

Dear Editor,

Please find enclosed manuscript entitled “High-throughput label-free cell classification using machine learning,” which we wish to be considered for publication in *Scientific Reports*. This paper is directly transferred from *Nature Communications* and suggested for immediate peer review at *Scientific Reports* by Dr. David Gevaux, senior editor of *Nature Communications*.

Conventional flow cytometry is a powerful tool for large-scale cell analysis due to its ability to measure anisotropic elastic light scattering of millions of individual cells as well as emission of fluorescent labels conjugated to cells. However, label-free cell analysis is recently in high demand for its use in personalized genomics, cancer diagnostics, and drug development as it avoids adverse effects of staining reagents on cellular viability or cell signaling. Currently available label-free cell assays mostly rely only on a single feature, namely on cell size or dielectric constant. They also have low throughput limiting the sample size that can be analyzed.

Here we report that by combining feature extraction and machine learning to high throughput quantitative phase imaging enabled by the photonic time stretch, label-free cells can be classified with record accuracy. Our system captures quantitative phase images in a flow-through microscope and extracts multiple biophysical features such as morphological parameters, optical loss characteristics, and protein concentration indicators of individual cells. These biophysical measurements form a hyperdimensional feature space in which supervised learning is performed for cell classification. As a validation of the enhanced classification sensitivity and specificity of our method, we show binary classification of *OT-II* white blood T-cells against *SW-480* colon cancer cells. Additionally, we show classification of lipid accumulating algal strains for biofuel production.

Our system opens up a new path to data-driven genotype-trained phenotypic diagnosis and better understanding of the heterogeneous gene expressions in cells and thus is expected to be a valuable tool for high-throughput label-free cell analysis in medical, biotechnological, and research applications. Hence we believe that our paper merits the attention of the broad audience of *Scientific Reports*.

Finally, we suggest Ming Li (Chinese Academy of Sciences, ml@semi.ac.cn), Ozdal Boyraz (University of California, Irvine, oboyraz@uci.edu), Fangxin Li (University of Toronto, fangxinlee.li@utoronto.ca), Chao Wang (University of Kent, c.wang@kent.ac.uk), and Jason Chou (Lawrence Livermore National Laboratory, chou8@llnl.gov) as the reviewers of the manuscript for their extensive knowledge on the subject matter. We would really appreciate it if Andrew Weiner (Purdue University, amw@ecn.purdue.edu), Alexander Gaeta (Cornell University, alg3@cornell.edu), Kevin Tsia (University of Hong Kong, tsia@hku.hk), Rick Trebino (Georgia Tech, rick.trebino@physics.gatech.edu), and Kenneth Wong (University of Hong Kong, kywong@eee.hku.hk) could be excluded from the review process because of the conflict of interests.

Thank you for your time and consideration.

Sincerely,

Claire Lifan Chen, Ata Mahjoubfar, and Bahram Jalali
Departments of Electrical Engineering
California NanoSystems Institute
University of California, Los Angeles (UCLA)