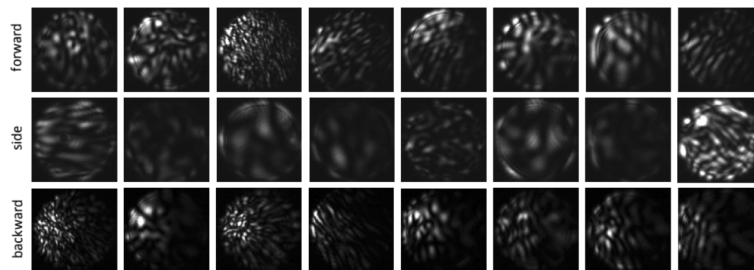
Fig. 5. Scattered light patterns from non-treated SH-SY5Y with red laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).
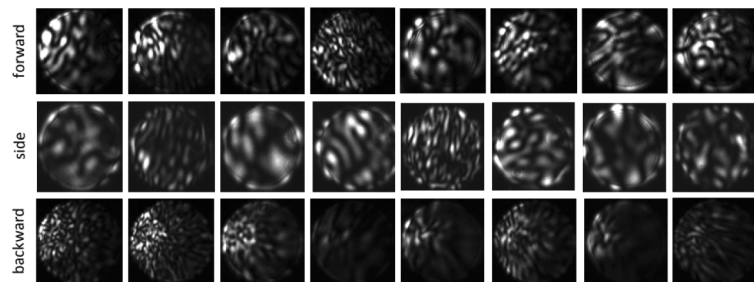
Fig. 6. Scattered light patterns from treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).

Fig. 7. Scattered light patterns from non-treated SH-SY5Y with green laser. The angular ranges corresponding to the three directions of measurements: 18°- 42° (forward), 79°-101° (side), and 141°-159° (backward).

Such interpretations of scattering spectra are enabled by the analytic solution, i.e. Mie theory, for the scattering from dielectric spheres, or numerical simulations involving modelling of cells as inhomogeneous dielectrics and solving Maxwell equations [8] describing the laser probe and scattered light. Motivated by the measurements with SH-SY5Y cells, this procedure of numerical modelling of the scattering experiments will be explored for the effect of the cell surface roughness on the scattering spectra.

Experimental samples of scattering spectra are shown in Figs. 4–7. The small fraction of images, approximately 10%, in the forward and side angular ranges display regular fringe patterns similar to Mie theory results for spherical beads. Because the same relative number of such spectra was observed in the treated and non-treated samples, we removed these spectra from further analysis.

Scattering spectra from red laser illumination are shown in Figs. 4 and 5, while those from green laser illumination are shown in Figs. 6 and 7. The top, middle and bottom rows show scattering patterns measured in forward, side and backward direction, respectively.

4.2. Models of cells and the effect of surface roughness on the scattering angular spectra

Rather than the regular fringe patterns shown in the bead experiment, the scattered light spectra of cells contain complicated patterns with dominant contributions from speckle-like distributions of scattered light intensities. The two types of cells, treated and non-treated SH-SY5Y cells, contribute to speckle-like distribution of the scattered light in Figs. 4–7. As it was established before in simulations and experiments [8,14–16], small organelles with relatively high indices of refraction can be responsible for the transformation of the regular Mie theory fringe patterns into randomly distributed local maxima of laser intensity in the angular scattering spectra.

In the following, we will first examine images of SH-SY5Y cells obtained by scanning electron microscopy and next introduce the theoretical model of the cell roughness. In addition to the effects of small-scale organelles inside the cells we will also show that scattering from cells with an increasing degree of surface roughness results in a transition from fringes to complicated angular distributions of scattered light intensity that can be observed in Figs. 4–7.

The scanning electron microscopy (SEM) images for non-treated and staurosporine-treated SH-SY5Y cells shown in Fig. 8 were taken to better understand their morphology.
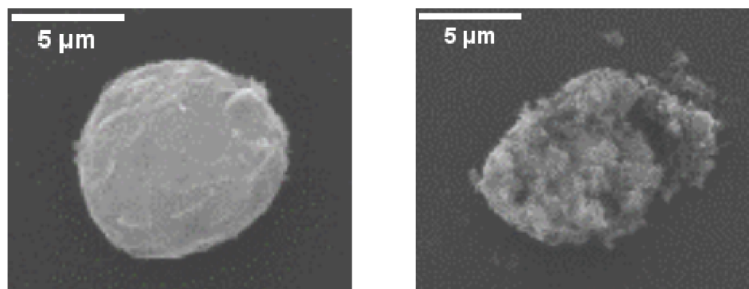
Fig. 8. SEM images of non-treated (left) and staurosporine-treated (right) SH-SY5Y cells. The two images correspond to cells of the similar size approximately 8 μm in diameter.

There is an obvious difference between SEM images of treated and non-treated SH-SY5Y cells. The non-treated cells in Fig. 8 are characterized by a relatively smooth surface and quasi-spherical shapes while the treated cells have rough, almost sponge-like surfaces and various shapes. The apparent similarities between the scattering spectra of these two kinds of cells despite such different physical features that are displayed in Fig. 8 contribute to difficulties in identifying the main components contributing to the scattered light spectra. We believe that the internal

components of the non-treated cells, such as mitochondria [33] produce speckle distributions of the scattered light, while in the case of treated cells the additional source of the complicated speckle-like spectra is the surface roughness.

We have developed a computer model of a spherical dielectric with a rough surface and have simulated scattering spectra using our Finite Difference Time Domain (FDTD) code, AETHER [8]. The surface roughness is characterized by two parameters, the amplitude of the surface modulations σ and the correlation length of the random surface perturbation, Λ [34]. The two parameters are introduced into the model by first dividing the sphere into many statistically independent circular thin slices. For each circular slice the radius is modified by small variations, h, such that the averages satisfy:

$$\langle h \rangle = 0 \tag{3}$$

$$\langle h^2 \rangle = \sigma^2 \tag{4}$$

The distribution of the modulations along the circle satisfies the probability distribution:

$$p(h) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-h^2}{2\sigma^2}\right) \tag{5}$$

The spacing of the surface modulations along the circumference of each slice is given by the correlation function:

$$C(R) = \frac{\langle h(\rho)h(\rho + R)\rangle}{\sigma^2} = \exp\left(\frac{-R^2}{\Lambda^2}\right) \tag{6}$$

where R is the length of an arc along the circumference of each slice of the sphere. The speckle-like simulated light scattered pattern for a spherical cell with rough surface is shown in Fig. 9. The simulation results support our conjecture that the speckles can come from the surface roughness of the cells. The analysis of the speckles can conversely give information of
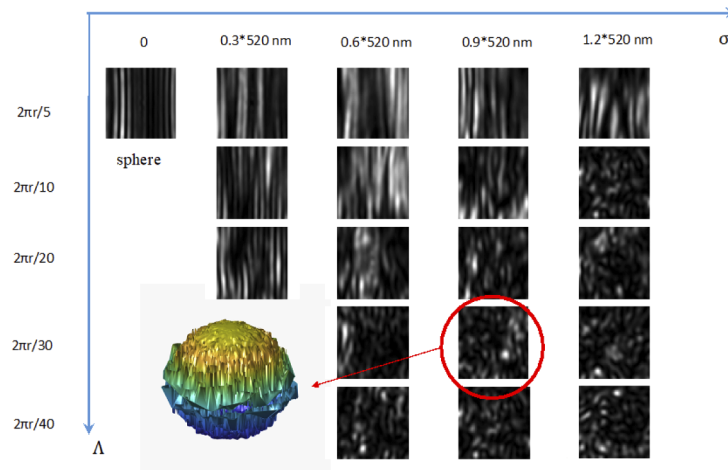
Fig. 9. Simulated scattered light pattern in the side direction from a spherical cells with changing surface roughness. The vertical axis shows the decreasing correlation lengths, Λ, expressed as the fraction of the circumference of different horizontal slices of the sphere. The horizontal axis indicates the increasing amplitude of the surface modulations. The wavelength of the laser probe is 520 nm. The 3D image of the sphere surface corresponds to the scattered spectrum inside red circle. The radius of the sphere is r=5 μm, index of refraction of a dielectric is n=1.57. The sphere is immersed in water with refraction index 1.334.

the cell surface. Potentially, we can distinguish the cells with different surfaces by doing the statistical analysis of the scattering patterns. As we proceed, clear identification of the source of the irregular scattering patterns, e.g. small organelles or the surface roughness, will require the application of pattern recognition and machine learning techniques.

5.   Discrimination between treated and non-treated SH-SY5Y cells

5.1.   Two-parameter analysis of 2D light scattering patterns

In our previous studies [14,15], we were able to use two observables, number of speckles and average area of their cross-sections, as parameters for cell identification. These speckles are produced as a result of interference of scattered light from small and optically dense mitochondria and as discussed above on random and deep perturbations of the cell surfaces. Here we perform similar analysis for scattered light patterns from non-treated (NT) and staurosporine-treated (ST) SH-SY5Y cell samples. Only the speckle and irregular patterns were analyzed as the fringe Mie scattering patterns were already eliminated. Figure 10 shows an example of the speckle patterns from treated cells. In our analysis, we first found each local maximum in the pattern, and then starting from each local maximum scanned pixel values in four directions (upward, downward, leftward and rightward) until reaching the half value of each local maximum pixel value. The scanning area around each local maximum is defined as the local cross-section area. As Fig. 10 shows, each local maximum is labeled with green colour, and each local cross-section area is represented by a red quadrilateral. All speckle patterns were smoothed by using a Gaussian smoothing filter to eliminate high-frequency noise before applying the method of finding the local maximum. The average cross-sectional area is defined as the summation of local cross-section areas divided by the number of local maxima.
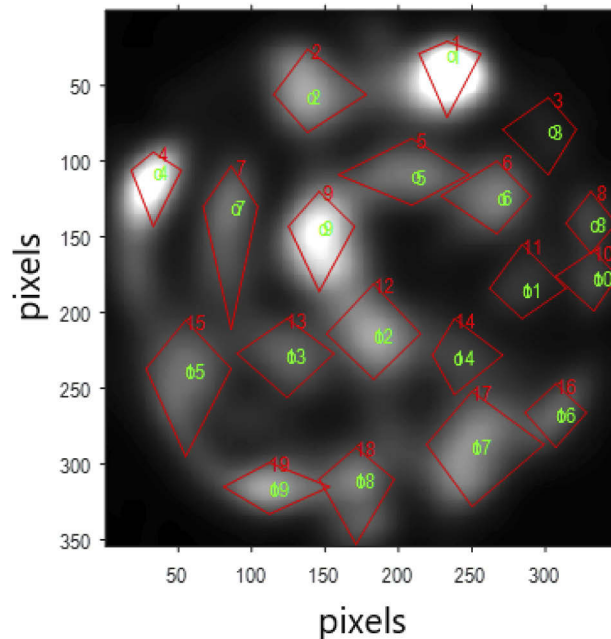
Fig. 10. A typical scattered light pattern from a single SH-SY5Y cell. Each local maximum is labeled with a green circle and each local cross-section area is labeled with a red quadrilateral shape.

This example illustrates the method of physical cytometry based on the number and average size of speckles in the 2D scattered light intensity distribution. In our previous studies [14,15] the two parameters, i.e., the average local cross-section areas and number of speckle spots, for different types of blood cells do not overlap with each other and thus can be used as fingerprints to identify the cell types. In the current study, the spreads of the two parameters for ST and NT SH-SY5Y cells result in them significantly overlapping with each other. The dataset including the two wavelengths and three observation directions was investigated to examine the behavior of the two parameters. It was found that the average number of speckle spots for NT cells are larger than ST cells and the average sizes of the speckles for NT cells are smaller than the ST cells [35]. The differences between the average numbers of speckle spots for NT cells and those for the ST cells are larger for red laser illumination than for green laser illumination, in all of the three observation directions, with no significant directional dependency [35]. Based on the observations, analysis is focused on the data in the forward direction with red laser illumination. The two parameters were determined from the scattered light patterns in the forward direction with red laser illumination for 31 ST cells (red solid circles) and 38 NT cells (black hollow circles), as shown in Fig. 11(a). Even though the two parameters for ST and NT cells overlap significantly with each other, it is more likely to find a NT cell with large number of spots than a ST cell. The average numbers of spots are 23 and 26 for ST and NT cells respectively. In Fig. 11(a) we found that the number of ST cells having 25 or more intensity maxima was 9 out of 31, while the number of NT cells having 25 or more speckles was 22 out of 38. Thus, this training data set gives the probabilities 0.58 and 0.29 for a NT cell and a ST cell, respectively, having 25 or more maxima. The probabilities from the training sample can be used to estimate the fraction of treated and non-treated cells in a sample where staurosporine was applied to SH-SY5Y cells to study the effectiveness of such treatment. An experiment was carried out to verify this approach using a mixed solution with 4:1 ratio of NT to ST cells as a probing data set. Scattered light patterns in the forward direction with red laser illumination from 27 cells in the mixed solution of probing data set were collected and analyzed as shown in Fig. 11(b). Of the 27 patterns analyzed, 17 were found having 25 or more speckle maxima. The result is consistent with the fact that most of the cells in the mixture are NT. Since the errors of measurements typically decrease proportionally to the inverse of the square root of the number of data, it is expected that the accuracy of the prediction can be improved with larger training and probing data sets. Thus, the above result would point to the fact that a single criterion, i.e., the number of speckle spots, can be used for the estimation of fractions of the two groups of cells in a mixed solution providing sufficient large set of training and probing data sets are available.
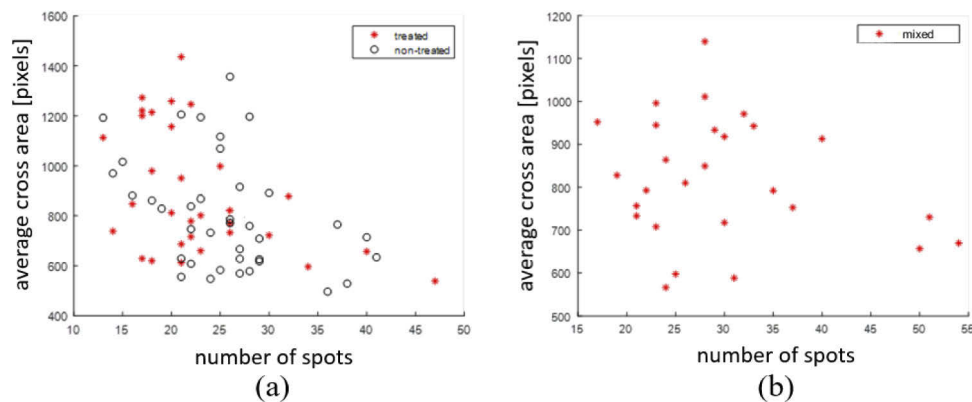

Fig. 11. Two-parameter plots, total local cross areas and number of speckle spots, for scattered light patterns from (a) "training" and (b) "probing" data sets.

However, to achieve robust identification and determination of the fraction of different cell types in a mixed solution with relatively small data sets, more sophisticated methods will need to be developed. In the next section, we describe a ML method developed for robust identification of the two groups of SH-SY5Y cells.

6. Classification of non-treated and treated SH-SY5Y cells using machine learning algorithms

In this section, we present the ML technique used in this paper for classification of scattered light patterns. The data is collected in the forward direction with red laser. The objective is to classify the patterns into two classes: ST (staurosporine-treated) and NT (non-treated).

The overall schematic of the ML module is shown in Fig. 12. The input images are resized to 224x224 pixels and are fed to a pretrained CNN, DenseNet-201 [20], for feature extraction. Note that the DenseNet-201 consists of 201 layers, including convolutional (Conv), ReLU activation (ReLU), batch normalization (BN), pooling layers, etc. A feature vector of length 1920 is obtained at the GAP (Global Average Pooling) layer of the DenseNet for each image. The feature vectors are then fed to an RBF-kernel SVM that outputs a real value score in the range [0, 1]. The score determines the probability of the image class (i.e., the probability of being an NT or ST). As the pretrained DenseNet is used only as a feature extractor, there is no training required for the DenseNet, and only the SVM needs to be trained to obtain the classification output.
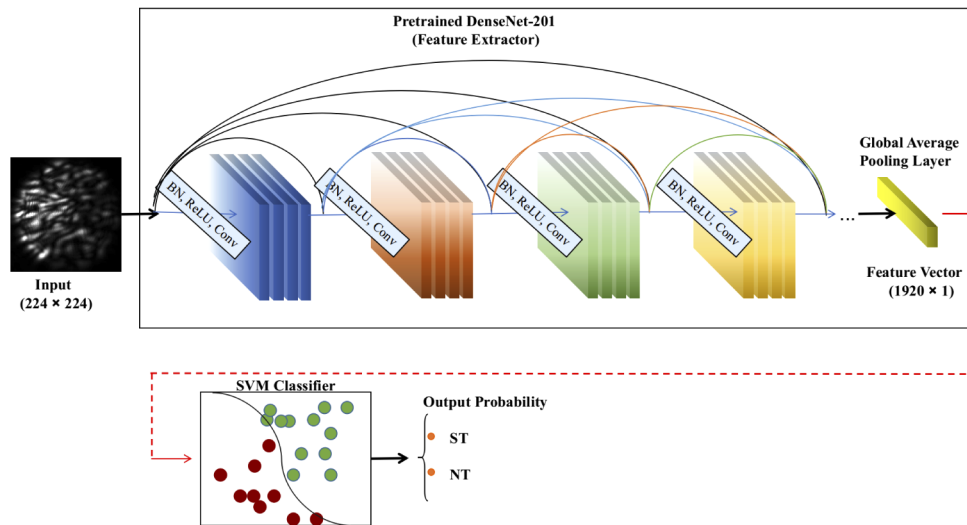
Fig. 12. Schematic of the machine learning module.

To evaluate the performance of the machine learning algorithm, we use a dataset of 360 images that consists of 180 NT and 180 ST cells. The average performance of 5-fold cross-validation is used as the final result. For 5-fold cross-validation, the entire dataset is divided into 5 non-overlapping subsets (with equal number of ST and NT cells in each subset). For each fold, one data subset (i.e., 20% of the data) is used as the testing data, and the remaining 4 subsets (i.e., 80% of data) are used as the training data. This process is repeated 5 times until each subset has been used as testing data. To select the best parameters of SVM, 5-fold cross-validation is performed on the training set within each fold. We also perform the same parameter selection step for the other hand-crafted features used as comparison in this paper, so as to make a fair comparison.

The classification performance is evaluated in terms of Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), and Area Under the ROC Curve (AUC), which are defined below.

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \tag{7}$$

$$SEN(TPR) = \frac{TP}{TP + FN} \tag{8}$$

$$SPE = \frac{TN}{TN + FP} \tag{9}$$

$$AUC = \int_0^1 TPR(F_0)\, d(F_0) \tag{10}$$

In this work, the NT cells are considered as the positive samples and the ST cells are considered as the negative samples for the SVM. In this case, TP (True Positive) is the number of the correctly classified NT cells, TN (True Negative) is the number of the correctly classified ST cells, FP (False Positive) is the number of the ST cells (mistakenly) classified as the NT cells and FN (False Negative) is the number of the NT cells being (mistakenly) classified as ST cells. The SEN, also known as True Positive Rate (TPR), shows how many among the NT cells are correctly classified as the NT cells. The SPE shows how many among the ST cells are correctly classified as ST cells. Note that the AUC calculates the area under the TPR-FPR curve (also known as the ROC curve), where TPR(F_0) is the value of the TPR when the false positive rate, defined as FPR=FP/(TN+FP) is F_0.

The performance of the proposed ML technique is also compared with other state-of-the-art scattering analysis techniques [14,15,18,19]. The SF technique [14,15] uses the two speckle features: number and area of speckles. For calculating the number and average area of speckles, we first perform the segmentation on the scattering patterns, and calculate the two features using the foreground mask. An example of a segmented mask is shown in Fig. 13. The original image first goes through a Gaussian filter to remove the noise and smooth the image. The GLCM technique [18] uses the gray-level co-occurrence matrix features. We set the gray level range to be [0,255], the pixel pair distance as 1, and use four directions of 0°, 45°, 90°, 135° to calculate the GLCM. For each of the four directions, 22 statistical features are calculated from the GLCM. Overall, an 88-dimensional feature vector is obtained by concatenating features obtained from the 4 directions. Note that before calculating the GLCM features, Gaussian filtering is performed on each image to reduce the impact of noise. The HOG technique [19] uses the histograms of oriented gradients as features. Here, the original images are divided into non-overlapping 16x16 pixel cells, and 9 bins are used for the histograms. The dimension of the extracted HOG features is 1764 for each image. The Hand-Crafted technique uses all three hand crafted features SF, GLCM and HOG. For all four techniques, SVM is used for pattern classification. 5-fold cross validation is used to obtain the classification performance. In this paper, SVM with RBF (radial basis function) kernel is used as it provided superior performance.

The classification performance (classes: NT and ST) of the proposed machine learning method along with the existing methods is shown in Table 1. It is observed that the deep features-based classification significantly outperforms the hand-crafted features. Among all these features, the SF technique provides the least satisfactory performance with ACC of 61.67%. One possible reason may be that the SF uses only a 2-dimensional feature vector, which contains less information. Therefore, it has limited power in characterizing the characteristics of NT and ST cells resulting in low discriminative ability. However, it still achieves an ACC over 50%, which indicates that useful information is contained in the speckle features. The GLCM features obtain an improvement of about 10% accuracy compared with the speckle features. This indicates that the co-occurrence matrix can capture more useful texture information and is beneficial for the
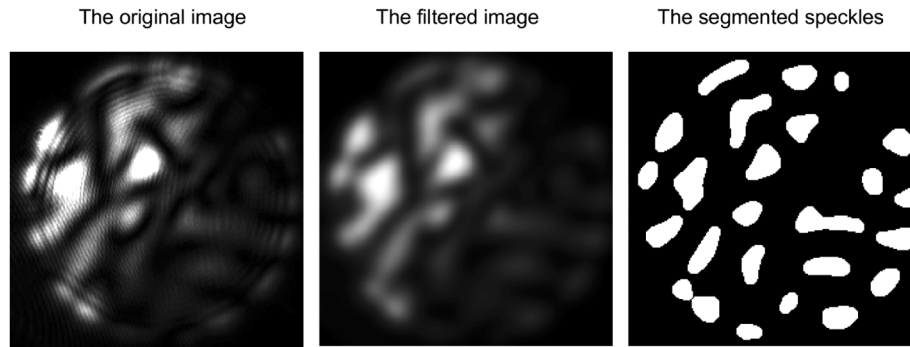
Fig. 13. Segmented speckles for pattern analysis. Note that speckles (in the left image) with very low contrast have also been identified and segmented.

distinction of the NT and the ST cells. Among the handcrafted features, the HOG features present the best performance. The combination of all three hand-crafted features provides further improved performance, since it captures more diverse and useful information, with 77.78% ACC, 82.78% SEN, 72.78% SPE, and 86.62% AUC.

Table 1. Classification Performance of the machine learning method. All techniques use SVM as the classifier.

Classification Method	ACC	SEN	SPE	AUC
SF [14,15]	61.67	63.33	60.00	64.86
GLCM [18]	71.11	70.56	71.67	77.19
HOG [19]	75.28	76.11	74.44	83.15
Hand-Crafted	77.78	82.78	72.78	86.62
Proposed ML	91.11	92.78	89.44	96.95

As observed in Table 1, the proposed ML technique provides the best performance among all techniques with over 10% improvement in ACC, SEN, SPE and AUC over the Hand-Crafted technique. This shows that, although the manually designed hand-crafted feature contains useful statistical features of image textures, it has limited discriminative power. It fails when the task is complex (The texture patterns of the non-treated and treated SH-SY5Y cells are visually similar, and can be regarded as the complex situation in this case). As a contrast, the CNN methods can extract semantic-meaningful features with the deep architecture. Texture features (such as edges, corners) can be extracted in the bottom layers of a CNN model. As the layers go deeper, more semantic-meaningful and rich features can be extracted (such as parts, shapes, patterns). In this paper, we used the feature vector in the GAP layer as the feature representation of an input image, which is a high-level and semantic-meaningful feature representation. Compared with the hand-crafted features, the extracted CNN features are not data dependent and can extract semantic-meaningful features, which are discriminative and hence more powerful for the classification task. The proposed ML technique can tackle the complex situation and obtain superior performance compared with the technique using hand-crafted features.

The proposed ML technique is evaluated on the Desktop with Intel i7-7700 4.2 GHz CPU with 32 GB memory and a GPU of Nvidia GeForce GTX 1080Ti with 11GB memory. The features are extracted using pretrained Densenet in Pytorch and the classification of SVM method is implemented using Matlab. The inference time for a single cell image is about 0.033 seconds on average, and the majority of the time is spent on feature extraction, which is about 0.0326 seconds. It worth noting that, the feature extraction time can be sped up by using compression

technology of deep neural networks or some more advanced CNN methods, which are more lightweight, efficient and without the loss of accuracy. The inference time is very fast and has the potential for real-time cell classification.

In some medical applications, the percentage of ST cells (PSTC) defined below, is a diagnostic factor.

$$PSTC = \frac{No.ofSTcells}{TotalNo.ofcells} \times 100\% = \frac{TN + FN}{TP + FP + TN + FN} \times 100\% \tag{11}$$

Here, we present the accuracy of estimating the PSTC using ML based classification of the ST and NT cells. In this experiment, the same testing dataset is used but with different numbers of ST cells (selected randomly). In this way, testing set with different concentration of ST cells are constructed. The process is repeated 10 times, and the average PSTC value is calculated. Three experiments were carried out using ground truth (i.e., calibrated dataset) PSTC of 50%, 33% and 20%. Table 2 shows the experimental results. The last column of Table 2 shows the predicted PSTC value. It is observed the predicted PDC value is very close to the ground truth PDC value when the dataset includes roughly similar number of NT and ST cells. However, when the number of NT cells exceeds significantly the number of ST cells, the predicted PSTC value deviates slightly more and is usually larger than the ground truth value. However, the predicted percentage is within 5% of the ground truth percentage, which may be acceptable for making diagnostic decision.

Table 2. Prediction of PSTC (percentage of ST cells) value.

Testing Dataset	Ground Truth Percentage	Predicted Percentage
36 ST, 36 NT	50.0	49.1
18 ST, 36 NT	33.3	35.2
9 ST, 36 NT	20.0	24.2

As discussed in Section 4, the sources of the observed speckle-like patterns for NT and ST SH-SY5Y cells can be from small organelles and cell surface roughness. The ML method developed here was able to effectively classify these two groups of SH-SY5Y cells and thus potentially can be used to pinpoint the origins of the scattered light patterns by studying simulated scattering patterns from various combinations of small organelles with different distributions and cell surface with different degree of roughness. This can potentially be a useful tool for studying SH-SY5Y cells treated with staurosporine at various concentrations which may affect the internal organelles and change the cell surface roughness.

7. Summary and conclusions

We have described in this paper experiments and theoretical analysis of light scattering from neuroblastoma cells. The main goal of this study is to examine different components, from label-free light scattering to machine learning image recognition, comprised in a procedure that allows accurate and reliable measurement of the concentration of different cells in biological samples. We have demonstrated the effectiveness of our method by measuring the relative concentration of the non-treated vs staurosporine treated SH-SY5Y neuroblastoma cells in a sample. Applications of this method to measurements of relative concentrations of cells in laboratory and clinical settings could offer a useful alternative to techniques currently used and based on the fluorescent cytometry and manual analysis of scattering spectra.

Our group has demonstrated over the years [9–16] that the measurement of the angular intensity distribution of the scattered light provides a non-invasive way of cell characterization that can be employed in cell discrimination and the recognition of various physiological states in cells of the same kind. Further advancements in the light scattering experiments in this paper involved

measurements in three directions (forward, side and backward) with the application of two laser sources of different wavelengths, 520 nm and 632.8 nm.

When combined with numerical modelling using our FDTD code AETHER [8], label-free scattering experiments led to the method of rapid cell size determination [11] and differentiation of the hematopoietic stem cells [10,12] based on the mitochondria distribution. Mitochondria represent randomly distributed scattering centers that give rise to speckle patterns in the scattered light distribution from the random interference of coherent light of the laser probe. Speckle patterns were described using two parameters statistical analysis [10,12] in terms of the number of speckles in 2D intensity distributions and their average cross-sections. This procedure was applied here to distinguish the non-treated and staurosporine treated SH-SY5Y neuroblastoma cells.

The SEM images revealed that application of the staurosporine leads to a change of the cell morphology and an increase in the cell surface roughness. Our theoretical modelling of the cell roughness and shape modification demonstrated in Sec. 4.2 that large amplitude and short correlation length modifications in the cell roughness result in the speckle pattern in the scattering light similar to the effect of scattering on mitochondria. The cumulative effect on the scattered light of the internal organelles and cell roughness reduced the accuracy of our two-parameter speckle analysis and prompted the application of machine learning techniques.

It was observed that a Deep Neural Network-based method developed in this study has the best classification performance with Accuracy (ACC) of 91%, Sensitivity (SEN) of 93%, Specificity (SPE) of 89% and Area Under the ROC Curve (AUC) of 97%. The robust classification was applied to determine the fraction of staurosporine-treated SH-SY5Y cells in a mixed solution consisting of staurosporine-treated and normal SH-SY5Y cells. Experimental results show very good prediction accuracy (within 5% of the ground truth fraction value).

In summary, our label-free cytometry has the potential for real-time detection and diagnosis of different cells. In this study, we have shown that it can be used for studying the effect of staurosporine on human neuroblastoma cells which has potential applications including fundamental apoptosis studies in neuroscience and neuroprotection research.

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. J. Picot, C. L. Guerin, C. Le Van Kim, and C. M. Boulanger, "Flow cytometry: retrospective, fundamentals and recent instrumentation," Cytotechnology **64**(2), 109–130 (2012).
2. K. Hiramatsu, T. Ideguchi, Y. Yonamine, S. Lee, Y. Luo, K. Hashimoto, T. Ito, M. Hase, J.-W. Park, Y. Kasai, S. Sakuma, T. Hayakawa, F. Arai, Y. Hoshino, and K. Goda, "High-throughput label-free molecular fingerprinting flow cytometry," Sci. Adv. **5**(1), eaau0241 (2019).
3. Y.-L. Pan, M. J. Berg, S. S.-M. Zhang, H. Noh, H. Cao, R. K. Chang, and G. Videen, "Measurement and autocorrelation analysis of two-dimensional light-scattering patterns from living cells for label-free classification," Cytometry, Part A **79A**(4), 284–292 (2011).
4. D. Watson, N. Hagen, J. Diver, P. Marchand, and M. Chachisvilis, "Elastic light scattering from single cells: orientational dynamics in optical trap," Biophys. J. **87**(2), 1298–1306 (2004).
5. L. Zhang, Y. Qin, K.-X. Li, X. Zhao, Y.-F. Xing, H. Zhao, Z. Jiang, W. Chen, N.-L. Yao, and L. Yuan, "Light scattering properties in spatial planes for label free cells with different internal structures," Opt. Quantum Electron. **47**(5), 1005–1025 (2015).
6. G. I. Ruban, V. V. Berdnik, D. V. Marinitch, N. V. Goncharova, and V. A. Loiko, "Light scattering and morphology of the lymphocyte as applied to flow cytometry for distinguishing healthy and infected individuals," J. Biomed. Opt. **15**(5), 057008 (2010).

7. J. R. Mourant, M. Conoplat, C. Brocker, O. Esponda-Ramos, T. M. Johnson, A. Matanock, K. Stetter, and J. P. Freyer, "Light scattering from cells: the contribution of the nucleus and the effects of proliferative status," J. Biomed. Opt. **5**(2), 131 (2000).

8. C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," J. Biomed. Opt. **10**(1), 014007 (2005).

9. K. Singh, X. Su, C. Liu, C. E. Capjack, W. Rozmus, and C. J. Backhouse, "A miniaturized wide-angle 2D cytometer," Cytometry, Part A **69A**(4), 307–315 (2006).

10. X. Su, C. E. Capjack, W. Rozmus, and C. J. Backhouse, "2D light scattering patterns of mitochondria in single cells," Opt. Express **15**(17), 10562 (2007).

11. X. Su, K. Singh, C. E. Capjack, J. Petracek, C. J. Backhouse, and W. Rozmus, "Measurements of light scattering in an integrated microfluidic waveguide cytometer," J. Biomed. Opt. **13**(2), 024024 (2008).

12. X. Su, K. Singh, W. Rozmus, C. J. Backhouse, and C. E. Capjack, "Light scattering characterization of mitochondrial aggregation in single cells," Opt. Express **17**(16), 13381 (2009).

13. X. Su, W. Rozmus, and Y. Y. Tsui, "Wide angle light scattering differentiation of organelle-size particle distributions in whole cells," Cytometry, Part A **77A**(6), 580–584 (2010).

14. X. Su, S. E. Kirkwood, M. Gupta, L. Marquez-Curtis, Y. Qiu, A. Janowska-Wieczorek, W. Rozmus, and Y. Y. Tsui, "Microscope-based label-free microfluidic cytometry," Opt. Express **19**(1), 387 (2011).

15. X. Su, M. Gupta, C. E. Capjack, Y. Y. Tsui, Y. Qiu, L. Marquez-Curtis, A. Janowska-Wieczorek, and W. Rozmus, "Label-free and non-invasive optical detection of the distribution of nanometer-size mitochondria in single cells," J. Biomed. Opt. **16**(6), 067003 (2011).

16. H. Shahin, M. Gupta, A. Janowska-Wieczorek, W. Rozmus, and Y. Y. Tsui, "Physical characterization of hematopoietic stem cells using multidirectional label-free light scatterings," Opt. Express **24**(25), 28877 (2016).

17. X. Su, S. Liu, X. Qiao, Y. Yang, K. Song, and B. Kong, "Pattern recognition cytometry for label-free cell classification by 2D light scattering measurements," Opt. Express **23**(21), 27558 (2015).

18. J. Liu, Y. Xu, W. Wang, Y. Wen, H. Hong, J. Q. Lu, P. Tian, and X.-H. Hu, "Machine learning of diffraction image patterns for accurate classification of cells modeled with different nuclear sizes," J. Biophotonics **13**(9), e202000036 (2020).

19. X. Su, T. Yuan, Z. Wang, K. Song, R. Li, C. Yuan, and B. Kong, "Two-dimensional light scattering anisotropy cytometry for label-free classification of ovarian cancer cells via machine learning," Cytometry, Part A **97**(1), 24–30 (2020).

20. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 2261–2269.

21. L. Liu, L. Mou, X. X. Zhu, and M. Mandal, "Automatic skin lesion classification based on mid-level feature learning," Comput. Med. Imaging Graph. **84**, 101765 (2020).

22. H. Xicoy, B. Wieringa, and G. J. M. Martens, "The SH-SY5Y cell line in Parkinson's disease research: a systematic review," Mol. Neurodegener. **12**(1), 10 (2017).

23. "Neurological Disorders," *Johns Hopkins Medicine*, https://www.hopkinsmedicine.org/health/conditions-and-diseases/neurological-disorders.

24. J. Kovalevich and D. Langford, "Considerations for the use of SH-SY5Y neuroblastoma cells in neurobiology," Methods Mol. Biol. **1078**, 9–21 (2013).

25. H-R. Xie, L.-S. Hu, and G. Y. Li, "SH-SY5Y human neuroblastoma cell line: In vitro cell model of dopaminergic neurons in Parkinson's disease," Chinese Medical Journal **123**, 1086–1092 (2010).

26. E. López and I. Ferrer, "Staurosporine- and H-7-induced cell death in SH-SY5Y neuroblastoma cells is associated with caspase-2 and caspase-3 activation, but not with activation of the FAS/FAS-L-caspase-8 signaling pathway," Mol. Brain Res. **85**(1-2), 61–67 (2000).

27. J. C. Fitzgerald, C. Ufer, L. A. De Girolamo, H. Kuhn, and E. Ellen Billett, "Monoamine oxidase-A modulates apoptotic cell death induced by staurosporine in human neuroblastoma cells," J. Neurochem. **103**(6), 2189–2199 (2007).

28. D. Jantas, M. Pytel, J. W. Mozrzymas, M. Leskiewicz, M. Regulska, L. Antkiewicz-Michaluk, and W. Lason, "The attenuating effect of memantine on staurosporine-, salsolinol- and doxorubicin-induced apoptosis in human neuroblastoma SH-SY5Y cells," Neurochem. Int. **52**(4-5), 864–877 (2008).

29. M. Leskiewicz, M. Regulska, B. Budziszewska, D. Jantas, L. Jaworska-Feil, A. Basta-Kaim, M. Kubera, G. Jagla, W. Nowak, and W. Lason, "Effects of neurosteroids on hydrogen peroxide- and staurosporine-induced damage of human neuroblastoma SH-SY5Y cells," J. Neurosci. Res. **86**(6), 1361–1370 (2008).

30. M. Barrachina, J. Secades, R. Lozano, C. Gómez-Santos, S. Ambrosio, and I. Ferrer, "Citicoline increases glutathione redox ratio and reduces caspase-3 activation and cell death in staurosporine-treated SH-SY5Y human neuroblastoma cells," Brain Res. **957**(1), 84–90 (2002).

31. H. Shahin, "Forward and side two dimensional scattered light patterns studies of single cells for label-free flow cytometry," M.S. thesis (Department of Electrical and Computer Engineering, University of Alberta, 2016).

32. P. Laven, "Simulation of rainbows, coronas, and glories by use of Mie theory," Appl. Opt. **42**(3), 436 (2003).

33. K. M. Lyser, "Low- and high-voltage electron microscopy of a human neuroblastoma in long-term organ culture," Cancer Res. **34**, 594 (1974).

34. N. Garcia and E. Stoll, "Monte Carlo calculations for electromagnetic-wave scattering from random rough surfaces," Phys. Rev. Lett. **52**(20), 1798–1801 (1984).
35. Y. Wan, "Non-invasive Label-free Cytometry Based on Laser Light Scattering Technique," M.S. thesis (Department of Electrical and Computer Engineering, University of Alberta, 2018).