



Label-free discrimination and quantitative analysis of oxidative stress induced cytotoxicity and potential protection of antioxidants using Raman micro-spectroscopy and machine learning

Wei Zhang^{a,1}, Jake S. Rhodes^{b,1}, Ankit Garg^c, Jon Y. Takemoto^d, Xiaojun Qi^e, Sitaram Harihar^f, Cheng-Wei Tom Chang^g, Kevin R. Moon^{b,**}, Anhong Zhou^{a,*}

^a Department of Biological Engineering, Utah State University, Logan, UT, 84322, USA

^b Department of Mathematics and Statistics, Utah State University, Logan, UT, 84322, USA

^c Logan High School, 162 W 100 S Logan, UT, 84321, USA

^d Department of Biology, Utah State University, Logan, UT, 84322, USA

^e Department of Computer Science, Utah State University, Logan, UT, 84322, USA

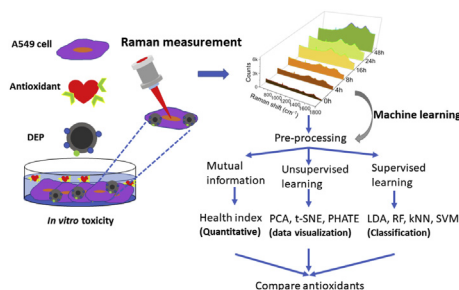
^f Department of Genetic Engineering, SRM Institute of Science and Technology, Kattankulathur, 603203, Tamil Nadu, India

^g Department of Chemistry and Biochemistry, Utah State University, Logan, UT, 84322, USA

HIGHLIGHTS

- Apply new algorithms (PHATE and MI) to visualize the Raman spectral data.
- Raman spectroscopy was utilized to monitor cellular responses to oxidative stress.
- The health index was proposed to quantitatively assess antioxidants protection.
- A number of machine learning algorithms were applied to analyze Raman spectral data.
- Correlation between Raman spectra and cytokine level was analyzed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 28 March 2020

Received in revised form

25 May 2020

Accepted 30 June 2020

Available online 12 July 2020

Keywords:

Raman spectroscopy

Machine learning

PHATE

ABSTRACT

Diesel exhaust particles (DEPs) are major constituents of air pollution and associated with numerous oxidative stress-induced human diseases. *In vitro* toxicity studies are useful for developing a better understanding of species-specific *in vivo* conditions. Conventional *in vitro* assessments based on oxidative biomarkers are destructive and inefficient. In this study, Raman spectroscopy, as a non-invasive imaging tool, was used to capture the molecular fingerprints of overall cellular component responses (nucleic acid, lipids, proteins, carbohydrates) to DEP damage and antioxidant protection. We apply a novel data visualization algorithm called PHATE, which preserves both global and local structure, to display the progression of cell damage over DEP exposure time. Meanwhile, a mutual information (MI) estimator was used to identify the most informative Raman peaks associated with cytotoxicity. A health index was defined to quantitatively assess the protective effects of two antioxidants (resveratrol and mesobiliverdin IX α) against DEP induced cytotoxicity. In addition, a number of machine learning

* Corresponding author.

** Corresponding author.

E-mail addresses: Kevin.Moon@usu.edu (K.R. Moon), Anhong.Zhou@usu.edu (A. Zhou).

¹ These authors contributed equally to this work.

Mutual information
Antioxidant

classifiers were applied to successfully discriminate different treatment groups with high accuracy. Correlations between Raman spectra and immunomodulatory cytokine and chemokine levels were evaluated. In conclusion, the combination of label-free, non-disruptive Raman micro-spectroscopy and machine learning analysis is demonstrated as a useful tool in quantitative analysis of oxidative stress induced cytotoxicity and for effectively assessing various antioxidant treatments, suggesting that this framework can serve as a high throughput platform for screening various potential antioxidants based on their effectiveness at battling the effects of air pollution on human health.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Air pollution is a major concern of the modern world, and has a serious toxicological impact on human health and the environment [1]. Diesel exhaust particles (DEPs) are one of the major components of air pollution and are associated with numerous human diseases, including cardiovascular diseases [2]; lung cancer [3]; vascular dysfunction which leads to thromboembolic disease and hypertension [4]; neurotoxicity and strokes [5]; prenatal health problems [6] and infertility [7]. At the cellular level, $10 \mu\text{g ml}^{-1}$ DEPs can substantially increase the production of reactive oxygen species (ROS) and reduce mitochondrial activity [8], which can induce oxidative damage to DNA, proteins and lipids [9]. *In vitro* toxicity studies are species-specific, simpler than *in vivo* studies, and can lead to a better understanding of *in vivo* conditions [10]. Conventional *in vitro* cytotoxicity assessments are based on analyzing cellular oxidative stress biomarkers, such as 8-oxodG for DNA damage [11], thiobarbituric acid reactive substances for lipid peroxidation [12], and 2,4-dinitrophenylhydrazine for protein peroxidation [12]. These assays are routinely utilized but are destructive (e.g., via DNA/protein extraction or dye labelling) and incapable of real-time measurement of cell behavior at the single cell level. Therefore, there is an urgent need to assess cytotoxicity using rapid, non-invasive methods and to develop new pharmaceutical drugs which will prevent the damage from air pollution to human health.

Raman spectroscopy has been demonstrated as a non-invasive and highly sensitive technique and has been widely applied to study biological samples [13,14]. With the unique ability of single living cell analysis in medium, Raman spectroscopy is specifically suitable for detecting molecular variation [15] by monitoring the spectral fingerprints of nucleic acids, proteins, lipids, and carbohydrates in the exposed cells [16]. This capability of Raman spectroscopy offers a new instrumental approach to assess *in vitro* cytotoxicity. The application of Raman spectroscopy in evaluating cell biochemical changes has been achieved in normal DEP-exposed human lung cells [17], anticancer drug-exposed nasopharyngeal carcinoma cells [18], and nanoparticle-induced A549 cells [19]. Thus, it is hypothesized that the oxidative stress induced on the cells via exposure to toxic substrates (e.g., DEPs) can be detected by monitoring the specific Raman spectral signatures (usually designated as characteristic peaks) of these cellular components.

Natural products, as an alternative to pharmaceutical compounds, have recently shown potential protection from particulate matter (PM_{2.5}, PM₁₀)-induced toxicity [20]. Resveratrol (3, 5, 4'-trihydroxystilbene, RES) can be naturally extracted from red grapes, berries, and other plants. It has been shown to exhibit antioxidant and anti-tumor activities [21]. Manufacturers of RES commercial supplements claim that they have various health benefits against numerous diseases and have displayed nearly no adverse effects in clinical studies [22]. Heme-derived bilins and bilirubin are natural

antioxidants [23]. As products of heme oxygenase-1, these bilins as well as expression of the producing enzyme cytoprotect against acute and chronic inflammatory conditions [24]. Mesobiliverdin IX α (MesoBV), an analog of biliverdin IX α and derived from photosynthetic cyanobacteria, was observed to inhibit lipid peroxidation and to cytoprotect pancreatic islets in a diabetic rat model for islet xenotransplantation [25]. Given these properties, our study examines the protective effects of RES and MesoBV against DEPs *in vitro*.

Machine learning (ML) methods are broadly divided into two categories: unsupervised methods, such as dimensionality reduction (e.g. principal component analysis or PCA, t-distributed stochastic neighbor embedding or t-SNE) and clustering (e.g. K-means and hierarchical cluster analysis); and supervised methods which include classification or regression. Common supervised methods include linear discriminant analysis (LDA), multiple linear regression, principal component regression, partial least squares (PLS), k-nearest neighbors (kNN), support vector machines (SVM), and random forests (RF). Unsupervised learning is often used as a precursor to supervised methods when working on large data sets, to either reduce the number of data points via clustering [26] or to reduce the dimensionality of the data [27]. PCA and t-SNE are unsupervised methods commonly used for data pre-processing or visualization. PCA makes use of an orthogonal transformation as a means of eliminating collinearity and reducing the dimensionality of the dataset. The scores may be used for cluster analysis (visualization) or regression. But biological data are often highly nonlinear [28], requiring the use of nonlinear methods for visualization, analyzing variable dependence, and dimensionality reduction. t-SNE assumes a non-linear manifold to find local relationships between datapoints and has been effectively used for high-dimensional visualization [29]. However, t-SNE plots do not capture the global structure and have challenges in performing visualization of very large data set, yielding a suboptimal and potentially misleading visualization [30]. In addition, recent applications [31–35] of ML in Raman spectra classification have shown its effectiveness in discriminating those samples with desired high sensitivity and specificity. Nevertheless, little attention has been paid to develop algorithms for visualization of Raman data, especially for the spectra collected from the progression of biological process.

To overcome shortcomings from conventional algorithms, we introduce two new methods in the context of Raman spectra data analysis: a nonlinear dimensionality reduction and visualization method called Potential of Heat-diffusion for Affinity-based Transition Embedding (PHATE) [27], and the GENIE estimator of mutual information (MI) [36,37]. PHATE is well-suited to preserve progression in the data structure, such as the progression of stem cells into different cell types as measured with single-cell RNA-sequencing [27]. MI measures the general dependence of random variables without making any assumptions about the nature of their underlying relationships. MI was an effective feature selection

criterion when applied to a database extracted from a non-small cell lung cancer clinical dataset to effectively reduce the dimensions of a multidimensional time series for clinical data [38].

In our analysis of cellular Raman spectroscopy data, we use PHATE and MI in addition to conventional ML methods to investigate the effects of DEP exposure and the protectants RES and MesoBV on human lung cells. As the flow chart shows (Fig. 1), PCA, t-SNE and PHATE are used to visualize the spectral transition with DEP exposure time. We use the MI to identify the most informative Raman peaks to define a cell-health index for differentiating between the antioxidative effects of RES and MesoBV. Conventional classification algorithms (LDA, kNN, RF and SVM) are also performed.

2. Methods

2.1. Cell culture

A549 cells were purchased from ATCC (Manassas, VA, USA) and cultured in F-12k medium containing 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA, USA) at 37 °C with 5% CO₂ in a humidified atmosphere. DEPs (10 µg ml⁻¹) were mixed with 2 mL culture medium and vortexed for 10 s, and subsequently sonicated for 20 min at room temperature. A549 cells underwent pretreatment with plain medium or medium containing RES or MesoBV at 10 µM for 24 h. Subsequently, cells were treated with a DEP solution for 0 (control), 4, 8, 16, 24 or 48 h. MesoBV was produced from phycocyanobilin recovered from lyophilized powders of the cyanobacterium *Spirulina platensis* [25]. The quantification of RES from an anonymous commercial supplement product was performed by high performance liquid chromatography (HPLC, Fig. S1).

2.2. Raman spectroscopy

The Raman spectra were measured by a Renishaw inVia Raman spectrometer (controlled by WiRE 3.4 software, Renishaw, UK) connected to a Leica microscope (Leica DMLM, Leica Microsystems, Buffalo Grove, IL, USA), equipped with a 785 nm near-infrared (IR) laser which was focused through a 63 × NA = 0.90 water immersion objective (Leica Microsystems, USA). The standard calibration peak for the spectrometer with silicon mode at a static spectrum was 520.5 ± 0.1 cm⁻¹. A549 samples were cultured on MgF₂ (United Crystals Co., Port Washington, NY, USA) and imaged in Earle's

balanced salt solution (EBSS). Raman spectra between 600 and 1800 cm⁻¹ wavenumbers were recorded for 1 accumulation of 10 s laser exposure in static mode. After cells were pretreated with plain medium or medium containing RES or MesoBV (10 µM) for 24 h, DEPs were introduced into the 48hr exposure group (48 h prior to measurement), followed by the 24, 16, 8 and 4 h exposure groups (hours mean the time prior to measurement, respectively) with the control (0 h) group last. All spectra of the five exposure and one control groups were collected at the same time (within 30min). Four different treatments (Control, DEP, RES + DEP and MesoBV + DEP) were included in Raman measurements. Five points on each cell were randomly selected for measurements. 75 spectra (five points for each cell and fifteen cells) in each time point for the DEP group and 50 spectra (five points for each cell and ten cells) in each time point for other groups were collected for further analysis. Raw Raman spectra were first baseline corrected by the asymmetric least squares smoothing method in Origin 2018 (asymmetric factor: 0.001, threshold: 0.05, smoothing factor: 5, number of iterations: 10). One-way ANOVA was performed via Origin 2018 to examine results from the study groups with P < 0.05 set as indicating a statistically significant difference.

2.3. Conventional classification

PCA is commonly used for unsupervised dimensionality reduction while PLS is a supervised dimensionality reduction technique. Dimensionality reduction was performed using either PCA or PLS before applying other ML methods. PCA-LDA, PLS-LDA were processed in Origin 2018. PHATE, MI estimation, kNN, RF, and SVM were performed using Matlab (version: 2019A). RF were computed using Matlab's TreeBagger function (80 trees); fitcknn was used for kNN classification, using k-values between 1 and 15. More details on the machine learning algorithms are provided in the Supplementary materials.

2.4. PHATE

PHATE is an unsupervised ML algorithm designed for visualizing non-linear, high-dimensional data which accounts for both local and global relationships in the data; in addition, PHATE denoises the data and can visualize progression (such as changes over time) for appropriate data sets [27]. The detailed algorithm can be seen in Ref. [27]. The PHATE visualizations used default settings, 90%, 95% and 99% variance of PCA components, and the optimal *t*-diffusion parameter as computed via the algorithm using the von Neumann entropy. The von Neumann entropy is computed using values of *t* up to 100 (by default) and a “knee” point is used to determine the optimal *t*. The custom code to estimate mutual information was modified from the algorithm described in Ref. [36].

2.5. Mutual information

MI quantifies the level of information obtained about a random variable through observations of another random variable [39]. Thus, MI can be viewed as a generalization of the correlation coefficient where the strength of nonlinear relationships is also measured, whereas the correlation coefficient only measures the strength of linear relationships. MI has many applications, one of which is feature selection. We use MI in a feature selection context to determine which wavenumbers might be used as a healthy status indicator for cells. The GENIE estimator uses kernel density estimators (KDE) to estimate the MI between the wavenumbers (*X*) and the time of measurements (*Y*). We estimate the Rényi- α MI [36] using 10-fold cross validation and then rank the wavenumbers based on the estimates.

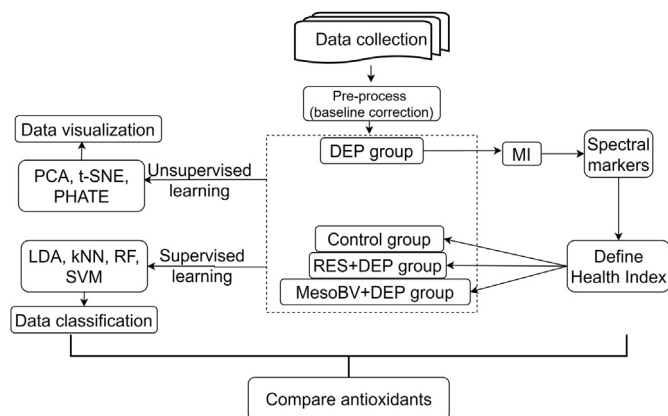


Fig. 1. Flow chart of the machine learning-based Raman spectra analysis process. PCA: principle component analysis; t-SNE: t-distributed stochastic neighbor embedding; PHATE [27]; LDA: linear discriminant analysis; kNN: k nearest neighbors; RF: random forests; SVM: support vector machine; MI: mutual information.

2.6. Cross-validation

PCA-LDA or PLS-LDA algorithms together with leave-one-out cross validation are well documented in the literature. For example, cross-validation was used in the diagnosis of nasopharyngeal cancers [40] and chronic lymphocytic leukemia [41] with satisfactory results. Here 10-fold cross validation was used for the MI analysis and the classifiers in Section 2.3. Raman spectra at each time point were randomly separated into ten groups with nine groups reserved for training data and one group (10% of total data) as testing data. Each group served as the test data once. In MI analysis, the top wavenumbers that showed up at least nine times in ten rotations were considered as the informative peaks. Higher MI corresponds to a stronger relationship. Using this MI algorithm, we are able to choose the number of top wavenumbers to view. In classification analysis, the accuracy across all tested classifiers are evaluated.

3. Results

3.1. Raman spectra in oxidative damage and antioxidant protection

Four groups of A549 cells including control without antioxidant nor DEP treatment, DEP, RES + DEP, and MesoBV + DEP were non-invasively measured by Raman spectroscopy at six DEP exposure time points (0, 4, 8, 16, 24 and 48 h). The averaged raw Raman spectra at different time points (Fig. 2) were initially pre-processed by baseline correction as shown in Fig. S2. The spectra are relatively constant across time in control groups with numerous peaks, indicating the cells were consistent at the molecular composition levels in normal culture systems. After the cells were exposed to DEPs without antioxidant pretreatment (Fig. 2D), the spectra pattern changed significantly with fewer characteristic peaks throughout the 48-hrs of DEP exposure. This clearly demonstrates that some important molecular structures in A549 cells were disrupted by the introduction of DEPs. In contrast, the spectra in RES + DEP and MesoBV + DEP groups (Fig. 2F&H) showed “recovered” characteristic peaks but in different spectra patterns. This may imply that the protective effect of RES and MesoBV on cellular composition are exerted in different ways.

3.2. Raman data visualization

Using PCA, t-SNE and PHATE, we performed data visualization on the Raman spectra to identify broader patterns in the data. The PCA plots mostly overlap from 0 h to 48 h in the Raman spectra of four treatment groups (Fig. 3A). The t-SNE plot shows a better cluster (Fig. 3B) but is not able to distinguish or visualize the trend of spectral changes, which contrasts with PHATE (Fig. 3C). In the PHATE plots (90% variance of PC components), the measurements from the DEP group (Fig. 3C, first row) tend to shift to the right branch of the graph as time progresses, while the measurements from the control group remain on the left branch of the scatterplot (Fig. 3C, second row). This suggests that information about the relative health of the cells is contained within the Raman spectra where the left branch represents healthy cells, and the right branch represents damaged cells. Although the axis is arbitrary and dependent on the particular data set used, a threshold of 0.01 (light blue plane) on the PHATE1 axis appears to be indicative of healthy cells. The percentage of healthy cells in the control group that were located left of this threshold throughout the whole period was closed to 100% (Fig. 3D), while cells exposed to DEPs tended to be unhealthy over time. The cells protected with RES or MesoBV displayed a different trend. The largest number of protected cells that appeared to recover to the healthy side of the threshold occurred at 16 h. At later time points, fewer protected cells are located on the healthy side. At 48 h, the percentage of healthy cells (as determined by this threshold) reached 54.34% in the RES + DEP group and 42.86% in MesoBV + DEP group, compared to 28.79% in DEP group. In addition, similar results were obtained using 95% or 99% variance of PC components (Figs. S3–4), suggesting that these antioxidants provide some protection against DEP exposure.

3.3. Raman feature extraction via mutual information estimation

To quantify the level of cell health during the 48 h of DEP exposure, principal components (from PCA) that explain 90% of variance were analyzed using LDA, but LDA was unable to accurately classify the 6 time points of spectra in the DEP group (Fig. S5A). We also analyzed the first 15 PLS factors that can explain 70% of the variance using LDA. In contrast, this approach shows

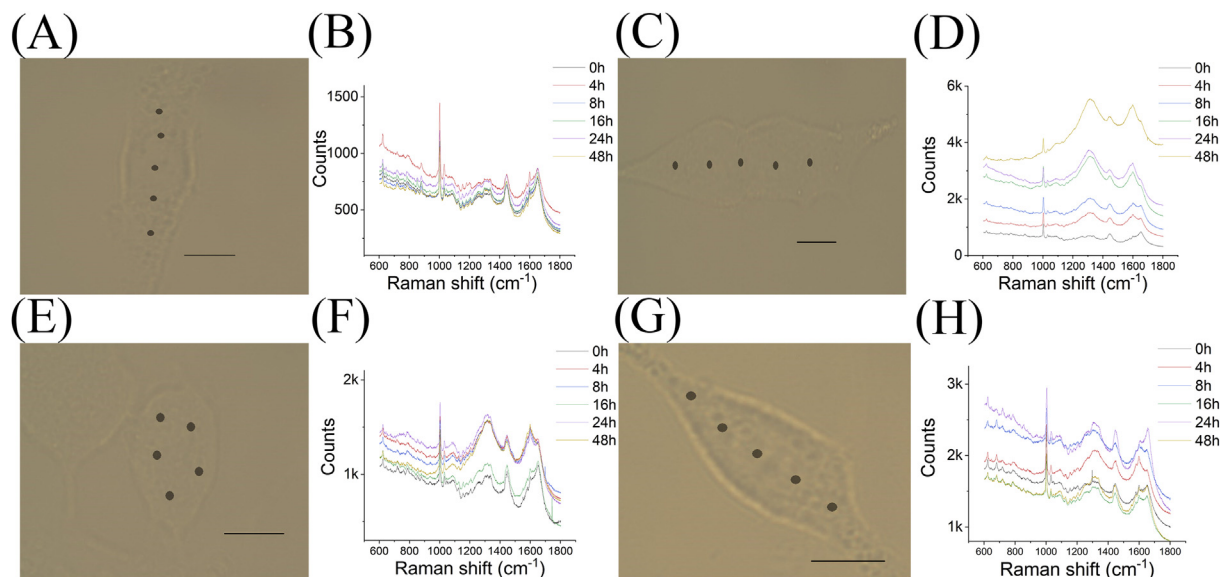


Fig. 2. Representative light images of single cells and corresponding averaged Raman spectra in the control (A–B), DEP (C–D), RES + DEP (E–F), and MesoBV + DEP groups (G–H). All spectra are averaged from at least 50 measurements at each time point. Black dots indicate the spots where five spectra per cell were performed. Scale bar = 10 μ m.

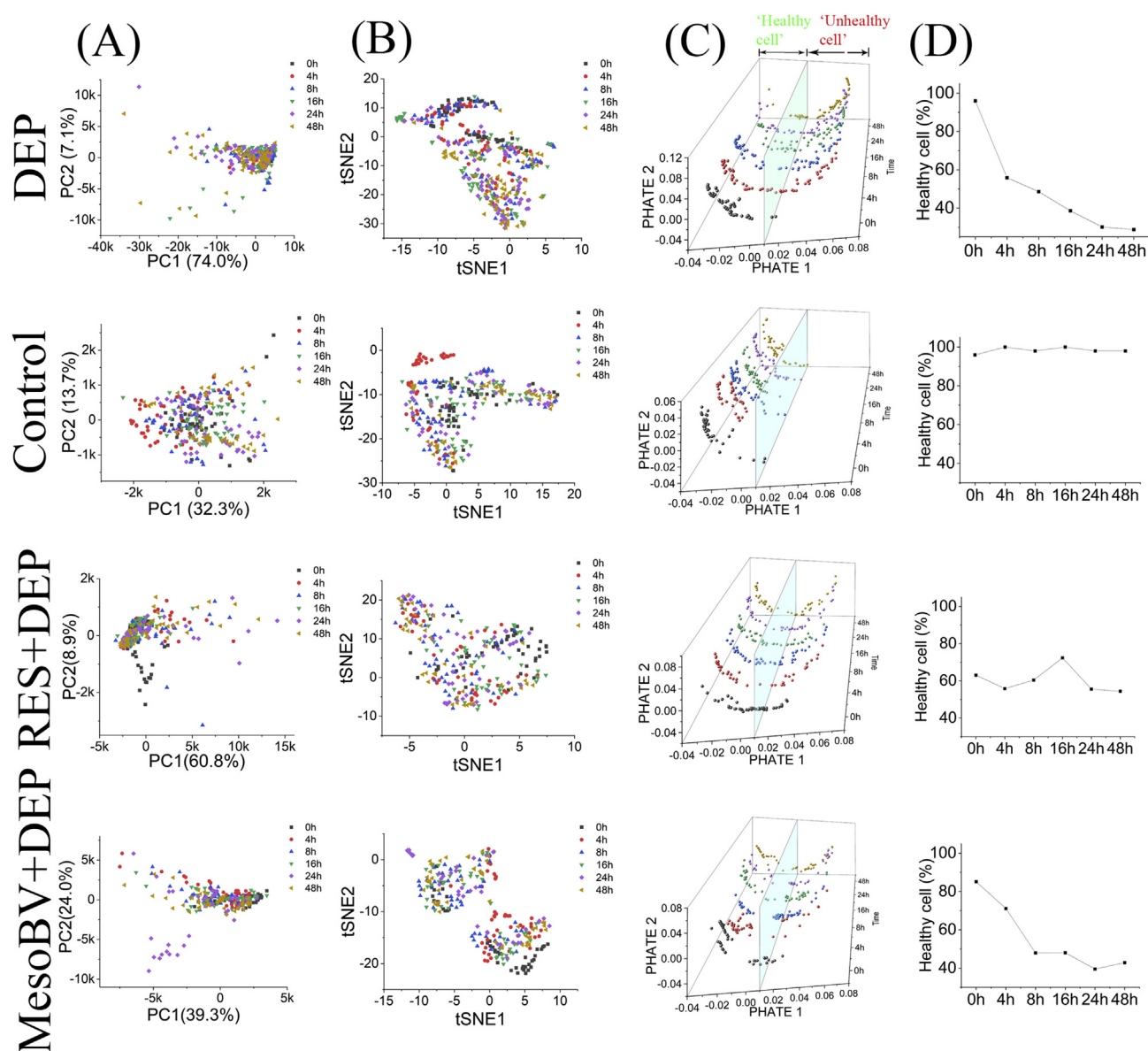


Fig. 3. Data visualization by PCA, t-SNE, and PHATE. Score plots of (A) PCA, (B) t-SNE and (C) PHATE analysis. The threshold 0.01 of PHATE 1 (light blue plane) was set as the criterion to separate healthy (left region) and unhealthy cells (right region). The percentage of healthy cells over time based on this threshold is shown in (D). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

distinguishable group clusters (Fig. S5B), corroborating the results obtained from PHATE. With increasing DEP exposure time, the clusters separate further, indicating gradually deteriorating cell structure during the 48-h period. However, PLS-LDA can only qualitatively discriminate each damaging group. To identify spectral markers that are most deterministic for indicating the cell health, we used the GENIE estimator for MI. We hypothesize that a suitable health index should decrease as DEP exposure time increases in the unprotected cells while remaining relatively constant in the control group over time. Therefore, we estimated the MI between each measured spectrum (in wavenumber range 600–1800 cm^{-1}) and DEP exposure time. A wavenumber with a large MI value has a stronger relationship with exposure time than a wavenumber with a smaller MI value (Table S1). The top wavenumbers in the DEP group were concentrated near the 1600 cm^{-1} peak (Fig. 4A), with a monotonic trend of elevated intensity in these wavenumbers as DEP exposure time increases (Fig. 4B).

It is interesting to note that the highest peak (1002 cm^{-1}) in each spectrum assigned to phenylalanine was shown as one of the least informative peaks (Fig. S6). Furthermore, a statistical analysis of the 1600 cm^{-1} peak in DEP groups at different time points showed that this peak increases intensity with the most significant differences ($P < 0.05$) among most time points (Fig. 4C). From this analysis, we conclude that 1600 cm^{-1} is well-suited as the spectral marker in our study to identify the cell health status in control or antioxidant protected groups. We therefore propose to define the health index (HI) as the following:

$$\text{Health index (HI)} = 1 - (\text{intensity of } 1600\text{cm}^{-1} - 200.838) / 903.101,$$

where the intensity of 1600 cm^{-1} peak is the highest signal (in counts) between the range of 1598–1602 cm^{-1} wavenumbers in

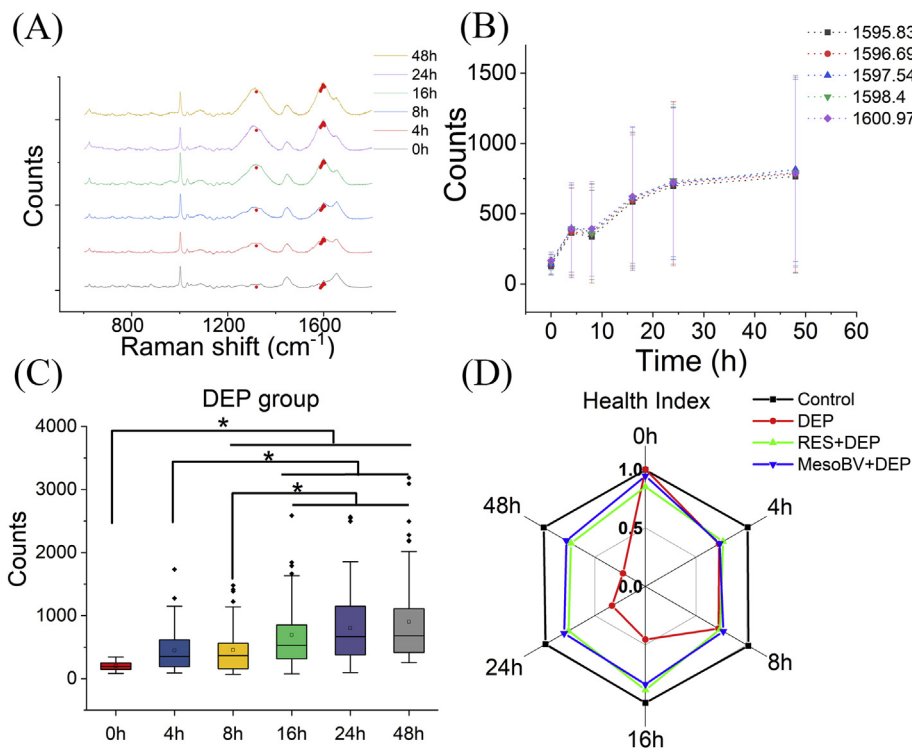


Fig. 4. Mutual Information analysis applied to the DEP group measurements: (A) top 15 wavenumbers based on the estimated mutual information noted by red spots on each spectrum; (B) variation of the top 5 wavenumbers over 48 h. (C) Statistical analysis of 1600 cm^{-1} peak in DEP groups at different time points. (D) Radar plot of the health index at each time point in all groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

each spectrum. 200.838 is the average intensity of the 1600 cm^{-1} peak at 0 h in the DEP group; 903.101 is the average intensity of the 1600 cm^{-1} peak at 48 h in the DEP group. By this definition, health index plots cross four treatments in terms of time point are shown in Fig. S7. In addition, a radar plot (Fig. 4D) clearly shows the HI trends in all four groups. Briefly, the HI values in the control group across all time points are near constant (very close to 1), running an outside clockwise circle from 0 h to 48 h . In DEP group, the circle falls into a “shrunk” cycle inside (unhealthy state); The antioxidant groups do show the circle expanding toward to outside circle of the control group (healthy state), indicating the protective effect from the antioxidants. In contrast, the HI values (Fig. 4D) exhibit a monotonic decreasing trend in the DEP group, indicating the cell damaging effect induced by DEP exposure. The protected groups have a similar HI value in the entire period.

3.4. Evaluation of health index in antioxidant protected cells

To determine the protective effects of each of the antioxidants, a statistical analysis among four groups at each time point were

performed (Table 1, Fig. S7). At 0 h , as expected, the difference in HI between the DEP and control groups (close to 1) is negligible. But both antioxidant pretreatments induced slight cell changes even without DEP exposure, as the RES + DEP group had an HI of 0.85 and the MesoBV + DEP group had an HI of 0.94 at 0 h . At 4 h and 8 h , there is no statistical difference (as the HI indicated) between the DEP and antioxidant treated groups. From 16 to 48 h , the cell health in the antioxidant treated groups (HI maintains 0.89 – 0.74) is profoundly better than that of the DEP groups (HI decreased from 0.45 to 0.22), showing a turning point at 16 h . There is no statistical difference between MesoBV + DEP group and RES + DEP group at most time points (excepting 0 h). In comparison to their own control group, the HI values in the RES + DEP group are significantly different across most time points (excepting 16 h), while the MesoBV + DEP group only show significant difference with the control group at 0 – 8 h . This implies that MesoBV may have a better protective effect against DEPs than RES at the same treatment concentration level. It is interesting to note that the PHATE visualization (Fig. 3) and the defined HI (Table 1) show the antioxidant protection effect at 16 h . However, it is unclear whether recovery at 16 h depends on DEP composition and/or DEP concentration.

3.5. Raman data classification

Conventional classification models were applied to evaluate the effect of DEPs and antioxidants on cell behavior. PCA (the number of components were chosen to explain 80% of variance) or PLS (components chosen to explain 80% of variance) were performed for dimensionality reduction. The results show a much clearer distinction of group clusters in PLS-LDA (Fig. S8) than those in PCA-LDA (Fig. S9). A number of other popular classification algorithms including RF, kNN and SVM (the basics of these algorithms are contained in Supplementary Materials) were also used to calculate

Table 1

Summary of the health index (HI) (mean \pm SD) at all time points across four treatment groups using 1600 cm^{-1} as the spectral marker.

Time (h)	Control	DEP	RES + DEP	MesoBV + DEP
0	1.0 ± 0.05	1.0 ± 0.07	0.85 ± 0.09^b	0.94 ± 0.10^b
4	1.0 ± 0.08	0.73 ± 0.37^b	0.77 ± 0.30^b	0.73 ± 0.31^b
8	1.0 ± 0.07	0.72 ± 0.39^b	0.74 ± 0.28^b	0.77 ± 0.18^b
16	1.0 ± 0.05	$0.45 \pm 0.55^{a,b}$	0.89 ± 0.16	0.84 ± 0.20
24	0.99 ± 0.08	$0.33 \pm 0.64^{a,b}$	0.76 ± 0.36^b	0.80 ± 0.21
48	1.0 ± 0.08	$0.22 \pm 0.76^{a,b}$	0.74 ± 0.36^b	0.78 ± 0.23

Note: ^a means significant difference versus DEP group at 0 h ($P < 0.05$). ^b means significant difference versus control group at same time point ($P < 0.05$).

the mean accuracy of classification at each time point (Table 2). The mean accuracy of classification at each time point were measured using 10-fold cross validation across each group. Two examples of calculation details for the accuracy for PLS-LDA and PLS-linear SVM are summarized in Tables S2–3. PLS-LDA provides a mean accuracy of about 75–85% at all time points, which is much better than PCA-LDA (around 60–72%). Compared to the LDA-based classifiers, the accuracy is generally improved for the kNN, RF and SVM classifiers. Among three different SVM kernels, the linear polynomial offered the best classification accuracy (all above 85%). Overall, the method with the highest accuracy is the kNN classifier, reaching around 90% in all time points.

3.6. Correlation analysis between whole Raman spectra and inflammatory responses

Furthermore, the inflammatory responses were assessed and correlatively analyzed with Raman spectra measurements. The typical Raman spectrum is shown in Fig. 5A (see detailed tentative assignments for these ten representative peaks in Table S4). The change in the mean difference of Raman peak intensity between treated and control cells in a given treatment group was calculated as D-value [42] and shown in Raman-barcode graph (Fig. 5B). As seen from the spectral barcode, the intensities of the 1313 (assigned to collagen/lipid) and 1600 cm^{-1} (assigned to protein) peaks clearly increased over time in the DEP group. It is interesting to note that the 1319 cm^{-1} identified by MI analysis is very close to the peak at 1313 cm^{-1} selected by Raman-barcode analysis. To test the effectiveness of using the 1313 cm^{-1} as another potential spectral marker, the peak counts at 1313 cm^{-1} are analyzed across all time points in the DEP group. A monotonic trend was also found (Fig. S10A). The health index (HI) values calculated using the 1313 cm^{-1} peak also showed the pattern of the changes across four treatment groups (Fig. S10B) that are visible in the 1600 cm^{-1} peak (Fig. 4D). The statistical analysis was summarized in Table S5. Overall, HI values calculated using 1313 or 1600 cm^{-1} show very similar results, except using 1313 cm^{-1} results in a relatively higher standard deviation. Thus, the 1600 cm^{-1} peak selected by both the MI algorithm and D-value calculation is an optimal peak that is used for our HI definition and calculation.

In addition to Raman spectra, the inflammatory responses of A549 cells were also assessed by human cytokine/chemokine panel (detailed methods in Supplementary Materials). Exposure of A549 cells to RES and MesoBV each in combination with DEP and DEP alone caused changes in the levels of pro-inflammatory (IL-6) [43] and inflammatory immunomodulatory cytokines and chemokines IL-1 β , IL-8, GRO α , I-309 [44–46] (Fig. S11). IL-1 β , GRO α , I-309 showed no or small elevations over 24 h following initial exposure to DEP. In all cases, MesoBV pretreatment caused significantly higher secretion rates of these cytokines or chemokines compared to DEP alone. IL-6 showed an initial decrease (4 h) upon DEP exposure and a subsequent gradual increase up to 48 h, but pretreatment with MesoBV resulted in IL-6 levels below those of DEP

alone at all time intervals measured. In contrast chemokine IL-8 was observed to increase after treatment with either RES or MesoBV. MesoBV elevated IL-8 higher than with RES across all time points except 24 h. Overall, RES and MesoBV exhibited different responses with IL-6 but similar responses with IL-8. These differences could be due to divergent regulatory signaling pathways. RES regulatory pathways involve PI3K/Akt and mitogen-activated protein kinase (MAPK) [47], while MesoBV pathways involve heme oxygenase 1, PKC- β II/TNF- α or MAPK/PI3K pathways [48]. RES pretreatment resulted in levels of IL-6, IL-8, IL-17, and GRO α , that exceeded levels with DEP alone at all time intervals up to 24 h, but no or little differences in levels of IL-1 β and I-309.

The averaged levels of cytokine/chemokines were paired with Raman spectra measurements, according to group and time points, to generate the 24×1300 matrix of correlation coefficients. The counter maps composed of Spearman correlations (Fig. 5C) demonstrates that most Raman peaks have relatively low correlation (-0.3 to 0.3) with cyto-/chemokine expression. Briefly, thymus and activation regulated chemokine (TARC), Regulated upon Activation, Normal T Cell Expressed and Presumably Secreted (RANTES), monocyte chemoattractant protein 2 (MCP-2) are most positively (-0.3 , red spots) correlated with wavenumber 678 cm^{-1} assigned to guanine (the circled region in Fig. 5C), followed by IFN γ , IL-23, IL-6, IL-4, IL-2. Whereas I-309 and MCP-1 are most negatively (~ -0.3) correlated with wavenumber 678 cm^{-1} . The Spearman correlation maps of individual groups are illustrated in Fig. S12, which are very similar to Pearson correlation (Fig. S13). In these correlation maps, it is interesting to note that the relationship between Raman spectra and cytokine/chemokine profiling exhibits the overall trend across three treatment groups: DEP group has more positive correlation (more red bars), followed by RES + DEP group showing more negative correlation (more blue bars) and MesoBV + DEP group (scattered red/blue bars), implying the difference in antioxidant activity in protecting DEP induced cytotoxicity.

4. Discussion

Raman spectroscopy is a universal technique for characterizing substances in life science, material research, and geological expeditions [49]. Univariate analysis of Raman spectra is typically not applicable, because of the overlapping of Raman peaks from different molecules. This shortcoming has been addressed by the application of numerical algorithms necessary for an accurate peak assignment and further emphasizing spectral variations among various treated samples [50]. A variety of ML algorithms [31–35,51,52] have been applied alongside Raman spectroscopy for this purpose and have gained popularity in life science applications. These methods basically read all of the spectra sequentially, extract only meaningful information from each spectrum, and then categorize the datasets based upon relevant information.

Our data shows the superior advantage of PHATE over PCA and t-SNE in visualizing Raman data, where PHATE is better able to identify the progression of cell behavior (Fig. 3). PCA scores capture

Table 2
The mean accuracy of different classification methods at each time point.

Time (h)	Classification methods						
	PCA-LDA	PLS-LDA	PLS-RF	PLS-kNN	PLS-SVM linear	PLS-SVM RBF	PLS-SVM linear polynomial
0	68.61%	74.62%	87.08%	90.18%	85.03%	91.23%	88.31%
4	72.51%	79.89%	85.86%	89.71%	89.86%	82.40%	85.01%
8	63.35%	84.26%	82.94%	87.45%	85.33%	77.81%	84.02%
16	63.65%	84.55%	90.74%	92.19%	90.65%	87.04%	89.64%
24	59.45%	70.27%	86.10%	91.07%	87.76%	77.49%	85.69%
48	65.33%	82.26%	83.42%	88.70%	88.55%	78.14%	85.26%

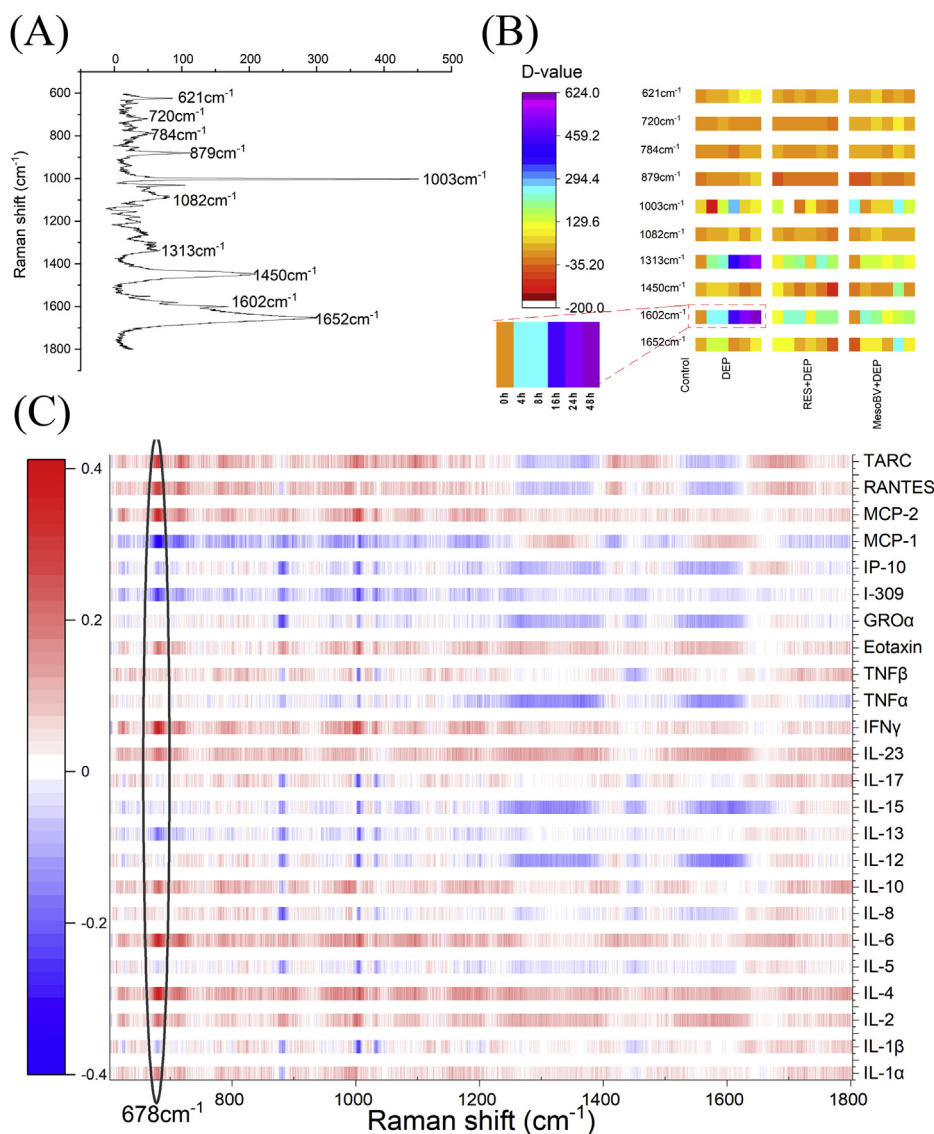


Fig. 5. Raman-barcode of cellular-response to stressors. (A) A representative Raman spectrum in the control group labeled with ten characteristic peaks, (B) D-value in three treatment groups (DEP, RES + DEP, MesoBV + DEP), and (C) Spearman correlation map to 24 cytokine/chemokines (including all groups and time points of Raman data). The change in Raman band intensity at particular wavenumbers were calculated as the difference (D-value) between treated and control cells.

the largest variance in the dataset (globally) but local relationships may remain hidden. However, *t*-SNE is based on a non-convex optimization problem which does not guarantee convergence to the global optimum. Additionally, the plotted *t*-SNE embeddings are sensitive to the tuning parameter and can be easily misinterpreted [53]. In contrast, PHATE produces a low-dimensional representation of high-dimensional data which preserves both the local and global structure. In addition, PHATE may be used to model transitions over time or space which is more difficult to achieve using PCA or *t*-SNE [27].

Among the numerous classification algorithms, kNN shows the best accuracy of classification in our Raman data (Table 2). For a given problem, the best algorithm may be affected by, among many other factors, the variance among explanatory variables, the number of explanatory variables, the number of noisy explanatory variables, and the number of observations [54]. LDA, as a linear classifier, makes classification decisions based on linear combinations of the training data. A major advantage of such methods is speed: given certain data, they can make comparable classification

accuracy as non-linear methods in a fraction of the time. However, if the data is not linearly separable, lower accuracy may be attained [55]. SVM finds the optimal soft-margin separating hyperplane to learn classification boundaries. The hyperplane may be constructed in the original data space or may be kernelized to better separate data which is not linearly separable. SVM relies on only a few support vectors and is thus not sensitive to outliers. Non-linear classifiers such as kNN and RF are commonly-used as non-parametric models which are likely to outperform linear models on data sets which are not linearly separable [55]. kNN and RF deal well with outliers and noisy data, which may have played a significant role in their high classification accuracy [56], especially in biological data, which is inherently noisy. Neither of these classifiers make any assumptions about the model by which the data was generated and are thus adapted to many datasets.

None of these above-mentioned classification algorithms quantitatively differentiate the cell behaviors in different treatments. We used MI estimation, which gives a quantitative measure of variable importance in Raman spectra analysis to successfully quantify the

cell health index in all situations. The MI analysis identified 1600 cm^{-1} as the representative peak of cell health, which is assigned to aromatic ring stretch ($\text{C}=\text{C}$), often referred as phenylalanine (protein). Raman spectral peaks are widely used to monitor the cell toxicity, as summarized in Table S6. Most peaks were reported to decrease as cells died, such as 788, 1320, 1342 and 1660 cm^{-1} . Some representative peaks have controversial responses to toxic treatment, including 1005 , 1231 cm^{-1} . Interestingly, two major peaks, 642 (assigned to nucleic acid) and $1395\text{--}1425\text{ cm}^{-1}$ (assigned to lipid), were found to increase in dead cells, implying the disordered structure of these macromolecules in unhealthy cells. It may not be always true that the 1600 cm^{-1} peak can serve as a cell health marker in other settings, but a similar MI analysis can help to identify the most valuable peak in a particular situation.

The ML methods used in this work enabled us to evaluate the protective effects of the two tested antioxidants. The classification analysis qualitatively shows the different effects of RES and MesoBV against DEPs (Figs. S8–9). PHATE (Fig. 3) shows an overall progression of cell status changes overtime. More importantly, our proposed health index shows a recovery from DEP damage with antioxidant protection that is well-maintained over a period of 48 h. The statistical analysis based on our HI suggests that MesoBV performs slightly better than RES at the same treatment concentration ($10\text{ }\mu\text{M}$). Both 1600 and 1313 cm^{-1} peak were identified by MI algorithm and D-value calculation, although 1600 cm^{-1} showed better performance than 1313 cm^{-1} as spectral marker. When combined with biological assessments, cell viability of all groups was not affected while both antioxidants show a significant inhibition of ROS production at all time points (Fig. S14) and promotion of multiple cytokine/chemokine responses (Fig. S11) against DEP exposure. Under our current experimental conditions, it seems the ROS production is not time-dependent (Fig. S14D). Similar phenomena were reported for other antioxidants against effects of a variety of toxic agents [57–59]. Our data show that Raman spectroscopic measurement of molecular fingerprints of cells exposed to oxidative stress (where the majority of the cells are alive) is more sensitive than ROS production assays (where the stained cells are dead). Further *in vitro* cell-based studies, such as dependence of concentration (DEP and antioxidant) and cell type (primary lung cells versus lung cancer cell) on cellular responses, are necessary to determine the health benefit of antioxidant-based supplement products as potential pharmaceuticals against air pollution.

5. Conclusion

We employed multiple ML algorithms to analyze high-volume Raman spectra data for different purposes. PHATE outperforms PCA and t-SNE, visualizing well the progression of cell damage by DEP exposure and indicates the trend of recovery of cell health in the presence of antioxidant. MI analysis identifies the most informative Raman peak that is used as a spectral marker to quantitatively predict the cell damage and antioxidant protection. This further demonstrates the possibility of the selection of spectral markers (defined as a health index in this study) to prescreen the protective effectiveness of various antioxidants, such as MesoBV and RES, against DEP exposure. Conventional classification algorithms including LDA, kNN, random forest, and SVM were used for discrimination among four treatment groups at the same time point. Additionally, the correlation analyses between Raman data and cytokine/chemokine responses (as immunomodulatory biomarkers) provides a proof-of-concept application, which may allow a deeper understanding of overall cellular response. Our findings provide a combination of Raman spectroscopy and ML tools, which can serve as a high throughput platform in screening various potential antioxidants in *in vitro* anti-toxicity studies.

CCrediT authorship contribution statement

Wei Zhang: Formal analysis, Writing - original draft. **Jake S. Rhodes:** Formal analysis, Writing - original draft. **Ankit Garg:** Investigation. **Jon Y. Takemoto:** Supervision, Writing - review & editing. **Xiaojun Qi:** Formal analysis. **Sitaram Harihar:** Conceptualization. **Cheng-Wei Tom Chang:** Conceptualization. **Kevin R. Moon:** Supervision. **Anhong Zhou:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank Yi Chen and Dr. Jixun Zhan from Utah State University for their help in HPLC measurements. We also would like to thank Dr. M. Ian Gilmour from the National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency (Research Triangle Park, NC) for generously providing DEP. Lastly, we would also like to thank the partial supports from Utah Agricultural Experiment Station.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.aca.2020.06.074>.

References

- [1] A. Ghorani-Azam, B. Riahi-Zanjani, M. Balali-Mood, Effects of air pollution on human health and practical measures for prevention in Iran, *J. Res. Med. Sci.* 21 (2016) 65.
- [2] A.J. Ghio, J.R. Sobus, J.D. Pleil, M.C. Madden, Controlled human exposures to diesel exhaust, *Swiss Med. Wkly.* 142 (2012) w13597.
- [3] A.J. Cohen, C.A. Pope 3rd, Lung cancer and air pollution, *Environ. Health Perspect.* 103 (Suppl 8) (1995) 219–224.
- [4] K.E. Cosselman, R.M. Krishnan, A.P. Oron, K. Jansen, A. Peretz, J.H. Sullivan, T.V. Larson, J.D. Kaufman, Blood pressure response to controlled diesel exhaust exposure in human subjects, *Hypertension* 59 (2012) 943–948.
- [5] S. Levesque, T. Taetzsch, M.E. Lull, U. Kodavanti, K. Stadler, A. Wagner, J.A. Johnson, L. Duke, P. Kodavanti, M.J. Surace, M.L. Block, Diesel exhaust activates and primes microglia: air pollution, neuroinflammation, and regulation of dopaminergic neurotoxicity, *Environ. Health Perspect.* 119 (2011) 1149–1155.
- [6] M. Wu, J.J. Ries, E. Proietti, D. Vogt, S. Hahn, I. Hoesli, Development of late-onset preeclampsia in association with road densities as a proxy for traffic-related air pollution, *Fetal Diagn. Ther.* 39 (2016) 21–27.
- [7] B. Fredricsson, L. Moller, A. Pousette, R. Westerholm, Human sperm motility is affected by plasticizers and diesel particle extracts, *Pharmacol. Toxicol.* 72 (1993) 128–133.
- [8] C.Y. Tseng, J.S. Wang, Y.J. Chang, J.F. Chang, M.W. Chao, Exposure to high-dose diesel exhaust particles induces intracellular oxidative stress and causes endothelial apoptosis in cultured *in vitro* capillary tube cells, *Cardiovasc. Toxicol.* 15 (2015) 345–354.
- [9] J.S. Wang, C.Y. Tseng, M.W. Chao, Diesel exhaust particles contribute to endothelial apoptosis via autophagy pathway, *Toxicol. Sci.* 156 (2017) 72–83.
- [10] J.R. Gillette, On the role of pharmacokinetics in integrating results from *in vivo* and *in vitro* studies, *Food Chem. Toxicol.* 24 (1986) 711–720.
- [11] P.H. Danielsen, S. Loft, P. Moller, DNA damage and cytotoxicity in type II lung epithelial (A549) cell cultures after exposure to diesel exhaust and urban street particles, *Part. Fibre Toxicol.* 5 (2008) 6.
- [12] R.M. Lasagni Vitar, J. Tau, N.S. Janzic, A.I. Tesone, A.G. Hvozda Arana, C.G. Reides, A. Berra, S.M. Ferreira, S.F. Llesuy, Diesel exhaust particles (DEP) induce an early redox imbalance followed by an IL-6 mediated inflammatory response on human conjunctival epithelial cells, *Exp. Eye Res.* 171 (2018) 37–47.
- [13] H.J. Butler, L. Ashton, B. Bird, G. Cinque, K. Curtis, J. Dorney, K. Esmonde-White, N.J. Fullwood, B. Gardner, P.L. Martin-Hirsch, M.J. Walsh, M.R. McAinsh, N. Stone, F.L. Martin, Using Raman spectroscopy to characterize biological materials, *Nat. Protoc.* 11 (2016) 664–687.
- [14] H.J. Byrne, P. Knief, M.E. Keating, F. Bonnier, Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells, *Chem. Soc.*

- Rev. 45 (2016) 1865–1878.
- [15] J.M. Surmacki, B.J. Woodhams, A. Haslehurst, B.A.J. Ponder, S.E. Bohndiek, Raman micro-spectroscopy for accurate identification of primary human bronchial epithelial cells, *Sci. Rep.* 8 (2018) 12604.
 - [16] R. Smith, K.L. Wright, L. Ashton, Raman spectroscopy: an evolving technique for live cell studies, *Analyst* 141 (2016) 3590–3600.
 - [17] W. Zhang, Q. Li, M. Tang, H. Zhang, X. Sun, S. Zou, J.L. Jensen, T.G. Liou, A. Zhou, A multi-scale approach to study biochemical and biophysical aspects of resveratrol on diesel exhaust particle-human primary lung cell interaction, *Sci. Rep.* 9 (2019) 18178.
 - [18] H. Huang, H. Shi, S. Feng, W. Chen, Y. Yu, D. Lin, R. Chen, Confocal Raman spectroscopic analysis of the cytotoxic response to cisplatin in nasopharyngeal carcinoma cells, *Anal. Methods* 5 (2013) 260–266.
 - [19] E. Efeoglu, A. Casey, H.J. Byrne, Determination of spectral markers of cytotoxicity and genotoxicity using in vitro Raman microspectroscopy: cellular responses to polyamidoamine dendrimer exposure, *Analyst* 142 (2017) 3848–3856.
 - [20] M. Kfoury, M. Borgia, A. Verdin, F. Ledoux, D. Courcot, L. Auezova, S. Fourmentin, Essential oil components decrease pulmonary and hepatic cells inflammation induced by air pollution particulate matter, *Environ. Chem. Lett.* 14 (2016) 345–351.
 - [21] N. Xia, A. Daiber, U. Forstermann, H. Li, Antioxidant effects of resveratrol in the cardiovascular system, *Br. J. Pharmacol.* 174 (2017) 1633–1646.
 - [22] C. Sergides, M. Chirila, L. Silvestro, D. Pitta, A. Pittas, Bioavailability and safety study of resveratrol 500 mg tablets in healthy male and female volunteers, *Exp. Ther. Med.* 11 (2016) 164–170.
 - [23] A.C. Bulmer, K. Ried, J.T. Blanchfield, K.H. Wagner, The anti-mutagenic properties of bile pigments, *Mutat. Res.* 658 (2008) 28–41.
 - [24] J.Y. Takemoto, C.-W.T. Chang, D. Chen, G. Hinton, Heme-derived bilins, *Isr. J. Chem.* 59 (2019) 378–386.
 - [25] T. Ito, D. Chen, C.W. Chang, T. Kenmochi, T. Saito, S. Suzuki, J.Y. Takemoto, Mesobiliverdin IXalpha enhances rat pancreatic islet yield and function, *Front. Pharmacol.* 4 (2013) 50.
 - [26] R. Gautam, S. Vanga, F. Ariese, S. Umapathy, Review of multidimensional data processing approaches for Raman and infrared spectroscopy, *EPJ Techn. Instrument.* 2 (2015) 8.
 - [27] K.R. Moon, D. van Dijk, Z. Wang, S. Gigante, D.B. Burkhardt, W.S. Chen, K. Yim, A.V.D. Elzen, M.J. Hirn, R.R. Coifman, N.B. Ivanova, G. Wolf, S. Krishnaswamy, Visualizing structure and transitions in high-dimensional biological data, *Nat. Biotechnol.* 37 (2019) 1482–1492.
 - [28] K.R. Moon, J.S. Stanley, D. Burkhardt, D. van Dijk, G. Wolf, S. Krishnaswamy, Manifold learning-based methods for analyzing single-cell RNA-sequencing data, *Curr. Opin. Struct. Biol.* 7 (2018) 36–46.
 - [29] W. Kong, S.A. Morris, Evaluation of Wu et al.: comprehending Global and Local Structure of Single-Cell Datasets, *Cell Syst.* 7 (2018) 565–566.
 - [30] D. Kobak, P. Berens, The art of using t-SNE for single-cell transcriptomics, *Nat. Commun.* 10 (2019) 5416.
 - [31] M. Muratore, Raman spectroscopy and partial least squares analysis in discrimination of peripheral cells affected by Huntington's disease, *Anal. Chim. Acta* 793 (2013) 1–10.
 - [32] Q. Mai, A review of discriminant analysis in high dimensions, *Wiley Interdiscipl. Rev.: Comput. Stat.* 5 (2013) 190–197.
 - [33] Q. Li, W. Li, J. Zhang, Z. Xu, An improved k-nearest neighbour method to diagnose breast cancer, *Analyst* 143 (2018) 2807–2811.
 - [34] R. Ullah, S. Khan, S. Javaid, H. Ali, M. Bilal, M. Saleem, Raman spectroscopy combined with a support vector machine for differentiating between feeding male and female infants mother's milk, *Biomed. Optic Express* 9 (2018) 844–851.
 - [35] S. Khan, R. Ullah, A. Khan, A. Sohail, N. Wahab, M. Bilal, M. Ahmed, Random forest-based evaluation of Raman spectroscopy for dengue fever analysis, *Appl. Spectrosc.* 71 (2017) 2111–2117.
 - [36] K.R. Moon, K. Sricharan, A.O. Hero, Ensemble Estimation of Mutual Information, 2017 IEEE International Symposium on Information Theory, (ISIT), 2017, pp. 3030–3034.
 - [37] K.R. Moon, K. Sricharan, A.O. Hero, Ensemble Estimation of Generalized Mutual Information with Applications to Genomics, arXiv, 2017.
 - [38] L. Fang, H. Zhao, P. Wang, M. Yu, J. Yan, W. Cheng, P. Chen, Feature selection method based on mutual information and class separability for dimension reduction in multidimensional time series for clinical data, *Biomed. Signal Process Contr.* 21 (2015) 82–89.
 - [39] R. Vicente, M. Wibral, M. Lindner, G. Pipa, Transfer entropy—a model-free measure of effective connectivity for the neurosciences, *J. Comput. Neurosci.* 30 (2011) 45–67.
 - [40] Y. Li, W. Huang, J. Pan, Q. Ye, S. Lin, S. Feng, S. Xie, H. Zeng, R. Chen, Rapid detection of nasopharyngeal cancer using Raman spectroscopy and multivariate statistical analysis, *Mol. Clin. Oncol.* 3 (2015) 375–380.
 - [41] M. Fere, C. Gobinet, L.H. Liu, A. Beljebbar, V. Untereiner, D. Gheldof, M. Chollat, J. Klossa, B. Chatelain, O. Piot, Implementation of a classification strategy of Raman data collected in different clinical conditions: application to the diagnosis of chronic lymphocytic leukemia, *Anal. Bioanal. Chem.* 412 (2020) 949–962.
 - [42] L. Teng, X. Wang, X. Wang, H. Gou, L. Ren, T. Wang, Y. Wang, Y. Ji, W.E. Huang, J. Xu, Label-free, rapid and quantitative phenotyping of stress response in *E. coli* via ramanome, *Sci. Rep.* 6 (2016) 34359.
 - [43] G. Mazzarella, F. Ferraraccio, M.V. Prati, S. Annunziata, A. Bianco, A. Mezzogiorno, G. Liguori, I.F. Angelillo, M. Cazzola, Effects of diesel exhaust particles on human lung epithelial cells: an in vitro study, *Respir. Med.* 101 (2007) 1155–1162.
 - [44] S.M. Opal, V.A. DePalo, Anti-inflammatory cytokines, *Chest* 117 (2000) 1162–1172.
 - [45] L. Tao, Q. Ding, C. Gao, X. Sun, Resveratrol attenuates neuropathic pain through balancing pro-inflammatory and anti-inflammatory cytokines release in mice, *Int. Immunopharm.* 34 (2016) 165–172.
 - [46] P.E. Schwarze, A.I. Totlandsdal, M. Lag, M. Refsnes, J.A. Holme, J. Ovreivik, Inflammation-related effects of diesel engine exhaust particles: studies on lung cells in vitro, *BioMed Res. Int.* 2013 (2013) 685142.
 - [47] S.S. Kulkarni, C. Canto, The molecular targets of resveratrol, *Biochim. Biophys. Acta* 1852 (2015) 1114–1123.
 - [48] T. Jansen, A. Daiber, Direct antioxidant properties of bilirubin and biliverdin. Is there a role for biliverdin reductase? *Front. Pharmacol.* 3 (2012) 30.
 - [49] J. Liu, M. Osadchy, L. Ashton, M. Foster, C.J. Solomon, S.J. Gibson, Deep convolutional neural networks for Raman spectrum recognition: a unified solution, *Analyst* 142 (2017) 4067–4074.
 - [50] S. Khan, R. Ullah, S. Shahzad, N. Anbreen, M. Bilal, A. Khan, Analysis of tuberculosis disease through Raman spectroscopy and machine learning, *Photodiagnosis Photodyn. Ther.* 24 (2018) 286–291.
 - [51] S. Khan, R. Ullah, S. Javaid, S. Shahzad, H. Ali, M. Bilal, M. Saleem, M. Ahmed, Raman spectroscopy combined with principal component analysis for screening nasopharyngeal cancer in human blood sera, *Appl. Spectrosc.* 71 (2017) 2497–2503.
 - [52] L. Tang, S. Peng, Y. Bi, P. Shan, X. Hu, A new method combining LDA and PLS for dimension reduction, *PLoS One* 9 (2014), e96944.
 - [53] M. Wattenberg, F. Viégas, I. Johnson, How to Use T-SNE Effectively, *Distill*, 2016.
 - [54] K. Kirasich, T. Smith, B. Sadler, Random forest vs logistic regression: binary classification for heterogeneous datasets, *SMU Data Sci. Rev.* 1 (2018) 9.
 - [55] B.H. Nayef, A comparison between linear and non-linear machine learning classifiers, *J. Al-Nahrain Univ. Sci.* 19 (2016) 145–153.
 - [56] J.A. Sáez, M. Galar, J. Luengo, F. Herrera, Tackling the problem of classification with noisy data using Multiple Classifier Systems: analysis of the performance and robustness, *Inf. Sci.* 247 (2013) 1–20.
 - [57] J.F. Lin, T.F. Tsai, S.C. Yang, Y.C. Lin, H.E. Chen, K.Y. Chou, T.I. Hwang, Benzyl isothiocyanate induces reactive oxygen species-initiated autophagy and apoptosis in human prostate cancer cells, *Oncotarget* 8 (2017) 20220–20234.
 - [58] A.C. Ruberte, S. Ramos-Inza, C. Aydllo, I. Talavera, I. Encio, D. Plano, C. Sanmartín, Novel N,N'-Disubstituted acylselenoureas as potential antioxidant and cytotoxic agents, *Antioxidants* 9 (2020) 55.
 - [59] G. Luis, M.A. Martin, R. Sonia, M. Raquel, B. Laura, A cell culture model for the assessment of the chemopreventive potential of dietary compounds, *Curr. Nutr. Food Sci.* 5 (2009) 56–64.