

Quantitative scoring of epithelial and mesenchymal qualities of cancer cells using machine learning and quantitative phase imaging

Van K. Lam
Thanh Nguyen
Vy Bui
Byung Min Chung
Lin-Ching Chang
George Nehmetallah
Christopher B. Raub

Quantitative scoring of epithelial and mesenchymal qualities of cancer cells using machine learning and quantitative phase imaging

Van K. Lam,^a Thanh Nguyen,^b Vy Bui,^b Byung Min Chung,^c
Lin-Ching Chang,^b George Nehmetallah,^b and Christopher B. Raub^{a,*}

^aThe Catholic University of America, Department of Biomedical Engineering,
Washington, DC, United States

^bThe Catholic University of America, Department of Electrical Engineering and
Computer Science, Washington, DC, United States

^cThe Catholic University of America, Department of Biology, Washington, DC, United States

Abstract

Significance: We introduce an application of machine learning trained on optical phase features of epithelial and mesenchymal cells to grade cancer cells' morphologies, relevant to evaluation of cancer phenotype in screening assays and clinical biopsies.

Aim: Our objective was to determine quantitative epithelial and mesenchymal qualities of breast cancer cells through an unbiased, generalizable, and linear score covering the range of observed morphologies.

Approach: Digital holographic microscopy was used to generate phase height maps of noncancerous epithelial (Gie-No3B11) and fibroblast (human gingival) cell lines, as well as MDA-MB-231 and MCF-7 breast cancer cell lines. Several machine learning algorithms were evaluated as binary classifiers of the noncancerous cells that graded the cancer cells by transfer learning.

Results: Epithelial and mesenchymal cells were classified with 96% to 100% accuracy. Breast cancer cells had scores in between the noncancer scores, indicating both epithelial and mesenchymal morphological qualities. The MCF-7 cells skewed toward epithelial scores, while MDA-MB-231 cells skewed toward mesenchymal scores. Linear support vector machines (SVMs) produced the most distinct score distributions for each cell line.

Conclusions: The proposed epithelial–mesenchymal score, derived from linear SVM learning, is a sensitive and quantitative approach for detecting epithelial and mesenchymal characteristics of unknown cells based on well-characterized cell lines. We establish a framework for rapid and accurate morphological evaluation of single cells and subtle phenotypic shifts in imaged cell populations.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JBO.25.2.026002](https://doi.org/10.1117/1.JBO.25.2.026002)]

Keywords: holography; quantitative phase; machine learning; epithelial; mesenchymal; cancer cells; support vector machine.

Paper 190339R received Sep. 25, 2019; accepted for publication Jan. 30, 2020; published online Feb. 18, 2020.

1 Introduction

Quantitative phase imaging (QPI) is a label-free optical imaging technique that measures the phase delay introduced when a coherent laser beam travels through a thin transparent specimen, such as cells.¹ The uncorrected optical pathlength yields information about cell morphology and geometric thickness of the specimen along its phase projection, as well as fluctuations in the

*Address all correspondence to Christopher B. Raub, E-mail: raubc@cua.edu

local index of refraction.² Reconstructed optical phase maps that result from QPI of cells possess phase texture in the variation of pixel values within the boundary of the cell,³ quantified through texture parameters.⁴ The pixel information from phase maps is relevant to cell structure and function. For example, a cell's refractive index can reveal its total protein concentration,⁵ organization, and distribution of subcellular organelles, which are distinctive for certain biological phenotypes.⁶ Hence, a collection of quantitative parameters from optical phase maps constitutes a cell type's phase signature, providing additional information beyond the cell shape.

Digital holographic microscopy (DHM) is one type of QPI that uses the holography principle to determine the amplitude and phase of the diffraction wavefront^{7,8} and is well-suited for imaging biological specimens. Telecentric DHM is able to optically compensate for phase aberrations introduced by an imaging objective, making the computational removal of additional optical aberrations easier.² Other key features of DHM include noninvasive imaging with low power density at the specimen, high temporal resolution, and rich quantitative pixel information. Recent applications of DHM to the assessment of living biological specimens include characterization of the global morphology of confluent cell layers,⁹ analysis of cell proliferation and morphology on various substrates,^{1,3,10} cell responses to drugs,^{11,12} determination of phase features relating to cell motility,^{4,13} and cell classification in flow cytometry.^{14,15}

Machine learning applications to QPI include rapid evaluation and classification of cell types and (patho)physiological states^{4,14-20} and improvements in reconstructed image quality.^{21,22} Phase texture from pixels, shape features of adherent cells,⁴ and dry mass/volumetric determination of cells in liquid suspension¹⁵ are quantitative parameters that served as predictors for binary classification of cancer and noncancerous cells. Support vector machines (SVMs) have been particularly successful in classification of cell lines using such quantitative phase parameters.^{4,15,23,24} Ensemble methods train multiple weak learners and then combine them to obtain better predictive performance for classification. For example, the Boosting ensemble method trains learners sequentially, focusing subsequent models on the previous models' misclassifications, while the Bagging ensemble method trains learners independently. A recent study suggested that ensemble methods could improve detection of clear cell renal cell carcinoma in kidney disease leading to improved diagnosis and treatment.²⁵

QPI has great potential to evaluate cells in thin sections and in cell-based screening assays. For example, machine learning classification from QPI compared favorably to manual scoring of the Gleason grade of prostate cancer from histology sections²⁶ or to conventional screening in terms of predicting pathological features of hematological diseases.²⁷ The use of parameters from QPI has recently been explored to assess shifts in population distributions of cell phase parameters, indicating altered phenotype, or to differentiate multiple bacteria species based on their single-cell profiling capability.^{28,29} The effects of cell seeding density,⁴ exposure to anti-cancer drugs,^{30,31} and other influences on cell phenotype³²⁻³⁴ have been robustly evaluated with QPI. Quantitative imaging and machine learning have the potential to save time, labor, and reduce human error in phenotypic profiling, which could help pathologists and scientists to accurately detect circulating tumor cells,³⁵ classify cancer cells,^{36,37} evaluate the metastatic potential of cancer cells,³⁸ and assess cancer drug resistance.³⁹ Thus, machine learning-assisted QPI has great power to aid in interpreting large-scale and high-dimensionality data from cells, potentially enhancing cancer diagnosis and treatment.

A key aspect of cancer relevant to disease outcomes is cancer cell morphology. Many cancers adopt either "epithelial" or "mesenchymal" morphologies, dependent on certain gene mutations, gene expression profiles influenced by the microenvironment, and epigenetic changes.⁴⁰ Indeed, the route to transformation for many precancers involves epithelial-to-mesenchymal transition in which cells switch from a quiescent phenotype with rounded morphology to an actively motile, invasive phenotype with elongated morphology.^{41,42} Complicating this picture, some cancer cell lines, such as MCF-7 cells, are rounded and form aggregates *in vitro* and yet are more invasive than cancer cell lines with single, elongated cell morphologies.⁴³ Another well-studied breast cancer cell line, MDA-MB-231, adopts elongated, mesenchymal, and rounded amoeboid morphologies as a bimodal invasion strategy to overcome microenvironmental barriers.⁴⁴ In previous studies, SVMs were used to classify rounded and elongated MDA-MB-231 cells³ and distinguish MCF-7 and MDA-MB-231 cells from noncancerous epithelial and mesenchymal cell

lines.⁴ These studies raised the question of whether a universal score could be developed to grade cells along the spectrum of epithelial to mesenchymal features.

Since results from previous studies classified cells based on textural and shape-based phase map features, we hypothesized that a quantitative score from machine learning algorithms trained on noncancerous epithelial and mesenchymal cell lines could be used to assign mesenchymal or epithelial morphological status to cancer cells. To test this hypothesis, a binary classifier of two noncancerous gingival cell lines, one epithelial and one fibroblast/mesenchymal, was evaluated. Then the algorithm trained on noncancerous cells was applied to two cancer cell lines of mixed morphology and an “epithelial–mesenchymal” (EM) score was derived. Results indicate that such an approach accurately classifies epithelial and mesenchymal cell lines and assigns cancer cells a phenotypic score on the EM axis consistent with observed morphology. We propose this approach of deriving morphological phenotypic scores from machine learning on archetypal cells as a generally useful and robust way to assess phenotypic characteristics of unknown cell populations and single cells, which holds promise for future clinical and research applications.

2 Materials and Methods

2.1 Cell Culture

Cell culture procedures were the same as in Ref. 4. For DHM imaging, cells were passaged when reaching 80% to 90% confluence and seeded on glass-bottomed Petri dishes. Immortalized human gingival keratinocytes (Gie-No3B11, abbreviated as GIE, derived from buccal gingiva),⁴⁵ immortalized human gingival fibroblasts (HGF, derived from American Type Culture Collection CRL-2014 primary gingival cells),^{46,47} and the breast cancer cell lines MCF-7⁴⁸ and MDA-MB-231,⁴⁹ both adenocarcinomas derived from pleural effusions, were seeded at respective densities of 60,000; 40,000; 40,000; and 30,000 cells in a 35-mm-diameter glass-bottomed Petri dish (Part #229632, CELLTREAT Scientific Products, Pepperell, Massachusetts). The different densities were estimated to produce a roughly equal number of cells per field of view after 24 h due to differences in growth rates and aggregation. Cancer cell lines were fed with Dulbecco’s modified Eagle’s medium (Lot # SLBW4140, Sigma-Aldrich, St. Louis, Missouri), supplemented with 10% Fetalgro (Rocky Mountain Biologicals, Missoula, Montana) and 1% penicillin-streptomycin (Corning Inc., Corning, New York). The HGF and GIE cell lines were cultured in Prigrow 3 and Prigrow 4, respectively (Applied Biological Materials, Inc., British Columbia, Canada). Nutrient media for gingival cell lines were supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin. Cells adherent after 24 h were fed with 200 μ l of fresh, prewarmed media and were covered with sterile cover slips. To avoid effects on cells from the ambient environment, each imaging session was performed over 15 to 20 min of total time out of the incubator.

2.2 Digital Holographic Microscopy Setup, Imaging, and Preprocessing

A detailed description of the telecentric DHM setup and image processing to optically compensate for phase aberrations is described in previously published studies.^{2,3,50} The telecentric DHM setup (Fig. 1) is based on a bitemporal configuration that optically cancels the bulk of the spherical aberrations caused by the microscope objectives (MOs).^{51–53} The lateral resolution was 1.2 μ m with 0.18 \times 0.18 μ m² pixel dimensions of the lateral reconstruction. A 632-nm-wavelength He-Ne laser was used to generate sample and reference beams that recombined at the camera sensor plane as holograms. The holograms were captured by a 1.3-MP CMOS camera (Lumenera Corporation, Inc., Ontario, Canada) and the reconstructed phase map was obtained using the Fresnel reconstruction algorithm.^{2,54}

Principal component analysis (PCA) was employed to cancel the main hologram phase aberrations. The following steps summarize the PCA algorithm: (1) perform singular value decomposition to obtain the first dominant principal component (PC), (2) obtain the linear and quadratic coefficients of the phase vectors from least square fitting of the two dominant singular

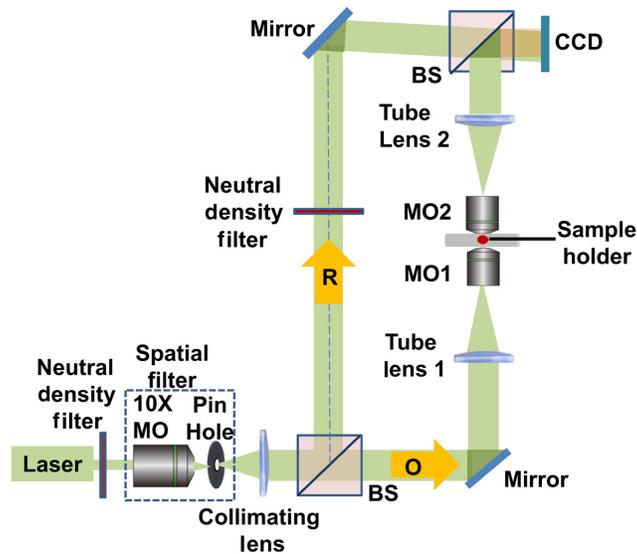


Fig. 1 The bitecentric DHM in transmission configuration, including MO, beamsplitters (BS), object beam (O), reference beam (R), and CMOS camera.

vectors, (3) use these coefficients to compute the phase at the camera sensor, $\varphi(k, l)$, and (4) multiply the conjugate $\varphi^*(k, l)$ with the hologram to obtain $\exp[-j\varphi_{\text{ob}}(k, l)]$, which is the phase due to the biological sample without contributions of MOs and tilt. Phase height was determined from the reconstructed optical pathlength by dividing by an assumed average index of refraction mismatch between cells and surrounding media of $\Delta n = 1.381 - 1.337 = 0.044$.

2.3 Machine Learning and Epithelial–Mesenchymal Score Generation

Machine learning algorithms were evaluated and used to classify gingival cells and for transfer learning on cancer cells to define an EM score (Fig. 2). Cells were segmented and 17 phase parameters were extracted from each of the four cell lines using a custom-written code in MATLAB (version R2015a), which was described previously.³⁴ Parameters are described in Table S1 in the [Supplementary material](#). In total, there were 1295 cells from four different cell lines, which were segmented throughout this study, including 332 cells of GIE, 309 cells of HGF, 307 cells of MCF-7, and 347 cells of MDA-MB-231.

Data were randomly partitioned at a ratio of 4:1 for training and testing. Training was performed on parameters from 252 and 229 GIE and HGF cells, respectively, following a fivefold

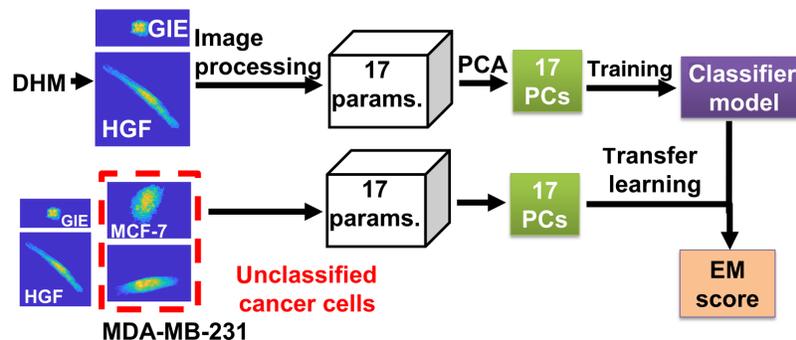


Fig. 2 Machine learning was performed on 1 to 17 features derived from phase maps reconstructed from adherent cells' holograms. Derived features from epithelial (GIE) and mesenchymal (HGF) cell types were used for training. For transfer learning, six PCs representing most of the variation in cell phase maps from two untrained cancer cell lines, MCF-7 and MDA-MB-231, were used for testing and to generate machine learning prediction scores as candidates for an EM score.

cross validation. First, the 17 phase parameter predictors were transformed into PCs using PCA, and the PCA-transformed data were used as inputs for training following fivefold cross validation. Training and cross validation using linear SVM were performed five times each, selecting 1 to 17 PCs as predictors. To evaluate the highest prediction accuracy during training, a one-factor repeated measures analysis of variance (ANOVA) was performed for accuracy on training performance resulting in 1 to 17 PCs used, with a Dunnett's *post-hoc* test to compare results to those of one PC. Then several single and ensemble methods were trained using the same number of PCs found to produce the highest accuracy from the linear SVM algorithm. Default settings in MATLAB were used for each classifier, including a cost parameter of 1 for misclassification. Accuracy was evaluated by comparing output labels to true cell line labels. These were compared to each other using a two-tailed Student's *t*-test. Results were reported as the mean \pm standard deviation. Plots of the first two PCs and receiver operating characteristic (ROC) curves for the best single and best ensemble method classifiers were constructed.

These most accurate single and ensemble algorithms in training were exported as two models in the MATLAB workspace using the classification Learner application. Each model was applied to the PCA-transformed data of 307 cells from MCF-7 and 347 cells from MDA-MB-231 combined with the 80 cells of GIE and 80 cells of the HGF cell lines used for testing. In addition to the classification accuracy, SVM scores, Boosted Trees (AdaBoost) scores, Bagged Trees scores, and SVM posterior probabilities, defined below, were also calculated. Cells from the two cancer cell lines MDA-MB-231 and MCF-7 were assigned as either mesenchymal or epithelial based on the binary classifier. All scores and posterior probabilities were plotted in histograms to evaluate the performance as an EM score. In addition, SVM scores and posterior probabilities were correlated to determine the relative sensitivity of the score and probability throughout their respective ranges. The SVM score s_j , the distance of the observation j to the decision boundary, was calculated as⁵⁵

$$s_j = \left(\frac{x_j}{s_k} \right)' \mathbf{\beta} + b, \quad (1)$$

where x_j is the predictor data of observation j , $s_k = 2.5196$ is the linear kernel scale, $\mathbf{\beta}$ is the vector of fitted linear coefficients, and b is the intercept of the hyperplane defining the separation. The posterior probability $P(s_j)$ was calculated as⁵⁶

$$P(s_j) = 1/(1 + \exp(As_j + B)), \quad (2)$$

where A and B are the fitted slope and intercept, respectively, of the sigmoid function. Meanwhile, the prediction score for AdaBoost, ranging from $-\infty$ to $+\infty$, was defined as⁵⁷

$$f(x) = \sum_{t=1}^T [a_t h_t(x)], \quad (3)$$

where $a_t = 0.5 \log[(1 - \varepsilon_t)/\varepsilon_t]$ are the weights of the sequential learners' hypotheses, ε_t is the weighted classification error of learner t , and $h_t(x)$ is the prediction of learner t for prediction data x for T total learners. The prediction scores for Boosted Trees are estimated posterior probabilities:⁵⁸

$$\hat{P}_{\text{bag}}(c|x) = \sum_{t=1}^T [a_t \hat{P}_t(c|x) I(t \in S)] / \sum_{t=1}^T [a_t I(t \in S)], \quad (4)$$

where $\hat{P}_t(c|x)$ is the estimated posterior probability of learner t for class c with given predictor data x , and $I(t \in S)$ is 1 when learner t is of the indices S from trees used in the prediction, otherwise it is 0.

3 Results

3.1 Cell Morphologies from Optical Phase Maps Vary Across and Within Cell Lines

Cell shapes from GIE [Fig. 3(a)] and HGF [Fig. 3(d)] cell lines resembled epithelial and mesenchymal morphologies, respectively. While GIE cells were more rounded and aggregated in clusters, HGF cells were more elongated with lower phase signals in pixels within the cell body. The cancer cell lines had morphologies in between GIE and HGF cells, with a more punctate phase texture [Figs. 3(b) and 3(c)]. Cells from the MCF-7 cell line form epithelial-like clusters with sharp cell and cluster boundaries [Fig. 3(b)]. Cells from the MDA-MB-231 cell line [Fig. 3(c)] appeared both rounded and elongated and were typically isolated.

3.2 Classification of Epithelial and Mesenchymal Cell Lines Is Highly Accurate

Binary classification was evaluated for multiple algorithms available in the MATLAB machine learning and statistics toolbox, using the training set of 481 cell phase maps ($n = 252$ from the GIE cell line; $n = 229$ from the HGF cell line), all PCs as predictors, with accuracies ranging from 82% to 96% and the highest for linear SVM. Tuning the hyperparameters of box constraint level and kernel scale did not improve training accuracy. Therefore, the number of PCs used as predictors to linear SVM was varied from 1 to 17 (Table 1). Linear SVM with 6, 8, and 17 PCs all produced higher training accuracies than 1 PC (ANOVA, $F=47.6$, $p < 0.001$, Dunnett's test versus 1 PC, $p < 0.001$). Six PCs were selected for use based on this statistical test and on previous models classifying cells based on phase features, which selected six PCs as the smallest

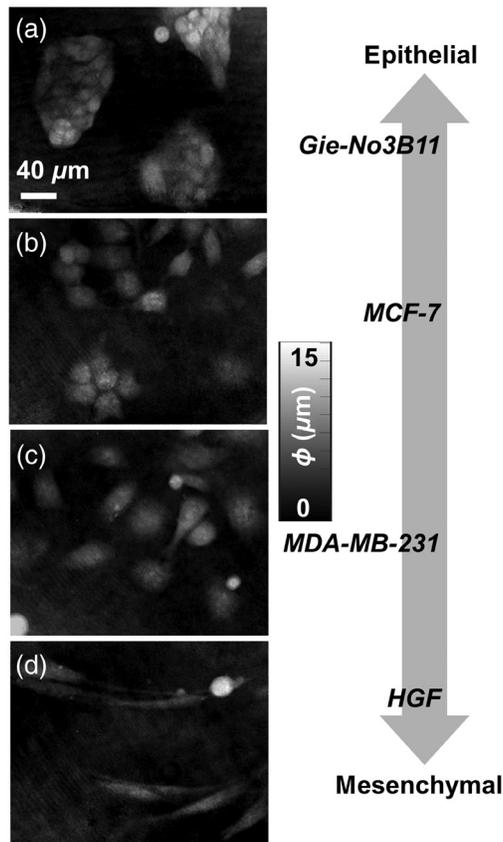


Fig. 3 Representative DHM phase maps from living, adherent cells of (a) Gie-No3B11 (GIE), (b) MCF-7, (c) MDA-MB-231, and (d) HGF cell lines. The cells are ordered qualitatively on an EM axis based on cell morphology apparent in phase maps. Scale and phase bars are indicated.

Table 1 Training accuracy of various machine learning algorithms to classify epithelial and mesenchymal cells.

Type	Method	Accuracy (% , $\mu \pm SD$)
Single	Linear SVM, 1 PC	93.0 \pm 0.2
	Linear SVM, 2 PCs	93.0 \pm 0.2
	Linear SVM, 4 PCs	92.8 \pm 0.3
	Linear SVM, 5 PCs	93.4 \pm 0.6
	Linear SVM, 6 PCs [*]	95.5 \pm 0.3
	Linear SVM, 8 PCs [*]	95.1 \pm 0.3
	Linear SVM, 17 PCs [*]	95.2 \pm 0.6
	Other SVMs, 6 PCs	95.4 \pm 3.9
	Decision trees, 6 PCs	91.6 \pm 0.2
	Nearest neighbor, 6 PCs	88.9 \pm 3.7
Ensemble	Boosted Trees, 6 PCs, AdaBoost ^{**}	94.3 \pm 0.5
	Bagged Trees, 6 PCs, Bag	95.2 \pm 0.4

^{*} $p < 0.001$, ANOVA and *post-hoc* Dunnett's test versus linear SVM, 1 PC.

^{**} $p < 0.001$, Student's *t*-test versus linear SVM, 6 PCs.

number producing no increase in area under the curve (AUC) of ROC curves. Linear SVM training resulted in an accuracy of 95.5% \pm 0.3%. Training using SVMs with different kernel functions (quadratic, cubic, Gaussian), decision trees, or $k = 1$ nearest-neighbor methods did not improve accuracy. The best ensemble method classifier was Bagged Trees (Bag ensemble method, 200 learners, and learning rate of 0.1), which did not improve the accuracy more than the best single method (*t*-test, $p = 0.25$). Boosted Trees (AdaBoost algorithm, 200 learners, 0.1 learning rate) produced lower accuracy than linear SVM, each trained on six PCs (*t*-test, $p < 0.01$). Figures 4(a)–4(d) provide scatterplots of PCs 1 versus 2 and ROC curves for the best performing single and ensemble methods from model training and validation.

Testing based on the linear SVM and Bagged Trees models on a naive dataset of $n = 80$ cells each from GIE and HGF cell lines produced error rates of 2.5% to 3.7% and 0%, respectively [Figs. 4(e), 4(f)]. Transfer learning using the linear SVM model classified 286/307 (87.0%) of MCF-7 cells as epithelial (GIE class) and 326/347 (93.9%) of MDA-MD-231 cells as mesenchymal (HGF class, data not shown). Transfer learning using the Bagged Trees model classified 262/307 (78.1%) of MCF-7 cells as epithelial (GIE class) and 329/347 (94.8%) of MDA-MD-231 cells as mesenchymal (HGF class, data not shown). Linear SVM, Bagged Trees, and Boosted Trees algorithms were used to calculate prediction scores for each cell of the test and transfer datasets (Fig. 5).

3.3 Binary Epithelial–Mesenchymal Classifier Prediction Scores Separate Cancer Cells by Morphology

The distributions of prediction scores from linear SVM as Euclidean distance from the classifying hyperplane [Fig. 5(a)], posterior probabilities [Fig. 5(b)], Boosted Trees [Fig. 5(c)], and Bagged Trees [Fig. 5(d)] were evaluated. Histograms of linear SVM prediction scores [Fig. 5(a), Eq. (1)] produced the most normal-appearing distributions for test data of GIE and HGF and transfer datasets of MCF-7 and MDA-MB-231 cells. Posterior probabilities from SVM [Fig. 5(b), Eq. (2)] and estimated posterior probabilities from Boosted Trees [Figs. 5(c), Eq. (3)] demonstrated excellent separation of classes, but weighted toward 0 and 1. The Boosted Trees predictions produced bimodal distributions of MCF-7 and MDA-MB-231 cell scores [Fig. 5(d), Eq. (4)]. Four scores

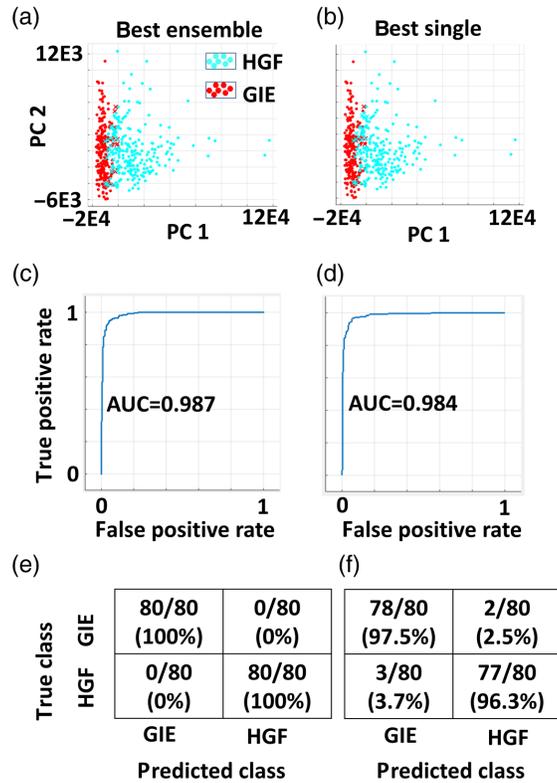


Fig. 4 Binary classifier training data (a) and (b) scatterplots of PC 1 versus 2, highlighting correctly classified GIE (red circle) and HGF (cyan circle) cells, and misclassified cells (red and cyan x's, with color representing the true class). (c) and (d) ROC curves from training data, with AUC listed. (e) and (f) Error tables from a test dataset for (e) Bagged Trees and (f) linear SVM.

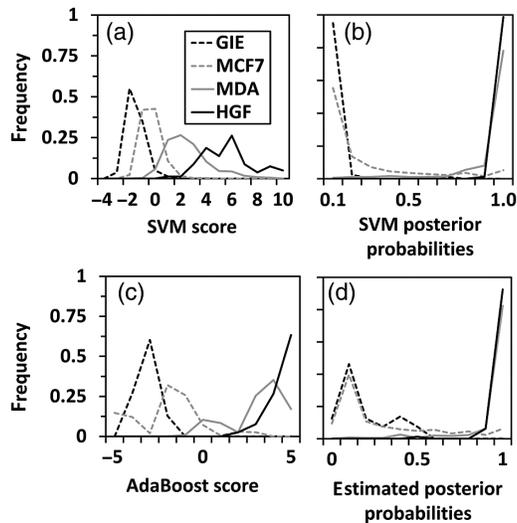


Fig. 5 Prediction scores for transfer learning of an EM classifier to cancer cells. Histograms of prediction score distributions for (a) SVM predictions, (b) SVM posterior probabilities, (c) AdaBoost predictions, and (d) estimated posterior probabilities from Boosted Trees for test datasets of $n = 80$ GIE cells (black dashed line) and $n = 80$ HGF cells (black solid line), and transfer learning datasets of $n = 307$ MCF-7 cells (gray dashed line) and $n = 347$ MDA-MB-231 cells (gray solid line).

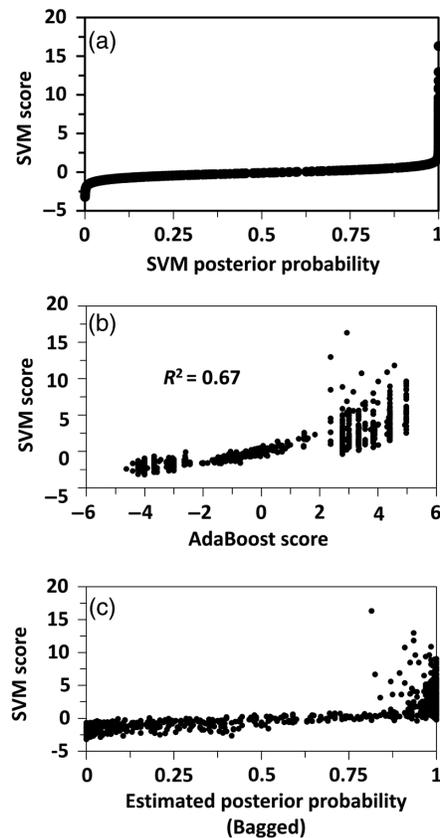


Fig. 6 Correlation plots of linear SVM prediction scores versus (a) posterior probability scores from SVM, (b) AdaBoost prediction scores, and (c) estimated posterior probability scores from Bagged Trees for the dataset of 814 cells defined previously from GIE, HGF, MCF-7, and MDA-MB-231 cell lines.

from HGF cells were outliers and were not included in the histograms. The outliers were extremely high SVM scores more than 5.6 standard deviations away from the population mean score. Correlations between linear SVM prediction scores versus posterior probabilities [Fig. 6(a)] and versus estimated posterior probabilities from Bagged Trees [Fig. 6(b)] were highly nonlinear for low and high scores of each. The correlation between linear SVM versus Boosted Trees prediction scores [Fig. 6(c)] was linear for central scores, but nonlinear overall, with discrete levels of Boosted Trees scores favored at low and high ends of the score range.

Cell phase maps representing the linear SVM (Fig. 7) and Boosted Trees (Fig. 8) prediction scores closest to minima, maxima, medians, and first and third quartiles demonstrated a graded appearance between epithelial and mesenchymal phenotypes, as represented by the cells nearest the median score from GIE and HGF cells, respectively (also depicted in Figs. 7 and 8). The selected cells are for the most part different (except for the MCF-7 Max and MDA-MB-231 Min, which were the same from the two scores), but reflect a trend of more mesenchymal morphology with higher score. Other features, including phase height (nm), area (μm^2), and eccentricity, were also included in each representative map demonstrating each cell's shape features. Cell phase height tended to decrease while area and eccentricity tended to increase when cells were more mesenchymal-like. Nevertheless, each geometrical feature itself did not completely correlate in rank order with the EM score derived from SVM [Eq. (1)] or from AdaBoost [Eq. (3)], for both MCF-7 [Figs. 7(b), 8(b)] and MDA-MD-231 [Figs. 7(c), 8(c)] cell lines.

4 Discussion

Machine learning algorithms applied to QPI of adherent cells in culture classify cell lines in a way useful for determining the functional phenotype on an EM axis. This study proposes a

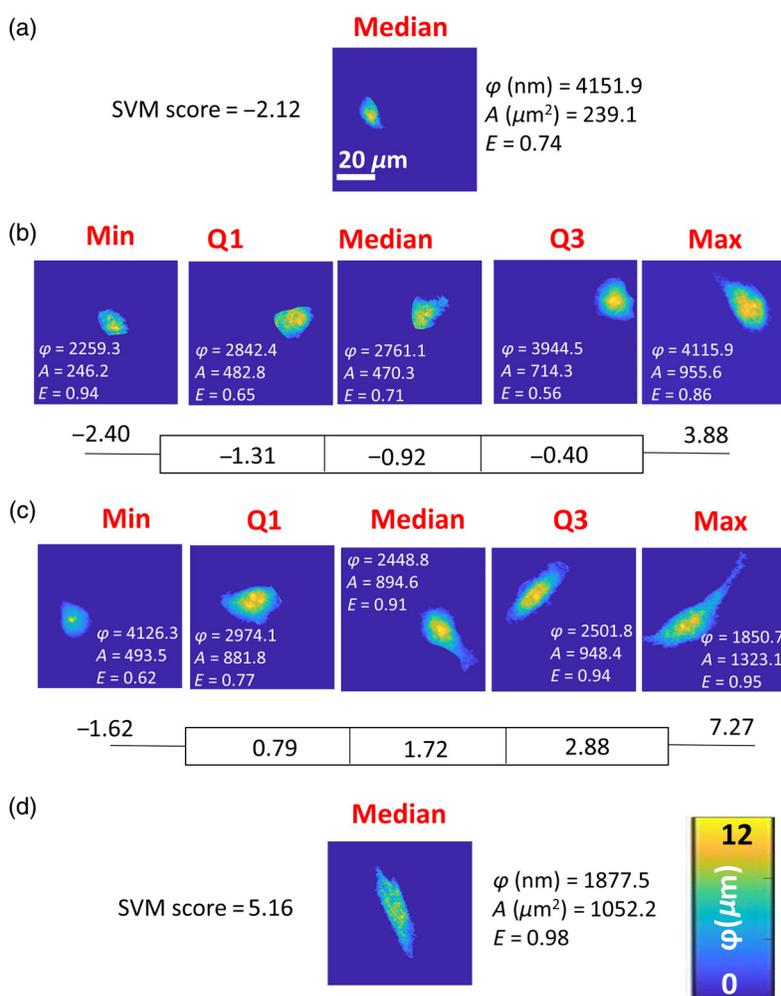


Fig. 7 Phase maps of epithelial, mesenchymal, and breast cancer cells representing the median SVM score of normal cell line (a) GIE and (d) HGF. The minimum, first quartile, median, second quartile, and maximum SVM score for cancer cell lines of (b) MCF-7 and (c) MDA-MB-231 are shown. SVM scores were derived from a binary classification SVM model trained on GIE and HGF cells, then tested on breast cancer cells to generate weighted classification scores. Phase height (φ) in nm, area (A) in μm^2 , and eccentricity (E) of each representative cell generated from phase maps are also listed in the figure near each cell.

transfer learning approach to define a graded phenotypic classification for breast cancer cells: train a binary classifier on known epithelial and mesenchymal cells, then test on the cancer cells of unknown phenotype, defining prediction scores for each unknown cell. The algorithms producing score distributions of cancer cells most evenly distributed between epithelial and mesenchymal extremes were linear SVM and Boosted Trees (AdaBoost) scores. The SVM score, the Euclidean distance to the linear hyperplane separating epithelial and mesenchymal classes, produced normal-appearing distributions within the cancer cell lines, easily interpretable as lying along an EM continuum. The Boosted Trees score also produced a prediction score capable of being interpreted as an EM continuum, but with bimodal score distributions for the cancer cell transfer learning predictions. These prediction scores from binary classifiers serve as robust, quantitative EM scores to define unknown cells with morphologies blended in between two morphological extremes.

This proof-of-concept study has several strengths in design and analyses, but also weaknesses related to the necessarily limited dataset. Strengths include the large numbers of cells imaged (>300 per cell line) and use of MCF-7 and MDA-MB-231 cell lines, well characterized as being more epithelial and mesenchymal in nature, respectively. QPI using DHM provides high

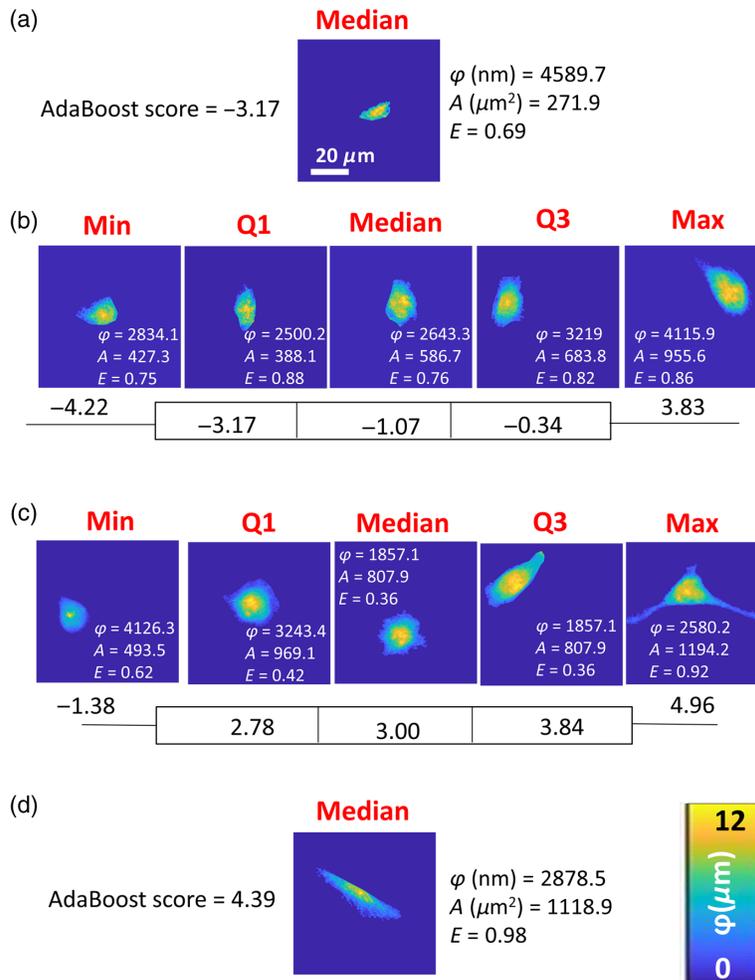


Fig. 8 Phase maps of epithelial, mesenchymal, and breast cancer cells representing the median AdaBoost score of normal cell line (a) GIE and (d) HGF. The minimum, first quartile, median, second quartile, and maximum AdaBoost score for cancer cell lines of (b) MCF-7 and (c) MDA-MB-231 are shown. AdaBoost scores were derived from a binary classification AdaBoost model trained on GIE and HGF cells, then tested on breast cancer cells to generate weighted classification scores. Phase height (φ) in nm, area (A) in μm^2 , and eccentricity (E) of each representative cell generated from phase maps are also listed in the figure near each cell.

accuracy of cell morphological measurements as well as pixel-level textural details.^{3,4} One potential concern is the consistency of the scoring results if different noncancerous epithelial and mesenchymal cells were used instead of the gingival cells available for this study. The smooth histogram in Fig. 5(a) and graded appearance of cells from low to high scores in Figs. 7 and 8 increase confidence in the broad applicability of the developed scores. Still, a classifier trained on syngeneic noncancerous cells derived from the same tissue as the cancer would likely be more patient-specific. A classifier trained on data from many normal epithelial and mesenchymal cell lines would arguably be more generalizable across many patients. The bimodal nature of AdaBoost prediction scores for breast cancer cells [Fig. 5(c)] and granularity of AdaBoost scores at low and high ends of the range [Fig. 6(b)] are a weakness of the AdaBoost predictions compared to linear SVM prediction scores. This is explained by the iterative AdaBoost training algorithm that trains more learners on data that is harder to classify, i.e., it is misclassified by initial learners in the ensemble method. This results in a finer resolution of scores in the middle of the score range, which are the harder to predict cases. The proposed EM score requires validation with additional cancer cells of different epithelial and mesenchymal morphologies from various breast cancer subtypes.⁵⁹ Despite these limitations of the current

training and test datasets, the transfer learning method proposed here quantitatively sorts individual cells along a putative morphological axis that produces well-ranked cells by visual inspection, a feat not reproducible using any single geometrical feature to rank the cells, including thickness, area, and eccentricity (Figs. 7 and 8).

Morphological evaluation using an EM score as proposed here could address a major issue in histology-based diagnostics derived from cell-to-cell heterogeneity. Such heterogeneity, especially in the absence of specific molecular biomarkers, makes risk stratification, diagnosis, and selection of treatment regimens less accurate.⁶⁰ Rapid morphological classification of biopsied cancer cell populations in slide sections using DHM could provide a first-order evaluation of cancer heterogeneity, especially if morphology were linked to clinical behavior. Interestingly, intravital microscopy suggests that cancer cells undergo EM transition to enable metastasis but can rapidly revert to epithelial phenotypes once in the metastatic site, a process termed EM plasticity.⁶¹ An EM score applied to individual cells might, therefore, be useful as an indicator of recent, concurrent, or impending metastasis within the sampled cancer cell population. Such a score could also track responses of the two cancer subpopulations to combined epithelial- and mesenchymal-directed therapies, such as inhibition of Wingless/Integrated (*Wnt*) and Yes-associated protein (*YAP*) signaling.⁶²

Quantitative sorting of adherent cells based on morphology is of potential utility in phenotypic screening and basic studies linking gene expression to phenotype and functional behavior. In phenotypic screening, DHM followed by assignment of a machine learning prediction score to individual cells would allow the detection of subtle morphological shifts in response to various treatments,⁶³ a task of increasing importance in drug discovery.⁶⁴ In this study, two breast cancer cell lines, MCF-7 and MDA-MB-231, were scored on an EM axis by linear SVM, consistent with their morphological appearance. These cell lines are well known for appearing with clustered epithelial-type and single, mesenchymal-type morphologies, respectively. Mesenchymal gene expression in MDA-MB-231 cells, including N-Cadherin, Snail, Slug, ZEB1 and 2, and Yes-associated protein 1 (*YAP1*), were downregulated after lentiviral insertion of E-cadherin, a marker of epithelial cells, which shifted the morphology of MDA-MB-231 cells to a more rounded, clustered epithelial type.⁶² Similarly, MCF-7 cells made to express Snail, a transcription factor typical of mesenchymal cells, become less round and experience an upregulation of mesenchymal-related genes and downregulation of epithelial-related genes.⁶⁵ Such basic studies have potential impact in defining the roles of epithelial and mesenchymal phenotypes in cancer behavior, leading to a better understanding of phenotypic transitions and plasticity in cancer. An EM score would aid such efforts by establishing the magnitude of phenotypic shifts with a given treatment.

An EM score has utility in interpreting qualitative morphological assignments. For example, breast cancer cells in three-dimensional views have been classified as “stellate,” “grape-like,” “mass,” or “round,” and unique gene expression profiles are linked to these classes.⁴⁰ A single, unified EM score applied to each of these classes may correlate with the expression of multiple key genes, linking morphology to gene expression profiles on a quantitative basis. One exciting future development of such an approach would be to determine the sensitivity of the EM score to differential expression of individual genes, something best achieved by direct comparison of parental and genetic knockout cell lines. There is some evidence that qualitative morphological classes do not correspond to invasiveness in all cases.⁴³ This finding is consistent with at least a subset of genes being responsible for invasiveness but not aggregate morphology, a hypothesis which is testable through sequential genetic knockout, EM scoring, and assessment of invasiveness *in vitro*. QPI of cancer cells in functional assays combined with classification scores such as the proposed EM score could aid such studies. In single-cell studies, phase images of cells of interest could guide laser-capture microdissection to link observed behavior, morphology, and gene expression at a single cell level. Indeed, advanced machine learning techniques, including deep learning,²⁹ have recently been applied to isolate cell subpopulations based on unique phase features⁶ and other phenotypic differences,^{66,37} including metastatic versus primary cancer⁶⁷ and different types of nonactivated lymphocytes.⁶⁸ The phase/morphology score concept described here could be applied to support decision-making in intelligent cell sorting systems, such as flow cytometry with QPI,^{69,34} to partition cells from a heterogeneous population into distinct morphological groups.⁷⁰

The proposed technique to generate EM scores offers greater robustness, adaptability, and flexibility than qualitative or single-parameter morphological characterization, but requires some interpretation. First, the SVM derived score is <0 for all cells classified as epithelial and >0 for all cells classified as mesenchymal, whatever their true origin. Robustness of the score derives from drawing upon multiple (six) PCs for classification and is demonstrated in Figs. 7 and 8 by the biophysical/geometrical parameters of phase height, area, and eccentricity, which are not in rank order for rank-ordered EM scores within the cancer cell lines. Whereas any single parameter suffers from heterogeneity from cell to cell or may lose sensitivity to some cells or cell responses,³⁹ multiple training features for machine learning classification regularly achieve higher performance than single-feature classification.¹⁸ Score adaptability stems from flexibility in defining the training dataset—different cell lines or primary cells could be used—as well as the cells to be scored by transfer learning. For example, the algorithm applied here to breast cancer cells could equally be applied to any cell type with a mixture of epithelial and mesenchymal qualities, such as cells undergoing epithelial-to-mesenchymal or mesenchymal-to-epithelial transition.⁷¹

5 Conclusion

This study proposes morphological scoring to sort unknown cells along a recognizable morphological axis using quantitative phase signatures and machine learning. As an example related to cancer cell phenotypes, phase features from well-characterized epithelial and mesenchymal cell lines were trained using SVM, producing a linear EM score applicable to cancer cells. This proposed morphological score has various future applications in characterizing individual cancer cells of unknown lineage and/or phenotype, and the general approach is applicable in comparing any cells to the morphologies of well-known, well-characterized cell lines.

Disclosures

The authors declared no conflicts of interest.

Acknowledgments

This work was supported by the U.S. National Institute of Biomedical Imaging and Bioengineering (Grant No. 1R03EB28017). The authors would like to thank Dr. Zaver Bhujwala (Johns Hopkins School of Medicine, Baltimore, Maryland) for the gift of the MDA-MB-231 cell line and Drs. Diane Bienek and Gili Kaufman (Volpe Research Center, American Dental Association Foundation) for the gift of HGF and Gie-No3B11 cell lines. In addition, we thank Thuc Phan (Department of Electrical Engineering and Computer Sciences, Catholic University of America) for assistance with data collection using digital holographic microscopy.

References

1. B. Kemper et al., “Label-free quantitative cell division monitoring of endothelial cells by digital holographic microscopy,” *J. Biomed. Opt.* **15**, 036009 (2010).
2. T. Nguyen et al., “Accurate quantitative phase digital holographic microscopy with single- and multiple-wavelength telecentric and nontelecentric configurations,” *Appl. Opt.* **55**, 5666–5683 (2016).
3. V. K. Lam et al., “Quantitative assessment of cancer cell morphology and motility using telecentric digital holographic microscopy and machine learning,” *Cytom. Part A* **93**(3), 334–345 (2018).
4. V. K. Lam et al., “Machine learning with optical phase signatures for phenotypic profiling of cell lines,” *Cytom. Part A* **95**(7), 757–768 (2019).
5. M. Born and E. Wolf, “Interference and diffraction with partially coherent light,” Chapter 13 in *Principles of Optics*, pp. 491–555, Pergamon Press, Elmsford, New York (1980).

6. P. Memmolo et al., "Identification and classification of biological micro-organisms by holographic learning," *Proc. SPIE* **11060**, 110600H (2019).
7. D. Gabor, "A new microscopic principle," *Nature* **161**, 777–778 (1948).
8. D. Gabor, "Microscopy by reconstructed wave fronts: II," *Proc. Phys. Soc. Sect. B* **64**(6), 449–469 (1951).
9. B. Kemper et al., "Quantitative phase imaging-based concepts for the analysis of global morphology changes in confluent cell layers," *Proc. SPIE* **10887**, 108871M (2019).
10. N. Pavillon et al., "Early cell death detection with digital holographic microscopy," *PLoS One* **7**(1), e30912 (2012).
11. R. Cao et al., "Quantitative observations on cytoskeleton changes of osteocytes at different cell parts using digital holographic microscopy," *Biomed. Opt. Express* **9**(1), 72–85 (2018).
12. Y. Li et al., "Digital holographic microscopy for longitudinal volumetric imaging of growth and treatment response in three-dimensional tumor models," *J. Biomed. Opt.* **19**(11), 116001 (2014).
13. D. Bettenworth et al., "Quantitative phase microscopy for evaluation of intestinal inflammation and wound healing utilizing label-free biophysical markers," *Histol. Histopathol.* **33**(5), 417–432 (2018).
14. T. Blasi et al., "Label-free cell cycle analysis for high-throughput imaging flow cytometry," *Nat. Commun.* **7**, 10256 (2016).
15. D. Roitshtain et al., "Quantitative phase microscopy spatial signatures of cancer cells," *Cytom. Part A* **91**(5), 482–493 (2017).
16. Z. Han et al., "Breast cancer multi-classification from histopathological images with structured deep learning model," *Sci. Rep.* **7**, 4172 (2017).
17. D. A. Van Valen et al., "Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments," *PLoS Comput. Biol.* **12**(11), e1005177 (2016).
18. C. L. Chen et al., "Deep learning in label-free cell classification," *Sci. Rep.* **6**, 21471 (2016).
19. Y. Rivenson et al., "PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning," *Light Sci. Appl.* **8**, 23 (2019).
20. R. Li et al., "Deep learning segmentation of optical microscopy images improves 3-D neuron reconstruction," *IEEE Trans. Med. Imaging* **36**(7), 1533–1541 (2017).
21. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Express* **26**(18), 22603–22614 (2018).
22. Y. J. Jo et al., "Quantitative phase imaging and artificial intelligence: a review," *IEEE J. Sel. Top. Quantum Electron.* **25**(1), 1–14 (2018).
23. T. C. Nguyen et al., "Quantitative assessment of cancer cell morphology and movement using telecentric digital holographic microscopy," *Proc. SPIE* **10074**, 100740U (2017).
24. Y. Ozaki et al., "Label-free classification of cells based on supervised machine learning of subcellular structures," *PLoS One* **14**, e0211347 (2019).
25. K. H. Park et al., "Efficient ensemble methods for classification on clear cell renal cell carcinoma clinical dataset," *Lect. Notes Comput. Sci.* **10752**, 235–242 (2018).
26. E. Arvaniti et al., "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Sci. Rep.* **8** (2018).
27. G. Kim et al., "Learning-based screening of hematologic disorders using quantitative phase imaging of individual red blood cells," *Biosens. Bioelectron.* **123**, 69–76 (2019).
28. N. Pavillon et al., "Noninvasive detection of macrophage activation with single-cell resolution through machine learning," *Proc. Natl. Acad. Sci. U. S. A.* **115**, E2676–E2685 (2018).
29. G. Kim et al., "Rapid and label-free identification of individual bacterial pathogens exploiting three-dimensional quantitative phase imaging and deep learning," bioRxiv (2019).
30. M. Mir et al., "Highly sensitive quantitative imaging for monitoring single cancer cell growth kinetics and drug response," *PLoS One* **9**, e89000 (2014).
31. B. Rappaz et al., "Digital holographic microscopy: a quantitative label-free microscopy technique for phenotypic screening," *Comb. Chem. High Throughput Screen.* **17**, 80–88 (2014).
32. Y. K. Park et al., "Measurement of red blood cell mechanics during morphological changes," *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6731–6736 (2010).

33. K. C. M. Lee et al., "Multi-ATOM: ultrahigh-throughput single-cell quantitative phase imaging with subcellular resolution," *J. Biophotonics* **12**, e201800479 (2019).
34. K. C. M. Lee et al., "Quantitative phase imaging flow cytometry for ultra-large-scale single-cell biophysical phenotyping," *Cytom. Part A* **95**, 510–520 (2019).
35. D. K. Singh et al., "Label-free fingerprinting of tumor cells in bulk flow using inline digital holographic microscopy," *Biomed. Opt. Express* **8**(2), 536 (2017).
36. S. Hakim, M. Yamaguchi, and F. Kimura, "Application of digital holography on diagnosis of malignant lymphoma," in *14th Workshop Info. Opt., 2015* (2015).
37. J. Mangal et al., "Unsupervised organization of cervical cells using bright-field and single-shot digital holographic microscopy," *J. Biophotonics* **12**, e201800409 (2019).
38. V. L. Calin et al., "Evaluation of the metastatic potential of malignant cells by image processing of digital holographic microscopy data," *FEBS Open Bio* **7**, 1527–1538 (2017).
39. T. Yao et al., "An optical study of drug resistance detection in endometrial cancer cells by dynamic and quantitative phase imaging," *J. Biophotonics* **12**, e201800443 (2019).
40. P. A. Kenny et al., "The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression," *Mol. Oncol.* **1**, 84–96 (2007).
41. J. P. Thiery, "Epithelial–mesenchymal transitions in tumour progression," *Nat. Rev. Cancer* **2**, 442–454 (2002).
42. J. Zavadil et al., "Epithelial-mesenchymal transition," *Cancer Res.* **68**(23), 9574–9577 (2008).
43. M. J. Ziperstein, A. Guzman, and L. J. Kaufman, "Breast cancer cell line aggregate morphology does not predict invasive capacity," *PLoS One* **10**, e0139523 (2015).
44. K. Paňková et al., "The molecular mechanisms of transition between mesenchymal and amoeboid invasiveness in tumor cells," *Cell. Mol. Life Sci.* **67**, 63–71 (2010).
45. S. Gröger, J. Michel, and J. Meyle, "Establishment and characterization of immortalized human gingival keratinocyte cell lines," *J. Periodontol. Res.* **43**, 604–614 (2008).
46. S. Vardar-Sengul et al., "Expression profile of human gingival fibroblasts induced by interleukin-1 β reveals central role of nuclear factor-kappa B in stabilizing human gingival fibroblasts during inflammation," *J. Periodontol.* **80**, 833–849 (2009).
47. R. P. Illeperuma et al., "Immortalized gingival fibroblasts as a cytotoxicity test model for dental materials," *J. Mater. Sci. Mater. Med.* **23**, 753–762 (2012).
48. H. D. Soule et al., "A human cell line from a pleural effusion derived from a breast carcinoma 1, 2," *J. Natl. Cancer Inst.* **51**, 1409–1416 (1973).
49. R. Cailleau et al., "Breast tumor cell lines from pleural effusions," *J. Natl. Cancer Inst.* **53**, 661–674 (1974).
50. C. Zuo et al., "Phase aberration compensation in digital holographic microscopy based on principal component analysis," *Opt. Lett.* **58**, 389–397 (2013).
51. A. Doblas et al., "Accurate single-shot quantitative phase imaging of biological specimens with telecentric digital holographic microscopy," *J. Biomed. Opt.* **19**, 046022 (2014).
52. E. Sánchez-Ortiga et al., "Digital holographic microscopy with pure-optical spherical phase compensation," *J. Opt. Soc. Am. A* **28**, 1410–1417 (2011).
53. E. Sánchez-Ortiga et al., "Aberration compensation for objective phase curvature in phase holographic microscopy: comment," *Opt. Lett.* **39**, 417 (2014).
54. E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Appl. Opt.* **38**, 6994–7001 (1999).
55. N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," Cambridge University Press, Cambridge (2000).
56. J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Adv. Large Margin Classifiers*, pp. 61–74, MIT Press, Cambridge, Massachusetts (2000).
57. J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Ann. Stat.* **29**(5), 1189–1232 (2001).
58. L. Breiman, "Bagging predictors," *Mach. Learn.* **24**(2), 123–140 (1996).
59. X. Dai et al., "Breast cancer cell line classification and its relevance with breast tumor subtyping," *J. Cancer* **8**, 3131–3141 (2017).

60. G. R. Sant, K. B. Knopf, and D. M. Albala, "Live-single-cell phenotypic cancer biomarkers-future role in precision oncology?," *NPJ Precis. Oncol.* **1**, 21 (2017).
61. E. Beerling et al., "Plasticity between epithelial and mesenchymal states unlinks EMT from metastasis-enhancing stem cell capacity," *Cell Rep.* **14**, 2281–2288 (2016).
62. A. Sulaiman et al., "Dual inhibition of Wnt and Yes-associated protein signaling retards the growth of triple-negative breast cancer in both mesenchymal and epithelial states," *Mol. Oncol.* **12**, 423–440 (2018).
63. A. V. Belashov et al., "Quantitative assessment of changes in cellular morphology at photodynamic treatment *in vitro* by means of digital holographic microscopy," *Biomed. Opt. Express* **10**(10), 4975 (2019).
64. W. Zheng, N. Thorne, and J. C. McKew, "Phenotypic screens as a renewed approach for drug discovery," *Drug Discov. Today* **18**, 1067–1073 (2013).
65. R. Mezencev et al., "Snail-induced epithelial-to-mesenchymal transition of MCF-7 breast cancer cells: systems analysis of molecular changes and their effect on radiation and drug sensitivity," *BMC Cancer* **16**, 236 (2016).
66. C. Brasko et al., "Intelligent image-based *in situ* single-cell isolation," *Nat. Commun.* **9**, 226 (2018).
67. M. Rubin et al., "TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set," *Med. Image Anal.* **57**, 176–185 (2019).
68. J. Yoon et al., "Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning," *Sci. Rep.* **7**, 6654 (2017).
69. J. Min et al., "Quantitative phase imaging of cells in a flow cytometry arrangement utilizing Michelson interferometer-based off-axis digital holographic microscopy," *J. Biophotonics* **12**, e201900085 (2019).
70. N. Nitta et al., "Intelligent image-activated cell sorting," *Cell* **175**, 266–276.e13 (2018).
71. M. A. Nieto, "The ins and outs of the epithelial to mesenchymal transition in health and disease," *Annu. Rev. Cell Dev. Biol.* **27**, 347–376 (2011).

Van K. Lam received her MS degree in biomedical engineering from the Catholic University of America, Washington DC, in 2017. Currently, she is pursuing her PhD in the optics and tissue engineering. Her research interest focuses on optical imaging techniques and machine learning applications to classify cancer cells and assess cancer invasiveness and the extracellular matrices. She is a member of OSA, SPIE, and BMES.

Thanh Nguyen is currently a postdoctoral researcher in the EECS department at the Catholic University of America and a science collaborator at NASA Goddard Space Flight Center. He received his MS degree in electrical engineering in 2013 and completed his PhD in 2018. His research interests include two-dimensional/three-dimensional imaging and deep learning. He is a reviewer for *Light: Science & Applications*, *Applied Optics*, *Optics Express*, *Optics Letters*, and *IEEE*. He is a member of OSA and SPIE.

Vy Bui received her BS and MS degrees in EECS from the Catholic University of America, Washington DC, in 2012 and 2014, and is currently pursuing her PhD in electrical engineering. She has been working on medical image computing research at the Cardiovascular CT Laboratory at the National Heart, Lung, and Blood Institute since 2016. Her research interests include image processing, computer vision, and deep learning for medical and biomedical image analysis.

Byung Min Chung is currently an assistant professor in the biology department at the Catholic University of America. His laboratory is studying molecular mechanisms underlying breast cancer pathogenesis to identify innovative and effective therapeutic targets and strategies. The study uses human cancer cells to characterize the role of a cytoskeletal protein, keratin 19 (K19), in cell cycle progression. To address this, multidisciplinary approaches, well-established, and cutting-edge techniques in molecular and cell biology, biochemistry, and biophysics are involved.

Lin-Ching Chang received her DSc degree in computer science from George Washington University in 1998. She is an associate professor in the Department of Electrical Engineering

and Computer Science and the director of the Data Analytics Program at the Catholic University of America. Prior to that, she was an IRTA postdoctoral fellow at NIH working on computational neuroscience projects. Her main research interests include machine learning, data analytics, computer vision, and medical informatics.

George Nehmetallah is currently an associate professor in the EECS department at the Catholic University of America. He has served as a PI and a co-PI for several projects funded by NSF, NASA, Air Force, Army, and DARPA, and was one of the 11 recipients awarded the Army SBIR Achievement in 2011. His research interests are in 3-D imaging, digital holography, spectroscopy, and metamaterials. He is a senior member of OSA and SPIE.

Christopher B. Raub is currently an assistant professor in biomedical engineering at the Catholic University of America, with research interests in the use of endogenous optical signals to understand interactions between cells and the tissue microenvironment. Currently, he is a PI on a project funded by NIBIB to assess microenvironmental determinants of cancer cell invasiveness using quantitative phase and polarized light microscopy. He is a member of BMES and the Orthopedic Research Society.